SCIENCE B-16

HISTORY OF THE EARTH AND OF LIFE

THE GEOLOGICAL TIME SCALE

Eon	Era	Period (with Cenozoic epochs)	Age (Ma)
Phanerozoic	Cenozoic	Quaternary	
		Holocene	
			0.01
		Pleistocene	
			1.7
		Tertiary	
		Pliocene	
			5
		Miocene	
			24.5
		Oligocene	
			38
		Eocene	
			55
		Paleocene	
			65
	Mesozoic	Cretaceous	
			144
		Jurassic	
			213
		Triassic	
			248
	Paleozoic	Permian	
			286
		Pennsylvanian	
		] Carboniferous	320
		Mississippian	
			360
		Devonian	
			408
		Silurian	
			438
		Ordovician	
			505
		Cambrian	
			590

Proterozoic (590 - 2500 Ma)

Archean (> 2500 Ma)

Professor Stephen J. Gould
Fall Term, 1986

SCIENCE B-16

TABLE OF CONTENTS

ow giddy by looking so far into the.

told the Royal Society that it was
orm some estimate with regard to
e of this Earth has existed." But
d Siccar Point what estimate could
orld which we inhabit is composed
ot of the earth which was the im-
sor of the present but of the earth
preceded the land that was above
sea while our present land was yet
of the ocean," he wrote. "Here are
ccessive periods of existence, and
n our measurement of time, a thing
tion. . . . The result, therefore, of
iry is, that we find no vestige of a
spect of an end."

Excerpt reprinted by permission
of Farrar, Straus and Giroux, Inc
from BASIN AND RANGE by John
McPhee. Copyright (c) 1980,1981
by John McPhee. This material
first appeared in The New Yorker

The Old Red Sandstone was put down by rivers flowing southward to a sea where marine strata were accumulating in the region that is now called Devon. The size, speed, and direction of the rivers—their islands, pitches, and bends—are not just inferable but can almost be seen, in structures in the Old Red Sandstone: gravel bars, point bars, ripples of the riverbeds, migrating channels, "waves" that formed of sand. The sea into which those rivers spilled ran all the way to Russia, but it was in the rock of Devonshire that geologists in the eighteen-thirties found cup corals—fossilized skeletons, cornucopian in shape—that were not of an age with corals they had found before. They had found related corals that were obviously less developed than these, and they had found corals that were more so.

The less developed corals had been in rock that lay under the Old Red Sandstone. The more developed corals had been in rock above the Old Red Sandstone. Therefore, it was inferred (correctly) that the Old Red Sandstone of North Britain and the marine limestone of Devon were of the same age, and that henceforth any rock of that age anywhere in the world—in downtown Iowa City; on Pequop Summit, in Nevada; in Stroudsburg, Pennsylvania; in Sandusky, Ohio—would be called Devonian. It was a name given, although they did not know it then, to fifty million years. They still had no means of measuring the time involved. They also had no way of knowing that those fifty million years had ended a third of a billion years ago. All they had was their new and expanding insight that they were dealing with time in quantities beyond comprehension. Devonian—395 to 345 million years before the present.

Geologists did not have to look long at the coal seams of Europe—the coals of the Ruhr, the coals of the Tyne—to decide that the coals were of an age, which they labelled Carboniferous. The coal and related strata lay on top of the Old Red Sandstone. So, in the succession of time, the Carboniferous period (eventually subdivided into Mississippian and Pennsylvanian in the United States) would follow the Devonian, coupling on, as the science would eventually determine, another sixty-five million years—345 to 280 million years before the present.

In this manner—with their fossil assemblages and faunal successions, their hammers decoding rock—geologists in the first eighty years of the nineteenth century constructed their scale of time. It was based on organic evolution, and, crossing the century, it both anticipated and confirmed Darwin. When the Devonian was defined in the light of the changes in corals, Darwin was obscure and not long off the Beagle, with twenty years to go before *The Origin of Species.* Meanwhile, the geologists were out correlating strata and reading there a record less of rock than of life. The rock had been recycled, and sandstones of one era could be indistinguishable from the sandstones of another, but organic evolution had not occurred in cycles, so it was through the antiquity of fossils that geologists worked out the comparative ages of the rock in which the fossils were preserved. Some creatures were more useful than others. Oysters and horseshoe crabs, for example, were of marginal assistance. Oysters had appeared in the Triassic, horseshoe crabs in the Cambrian. Both had evolved minimally and had obviously avoided extinction. Some creatures, on the other hand, had appeared suddenly, had evolved quickly, had become both abundant and geographically widespread, and then had died out, or died down, abruptly. Geologists canonized them as "index fossils" and studied them in groups. Experience proved that the surest method of working out relative ages of rock was not through individual creatures but through the relating of successive strata to whole collections of creatures whose fossils were contained therein—a painstaking comparison of arrivals and extinctions that helped to characterize the divisions of the time scale and define its boundaries with precision.

Imagine an E. L. Doctorow novel in which Alfred Tennyson, William Tweed, Abner Doubleday, Jim Bridger, and Martha Jane Canary sit down to a dinner cooked by Rutherford B. Hayes. Geologists would call that a fossil assemblage. And, without further assistance from Doctorow, a geologist could quickly decide—as could anyone else—that the dinner must have occurred in the middle eighteen-seventies, because Canary was eighteen when the decade began, Tweed became extinct in 1878, and the biographies of the others do not argue with these limits. In progressive refinements, geologists with their fossil assemblages established their systems and series and stages of rock, their eras and periods and epochs of time. But, unlike Doctorow, who deals with a mere half-dozen people around a dinner table, the geologists would assemble from one set of strata hundreds and even thousands of species from all over

the food chain, and by lining up their genetic histories side by side establish with near-certainty points in comparative time.

Some of these time lines were bolder than others, and none more so than the one that underlined the first appearance of fossils in abundance in the world. It marked a great and sudden explosion of life, all the major phyla having developed more or less at the same time and now acquiring skeletons and shells and teeth and other hard components that allowed them individually to be reported to the future. Because rock that held these early fossils was first studied on Harlech Dome and adjacent Welsh terrains, geologists named the system Cambrian, after the Roman name for Wales. They then named the Silurian for a Welsh tribe that bitterly defied the Romans. After some years and more comparative study, an argument broke out over the Cambro-Silurian line, a scientific battle royal in which the Cambrian forces tried to move their banner forward through time and the Silurian proponents attempted to push theirs back. The disputed block of time became a sort of demilitarized zone. Friendships came unstuck. The standoff lasted for decades, until some genius in scientific diplomacy suggested that the disputed time had enough characteristics of its own to be given the status of a discrete period, an appropriate name for which—in honor of another tribe of intractable Welsh belligerents—would be Ordovician. There was a lot of room for generosity. There was plenty of time for all. Cambrian—570 to 500. Ordovician—500 to 435. Silurian—435 to 395 million years before the present.

A British geologist went to Russia and after a season or two's tapping at the Urals named still another period in time, and system of rock, for the upland oblast of Perm. There were formations in Perm with a fossil story distinctly their own that were superimposed—as they happen to be in Pennsylvania, as they happen to be at the rim of the Grand Canyon—upon the Carboniferous. What was distinct about the character of the Permian assemblages was not only the forms to which they had evolved but also their absence in great numbers from higher, younger strata. There had evidently been a wave of death, in which thousands of species had vanished from the world. No one has explained what happened—at least not to the general satisfaction. A drastic retreat of shallow seas may have destroyed innumerable environments. A change of ocean salinity may have ended a lot of life through osmotic shock. The cause may have been extraterrestrial—lethal radiation from a supernova dying nearby. None of these hypotheses has attracted enough concurrence to be dressed out in full as a theory, but, whatever the cause, no one argues that at least half

the fish and invertebrates and three-quarters of all amphibians—perhaps as much as ninety-six per cent of all marine faunal species—disappeared from the world in what has come to be known as the Permian Extinction.

It was an extinction of a magnitude that would be approached only once in subsequent history, or—to express that more gravely—only once before the present day. The sharp line of creation at the outset of the Cambrian had an antiphonal parallel in the Permian Extinction, and the whole long stretch between the one and the other was set apart in history as the Paleozoic era. It was a unit—well below the surface but far above the bottom—just hanging there suspended in the formless pelagics of time. No vestige of a beginning. No prospect of an end. The Paleozoic—570 to 230 million years before the present, a thirteenth of the history of the earth. Cambrian, Ordovician, Silurian, Devonian, Mississippian, Pennsylvanian, Permian. When I was seventeen, I used to accordion-pleat those words, mnemonically capturing the vanished worlds of "Cosdmpp," the order of the periods, the sequence of the systems. It was either that or write them in the palm of one hand.

Lyell, Cuvier, Conybeare, Phillips, von Alberti, von Humboldt, Desnoyers, d'Halloy, Sedgwick, Murchison, Lapworth, Smith (William "Strata"

Smith): the geologists who extended Hutton's insight and built this time scale conjoined their names in the history of the science in a way that would not be repeated for more than a hundred years, until a roster of comparable length—Hess, Heezen, McKenzie, Morgan, Wilson, Matthews, Vine, Parker, Sykes, Ewing, Le Pichon, Cox, Menard—would effect the plate-tectonics revolution. The system of rock immediately above the Paleozoic, in which all that Permian life failed to reappear, was typified by three formations in Germany—certain sandstones, limestones, and marly shales—that ran like a striped flag through the Black Forest, the Rhine Valley, and lent the name Triassic to thirty-five million years. In the Triassic, the earliest subdivision of the Mesozoic era, two families of reptiles that had survived the Permian Extinction began to show patterns of unprecedented growth. This would continue for a hundred and fifty million years—through the Jurassic and out to the end of Cretaceous time, when the "fearfully great lizards," on the point of disappearance, would reach their greatest size, not to be surpassed until epochs that followed the Eocene development of whales. European geologists studying the massive limestones of the Jura—the gentle mountains of the western cantons of Switzerland and of Franche-Comté—related the copious displays of ancient life there to comparable assemblages elsewhere in the world, and called them all Jurassic. The first bird appeared in the Jurassic. It had claws on its wings and teeth in its bill and a reptile's long tail sprouting feathers. Its complete performance envelope as a flier was to climb a tree and jump.

Physicists, chemists, and mathematicians, taking note of all the nomenclatural inconsistencies—of time named for mountain ranges, time named for savage tribes, time named for a country here, a county there, an oblast in the Urals—have politely, gently, suggested that, in this one sense only, the time scale seems archaic, seems, if one may say so, out of date. Geology might be better served by a straightforward system of numbers. The reaction of geologists, by and large, has been to look upon this suggestion as if it had come over a bridge that exists between two cultures. A Continental geologist, in 1822, named seventy-six million years for the white cliffs of Dover, for the downs of Kent and Sussex, for the chalky ground of Cognac and Champagne. Related strata were spread out through Holland, Sweden, Denmark, Germany, and Poland. He called it Le Terrain Crétacé. If that name was apt, his own was irresistible. He was J. J. d'Omalius d'Halloy. Triassic, Jurassic, Cretaceous. When the Cretaceous ended, the big marine reptiles had disappeared, the flying reptiles, the dinosaurs, the rudistid clams, and many species of fish, not to mention the total elimination or severe reduction of countless smaller species from the sea. Once in a great while, the earth moves through cosmic dust that collects near the arms of the galaxy. It has been suggested that this dust might have deflected enough sunlight to bring on the biotic catastrophe. An ocean gone stagnant would have done the same, killing phytoplankton, which prosper in the currents of mixed-up seas. Break the food chain and creatures die out above the break. Phytoplankton are the base of the food chain. The Arctic Ocean, surrounded by continents that had drifted together, might have become in the Cretaceous the greatest lake in all eternity, and when the North Atlantic opened up enough to let the water flood the southern seas the life in them would have suffered a cold osmotic shock. Drastic fluctuations of sea level—also related, perhaps, to the separation of continents—might have caused changes in air temperature and ocean circulation that were enough to sunder the food chain. At the end of 1979, a small group at the Lawrence Radiation Laboratory, in Berkeley—among them the physicist Luis Alvarez, winner of a Nobel Prize, and his son Walter, who is a geologist—brought forth a piece of science in which they present the catastrophe as the effect of an Apollo Object colliding with the earth. An Apollo Object is an "earth-orbit-crossing" asteroid that is at least a kilometre in

diameter and is in the category of asteroids that have pockmarked the surface of Mercury, Mars, and the moon, and the surface of the earth as well, although most of the evidence has been obscured here by erosion. Like the general run of meteorites, an Apollo Object could be expected to contain a percentage of iridium and other platinum-like metals at least a thousand times greater than the concentration of the same metals in the crust of the earth. In widely separated parts of the world—Italy, Denmark, New Zealand—the Berkeley researchers have found a thin depositional band, often just a centimetre thick, that contains unearthly concentrations of iridium. Below that sharp line are abundant Cretaceous fossils, and above it they are gone. It marks precisely the end of Cretaceous time. The Berkeley calculations suggest an asteroid about six miles in diameter hitting the earth with a punch of a hundred million megatons, making a crater a hundred miles wide. Such an occurrence—which could repeat itself tomorrow afternoon, there being something like seven hundred big asteroids out there in threatening orbits—would have sent up a mushroom cloud containing some thirty thousand cubic kilometres of pulverized asteroid and terrestrial crust, part of which would have gone into the stratosphere and spread quickly over the earth, keeping sunlight off the lands and seas and suppressing photosynthesis. On August 26 and 27, 1883, when the island Krakatoa, in the Sunda Strait, exploded with great violence, it sent less than twenty cubic kilometres of material into the air, but within a few days dust had spread above the whole earth, turning daylight into dusk. It made exceptionally brilliant sunsets for two and a half years. Edmund Halley, who died when James Hutton was fifteen, once wrote a paper suggesting that the way God started Noah's Flood was by directing a big comet into collision with the earth. The Cretaceous Extinction, whatever its cause, was one of the two most awesome annihilations of life in the history of the world. With the Permian Extinction before it, it framed the Mesozoic, an era of burgeoning creation within deadly brackets of time.

For establishing our bearings through time, we obviously owe an incalculable debt to vanished and endangered species, and if the condor, the kit fox, the human being, the black-footed ferret, and the three-toed sloth are at the head of the line to go next, there is less cause for dismay than for placid acceptance of the march of prodigious tradition. The opossum may be Cretaceous, certain clams Devonian, and oysters Triassic, but for each and every oyster in the sea, it seems, there is a species gone forever. Be a possum is the message, and you may outlive God. The Cenozoic era—coming just after the Cretaceous Extinction, and extending as it does to the

latest tick of time—was subdivided in the eighteen-thirties according to percentages of molluscan species that have survived into the present. From the Eocene, for example, which ended some thirty-eight million years ago, roughly three and a half per cent have survived. Eocene means "dawn of the recent." The first horse appeared in the Eocene. Looking something like a toy collie, it stood three hands high. From the Miocene ("moderately recent"), some fifteen per cent of molluscan species survive; from the Pliocene ("more recent"), the number approaches half. As creatures go, mollusks have been particularly hardy. Many species of mammals fell in the Pliocene as prairie grassland turned to tundra and ice advanced from the north. From the Pleistocene ("most recent"), more than ninety per cent of molluscan species live on. The Pleistocene has also been traditionally defined by four great glacial pulsations, spread across a million years—the Nebraskan ice sheet, the Kansan ice sheet, the Illinoian and Wisconsinan ice sheets. It now appears that these were the last of many glacial pulsations that have occurred in relatively recent epochs, beginning probably in the Miocene and reaching a climax in the ice sheets of Pleistocene time. The names of the Cenozoic epochs were proposed by Charles Lyell, whose *Principles of Geology* was the standard text through much of the nineteenth century. To settle problems here and there, the Oligocene ("but a little recent") was inserted in the list, and the Paleocene ("old recent") was sliced off the beginning. Paleocene, Eocene, Oligocene, Miocene, Pliocene, Pleistocene—sixty-five million to ten thousand years before the present. Divisions grew shorter in the Cenozoic—the epochs range from eighteen million years to less than two million—because so much remains on earth of Cenozoic worlds.

Ignoring its geology, I guess I don't know a paragraph in literature that I prefer to the one Joseph Conrad begins by saying, "Going up that river was like travelling back to the earliest beginnings of the world, when vegetation rioted on the earth and the big trees were kings." He says, moments later, "This stillness of life did not in the least resemble a peace. It was the stillness of an implacable force brooding over an inscrutable intention. It looked at you with a vengeful aspect. I got used to it afterwards; I did not see it anymore; I had no time. I had to keep guessing at the channel; I had to discern, mostly by inspiration, the signs of hidden banks; I watched for sunken stones." Metaphorically, he travelled back to the Carboniferous, when the vegetal riot occurred, but scarcely was that the beginning of the world. The first plants to appear on land, ever, appeared in the Silurian. Through the Ordovician and the Cambrian, there had been no terrestrial vegetation at all. And in the deep shadow below the Cambrian were seven years for every one in all subsequent time. There were four billion years back there—since the earliest beginnings of the world. There were scant to nonexistent fossils. There were the cores of the cratons, the rock of the continental shields, the rock of the surface of the moon. There were the reefs of the Witwatersrand. There was the rock that would become the Adirondack Mountains, the Wind River summits, the Seward Peninsula, Manhattan Island. But so little is known of this seven-eighths of all history that in a typical two-pound geological textbook there are fourteen pages on Precambrian time. The Precambrian has attracted geologists of exceptional imagination, who see families of mountains in folded schists. Uranium-lead and potassium-argon radiometric dating have helped them to sort out their Kenoran, Hudsonian, Elsonian Orogenies, their Aphebian, Hadrynian, Paleohelikian time. Isolating the first two billion years of the life of the earth, they called it the Archean Eon. In the Middle Archean, photosynthesis began. Much later in the Precambrian, somewhere in Helikian or Hadrynian time, aerobic life appeared. There is no younger rock in the United States than the travertine that is forming in Thermopolis, Wyoming. A 2.7-billion-year-old outcrop of the core of the continent is at the head of Wind River Canyon, twenty miles away. Precambrian—4,600 to 570 million years before the present.

At the other end of the scale is the Holocene, the past ten thousand years, also called the Recent—Cro-Magnon brooding beside the melting ice. (The Primitive and Secondary eras of eighteenth-century geology are long since gone from the vocabulary, but oddly enough the Tertiary remains. The term, which is in general use, embraces nearly all of the Cenozoic, from the Cretaceous Extinction to the end of the Pliocene, while the relatively short time that follows—the Pleistocene plus the Holocene—has come to be called the Quaternary. The moraines left by ice sheets are Quaternary, as are the uppermost basin fillings in the Basin and Range.) It was at some moment in the Pleistocene that humanity crossed what the geologist-theologian Pierre Teilhard de Chardin called the Threshold of Reflection, when something in people "turned back on itself and so to speak took an infinite leap forward. Outwardly, almost nothing in the organs had changed. But in depth, a great revolution had taken place: consciousness was now leaping and boiling in a space of super-sensory relationships and representations; and simultaneously consciousness was capable of perceiving itself in the concentrated simplicity of its faculties. And all this happened for the first time." Friars of another sort—evangelists of the so-called Environmental Movement—have often made use of the geologic time scale to place in perspective that great "leap forward" and to suggest what our reflective capacities may have meant to Mother Earth. David Brower, for example, the founder of Friends of the Earth and emeritus hero of the Sierra Club, has tirelessly travelled the United States for thirty years delivering what he himself refers to as "the sermon," and sooner or later in every talk he invites his listeners to consider the six days of Genesis as a figure of speech for what has in fact been four and a half billion years. In this adjustment, a day equals something like seven hundred and fifty million years, and thus "all day Monday and until Tuesday noon creation was busy getting the earth going." Life began Tuesday noon, and "the beautiful, organic wholeness of it" developed over the next four days. "At 4 P.M. Saturday, the big reptiles came on. Five hours later, when the redwoods appeared, there were no more big reptiles. At three minutes before midnight, man appeared. At one-fourth of a second before midnight, Christ arrived. At one-fortieth of a second before midnight, the Industrial Revolution began. We are surrounded with people who think that what we have been doing for that one-fortieth of a second can go on indefinitely. They are considered normal, but they are stark raving mad." Brower holds up a photograph of the world—blue, green, and swirling white. "This is the sudden insight from Apollo," he says. "There it is. That's all. We see through the eyes of the astronauts how fragile our life really is." Brower has computed that we are driving through the earth's resources at a rate comparable to a man's driving an automobile a hundred and twenty-eight miles an hour—and he says that we are accelerating.

In like manner, geologists will sometimes use the calendar year as a unit to represent the time scale, and in such terms the Precambrian runs from New Year's Day until well after Halloween. Dinosaurs appear in the middle of December and are gone the day after Christmas. The last ice sheet melts on December 31st at one minute before midnight, and the Roman Empire lasts five seconds. With your arms spread wide again to represent all time on earth, look at one hand with its line of life. The Cambrian begins in the wrist, and the Permian Extinction is at the outer end of the palm. All of the Cenozoic is in a fingerprint, and in a single stroke with a medium-grained nail file you could eradicate human history. Geologists live with the geologic scale. Individually, they may or may not be alarmed by the rate of exploitation of the things they discover, but, like the environmentalists, they use these repetitive analogies to place the human record in perspective—to see the Age of Reflection, the last few thousand years, as a small bright sparkle at the

end of time. They often liken humanity's presence on earth to a brief visitation from elsewhere in space, its luminous, explosive characteristics consisting not merely of the burst of population in the twentieth century but of the whole millennial moment of people on earth—a single detonation, resembling nothing so much as a nuclear implosion with its successive neutron generations, whole generations following one another once every hundred-millionth of a second, temperatures building up into the millions of degrees and stripping atoms until bare nuclei are wandering in electron seas, pressures building up to a hundred million atmospheres, the core expanding at five million miles an hour, expanding in a way that is quite different from all else in the universe, unless there are others who also make bombs.

The human consciousness may have begun to leap and boil some sunny day in the Pleistocene, but the race by and large has retained the essence of its animal sense of time. People think in five generations—two ahead, two behind—with heavy concentration on the one in the middle. Possibly that is tragic, and possibly there is no choice. The human mind may not have evolved enough to be able to comprehend deep time. It may only be able to measure it. At least, that is what geologists wonder sometimes, and they have imparted the questions to me. They wonder to what extent they truly sense the passage of millions of years. They wonder to what extent it is possible to absorb a set of facts and move with them, in a sensory manner, beyond the recording intellect and into the abyssal eons. Primordial inhibition may stand in the way. On the geologic time scale, a human lifetime is reduced to a brevity that is too inhibiting to think about. The mind blocks the information. Geologists, dealing always with deep time, find that it seeps into their beings and affects them in various ways. They see the unbelievable swiftness with which one evolving species on the earth has learned to reach into the dirt of some tropical island and fling 747s into the sky. They see the thin band in which are the all but indiscernible stratifications of Cro-Magnon, Moses, Leonardo, and now. Seeing a race unaware of its own instantaneousness in time, they can reel off all the species that have come and gone, with emphasis on those that have specialized themselves to death.

In geologists' own lives, the least effect of time is that they think in two languages, function on two different scales.

"You care less about civilization. Half of me gets upset with civilization. The other half does not get upset. I shrug and think, So let the cockroaches take over."

"Mammalian species last, typically, two million years. We've about used up ours. Every time Leakey

finds something older, I say, 'Oh! We're overdue.' We will be handing the dominant-species-on-earth position to some other group. We'll have to be clever not to."

"A sense of geologic time is the most important thing to suggest to the nongeologist: the slow rate of geologic processes, centimetres per year, with huge effects, if continued for enough years."

"A million years is a short time—the shortest worth messing with for most problems. You begin tuning your mind to a time scale that is the planet's time scale. For me, it is almost unconscious now and is a kind of companionship with the earth."

"It didn't take very long for those mountains to come up, to be deroofed, and to be thrust eastward. Then the motion stopped. That happened in maybe ten million years, and to a geologist that's really fast."

"If you free yourself from the conventional reaction to a quantity like a million years, you free yourself a bit from the boundaries of human time. And then in a way you do not live at all, but in another way you live forever."

Gould, S.J.
"Quick lives and quirky
changes"
Natural History
American Museum of Nat. Hist.
Used by permission

QUICK LIVES AND QUIRKY CHANGES

Solitary Origins of Some Societies

The "wondrous indefiniteness of evolution" includes many quirky shifts in function

by Stephen Jay Gould

Posthumous triumph is hollow, however abstractly rewarding. Nanki-Poo refused Ko-Ko's inducement to undergo a ceremonious public beheading rather than a private suicide: "There'll be a procession—bands—dead march—bells tolling . . . then, when it's all over, general rejoicings, and a display of fireworks in the evening. You won't see them, but they'll be there all the same." And I never could figure out why America's premier nineteenth-century anthropologists J.W. Powell and W J McGee made a bet about who had the larger brain—to be settled by autopsy when the joy of victory could no longer be savored.

Nonetheless, I just made a dumb bet with a female jogging enthusiast: that no woman would win the Boston marathon in my lifetime. I'd rather lose, but expect I won't. Still, if superior average speed of running males is among the few insignificant, but genuinely biological, differences between human sexes, I can only respond to charges of gloating (for the abstraction I represent, but not, alas, for me and my huffing eight-minute miles) with a statement of genuine regret. How gladly would I trade this useless advantage for the most precious benefit of being female—several extra years of average life.

I do not know whether shorter male life is a generality in nature—and whether we should therefore add to smaller average size (see July's column) another biological strike against machismo—but I just learned (with thanks to Martin L. Adamson) about an instructive extreme case.

In 1962, James H. Oliver, Jr., traced the life cycle of a mite that parasitizes the cocoons of earthworms (*Journal of Parasitology*, vol. 48, p. 120). Both males and females of *Histiostoma murchiei* pass through an egg and three juvenile stages before molting into an adult. In addition, the female intercalates one additional stage—euphoniously named the hypopus—between the second and third preadult phases. Females develop at a leisurely pace for such a small creature. Discounting the hypopus, the passage from egg to adult, through stages held in common with males, takes one to three weeks. The additional hypopus may extend female life greatly—for these mites find and infest other cocoons only while in the hypopal stage (males always stay at home). The hypopus may, first of all, remain dormant for long periods within the skin of the previous juvenile stage, awaiting (so to speak) favorable conditions for emergence and movement to another cocoon. When the hypopus does emerge, it may then live for a long time, moving about in its own cocoon (and sometimes becoming dormant again) or moving out in search of a new home.

Males, by contrast, race through the same stages (minus the hypopus) with a celerity that should inspire Bill Rodgers as he trudges up Heartbreak Hill next Patriot's Day. "Adult males," Oliver writes, "have been observed copulating with their mother within three to four days after being laid as eggs," and they die soon after this bout of incestuous joy. Why this outstanding difference in life span between the sexes? And what has it to do with the Oedipal habits of these mites? A further look at the unusual reproductive biology of these parasites seems to provide the answer.

When a hypopus finds a new cocoon, it lays two to nine eggs within two days after molting into an adult—and without benefit of fertilization. All these eggs develop into males, the only source of potential husbands. What better evolutionary rationale for rapid male development could we hope to find? The females of most species must seek their husbands. These mites make them from scratch and then wait. Males of *H. murchiei* are little more than sources of sperm; the sooner they can perform, the better.

Two days after her incestuous mating, the female begins to lay eggs again and may continue for two to five days, producing as many as 500 offspring—all female this time.

Yet in solving one problem—the differential speed of development between sexes—we have only encountered more curious questions: How can this system work in the first place? How can an unmated female, alone in a new cocoon, produce a generation of husbands? And why are the offspring of her next reproductive bout all female?

The answers lie in the unfamiliar style of sex determination in these mites. In most animals, both males and females have paired chromosomes, and the status of one pair determines the sex of its bearer. Human females, for example, have two large sex chromosomes (designated XX), while males have one large (X) and one small (Y) chromosome in their determining pair. All unfertilized egg cells carry a single X, while sperm carry either an X or a Y. We each owe our sex to the good fortune of one sperm among the millions per ejaculate. Animals with paired chromosomes in both sexes are called diploid.

Some animals use a different system of sex determination. Females are diploid, but males have only one chromosome for each female pair and are called haploid (for half the diploid number). In other words, males—ironic as this may seem—develop from unfertilized eggs and have no fathers. Fertilized eggs produce diploid females. Animals using this system are called haplodiploid (because males are haploid and females diploid).

H. murchiei is haplodiploid. Hence, the unmated female in a new cocoon raises a generation of males from unfertilized eggs, and a subsequent generation of females from the resultant incest.

Haplodiploidy, a fascinating phenomenon rich in implication, has circulated through these columns in various contexts for years. It helped to explain the origin of social systems in ants and bees (November 1976), and it underlay the habits of a male mite that fertilizes several sisters within his mother's body and dies before "birth" (March 1980). It also circulates widely through the animal kingdom. Haplodiploid species have been found in rotifers, nematodes, mites, and in four separate orders of insects—the Thysanoptera (thrips), the Homoptera (aphids, cicadas, and their allies), the Coleoptera (beetles), and the Hymenoptera (ants, bees, and wasps). These groups are not closely related and their presumed common ancestors are diploid. Thus, haplodiploidy has arisen independently—and often many times—within each group. Although most of these groups contain only a few haplodiploid species amidst a host of ordinary diploids, the Hymenoptera, with more than 100,000 named species, are exclusively haplodiploid. Since vertebrates only include some 50,000 species, as Oliver reminds us, our chauvinistic impression that haplodiploidy is curious or rare should also be revised. At least 10 percent of all named animal species are haplodiploid, and those 100,000 hymenopterans form a good chunk of the 700,000 or so insect species that taxonomists have recognized.

Within the last decade, haplodiploidy has figured most prominently in the news (both general and scientific) for its role in an ingenious Darwinian explanation of an old biological mystery: the origin of sociality in Hymenoptera, particularly the existence of sterile "worker" castes, invariably female, in ants and bees. Since sociality evolved several times within the Hymenoptera, the invariant system of sterile female castes demands a general explanation. The larger problem is even more puzzling: Why, in a presumably Darwinian world filled with organisms acting only for their personal reproductive success, should large numbers of females "forgo" their own reproduction to help their mother (the queen) raise more sisters?

The ingenious explanation (in brief; see my previous column for more details) relies upon the peculiar asymmetries of genetic relationship between sexes in haplodiploid animals. In both diploids and haplodiploids, mothers pass half their genetic material (one set of chromosomes in each egg cell) to each offspring. They are therefore equally related (by half of their genetic selves) to both sons and daughters. A female in diploid species also shares approximately half her genes with both brothers and sisters. But a female in haplodiploid species shares three-fourths of her genes with sisters and only one-fourth with brothers, for the following reason. Consider any gene (on a single chromosome) in sisters. What is the probability that a brother will share it? If the gene is on a paternal chromosome, then the brother has zero probability of sharing it, for he has no paternal chromosomes. If the gene is on a maternal chromosome, then he has a 50 percent chance of sharing it with his sister because he either received the same chromosome from his mother or the other member of the pair. Thus, summing over all genes, the relationship between brother and sister is the average between zero (for paternal genes of sisters, necessarily absent in brothers) and 50 percent (for maternal genes), or 25 percent.

What then is the probability that a sister will share the same gene? If it is a paternal gene, the sister must share it since fathers have only one set of chromosomes and they pass their entire genetic program to each daughter. If it is a maternal gene, the chance is 50 percent by the same argument advanced for brothers. The total relationship between sisters is therefore the average between 100 percent (for paternal genes) and 50 percent (for maternal genes), or 75 percent.

Females are therefore more closely related to their sisters (by three-fourths) than either to their mothers (by one-half) or to their own potential offspring (also by one-half). If the Darwinian imperative leads organisms to maximize the numbers of their own genes in future generations, then females will do better by helping their mother raise sisters (as sterile workers do) than by producing their own offspring. Thus, the asymmetry of genetic relationship in haplodiploids may explain why worker castes of social Hymenoptera are invariably female and why sociality in this style has evolved many times in the Hymenoptera but not in the much larger array of diploid organisms. (As always, our complex world provides an exception —the diploid termites, relatives of cockroaches, which at least include both males and females in their worker castes.)

This explanation of an old mystery has so intrigued biologists that a subtle reversal of causality has crept into some accounts. The very existence of haplodiploidy is linked with force and elegance to the evolution of sociality, and we are almost led to believe that this mode of sex determination arose "for," or at least in the context of, the marvelous social organization of ants and bees. Yet a moment's explicit reflection assures us that this cannot be so, for two reasons.

First, *all* hymenopterans are haplodiploid, but only a few lineages within the group have developed complex social systems (most hymenopterans are asocial or minimally social wasps). The common ancestor of living hymenopterans must have been haplodiploid, but it was certainly not fully social since the complex society of highly derived bees and ants has evolved as a phyletic afterthought in several independent lineages. Causality must run in the other direction. Haplodiploidy does not exist for sociality unless the future can control the past. Rather, haplodiploidy arose for other reasons and then permitted, by good and unplanned fortune, the later evolution of this wonderfully complex and successful mode of sociality. But what other reasons? Which brings me, finally, to the point of this column, to the main reason for my fascination with *H. murchiei*, and more immediately, to the second item.

When we consider the usual ecological context of haplodiploidy in a broad range of animals that may have evolved it directly (and not merely co-opted it for another use), an interesting pattern emerges. *H. murchiei* shares a mode of life with the mites that die before birth and with many other haplodiploid animals in distantly related groups: all are "colonizers," species that survive by seeking rare but rich resources and then reproducing as fast as they can when uncommon fortune rewards their search (the vast majority of *Histiostoma*'s hypopi die before finding a fresh earthworm cocoon). Haplodiploidy provides several advantages in this chancy approach to survival. Successful colonization does not require two separate migrations of a male and a female or even that a single migrating female be fertilized before her search for a new resource begins. Any unmated female, even a juvenile, becomes a potential source of new colonies, since she can make a generation of males all by herself and then mate with them to begin a generation of females—the strategy evolved by *Histiostoma*.

When colonizers find a rich but ephemeral resource, haplodiploidy may enhance the speed of raising new generations by permitting fertilized females to control the sex ratio of their offspring. As I argued in my column on "death before birth," when brothers mate with sisters, more offspring will people the next generation if mothers can put most of their limited reproductive energy into making females and produce only a minimal number of males (one will often do). Since one male may fertilize many females, and the reproductive rate of a population is limited by the number of eggs, not by sperm, why make vast numbers of superfluous males? The principle is fine in theory, but most animals cannot easily control the sex ratio of their offspring. Despite prayers and entreaties for boys in many sexist human societies, girls continue to assert their birthright (and birth rate) of nearly 50 percent.

But many haplodiploids can control the sex ratio of their offspring. If a female stores sperm within her body after mating, any eggs that bypass the storage area become males, while those that contact it become females. Haplodiploid mites with highly unequal sex ratios often produce a brood of female eggs and then shut off the sperm supply to add a male or two right at the end.

This complex of associated features—a colonizing life style, rare and ephemeral resources, rapid reproduction, and ease of rearing new generations in strange places—seems to define the original context of advantage for haplodiploidy. If we assume, as a hypothesis only, that haplodiploidy usually arises as an adaptation for life in this uncertain world, then it must be interpreted as a lucky accident with respect to its later utility in the development of sociality in ants and bees.

Now what could be more different, in our usual biological thinking, than the chancy life of a solitary female colonizer (whose offspring can hardly become very social on a resource that doesn't last more than a generation or two), and the complexity, stability, and organization of ant and bee societies? Is it not peculiar in the extreme that haplodiploidy, a virtual prerequisite for the evolution of hymenopteran societies, probably first evolved as an adaptation for a life style almost diametrically opposed (at least in its metaphorical implications)? If I can convince you that it is not peculiar at all, but an example of a basic principle that distinguishes evolutionary biology from a common stereotype about science in general, then this essay has succeeded.

To assume that the current utility of a feature permits an inference about the reasons for its evolutionary origin is a lamentably common error. Current utility and historical origin are different subjects. Any feature, regardless of how or why it first evolved, becomes available for co-optation to other roles, often strikingly different. Complex features are bursting with potentialities; their conceivable use is not confined to their original function (I confess that I have used a credit card to force a door). And these evolutionary shifts in function can be as quirky and unpredictable as the potentials of complexity are vast. It happens all the time; it virtually defines the wondrous indefiniteness of evolution.

The balancing fins of fish become the propulsive limbs of terrestrial vertebrates, while the propulsive tail becomes an organ that often aids in balance. The bone that suspended an ancestral fish's upper jaw to its cranium becomes the bone that transmits sound to the ears of reptiles. Two bones that articulated the jaws of that reptile then become the other two sound-transmitting bones of the mammalian middle ear. When we see how beautifully our hammer, anvil, and stirrup function in hearing, who would imagine that one bone first suspended jaw to cranium, while two others articulated the jaws? And a mode of sex determination that may first have aided a lonely female colonizer apparently became the basis of social systems only rivaled by our own for their complexity.

As we probe deeper and further back, the unpredictabilities mount. I discussed the quirkiness of a functional shift toward support of sociality by a feature probably evolved as an aid to colonization. But what about the larger reason for our imperfect and unpredictable world: structural limits imposed by features evolved for other reasons? Social systems like those of ants and bees might be of enormous advantage to hosts of other creatures. But perhaps they do not evolve simply because it is so difficult to get them started in diploid organisms (only termites have succeeded), while haplodiploid hymenopterans develop them again and again. And going one step further back (I promise to stop here), what about constraints on the evolution of haplodiploidy itself? Haplodiploidy might be a wonderful adaptation to a host of ecologies, but it cannot always be easily evolved.

Assuming that haplodiploids generally arise from diploids, what does it take to turn a haploid creature into a male? Under some systems of diploid sex determination, male haploids cannot easily evolve. A haploid human would not be male, for a single X chromosome induces the development of a sterile female. But other diploids have a so-called XX-XO system of sex determination, in which females have two X chromosomes and males have a single X with no accompanying Y (but all other chromosomes in pairs). In such systems, a haploid organism might develop easily and directly into a male. (The XX-XO system is not an absolute prerequisite for haplodiploidy as more complex modifications also can produce male haploids from other modes of diploid sex determination.)

In short, modes of sex determination limit haplodiploidy, haplodiploidy limits sociality, and sociality requires a quirky shift in the adaptive significance of haplodiploidy. What order can we find in evolution amidst such a crazy quilt of limits to a sensibly perfect and predictable world?

Some might be tempted to read an almost mystical message into this theme—that evolution imposes an ineffable unknowability upon nature. I would strongly reject such an implication: knowledge and prediction are different phenomena. Others might try to read a sad or pessimistic message—that evolution isn't a very advanced science or isn't a science at all if it can't predict the course of an imperfect world. Again, I would reject any such reading of my words about constraint and quirky functional shift.

The problem lies with our simplistic and stereotyped view of science as a monolithic phenomenon based on regularity, repetition, and ability to predict the future. Sciences that deal with objects less complex and less historically bound than life may follow this formula. Hydrogen and oxygen, mixed in a certain way, make water today, made water billions of years ago, and presumably will make water for a long time to come. Same water, same chemical composition. No indication of time, no constraints imposed by a history of previous change.

Organisms, on the other hand, are directed and limited by their past. They must remain imperfect in their form and function, and to that extent unpredictable, since they are not optimal machines. We cannot know their future with certainty if only because a myriad of quirky functional shifts lie within the capacity of any feature, however well adapted it may be to a present role.

The science of complex historical objects is a different, not a lesser, enterprise. It seeks to explain the past, not predict the future. It searches for principles and regularities underlying the uniqueness of each species and interaction, while treasuring that irreducible uniqueness and describing all its glory. Notions of science must bend (and expand) to accommodate life. The art of the soluble, Peter Medawar's definition of science, must not become short-sighted, for life is long.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

This View of Life **by Stephen Jay Gould**

Our Allotted Lifetimes

Gould, S.J.
"Our allotted lifetimes"
Natural History
American Museum of Nat. Hist.
Used by permission

During an average life, all mammals have nearly the same number of breaths and heartbeats

Meeting with Henry Ford in E. L. Doctorow's *Ragtime*, J. P. Morgan praises the assembly line as a faithful translation of nature's wisdom:

> Has it occurred to you that your assembly line is not merely a stroke of industrial genius but a projection of organic truth? After all, the interchangeability of parts is a rule of nature. . . . All mammals reproduce in the same way and share the same designs of self-nourishment, with digestive and circulatory systems that are recognizably the same, and they enjoy the same senses. . . . Shared design is what allows taxonomists to classify mammals as mammals.

An imperious tycoon should not be met with equivocation; nonetheless, I can only reply "yes, and no" to Morgan's pronouncement. Morgan was wrong if he thought that large mammals are geometric replicas of small ones. Elephants have relatively smaller brains and thicker legs than mice, and these differences record a general rule of mammalian design, not the idiosyncracies of particular animals.

Morgan was right in arguing that large animals are essentially similar to small members of their group. The similarity, however, does not lie in a constant shape. As I have argued in several columns, the basic laws of geometry dictate that animals must change their shape in order to perform the same function at different sizes. I remind readers of the classical example, first discussed by Galileo in 1638: the strength of an animal's leg is a function of its cross-sectional area (length × length); the weight that the leg must support varies as the animal's volume (length × length × length). If a mammal did not alter the relative thickness of its legs as it got larger, it would soon collapse since body weight would increase much faster than the supporting strength of limbs. Instead, large mammals have relatively thicker leg bones than small mammals. To remain the same in function, animals must change their form.

The study of these changes in form is called "scaling theory." Scaling theory has uncovered a remarkable regularity of changing shape over the 25-millionfold range of mammalian weight from shrew to blue whale. If we plot brain weight versus body weight for all mammals on the so-called mouse-to-elephant (or shrew-to-whale) curve, very few species deviate far from a single line expressing the general rule: brain weight increases only two-thirds as fast as body weight as we move from small to large mammals. (We share with bottle-nosed dolphins the honor of greatest deviance from the curve.)

We can often predict these regularities from the physicial behavior of objects. The heart, for example, is a pump. Since all mammalian hearts are similar in function, small hearts will pump considerably faster than large ones (imagine how much faster you could work a finger-sized toy bellows than the giant model that fuels a blacksmith's large forge). On the mouse-to-elephant curve for mammals, the length of a heartbeat increases between one-fourth and one-third as fast as body weight as we move from small to large mammals. The generality of this conclusion has just been affirmed in an interesting study by J. E. Carrel and R. D. Heathcote on the scaling of heart rate in spiders. They used a cool laser beam to illuminate the hearts of resting spiders and drew a crab spider-to-tarantula curve for eighteen species spanning nearly a thousandfold range of body weight. Again, scaling is very regular with heart rate increasing four-tenths as fast as body weight (or .409 times as fast, to be exact).

We may extend this conclusion for hearts to a very general statement about the pace of life in small versus large animals. Small animals tick through life far more rapidly than large ones—their hearts work more quickly, they breathe more frequently, their pulse beats much faster. Most importantly, metabolic rate, the so-called fire of life, scales only three-fourths as fast as body weight in mammals. Large mammals generate much less heat per unit of body weight to keep themselves going. Tiny shrews move frenetically, eating nearly all their waking lives to keep their metabolic fire burning at its maximal rate among mammals; blue whales glide majestically, their hearts beating the slowest rhythm among active, warmblooded creatures.

If we consider the scaling of lifetime among mammals, an intriguing synthesis of these disparate data seems to suggest itself. We have all had enough experience with mammalian pets of various sizes to understand that small mammals tend to live for a shorter time than large ones. In fact, the scaling of mammalian lifetime follows a regular curve at about the same rate as heartbeat and breath time—between one-fourth and one-third as fast as body weight as we move from small to large animals. (Again, *Homo sapiens* emerges as a very peculiar animal. We live far longer than a mammal of our body size should. In previous columns, I have argued that humans evolved by a process called "neoteny"—the retention of shapes and growth rates that characterize juvenile stages of our primate ancestors. I also believe that neoteny is responsible for our elevated longevity. Compared with other mammals, all stages of human life—from juvenile features to adulthood—arise "too late." We are born as helpless embryos after a long gestation; we mature late after an extended childhood; we die, if fortune be kind, at ages otherwise reached only by the very largest warmblooded creatures.)

Usually, we pity the pet mouse or gerbil that lived its full span of a year or two at most. How brief its life, while we endure for the better part of a century. As the main theme of this column, I want to argue that such pity is misplaced (our personal grief, of course, is quite another matter; with this, science does not deal). J. P. Morgan of *Ragtime* was right—small and large mammals are essentially similar. Their lifetimes are scaled to their life's pace, and all endure for approximately the same amount of biological time. Small mammals tick fast, burn rapidly, and live for a short time; large ones live long at a stately pace. Measured by their own internal clocks, mammals of different sizes tend to live for the same amount of time.

Yet we are prevented from grasping this important and comforting concept by a deeply ingrained habit of Western thought. We are trained from earliest memory to regard absolute Newtonian time as the single valid measuring stick in a rational and objective world. We impose our kitchen clock, ticking equably, upon all things. We marvel at the quickness of a mouse, express boredom at the torpor of a hippopotamus. Yet each is living at the appropriate pace of its own biological clock.

I do not wish to deny the importance of absolute, astronomical time to organisms. Animals must measure it to lead successful lives. Deer must know when to regrow their antlers, birds when to migrate. Animals track the day–night cycle with their circadian rhythms; jet lag is the price we pay for moving much faster than nature intended. The bamboos of my April column somehow count 120 years before flowering again.

But absolute time is not the appropriate measuring stick for all biological phenomena. Consider the song of the humpback whale. These magnificent animals sing with such volume that their sounds travel through water for thousands of miles, perhaps even around the world, as their leading student Roger S. Payne has suggested. E. O. Wilson has described the awesome effect of these vocalizations: "The notes are eerie yet beautiful to the human ear. Deep basso groans and almost inaudibly high soprano squeaks alternate with repetitive squeals that suddenly rise or fall in pitch." We do not know the function of these songs. Perhaps they enable whales to find each other and to stay together during their annual transoceanic migrations.

Each whale has its own characteristic song; the highly complex patterns are repeated over and over again with great faithfulness. No scientific fact that I have learned in the last decade struck me with more force than Payne's report that the length of some songs may extend for more than half an hour. I have never been able to memorize the five-minute first Kyrie of the B-minor Mass (and not for want of trying); how could a whale sing for thirty minutes and then repeat itself accurately? Of what possible use is a thirty-minute repeat cycle—far too long for a human to recognize; we would never grasp it as a single song (without Payne's recording machinery and much study after the fact). But then I remembered the whale's metabolic rate, the enormously slow pace of its life compared with ours. What do we know about a whale's perception of thirty minutes? A humpback may scale the world to its own metabolic rate: its half-hour song may be our minute waltz. From any point of view, the song is spectacular; it is the most elaborate single display so far discovered in any animal. I merely urge the whale's point of view as an appropriate perspective.

We can provide some numerical precision to support the claim that all mammals, on average, live for the same amount of biological time. In a method developed by W. R. Stahl, B. Gunther, and E. Guerra in the late 1950s and early 1960s, we search the mouse-to-elephant equations for biological properties that scale at the same rate against body weight. For example, Gunther and Guerra give the following equations for mammalian breath time and heartbeat time versus body weight.

$$\text{breath time} = .0000470\ \text{body}^{0.28}$$
$$\text{heartbeat time} = .0000119\ \text{body}^{0.28}$$

(Nonmathematical readers need not be overwhelmed by the formalism. The equations simply mean that both breath time and heartbeat time increase about .28 times as fast as body weight as we move from small to large mammals.) If we divide the two equations, body weight cancels out because it is raised to the same power.

$$\frac{\text{breath time}}{\text{heartbeat time}} = \frac{.0000470\ \cancel{\text{body}^{0.28}}}{.0000119\ \cancel{\text{body}^{0.28}}} = 4.0$$

This says that the ratio of breath time to heartbeat time is 4.0 in mammals of any body size. In other words, all mammals, whatever their size, breathe once for each four heartbeats. Small animals breathe and beat their hearts faster than large animals, but both breath and heart slow up at the same relative rate as mammals get larger.

Lifetime also scales at the same rate to body weight (.28 times as fast as we move from small to large mammals). This means that the ratio of both breath time and heartbeat time to lifetime is also constant over the whole range of mammalian size. When we perform an exercise similar to that above, we find that all mammals, regardless of their size, tend to breathe about 200 million times during their lives (their hearts, therefore, beat about 800 million times). Small mammals breathe fast, but live for a short time. Measured by the sensible internal clocks of their own hearts or the rhythm of their own breathing, all mammals live about the same time. (Astute readers, having counted their breaths, may have calculated that they should have died long ago. But *Homo sapiens* is a markedly deviant mammal in more ways than braininess alone. We live about three times as long as mammals of our body size "should," but we breathe at the "right" rate and thus live to breathe about three times as much as an average mammal of our body size.)

The mayfly lives but a day as an adult. It may, for all I know, experience that day as we live a lifetime. Yet all is not relative in our world, and such a short glimpse of it must invite distortion in interpreting events ticking on longer scales. In a brilliant metaphor, the pre-Darwinian evolutionist Robert Chambers spoke of a mayfly watching the metamorphosis of a tadpole into a frog (from *Vestiges of the Natural History of Creation, 1844*):

> Suppose that an ephemeron [a mayfly], hovering over a pool for its one April day of life, were capable of observing the fry of the frog in the waters below. In its aged afternoon, having seen no change upon them for such a long time, it would be little qualified to conceive that the external branchiae [gills] of these creatures were to decay, and be replaced by internal lungs, that feet were to be developed, the tail erased, and the animal then to become a denizen of the land.

Human consciousness arose but a minute before midnight on the geologic clock. Yet we mayflies, ignorant perhaps of the messages buried in earth's long history, try to bend an ancient world to our purposes. Let us hope that we are still in the morning of our April day.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

AMERICAN SCIENTIST, 56, 4, pp. 400–413, 1968

Went, A.F.
"The size of man"
American Scientist
Published by Sigma Xi
Used by permission

THE SIZE OF MAN

By F. W. WENT

IN THE following article, I want to show how important his physical size is for man and how many of his attainments, such as the development of a technology, were possible only because of his specific size. In the course of these considerations, it will become clear that many physiological and mechanical processes have grave limitations in relation to size. We cannot indiscriminately magnify or reduce processes from the microscopic to the macroscopic level or the reverse without physiological distortion.

To emphasize the significance of size, frequent comparisons will be made between man and ant. In spite of the enormous difference in size (linearly, three orders of magnitude; on the basis of mass, 8 orders of magnitude), both have developed highly complex social systems, and ants (like bees, termites, and other social insects) have also effective communications systems. Whereas the social insects developed millions of years ago to their high degree of specialization, they never developed a technology with which they could have dominated the earth. On the other hand once he attained a certain degree of social differentiation, man explosively developed to his present state in the course of only 10,000 years. Whereas I cannot possibly attempt to discuss all contributing factors to man's emergence, I want to point out that his physical size was a dominating factor.

The relative importance of size was completely unknown until man started to probe into the microscopic world with lenses and microscopes.

When, in the second half of the seventeenth century, man looked for the first time through a microscope, he suddenly became aware of an entirely new world—until then almost unsuspected, the micro-world. The genius of these first microscopists made them realize what an important step they were taking, and the showmanship of some of them, notably Leeuwenhoek, brought the significance of this new world also to the attention of nonscientists. The common man, and royalty, flocked to the door of Leeuwenhoek's simple house in Delft to catch a glimpse of this new world, so full of unsuspected wonders. Suddenly, highly important principles, such as that of preformation and *generatio spontanea*, could be approached visually and experimentally. Curiously enough the basis not only for new understanding but also for a schism was laid.

One of the first men who speculated on the remarkable possibilities which magnification or diminution of physical dimensions provides was Jonathan Swift, who, in *Gulliver's Travels*, drew some conclusions as to what dwarfs and giants would really look like, and what sociological consequences size would have. Some time ago Florence Moog (*Scientific American*, November 1948) showed that Swift was a "bad biologist," or Gulliver a "poor liar." She showed that a linear reduction in size would carry with it a reduction in the number of brain cells, and hence a reduction in intellectual capacity in Lilliputians, whereas the enormous Brobdingnagians were physically impossible; they could have had physical reality only if their necks and legs had been short and thick. These 90-ton monsters could never have walked on dry land, nor could their tremendous weight have been carried on proportionately-sized feet.

Even though Swift, in his phantasy, committed a number of physical errors, because he was not sufficiently aware of the fact that some physical properties of a body are proportional to the linear dimensions (height), whereas others vary with the third power of linear size (such as weight and cell number), yet he surpassed his medieval predecessors in many respects and drew a number of excellent conclusions, bringing both giants and dwarfs close to physical reality.

Since I want to make an argument, that between the macroscopic world (visible by the naked eye) and the microworld (of microscopic dimensions) a definite discontinuity exists, we first will consider whether the opposite view of a gradual transition has any merit. For instance, the microscopic embryo already has a number of features which the young or mature individual shows; sand grains under a microscope are like rocks, and a copepod seems nothing more than a miniature shrimp. And yet, it is curious that commonly there is not just a quantitative, but also a qualitative difference where we cross the border between macro- and microworld. For instance, all animals in the macro-world have hearts or more primitive means of circulation of body fluids. On the other hand, the cell, the unit of life activity, is almost without exception a microscopic structure. Similarly, the legs of animals, and walking, are achievements of the macro-world, whereas chemical cell reactions can be understood only on the molecular level.

The specific size of each organism is partly determined by its life habits. An elephant who eats branches (which are the food storage organs of many trees) must have enormous teeth to grind them, whereas harvester ants, living on tiny seeds gleaned from the soil, could not possibly be large and still collect enough food. Similarly parasites must be smaller than the organisms they grow on.

In spite of the enormous differences in size between species, cells of all organisms are equal in size and have an average volume of $10^3\mu^3$ in animals and $10^5\mu^3$ in plants. Therefore, size differences in plants and animals are almost exclusively due to differences in cell number. This suggests that fundamentally a cell functions well only within the size range which normally occurs. This is even more clearly indicated when we compare cell sizes within a single organism. In plants the most active cells, which are growing and dividing and have the greatest metabolic rate, are those in the growing point of stem and root tips, and these are much smaller than the mature cells and more nearly equal in size. They range from 1 to 3 $\times 10^3\mu^3$ for the meristematic cells of the fastest growing plants, and are in the $10^4\mu^3$ range for the slower growing ones.

This uniform cell size suggests that the most fundamental cell activities are dependent upon size. This conclusion is strengthened by the consideration that the most active cells are smallest and are most nearly alike in size, both in plants and animals. Actually the fastest growing cells are those of bacteria, with a generation time as short as 20 minutes. For cells in the growing point of a plant, the time elapsed between successive divisions is 24 hours or more, or about 100 times as long, whereas the linear dimensions are in the proportion of 1:10. Therefore, we might conclude that the division rate and the activity of the cell are controlled by a process which decreases in rate with the square of the diameter of the cell. This is typical for a diffusion process as shown in Figure 1. Mass reactions would presumably not be much influenced by the degree to which the reaction mixture is divided into smaller cells. From Figure 1 we can also conclude that if a diffusion process in the cell is rate-limiting, this must be a process involving very large molecules, for small molecules, like sugar, become equally divided within a cell by diffusion terms of seconds. If, on the other hand, food distribution in the body nan depended on diffusion, it would take a lifetime to get sugar fed stomach to diffuse into the feet and hands, and it would be

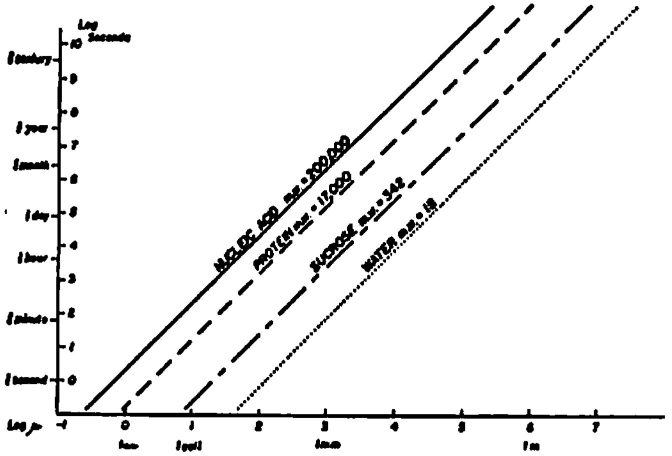


FIG. 1. Relationship between diffusion time (ordinate) and distance where half saturation is reached (abscissa) for a number of substances with different molecular weights. Over short distances diffusion is very fast, whereas over long distances diffusion is fantastically slow. Data based on calculations by Dr. K. Sevier.

possible to take only a few steps in a lifetime. This shows that, depending on size, different processes are differentially controlled. At the size of a cell, mixing of even larger molecules is almost instant, and protoplasmic streaming would, for instance, not speed up the equalization of sugar distribution in a cell of about 0.1 mm in diameter. On the other hand, streaming and mass flow become almost imperative to get food distributed in an organism, when it is more than a millimeter in size.

In this connection it is interesting to note that, almost without exception, active animals of more than 1 mm size have a heart or other organ for circulation of substances through the body: above 1 mm size hydraulic flow has to substitute for food distribution by diffusion. And a major reason why insects are all small when compared with the higher animals, is that they depend for oxygen supply to their tissues mainly upon diffusion through tracheae. Some of the larger insects do have ventilation in their tracheae and do not get oxygen only by diffusion. For any larger sized animal, lungs or gills with active ventilation are essential. The exo-skeleton of insects would make lung ventilation very difficult. Arthropods with gills past which they pump water have become considerably larger (crabs, lobsters).

TABLE 1

ACTIVITIES AND ACCOMPLISHMENTS OF ANIMALS COMPARABLE WITH THOSE OF MAN

Tools:	California Sea Otter carries a stone as anvil to crack shellfish
	Birds swallow gravel to grind food in stomach
	Lobsters use sand as otoliths in antennae
	Ant lion pelts victims with sand to let them slide into his sand trap
	The fish, Toxotes jaculator, aims water-drops at insects above the water which then drop into water and are eaten
	Monkeys use sticks to reach fruit
Structures:	Nests (of birds, bees, wasps, ants, termites, Weaver birds)
	Webs and trapdoors of spiders
	Hiding structure of Octopus made with stones
	Container made from sand grains by the protist Difflugia
Cultivation:	Fungi (leaf cutter ant)
	Flowering plants (epiphytic ant gardens)
	Aphids (ants)
Communications:	Sound (mammals and birds)
	Smell (ants)
	Touch (ants)
	Vision: Dance (bees)
	Radar (bats)
Societies:	Ants, bees, termites
	Monkeys
	Beavers
	Weaver birds
	Ground hogs
Food Storage:	Squirrels and many other rodents
	Ants, bees, termites
Air Conditioning:	Ventilation of Atta nests
	Burrows of desert animals

The emergence of a man as a thinking creature with a full-fledged technology, from among the tens of millions of animals which have lived and still are living on earth, has been attributed to a variety of factors, anatomical and mechanical. A favorite among these is his brain volume; others are his erect posture and the development of hands. Then again his use of fire is considered the main factor which differentiates man from beast. I believe that as good a case (or perhaps even a better one) can be made by considering man's physical size as the critical factor which made it possible for him to develop a technology and to use fire.

In most respects there are no differences in principle between man and animals. We will not consider any actions for which the animal body shows special adaptation, or any instinct actions. The accompanying Table 1 shows that many animals have learned the use of tools, make elaborate structures (usually nests), cultivate plants and animals, have means of communicating with each other, develop societies (with hierarchies of authority and division of labor), store food, and make use of air-conditioning.

In most of these respects, however, there are quantitative differences between man and the animals. Man's tools are infinitely more refined, his structures more sophisticated, his communication system incomparably more complex, his food storage and air-conditioning of a completely different order of magnitude. But there are also qualitative differences between man and all other animals. Among these I want to stress:

(1) the use of fire
(2) the capture and use of animal and mechanical energy (horsepower, wind and water power, the steam and internal combustion engines, atomic energy)
(3) the use of the wheel
(4) the production and use, first of stone implements, and later of metals
(5) the development of script
(6) the storage of information, available for education and development.

When we analyze these six functions, we find that each one of them could not have been developed or used by ants or other small animals, no matter how far advanced their society might have become. This is a result of the principles of critical mass and critical size.

Just as an example of what is meant by the latter statement, let us consider for a moment the dimensional limitations of fire. A flame cannot be smaller than several millimeters length (and even then is relatively unstable), and requires a volatile combustible material, such as gas or alcohol. Nonvolatile combustible material such as wood or coal has a much larger critical mass for combustion. The reason for the lower limit of flame size is that the ignition point of gases and vapors lies rather high, usually many hundreds of degrees centigrade. When the temperature of the flame decreases below the ignition point, the flame extinguishes. Since the flame has to be maintained with the oxygen of the air, and this air is cold, the flame must be large enough to heat the inrushing air to the ignition point, and provide enough heat to volatilize the fuel to be burned. When there is wind, the critical flame size increases because of the strong cooling by the inrushing air. Then candles are impractical, unless shielded from the wind, and torches are required with flames of sufficient size so that the air brought in by the wind can be heated to the ignition point of the burning material.

The critical size of a wood or coal fire is considerably greater than for a flame, because not only the air, but also the basic mass producing the combustible materials has to be heated to the point that either volatile materials are liberated or that the ignition point of the solid material is reached. At that temperature so much heat is radiated that such a fire can only be maintained if the burning piece of wood or coal is surrounded by other burning pieces. Interestingly enough, a wood or coal fire above the critical size produces just the right amount of heat to warm man in a cave, or a room, or a camping site. But ants or small rodents would have to keep too far away to make a fire economical, or rather, they would be unable to bring up enough wood to keep the fire going. Therefore in an ant society fire is not an economical possibility, and they have developed without its benefits, by operating only while outside temperatures are within the physiological range. Man on the other hand has been able to move into very cold areas by using fire.

Let us consider another significant difference between the world of the ant and the world man lives in. This is also a consequence of dimension. Mass and weight decrease with the third power of linear size. The weight of a body in the macro-world of the earth crust has practically nothing to do with its position in relation to other objects since the gravitational pull of the earth completely overshadows all other forces such as the gravitational effects of the surrounding objects.

But as the dimensions decrease another force becomes important, and this is molecular attraction or cohesion, also manifested in absorption phenomena. In Figure 2 we see how gravitational attraction increases with the third power of the linear dimension of an object. Molecular attraction on the other hand increases with the second power of the linear dimension, and decreases with the square root of the distance between the surfaces. Since with a decrease in size there generally is also a decrease in effective distance between bodies in physical contact, the force with which touching bodies are attracted one to another increases approximately with the linear size whereas mass is increasing with the third power of the diameter. This means that a 1 mm^3 cube with a reasonably flat surface (such as a wooden cube cut with a razor blade) does not drop off but can be supported by adhesion to another flat surface. In this case the gravitational force is barely balanced by adhesion. On the other hand, for a 1 cm³ block, gravity is fifty times greater than adhesion, whereas for a cube of a 0.1 mm rib size adhesion is 50 times stronger than gravity. At still smaller sizes, gravity loses all significance in relation to molecular attraction, regardless of the shape of the body, and particles of dust will stick anywhere to any surface. Basically, electrostatic attraction is similar to molecular attraction, since it is proportional to the surface of the charged bodies and inversely proportional to the square of their distance. The absolute value of electrostatic attraction is a function of the charge and is potentially more than a thousandfold greater than molecular attraction. But when bodies are in contact with each other, they lose their charge, and only molecular attraction remains.

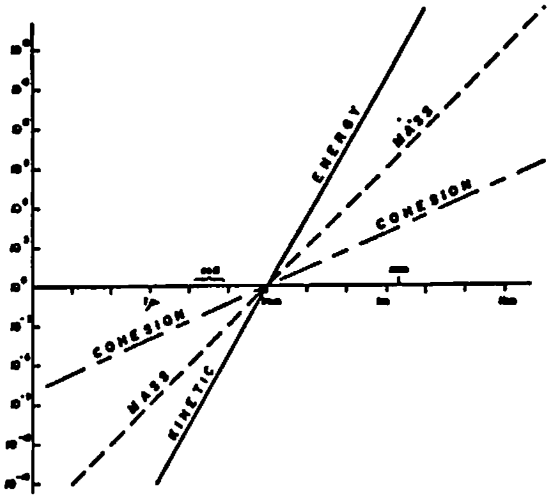


FIG. 2. Dependence of (1) cohesion between surfaces, (2) mass (weight), and (3) kinetic energy (relative values indicated on ordinate) upon linear dimension (abscissa). Both cohesion and kinetic energy values are only approximate, based on average roughness of surface and on average velocities of masses of various sizes.

The foregoing considerations make it clear why man cannot walk up against a vertical surface since at his size gravitational attraction exceeds molecular adhesion more than a millionfold. A fly, on the other hand, with its flat sole surfaces (not suction cups) is just in the size range, where Figure 2 shows that gravity and adhesion are of the same order of magnitude. Therefore it can walk against surfaces in any position, even upside down against the ceiling. Consequently we find a complete change in the physical world of organisms, between the size where gravitational forces exceed molecular forces, and the size where molecular forces predominate. I propose to name these two worlds: The macro-world, or Newtonian world where gravitational forces hold sway; and the micro-world or molecular world, where not classical mechanics but thermodynamic and statistical laws are dominant. A good criterion as to whether a process belongs in the classical, mechanical, or thermodynamical category is the degree to which it is dependent upon temperature. In general, we can say that gravitational attraction, or the velocity of an object is largely independent of temperature. Chemical reactions, or osmotic water movement, on the other hand, depend on the activity or specific temperature of the participating molecules.

Considerations about kinetic energy show us another fundamental difference between the macro-world of man and the micro-world of insects and smaller creatures. The numerical values of kinetic energy actually give us a good clue as to the optimal size of man. A 2 m tall man, when tripping, will have a kinetic energy upon hitting the ground 20–100 times greater than a small child who learns to walk. This explains why it is safe for a child to learn to walk; whereas adults occasionally break a bone when tripping, children never do. If a man were twice as tall as he is now, his kinetic energy in falling would be so great (32 times more than at normal size) that it would not be safe for him to walk upright. Consequently man is the tallest creature which could reasonably walk upright on two legs. The larger mammals can become taller, because they are more stable on their four legs. Yet, they break bones more easily when they fall. These are all cases indicating the dangers involved in living in the macro-world, and the limitations it imposes on the macro-animals. The micro-animals, such as insects and micro-organisms live in a totally different world, unattended by the forces (and dangers) of gravity and kinetic energy. However, they depend upon the equally exacting forces of thermodynamics. Perhaps we should call these Gibbsian forces in contradistinction to Newtonian forces because of the emphasis Gibbs placed on the forces and laws of thermodynamics.

From these considerations it seems clear why man could not be much larger than he is, nor could an animal of the size of an ant have conquered the world with a technology. But the question remains open, why could not an animal of rat or rabbit size have come to dominate the world? At that size they could have developed fire, as the amount of wood collected by beavers attests. And from that, every other technological development might conceivably have sprung as it did in the case of man. However, there are other primitive activities for which man-size is required, and without which man could not have reached and crossed the brink of technology.

The earliest tools and weapons of man all are limited in their effectiveness by size. A club of a size to kill a larger animal could not be swung effectively by a small child. Similarly the usability of spears and arrows is limited by size, which determines their momentum. If we assume that a spear or a club of a size proportionate to body size is handled by a 7 ft giant, its impact would be 4 times greater than when handled by an ordinary 5 ft 8 in. man, whereas the wallop of a 5 ft man would be just half. But compared with an ordinary man, the blow of a 3-year old child or a 3 ft creature in general could produce only one twenty-fifth of the energy, utterly insufficient to kill prey or hunt larger animals.

The same differences would be found when comparing the output of an axe-blow or mining chisel. A 3-ft man could neither cut lumber nor excavate a mine in solid rock. And here we find one of the more basic reasons for the size of man. Man has adjusted his activities to his particular size; this happened to be sufficient for exploiting fire, hunting larger animals, cutting and splitting wood, and mining minerals. Now that with technology man can produce energy many fold what nature has endowed him with, he can move mountains and dig canals with steam shovel and dynamite, he can build palaces and skyscrapers with rivet and crane, and multiply his speed 10 to 100 times by car and airplane. But man himself, was the prime mover in all this development, for which nature provided his body structure and specifically his size.

Before the microscope came into general use, scientists studied the interrelationships and interactions of large objects. Started by the Greeks and powerfully advanced by Galileo, classical mechanics developed a large body of knowledge, which received its theoretical basis through Newton. He recognized in gravity a force by which bodies could act on each other at a distance; gravitational force increases with the mass of a body. Gravity is the basis for classical mechanics, it is the framework for the understanding of what we see of the world around us, the macro-world. Mass movement such as hydraulic flow or wind all can be approached by classical mechanics. Especially the behavior of our solar system provides a spectacular and triumphal confirmation of Newton's basic ideas.

Yet there were a number of phenomena, usually exhibited by small bodies, which indicated that forces other than gravity and kinetic energy played a role in our physical world. There were in the first place the chemical forces exhibited by the elements, and there was the phenomenon of life which seemed to be outside the realm of known physical phenomena. But there were physical facts, too, which required explanation on a basis other than classical mechanics, such as expansion by heating, absorption, and Brownian movement. This all heralded the molecular theory which now holds sway in the domain of exceedingly small dimensions. The Newtonian world was one of large bodies in which individual molecules and the forces between them vanished into insignificance. In this macro-world the main forces are gravity and kinetic energy; they are proportional to the mass of an object, and its chemical or molecular structure is of no consequence. Since it is this macro-world we see with our eyes, feel with our fingers, and measure with our yardsticks and scales, it is this world of which early scientists discovered the laws, which we then used to interpret living organisms. Therefore, our Classical Biology describes organisms and their actions and functions in terms of classical mechanics. If we make up the balance of what biologists have accomplished working just in the Newtonian macro-world, the list is impressive, and includes the discovery and analysis of such principles as evolution, development, growth, differentiation; it comprises knowledge about water translocation and blood circulation in organisms, the mechanics of movement and response; a good part of the living world has been described, measured, and brought under human control; malfunctioning has been related to parasitism and mechanical and chemical interference, etc.

In addition to all this, a new field is opening up in biology; it is situated in the micro- or molecular world. Here molecular forces hold sway, acting over very short distances and reaching maximum values between adjacent molecules. Gravitational forces or kinetic energy disappear as controlling factors because they decrease so rapidly with decreasing mass. This field, appropriately named Molecular Biology, uses entirely different techniques and solves different problems; it is supplementary to Classical Biology, but cannot possibly replace it.

It is most interesting that man appears on the threshold between these worlds. With his body as a whole, with most of his actions, he stands and lives in the macro-world. An ant, on the other hand, lives mostly in the micro-world, where gravity and kinetic energy are of secondary importance. Only through some of its most ambitious structures, the ant elevates himself into the macro-world. Therefore the ant did not develop a technology reaching into the world of Newtonian forces. A wheel, for instance, is possible only in the macro-world, and so are ropes and gears. Fire, as explained earlier, is possible only with a sufficiently large mass of combustible material which happens to be just correct for use by agents or devices on a scale of human dimensions. Even the smaller mammals would be unable to use fire effectively. And books are possible in a macro-world only. Even if gold leaf could be used to make books which would decrease their thickness a thousandfold, adhesion would keep the gold leaves together, making it impossible to turn a leaf. And decreasing the letters a thousandfold would bring them to the limits of visibility since visible light cannot resolve shapes under 1μ size. If ants wanted mechanically to record information they would have to use tablets like the Assyrians. Yet, they would be unable to chisel characters into stone surfaces because, at their size, hammers are ineffective, packing hardly any kinetic energy.

Although most of the differences between the macro- and micro-worlds are matters of common knowledge, the implications of these differences are often forgotten and thus an amazing number of incorrect interpretations are suggested by magnification of the micro-world into the dimensions of the macro-world. Let me enumerate some of these differences.

In the macro-world we can weigh objects by measuring the gravitational pull on them. It is easy to compare weights in the pound or kilogram class, but when we come below the milligram, weight loses its significance, for molecular attraction so much exceeds gravitational pull that weight is an abstraction rather than a reality. Then we prefer to speak in molecular equivalents.

The same thing holds for length. In the macro-world a meter has a very concrete meaning; it is a definite and fixed distance to be measured. But as we reduce size, the wave length of the used radiation starts to limit the accuracy of our measurements. At a still smaller scale, thermal motion of molecules and molecular aggregates make distance a statistical average; it is no longer a reality.

Our time scale is equally precise in both worlds but the dimensions are entirely different. In the macro-world we are dealing with seconds to millions of light years; in the micro-world we frequently are confronted with milli- or microseconds or less.

When we are considering volume in the micro-world, gas bubbles cannot exist any more. A gas bubble of 1μ diameter in water can exist only when the gas is under pressure of three atmospheres; a bubble of 0.1μ is physically almost impossible because of the extreme surface tension of the water causing a pressure of 30 atm inside the bubble. This results in the well-known property of cohesion of water, preventing it from pulling away from a thoroughly soaked hydrophilic surface such as the wall of a tracheid or vessel in wood.

There has been much discussion during the last years whether water cohesion could actually account for the enormous negative tensions necessary to move water into the highest branches of the tallest trees. In physical models it never was possible to reach through cohesion negative tensions of several hundred atmospheres, as are claimed to exist in, for instance, drying annulus cells of fern sporangia. However, here the data of Figure 2 give the solution. In a microscopic system, the kinetic energy of the mass of water enclosed in a cell remains very small even after a sharp shock of the cell. In a larger system, such as a glass tube used in a physical model, slight vibrations can build up kinetic energies of the water in the tube a million times as great as that in individual plant cells, such as tracheids. Since this kinetic energy can temporarily exceed the cohesive force of the water, under tension, a water column breaks much sooner in a macroscopic than in a microscopic system.

This provides another striking difference between the macro- and micro-world. If a tree had a plumbing system like that of a house, consisting of fairly large and wide tubes, the cohesion of water would be insufficient to raise water to more than 10 m (30 ft) height, and without a pressure system, pushing water beyond the 10 m limit, trees could not reach any greater height. As a matter of fact, without a hydraulic water transport system, no land plants could be more than a few millimeters high nor could they be very different in form from a liverwort or a fern prothallium since osmotic and other thermodynamic systems become practically ineffective if diffusion or osmosis has to move water molecules beyond the microscopic range as seen in Figure 1. Therefore the first successful landplants, raising themselves even slightly above the surface of the wet ground, had to augment their osmotic or, more generally speaking, thermodynamic water moving system with a hydraulic system; or, in other words, they had to graduate from the micro- to the macro-world.

All these examples show the chasm which exists between macro- and micro-world, a chasm which as biologists we have to cross continuously in both directions without having been alerted to the pitfalls of the crossing. For instance, when we observe a cell under the microscope, the magic of wave-mechanics transposes the basic unit of life from the micro-world into a body with the apparent size of a pea or a football, and in our mind it becomes part of the macro-world. Its practical weightlessness, its diminutive size disappear and we think of it carrying out its functions on a macro scale. Let us consider how a sugar crystal introduced in the football-size cell would dissolve and mix through diffusion with the cell contents. It would require weeks before concentration equality would be reached. But a proportionately small sugar particle in the micro-sized cell would dissolve and reach diffusion equilibrium in a minute. This same cell when flaccid can take up the necessary water to become fully turgid in a matter of minutes. But a plasmolyzed football would require days to become turgid again.

Less extreme cases would involve the performance of an ant which can easily lift ten times its own weight which, transposed in human dimensions, would make it simple for milady shopping to carry her compact car off the street into the parking lot. Or a leafcutter ant cutting off a leaf segment high in a tree hundreds of times its own height will fall down with it without batting an eyelid (they do not have any), in contrast to a steeplejack a thousand feet up in a tower. We also often compare a wheat plant 1½ m high, having a stem diameter of 7 mm, with a 2 m wide tower 250 times as high (400 m, about the height of the Empire State Building). The latter example shows that we do not need to move from micro- to macro-world to make impermissible comparisons by proportioning an object from one into another dimension.

To indicate how man is continuously weaving his way between Newtonian and Gibbsian forces, let us consider his first experience in the morning: his shower. As he opens the tap, the hydraulic pressure of the water in the pipe forces it out of the nozzle. There its potential energy (pressure) is translated into kinetic energy, and hundreds of threads of water fight their way through the air resistance (molecular), and guided by kinetic energy, modified by gravity, the water rushes toward you. Surface tension (molecular force) tends to break the threads of water into droplets. These are then mechanically intercepted by your body. Then gravity pulls the larger droplets down; the smaller ones are held by adhesion (molecular surface forces) against your body. Soap, increasing adhesion of the water to the body, aids the solution of sweat and soluble substances from the body surface, and by emulsifying fats decreases adhesion of dirt particles. These are all phenomena on the molecular level.

When we turn the tap off, we first aid gravitational forces by centrifugal and momental ones (by shaking) to remove the larger water droplets, and then take recourse to molecular forces again by using a towel, which partitions the adhering water between the body and fiber surfaces. Since the fiber surface of a large towel is about a hundredfold that of our body, 99 per cent of the water remains in the towel. The rest evaporates rapidly and we are dry again, except for our hair. This has, like a towel, from 10 to 100 times the body surface, and therefore thermodynamics tells us that a considerable portion of the water will remain on the hair which just has to evaporate gradually. Now an ant could not possibly take a shower, or even wash himself with water. The water droplets into which a shower stream breaks up would hit him like rocks and envelop him if the chitin of his body were not water repellent. He could not breathe any more when the openings of his trachea were closed by the water droplet in which he is caught. If he tried to dry himself, he would be sucked against the towel. Therefore he has to dry-clean himself; rub dust and other particles off his head and body.

After his dry-cleaning, an ant could not possibly dress: Molecular forces (adhesion) would hold sufficiently pliable clothes tightly squeezed against his body and he would be unable to remove them again. Then he has to forego a cup of coffee for with his dimensions he could not pour a liquid. He could at most squeeze a tiny droplet onto a stone surface and suck from this droplet. Actually, living in a world in which gravity is not an effective force, he could develop only techniques like those employed by astronauts in space. A newspaper would be out of question: he never would be able to pull the pages apart, stuck together by adhesion. Besides, as explained earlier, the letters would be physically unreadable.

After finishing his imaginary breakfast, the working ant would not kiss his wife goodbye since he is a sterile female, and the children are not his (or rather hers): they are communal property of the whole colony. Lighting a cigarette on his way to work is out of the question: not only can an ant not use fire, but he has no lungs to suck in the smoke. The lack of wheels necessitates walking to the place of work and all activities of an ant colony have to occur within walking distance.

Although it is conceivable that sounds (at a very high pitch) could be made by ants (such as crickets do), they could not speak since they have no air stream from lungs to activate vocal chords. Therefore they communicate by touch and smell.

Yet the life of an ant must be a comparatively comfortable one. Take for instance the harvester ants of the desert. They dig nests in the ground over ten feet deep where the temperature never changes, and where it is pleasantly warm all year long. The humidity at such a depth is quite high so that the ants lose little water by transpiration. They do not operate in the heat of the day, but work above ground only while the temperature is comfortable. Because they collect and store large quantities of seeds in their nests, they do not need to work all day long, but can restrict their above-ground activities to the more pleasant periods of the day and live an air-conditioned life without requiring the complicated mechanical equipment man needs to live comfortably in the desert.

So when we come to think of it, a life mostly at the molecular level is nothing to be sneezed at. It does not have the comforts which a technology on the Newtonian level can provide, and it lacks the possibilities for further development through information storage. But there are enough possibilities, and I can easily imagine a medieval type of society in an ant hill, perhaps even with minstrels, who would not strum lutes, but would waft pleasant sequences of aromas (their method of communication) for the assembled workers, reigned over by their queen.

There is a moral to this perhaps too fanciful discourse. There exists a schism in Biology, between the "Classical Biologists" and the "Molecular Biologists." It is sometimes claimed that only Biology on the molecular level is able to approach the problems of life and living organisms. But I hope to have shown, that an understanding of the living world is only possible if *both* classical-mechanical *and* thermodynamic approaches are used.

Berry,L.G., Mason, B. and
Dietrich, R.V.
Mineralogy 2nd edition 1983
W.H. Freeman
Used by permission

15 Class VIII: Silicates

The silicates include about a third of all mineral species. Although many of these minerals are quite rare, others make up a large part of the earth's crust. In fact, the crust has been estimated to consist of about 95 percent silicate minerals, of which some 60 percent is feldspar and 12 percent quartz. The predominance of silicates and aluminosilicates reflects the abundance of oxygen, silicon, and aluminum, which are the most common elements in the crust (see page 179).

The number of elements that form silicate minerals is comparatively few. Some additional elements, such as Rb and Sr, do not form specific silicate minerals; instead, they are dispersed in small amounts, replacing more abundant elements in several, for the most part fairly common, silicate minerals. Still other elements that do not form common independent silicates occur associated with other cations (e.g., boron, which forms a number of borosilicates).

The great multiplicity of silicate minerals is due to the variety of silicates that can be formed from the same elements. Many have complex and variable compositions—complexities that hindered the development of a satisfactory classification of silicate minerals until X-ray investigations provided the means for determining their crystal structures.

Prior to the elucidation of their crystal structures, the composition of the silicates was generally interpreted in terms of hypothetical silicic acids, all of which were derived from a theoretical orthosilicic acid, H_4SiO_4. Some of the hypothetical silicic acids were as follows:

Orthosilicic acid	H_4SiO_4
Metasilicic acid	$H_2SiO_3 = (H_4SiO_4 - H_2O)$
Orthodisilicic acid	$H_6Si_2O_7 = (2H_4SiO_4 - H_2O)$
Metadisilicic acid	$H_2Si_2O_5 = (2H_4SiO_4 - 3H_2O)$
Trisilicic acid	$H_4Si_3O_8 = (3H_4SiO_4 - 4H_2O)$

Thence, ratios of silicon to oxygen could be derived to fit any composition. The silicic acid basis had some success in the interpretation of some of the simpler compounds, but it led to manifest absurdities when it was applied to more complex minerals. (For example, the solid-solution end-members albite, $NaAlSi_3O_8$, and anorthite, $CaAl_2Si_2O_8$, fell into different silicate groups.) Consequently, the silicic acid classification of the silicate minerals has been superseded by a classification, outlined below, that is based on crystal structure.

THE STRUCTURE AND CLASSIFICATION OF THE SILICATES

In nearly all silicate structures, the silicon atoms are in fourfold coordination with oxygen. The bonds between silicon and oxygen are so strong that the four oxygens are always at the corners of a tetrahedron of nearly constant dimensions and regular shape, whatever the configuration of the rest of the structure (Figure 15-1). Most of the common silicate minerals can be classified as belonging to one of the different silicate types (see Table 15-1 and Figures 15-2 through 15-5); as shown, the tetrahedra may exist as separate and distinct units or they may be linked by sharing their corners (i.e., their oxygens). Silicate classification for most silicate minerals is based on the types of linkages as follows:

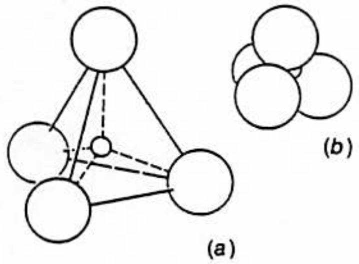
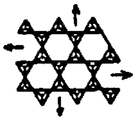


Figure 15-1. The silica tetrahedron. (a) Expanded view showing relatively large oxygen ions at corners and a small silicon ion at the center of the tetrahedron; short dashed lines represent bonds between silicon and surrounding oxygen ions. (b) Tetrahedron with correct spatial arrangement, dashed circle indicates central silicon ion.

1. *Independent tetrahedral groups:* In this class, the silicon-oxygen tetrahedra are present as separate entities (Figure 15-2*a*). The resultant composition is SiO_4, which gives such minerals as forsterite, Mg_2SiO_4. This division of the silicates, with the Si:O ratio of 1:4, is known as the *nesosilicates*, from the Greek root meaning island.

2. *Double tetrahedra structures:* Two silicon-oxygen tetrahedra are linked by the sharing of one oxygen between them; the resulting composition is Si_2O_7, with an Si:O ratio of 2:7 (Figure 15-2*b*); a representative mineral is hemimorphite, $Zn_4Si_2O_7(OH)_2 \cdot H_2O$. Such silicates are known as *sorosilicates*, from the Greek root meaning group.

Table 15-1. Structural classification of the silicates

Classification	Structural Arrangement[a]	Formula of Complex Anion	Si:O	Example
Nesosilicates	Independent tetrahedra 	$(SiO_4)^{-4}$	1:4	Forsterite, $Mg_2(SiO_4)$
Sorosilicates	Two tetrahedra sharing one oxygen 	$(Si_2O_7)^{-6}$	2:7	Hemimorphite, $Zn_4(Si_2O_7)(OH)_2 \cdot H_2O$
Cyclosilicates	Closed rings of tetrahedra, each sharing two oxygens			
		$(Si_3O_9)^{-6}$	1:3	Benitoite, $BaTi(Si_3O_9)$
		$(Si_4O_{12})^{-6}$		Axinite, $Ca_3Al_2(BO_3)(Si_4O_{12})(OH)$
		$(Si_6O_{18})^{-6}$		Beryl, $Be_3Al_2(Si_6O_{18})$
Inosilicates	Continuous chains[b] of tetrahedra		1:3 and 4:11	
	Single chain, each tetrahedron sharing two oxygens 	$(SiO_3)_m^{-2}$		Enstatite, $Mg_2(SiO_3)_2$
	Double chain, alternate tetrahedra sharing two and three oxygens 	$(Si_4O_{11})_m^{-6}$		Anthophyllite, $Mg_7(Si_4O_{11})_2(OH)_2$
Phyllosilicates	Continuous sheets of tetrahedra, each sharing three oxygens 	$(Si_2O_5)_m^{-2}$	2:5	Kaolinite, $Al_4(Si_2O_5)_2(OH)_8$
Tektosilicates	Three-dimensional framework of tetrahedra, each sharing all four oxyens (see Figure 15-5)	$(SiO_2)^0$	1:2	Quartz, SiO_2

---

[a] By sharing an oxygen, silica tetrahedra can polymerize to form larger, complex anions. In the diagrams in this table, oxygens and silicons are not shown as such; an oxygen would be at each apex and a silicon would be at the center of each tetrahedron. The complex anionic groups are drawn as viewed from above; solid lines are in or above, whereas broken lines are below, the plane of the paper.

[b] There is also a triple chain arrangement, giving $(Si_3O_8)_n^{-4}$, that is fundamental to a few macroscopically indistinguishable minerals—e.g., chesterite and jimthompsonite.

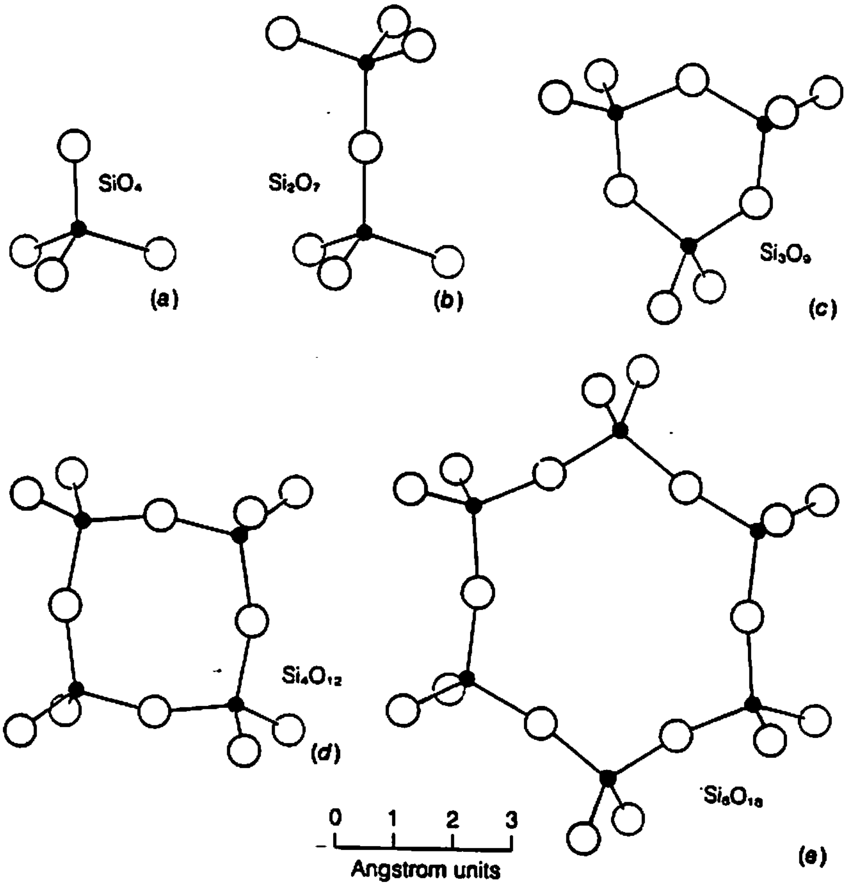


Figure 15-2. Types of linkage of silicon-oxygen tetrahedra. (a) Independent tetrahedra. (*b*) Double tetrahedra. (c), (*d*), and (e) Ring structures.

3. *Ring structures:* In this class, two of the oxygens of each tetrahedron are shared with neighboring tetrahedra, and the angular positions of the tetrahedra are such that closed units of a ringlike structure result. Rings of three, four, and six tetrahedra are known (Figures 15-2c, *d*, and e). Typical examples, each with an Si:O ratio of 1:3, are benitoite, $BaTiSi_3O_9$, with three linked tetrahedra; axinite, $(Ca,Mn,Fe)_3Al_2(BO_3)Si_4O_{12}(OH)$, with four; and beryl, $Be_3Al_2Si_6O_{18}$, with six. This division of the silicates is known as the *cyclosilicates* (Greek *cyclo*, ring).

4. *Chain structures:* In chain structures, tetrahedra are joined together to produce chains of indefinite extent. There are two principal modifications of this structure yielding somewhat different compositions: (a) single chains, in which the Si:O ratio is 1:3 (Figure 15-3a), as characterized by the pyroxenes and pyroxenoids; and (b) double chains, in which alternate tetrahedra in two parallel single chains are cross-linked and the Si:O ratio is 4:11 (Figure 15-3*b*), as characterized by the amphiboles. The chains may be indefinite in extent, are generally elongated in the c direction of the crystal, and

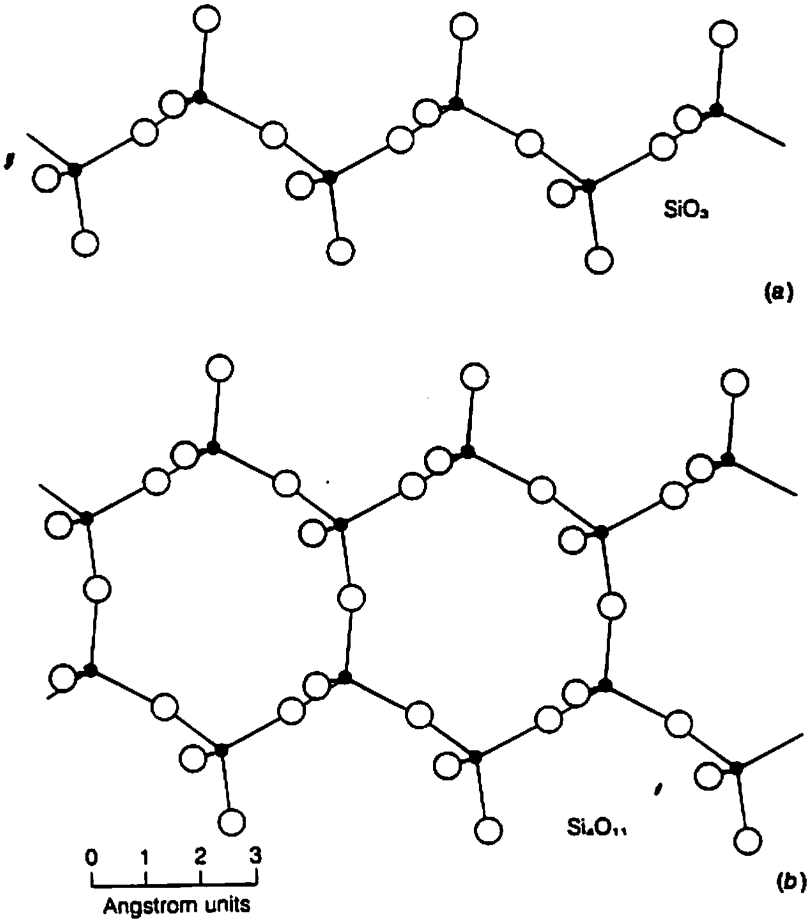


Figure 15-3. Types of linkage of silicon-oxygen tetrahedra. (a) Single chains. (b) Double chains.

are bonded to each other by the metallic elements. This division of the silicates is known as *inosilicates* (Greek *ino*, thread or chain).

5. *Sheet structures:* In sheet structures, three oxygens of each tetrahedron are shared with adjacent tetrahedra to form flat sheets of indefinite extent. In essence, this structure is the double-chain inosilicate structure extended indefinitely in two directions instead of only one. The linkage, which gives an Si:O ratio of 2:5 (Figure 15-4), is the fundamental unit in all mica and clay structures. The sheets form a hexagonal planar network that is responsible for the principal characteristics of minerals of this type—their pronounced pseudo-hexagonal habit and their perfect basal cleavage parallel to the planes of the sheets. This division of the silicates is known as *phyllosilicates* (Greek *phyllo*, leaf or sheet).

6. *Three-dimensional networks:* In three-dimensional networks, every SiO_4 tetrahedron shares all of its corners with other tetrahedra,

giving a three-dimensional network in which the Si:O ratio is 1:2 (Figure 15-5). The forms of silica—quartz, tridymite, cristobalite—have this arrangement. In SiO_2, the positive and negative charges balance. In most silicates of this type, the silicon is partly replaced by aluminum, so the composition is $(Si,Al)O_2$ and the substitution of Al^{+3} for Si^{+4} requires additional positive ions to restore electrical neutrality [e.g., nepheline, $Na(AlSiO_4)$]. The feldspars and the zeolites are also examples of this division of the silicates, which is known as the *tektosilicates* (from the Greek root meaning framework).

Consequently, most of the common silicate minerals can be classified as summarized in Table 15-1. Some minerals, however, contain groupings of more than one type or belong to other types (e.g., the xonotlite $[Ca_6Si_6O_{17}(OH)_2]$ type of chain structure and the double ring structure of osumilite $[(K,Na)(Fe,Mg)_2(Al,Fe)_3(Si,Al)_{12}O_{30} \cdot H_2O]$).

The other constituents of a silicate structure—such as additional oxygen atoms, hydroxyl groups, water molecules, and cations—are arranged with the silicate groups in such a way as to produce mechanically stable and electrically neutral structures. Aluminum, after silicon the most abundant cation in the earth's crust, plays an especially noteworthy role. As we mentioned earlier, Al is stable in both fourfold and sixfold coordination. Consequently, it can replace silicon in the SiO_4 groups, and it can also have the same role as the common six-coordination cations Mg^{+2}, Fe^{+2}, Fe^{+3}, etc. In many mineral groups

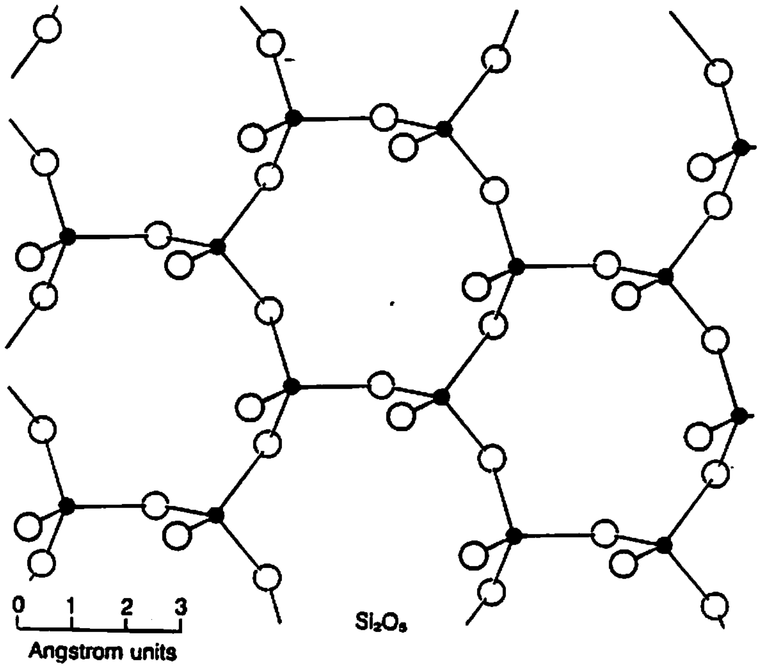


Figure 15-4. Types of linkage of silicon-oxygen tetrahedra. Sheet structure.

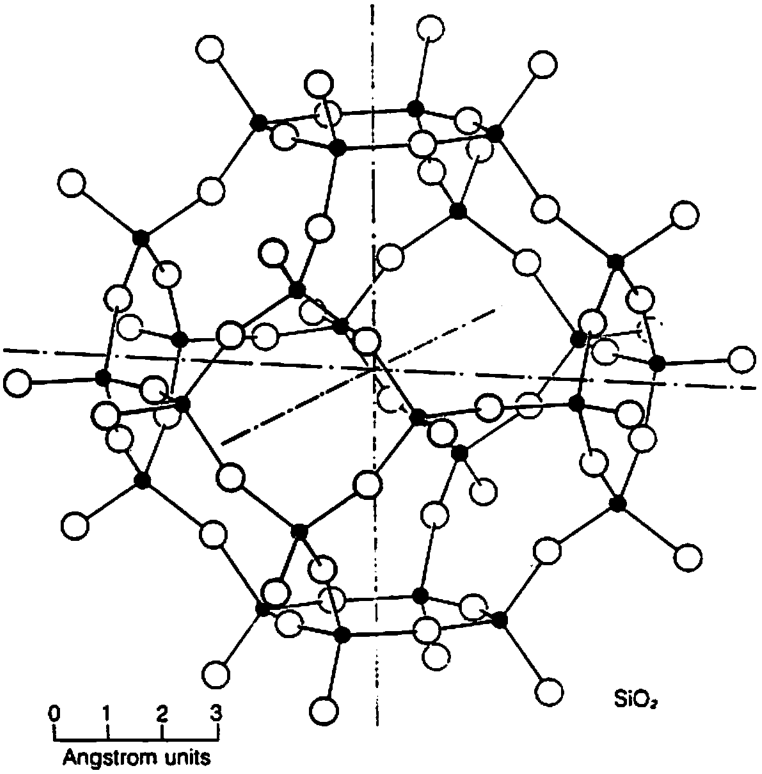


Figure 15-5. Types of linkage of silicon-oxygen tetrahedra. Three-dimensional network.

(e.g., the garnets and the feldspars), aluminum is present entirely in a single coordination; in others (e.g., the amphiboles, the pyroxenes, and the micas), it may be present in both coordinations.

The valence charge on the silicate unit, which determines the number and charge of the other ions present in a structure, can be calculated easily for any unit if you only remember that each silicon has a positive charge of 4 and that each oxygen has a negative charge of 2. Thus, the charge on a single SiO_4 unit is $[1(+4) + 4(-2) =] -4$; on an Si_2O_7 unit, $[2(+4) + 7(-2) =] -6$; on an SiO_3 unit, $[1(+4) + 3(-2) =] -2$; on an Si_4O_{11} unit, $[4(+4) + 11(-2) =] -6$; on an Si_2O_5 unit, $[2(+4) + 5(-2) =] -2$; and on SiO_2, $[1(+4) + 2(-2) =] 0$.

As might be expected, different compositions, habits, and physical properties correspond to different linkages of the tetrahedra in the various types of silicates. Several examples are rather obvious: the bonding within the Si-O framework is much stronger than the bonding between the metal cations and the framework, so the cleavage planes of the silicates are parallel to Si-O chains or sheets. The sheet structure, for instance, produces the platy form of the micas, the chlorites, the kaolins, and several other minerals. The chain structures produce prismatic or fibrous crystals, as exemplified by the pyroxenes and amphiboles. The three-dimensional network structures are commonly manifested by more-or-less equidimensional crystals, and there are characteristic density ranges and refractive index limits that correspond to differences of structure. In general, increased complexity of silicate linkage results in looser packing of the ions, which results in a trend toward lower densities and refractive indices for inosilicates than for, for example, nesosilicates, etc. (For comparable compositions, see Table 15-2.)

Table 15-2. Relationship between structure type and specific gravity in the magnesium silicate minerals

Structure Type	Mineral Example	Specific Gravity
Nesosilicate	Forsterite, Mg_2SiO_4	3.22
Inosilicate (single chain)	Enstatite, $Mg_2(SiO_3)_2$	3.18
Inosilicate (double chain)	Anthophyllite, $Mg_7(Si_4O_{11})_2(OH)_2$	2.86
Phyllosilicate	Talc, $Mg_3(Si_2O_5)_2(OH)_2$	2.82

In discussing the composition and structure of the silicate minerals, we often find it convenient to use a general formula that fits all species within a given group. Several of the general formulas in the mineral descriptions that follow this introduction make use of the following symbols:

W = large cations having coordination number greater than six; principally Ca, Na, and K.
X = medium-sized bivalent cations (and lithium) in sixfold coordination; principally Mg and Fe^{+2}.
Y = medium-sized trivalent and quadrivalent cations in sixfold coordination; these include Al, Fe^{+3}, and Ti^{+4}.
Z = small cations in four-coordination; principally Si, commonly replaced by Al (and also by B in a few minerals).

The silicate minerals that are described in detail in this book are:

Subclass Tektosilicates	
Silica Group:	
Quartz	SiO_2
Tridymite	SiO_2
Cristobalite	SiO_2
Opal	$SiO_2 \cdot nH_2O$
Feldspar Group:	WZ_4O_8
Sanidine	$KAlSi_3O_8$
Orthoclase	$KAlSi_3O_8$
Microcline	$KAlSi_3O_8$

THE
THEORY
OF THE
EARTH

Book I

Concerning the Deluge, and the Dissolution of the Earth.

CHAPTER I

THE INTRODUCTION

An Account of the whole Work; of the Extent and general Order of it.

SINCE I was first inclin'd to the Contemplation of Nature, and took pleasure to trace out the Causes of Effects, and the dependance of one thing upon another in the visible Creation, I had always, methought, a particular curiosity to look back into the first Sources and ORIGINAL of Things; and to view in my mind, so far as I was able, the Beginning and Progress of a RISING WORLD.

And after some Essays of this Nature, and, as I thought, not unsuccessful, I carried on my enquiries further, to try whether this *Rising World*, when form'd and finisht, would continue always the same; in the same form, structure, and consistency; or what changes it would successively undergo, by the continued action of the same Causes that first produc'd it; And, lastly, what would be its final Period and Consummation. This whole Series and compass of things taken together, I call'd a COURSE OF NATURE, or a SYSTEM OF NATURAL PROVIDENCE; and thought there was nothing belonging to the External World more fit or more worthy our study and meditation, nor any thing that would conduce more to discover the ways of Divine Providence, and to show us the grounds of all true knowledge concerning Nature. And therefore to clear up the several parts of this Theory, I was willing to lay aside a great many other Speculations, and all those dry subtleties with which the Schools, and the Books of Philosophers, are usually fill'd.

But when we speak of a *Rising World*, and the Contemplation of it, we do not mean this of the *Great Universe*; for who can describe the Original of that? But we speak of the *Sublunary World*, This Earth and its dependencies, which rose out of a Chaos about six thousand years ago; And seeing it hath faln to our lot to act upon this Stage, to have our present home and residence here, it seems most reasonable, and the place design'd by Providence, where we should first imploy our thoughts to understand the works of God and Nature. We have accordingly therefore design'd in this Work to give an account of the Original of the Earth, and of all the great and general changes that it hath already undergone, or is hence forwards to undergo, till the Consummation of all things. For if from those Principles we have here taken, and that Theory we have begun in these two first Books, we can deduce with success and clearness the Origin of the Earth, and those States of it that are already past; Following the same Thred, and by the conduct of the same Theory, we will pursue its Fate and History through future Ages, and mark all the great Changes and Conversions that attend it *while Day and Night shall last*; that is, so long as it continues an Earth.

By the States of the Earth that are already past, we understand chiefly *Paradise* and the *Deluge*; Names well known, and as little known in their Nature. By the Future States we understand the *Conflagration*, and what new Order of Nature may follow upon that, till the whole Circle of Time and Providence be compleated. As to the first and past States of the Earth, we shall have little help from the Ancients, or from any of the Philosophers, for the discovery or description of them; We must often tread unbeaten paths, and make a way where we do not find one; but it shall be always with a Light in our hand, that we may see our steps, and that those that follow us may not follow us blindly. There is no Sect of Philosophers that I know of that ever gave an account of the Universal Deluge, or discover'd, from the contemplation of the Earth, that there had been such a thing already in Nature. 'Tis true, they often talk of an alternation of *Deluges* and *Conflagrations* in this Earth, but they speak of them as things to come; at least they give no proof or argument of any that hath already destroyed the World. As to *Paradise*, it seems to be represented to us by the *Golden Age*; whereof the Ancients tell many stories, sometimes very luxuriant, and sometimes very defective: For they did not so well understand the difference betwixt the New-made Earth and the Present, as to see what were the just grounds of the Golden Age, or of Paradise: Tho' they had many broken Notions concerning those things. As to the *Conflagration* in particular, This hath always been reckon'd One amongst the Opinions or Dogmata of the Stoicks, *That the World was to be destroy'd by Fire*, and their Books are full of this Notion; but yet they do not tell us the Causes of the Conflagration, nor what preparations there are in Nature, or will be, towards that great Change. And we may generally observe this of the *Ancients*, that their Learning or Philosophy consisted more in Conclusions, than in Demonstrations; They had many truths among them, whereof they did not know themselves the premisses or the proofs: Which is an argument to me, that the knowledge they had, was not a thing of their own invention, or which they came to by fair reasoning and observations upon Nature, but was delivered to them from others by Tradition and Ancient fame, sometimes more publick, sometimes more secret: These Conclusions they kept in mind, and communicated to those of their School, or Sect, or Posterity, without knowing, for the most part, the just grounds and reasons of them.

'Tis the Sacred writings of Scripture that are the best monuments of Antiquity, and to those we are chiefly beholden for the History of the first Ages, whether Natural History or Civil. 'Tis true, the Poets, who were the most ancient Writers amongst the *Greeks*, and serv'd them both for Historians, Divines, and Philosophers, have deliver'd some things concerning the first Ages of the World, that have a fair resemblance of truth, and some affinity with those accounts that are given of the same things by sacred Authors, and these may be of use in due time and place; but yet, lest any thing fabulous should be mixt with them, as commonly there is, we will never depend wholly upon their credit, nor assert any thing upon the authority of the Ancients which is not first prov'd by natural Reason, or warranted by Scripture.

It seems to me very reasonable to believe, that besides the precepts of Religion, which are the principal subject and design of the Books of holy Scripture, there may be providentially conserv'd in them the memory of things and times so remote, as could not be retriev'd, either by History, or by the light of Nature; and yet were of great importance to be known, both for their own excellency, and also to rectifie the knowledge of men in other things consequential to them: Such points may be, *Our great Epocha* or the Age of the Earth, The Origination of mankind, The first and Paradisiacal state, The destruction of the Old World by an universal Deluge, The longevity of its inhabitants, The manner of their preservation, and of their peopling the Second Earth; and lastly, The Fate and Changes it is to undergo. These I always lookt upon as the Seeds of great knowledge, or heads of Theories fixt on purpose to give us aim and direction how to pursue the rest that depend upon them. But these heads, you see, are of a mixt order, and we propose to our selves in this Work only such as belong to the Natural World; upon which I believe the trains of Providence are generally laid; And we must first consider how God hath order'd Nature, and then now the Oeconomy of the Intellectual World is adapted to it; for of these two parts consists the full System of Providence. In the mean time, what subject can be more worthy the thoughts of any serious person, than to view and consider the Rise and Fall, and all the Revolutions, not of a Monarchy or an Empire, of the *Grecian* or *Roman* State, but of an intire World.

The obscurity of these things, and their remoteness from common knowledge will be made an argument by some, why we should not undertake them; And by others, it may be, the very same thing will be made an argument why we should; for my part I think *There is nothing so secret that shall not be brought to Light*, within the compass of *Our World*; for we are not to understand that of the whole Universe, nor of all Eternity, our capacities do not extend so far; But whatsoever concerns this Sublunary World in the whole extent of its duration, from the Chaos to the last period, this I believe Providence hath made us capable to understand, and will in its due time make it known. All I say, betwixt the first Chaos and the last Completion of Time and all things temporary, This was given to the disquisitions of men; On either hand is Eternity, before the World and after, which is without our reach: But that little spot of ground that lies betwixt those two great Oceans, this we are to cultivate, this we are Masters of, herein we are to exercise our thoughts, to understand and lay open the treasures of the Divine Wisdom and Goodness hid in this part of Nature and of Providence.

As for the difficulty or obscurity of an argument, that does but add to the pleasure of contesting with it, when there are hopes of victory; and success does more than recompence all the pains. For there is no sort of joy more grateful to the mind of man, than that which ariseth from the invention of Truth; especially when 'tis hard to come by. Every man hath a delight suited to his Genius, and as there is pleasure in the right exercise of any faculty, so especially in that of Right-reasoning; which is still the greater, by how much the consequences are more clear, and the chains of them more long: There is no Chase so pleasant, methinks, as to drive a Thought, by good conduct, from one end of the World to the other; and never to lose sight of it till it fall into Eternity, where all things are lost as to our knowledge.

This Theory being chiefly Philosophical, Reason is to be our first Guide; and where that falls short, or any other just occasion offers itself, we may receive further light and confirmation from the Sacred writings. Both these are to be lookt upon as of Divine Original, God is the Author of both; He that made the Scripture made also our Faculties, and 'twere a reflection upon the Divine Veracity, for the one or the other to be false when rightly us'd. We must therefore be careful and tender of opposing these to one another, because that is, in effect, to oppose God to himself. As for Antiquity and the Testimonies of the Ancients, we only make general reflections upon them, for illustration rather than proof of what we propose; not thinking it proper for an English Treatise to multiply citations out of Greek or Latin Authors.

I am very sensible it will be much our interest, that the Reader of this Theory should be of an ingenuous and unprejudic'd temper; neither does it so much require Book-learning and Scholarship, as good natural sence to distinguish *True* and *False*, and to discern what is well prov'd, and what is not. It often happens that Scholastick Education, like a Trade, does so fix a man in a particular way, that he is not fit to judge of any thing that lies out of that way; and so his Learning becomes a clog to his natural parts, and makes him more indocile, and more incapable of new thoughts and new improvements, than those that have only the Talents of Nature. As Masters of exercise had rather take a Scholar that never learn'd before, than one that hath had a bad Master; so generally one would rather chuse a Reader without art, than one ill-instructed; with learning, but opinionative and without judgment: yet it is not necessary they should want either, and Learning well plac'd strengthens all the powers of the mind. To conclude, just reasoning and a generous love of Truth, whether with or without Erudition, is that which makes us most competent Judges what is true; and further than this, in the perusal and examination of this Work, as to the Author as much candor as you please, but as to the Theory we require nothing but attention and impartiality.

CHAPTER II

A general account of Noah's *Flood; A computation what quantity of Water would be necessary for the making of it; that the common Opinion and Explication of that Flood is not intelligible.*

'TIS now more than Five Thousand years since our World was made, and though it would be a great pleasure to the mind, to recollect and view at this distance those first Scenes of Nature: what the face of the Earth was when fresh and new, and how things differ'd from the state we now find them in, the speculation is so remote, that it seems to be hopeless, and beyond the reach of Humane Wit. We are almost the last Posterity of the first Men, and faln into the dying Age of the World; by what footsteps, or by what guide, can we trace back our way to those first Ages, and the first order of things? And yet, methinks, it is reasonable to believe, that Divine Providence, which sees at once throughout all the Ages of the World, should not be willing to keep Mankind finally and fatally ignorant of that part of Nature, and of the Universe, which is properly their Task and Province to manage and understand. We are the Inhabitants of the Earth, the Lords and Masters of it; and we are endow'd with Reason and Understanding; doth it not then properly belong to us to examine and unfold the works of God in this part of the Universe, which is faln to our lot, which is our heritage and habitation? And it will be found, it may be, upon a stricter Enquiry, that in the present form and constitution of the Earth, there are certain marks and Indications of its first State; with which if we compare those things that are recorded in Sacred History, concerning the first Chaos, Paradise, and an universal Deluge, we may discover, by the help of those Lights, what the Earth was in its first Original, and what Changes have since succeeded in it.

And though we shall give a full account of the Origin of the Earth in this Treatise, yet that which we have propos'd particularly for the Title and Subject of it, is to give an account of the primæval PARADISE, and of the universal DELUGE: Those being the two most important things that are explain'd by the Theory we propose. And I must beg leave in treating of these two, to change the order, and treat first of the Deluge, and then of Paradise: For though the State of Paradise doth precede that of the Flood in Sacred History, and in the nature of the thing, yet the explication of both will be more sensible, and more effectual, if we begin with the Deluge; there being more Observations and Effects, and those better known to us, that may be refer'd to this, than to the other; and the Deluge being once truly explain'd, we shall from thence know the form and Quality of the Ante-diluvian Earth. Let us then proceed to the explication of that great and fatal Inundation, whose History is well known; and according to *Moses*, the best of Historians, in a few words is this—

Sixteen Hundred and odd years after the Earth was made, and inhabited, it was over-flow'd, and destroy'd in a Deluge of water. Not a Deluge that was National only, or over-run some particular Country or Region, as *Judea* or *Greece*, or any other, but it overspread the face of the whole Earth, from Pole to Pole, and from East to West, and that in such excess, that the Floods over-reacht the Tops of the highest Mountains; the Rains descending after an unusual manner, and the fountains of the *Great Deep* being broke open; so as a general destruction and devastation was brought upon the Earth, and all things in it, Mankind and other living Creatures; excepting only *Noah* and his Family, who by a special Providence of God were preserv'd in a certain Ark, or Vessel made like a Ship, and such kinds of living Creatures as he took in to him. After these waters had rag'd for some time on the Earth, they began to lessen and shrink, and the great waves and fluctuations of this *Deep* or *Abysse*, being quieted by degrees, the waters retir'd into their Chanels and Caverns within the Earth; and the Mountains and Fields began to appear, and the whole habitable Earth in that form and shape wherein we now see it. Then the World began again, and from that little Remnant preserv'd in the Ark, the present race of Mankind, and of Animals, in the known parts of the Earth, were propagated. Thus perisht the old World, and the present arose from the ruines and remains of it.

This is a short story of the greatest thing that ever yet hapned in the world, the greatest revolution and the greatest change in Nature; and if we come to reflect seriously upon it, we shall find it extremely difficult, if not impossible, to give an account of the waters that compos'd this Deluge, whence they came or whither they went. If it had been only the Inundation of a Country, or of a Province, or of the greatest part of a Continent, some proportionable causes perhaps might have been found out; but a Deluge overflowing the whole Earth, the whole Circuit and whole Extent of it, burying all in water, even the greatest Mountains, in any known parts of the Universe, to find water sufficient for this Effect, as it is generally explained and understood, I think is impossible. And what we may the better judge of the whole matter, let us first compute how much water would be requisite for such a Deluge, or to lay the Earth, consider'd in its present form, and the highest Mountains, under water. Then let's consider whether such a quantity of water can be had out of all the stores that we know in Nature: And from these two we will take our Ground and Rise, and begin to reflect, whether the World hath not been hitherto mistaken in the common opinion and explication of the general Deluge.

To discover how much water would be requisite to make this Deluge, we must first suppose enough to cover the plain surface of the Earth, the Fields and lower Grounds; then we must heap up so much more upon this as will reach above the tops of the highest Mountains; so as drawing a Circle over the tops of the highest Mountains quite round the Earth, suppose from Pole to Pole, and another to meet it about the middle of the Earth, all that space or capacity contain'd within these Circles is to be fill'd up with water. This I confess will make a prodigious mass of water, and it looks frightfully to the imagination; 'tis huge and great, but 'tis extravagantly so, as a great Monster: It doth not look like the work of God or Nature: However let's compute a little more particularly how much this will amount to, or how many Oceans of water would be necessary to compose this great Ocean rowling in the Air, without bounds or banks.

If all the Mountains were par'd off the Earth, and so the surface of it lay even, or in an equal convexity every where with the surface of the Sea, from this surface of the Sea let us suppose that the height of the Mountains may be a mile and an half; or that we may not seem at all to favour our own opinion or calculation, let us take a mile only for the perpendicular height of the Mountains. Let us on the other side suppose the Sea to cover half the Earth, as 'tis generally believ'd to do; and the common depth of it, taking one place with another, to be about a quarter of a mile or 250 paces. I say, taking one place with another, for though the middle Chanel of the great Ocean be far deeper, we may observe, that there is commonly a descent or declivity from the shore to the middle part of the Chanel, so that one comes by degrees into the depth of it; and those shory parts are generally but some fathoms deep. Besides, in arms of the Sea, in Straits and among Islands, there is commonly no great depth, and some places are plain shallows. So as upon a moderate computation, one place compar'd with another, we may take a quarter of a mile, or about an hundred fathoms, for the common measure of the depth of the Sea, if it were cast into a Chanel of an equal depth every where. This being suppos'd, there would need four Oceans to lie upon this Ocean, to raise it up to the top of the Mountains, or so high as the waters of the Deluge rise; then four Oceans more to lie upon the Land, that the water there might swell to the same height; which together make eight Oceans for the proportion of the water requir'd in the Deluge.

'Tis true, there would not be altogether so much water requir'd for the Land as for the Sea, to raise them to an equal height; because Mountains and Hills would fill up part of that space upon the Land, and so make less water requisite. But to compensate this, and confirm our computation, we must consider in the first place, that we have taken a much less height of the Mountains than is requisite, if we respect the Mediterraneous Mountains, or those that are at a great distance from the Sea; For their height above the surface of the Sea, computing the declivity of the Land all along from the Mountains to the Sea-side (and that there is such a declivity is manifest from the course and descent of the Rivers) is far greater than the proportion we have taken: For the height of Mountains is usually taken from the foot of them, or from the next plain, which if it be far from the Sea, we may reasonably allow as much for the declension of the Land from that place to the Sea, as for the immediate height of the Mountain; So, for instance, the Mountains of the Moon in *Africa*, whence the *Nile* flows, and after a long course falls into the Mediterranean Sea by *Egypt*, are so much higher than the surface of that Sea, first, as the Ascent of the Land is from the Sea to the foot of the Mountains, and then as the height of the Mountains is from the bottom to the top: For both these are to be computed when you measure the height of a Mountain, or of a mountainous Land, in respect of the Sea: And the height of Mountains to the Sea being thus computed, there would be need of six or eight Oceans to raise the Sea alone as high as the highest In-land Mountains; And this is more than enough to compensate the less quantity of water that would be requisite upon the Land. Besides, we must consider the Regions of the Air upwards to be more capacious than a Region of the same thickness in or near the Earth, so as if an Ocean pour'd upon the surface of the dry Land, supposing it were all smooth, would rise to the height of half a quarter of a mile every where; the like quantity of water pour'd again at the height of the Mountains, would not have altogether the same effect, or would not there raise the mass half a quarter of a mile higher; for the surfaces of a Globe, the farther they are from their Center, are the greater; and so accordingly the Regions that belong to them. And, lastly, we must consider that there are some Countries or Valleys very low, and also many Caverns or Cavities within the Earth, all which in this case were to be first fill'd with water. These things being compar'd and estimated, we shall find that notwithstanding the room that Hills and Mountains take up on the dry Land, there would be at least eight Oceans requir'd, or a quantity of water eight times as great as the Ocean, to bring an universal Deluge upon the Earth, as that Deluge is ordinarily understood and explained.

The proportion of water for the Deluge being thus stated, the next thing to be done, is to enquire where this water is to be found; if any part of the Sublunary World will afford us so much: Eight Oceans floating in the Air, make a great bulk of water, I do not know what possible Sources to draw it from. There are the Clouds above, and the Deeps below, and in the bowels of the Earth; and these are all the stores we have for water; and *Moses* directs us to no other for the causes of the Deluge. *The Fountains* (he saith) *of the great Abysse were broken up, or burst asunder,* and the Rain descended for forty days, the *Cataracts* or *Floodgates* of Heaven being open'd. And in these two, no doubt, are contain'd the causes of the great Deluge, as according to *Moses,* so also according to reason and necessity; for our World affords no other treasures of water. Let us therefore consider how much this Rain of forty days might amount to, and how much might flow out of the Abysse, that so we may judge whether these two in conjunction would make up the Eight Oceans which we want.

As for the Rains, they would not afford us one Ocean, nor half an Ocean, nor the tenth part of an Ocean, if we may trust to the Observations made by others concerning the quantity of water that falls in Rain. *Mersennus* gives us this account of it. "It appears by our Observations, that a Cubical Vessel of Brass, "whereof we made use, is fill'd an inch and an half in half an hours time; but "because that sucks up nothing of the moisture as the Earth doth, let us take an "inch for half an hours Rain; whence it follows, that in the space of 40 days and "nights Rain, the waters in the Deluge would rise 160 feet, if the Rains were "constant and equal to ours, and that it rain'd at once throughout the face of the "whole Earth. But the Rain of the Deluge, saith he, should have been 90 times greater than this, to cover, for instance, the Mountains of *Armenia,* or to reach 15 Cubits above them. So that according to his computation, the 40 days Rain would supply little more than the hundredth part of the water requisite to make the Deluge. 'Tis true, he takes the heighth of the Mountains higher than we do; but, however, if you temper the Calculation on all sides as much as you please, the water that came by this Rain would be a very inconsiderable part of what was necessary for a Deluge. If it rain'd 40 days and 40 nights throughout the face of the whole Earth, in the Northern and Southern Hemisphere all at once, it might be sufficient to lay all the lower grounds under water, but it would signifie very little as to the overflowing of the Mountains. Whence another Author upon the same occasion hath this passage. "If the Deluge had been made by Rains only, "there would not have needed 40 days, but 40 years Rain to have brought it to "pass. And if we should suppose the whole middle Region condens'd into water, it would not at all have been sufficient for this effect, according to that proportion some make betwixt Air and Water; for they say, Air turn'd into Water takes up a hundred times less room than it did before. The truth is, we may reasonably suppose, that all the vapours of the middle Region were turn'd into water in this 40 days and 40 nights Rain, if we admit, that this Rain was throughout the whole

g. Phys.
ach., p. 221.

4 *feet in* 24
ri.

ct. cat. in
n. 7. 3.

Earth at once, in either Hemisphere, in every Zone, in every Climate, in every Country, in every Province, in every Field; and yet we see what a small proportion all this would amount to.

Having done then with these Superiour Regions, we are next to examine the Inferiour, and the treasures of water that may be had there. *Moses* tells us, that the Fountains of the great Abysse were broke open, or *clove asunder*, as the word there us'd doth imply; and no doubt in this lay the great mystery of the Deluge, as will appear when it comes to be rightly understood and explain'd; but we are here to consider what is generally understood by the *great Abysse*, in the common explication of the Deluge; and 'tis commonly interpreted either to be the Sea, or Subterraneous waters hid in the bowels of the Earth: These, they say, broke forth and rais'd the waters, caus'd by the Rain, to such an height, that together they overflowed the highest Mountains. But whether or how this could be, deserves to be a little examin'd.

And in the first place; the Sea is not higher than the Land, as some have formerly imagin'd; fansying the Sea stood, as it were, upon a heap, higher than the shore; and at the Deluge a relaxation being made, it overflow'd the Land. But this conceit is so gross, and so much against reason and experience, that none I think of late have ventur'd to make use of it. And yet on the other hand, if the Sea lie in an equal convexity with the Land, or lower generally than the shore, and much more than the mid-land, as it is certainly known to do, what could the Sea contribute to the Deluge? It would keep its Chanel, as it doth now, and take up the same place. And so also the Subterraneous waters would lie quiet in their Cells; whatsoever Fountains or passages you suppose, these would not issue out upon the Earth, for water doth not ascend, unless by force. But let's imagine then that force us'd and appli'd, and the waters both of the Sea and Caverns under ground drawn out upon the surface of the Earth, we shall not be any whit the nearer for this; for if you take these waters out of their places, those places must be fill'd again with other waters in the Deluge; so as this turns to no account upon the whole. If you have two Vessels to fill, and you empty one to fill the other, you gain nothing by that, there still remains one Vessel empty; you cannot have these waters both in the Sea and on the Land, both above ground and under; nor can you suppose the Chanel of the Sea would stand gaping without water, when all the Earth was overflow'd, and the tops of the Mountains cover'd. And so for Subterraneous Cavities, if you suppose the water pumpt out, they would suck it in again when the Earth came to be laid under water; so that upon the whole, if you thus understand the *Abysse* or *great Deep*, and the breaking open its *Fountains* in this manner, it doth us no service as to the Deluge, and where we expected the greatest supply, there we find none at all.

What shall we do then? whither shall we go to find more than seven Oceans of water that we still want? We have been above and below; we have drain'd the whole middle Region, and we have examin'd the Deeps of the Earth; they must want for themselves, they say, if they give us any; And, besides, if the Earth should disgorge all the water that it hath in its bowels, it would not amount to above half an Ocean, which would not at all answer our occasions. Must we not then conclude, that the common explication of the Deluge makes it impossible? there being no such quantity of water in Nature as they make requisite for an universal Deluge. Yet to give them all fair play, having examin'd the waters above the Earth, or in the Air, the waters upon the Earth, and the waters under the Earth; let us also consider if there be not waters above the Heavens, and if those might not be drawn down for the Deluge. *Moses* speaks of waters above the firmament, which though it be generally understood of the middle Region of the Air, especially as it was constituted before the Deluge, yet some have thought those to be waters plac'd above the highest Heavens, or Super-celestial waters: and have been willing to make use of them for a supply, when they could not find materials enough under the Heavens to make up the great mass of the Deluge. But the Heavens above, where these waters lay, are either solid, or fluid; if solid, as Glass or Crystal, how could the waters get through 'em to descend upon the Earth? If fluid, as the Air or Æther, how could the waters rest upon them? For Water is heavier than Air or Æther; So that I am afraid those pure Regions will prove no fit place for that Element, upon any account. But supposing these waters there, how imaginary soever, and that they were brought down to drown the World in that vast quantity that would be necessary, what became of them when the Deluge ceas'd? Seven or eight Oceans of water, with the Earth wrapt up in the middle of them, how did it ever get quit of them? how could they be dispos'd of when the Earth was to be dri'd, and the World renew'd? It would be a hard task to lift them up again among the Spheres, and we have no room for them here below. The truth is, I mention this opinion of the Heavenly waters, because I would omit none that had ever been made use of to make good the common explication of the Deluge; but otherwise, I think, since the System of the World hath been better known, and the Nature of the Heavens, there are none that would seriously assert these Super-celestial waters, or, at least, make use of them so extravagantly, as to bring them down hither for causes of the Deluge.

We have now employ'd our last and utmost endeavours to find out waters for the vulgar Deluge, or for the Deluge as commonly understood; and you see with how little success; we have left no corner unsought, where there was any appearance or report of water to be found, and yet we have not been able to collect the eighth part of what was necessary upon a moderate account. May we not then with assurance conclude, that the World hath taken wrong measures hitherto in their notion and explication of the general Deluge? They make it impossible and unintelligible upon a double account, both in requiring more water than can be found, and more than can be dispos'd of, if it was found: or could any way be withdrawn from the Earth when the Deluge should cease. For if the Earth was encompass'd with eight Oceans of water heapt one upon another, how these should retire into any Chanels, or be drain'd off, or the Earth any way disengag'd from them, is not intelligible; and that in so short a time as some months: For the violence of the Deluge lasted but four or five months, and in as many months after the Earth was dry and habitable. So as upon the whole enquiry, we can neither find source nor issue, beginning nor ending, for such an excessive mass of waters as the Vulgar Deluge requir'd; neither where to have them, nor if we had them, how to get quit of them. And I think men cannot do a greater injury or injustice to Sacred History, than to give such representations of things recorded there, as to make them unintelligible and incredible; As on the other hand, we cannot deserve better of Religion and Providence, than by giving such fair accounts of all things propos'd by them, or belonging to them, as may silence the Cavils of Atheists, satisfie the inquisitive, and recommend them to the belief and acceptance of all reasonable persons.

CHAPTER III

All Evasions answered; That there was no new Creation of waters at the Deluge: And that it was not particular or National, but extended throughout the whole Earth. A prelude and preparation to the true Account and Explication of it: The method of the first Book.

THOUGH in the preceeding Chapter we may seem to have given a fair trial to the common opinion concerning the state of the Deluge, and might now proceed to sentence of condemnation; yet having heard of another plea, which some have us'd in its behalf, and another way found out by recourse to the Supream Power, to supply all defects, and to make the whole matter intelligible, we will proceed no further till that be consider'd; being very willing to examine whatsoever may be offer'd, in that or any other way, for resolving that great difficulty which we have propos'd, concerning *the quantity of water requisite for such a Deluge*. And to this they say in short, that *God Almighty created waters on purpose to make the Deluge, and then annihilated them again when the Deluge was to cease*; And this, in a few words, is the whole account of the business. This is to cut the knot when we cannot loose it; They show us the naked arm of Omnipotency; such Arguments as these come like lightning, one doth not know what Armour to put on against them, for they pierce the more, the more they are resisted: We will not therefore oppose any thing to them that is hard and stubborn, but by a soft answer deaden their force by degrees.

And I desire to mind those persons in the first place of what St. *Austin* hath said upon a like occasion, speaking concerning those that disprov'd the opinion of waters above the Heavens (which we mentioned before) by natural Reasons. "We are not, saith he, to refute those persons, by saying, that according to the "Omnipotence of God, to whom all things are possible, we ought to believe "there are waters there as heavy as we know and feel them here below; for our "business is now to enquire according to his Scripture, how God hath constituted "the Nature of things, and not what he could do or work in these things by a "miracle of Omnipotency. I desire them to apply this to the present argument for the first answer.

Secondly, let them consider, that *Moses* hath assign'd causes of the Deluge; *Forty days Rain, and the disruption of the Abysse*; and speaks nothing of a new nature; ſo that they but retard for a while the progreſs by which they are all reſolved into duſt, and ſooner or later committed to the boſom of the deep.

117. We are not, however, to imagine, that there is no where any means of repairing this waſte; for, on comparing the concluſion at which we are now arrived, viz. that the preſent continents are all going to decay, and their materials deſcending into the ocean, with the propoſition firſt laid down, that theſe ſame continents are compoſed of materials which muſt have been collected from the decay of former rocks, it is impoſſible not to recogniſe two correſponding ſteps of the ſame progreſs; of a progreſs, by which mineral ſubſtances are ſubjected to the ſame ſeries of changes, and alternately waſted away and renovated. In the ſame manner, as the preſent mineral ſubſtances derive their origin from ſubſtances ſimilar to themſelves; ſo, from the land now going to decay, the ſand and gravel forming on the ſea-ſhore, or in the beds of rivers; from the ſhells and corals which in ſuch enormous quantities are every day accumulated in the boſom of the ſea; from the drift wood, and the multitude of vegetable and animal remains continually depoſited in the ocean: from all theſe we cannot doubt, that ſtrata are now forming in thoſe regions, to which nature ſeems to have confined the powers of mineral reproduction; from which, after being conſolidated, they are again deſtined to emerge, and to exhibit a ſeries of changes ſimilar to the paſt *.

118. How often theſe viciſſitudes of decay and renovation have been repeated, is not for us to determine: they conſtitute a ſeries, of which, as the author of this theory has remarked, we neither ſee the beginning nor the end; a circumſtance that accords well with what is known concerning other parts of the economy of the world. In the continuation of the different ſpecies of animals and vegetables that inhabit the earth, we diſcern neither a beginning nor an end; and, in the planetary motions, where geometry has carried the eye ſo far both into the future and the paſt, we diſcover no mark, either of the commencement or the termination of the preſent order †. It is unreaſonable, indeed, to ſuppoſe, that ſuch marks ſhould any where exiſt. The Author of nature has not given laws to the univerſe, which, like the inſtitutions of men, carry in themſelves the elements of their own deſtruction. He has not permitted, in his works, any ſymptom of infancy or of old age, or any ſign by which we may eſtimate either their future or their paſt duration. He may put an end, as he no doubt gave a beginning,

---

* Note xix.

† Note xx.

ing, to the preſent ſyſtem, at ſome determinate period; but we may ſafely conclude, that this great *cataſtrophe* will not be brought about by any of the laws now exiſting, and that it is not indicated by any thing which we perceive.

119. To aſſert, therefore, that, in the economy of the world, we ſee no mark, either of a beginning or an end, is very different from affirming, that the world had no beginning, and will have no end. The firſt is a concluſion juſtified by common ſenſe, as well as ſound philoſophy; while the ſecond is a preſumptuous and unwarrantable aſſertion, for which no reaſon from experience or analogy can ever be aſſigned. Dr Hutton might, therefore, juſtly complain of the uncandid criticiſm, which, by ſubſtituting the one of theſe aſſertions for the other, endeavoured to load his theory with the reproach of atheiſm and impiety. Mr Kirwan, in bringing forward this harſh and ill founded cenſure, was neither animated by the ſpirit, nor guided by the maxims of true philoſophy. By the ſpirit of philoſophy, he muſt have been induced to reflect, that ſuch poiſoned weapons as he was preparing to uſe, are hardly ever allowable in ſcientific conteſt, as having a leſs direct tendency to overthrow the ſyſtem, than to hurt the perſon of an adverſary, and to wound, perhaps incurably, his mind, his reputation, or his peace. By the maxims of philoſophy, he muſt have been reminded, that, in no part of the hiſtory of nature, has any mark been diſcovered, either of the beginning or the end of the preſent *order;* and that the geologiſt ſadly miſtakes, both the object of his ſcience and the limits of his underſtanding, who thinks it his buſineſs to explain the means employed by INFINITE WISDOM for eſtabliſhing the laws, which now govern the world.

By attending to theſe obvious conſiderations, Mr Kirwan would have avoided a very illiberal and ungenerous proceeding; and, however he might have differed from Dr Hutton as to the *truth* of his opinions, he would not have cenſured their *tendency* with ſuch raſh and unjuſtifiable ſeverity.

But, if this author may be blamed for wanting the temper, or neglecting the rules, of philoſophic inveſtigation, he is hardly leſs culpable, for having ſo ſlightly conſidered the ſcope and ſpirit of a work which he condemned ſo freely. In that work, inſtead of finding the world repreſented as the reſult of neceſſity or chance, which might be looked for, if the accuſations of atheiſm or impiety were well founded, we ſee every where the utmoſt attention to diſcover, and the utmoſt diſpoſition to admire, the inſtances of wiſe and beneficent deſign,

ſign manifeſted in the ſtructure, or economy of the world. The enlarged views of theſe, which his geological ſyſtem afforded, appeared to Dr Hutton himſelf as its moſt valuable reſult. They were the parts of it which he contemplated with greateſt delight; and he would have been leſs flattered, by being told of the ingenuity and originality of his theory, than of the addition which it had made to our knowledge of *final cauſes*. It was natural, therefore, that he ſhould be hurt by an attempt to accuſe him of opinions, ſo different from thoſe which he had always taught; and if he anſwered Mr Kirwan's attack with warmth or aſperity, we muſt aſcribe it to the indignation excited by unmerited reproach.

120. But to return to the natural hiſtory of the earth: Though there be in it no *data*, from which the commencement of the preſent order can be aſcertained, there are many by which the exiſtence of that order may be traced back to an antiquity extremely remote. The beds of primitive ſchiſtus, for inſtance, contain ſand, gravel, and other materials, collected, as already ſhewn, from the diſſolution of mineral bodies; which bodies, therefore, muſt have exiſted long before the oldeſt part of the preſent land was formed. Again, in this gravel we ſometimes find pieces of ſandſtone, and of other compound rocks, by which we are of courſe carried back a ſtep farther, ſo as to reach a ſyſtem of things, from which the preſent is the third in ſucceſſion; and this may be conſidered as the moſt ancient epocha, of which any memorial exiſts in the records of the foſſil kingdom.

121. Next in the order of time to the conſolidation of the primary ſtrata, we muſt place their elevation, when, from being horizontal, and at the bottom of the ſea, they were broken, ſet on edge, and raiſed to the ſurface. It is even probable, as formerly obſerved, that to this ſucceeded a depreſſion of the ſame ſtrata, and a ſecond elevation, ſo that they have twice viſited the ſuperior, and twice the inferior regions. During the ſecond immerſion, were formed, firſt the great bodies of pudding-ſtone, that in ſo many inſtances lie immediately above them and next were depoſited the ſtrata that are ſtrictly denominated ſecondary.

122. The third great event, was the raiſing up of this compound body of old and new ſtrata from the bottom of the ſea, and forming it into the dry land, or the continents, as they now exiſt *. Contemporary with this, we muſt ſuppoſe the injection of melted matter among the ſtrata, and the conſequent formation of the cryſtallized and unſtratified rocks, namely, the granite, metallic veins, and veins of porphyry and whinſtone.

---

* Note xxi.

whinſtone. This, however, is to be conſidered as embracing a period of great duration; and it muſt always be recollected, that veins are found of very different formation; ſo that when we ſpeak generally, it is perhaps impoſſible to ſtate any thing more preciſe concerning their antiquity, than that they are poſterior to the ſtrata, and that the veins of whinſtone ſeem to be the moſt recent of all, as they traverſe every other.

123. In the fourth place, with reſpect to time, we muſt claſs the facts that regard the detritus and waſte of the land, and muſt carefully diſtinguiſh them from the more ancient phenomena of the mineral kingdom. Here we are to reckon the ſhaping of all the preſent inequalities of the ſurface; the formation of hills of gravel, and of what have been called tertiary ſtrata, conſiſting of looſe and unconſolidated materials; alſo collections of ſhells not mineralized, like thoſe in Turaine; ſuch petrifactions as thoſe contained in the rock of Gibraltar, on the coaſt of Dalmatia, and in the caves of Bayreuth. The bones of land animals found in the ſoil, ſuch as thoſe of Siberia, or North America, are probably more recent than any of the former*.

124. Theſe phenomena, then, are all ſo many marks of the lapſe of time, among which the principles of geology enable us to diſtinguiſh a certain order, ſo that we know ſome of them to be more, and others to be leſs diſtant, but without being able to aſcertain, with any exactneſs, the proportion of the immenſe intervals which ſeparate them. Theſe intervals admit of no compariſon with the aſtronomical meaſures of time; they cannot be expreſſed by the revolutions of the ſun or of the moon; nor is there any ſynchroniſm between the moſt recent epochas of the mineral kingdom, and the moſt ancient of our ordinary chronology.

125. On what is now ſaid is grounded another objection to Dr Hutton's theory, namely, that the high antiquity aſcribed by it to the earth, is inconſiſtent with that ſyſtem of chronology which reſts on the authority of the Sacred Writings. This objection would no doubt be of weight, if the high antiquity in queſtion were not reſtricted merely to the globe of the earth, but were alſo extended to the human race. That the origin of mankind does not go back beyond ſix or ſeven thouſand years, is a poſition ſo involved in the narrative of the Moſaic books, that any thing inconſiſtent with it, would no doubt ſtand in oppoſition to the teſtimony of thoſe ancient records. On this ſubject, however, geology is ſilent; and the hiſtory of arts and ſciences, when traced as high as any authentic monuments extend, refers the

---

* Note xxii.

the beginnings of civilization to a date not very different from that which has juſt been mentioned, and infinitely within the limits of the moſt recent of the epochas, marked by the phyſical revolutions of the globe.

On the other hand, the authority of the Sacred Books ſeems to be but little intereſted in what regards the mere antiquity of the earth itſelf; nor does it appear that their language is to be underſtood literally concerning the *age* of that body, any more than concerning its *figure* or its *motion*. The theory of Dr Hutton ſtands here preciſely on the ſame footing with the ſyſtem of Copernicus; for there is no reaſon to ſuppoſe, that it was the purpoſe of revelation to furniſh a ſtandard of geological, any more than of aſtronomical ſcience. It is admitted, on all hands, that the Scriptures are not intended to reſolve phyſical queſtions, or to explain matters in no way related to the morality of human actions; and if, in conſequence of this principle, a conſiderable latitude of interpretation were not allowed, we ſhould continue at this moment to believe, that the earth is flat; that the ſun moves round the earth; and that the circumference of a circle is no more than three times its diameter.

It is but reaſonable, therefore, that we ſhould extend to the geologiſt the ſame liberty of ſpeculation, which the aſtronomer and mathematician are already in poſſeſſion of; and this may be done, by ſuppoſing that the chronology of Moses relates only to the human race. This liberty is not more neceſſary to Dr Hutton than to other theoriſts. No ingenuity has been able to reconcile the natural hiſtory of the globe with the opinion of its recent origin; and accordingly the coſmologies of Kirwan and De Luc, though contrived with more mineralogical ſkill, are not leſs forced and unſatisfactory than thoſe of Burnet and Whiſton.

126. It is impoſſible to look back on the ſyſtem which we have thus endeavoured to illuſtrate, without being ſtruck with the novelty and beauty of the views which it ſets before us. The very plan and ſcope of it diſtinguiſh it from all other theories of the earth, and point it out as a work of great and original invention. The ſole object of ſuch theories has hitherto been, to explain the manner in which the preſent laws of the mineral kingdom were firſt eſtabliſhed, or began to exiſt, without treating of the manner in which they now proceed, and by which their continuance is provided for. The authors of theſe theories have accordingly gone back to a ſtate of things altogether unlike the preſent, and have confined their reaſonings, or

their

their fictions, to a criſis which never has exiſted but once, and which never can return. Dr Hutton, on the other hand, has guided his inveſtigation by the philoſophical maxim, *Cauſam naturalem et aſſiduam quærimus, non raram et fortuitam.* His theory, accordingly, preſents us with a ſyſtem of wiſe and provident economy, where the ſame inſtruments are continually employed, and where the decay and renovation of foſſils being carried on at the ſame time in the different regions allotted to them, preſerve in the earth the conditions eſſential for the ſupport of animal and vegetable life. We have been long accuſtomed to admire that beautiful contrivance in nature, by which the water of the ocean, drawn up in vapour by the atmoſphere, imparts, in its deſcent, fertility to the earth, and becomes the great cauſe of vegetation and of life; but now we find, that this vapour not only fertilizes, but creates the ſoil; prepares it from the ſolid rock, and, after employing it in the great operations of the ſurface, carries it back into the regions where all its mineral characters are renewed. Thus, the circulation of moiſture through the air, is a prime mover, not only in the annual ſucceſſion of the ſeaſons, but in the great geological cycle, by which the waſte and reproduction of entire continents is circumſcribed. Perhaps a more ſtriking view than this, of the wiſdom

that preſides over nature, was never preſented by any philoſophical ſyſtem, nor a greater addition ever made to our knowledge of final cauſes. It is an addition which gives conſiſtency to the reſt, by proving, that equal foreſight is exerted in providing for the whole and for the parts, and that no leſs care is taken to maintain the conſtitution of the earth, than to preſerve the tribes of animals and vegetables which dwell on its ſurface. In a word, it is the peculiar excellence of this theory, that it aſcribes to the phenomena of geology an order ſimilar to that which exiſts in the provinces of nature with which we are beſt acquainted; that it produces ſeas and continents, not by accident, but by the operation of regular and uniform cauſes; that it makes the decay of one part ſubſervient to the reſtoration of another, and gives ſtability to the whole, not by perpetuating individuals, but by reproducing them in ſucceſſion.

127. Again, in the detail of this theory, and the ample induction on which it is founded, we meet with many facts and obſervations, either entirely new, or hitherto very imperfectly underſtood. Thus, the veins which proceed from maſſes of granite, and penetrate the incumbent ſchiſtus, had either eſcaped the obſervation of former mineralogiſts, or the importance of the phenomenon had been entirely overlooked. Dr

Hutton has deſcribed the appearances with great accuracy, and drawn from them the moſt intereſting concluſions. At the junction of the primary and ſecondary ſtrata, the facts which he has noted had been obſerved by others; but no one I think had ſo fully underſtood the language which they ſpeak, or had ſo clearly perceived the conſequences that neceſſarily follow from them. He is the firſt who diſtinctly pointed out the characters which diſtinguiſh whinſtone from lava, and who explained the true relation that ſubſiſts between theſe ſubſtances He alſo diſcovered the induration of the ſtrata, in contact with veins of whin, and the charring of the coal in their vicinity. His theory alſo enabled him to determine the affinity of whinſtone and granite to one another, and their relation to the other great bodies of the mineral kingdom.

To the obſervations of the ſame excellent geologiſt, we are indebted for the knowledge of the general and important fact, that all the hard ſubſtances of the mineral kingdom, when elevated into the atmoſphere, have a tendency to decay, and are ſubject to a diſintegration and waſte, to which no limit can be ſet but that of their entire deſtruction; that no proviſion is made on the ſurface for repairing this waſte, and that there, no new foſſil is produced; that the formation of all the varied ſcenery which the ſurface of the earth exhibits, depends on the operation of cauſes, the momentary exertions of which are familiar to us, though we knew not before the effects which their accumulated action was able to produce. Theſe are facts in the natural hiſtory of the earth, the diſcovery of which is due to Dr Hutton; and, ſhould we lay all further ſpeculation aſide, and conſider the theory of the earth as a work too great to be attempted by man, we muſt ſtill regard the phenomena and laws juſt mentioned, as forming a ſolid and valuable addition to our knowledge.

128. If we would compare this theory with others, as to the inviſible agents which it employs, we muſt conſider, that fire and water are the two powers which all of them muſt make uſe of, ſo that they can differ from one another only by the way in which they combine theſe powers. In Dr Hutton's ſyſtem, water is firſt employed to depoſite and arrange, and then fire to conſolidate, mineralize, and laſtly, to elevate the ſtrata; but, with reſpect to the unſtratified or cryſtallized ſubſtances, the action of fire only is recogniſed. The ſyſtem having leaſt affinity to this is the Neptunian, which aſcribes the formation of all minerals to the action of water alone, and extends this hypotheſis even to the unſtratified rocks. Here, therefore, the action of fire is entirely excluded; and the Neptuniſts have

have certainly made a great ſacrifice to the love of truth, or of paradox, in rejecting the aſſiſtance of ſo powerful an auxiliary *.

129. In the ſyſtems which employ the agency of the latter element, we are to look for a greater reſemblance to that of Dr Hutton, though many and great marks of diſtinction are eaſily perceived. In the coſmologies, for example, of LEIBNITZ and BUFFON, fire and water are both employed, as well as in this; but they are employed in a reverſe order. Theſe philoſophers introduce the action of fire firſt, and then the action of water, which is to invert the order of nature altogether, as the conſolidation of the rocks muſt be poſterior to their ſtratification. Indeed, the theory of Buffon is ſingularly defective: beſides inverting the order of the two great operations of ſtratification and conſolidation, and of courſe giving no real explanation of the latter, it gives no account of the elevation, or highly inclined poſition of the ſtrata; it makes no diſtinction between ſtratified and unſtratified bodies, nor does it offer any but the moſt unſatisfactory explanation of the inequalities of the earth's ſurface. This ſyſtem, therefore, has but a very diſtant reſemblance to the Huttonian theory †.

130. The ſyſtem of LAZZARO MORO has been remarked as approaching nearer to this theory than any other; and it is certain, that one very important principle is common to them both. The theory of the Italian geologiſt was chiefly directed to the explanation of the remains of marine animals, which are found in mountains far from the ſea; and it appears to have been ſuggeſted to him by the phenomena of the *Campi Phlegræi*, and by the production of the new iſland of *Santorini* in the Archipelago. He accordingly ſuppoſes, that the iſlands and continents have been all raiſed up, like the above-mentioned iſland, from the bottom of the ſea, by the force of volcanic fire: that theſe fires began to burn under the bottom of the ocean ſoon after the creation of the world, when as yet the ocean covered the whole earth: that they at firſt elevated a portion of the land; and in this primitive land no ſhells are found, as the original ocean was deſtitute of fiſh. The volcanoes continuing to burn, under the ſea, after the creation of animated nature, the ſtrata that were then raiſed up by their action were full of ſhells and other marine objects; and, from the violence with which they were elevated, aroſe the contortions and inclined poſition which they frequently poſſeſs *.

I 3 This

---

* NOTE XXIII.

† NOTE XXIV.

* De' Croſtacei, et degli altri Marini Corpi, che ſi trovano ſu' Monti: di Ant. Lazzaro Moro. Vinezia. 1740.

This ſyſtem is imperfect, as it makes no peculiar proviſion for the conſolidation of the ſtrata, which, according to it, as well as the Neptunian ſyſtem, muſt be aſcribed to the action, not of fire, but of water. No account is given of the mineralization of the ſhells found in the ſtrata, or of the difference between them and the ſhells found looſe at the bottom of the ſea; and no diſtinction is made between ſtratified and unſtratified ſubſtances. But, with all this, Lazzaro Moro has certainly the merit of having perceived, that ſome other power than that which depoſited the ſtrata, muſt have been employed for their elevation, and that they have endured the action of a diſturbing force.

131. From this compariſon it appears, that Dr Hutton's theory is ſufficiently diſtinct, even from the theories which approach to it moſt nearly, to merit, in the ſtricteſt ſenſe, the appellation of *new* and *original*. There are indeed few inventions or diſcoveries, recorded in the hiſtory of ſcience, to which nearer approaches were not made before they were fully unfolded. It therefore very well deſerves to be diſtinguiſhed by a particular name; and, if it behoves us to follow the analogy obſerved in the names of the two great ſyſtems, which at preſent divide the opinions of geologiſts, we may join Mr Kirwan in calling this the PLUTONIC SYSTEM. For my own part, I would rather have it characterized by a leſs ſplendid, but juſter name, that of the HUTTONIAN THEORY.

132. The circumſtance, however, which gives to this theory its peculiar character, and exalts it infinitely above all others, is the introduction of the principle of preſſure, to modify the effects of heat when applied at the bottom of the ſea. This is in fact the key to the grand enigma of the mineral kingdom, where, while one ſet of phenomena indicates the action of fire, another ſet, equally remarkable, ſeems to exclude the poſſibility of that action, by preſenting us with mineral ſubſtances, in ſuch a ſtate as they could never have been brought into by the operation of the fires we ſee at the ſurface of the earth. Theſe two claſſes of phenomena are reconciled together, by admitting the power of compreſſion to confine the volatile parts of bodies when heat is applied to them, and to force them, in many inſtances, to undergo fuſion, inſtead of being calcined or diſſipated by burning or inflammation. In this hypotheſis, which ſome affect to conſider as a principle gratuitouſly aſſumed, there appears to me nothing but a very fair and legitimate generalization of the properties of heat. Combuſtion and inflammation are chemical proceſſes, to which other conditions are required, beſides the preſence of a high temperature. The

ſtate

By parity of reasoning we can hardly expect that any hypogene rocks of the tertiary periods will have been brought within the reach of human observation, seeing that the emergence of such rocks must always be so long posterior to the date of their origin, and still less can formations of this class become generally visible until so much time has elapsed as to confer on them a high relative antiquity. Extensive denudation must also combine with upheaval before they can be displayed at the surface throughout wide areas.

All geologists who reflect on subterranean movements now going on, and the eruptions of active volcanos, are convinced that great changes are now continually in progress in the interior of the earth's crust far out of sight. They must be conscious, therefore, that the inaccessibility of the regions in which these alterations are taking place, compels them to remain in ignorance of a great part of the working of existing causes, so that they can only form vague conjectures in regard to the nature of the products which volcanic heat may elaborate under great pressure.

But when they find in mountain chains of high antiquity that what was once the interior of the earth's crust has since been forced outwards and exposed to view, they will naturally expect, in the examination of those mountainous regions, to have an opportunity of gratifying their curiosity by obtaining a sight not only of the superficial strata of remote eras, but also of the contemporaneous nether-formed rocks. Having recognized therefore in such mountain chains some ancient rocks of aqueous and volcanic origin, corresponding in character to superficial formations of modern date, they will regard any other class of ancient rocks, such

as granite and gneiss, as the *residual phenomena* of which they are in search. These latter rocks will not answer the expectations previously formed of their probable nature and texture, unless they wear a foreign and mysterious aspect, and have in some places been fused or altered by subterranean heat; in a word, unless they differ wholly from the fossiliferous strata deposited at the surface, or from the lava and scoriæ thrown out by volcanos in the open air. It is the total distinctness, therefore, of crystalline formations, such as granite, hornblende-schist, and the rest, from every substance of which the origin is familiar to us, that constitutes their claim to be regarded as the effects of causes now in action in the subterranean regions. They belong not to an order of things which have passed away; they are not the monuments of a primeval period, bearing inscribed upon them in obsolete characters the words and phrases of a dead language; but they teach us that part of the living language of nature which we cannot learn by our daily intercourse with what passes on the habitable surface.

Start Here

Concluding remarks on the identity of the ancient and present system of terrestrial changes.— I shall now conclude the discussion of a question with which we have been occupied since the beginning of the fifth chapter; namely, whether there has been any interruption, from the remotest periods, of one uniform system of change in the animate and inanimate world. We were induced to enter into that inquiry by reflecting how much the progress of opinion in Geology had been influenced by the assumption that the analogy was slight in kind, and still more slight in degree, between the causes which produced the former revolutions of the globe, and those now in every day operation. It appeared clear that the earlier geologists had not only a scanty acquaintance with existing changes, but were singularly unconscious of their ignorance. With the presumption naturally inspired by this unconsciousness, they had no hesitation in deciding at once that time could never enable the existing powers of nature to work out changes of great magnitude, still less such important revolutions as those which are brought to light by Geology. They, therefore, felt themselves at liberty to indulge their imaginations in guessing at what *might be*, rather than in inquiring *what is*; in other words, they employed themselves in conjecturing what might have been the course of nature at a remote period, rather than in the investigation of what was the course of nature in their own times.

It appeared to them more philosophical to speculate on the possibilities of the past, than patiently to explore the realities of the present; and having invented theories under the influence of such maxims, they were consistently unwilling to test their validity by the criterion of their accordance with the ordinary operations of nature. On the contrary, the claims of each new hypothesis to credibility appeared enhanced by the great contrast, in kind or intensity, of the causes referred to and those now in operation.

Never was there a dogma more calculated to foster indolence, and to blunt the keen edge of curiosity, than this assumption of the discordance between the ancient and existing causes of change. It produced a state of mind unfavourable in the highest degree to the candid reception of the evidence of those minute but incessant alterations which every part of the earth's

surface is undergoing, and by which the condition of its living inhabitants is continually made to vary. The student, instead of being encouraged with the hope of interpreting the enigmas presented to him in the earth's structure, — instead of being prompted to undertake laborious inquiries into the natural history of the organic world, and the complicated effects of the igneous and aqueous causes now in operation, was taught to despond from the first. Geology, it was affirmed, could never rise to the rank of an exact science, — the greater number of phenomena must for ever remain inexplicable, or only be partially elucidated by ingenious conjectures. Even the mystery which invested the subject was said to constitute one of its principal charms, affording, as it did, full scope to the fancy to indulge in a boundless field of speculation.

The course directly opposed to this method of philosophizing consists in an earnest and patient inquiry how far geological appearances are reconcileable with the effect of changes now in progress, or which may be in progress in regions inaccessible to us, and of which the reality is attested by volcanos and subterranean movements. It also endeavours to estimate the aggregate result of ordinary operations multiplied by time, and cherishes a sanguine hope that the resources to be derived from observation and experiment, or from the study of nature such as she now is, are very far from being exhausted. For this reason all theories are rejected which involve the assumption of sudden and violent catastrophes and revolutions of the whole earth, and its inhabitants, — theories which are restrained by no reference to existing analogies, and in which a desire is manifested to cut, rather than patiently to untie, the Gordian knot.

We have now, at least, the advantage of knowing, from experience, that an opposite method has always put geologists on the road that leads to truth, — suggesting views which, although imperfect at first, have been found capable of improvement, until at last adopted by universal consent; while the method, of speculating on a former distinct state of things and causes, has led invariably to a multitude of contradictory systems, which have been overthrown one after the other, — have been found incapable of modification, — and which have often required to be precisely reversed.

The remainder of this work will be devoted to an investigation of the changes now going on in the crust of the earth and its inhabitants. The importance which the student will attach to such researches will mainly depend in the degree of confidence which he feels in the principles above expounded. If he firmly believes in the resemblance or identity of the ancient and present system of terrestrial changes, he will regard every fact collected respecting the causes in diurnal action as affording him a key to the interpretation of some mystery in the past. Events which have occurred at the most distant periods in the animate and inanimate world will be acknowledged to throw light on each other, and the deficiency of our information respecting some of the most obscure parts of the present creation will be removed. For as, by studying the external configuration of the existing land and its inhabitants, we may restore in imagination the appearance of the ancient continents which have passed away, so may we obtain from the deposits of ancient seas and lakes an insight into the nature of the subaqueous processes

now in operation, and of many forms of organic life, which, though now existing, are veiled from sight. Rocks, also, produced by subterranean fire in former ages, at great depths in the bowels of the earth, present us, when upraised by gradual movements, and exposed to the light of heaven, with an image of those changes which the deep-seated volcano may now occasion in the nether regions. Thus, although we are mere sojourners on the surface of the planet, chained to a mere point in space, enduring but for a moment of time, the human mind is not only enabled to number worlds beyond the unassisted ken of mortal eye, but to trace the events of indefinite ages before the creation of our race, and is not even withheld from penetrating into the dark secrets of the ocean, or the interior of the solid globe; free, like the spirit which the poet described as animating the universe,

> ——— ire per omnes
> Terrasque, tractusque maris, cœlumque profundum.

began at a later period; and in every step which it has hitherto made towards sound theoretical principles, it has had to contend against more violent prepossessions. The practical advantages already derived from it have not been inconsiderable: but our generalizations are yet imperfect, and they who come after us may be expected to reap the most valuable fruits of our labour. Meanwhile the charm of first discovery is our own; and, as we explore this magnificent field of inquiry, the sentiment of a great historian of our times may continually be present to our minds, that "he who calls what has vanished back again into being, enjoys a bliss like that of creating."*

---

* Niebuhr's Hist. of Rome, vol. i. p. 5. Hare and Thirlwall's translation.

CHAPTER V.

PREJUDICES WHICH HAVE RETARDED THE PROGRESS OF GEOLOGY.

Prepossessions in regard to the duration of past time — Prejudices arising from our peculiar position as inhabitants of the land — Of those occasioned by our not seeing subterranean changes now in progress — All these causes combine to make the former course of Nature appear different from the present — Objections to the doctrine that causes similar in kind and energy to those now acting, have produced the former changes of the earth's surface, considered.

IF we reflect on the history of the progress of geology, as explained in the preceding chapters, we perceive that there have been great fluctuations of opinion respecting the nature of the causes to which all former changes of the earth's surface are referable. The first observers conceived the monuments which the geologist endeavours to decipher to relate to an original state of the earth, or to a period when there were causes in activity, distinct, in kind and degree, from those now constituting the economy of nature. These views were gradually modified, and some of them entirely abandoned in proportion as observations were multiplied, and the signs of former mutations more skilfully interpreted. Many appearances, which had for a long time been regarded as indicating mysterious and extraordinary agency, were finally recognized as the necessary result of the laws now governing the material world; and the discovery of this unlooked-for conformity has at length induced some philosophers to infer, that, during the ages contemplated in geology, there has never been any interruption to the agency of the same uniform laws of change. The same assemblage of general causes, they conceive, may have been sufficient to produce, by their various combinations, the endless diversity of effects, of which the shell of the earth has preserved the memorials; and, consistently with these principles, the recurrence of analogous changes is expected by them in time to come.

Whether we coincide or not in this doctrine, we must admit that the gradual progress of opinion concerning the succession of phenomena in very remote eras, resembles, in a singular manner, that which has accompanied the growing intelligence of every people, in regard to the economy of nature in their own times. In an early stage of advancement, when a great number of natural appearances are unintelligible, an eclipse, an earthquake, a flood, or the approach of a comet, with many other occurrences afterwards found to belong to the regular course of events, are regarded as prodigies. The same delusion prevails as to moral phenomena, and many of these are ascribed to the intervention of demons, ghosts, witches, and other immaterial and supernatural agents. By degrees, many of the enigmas of the moral and physical world are explained, and, instead of being due to extrinsic and irregular causes, they are found to depend on fixed and invariable laws. The philosopher at last becomes convinced of the undeviating uniformity of secondary causes; and, guided by his faith in this principle, he determines the probability of accounts transmitted to

him of former occurrences, and often rejects the fabulous tales of former times, on the ground of their being irreconcilable with the experience of more enlightened ages.

Prepossessions in regard to the duration of past time. — As a belief in the want of conformity in the causes by which the earth's crust has been modified in ancient and modern periods was, for a long time, universally prevalent, and that, too, amongst men who were convinced that the order of nature had been uniform for the last several thousand years, every circumstance which could have influenced their minds and given an undue bias to their opinions deserves particular attention. Now the reader may easily satisfy himself, that, however undeviating the course of nature may have been from the earliest epochs, it was impossible for the first cultivators of geology to come to such a conclusion, so long as they were under a delusion as to the age of the world, and the date of the first creation of animate beings. However fantastical some theories of the sixteenth century may now appear to us, — however unworthy of men of great talent and sound judgment, — we may rest assured that, if the same misconception now prevailed in regard to the memorials of human transactions, it would give rise to a similar train of absurdities. Let us imagine, for example, that Champollion, and the French and Tuscan literati lately engaged in exploring the antiquities of Egypt, had visited that country with a firm belief that the banks of the Nile were never peopled by the human race before the beginning of the nineteenth century, and that their faith in this dogma was as difficult to shake as the opinion of our ancestors, that the earth was never the abode of living beings until the creation of the present continents, and of the species now existing,—it is easy to perceive what extravagant systems they would frame, while under the influence of this delusion, to account for the monuments discovered in Egypt. The sight of the pyramids, obelisks, colossal statues, and ruined temples, would fill them with such astonishment, that for a time they would be as men spell-bound—wholly incapable of reasoning with sobriety. They might incline at first to refer the construction of such stupendous works to some superhuman powers of a primeval world. A system might be invented resembling that so gravely advanced by Manetho, who relates that a dynasty of gods originally ruled in Egypt, of whom Vulcan, the first monarch, reigned nine thousand years; after whom came Hercules and other demigods, who were at last succeeded by human kings.

When some fanciful speculations of this kind had amused their imaginations for a time, some vast repository of mummies would be discovered, and would immediately undeceive those antiquaries who enjoyed an opportunity of personally examining them; but the prejudices of others at a distance, who were not eye-witnesses of the whole phenomena, would not be so easily overcome. The concurrent report of many travellers would, indeed, render it necessary for them to accommodate ancient theories to some of the new facts, and much wit and ingenuity would be required to modify and defend their old positions. Each new invention would violate a greater number of known analogies; for if a theory be required to embrace some false principle, it becomes more visionary in proportion as facts are multiplied, as would be the case if geometers were now required to form an astronomical

system on the assumption of the immobility of the earth.

Amongst other fanciful conjectures concerning the history of Egypt, we may suppose some of the following to be started. "As the banks of the Nile have been so recently colonized for the first time, the curious substances called mummies could never in reality have belonged to men. They may have been generated by some *plastic virtue* residing in the interior of the earth, or they may be abortions of nature produced by her incipient efforts in the work of creation. For if deformed beings are sometimes born even now, when the scheme of the universe is fully developed, many more may have been 'sent before their time, scarce half made up,' when the planet itself was in the embryo state. But if these notions appear to derogate from the perfection of the Divine attributes, and if these mummies be in all their parts true representations of the human form, may we not refer them to the future rather than the past? May we not be looking into the womb of Nature, and not her grave? May not these images be like the shades of the unborn in Virgil's Elysium—the archetypes of men not yet called into existence?"

These speculations, if advocated by eloquent writers, would not fail to attract many zealous votaries, for they would relieve men from the painful necessity of renouncing preconceived opinions. Incredible as such scepticism may appear, it has been rivalled by many systems of the sixteenth and seventeenth centuries, and among others by that of the learned Falloppio, who regarded the tusks of fossil elephants as earthy concretions, and the pottery or fragments of vases in the Monte Testaceo, near Rome, as works of nature, and not of art. But when one generation had passed away, and another, not compromised to the support of antiquated dogmas, had succeeded, they would review the evidence afforded by mummies more impartially, and would no longer controvert the preliminary question, that human beings had lived in Egypt before the nineteenth century: so that when a hundred years perhaps had been lost, the industry and talents of the philosopher would be at last directed to the elucidation of points of real historical importance.

But the above arguments are aimed against one only of many prejudices with which the earlier geologists had to contend. Even when they conceded that the earth had been peopled with animate beings at an earlier period than was at first supposed, they had no conception that the quantity of time bore so great a proportion to the historical era as is now generally conceded. How fatal every error as to the quantity of time must prove to the introduction of rational views concerning the state of things in former ages, may be conceived by supposing the annals of the civil and military transactions of a great nation to be perused under the impression that they occurred in a period of one hundred instead of two thousand years. Such a portion of history would immediately assume the air of a romance; the events would seem devoid of credibility, and inconsistent with the present course of human affairs. A crowd of incidents would follow each other in thick succession. Armies and fleets would appear to be assembled only to be destroyed, and cities built merely to fall in ruins. There would be the most violent transitions from foreign or intestine war to periods of profound peace, and the works effected during the years of disorder or

tranquillity would appear alike superhuman in magnitude.

He who should study the monuments of the natural world under the influence of a similar infatuation, must draw a no less exaggerated picture of the energy and violence of causes, and must experience the same insurmountable difficulty in reconciling the former and present state of nature. If we could behold in one view all the volcanic cones thrown up in Iceland, Italy, Sicily, and other parts of Europe, during the last five thousand years, and could see the lavas which have flowed during the same period; the dislocations, subsidences, and elevations caused by earthquakes; the lands added to various deltas, or devoured by the sea, together with the effects of devastation by floods, and imagine that all these events had happened in one year, we must form most exalted ideas of the activity of the agents, and the suddenness of the revolutions. Were an equal amount of change to pass before our eyes in the next year, could we avoid the conclusion that some great crisis of nature was at hand! If geologists, therefore, have misinterpreted the signs of a succession of events, so as to conclude that centuries were implied where the characters imported thousands of years, and thousands of years where the language of nature signified millions, they could not, if they reasoned logically from such false premises, come to any other conclusion than that the system of the natural world had undergone a complete revolution.

We should be warranted in ascribing the erection of the great pyramid to superhuman power, if we were convinced that it was raised in one day; and if we imagine, in the same manner, a mountain-chain to have been elevated, during an equally small fraction of the time which was really occupied in upheaving it, we might then be justified in inferring, that the subterranean movements were once far more energetic than in our own times. We know that one earthquake may raise the coast of Chili, for a hundred miles to the average height of about three feet. A repetition of two thousand shocks, of equal violence, might produce a mountain-chain one hundred miles long, and six thousand feet high. Now, should one or two only of these convulsions happen in a century, it would be consistent with the order of events experienced by the Chilians from the earliest times: but if the whole of them were to occur in the next hundred years, the entire district must be depopulated, scarcely any animals or plants could survive, and the surface would be one confused heap of ruin and desolation.

One consequence of undervaluing greatly the quantity of past time, is the apparent coincidence which it occasions of events necessarily disconnected, or which are so unusual, that it would be inconsistent with all calculation of chances to suppose them to happen at one and the same time. When the unlooked-for association of such rare phenomena is witnessed in the present course of nature, it scarcely ever fails to excite a suspicion of the preternatural in those minds which are not firmly convinced of the uniform agency of secondary causes; — as if the death of some individual in whose fate they are interested happens to be accompanied by the appearance of a luminous meteor, or a comet, or the shock of an earthquake. It would be only necessary to multiply such coincidences indefinitely, and the mind of every philosopher would be disturbed. Now it would be difficult to exaggerate the number of physical events, many of them most rare and unconnected in their nature, which were imagined by the Woodwardian hypothesis to have happened in the course of a few months: and numerous other examples might be found of popular geological theories, which require us to imagine that a long succession of events happened in a brief and almost momentary period.

Another liability to error, very nearly allied to the former, arises from the frequent contact of geological monuments referring to very distant periods of time. We often behold, at one glance, the effects of causes which have acted at times incalculably remote, and yet there may be no striking circumstances to mark the occurrence of a great chasm in the chronological series of Nature's archives. In the vast interval of time which may really have elapsed between the results of operations thus compared, the physical condition of the earth may, by slow and insensible modifications, have become entirely altered; one or more races of organic beings may have passed away, and yet have left behind, in the particular region under contemplation, no trace of their existence.

To a mind unconscious of these intermediate events, the passage from one state of things to another must appear so violent, that the idea of revolutions in the system inevitably suggests itself. The imagination is as much perplexed by the deception, as it might be if two distant points in space were suddenly brought into immediate proximity. Let us suppose, for a moment, that a philosopher should lie down to sleep in some arctic wilderness, and then be transferred by a power, such as we read of in tales of enchantment, to a valley in a tropical country, where, on awaking, he might find himself surrounded by birds of brilliant plumage, and all the luxuriance of animal and vegetable forms of which Nature is so prodigal in those regions. The most reasonable supposition, perhaps, which he could make, if by the necromancer's art he was placed in such a situation, would be, that he was dreaming; and if a geologist form theories under a similar delusion, we cannot expect him to preserve more consistency in his speculations, than in the train of ideas in an ordinary dream.

It may afford, perhaps, a more lively illustration of the principle here insisted upon, if I recall to the reader's recollection the legend of the Seven Sleepers. The scene of that popular fable was placed in the two centuries which elapsed between the reign of the emperor Decius and the death of Theodosius the younger. In that interval of time (between the years 249 and 450 of our era) the union of the Roman empire had been dissolved, and some of its fairest provinces overrun by the barbarians of the north. The seat of government had passed from Rome to Constantinople, and the throne from a Pagan persecutor to a succession of Christian and orthodox princes. The genius of the empire had been humbled in the dust, and the altars of Diana and Hercules were on the point of being transferred to Catholic saints and martyrs. The legend relates "that when Decius was still persecuting the Christians, seven noble youths of Ephesus concealed themselves in a spacious cavern in the side of an adjacent mountain, where they were doomed to perish by the tyrant, who gave orders that the entrance should be firmly secured with a pile of huge stones. They immediately fell into a deep slumber, which was miraculously prolonged, without injuring the powers of life, during a period of 187 years.

At the end of that time the slaves of Adolius, to whom the inheritance of the mountain had descended, removed the stones to supply materials for some rustic edifice: the light of the sun darted into the cavern, and the seven sleepers were permitted to awake. After a slumber, as they thought, of a few hours, they were pressed by the calls of hunger, and resolved that Jamblichus, one of their number, should secretly return to the city to purchase bread for the use of his companions. The youth could no longer recognize the once familiar aspect of his native country, and his surprise was increased by the appearance of a large cross triumphantly erected over the principal gate of Ephesus. His singular dress and obsolete language confounded the baker, to whom he offered an ancient medal of Decius as the current coin of the empire; and Jamblichus, on the suspicion of a secret treasure, was dragged before the judge. Their mutual inquiries produced the amazing discovery, that two centuries were almost elapsed since Jamblichus and his friends had escaped from the rage of a Pagan tyrant."*

This legend was received as authentic throughout the Christian world before the end of the sixth century, and was afterwards introduced by Mahomet as a divine revelation into the Koran, and from hence was adopted and adorned by all the nations from Bengal to Africa who professed the Mahometan faith. Some vestiges even of a similar tradition have been discovered in Scandinavia. "This easy and universal belief," observes the philosophical historian of the Decline and Fall, "so expressive of the sense of mankind, may be ascribed to the genuine merit of the fable itself. We

---

* Gibbon, Decline and Fall, chap. xxxiii.

imperceptibly advance from youth to age, without observing the gradual, but incessant, change of human affairs; and even in our larger experience of history, the imagination is accustomed, by a perpetual series of causes and effects, to unite the most distant revolutions. But if the interval between two memorable eras could be instantly annihilated; if it were possible, after a momentary slumber of two hundred years, to display the new world to the eyes of a spectator who still retained a lively and recent impression of the old, his surprise and his reflections would furnish the pleasing subject of a philosophical romance."*

Prejudices arising from our peculiar position as inhabitants of the land.—The sources of prejudice hitherto considered may be deemed peculiar for the most part to the infancy of the science, but others are common to the first cultivators of geology and to ourselves, and are all singularly calculated to produce the same deception, and to strengthen our belief that the course of nature in the earlier ages differed widely from that now established. Although these circumstances cannot be fully explained without assuming some things as proved, which it will be the object of another part of this work to demonstrate, it may be well to allude to them briefly in this place.

The first and greatest difficulty, then, consists in an habitual unconsciousness that our position as observers is essentially unfavourable, when we endeavour to estimate the magnitude of the changes now in progress. In consequence of our inattention to this subject, we are liable to serious mistakes in contrasting the present with former states of the globe.

---

* Gibbon, Decline and Fall, chap. xxxiii.

As dwellers on the land, we inhabit about a fourth part of the surface; and that portion is almost exclusively a theatre of decay, and not of reproduction. We know, indeed, that new deposits are annually formed in seas and lakes, and that every year some new igneous rocks are produced in the bowels of the earth, but we cannot watch the progress of their formation; and as they are only present to our minds by the aid of reflection, it requires an effort both of the reason and the imagination to appreciate duly their importance. It is, therefore, not surprising that we estimate very imperfectly the result of operations thus invisible to us; and that, when analogous results of former epochs are presented to our inspection, we cannot immediately recognize the analogy. He who has observed the quarrying of stone from a rock, and has seen it shipped for some distant port, and then endeavours to conceive what kind of edifice will be raised by the materials, is in the same predicament as a geologist, who, while he is confined to the land, sees the decomposition of rocks, and the transportation of matter by rivers to the sea, and then endeavours to picture to himself the new strata which Nature is building beneath the waters.

Prejudices arising from our not seeing subterranean changes.—Nor is his position less unfavourable when, beholding a volcanic eruption, he tries to conceive what changes the column of lava has produced, in its passage upwards, on the intersected strata; or what form the melted matter may assume at great depths on cooling; or what may be the extent of the subterranean rivers and reservoirs of liquid matter far beneath the surface. It should, therefore, be remembered, that the task imposed on those who study the earth's history requires no ordinary share of discretion; for we are precluded from collating the corresponding parts of the system of things as it exists *now*, and as it existed at former periods. If we were inhabitants of another element—if the great ocean were our domain, instead of the narrow limits of the land, our difficulties would be considerably lessened; while, on the other hand, there can be little doubt, although the reader may, perhaps, smile at the bare suggestion of such an idea, that an amphibious being, who should possess our faculties, would still more easily arrive at sound theoretical opinions in geology, since he might behold, on the one hand, the decomposition of rocks in the atmosphere, or the transportation of matter by running water; and, on the other, examine the deposition of sediment in the sea, and the imbedding of animal and vegetable remains in new strata. He might ascertain, by direct observation, the action of a mountain torrent, as well as of a marine current; might compare the products of volcanos poured out upon the land with those ejected beneath the waters; and might mark, on the one hand, the growth of the forest, and on the other that of the coral reef. Yet, even with these advantages, he would be liable to fall into the greatest errors when endeavouring to reason on rocks of subterranean origin. He would seek in vain, within the sphere of his observation, for any direct analogy to the process of their formation, and would therefore be in danger of attributing them, wherever they are upraised to view, to some "primeval state of nature."

But if we may be allowed so far to indulge the imagination, as to suppose a being entirely confined to the nether world—some "dusky melancholy sprite," like Umbriel, who could "flit on sooty pinions to the central earth," but who was never permitted to "sully the fair face of light," and emerge into the regions of water and of air; and if this being should busy himself in investigating the structure of the globe, he might frame theories the exact converse of those usually adopted by human philosophers. He might infer that the stratified rocks, containing shells and other organic remains, were the oldest of created things, belonging to some original and nascent state of the planet. "Of these masses," he might say, "whether they consist of loose incoherent sand, soft clay, or solid stone, none have been formed in modern times. Every year some part of them are broken and shattered by earthquakes, or melted by volcanic fire; and when they cool down slowly from a state of fusion, they assume a new and more crystalline form, no longer exhibiting that stratified disposition, and those curious impressions and fantastic markings, by which they were previously characterized. This process cannot have been carried on for an indefinite time, for in that case all the stratified rocks would long ere this have been fused and crystallized. It is therefore probable that the whole planet once consisted of these mysterious and curiously bedded formations at a time when the volcanic fire had not yet been brought into activity. Since that period there seems to have been a gradual development of heat; and this augmentation we may expect to continue till the whole globe shall be in a state of fluidity and incandescence."

Such might be the system of the Gnome at the very time that the followers of Leibnitz, reasoning on what they saw on the outer surface, might be teaching the opposite doctrine of gradual refrigeration, and averring that the earth had begun its career as a fiery comet, and might be destined hereafter to become a frozen mass. The tenets of the schools of the nether and of the upper world would be directly opposed to each other, for both would partake of the prejudices inevitably resulting from the continual contemplation of one class of phenomena to the exclusion of another. Man observes the annual decomposition of crystalline and igneous rocks, and may sometimes see their conversion into stratified deposits; but he cannot witness the reconversion of the sedimentary into the crystalline by subterranean fire. He is in the habit of regarding all the sedimentary rocks as more recent than the unstratified, for the same reason that we may suppose him to fall into the opposite error if he saw the origin of the igneous class only.

It was not an impossible contingency, that astronomers might have been placed at some period in a situation much resembling that in which the geologist seems to stand at present. If the Italians, for example, in the early part of the twelfth century, had discovered at Amalfi, instead of the pandects of Justinian, some ancient manuscripts filled with astronomical observations relating to a period of three thousand years, and made by some ancient geometers who possessed optical instruments as perfect as any in modern Europe, they would, probably, on consulting these memorials, have come to a conclusion that there had been a great revolution in the solar and sidereal systems. "Many primary and secondary planets," they might say, "are enumerated in these tables, which exist no longer. Their positions are assigned with such precision, that we may assure ourselves that there is nothing in their place at present but the blue ether. Where one star

is visible to us, these documents represent several thousands. Some of those which are now single consisted then of two separate bodies, often distinguished by different colours, and revolving periodically round a common centre of gravity. There is nothing analogous to them in the universe at present; for they were neither fixed stars nor planets, but seem to have stood in the mutual relation of sun and planet to each other. We must conclude, therefore, that there has occurred, at no distant period, a tremendous catastrophe, whereby thousands of worlds have been annihilated at once, and some heavenly bodies absorbed into the substance of others."

When such doctrines had prevailed for ages, the discovery of some of the worlds, supposed to have been lost, (the satellites of Jupiter, for example,) by aid of the first rude telescope invented after the revival of science, would not dissipate the delusion, for the whole burden of proof would now be thrown on those who insisted on the stability of the system from a remote period, and these philosophers would be required to demonstrate the existence of *all* the worlds said to have been annihilated.

Such popular prejudices would be most unfavourable to the advancement of astronomy; for, instead of persevering in the attempt to improve their instruments, and laboriously to make and record observations, the greater number would despair of verifying the continued existence of the heavenly bodies not visible to the naked eye. Instead of confessing the extent of their ignorance, and striving to remove it by bringing to light new facts, they would indulge in the more easy and indolent employment of framing imaginary theories concerning catastrophes and mighty revolutions in the system of the universe.

For more than two centuries the shelly strata of the Subapennine hills afforded matter of speculation to the early geologists of Italy, and few of them had any suspicion that similar deposits were then forming in the neighbouring sea. They were as unconscious of the continued action of causes still producing similar effects, as the astronomers, in the case above supposed, of the existence of certain heavenly bodies still giving and reflecting light, and performing their movements as of old. Some imagined that the strata, so rich in organic remains, instead of being due to secondary agents, had been so created in the beginning of things by the fiat of the Almighty. Others, as we have seen, ascribed the imbedded fossil bodies to some plastic power which resided in the earth in the early ages of the world. In what manner were these dogmas at length exploded? The fossil relics were carefully compared with their living analogues, and all doubts as to their organic origin were eventually dispelled. So, also, in regard to the nature of the containing beds of mud, sand, and limestone: those parts of the bottom of the sea were examined where shells are now becoming annually entombed in new deposits. Donati explored the bed of the Adriatic, and found the closest resemblance between the strata there forming, and those which constituted hills above a thousand feet high in various parts of the Italian peninsula. He ascertained by dredging that living testacea were there grouped together in precisely the same manner as were their fossil analogues in the inland strata; and while some of the recent shells of the Adriatic were becoming incrusted with

calcareous rock, he observed that others had been newly buried in sand and clay, precisely as fossil shells occur in the Subapennine hills. This discovery of the identity of modern and ancient submarine operations was not made without the aid of artificial instruments, which, like the telescope, brought phenomena into view not otherwise within the sphere of human observation.

In like manner, the volcanic rocks of the Vicentin had been studied in the beginning of the last century; but no geologists suspected, before the time of Arduino, that these were composed of ancient submarine lavas. During many years of controversy, the popular opinion inclined to a belief that basalt and rocks of the same class had been precipitated from a chaotic fluid, or an ocean which rose at successive periods over the continents, charged with the component elements of the rocks in question. Few will now dispute that it would have been difficult to invent a theory more distant from the truth; yet we must cease to wonder that it gained so many proselytes, when we remember that its claims to probability arose partly from the very circumstance of its confirming the assumed want of analogy between geological causes and those now in action. By what train of investigation were geologists induced at length to reject these views, and to assent to the igneous origin of the trappean formations? By an examination of volcanos, now active, and by comparing their structure and the composition of their lavas with the ancient trap rocks.

The establishment, from time to time, of numerous points of identification, drew at length from geologists a reluctant admission, that there was more correspondence between the condition of the globe at remote eras and now, and more uniformity in the laws which have regulated the changes of its surface, than they at first imagined. If, in this state of the science, they still despaired of reconciling every class of geological phenomena to the operations of ordinary causes, even by straining analogy to the utmost limits of credibility, we might have expected, at least, that the balance of probability would now have been presumed to incline towards the close analogy of the ancient and modern causes. But, after repeated experience of the failure of attempts to speculate on geological monuments, as belonging to a distinct order of things, new sects continued to persevere in the principles adopted by their predecessors. They still began, as each new problem presented itself, whether relating to the animate or inanimate world, to assume an original and dissimilar order of nature; and when at length they approximated, or entirely came round to an opposite opinion, it was always with the feeling, that they were conceding what they had been justified *à priori* in deeming improbable. In a word, the same men who, as natural philosophers, would have been most incredulous respecting any extraordinary deviations from the known course of nature, if reported to have happened *in their own time*, were equally disposed, as geologists, to expect the proofs of such deviations at every period of the past.

I shall now proceed to enumerate some of the principal difficulties still opposed to the theory of the uniform nature and energy of the causes which have worked successive changes in the crust of the earth, and in the condition of its living inhabitants. The discussion of so important a question on the present occasion may appear premature, but it is one which

naturally arises out of a review of the former history of the science. It is, of course, impossible to enter into such speculative topics, without occasionally carrying the novice beyond his depth, and appealing to facts and conclusions with which he is as yet unacquainted; but his curiosity cannot fail to be excited by having his attention at once called to some of the principal points in controversy, and after reading the second, third, and fourth books, he will return, it is hoped, to these preliminary essays with increased interest and profit.

Proofs of former revolutions in climate, as deduced from fossil remains have afforded one of the most popular objections to the theory which endeavours to explain all geological changes by reference to those now in progress on the earth. The probable causes, therefore, of fluctuations in climate, will first be treated of in the following chapters.

The Great Scablands Debate

Gould, S.J.
"The great scabland debate"
Natural History
American Museum of Nat. Hist.
Used by permission

How could so vast an area of eastern Washington change so much in so little time?

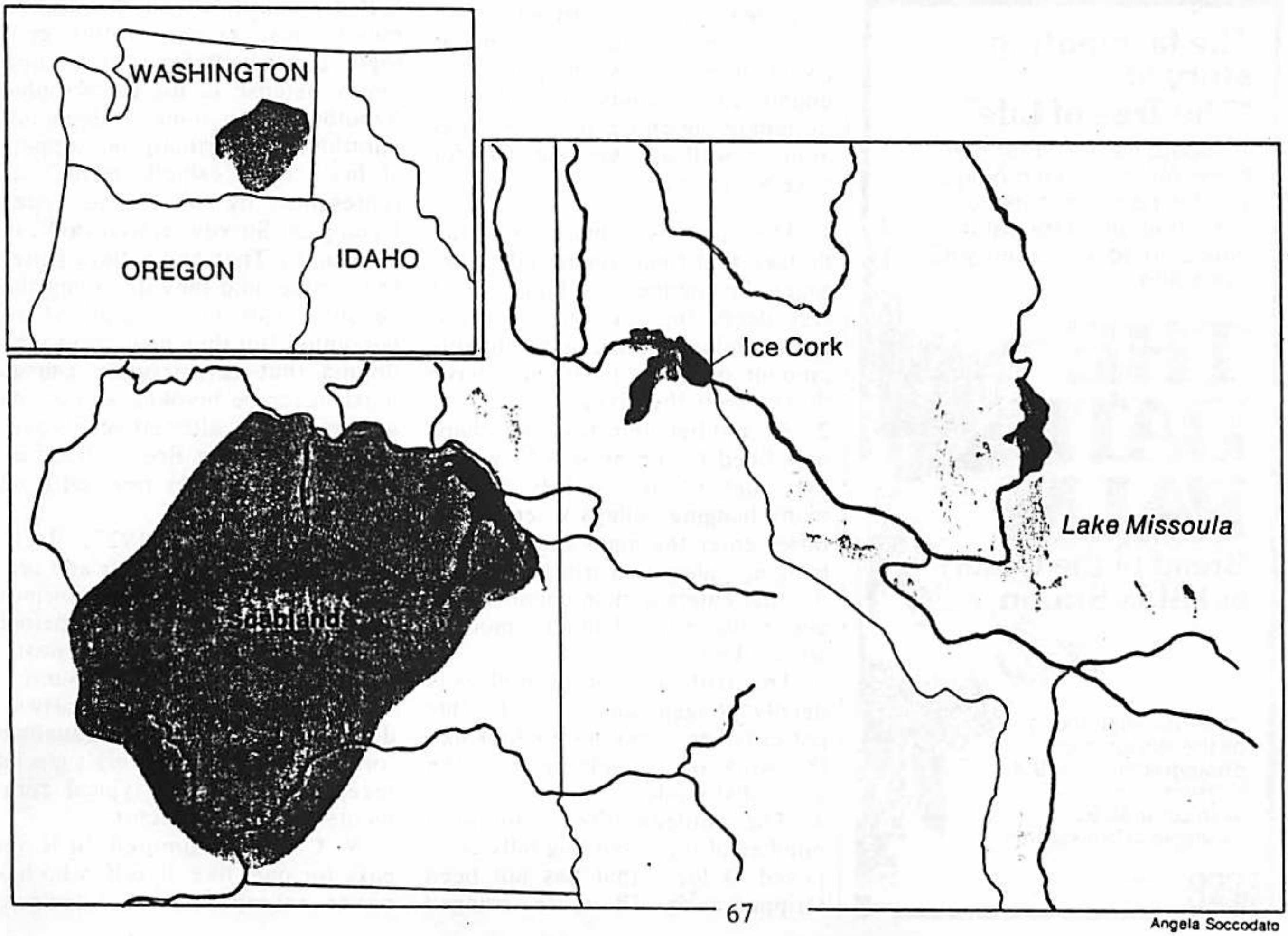
The introductory paragraphs of popular guidebooks usually tout prevailing orthodoxy in its purest form—dogma unadulterated by the "howevers" of professional writing. Consider the following from our National Park Service's auto tour of Arches National Park:

> The world and all it contains is in a continuous process of change. Most of the changes in our world are very tiny and so escape our notice. They are real, however, and over an immense span of time their combined effect is to bring about great change. If you stand at the base of a canyon wall and rub your hand on the sandstone, hundreds of grains of sand are dislodged. It seems like an insignificant change, but that's how the canyon was formed. Various forces have dislodged and carried away grains of sand. Sometimes the process is "very fast" (as when you rub the sandstone) but most of the time it is much slower. If you allow sufficient time, you can tear down a mountain or create a canyon—a few grains at a time.

As the primary lesson of geology, this pamphlet proclaims that big results arise as the accumulated effect of tiny changes. Time, geology's inexhaustible resource, performs all the miracles.

Yet, when the pamphlet turns to details, we encounter a different scenario for erosion in Arches. We learn that a balanced rock known as "Chip Off the Old Block" fell during the winter of 1975–76. Before and after photographs of the magnificent Skyline Arch receive the following commentary: "It remained thus for as long as man knew the arch, until, late in 1940, the block of stone fell, and Skyline was suddenly twice its former size." The arches form by sudden, intermittent collapse and toppling, not by imperceptible removal of sand grains. Yet gradualist orthodoxy is so entrenched that the authors of this pamphlet failed to note the inconsistency between their own factual account and the stated theory of their introduction.

I have questioned the gradualist



Angela Soccodato

orthodoxy before in this column (May and June-July 1977), in general and biological terms. I have argued that gradualism is a culturally conditioned prejudice, not a fact of nature, and I have made a plea for pluralism in concepts of rate. Punctuational change is at least as important as imperceptible accumulation.

In this column, I tell a local, geologic story. But it conveys the same message—that dogmas play their worst role when they lead scientists to reject beforehand a counterclaim that could be tested in nature.

Flow basalts of volcanic origin blanket most of eastern Washington. These basalts are often covered by a thick layer of loess, a fine-grained, loosely packed sediment blown in by winds during the ice ages. In the area between Spokane and the Snake and Columbia rivers to the south and west, many spectacular, elongate, subparallel channelways are gouged through the loess and deeply into the hard basalt itself. These coulees, to use the local name, must have been conduits for glacial meltwaters, for they run down gradient from an area near the southern extent of the last glacier into the two major rivers of eastern Washington. The channeled scablands—as geologists designate the entire area—are puzzling as well as awesome, and for several reasons:

1. The channels connect across tall divides that once separated them. Since the channels are hundreds of feet deep, this extensive anastomosis indicates that a prodigious amount of water must once have flowed over the divide.
2. As another item favoring channels filled to the brim with water, the sides of the coulees contain many hanging valleys where tributaries enter the main channels. (A hanging valley is a tributary channel that enters a main channel high above the main channel's modern stream bed.)
3. The hard basalt of the coulees is deeply gouged and scoured. This pattern of erosion does not look like the work of a gentle river in the gradualist mode.
4. The coulees often contain a number of high-standing hills composed of loess that has not been stripped away. These are arranged as if they were once islands in a gigantic braided stream.
5. The coulees contain discontinuous deposits of basaltic stream gravel, often composed of rock foreign to the local area.

Just after World War I, Chicago geologist J Harlen Bretz advanced an unorthodox hypothesis to account for this unusual topography (yes, that's J without a period, and don't ever let one slip in, for his wrath can be terrible). He argued that the channeled scablands had been formed all at once by a single, gigantic flood of glacial meltwater. This local catastrophe filled the coulees, cut through hundreds of feet of loess and basalt, and then receded in a matter of days. He ended his major work of 1923 with these words:

> Fully 3,000 square miles of the Columbia Plateau were swept by the glacial flood, and the loess and silt cover removed. More than 2,000 square miles of this area were left as bare, eroded rock-cut channel floors, now the scablands, and nearly 1,000 square miles carry gravel deposits derived from the eroded basalt. It *was* a debacle which swept the Columbia Plateau.

Bretz's hypothesis became a minor *cause célèbre* within geologic circles. Bretz's stout and lonely defense of his catastrophic hypothesis won some grudging admiration, but virtually no support at first. The "establishment," as represented by the United States Geological Survey, closed ranks in opposition. They had nothing better to propose, and they did admit the peculiar character of scabland topography. But they held firm to the dogma that catastrophic causes must never be invoked so long as any gradualist alternative existed. Instead of testing Bretz's flood on its own merits, they rejected it on general principles.

On January 12, 1927, Bretz bearded the lion in its lair and presented his views at the Cosmos Club, in Washington, D.C., before an assembled group of scientists, many from the Geological Survey. The published discussion clearly indicates that a priori gradualism formed the basis for Bretz's glacial reception. I include typical comments from all detractors.

W. C. Alden admitted "it is not easy for one, like myself, who has never examined this plateau to supply offhand an alternative explanation of the phenomena." Nonetheless, undaunted, he continued: "The main difficulties seem to be: (1) The idea that all the channels must have been developed simultaneously in a very short time; and (2) the tremendous amount of water that he postulates. . . . The problem would be easier if less water was required and if longer time and repeated floods could be allotted to do the work."

James Gilluly, this century's chief apostle of geologic gradualism, ended a long comment by noting "that the actual floods involved at any given time were of the order of magnitude of the present Columbia's or at most a few times as large, seems by no means excluded by any evidence as yet presented."

E. T. McKnight offered a gradualist alternative for the gravels: "This writer believes them to be the normal channel deposits of the Columbia during its eastward shift over the area in preglacial, glacial, and postglacial times."

G. R. Mansfield doubted that "so much work could be done on basalt in so short a time." He also proposed a calmer explanation: "The scablands seem to me better explained as the effects of persistent ponding and overflow of marginal glacial waters, which changed their position or their places of outlet from time to time through a somewhat protracted period."

Finally, O. E. Meinzer admitted that "the erosion features of the region are so large and bizarre that they defy description." They did not, however, defy gradualist explanation: "I believe the existing features can be explained by assuming normal stream work of the ancient Columbia River." Then, more baldly than most of his colleagues, he proclaimed his faith: "Before a theory that requires a seemingly impossible quantity of water is fully accepted, every effort should be made to account for the existing features without employing so violent an assumption."

The story has a happy ending, at least from my point of view, for Bretz was delivered from the lion's lair by later evidence. Bretz's hypothesis has prevailed, and virtually all geologists now believe that catastrophic floods cut the channeled scablands. Bretz had found no adequate source for his floodwaters. He knew that the glaciers had advanced as far as Spokane, but neither he nor anyone else could imagine a reasonable way to melt so much water so rapidly. Indeed, we still have no mechanism for such an episodic melting.

The solution came from another direction. Geologists found evidence for an enormous, ice-dammed glacial lake in western Montana. This lake emptied catastrophically when the glacier retreated and the dam broke. The spillway for its waters leads right into the channeled scablands.

Bretz had presented no really direct evidence for deep, surging water. Gouging might have proceeded sequentially, rather than all at once; anastomosis and hanging valleys might reflect filled coulees with gentle, rather than raging, flow. But when the first good aerial photographs of the scablands were taken, geologists noticed that several areas on the coulee floors are covered with giant stream bed ripples, up to 22 feet high and 425 feet long. Bretz, like an ant on a Yale bladderball, had been working on the wrong scale. He had been walking over the ripples for decades but had been too close to see them. They are, he wrote quite correctly, "difficult to identify at ground level under a cover of sagebrush." Observations can only be made at appropriate scales.

Hydraulic engineers can infer the character of flow from the size and shape of ripples on a stream bed. V. R. Baker estimates a maximum discharge of 752,000 cubic feet per second in the scabland flow channels. Such a flood could have moved 36-foot boulders.

I could end here with a cardboard version of the story much to my liking: Perceptive hero suppressed by blinded dogmatists stands firm, expresses his allegiance to fact over received opinion, and eventually prevails by patient persuasion and overwhelming documentation. The outline of this tale is surely valid: gradualist bias *did* lead to an out-of-hand rejection of Bretz's catastrophic hypothesis, and Bretz (apparently) was right. But, as I read through the original papers, I realized that this good guy–bad guy scenario must yield to a more complex version. Bretz's opponents were not benighted dogmatists.

They did have a priori preferences, but they also had good reasons to doubt catastrophic flooding based on Bretz's original arguments. Moreover, Bretz's style of scientific inquiry virtually guaranteed that he would not triumph with his initial data.

Bretz proceeded in the classic tradition of strict empiricism. He felt that adventurous hypotheses could only be established by long and patient collecting of information in the field. He eschewed theoretical discussion and worried little about the valid conceptual problem that so bothered his adversaries: Where could so much water come from so suddenly?

Bretz tried to establish his hypothesis by toting up evidence of erosion in the field, piece by patient piece. He seemed singularly uninterested in finding the missing item —a source for the water—that would render his story coherent. For this attempt might involve speculation without direct evidence, and Bretz relied only upon fact. When Gilluly challenged him on the absence of a source for the water, Bretz simply replied: "I believe that my interpretation of channeled scabland should stand or fall on the scabland phenomena themselves."

But why should an opponent be converted by such an incomplete theory? Bretz believed that the southern end of the glacier had melted precipitously, and no scientist could imagine a way to melt ice so quickly. (Bretz tentatively suggested volcanic activity under the ice, but quickly abandoned the theory when Gilluly attacked.) Bretz stayed in the scablands, while the answer sat in western Montana. Glacial Lake Missoula had been in the literature since the 1880s, but Bretz did not make the connection—he was working in other ways. His opponents were right. We still do not know a way to melt so much ice so quickly. But the premise shared by all participants was wrong: the source of the water was water.

Events that "cannot happen" according to received wisdom rarely gain respectability by a simple accumulation of evidence for their occurrence; they require a mechanism to explain how they *can* happen. Early supporters of continental drift ran into the same difficulty that Bretz encountered. Their evidence of faunal and lithological similarities between continents now widely separated strikes us today as overwhelming, but it failed in their time because no reasonable mechanism had been proposed for moving continents. The theory of plate tectonics has since provided a mechanism and established the idea of continental drift.

Moreover, Bretz's opponents did not rest their case entirely on the unorthodox character of Bretz's hypothesis. They also marshaled some specific facts on their side, and they were partly right. Bretz originally insisted upon a single flood, while his opponents cited much evidence to show that the scablands had not formed all at once. We now know that Lake Missoula formed and re-formed several times as the glacial margin fluctuated. In his latest work, Bretz called for eight separate episodes of catastrophic flooding. Bretz's opponents were wrong in inferring gradual change from the evidence of temporal spread: catastrophic episodes can be separated by long periods of quiescence. But Bretz was also wrong in attributing the formation of the scablands to a single flood.

I prefer heroes of flesh, blood, and fallibility, not of tinseled cardboard. Bretz is inscribed on my ledger because he stood against a firm, highly restrictive dogma that never had made any sense: the emperor had been naked for a century. Charles Lyell, the godfather of geologic gradualism, had pulled a fast one in establishing the doctrine of imperceptible change (see my column of February 1975). He had argued, quite rightly, that geologists must invoke the invariance (uniformity) of natural law through time in order to study the past scientifically. He then applied the same term—uniformity—to an empirical claim about rates of processes, arguing that change must be slow, steady, and gradual, and that big results can only arise as the accumulation of small changes.

But the uniformity of law does not preclude *natural* catastrophes, particularly on a local scale. Perhaps some invariant laws operate to produce infrequent episodes of sudden, profound change. Bretz may not have cared for this brand of philosophical waffling. He probably would brand it as vacuous nonsense preached by an urban desk man. But he had the independence and gumption to live by a grand old slogan from Horace, often espoused by science but not often followed: *Nullius addictus jurare in verba magistri,* "I am not bound over to swear allegiance to any master."

My tale ends with two happy postscripts. First, Bretz's hypothesis that channeled scabland reflects the action of catastrophic flooding has been fruitful far beyond Bretz's local area. Scablands have been found in association with other western lakes, most notably Lake Bonneville, the large ancestor of a little puddle in comparison—Great Salt Lake, Utah. Other applications have ranged about as far as they can go. Bretz has become the darling of planetary geologists who find in the channelways of Mars a set of features best interpreted by Bretz's style of catastrophic flooding.

Second, Bretz did not share the fate of Alfred Wegener, dead on the Greenland ice while his theory of continental drift lay in limbo, or of Schubert, buried in a pauper's grave so long before Fischer-Dieskau could sing his songs to an adoring world. J Harlen Bretz presented his hypothesis sixty years ago, but he has lived to enjoy his vindication. He is now well into his nineties, feisty as ever and justly pleased with himself. In 1969, he published a forty-page paper summarizing a half century of controversy about the channeled scablands of eastern Washington. He closed with this statement:

> The International Association for Quaternary Research held its 1965 meeting in the United States. Among the many field excursions it organized was one in the northern Rockies and the Columbia Plateau in Washington. . . . The party . . . traversed the full length of the Grand Coulee, part of the Quincy basin and much of the Palouse-Snake scabland divide, and the great flood gravel deposits in the Snake Canyon. The writer, unable to attend, received the next day a telegram of "greetings and salutations" which closed with the sentence, "We are now all catastrophists."

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

Baker, B.R. and Nummedal, D.
The Channeled Scabland- Field
Conference Guide copyright 1978
Published by N.A.S.A.
Used by permission

The Channeled Scabland

A Guide to the Geomorphology of the
Columbia Basin, Washington
Prepared for the Comparative Planetary Geology
Field Conference held in the Columbia Basin
June 5-8, 1978

edited by

VICTOR R. BAKER
Department of Geological Sciences
The University of Texas at Austin
Austin, Texas 78712

and

DAG NUMMEDAL
Department of Geology
The University of South Carolina
Columbia, S. C. 29208

Sponsored by
Planetary Geology Program
Office of Space Science
National Aeronautics and Space Administration
Washington, D. C. 20546

Preface

As the Mariner and Viking spacecrafts photographed the large channels on Mars, they generated a renewed interest in the erosional effects of large quantities of running water. As the debate continues whether these channels were caused by catastrophic floods, wind, lava or mudflows, it is imperative to examine, in the field, the morphologic criteria which appears to be diagnostic of catastrophic flooding. Hence this field conference to the Channeled Scabland.

Although we now propose to use the Channeled Scabland as a testing area for the Martian flood hypotheses, the flood origin of the scabland itself has a history of intense controversy. Ironically, many of the arguments against catastrophic floods on Mars are the same as those advanced against Bretz' "outrageous hypothesis" for the scabland origin. These include the source of such large quantities of water, the mechanism of its sudden release and the magnitude and nature of subfluvial erosion. Also, some of the alternative fluids proposed for channel generation on Mars were advanced in the debate about the origin of the Channeled Scabland.

THE CHANNELED SCABLAND intends to provide a thorough description of the morphology of the region and a discussion of its inferred cause, including the hydrodynamics of high-velocity flood erosion. It is designed as a guide, written specifically for the Comparative Planetology Field Conference in the Columbia Basin in June, 1978. The editors hope, however, that the book will provide a useful guide also for future field trips and an introduction to the area for geologists who intend to launch their own studies on specific attributes of this fascinating landscape.

The book is dedicated to J Harlen Bretz in recognition of his pioneering advocacy of the flood origin of the scablands. It is also an expression of our appreciation of the many other scientists who directly or indirectly have contributed to our understanding of the scabland origin. The ideas on the Channeled Scabland set forth in this book have matured through discussions with numerous colleagues, including Jon C. Boothroyd, William C. Bradley, J Harlen Bretz, Robert K. Fahnestock, Roald Fryxell, Harold E. Malde, George E. Neff, Peter C. Patton, Russell G. Shepherd, Richard B. Waitt, Jr. and Paul L. Weis.

The field studies and analytical work on the problem involved the assistance of Pauline M. Baker, Frances A. Heaton, R. Craig Kochel, and John M. MacGregor. Field work by V. Baker was initially supported by National Science Foundation Grant GA-21478. Subsequent studies were supported by NASA Grants NGR 44-012194 and NSG-7326, Planetology Program Office, and by The Geology Foundation, The University of Texas at Austin. Field studies by D. Nummedal have been supported by NASA Grant NSG-7076, Planetology Program Office, and the University of South Carolina.

P. Jeffrey Brown and Duane T. Eppler, the University of South Carolina, provided great help in guidebook editing, logistics planning and coordination of the field trip.

We extend our sincere appreciation to John S. Shelton for providing numerous illustrations for the book and to Henry D. Neumann for assistance in obtaining the photos for the color plates.

Neither the book nor the field trip would have been possible without the enthusiastic support by Stephen E. Dwornik, translated into financial assistance through grants to the participating principal investigators in NASA's planetary geology program.

We also thank Priscilla Ridgell and Carleen Sexton for the typing, Burk Scheper for the phototechnical work and Nanette Muzzy for the graphics.

D.N.
V.R.B.

Biographical Sketch of J Harlen Bretz

J Harlen Bretz was born on September 2, 1882, in Saranac, Michigan. He attended Albion College, where he studied biology and met classmate Fanny Challis, whom he married in 1906. As a high school teacher in Flint, Michigan and later Seattle, Washington, he developed an avid interest in geology. His spare time was consumed with studies of glacial geology in the Puget Sound region. His hobby became his profession when the Puget Sound study became the basis for a Ph.D. summa cum laude at the University of Chicago. The entire dissertation was published (Bretz, 1913) and is distinguished for his revisions of the area's glacial history which had been described earlier by Bailey Willis and George Otis Smith. Bretz was also the first to recognize the significance of the great Osceola mudflow from Mount Rainier.

Bretz next spent a year on the faculty of the University of Washington. Unfortunately, his colleagues there did not share his zeal for field work, and Bretz accepted an invitation from R. D. Salisbury to return to the University of Chicago. He began teaching a field course at Devil's Lake, Baraboo, Wisconsin. His love for field work soon brought him back to Washington for summer field courses. Starting in the Columbia Gorge between Oregon and Washington, his field course soon moved into the Channeled Scabland as Bretz formulated his flood hypothesis (Baker, Ch. 1, this volume).

While a faculty member at the University of Chicago for over 30 years, 300 students completed their field geology training with "Doc" Bretz. The graduates include E. Dorf, M. K. Hubbert; W. C. Krumbein, J. L. Hough, A. N. Strahler, and H. S. Yoder, to name just a few.

He wrote 20 major articles and monographs on the Channeled Scabland, served as an associate editor of the Journal of Geology and was associated with the Washington, Illinois and Missouri Geological Surveys. In the 1930's, his research turned to physiographic studies in Greenland and the geology of the Chicago region. His two monographs on Chicago contain an ingenious analysis of the draining of Glacial Lake Chicago (a predecessor of modern Lake Michigan). From 1938 to 1961, Bretz established one of the most important American schools of thought on the origin of limestone caverns. His studies of caves in 17 states, Mexico, and Bermuda placed speleology on a firm scientific base. His insights and energy are a major reason for the modern resurgence of karst geomorphic and hydrologic studies in the United States.

Dr. Bretz officially retired from the University of Chicago in 1947. However, his years as Professor Emeritus were nearly as productive scientifically as those on the active faculty. With Leland Horberg, he published the first modern geological analysis of petrocalcic soil formation (caliche). Perhaps the most extensive survey of caves in one state was his book "Caves of Missouri" (1956). In the 1960's, Bretz' monumental analysis of geomorphic history in the Ozarks of Missouri had him questioning the new geomorphic paradigm of "dynamic equilibrium." He is a member of the American Association for the Advancement of Science and a fellow of the Geological Society of America.

Today, at age 95, Dr. Bretz still maintains a vigorous correspondence with former students and colleagues from his home at 2114 Cedar Road, Homewood, Illinois 60430.

Contents

Introduction

J HARLEN BRETZ
Professor Emeritus of Geology
The University of Chicago
Chicago, Illinois 60637

The publication in 1923 of a geomorphic study of "The Channeled Scabland of the Columbia Plateau" in southeastern Washington state launched a controversy that lasted for decades. A map accompanying the paper depicted a pattern of abandoned erosional waterways, many of them streamless canyons (coulees) with former cataract cliffs and plunge basins, potholes and deep rock basins, all eroded in the underlying basalt of the gently southwestward dipping slope of that part of the Columbia Plateau. The pattern of dry stream ways was described as at network, a plexus, an anastomosis; totally unlike any other drainage pattern on earth. A debacle was asked for, the volume of which filled existing normal stream valleys to overflowing. This great flood spilled over former divides, eroding their summits to complete the network. Associated with the enormously enlarged drainage ways in favorable places were similarly huge mounds of stream gravel which the writer called great river bars. Huge stream-rolled boulders occurred in these bars. The boulders were obviously plucked from the columnar basalt bedrock by the postulated high-velocity currents.

The term valley would no longer suffice. The abandoned rock-bound former waterways were called channels, and the entire composite was named "Channeled Scabland." The total area involved was 18 townships wide by 22 townships long (approximately 40,000 km^2).

The Cordilleran ice sheet had advanced to make contact with the heads of the dominant channel ways. Obviously the plexus of channels was in some way genetically related to the northern ice sheet. But where on our planet had glacial wastage produced a record of flooding and the total remaking of existing topography? The writer called for a catastrophic event.

Catastrophism had virtually vanished from geologic thinking when Hutton's concept of "the Present is the key to the Past" was accepted and Uniformitarianism was born. Was not this debacle that had been deduced from the Channeled Scabland simply a return, a retreat to catastrophism, to the dark ages of geology? It could not, it must not be tolerated.

This, the writer of the 1923 article learned when, in 1927, he was invited to lecture on his finding and thinkings before the Geological Society of Washington, D.C. an organization heavily manned by the staff of the United States Geological Survey. A discussion followed the lecture, and six elders spoke their prepared rebuttals. They demanded, in effect, a return to sanity and Uniformitarianism.

The upstart theorist was not upset nor silenced. Despite his knowledge that the country was full of other dissenters to his flood theory, he proceeded to publish more papers on his favorite topic, now named the Spokane Flood. He described other features of the afflicted plateau which he claimed were inexplicable without his flood of glacially derived meltwater. His apostacy would not be corrected as advised by the elders. The one-man rebellion was still alive.

1924 saw two new papers, 1925 saw one more and 1927, three more. One of the three contained the lecture and discussion already noted. Another paper in 1928 was the writer's reply to all alternative hypotheses thus far suggested. Also in 1928, he traced his flood down the Columbia as far as Portland, Oregon, adding a 200 square mile delta in the Willamette Valley.

By 1930, he had found a source for that immense discharge across the plateau. Clark Fork of the Columbia River, draining a large mountainous region of western Montana, had been dammed by the Cordilleran ice sheet at its traverse of the Idaho panhandle. This formed an immense glacial lake with an estimated volume of 500 cubic miles. The lake had been named some years before as Glacial Lake Missoula. The first geologist to describe the lake ironically was one of the six challenging elders in Washington in 1927 and the author of a short paper on problematical features, perhaps glacial in origin, in what came to be known as the Channeled Scabland.

If Lake Missoula had a properly located place for its ice dam and a clear route thence to the Channeled Scabland, then presto, we would have the big problem solved. Missoula's depth at the dam was known from its shorelines to have been 2000 feet, and there was a clear route to Spokane and the Scabland. A catastrophic failure of the dam would release 500 cubic miles of glacially derived water with adequate gradient to Spokane.

Late in the field study, a criterion of undeniable validity for the occurrence of a flood, or several such, came to light. Hidden largely by sagebrush were numerous occurrences of current ripple marks. They were discovered because the U.S. Bureau of Reclamation had taken aerial photographs of the area to be irrigated with Grand Coulee water. Then it became clear that some gravel surfaces, curiously humpy, were covered with giant current ripples. An investigator, standing between two humps, could not see over either one. Indeed, the size of these ripple ridges made them really small hills. Finally came the discovery of giant current ripples in parts of Lake Missoula where, in a catastrophic emptying, strong currents were formed. When investigated by a specialist in fluvial hydraulics and hydrodynamics there could be no question whatever of their origin, no other explanation for their rhythmic patterns than that of bedform development by amazingly deep, swift flood water.

Measurements of records for depths of water and gradients of water surfaces in channels with proper cross sections also yielded amazing figures for velocities. Plucking of channel bedrock to yield and transport huge boulders became understandable when this 1973 paleohydraulic and hydrodynamic study was made.

Back in 1927 geologic opinion still wavered. One man accepted flooding but did it with local iceberg dams in different channels. Another was a die-hard Uniformitarian who in 1938 still argued that no flood had occurred; that the erosional complex on the plateau was made by "leisurely streams no larger than the Snake in flood today." (But that was before the giant ripple marks were discovered.)

"Meticulously detailed" criticisms of the gross errors and assumptions in the 1938 non-flood interpretation appeared when three geologists spent much of the summer of 1952 (published in 1956) in studying chiefly the excavations made by the U.S. Bureau of Reclamation in diverting Columbia River water behind Grand Coulee dam. This water irrigates from a million acres of the Quincy Basin, an extensive fill made by scabland floods. There were repeated burstings of Lake Missoula dams and refillings when later advances of the ice front made other episodes in the lake's history. How many is not yet definitely determined, despite recent and continuing research on this problem.

When the Geological Society of America held its annual meeting in Seattle in November, 1977, and field trips were made to the flood-ravaged Columbia Plateau and valley, guidebooks made clear that the *catastrophic* history of plateau and master river valley is being pursued with enthusiasm.

Chapter 1

The Spokane Flood Controversy

VICTOR R. BAKER
Department of Geological Sciences
The University of Texas at Austin
Austin, Texas 78712

ABSTRACT

The Spokane Flood controversy is both a story of ironies and a marvelous exposition of the scientific method. In a brilliant series of papers between 1923 and 1932, J Harlen Bretz shocked the geological community with his studies of an enormous plexus of proglacial channels eroded into the loess and basalt of the Columbia Plateau, eastern Washington. This region, which he named the "Channeled Scabland," contained erosional and depositional features that were unique among fluvial phenomena. With painstaking field work, before the advent of aerial photographs and modern topographic maps, Bretz documented the field relationships of the region. He argued that the landforms could only be explained as the product of a relatively brief, but enormous flood, which he called the "Spokane Flood." Considering the nature and vehemence of the opposition to this outrageous hypothesis, the eventual triumph of that idea constitutes one of the most fascinating episodes in the history of modern geomorphology.

INTRODUCTION

The inimitable words of J Harlen Bretz (1928c, p. 446) describes the scene in eastern Washington:

> "No one with an eye for landforms can cross eastern Washington in daylight without encountering and being impressed by the "scabland". Like great scars marring the otherwise fair face of the plateau are these elongated tracts of bare, or nearly bare, black rock carved into mazes of buttes and canyons. Everybody on the plateau knows scabland. It interrupts the wheat lands, parceling them out into hill tracts less than 40 acres to more than 40 square miles in extent. One can neither reach them nor depart from them without crossing some part of the ramifying scabland. Aside from affording a scanty pasturage, scabland is almost without value. The popular name is an expressive metaphor. The scablands are wounds only partially healed—great wounds in the epidermis of soil with which Nature protects the underlying rock.
>
> With eyes only a few feet above the ground the observer today must travel back and forth repeatedly and must record his observations mentally, photographically, by sketch and by map before he can form anything approaching a complete picture. Yet long before the paper bearing these words has yellowed, the average observer, looking down from the air as he crosses the region, will see almost at a glance the picture here drawn by piecing together the ground-level observations of months of work. The region is unique: let the observer take the wings of the morning to the uttermost parts of the earth: he will nowhere find its likeness.
>
> Conceive of a roughly rectangular area of about 12,000 square miles, which has been tilted up along its northern side and eastern end to produce a regional slope approximately 20 feet to the mile. Consider this slope as the warped surface of a thick, resistant formation, over which lies a cover of unconsolidated materials a few feet to 250 feet thick. A slightly irregular dendritic drainage pattern in maturity has been developed in the weaker materials, but only the major stream ways have been eroded into the resistant underlying bed rock. Deep canyons bound the rectangle on the north, west, and south, the two master streams which occupy them converging and joining near the southwestern corner where the downwarping of the region is greatest.

Conceive now that this drainage system of the gently tilted region is entered by glacial waters along more than a hundred miles of its northern high border. The volume of the invading water much exceeds the capacity of the existing stream ways. The valleys entered become river channels, they brim over into neighboring ones, and minor divides within the system are crossed in hundreds of places. Many of these divides are trenched to the level of the preexisting valley floors, others have the weaker superjacent formations entirely swept off for many miles. All told, 2800 square miles of the region are scoured clean onto the basalt bedrock, and 900 square miles are buried in the debris deposited by these great rivers. The topographic features produced during this episode are wholly river-bottom forms or are compounded of river-bottom modifications of the invaded and over-swept drainage network of hills and valleys. Hundreds of cataract ledges, of basins and canyons eroded into bed rock, of isolated buttes of the bed rock, of gravel bars piled high above valley floors, and of island hills of the weaker overlying formations are left at the cessation of this episode. No fluviatile plains are formed, no lacustrine flats are deposited, almost no debris is brought into the region with the invading waters. Everywhere the record is of extraordinarily vigorous subfluvial action. The physiographic expression of the region is without parallel; it is unique, this channeled scabland of the Columbia Plateau."

A mere glance at a modern LANDSAT photograph of the Channeled Scabland (Fig 1.1)

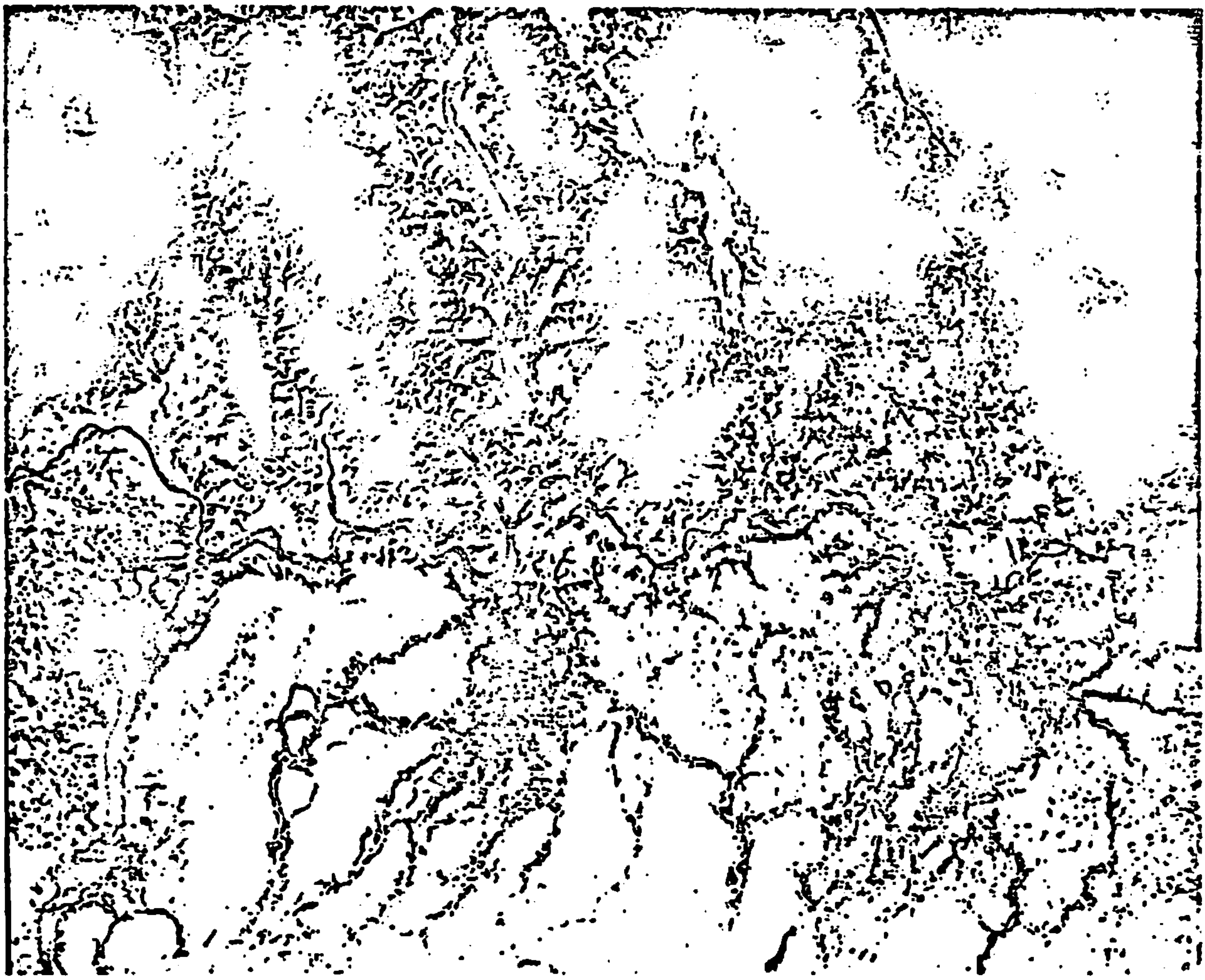


Figure 1.1. LANDSAT photograph of the northern part of the Channeled Scabland. Scabland channels form the dark-toned anastomosis that contrasts with the wheat farms on the light-toned Palouse loess. The Columbia and Spokane Rivers occur at the top (north) of the photograph. The far left scabland complex is the Grand Coulee-Hartline Basin-Lenore Canyon tract. At the center is the Telford-Crab Creek Scabland complex. At the right (east) is the Cheney-Palouse scabland tract (LANDSAT E-1003-18150 composite, 26 July 1972).

will show the features that Bretz, studying from the ground, developed as the basis of his flood hypothesis. The extensive wheat cultivation on the loess presents a vivid contrast to the flood-scared basalt exposed in the channel ways.

The unique character of the dry river courses ("coulees") of the Channeled Scabland was appreciated by the first scientific observers of the region. Rev. Samuel Parker (1838) provided the first published statement on the Grand Coulee: "[it] was indubitably the former channel of the river [Columbia]." Lieutenant T. W. Symons (1882) of the U.S. Army traversed the Grand Coulee, stating that he, "went north through the coulee, its perpendicular walls forming a vista like some grand old ruined roofless hall, down which we traveled hour after hour." Symons (1882) initiated the widely held notion that during glacial episodes of the Pleistocene the Columbia had simply been diverted across the Columbia Plateau. Variations on this general theme were standard in the early literature (Russell, 1893; Dawson, 1898; Salisbury, 1901; Calkins, 1905).

The Grand Coulee gained international fame in 1912 when it was traversed by the American Geographical Society's Transcontinental Excursion. Karl Oestreich (1915) of the University of Utrecht described the coulee as "eines mächtigen Flusses Bett . . . ohne jede Spur von Zerfall der frischen Form." He provided an excellent description of significant features that required a special origin: exhumed granite hills, perpendicular walls, and the hanging valleys marginal to the upper Coulee. He ascribed these hanging valleys to glacial erosion and to deepening of the coulee by the glacial Columbia River. Moreover, he recognized that the upper Grand Coulee was carved through a preglacial divide, which he correctly located just north of Coulee City.

Another foreign observer on the American Geographical Society excursion was H. Baulig, University of Rennes. Baulig (1913) described the loess, coulees, dry falls ("cataracte desséchée de la Columbia"), rock basins, and plunge pools. The origin of these features was ascribed to a glacial diversion of the Columbia. Nevertheless, he marveled at the scale of erosion (Baulig, 1913, p. 159): "peutêtre unique du relief terrestre,—unique par ses dimensions, sinon par son origine."

Dr. O. E. Meinzer, the eminent hydrologist of the U.S. Geological Survey, took an early interest in the western part of the Channeled Scabland He observed (Meinzer, 1918) that the glacially diverted Columbia at Grand Coulee "cut precipitous gorges several hundred feet deep, developed three cataracts, at least one of which was higher than Niagara, . . . and performed an almost incredible amount of work in carrying boulders many miles and gouging out holes as much as two hundred feet deep." He implies that the great erosion occurred because the Columbia River was diverted across the steeply dipping basalt surface of the northern Columbia Plateau.

It was not until the studies of J Harlen Bretz (1923-1932) that the scientific study of this region began in earnest. Bretz interpreted the erosional and depositional features of the region as the product of a brief but enormous flood, which he called the Spokane Flood. For geology in the 1920's this was clearly an outrageous hypothesis. Olson (1969) has described the reception of the idea. "During its not always calm history, the story of the development of the Channeled Scabland was thought by some to have brushed beyond the dividing line in flaunting catastrophe too vividly in the face of the uniformity that had lent scientific dignity to interpretation of the history of the earth." The reaction of the scientific community was predictable, "this heresy must be gently but firmly stamped out" (Bretz and others, 1956, p. 961).

AN OUTRAGEOUS HYPOTHESIS

Because the Spokane Flood controversy is so tied to Bretz as its central figure, this review will consider part of his professional career during the years of his formulation of the flood hypothesis. The ensuing debates were not always marked by scientific objectivity, but their recounting is a fascinating example of the triumph of an outrageous hypothesis. Only in the last two decades has the flood hypothesis gained general acceptance. It is a measure of scientific maturity that in current studies of the Channeled Scabland, "the idea, but not the man has become central" (Olson, 1969).

While teaching at the University of Chicago, Bretz began conducting a summer field course in the wilds of the Columbia Gorge between Washington and Oregon. The idea for a study of the Channeled Scabland came during the summer of 1922. As he relates the story, "One summer I was out in Spokane. I saw a section of a topographic map of what is now called the Channeled Scabland, and from that I got the idea" (Quotation from Seattle Times, Sunday Magazine, July 11, 1971, p. 13).

Without the benefit of modern aerial photographs or even adequate topographic map coverage, Bretz began to take parties of advanced students into the region for month-long field studies. The work continued over the next 7 years. He soon revised an earlier notion that a marine submergence had occurred just downstream from the Channeled Scabland (Bretz, 1919). Nevertheless, the erratic granite boulders, which he had used as evidence for the submergence, were scattered about the basalt plateau far beyond the limits reached by Pleistocene glaciation. Bretz (1923a) named the glaciation responsible for these erratics the "Spokane Glaciation."

Although his first paper on the Channeled Scabland (actually the text of an oral presentation to the Geological Society of America) took care not to call upon cataclysmic origins, Bretz (1923a) provided a detailed description of physiographic relationships in the region. An example in his description (Bretz, 1923a, p. 601) of the pre-flood drainage line that was later enlarged to form the lower part of Moses Coulee: "The cliffs here are deeply notched by wide-open V-shaped tributary valleys. . . . These notches give the cliffs a striking resemblance to a series of great rounded gables in alignment. . . . Both widening and deepening in the basalt occurred and the tributaries were left hanging. They have since attained topograhpic adjustment by building large alluvial fans out on the canyon floor." He further noted that prodigious quantities of water were involved in the erosion. Referring to three outlets at the south end of the Hartline Basin (Dry Coulee, Lenore Canyon, and Long Lake Canyon), Bretz (1923a, p. 593-594) states, ". . . these are truly distributary canyons. They mark a distributive or braided course of the Spokane glacial flood over a basalt surface which possessed no adequate pre-Spokane valleys."

Bretz (1923a, p. 603) originally thought that the scabland gravels were organized into terrace remnants. However, after noting that they lacked a "sharp terrace form," this interpretation was quickly modified (Bretz, 1923b, p. 643): ". . . the evidence seems conclusive that all gravel deposits of the scablands are bars, built in favorable situations in the great streams which eroded the channels." With this conclusion he was forced to call upon catastrophic quantities of water. If the bars were over 100 feet in height, even greater water depths were required to form them. The second paper (Bretz, 1923b) also included the first detailed geomorphic map of the entire Channeled Scabland, showing the overall anastomosing pattern assumed by a great flood of water.

Bretz (1923b, p. 624-626) was the first to recognize the streamlined loess hills of the Cheney-Palouse scabland. He described them as follows: "A very striking and significant feature of the steepened slopes is their convergence at the northern ends of the groups to form great prows, pointing up the scabland's gradient. . . . The nose of a prow may extend as a sharp ridge from the scabland to the very summit of the hill. It is impossible to study these prow-pointed loessial hills, surrounded by the scarred and channeled basalt scablands, without seeing in them the result of a powerful eroding agent which attacked them about their bases and most effectively from the scabland's up-gradient direction."

Bretz knew that his interpretation would be controversial. He argued (Bretz, 1923b, p. 621), "All other hypotheses meet fatal objections. Yet the reader of the following more detailed descriptions, if now accepting the writer's interpretation, is likely to pause repeatedly and question that interpretation. The magnitude of the erosive changes wrought by these glacial streams is nothing short of amazing."

Bretz subsequently argued that the rugged scabland of anastomosing channels and rock basins cut into the basalt was the product of subfluvial quarrying. He described this process for the modern Columbia River near The Dalles, Oregon (Bretz, 1924). Moreover, he asserted that only large vigorous streams could produce such forms. The eventual conclusion from these varying lines of evidence was that so much glacial meltwater occupied the pre-existing valleys on the Columbia Plateau that it must have constituted a vast but short-lived flood, the "Spokane flood" (Bretz, 1925, p. 98). The flood spilled across pre-glacial stream divides, eroding the maturely dissected loess topography to form linear channels, and leaving a legacy of scoured loess scarps, hanging distributary valleys, and high-level fluvial deposits. It also built the huge constructional bars of gravel and then subsided so quickly that these bedforms were left almost unmodified by the subsiding water (Bretz, 1925, p. 105).

Bretz (1925) was able to trace the path of the great flood downstream through the Columbia Gorge to its debouchure into the Willamette lowland, where it built the "Portland delta." On this great subfluvial fan he recognized the significance of macroturbulence in accounting for certain flood features: "The Rocky Butte fosse is but the unfilled locus of an eddy caused by downward deflection where the current impinged on the east face of the butte. . . . The dependent terrace to the west was deposited in the slack water below the obstruction" (Bretz, 1925, p. 256).

Bretz (1925) even made the first estimate of the flood discharge. He chose Wallula Gap for this calculation because of the ponding effect of the constriction. His calculated maximum flow rate was 1.9 x 10^6 m^3/s (66.1 x 10^6 cfs), but he noted that this erred toward the low side. Nevertheless, he stated, "it represents the melting of about 42 cubic miles of ice daily" (Bretz, 1925, p. 258). He then notes that the insolation properties of ice and the total available ice mass north of the Channeled Scabland brings the whole concept into doubt. "The writer," he says (Bretz, 1925, p. 259), "has repeatedly been driven to this position of doubt, only to be forced by reconsideration of the field evidence to use again the conception of enormous volume. . . . These remarkable records of running water on the Columbia Plateau and in the valleys of the Snake and Columbia Rivers cannot be interpreted in terms of ordinary river action and ordinary valley development. . . . Enormous volume, existing for a very short time, alone will account for their existence."

Bretz (1925) then speculated on the somewhat obscure conditions that produced the Spokane Flood. He could only think of two possible explanations: (1) a very rapid and short-lived climatic amelioration, and (2) a gigantic glacier burst produced by volcanic activity beneath an ice cap. He noted severe objections to either hypothesis, but held that the great flood had occurred in spite of the problems in accounting for its source.

THE SPOKANE FLOOD DEBATE

In 1927 the Geological Society of Washington, D.C., invited Bretz to give a lecture "Channeled Scabland and the Spokane Flood." It was a purposeful invitation: a veritable phalanx of doubters had been assembled to debate the flood hypothesis. Bretz (1927a) presented the basic outline of his theory to date, citing the detailed field evidence which he could not explain by any hypothesis other than a great flood of water. The first discussant was W. C. Alden, who cautiously warned of the difficulties with the hypothesis. Lacking personal field experience in the region he suggested that the rock basins might be collapsed lava caves, but he realized that the major features indicated stream erosion. "It seems to me impossible that such part of the great ice fields as would have drained across the Columbia Plateau could, under any probable conditions, have yielded so much water as is called for in so short a time. . . . It appears that ice sheets of three distinct stages of glaciation invaded the borders of this region and may have afforded conditions of repeated floodings of much smaller volume" (Alden, 1927, p. 203).

O. E. Meinzer voiced a commonly held view of the Channeled Scabland, ". . . the Columbia River is a very large stream, especially in its flood stages, and it was doubtless still larger in the Pleistocene epoch. Its erosive work in the Grand Coulee . . . appears to me about what would be expected from a stream of such size when diverted from its valley and poured for a long time over a surface of considerable relief that was wholly unadjusted to it" (Meinzer, 1927, p. 207). He argued that the glacially swollen Columbia could have easily cut the Dry Falls and deposited the great gravel fan of the northern Quincy Basin. He described the Quincy Basin as containing an extensive series of terraces. Moreover, the high-level channels were explained by progressive abandonment as the glacial Columbia progressively cut down to lower levels.

One difficulty that Meinzer appreciated from his field work in the Quincy Basin (Schwennessen and Meinzer, 1918) was the fact that four great spillways led out of the region where water had obviously been ponded. Bretz (1923a) had shown that the upper limits of the torrents that poured through these spillways occurred at the same altitudes. Rather than ascribing this coincidence to contemporaneous operation, Meinzer actually published the idea that the spillways had been cut one at a time and subsequent minor earth movements had later brought them to an equivalent altitude. "This recent deformation may account to some extent for channels cut through ridges that can not otherwise be well explained except by assuming excessive depths of flood water" (Meinzer, 1927, p. 208).

E. T. McKnight was also a participant in the Washington discussions. He subsequently suggested (McKnight, 1927) that a glacially diverted Columbia River was a viable alternative to Bretz' hypothesis. In response Bretz (1927b) argued that the great flood channels and bars near Gable Mountain (in the Pasco Basin) were far too large to be ascribed to the Columbia River. He made his position quite clear (Bretz, 1927b, p. 468): "I think I am as eager as anyone to find an explanation for the Channeled Scabland of the Columbia Plateau which will fit all the facts and will satisfy geologists. I have put forth the flood hypothesis only after much hesitation and only when accumulating data seemed to offer no alternative."

Bretz continued to answer various criticisms of his flood hypothesis (Bretz, 1928a, 1928b), and he established some new lines of inquiry into the problem. He (Bretz, 1929) showed that each of the valleys entering the eastern margin of the scabland spillways contained flood deposits emplaced by phenomenally deep water flowing *up* the tributaries away from the scabland channels. Along the Snake River he traced these deposits to beyond Lewiston, Idaho, more than 85 miles upstream from the nearest scabland channel. The conclusion again defied conventional wisdom (Bretz, 1929, p. 509): "Upvalley currents of great depth and great vigor are essential. . . . No descending gradient of the valley floor can be held responsible. The gradient must have existed in the *surface* of that flood. The writer, forced by the field evidence to this hypothesis, though warned times without number that he will not be believed, must call for an unparalleled rapidity in the rise of the scabland rivers." Each subsequent study produced yet another affirmation of the flood theory. Bretz (1930b) writes: "The writer, at least normally sensitive to adverse criticism, has no desire to invite attention simply by advocating extremely novel views. Back of the repeated assertion of the verity of the Spokane Flood lies a unique assemblage of erosional forms and glacial water deposits; an assemblage which can be resolved into a genetic scheme only if time be very short, volume very large, velocity very high, and erosion chiefly by plucking of the jointed basalt."

Among the spectators at the Washington lecture was J. T. Pardee. Pardee (1922) also had written on the origin of the Channeled Scabland. W. C. Alden, who was Chief of Pleistocene Geology, U.S. Geological Survey, had sent Pardee to study the scablands. He published a brief article (Pardee, 1922) proposing that the Cheney-Palouse scablands tract had been created by glaciation of rather unusual character. Bretz later visited Pardee's field locations and found that his "glacial" deposits were flood bars (Bretz, 1974). Correspondence between Alden, Bretz, and Pardee suggests that Pardee was really considering a hypothesis that the scablands might be related to drainage from a large Pleistocene lake that he had studied in the western part of Montana (Fig. 1.2) (Pardee, 1910). It appears that Alden dissuaded him from that idea (Bretz, 1974). In his memorandum of September 25, 1922, to

Figure 1.2. Late Pleistocene strandlines of Lake Missoula at Missoula, Montana. The highest strandlines reach 1280 m (4200 feet).

David White, Chief Geologist of the U.S.G.S., Alden notes of Pardee's work: "... very significant phenomena were discovered in the region southwest of Spokane.... The results so far ... require caution in their interpretation. The conditions warn against premature publication." David White later asked Bretz if he knew what Alden's middle name was. When Bretz replied in the negative, White said, "It's Cautious, Bretz, Cautious."

It seems clear that the source of the great scabland floods was known even as Bretz was struggling to defend his hypothesis to doubters at the Washington meeting. One story has it that during the discussion Pardee leaned over to Kirk Bryan and said, "I know where Bretz' flood came from."

Bretz finally solved the source problem for the Spokane Flood in 1928. Although Harding (1929) without consultation or acknowledgement made the first announcement of Bretz' idea, Bretz (1930a) later published the discovery that scabland flooding resulted from an abrupt failure of the ice dam that retained Glacial Lake Missoula. Bretz (1932a) clearly illustrated the relationship of Lake Missoula to the Channeled Scabland.

James Gilluly was another of those at the Washington meeting who was upset with Bretz' hypothesis. Although he had not studied the Channeled Scabland in the field, he presented an imaginative and persuasive argument for the creation of the unusual landforms by the long-continued erosion of present-sized streams (Gilluly, 1927, p. 203-205). He took exception to a minor point concerning the use of talus heights as time indicators and then attacked the major weak point in the flood hypothesis. At that time the only two explanations offered for achieving the great volumes of flood water were (1) a very sudden climatic amelioration, and (2) subglacial volcanism and a resulting glacier burst. Some simple calculations demonstrate the inadequacy of either explanation in producing the required volumes of water in so short a time. He concluded, in essence, that Occam's razor did not apply to the Channeled Scabland and called for a more complex sequence of adjustments by rivers or floods not much larger than the Columbia. In reply Bretz (1927a) asked whether the lack of a documented source for the flood was proof that the flood had not occurred. He argued that the scabland phenomena themselves required the existence of a great flood.

Aaron Waters (*in* Bretz, 1972) relates that Gilluly was later to change his mind in this matter. Many years after the incident at the Washington Academy of Science Gilluly visited the Channeled Scabland on a field excursion. As he observed the Palouse-Snake divide crossing, a major scabland stream channel, his astonishment changed to a smiling comment, "How could *anyone* have been so wrong?" Nevertheless, the emotion of those days is evinced by the geologists who continued to deny the flood hypothesis and apparently never changed their minds on the matter: W. C. Alden, K. Bryan, W. H. Hobbs, F. Leverett, C. R. Mansfield, J. C. Merrian, O. E. Meinzer, and G. O. Smith.

The published record of the Spokane Flood debate is clear on one major point. Bretz repeatedly asked only that his flood hypothesis be considered not by emotion or intuition, but by the established principles of the scientific method. His detailed paper on the scabland bars contains the most eloquent expression of this plea (Bretz, 1928b, p. 701):

> "Ideas without precedent are generally looked on with disfavor and men are shocked if their conceptions of an orderly world are challenged. A hypothesis earnestly defended begets emotional reaction which may cloud the protagonist's view, but if such hypotheses outrage prevailing modes of thought the view of antagonists may also become fogged.
>
> On the other hand, geology is plagued with extravagant ideas which spring from faulty observation and misinterpretation. They are worse than "outrageous hypotheses," for they lead nowhere. The writer's Spokane Flood hypothesis may belong to the latter class, but it can not be placed there unless errors of observation and direct inference are demonstrated. The writer insists that until then it should not be judged by the principles applicable to valley formation, for the scabland phenomena are the product of river channel mechanics. If this is in error, inherent disharmonies should establish the fact, and without adequate acquaintance with the region, this is the logical field for critics."

THE REVISIONISTS

By the early 1930's the Channeled Scabland problem had become something of a sensation for American geology. Bretz (1932a, 1932b) had published the last of his field results, and he had embarked on new problems in Greenland and Alberta and ground-water studies in the U.S. His monumental but controversial field study was now open to the kind of attack that he himself had so strongly urged—new field studies.

Ira S. Allison (1933) was the first to enter the new foray. His view was not a denial of the Spokane Flood, but a modification. He argued that it was ice, rather than mere volume, that was the critical factor in the flood. He presented detailed evidence for the ponding of flood water all the way from the Columbia River gorge to the Wallula Gap. This ponding was produced ("in spite of the obvious difficulties of such an explanation") by a blockade of ice in the Columbia gorge. The blockade grew gradually headward until it extended into eastern Washington. As water was dammed to higher levels it spilled across secondary drainage divides creating the enigmatic hanging valleys, high-level gravels and widely distributed erratics. One of the key insights of Allison's motivation was in his last sentence, "perhaps this revision will make the idea of such a flood more generally acceptable" (Allison, 1933, p. 722).

Hodge (1934) published a brief interpretation of the Channeled Scabland involving mainly glacial processes. He hypothesized a complicated alternation of ice advances and drainage changes. The basalt was quarried by glacial erosion, and channel complexes in the basalt were produced by the diversion of meltwater streams around blocks of stagnant glacier ice and jams of berg ice. The theory was never adequately supported by published field evidence.

Perhaps the most serious alternative to the Spokane Flood hypothesis was posed by Richard Foster Flint (1938b). In many ways Flint's study is one of the most ironic in the annals of geology. He presented a carefully worded argument that cited a considerable amount of field data. He stated that the scabland gravel was relatively fine: "Gravel coarser than pebble size is common only in the northern part of the tract" (Flint, 1938b, p. 472). This description was combined with the observation of relatively good size sorting and fair to good rounding to suggest, "a picture of leisurely streams with normal discharge" (Flint, 1938b, p. 472). It is obvious from Flint's sedimentological descriptions that he was giving most of his attention to the slackwater facies of the Missoula flood deposits in the various scabland channels.

One of Flint's most important arguments was that the surface form of the scabland deposits was that of "non-paired, stream-cut terraces in various states of dissection" (Flint, 1938b, p. 475). It was an idea that Bretz had introduced (Bretz, 1923a) and subsequently rejected after closer field study. Flint thought that Bretz' revised interpretation of the deposits as constructional bar forms could explain some, but not all of the field relationships. He suggested that a sequence of channel aggradation by normal proglacial outwash was followed by dissection to leave remnants of fill that occasionally resembled bar forms.

Flint (1938b) accepted Bretz's (1928b) arguments that the flood gravel often (1) occurred in the lee of island-like areas, (2) had rounded upper surfaces, and (3) exhibited a parallelism of surface slopes with the dip of underlying foresets. He argued that "terraces" had been extensively dissected by a downstream base level reduction. The "terraces" were prefentially preserved in the lee of island-like areas. In addition, the low precipitation plus the high permeability of the gravel prevented gullying, so the gravel deposits developed rounded slopes by dry creep. Finally, he showed that many of the gravel slopes did indeed truncate the underlying bedding. As specific cases, he argued that Bretz' Willow Creek bar, Staircase Rapids bar, Palouse Canyon bar, Midcanyon bar, and Shoulder bar were all simply terrace remnants. Subsequent studies have shown that three of these bars have prominent giant current ripples on their upper surfaces (Fig. 1.3).

Flint also described multiple scarps and benches on the Palouse loess. Instead of recording the high-water mark of the Spokane Flood (Bretz, 1928b, p. 701), he interpreted these scarps as evidence of lateral planation by proglacial streams. Subsequent studies in the Cheney-Palouse scabland by Patton and Baker (Chap. 6, this volume) reveal that these scarps resulted from differential erosion of Palouse Formation paleosols and from the exposure of calichified gravel underlying local areas of Palouse loess.

Flint traced the coarse scabland deposits downstream into the Pasco Basin. There he found that the deposits changed from sand and gravel to silt and fine sand containing erratic stones. He named the fine-grained facies the "Touchet beds." The deposits had already been described by Bretz (1928a, p. 325-328; 1929, p. 516-536; 1930b, p. 414), who ascribed them to ponded flood water; and by Allison (1933), who ascribed them to water ponded by ice jams. The silts are recognized only to a uniform elevation of about 350 m. The stratification ranges from rhythmic parallel bedding to cut-and-fill. The included erratic stones are granite, basalt, and other crystalline lithologies. Intense folding, fracturing, and clastic dikes imply slumping and sliding of the water-saturated silt on gentle subaqueous slopes. Flint thought that these relationships were most consistent with a large lake, which he proposed was ponded by a landslide dam or glacier ice in the Columbia gorge. Following Symons (1882) he named this water body Lake Lewis.

At this point Flint had the necessary tools to erect his hypothesis. The proglacial meltwater streams of normal discharge overran the northern margin of the Cheney-Palouse tract. This flow was derived from lobes of ice at the heads of the Cheney-Palouse and Telford-Crab Creek scabland tracts. Flint thought water from Lake Missoula (Bretz, 1930a) need not be involved. Instead, he observed that the discharge "was less than that of the Snake River today" (Flint, 1938b, p. 515).

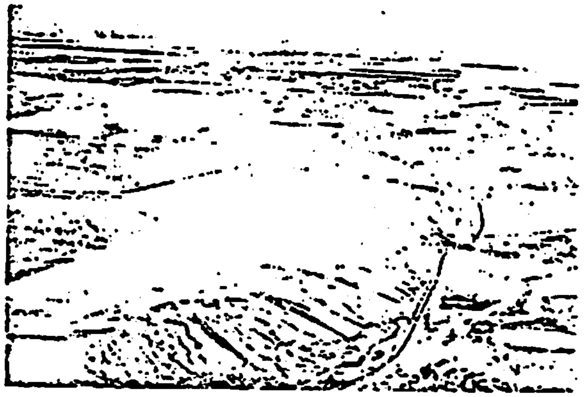


Figure 1.3. Oblique aerial photograph of Staircase Rapids bar. The bar is approximately 50 m high and composed of coarse flood gravel. The giant current ripples on the upper bar surface (left foreground) were actually first described by Flint (1938b) who did not recognize their origin. Bretz and others (1956, p. 1000-1002) later used these and other giant current ripple sets to demonstrate Flint's "faulty reasoning."

As Lake Lewis rose, the "leisurely" streams that Flint envisioned aggraded, forming a thick fill. This fill blocked preglacial tributaries to the Channeled Scabland, such as the Snake River, and formed marginal lakes which accumulated fine-grained sediments. The steep scarps on the Palouse loess were then cut by lateral planation of the streams flowing on this fill. When Lake Lewis finally drained, the streams gradually incised the fill to form terraces. Moreover, Flint was able to explain the enigmatic notched spurs and slotlike hanging canyons as the result of superposition of streams from the widespread fill rather than a consequence of divide crossing by catastrophic flood water.

Flint argued that the complex of anastomosing channel ways cut into basalt was a consequence of erosion by relatively small streams operating on various profiles. He stated that scabland-type erosion should occur wherever rock material with vertical planes of weakness is subjected to stream flow. As examples of such erosion he cited Red Rock Pass, Idaho, an outlet of pluvial Lake Bonneville (Gilbert, 1890). He also noted the scabland erosion at Twin Falls, Idaho, where the Snake River flows in a canyon nearly as spectacular as the scabland channels. He noted, "the . . . [basalt] flows yielded to the hydraulic force of the Snake River as similar flows on the Columbia Plateau yielded to the hydraulic force of proglacial streams, yet I am not aware that unusual floods have been held to have affected the upper Snake River" (Flint, 1938b, p. 492). These words were written 30 years too soon! Malde (1968) described the catastrophic outburst of Lake Bonneville that eroded the scabland forms at Red Rock Pass and Twin Falls.

In yet another ironic passage, Flint (1938b, p. 504-505) calculated the probable rate of filling for Lake Lewis at the modern discharge of the Columbia River. He stated, "the calculated time, 13 years 1 month, seems grossly inadequate for the deposition of the fill in the scabland tracts." He rationalized his interpretation, however, by referring back to the interpreted filling episode. Bretz' flood theory was so despicable that even circular reasoning could be employed to erect an alternative hypothesis.

A careful examination of Flint's (1938b) paper reveals that he observed and described the morphological feature which, more than any other, was absolutely incompatible with his elegant theory. On the surfaces of the scabland "terraces" he described an intricate microtopography of anastomosing channels, small depressions, and crescentic channels (Flint, 1938b, p. 475). In other areas he observed "mamillary undulatory topography." As an example he gives the precise location of the train of giant current ripples on the upper surface of Staircase Rapids Bar, 3 km north of Washtucna (Flint, 1938b, p. 486). Although the ripples that he describes are somewhat masked by overlying slackwater sediments, Flint (1938b, p. 499-500) even states the characteristic ripple magnitude: "The undulations are 20 to 100 feet long, and have amplitudes up to 10 feet. Their axes are generally transverse to the Snake River." How ironic that Flint was the first to accurately describe (without knowing what they were) the very feature that Bretz and others (1956) later presented as incontrovertible evidence for catastrophic flood flows (Fig. 1.3)!

It was Allison (1941) who published the first criticism of Flint's fill hypothesis for the origin of the Channeled Scabland. The first shortcoming noted was that the anastomosing channel patterns and deep rock basins could not have been eroded by "normal" streams. Second, Allison disputed Flint's correlation of the scabland gravels to the Touchet beds, suggesting that the Touchet sequence was younger than the gravels. Third, he agreed with Bretz that the peculiar shapes of the scabland deposits required extraordinary processes. The conclusion was that the complex jamming of various channels with ice was the only reasonable explanation for the unusual drainage patterns and depositional features.

Another example of the strong emotions evoked by the Spokane Flood controversy involves W. H. Hobbs, an eminent glacial geologist from the University of Michigan. He spent several weeks studying the terrain in southeastern Washington and prepared a paper explaining the landforms as the product of a "Scabland Glacial Lobe." Both Bretz and Flint reviewed the paper for the Geological Society of America, and both recommended rejection. The paper was then submitted to the American Philosophical Society, which had supplied part of the funds for the study. Bretz again reviewed the paper, and again it was rejected. Although a brief statement of the hypothesis was published (Hobbs, 1943), the main manuscript had to be published privately (Hobbs, 1947). The author expressed his feelings in the "Foreword" to his paper:

> "In the winter of 1942-43 I was listening with much interest to a lecture on the late geological history of the so-called Scabland area which is southwest of Spokane and close to the supposed southern front of the Pleistocene Cordilleran continental glacier. A map projected on the screen dozens of lakes, none of which transgressed its border, an almost sure indication that this lobate area had once been actually covered by a Pleistocene glacier lobe.
>
> Surrounding this lobe on the lecturer's map could be seen a broad apron of gravels, and enveloping the gravels were heavy deposits of silt. These relationships of glacier lobe to outwash and loess duplicated what I had observed in west Greenland. The lecturer explained, however, that the deposits represented upon his map had been laid down by a great flood of water of unknown origin, the "Spokane Flood."
>
> In the belief that my Greenland observations had given me an advantage in interpreting the evidence within the Scabland region, I then and there decided to make a personal study of it on the ground. Although two other very extended studies had already been made of it by Fellows of the Geological Society of America, and their conclusions had been published *in extenso* in its *Bulletin*, the Society provided me with a grant of money which made possible a new study of the area. This field investigation was carried out during two seasons, and the results and conclusions met with unusually enthusiastic general approval when they were presented to the Society in 1945 at its Pittsburgh meeting. Following tumultuous applause in the crowded section the discussion was throughout approving."

The Hobbs paper contains so many fundamental errors that one marvels at the absurd limits that were being stretched to find an alternative to catastrophic flooding as the cause of the Channeled Scabland. Hobbs (1947) argued that the scabland was a product of glacial scour and that the Palouse loess was deposited contemporaneous to this glaciation by anticyclonic winds off the ice that lay in the various "channels." He interpreted many scabland gravel deposits as moraine remnants modified by glacier-border drainage.

VINDICATION

At long last Pardee (1942) shared his observations of Glacial Lake Missoula that firmly indicated its role as the source of catastrophic floods through the Channeled Scabland. He noted that about 500 cubic miles of water were impounded behind a glacial lobe which occupied the basin of modern Lake Pend Oreille in northern Idaho. Pardee believed that this glacial dam had failed suddenly with a resultant rapid draining of the lake. Evidence for this failure included severely scoured constrictions in the lake basin, huge bars of current-transported debris (Fig. 1.4), and giant current ripple marks with heights of 50 feet and spacings of 500 feet (Fig. 1.5). Lake Missoula was the obvious source for the catastrophic flood flows required by Bretz' hypothetical origin of the Channeled Scabland (Fig. 1.6). Pardee did not state the connection, perhaps leaving that point generously to Bretz. Even Alden remained cautious to the end. His last published report on Lake Missoula observed (Alden, 1953, p. 155): "Abrupt release of water from lowering of the ice dam . . . might result in floods of great magnitude. . . . Each may, *perhaps*, have been the origin of many violent floods that are *supposed* to have swept over the scablands."

In the summer of 1952, Bretz, then nearly 70 years old, returned for his last summer of field work in the Channeled Scabland. The purpose was to investigate new data that had been obtained in surveys for the Bureau of Reclamation's Columbia Basin project. Professor H. T. U. Smith accompanied him, acting in the field as "skeptic for all identifications and interpretations" (Bretz and others, 1956). With the aid of Mr. George E. Neff of the Bureau of Reclamation that study (Bretz and others, 1956) answered with meticulous detail all previous criticisms of the flood hypothesis.

Central to the 1956 investigation was the study of the scabland depositional features. Extensive excavations for the irrigation project and new topographic maps proved that the gravel hills called bars by Bretz (1928b) were indeed that, subfluvial depositional bedforms. Most convincing of all was the presence of giant current ripples on the upper bar surfaces. These showed clearly that bars 30 m high were completely inundated by phenomenal flows of water. Numerous examples of giant current ripples were found on the same bars which Flint had interpreted as terraces. Such features could only have been produced by the flow velocities associated with truly catastrophic discharges. Bretz and others (1956) and Bretz (1959) modified Bretz' earlier interpretations to allow for several episodes of flooding. The central theme of their study, however, was that only a hypothesis involving flooding could account for all the features of the Channeled Scabland. More recent studies of the

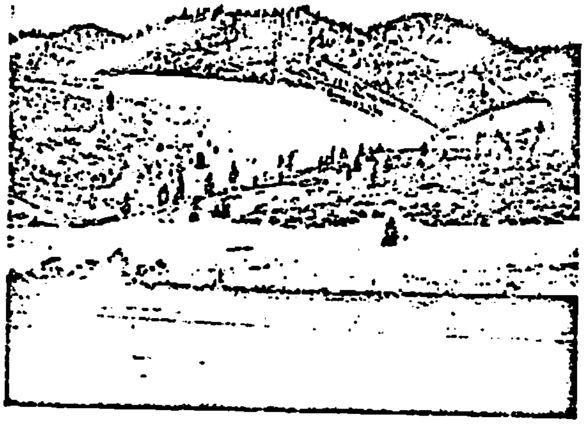


Figure 1.4. Large "gulch fill" formed at the mouth of a tributary canyon along the Flathead River, Perma, Montana. The deposit is an eddy bar (Baker, 1973a) formed during the rapid draining of glacial Lake Missoula. First recognized by Pardee (1942) this gravel deposit was later breached by a small stream to form the V-shaped notch visible at right. The low terrace in the foreground is composed of lacustrine silt.

Figure 1.5. Giant current ripples at Camas Prairie, north of Plains, Montana. The ripples are composed of gravel and consist of ridges up to 15 m high and spaced as much as 200 m apart. The ripples cover approximately 10 km² of the northern Camas Prairie. Faint strandlines of Lake Missoula are visible in the background.

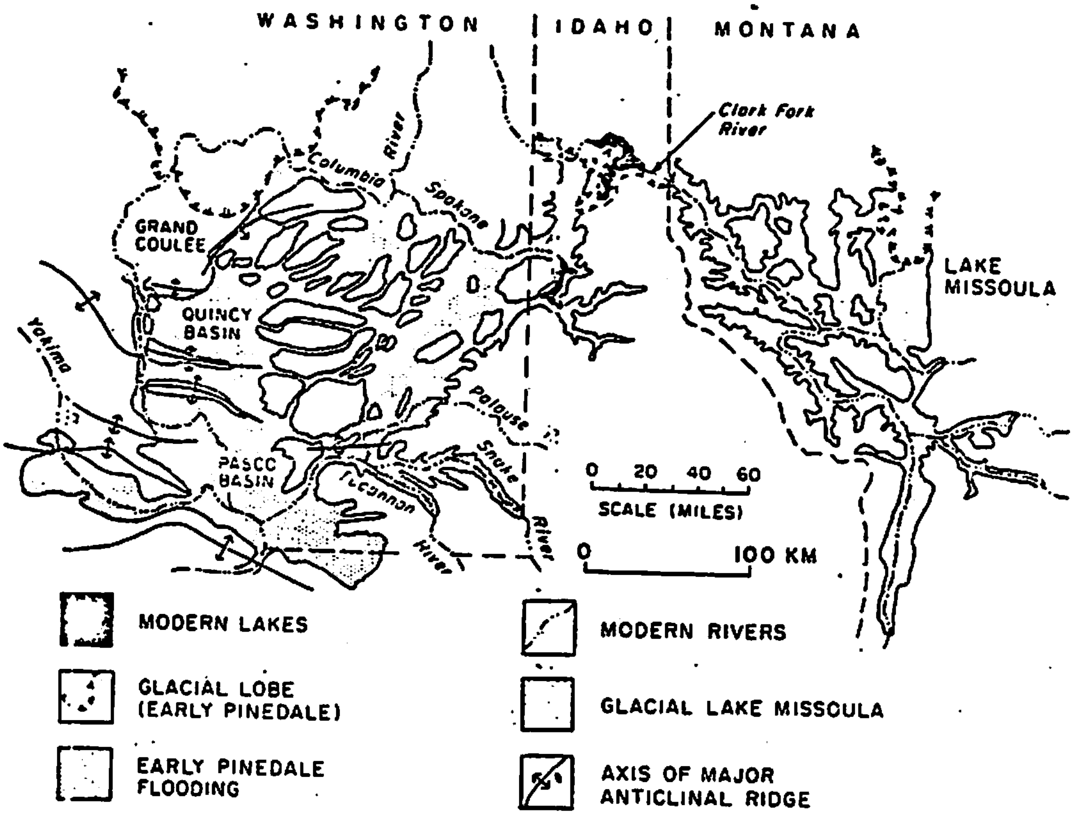


Figure 1.6. Relationship of glacial Lake Missoula to the Channeled Scabland of eastern Washington (Baker, 1973a).

Quaternary geology of eastern Washington have accepted this reasoning (Trimble, 1963; Fryxell and Cook, 1964; Richmond and others, 1965; Baker, 1973a).

Perhaps the final words on the Channeled Scabland controversy were delivered following a field trip, Field Conference E of the 7th Congress, International Association for Quaternary Research. During August, 1965, an international party of geologists observed the evidence in Montana for Lake Missoula's catastrophic outbursts. They then traveled through the Channeled Scabland studying the giant current ripples, flood gravel bars, and scabland erosion forms. Dr. Bretz was unable to attend the trip because of health. When the field party reached Pullman, they sent a long telegram to him at Homewood, Illinois. The telegram opened with "greetings and salutations" and closed with the sentence, "We are now all catastrophists" (Bretz, 1969, 1973).

DISCUSSION

When Bretz published his work on the Channeled Scabland, the paradigm of Geology was uniformity. The Spokane Flood hypothesis appeared to contradict the uniformitarian tradition that made geology a science in the nineteenth century. Indeed it was not until after 1840 that the flood theory fell into serious decline. The catastrophist idea of the Noachian debacle was finally laid to rest when Louis Agassiz showed that his glacial theory could explain erratics, striations, till, fluvioglacial activity, etc. Old ideas die hard, however, and catastrophist absurdities still appeared in the literature of the early 1900's (as they do even today). Little wonder then that Bretz' Spokane flood hypothesis appeared as an anathema to many of his contemporaries.

Simultaneously the Spokane Flood hypothesis established a conflict between two important cornerstones of geological philosophy: (1) the triumph of the glacial theory over diluvian myth, and (2) the scientific tolerance of outrageous hypotheses. It is a classic dilemma for the scientist to distinguish absurdity from outrage. A foolish idea is always self-evident, but not so with the rare, creative insight that happens to pass all reasonable bounds in the consensus of knowledge. The remarks of a former president of our society: "How narrowly limited is the special field, either in subject or locality, upon which a member of the Geological Society of America now ventures to address his colleagues. . . . I wonder sometimes if younger men do not find our meeting rather demure, not to say a trifle dull; and whether they would not enjoy a return to the livelier manners of earlier times . . . (Their) feeling of discouragement must often be shared by the chairman of a meeting when, after his encouraging invitation, 'This interesting paper is now open for discussion,' only silence follows. . . . We shall be indeed fortunate if geology is so marvelously enlarged in the next thirty years as physics has been in the last thirty. But to make such progress violence must be done to many of our accepted principles."

After speaking these words in 1926, William Morris Davis made a case for the value of outrageous geological hypotheses, even suggesting that geologists seriously consider "the Wegener outrage of wandering continents." He concluded by saying that the valuable outrage was that which encouraged the contemplation of other possible behaviors. Such outrages deserve contemplation followed not, he states, "by an off-hand verdict of 'impossible' or 'absurd', but a contemplation delibrate enough to seek out just what conditions would make the outrage seem permissible and reasonable."

Needless to say, W. M. Davis was one of the first to accept Bretz' interpretation in the 1920's. It is a commentary on those years that others were not so tolerant. "During all those years, I was fighting for my professional career." (Quotation of Dr. Bretz by the Seattle Times, July 11, 1971.) Bretz himself explored the consequences of his "outrage." His 1956 paper resoundingly confirmed the catastrophic flood theory by answering in meticulous detail all the previous objections to his grand hypothesis. It took over 30 years and the coming of a new generation of geologists for his theory to gain general acceptance.

The Spokane Flood controversy is both a story of ironies and a marvelous exposition of the scientific method. One cannot but be amazed at the spectacle of otherwise objective scientists twisting hypotheses to give a uniformitarian explanation to the Channeled Scabland. Undoubtedly these men thought they were upholding the very framework of geology as it had been established in the writings of Hutton, Lyell, and Agassiz. The final irony may be that Bretz' critics never really appreciated the scientific implications of Agassiz' famous dictum, "study nature, not books." Perhaps no geologist has understood and lived the spirit of those words more enthusiastically than J Harlen Bretz.

As the Viking spacecrafts were orbiting Mars in the summer of 1976, the cameras were trained on the great Martian channel systems. They revealed uplands streamlined by fluid flow, eroded scabland on the channel floor, and many other features that we now know to be diagnostic of bedrock erosion by catastrophic flooding. Fifty years after J Harlen Bretz' theory of scabland erosion on the Columbia Plateau was being denounced at an infamous meeting of the Washington Academy of Science, Viking scientists were using Bretz' well-documented studies of the Channeled Scabland as the major earth-analog to Martian channel erosion. Few geological concepts, born amid bitter controversy over a half century ago, have continued to have such relevance to our science.

16 March 1979, Volume 203, Number 4385 SCIENCE

Corliss, J.B., et. al.
"Submarine thermal springs on
the Galapgos Rift"
Science
Published by the A.A.A.S
Used by permission

Submarine Thermal Springs on the Galápagos Rift

John B. Corliss, Jack Dymond, Louis I. Gordon,
John M. Edmond, Richard P. von Herzen, Robert D. Ballard,
Kenneth Green, David Williams, Arnold Bainbridge,
Kathy Crane, Tjeerd H. van Andel

ummary. The submarine hydrothermal activity on and near the Galápagos Rift
been explored with the aid of the deep submersible *Alvin*. Analyses of water
iples from hydrothermal vents reveal that hydrothermal activity provides signifi-
or dominant sources and sinks for several components of seawater; studies of
ductive and convective heat transfer suggest that two-thirds of the heat lost from
oceanic lithosphere at the Galápagos Rift in the first million years may be vented
thermal springs, predominantly along the axial ridge within the rift valley. The
areas are populated by animal communities. They appear to utilize chemosyn-
is by sulfur-oxidizing bacteria to derive their entire energy supply from reactions
een the seawater and the rocks at high temperatures, rather than photosynthe-

he presence of active convective cir-
ulation of seawater through newly
ned oceanic crust along the Galá-
os Rift was proposed by Sclater and
gord (*1*). Such circulation was
ngly indicated by profiles of bottom
er temperatures and measurements
eat flow obtained by Williams *et al.*
nd by deep tow observations made
litgord and Mudie (*3*) on the *South-*
expedition in 1972. Water temper-
anomalies over the spreading axis
attributed to plumes of warm water
from hydrothermal vents. System-
fluctuations of heat flow in sedi-
s 5 to 30 kilometers south of the
ding axis, and linear arrays of
ds on these sediments, were inter-
d as resulting from cells of hydro-
al circulation in the underlying ba-
crust. Abundant microearthquake
ity and a fish kill, in which dead benthic fish floated to the surface, were attributed to volcanic eruptions or associated hydrothermal activity (*4*).

The existence of thermal springs associated with volcanism at midocean spreading centers was first suggested 12 years ago (*5*). Early speculations about the process were based on the occurrences of metalliferous deep-sea sediments (*6*), altered midocean ridge basalts (*7*), and anomalous heat flow associated with midocean ridges (*8*). Explicit models for submarine hydrothermal systems were based on features of the chemistry of dredged basalts with contrasting cooling histories (*9*), models of hydrothermal convective cooling of the oceanic crust (*10, 11*), and observations of exposed oceanic crust in ophiolite complexes (*12*). Recently, studies have been conducted on basalt-seawater interaction (*13-17*), physical models of seawater hydrothermal convective systems (*18, 19*), and global budgets of hydrothermal heat and metals (*20-22*). Further studies have been made of metalliferous sediments (*23-25*) and hydrothermal effects in ophiolites (*26-28*), and searches have been made at sea for active hydrothermal systems (*2, 4, 29*).

On the basis of the evidence gathered on the *Southtow* cruise, we proposed a diving program on the Galápagos Rift to the International Decade of Ocean Exploration Office of the National Science Foundation. A detailed bathymetric survey conducted by the U.S. Naval Oceanographic Office with a multibeam sounding (MBS) system (*30*), and results of the *Pleiades* cruise in 1976 (*31-37*), established the area as a prime site for a comprehensive study of midocean ridge hydrothermal activity with a submersible.

In February and March of 1977, we made a series of 24 dives in the deep submersible *Alvin* on the 2.5-kilometer-deep axis of the Galápagos Rift (Fig. 1A). These dives enabled us to make direct visual observations of the area, to make small-scale physical measurements, and to obtain samples of fluids and related deposits at thermal-spring vents on the sea floor. Our goal was not only to prove the existence of hydrothermal plumes by firsthand observation, but also to explore phenomena associated with the convective circulation of seawater beneath the ocean floor. We thus attempted to (i) locate active hydrothermal vents on the sea floor and relate their distribution to the local and regional crustal tectonic fabric; (ii) determine the heat and seawater budgets of hydrothermal systems; (iii) characterize and quantify the fluxes of matter from the oceanic crust into the deep-sea environment; (iv) study the in-

---

J. B. Corliss, J. Dymond, and L. I. Gordon are at the School of Oceanography, Oregon State University, Corvallis 97331. J. M. Edmond is in the Department of Earth and Planetary Sciences, Massachusetts Institute of Technology, Cambridge 02139. R. P. von Herzen, R. D. Ballard, and K. Green are at the Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543. D. Williams is at the U.S. Geological Survey, Federal Center, Denver, Colorado 80225. A. Bainbridge and K. Crane are at the Scripps Institution of Oceanography, La Jolla, California 92037 (K. Crane's present address is Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543). Tj. H. van Andel is in the Department of Geology and Geophysics, Stanford University, Stanford, California 94305.

teraction of these fluids with seawater, and the processes of precipitation, dispersal, and sedimentation of resultant solid phases; (v) characterize the hydrothermal fluids chemically, and to relate these observations to models of the kinetics and thermodynamics of water-rock interaction; and (vi) quantify the fluxes of matter from seawater into the rocks during alteration of the crust.

In the course of our explorations, we discovered extraordinary communities of organisms living in the thermal vent areas at the rift axis and we describe some of these animals later in this article.

Geology of the Rift Valley

During the *Alvin* diving expedition a towed camera sled (*Angus*) was used from the escort vessel RV *Knorr* to obtain about 70,000 color photographs of the bottom, along a series of closely spaced tracks (*38, 39*). Additional color photographs supplemented by visual observations were obtained on 11 dives of the submersible *Alvin*, in four of which we made long traverses of the terrain. The camera sled and submersible were precisely navigated within an acoustic transponder net tied into the 1976 deep tow grid. The precision of the *Alvin* and *Angus* navigation is approximately 10 to 20 meters.

The Galápagos divergent plate boundary at 86°W lies in a small rift valley, about 3 to 4 km wide, 200 to 250 m deep, and trending east-west (Fig. 1B) (*3, 32*). The valley resembles the inner rift of the Mid-Atlantic Ridge at 36°N (*40*) except the latter is contained in a 30-km-wide rift valley, whereas the Galápagos Rift is bordered by the gently outward sloping flanks of the ridge.

The Galápagos Rift valley is bordered on the north and south by steep escarpments produced by normal faulting. The north wall is straight and continuous; the south wall is replaced west of 86°W by a deep elongate trough resulting in a wall offset about 400 m to the south. The floor of the rift valley (Fig. 1B), at a mean depth of about 2450 m, contains a single, straight, east-west trending axial ridge, rising about 20 m above the mean depth of the valley. In addition, there are several long and narrow marginal ridges, similar in height and mean width to the axial ridge. The axial ridge is essentially free of sediment and consists of the youngest volcanics in the area. Toward the north and south, both the sediment cover and the degree of visible surficial alteration of the volcanics increase.

Two principal lava types can be distinguished in the *Angus* photography: pillow and sheet flows (*39*). The sheet flows

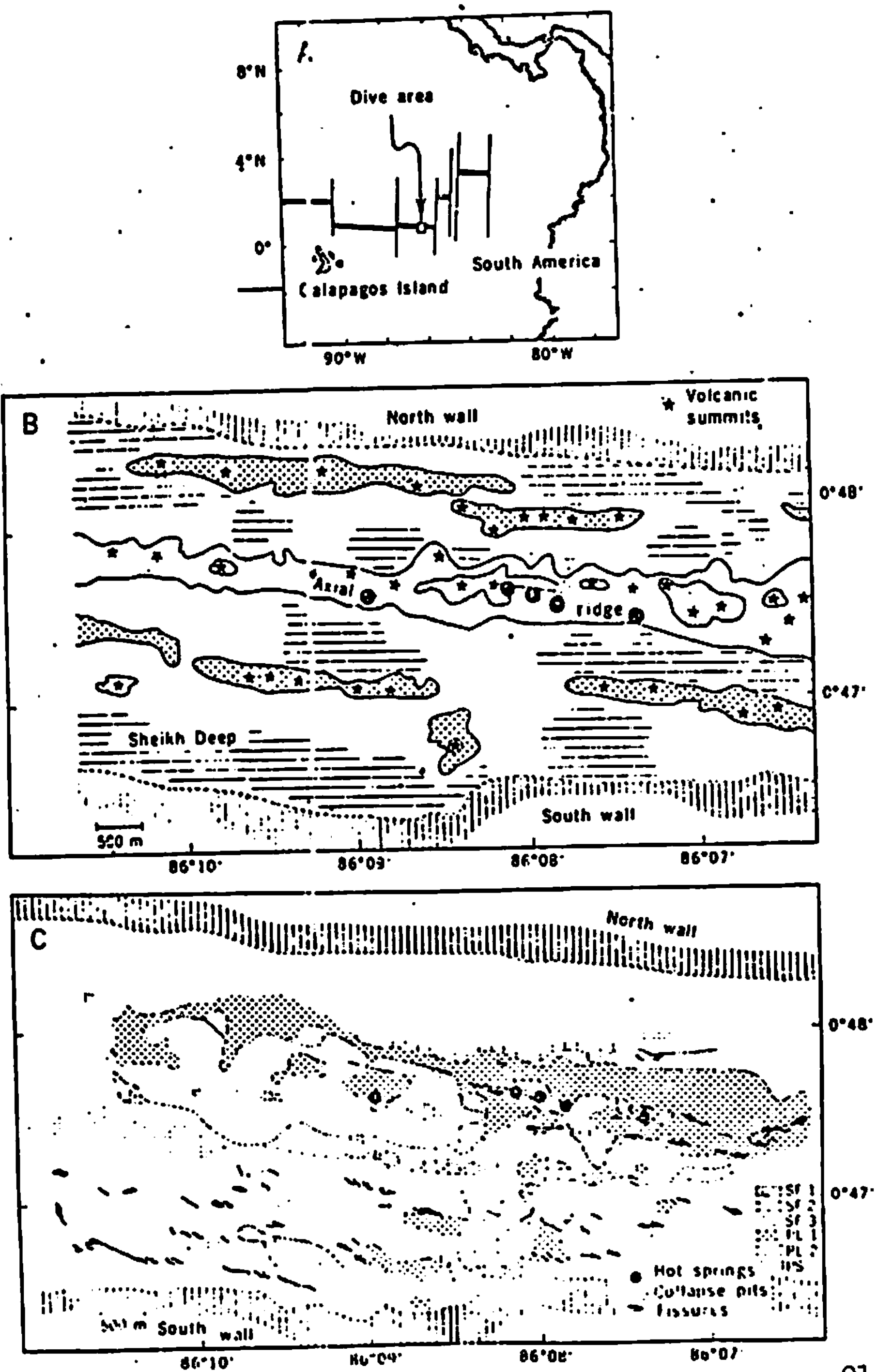


Fig. 1. (A) Location of the survey area. (B) Morphological units of the rift valley of the Galápagos Rift near 86°W, based on analysis of the multibeam bathymetric chart with 2 fathoms (3.65 m) contour interval. Axial ridge and marginal ridges are stippled, and main axial volcanoes are indicated with denser patterns. Marginal deeps are horizontally shaded, while the north and south wall are indicated with vertical bars. The depth of the axial ridge is about 2500 m. The vent areas are named, from west to east, Clambake, Clambake II, Dandelions, Oyster Bed, and Garden of Eden. (Clambake II appeared dead from *Alvin*, but a small area of living animals was found in the *Angus* photographs.) (C) Volcanic units of the rift valley of the Galápagos Rift near 86°W. Symbols: *SF1*, youngest volcanics consisting mainly of fresh glassy sheet flows free of sediment with minor, fresh, pillowed flows; *SF2*, slightly older sheet flow unit with associated pillow lavas, free of sediment; *SF3*, oldest sheet flow units, with altered glass and moderate sediment dusting; *PL1*, pillowed flows forming the main basement of the axial ridge, older than units *SF1* and *SF2*, and possibly approximately contemporaneous with *SF3*, moderately dusted with sediment; *PL2*, even older pillow basalts, not associated with present axial volcanism, possibly preceding sheet flows of *SF3*; and *PS*, marginal volcanic terrain (pillowed and sheet flows) ~75 percent covered with sediment. Geology is based mainly on towed camera runs and is somewhat generalized.

are generally flat and range from smooth to wrinkled and ropy. The pillow terrain shows much more relief, up to 10 m or more, and generally consists of large (up to 1 m or more) pillows. Extensive, blocky lava is common in sheet flows, but also occurs occasionally in pillowed terrain, usually near the flow fronts.

We have recognized a sequence of relative ages on the basis of (i) the degree of alteration of the basalt surface, (ii) the freshness of glass, (iii) the abundance of small glassy buds on the pillows, and (iv) the amount of sediment cover. The spreading rate of the Galápagos Rift, determined from magnetic anomalies, is approximately 35 millimeters per year (1, 3), whereas the sedimentation rate in the region at the ridge depth is about 5 centimeters per 1000 years, according to measurements of sediment thickness on the crust north and south of the spreading axis.

The virtual absence of sediments on the youngest units, which are commonly not even dusted with detritus, suggests a maximum age of less than 100 years. At the inner boundary of the sediment-covered valley margins, the sediment cover is thick enough so that only the tops of the larger pillows emerge, indicating a minimum age of 10,000 years or more at a distance of 500 m or so from the axis, in reasonable accord with a crustal age of about 14,000 years estimated from the spreading rate.

As shown in Fig. 1C, the center and western part of the axial ridge is covered with the youngest rocks; mainly sheet flows (SF 1 to SF 3) associated with a minor amount of equally fresh pillowed flows. The sheet flows surround and lap onto slightly older masses of pillowed terrain that are somewhat altered and have much less glass. The two larger hills on the axial ridge, as well as most summits, consist of this slightly older pillowed terrain that was inundated by a very recent sheet flow, perhaps 10 to 50 years ago.

The fresh sheet flows extend well

Fig. 2 (left). Clambake vent area. The water flows from the vent at ~12°C mixing immediately with ambient bottom water at ~2°C, precipitating in a milky white plume which served as a visual signal of nearby vents to observers in *Alvin*. The large clams and mussels dominate this vent area. The area of active venting is roughly circular, about 50 meters in diameter. The light-emitting diode display in the lower left corner provides depth, gyro heading, dive number, and time. [Photo by *Alvin's* automated camera] Fig. 3 (right). The Dandelions vents. The small spherical animals in the lower left corner have not been identified. Such animals were found in each vent area; at the Garden of Eden, they lived around the periphery of the vent area outside the area of venting water. They suspend themselves from the rocks above the bottom on very fine weblike "legs" not visible in this photograph. The absence here of large filter feeders common to other vent areas is very significant: because there is no hydrogen sulfide in the venting fluids here (presumably because of more extensive subsurface mixing), the sulfur-oxidizing bacteria, which are the normal primary producers supporting the filter feeders, are not present. [Photo by J. B. Corliss]

Fig. 4 (left). The Garden of Eden. This appears to be the youngest area, where water with temperatures to 17°C flows from the fissures bounding flat, lobate pillows of a very fresh-appearing lava flow. Limpets and Pogonophora are well established in these fissures. In this area *Alvin* was surrounded by the shimmering caused by mixing of the warm and cold water. [Photo by *Alvin's* automatic camera] Fig. 5 (right). Oyster Bed vent area. *Alvin's* mechanical arm inserts the water-sampling probe into a cluster of Pogonophora living in the warm water flowing from the rocks. The cancroid crab (center) was common in all vent areas; the round white animals attached to the rocks are members of a new family of limpets. Abundant mussels of the same species as those in Clambake, but smaller, are also present in this vent field. [Photo by J. Edmond] beyond the boundaries of the axial ridge into the marginal depressions. Locally, distinctions can be made between two successive phases of the youngest sheet flows. The older sheet flow sequence south of the axial ridge is marked by numerous collapse pits, some several hundred meters in diameter, up to 20 m deep, and floored with smooth pahoehoe. The lava surface is frequently encrusted with bright yellow efflorescences and closely resembles, in morphology and flow forms, the Halmaumau lava pits on the island of Hawaii. Exposures in the walls of the deeper pits suggest a thickness of the sheet flows in excess of 25 m. Collapse pits also occur in the youngest sheet flows, but these pits are generally smaller.

Fissuring is widespread in the marginal depression and ridges and affects both pillowed and sheet flows, but it is rather uncommon on the axial ridge. The youngest flow units are not fissured at all; fissuring is common only in the slightly older eastern pillow units. The fissures in the rift valley range from hairline cracks to fissures several meters wide; occasionally they are developed as small horst-and-graben structures with vertical offsets of a few meters. It is likely that many of the individual fissure crossings in our photographs form parts of long fissure systems. Crane (32), using a side-scan sonar, described fissure systems with lengths of 500 m to a few kilometers.

Hydrothermal Vents at the Rift Axis

The four active vent areas sampled (see Fig. 1) range in dimension from 30 to 100 m and are all located on or near the axial ridge in pillow basalt slightly older than the youngest sheet flow, near the contact with this unit. In Clambake and Dandelions, the pillows are typical bulbous forms (Figs. 2 and 3), whereas in the Garden of Eden (Fig. 4) the pillows are flat and lobate, transitional to a sheet flow. The Oyster Bed vents (Fig. 5) are located in high-relief, talus-covered terrain produced by faulting.

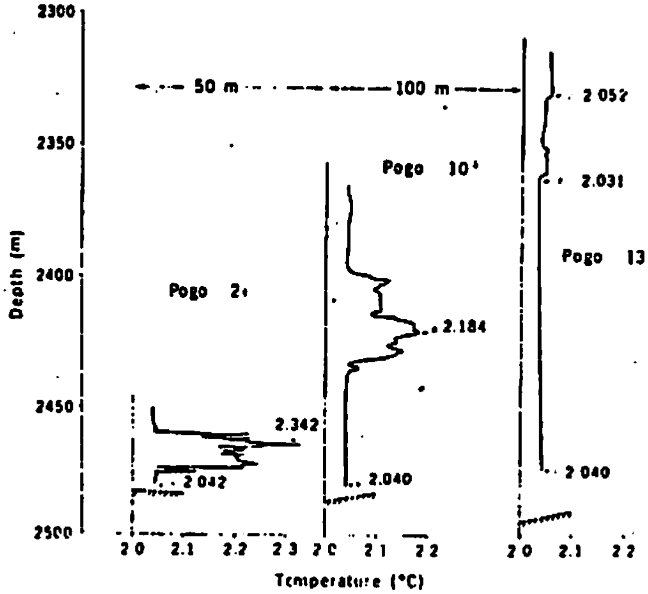
All vent areas are near (within 10 to 50 m) small vertical fault scarps or within grabens with offsets of a meter or so. Small fissures (2 to 10 cm wide) which cut across pillows can be traced for several meters near some vent areas. Large open fissures do not serve as vents for warm water; in numerous crossings of such features no flow of warm water was detected. In all cases, the hydrothermal fluids are streaming up from the normal openings between pillows found in a rough pillowed surface or between talus blocks in Oyster Bed.

The shimmering produced by mixing of the warm water with the cold bottom water was the most striking physical feature of the vents. In addition, a very faint milky white precipitate could be seen forming in the rising fluid less than a meter above the bottom, and the rocks were coated with a thin manganese oxide coating where the warm water was in contact with them. The maximum temperatures measured differed between vent areas, from 7°C in Dandelions to 17°C in the Garden of Eden, compared to ambient bottom water temperatures of 2.0°C.

A long-term temperature recorder with thermistor sensors was left at one of the Clambake vent sites for about 10 days. One sensor, placed in the vent at the sea floor, recorded a relatively steady and high temperature (10.5° ± 1.0°C), over the entire period. Another sensor about 50 cm above the other recorded lower and more variable temperatures ranging from 4° to 6°C with excursions to 8°C, reflecting the turbulent entrainment of 2°C ambient bottom water into the warm water.

Fluid flow rates were measured at three different "vents," two in the Garden of Eden and one in Clambake. The flow rates, estimated by visual observations from *Alvin* on a vane-type flowmeter placed over these individual vents, range from 2 to 10 liters per second. The "vents" were areas within the vent fields where a sufficient flow was concentrated to allow measurement of a vertical current. Within the vent fields, warm water was flowing from essentially every natural opening in the rocks at various rates, making estimates of total discharge based on direct measurements of flow rates difficult.

This warm water rising from the vent areas formed thick layers or plumes of



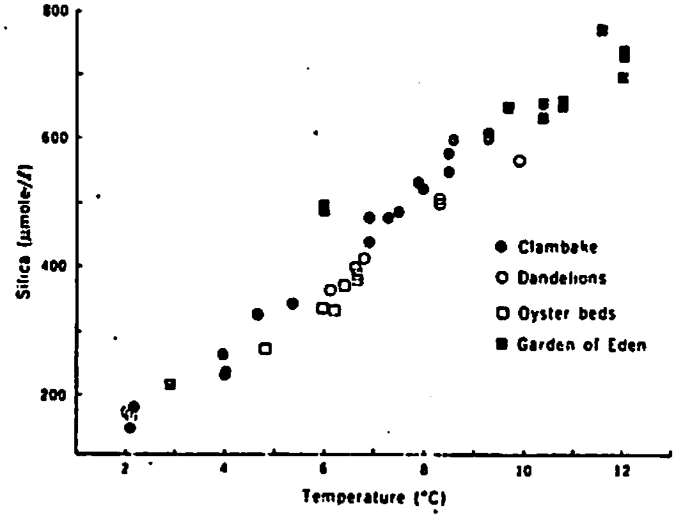


Fig. 6 (left). Clambake plume profiles. An acoustically navigated CTD (a device for measuring conductivity, temperature, and depth) was towed horizontally through the vent area and raised and lowered obliquely to provide a three-dimensional view of the plume. The departure from the vertical of each profile is given and bottom depths are marked for the start and end of each profile. Fig. 7 (right). Silica plotted against temperature: all the ridge crest data are utilized. The ambient values are 165 μmole per liter and 2.04°C. The extension of the mixing line intersects the quartz solubility curve at ~350°C. If the assumptions of the quartz geothermometer are valid, this indicates that the seawater in all vent areas last equilibrated with rocks at depth at this temperature. This "end-member" fluid rises, mixing with descending cold water which, because of its greater density and resulting pressure head, will readily flow laterally into the rising subsurface plume.

varying horizontal and vertical extent which were detected to 180 m above the bottom. These plumes were detected on the *Southtow* and *Pleiades* cruises, and on the diving expedition (*Knorr 64*) by thermistors mounted on the *Angus* camera sled. A transponder-navigated instrument (CTD) which recorded conductivity, temperature, depth, and dissolved oxygen, was towed through the vent areas to map the horizontal and vertical structure of the plumes (Fig. 6). Data of this type will be required to determine the total fluid and thermal flux of individual vent areas.

Chemistry of the Hydrothermal Fluids

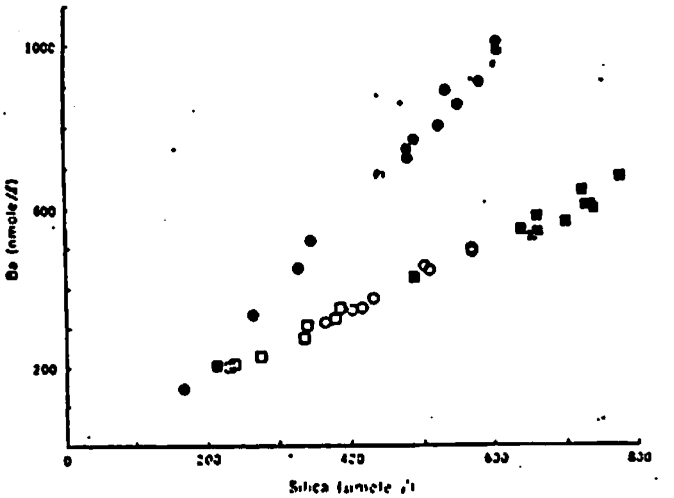
Our *Alvin* sample probe, with a thermistor attached, was inserted into the openings from which the water flowed, and the fluids were sampled by pumping the water through a variety of valves which could direct the flow through filters, if desired, and then into sample bottles. Downstream from the samplers, the water flowed past sensors for measuring temperature, conductivity, dissolved oxygen, and *p*H. These data were displayed and recorded continuously, along with other dive parameters, on magnetic tape within the submarine. Eighty-eight approximately 9-liter samples of hydrothermal fluids and associated ambient water were collected, and more than 2000 subsamples have been distributed for laboratory analysis on shore, to supplement the extensive shipboard analytical program.

The temperature and chemistry of the water samples can be considered the result of two mixing processes. First, the water flowing up through the rocks prior to venting through the sea floor can be considered a mixture of some initial hydrothermal fluid formed by water-rock interactions at depth, and cooler descending seawater entrained with this warm water as it rises through the rocks (subsurface mixing). Second, the water flowing from the rocks immediately begins mixing with ambient seawater, and varying amounts of this ambient water are entrained prior to sampling (surface mixing). The subsurface and surface mixing are essentially one single mixing process involving only two ultimate end-members: (i) the fluid produced at depth through interactions with the rocks and (ii) ambient seawater.

For conservative components (that is, those not added to or removed from the fluids during ascent and mixing), if the subsurface end-member has uniform properties within a single vent area or several vent areas, all samples will fall on a single mixing curve. This is clearly the case for silica in all vent areas (Fig. 7). The scatter may be due to the temperature uncertainty (≈ ± 0.5°C) caused by differing proportions of ambient waters being entrained during the sampling interval. Hence the silica data are probably a better measure of the degree of dilution; certainly graphs in which the other elements are plotted against silica show significantly less scattering than when they are plotted against temperature.

The extrapolation of the silica-temperature line intersects the quartz solubility curve for distilled water at 1000 atmospheres (*41–43*) at about 300°C; it coincides with the curve for opal in the range 50° to 70°C. Although experimental seawater-basalt hydrothermal systems at temperatures below 300°C achieve equilibrium with amorphous silica rather than quartz (*13, 16*), above this temperature the quartz grows in the solution (*14*), and quartz is a common phase in hydrothermally altered oceanic crustal rocks (*44, 45*). It is possible that there is a secondary reaction with the glassy basalts in the upper part of the section. In addition, this extrapolation does not reflect departures from linearity of the enthalpy of seawater above 200°C, or uncertainties in the solubility of silica in seawater at these temperatures. Resolution of these effects may be possible by comparison with other chemical and isotopic measures of solution temperatures.

Other elements have a variety of relationships with temperature and silica. The amount of barium increases strongly in all the vent fields (Fig. 8); however, the Ba/Si gradient in the Clambake area is more than twice that determined for the other three vent fields. The concentration of manganese (*46*) is strongly correlated with barium, with values as high as 31 micromoles per liter. For lithium, by contrast, the Clambake, Garden of Eden, and Dandelions waters fall on one line with Oyster Beds having a lower Li/Si gradient. The values range to more than twice ambient. These variations among vent fields reveal differences in subsurface high-temperature fluid com-



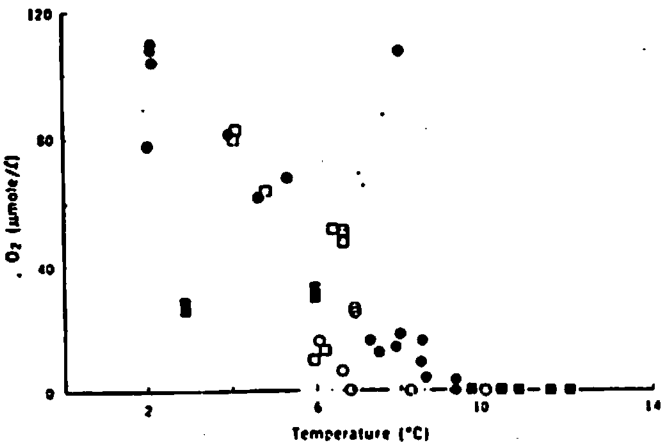


Fig. 8 (left). Barium plotted against silica. The two distinctive mixing curves indicate that the Clambake vent area has a different "end-member" fluid composition from the other vent areas. Clambake is 1000 m from Dandelions, the nearest of the three, which are separated from each other by less than 700 m. Manganese has the same relationship as barium with silica. Such springs are the major source of manganese in the oceans and are an important source of barium. Fig. 9 (right). Oxygen concentration plotted against temperature. The shipboard oxygen data suggest that the water is anoxic as it flows from the rocks, and that the oxygen present in our samples results from surface mixing with ambient oxygenated bottom water. Reaction with hydrogen sulfide during the interval between sampling and analysis may affect this relationship, but preliminary data in situ from the CTD-O_2 system on *Alvin* suggest that this effect is small.

positions that must be related to variations in physical and chemical conditions of the water-rock interactions.

The amount of calcium increases more than 1 percent, the Clambake area again having a higher concentration than the rest. Strontium is constant over all the vents to within the measurement precision (0.25 percent). Magnesium shows a uniform decrease by about 3 percent over the range sampled, in agreement with predictions that magnesium is removed from seawater during hydrothermal circulation in rock-dominated systems (*13*). The magnesium-temperature line extrapolates to a "zero magnesium" temperature of about 280° ± 20°C in agreement with the quartz geothermometer value. Though neither is conclusive, the coincidence of these temperatures tends to support their individual validity, suggesting that the seawater last equilibrated with the rocks at a temperature of about 300°C.

Hydrogen sulfide also shows distinctive distributions in each vent field. None was detected in Dandelions. The Clambake values range up to 20 μmole per liter, Garden of Eden to 60 μmole per liter, and Oyster Beds to 160 μmole per liter. The concentrations of elements which form extremely insoluable sulfides under these conditions are concomitantly low. Data for iron give a range of values equivalent to iron to manganese ratios of from three to several hundred. It appears that nickel, copper, and cadmium are completely depleted in the subsurface fluids and are added only by surface mixing with ambient water. Abundance patterns for the rare earth elements show a strong positive europium anomaly which may reflect reduction of Eu^{3+} to Eu^{2+} and preferential extraction from the rocks under these reducing conditions.

The effects of surface mixing of ambient water and vent water are clearly shown in Fig. 9. Dissolved oxygen from the ambient bottom water often coexists in the samples with hydrogen sulfide in disequilibrium. The trend toward zero oxygen at 8° to 10°C indicates either that the vent water contains essentially no oxygen prior to emerging from the rocks, or that extensive reaction between oxygen and hydrogen sulfide occurs in the samples subsequent to collection. The data from the oxygen probe in the sampling system on *Alvin* minimize the latter possibility since the difference between these data and the shipboard measurements on identical samples is less than 25 μmole per liter. Nitrate, another oxidized species found in disequilibrium with sulfide in the samples, is linearly correlated with oxygen, extrapolating to zero at approximately minus 25 μmole of oxygen per liter, and is presumably supplied by surface mixing with ambient water. Scanning-electron-microscope analysis of material filtered from the vent fluids shows the milky precipitate appearing in the vent fluids to be elemental sulfur, presumably forming by oxidation of hydrogen sulfide.

The acidity of subsurface fluids is reflected in the alkalinity changes during surface mixing. As the solutions mix with ambient water and cools the alkalinity decreases by 25 microequivalents per liter per degree Celsius. The total CO_2 increases by 25 μmole per liter per degree Celsius, indicating a significant flux of CO_2 from the mantle. Isotopic measurements should allow us to quantify the flux more precisely.

The extent of entrainment of descending ambient seawater appears to vary between vent fields, being greatest in Dandelions and probably least in Garden of Eden. The degree of subsurface mixing exerts a profound effect on the redox systems in general and on elements such as iron, copper, nickel, zinc, and cadmium in particular. Transport of these elements in solution in the presence of large amounts of sulfide (in the millimolar range at 300°C) requires acidic, high-temperature conditions (*47*). If, indeed, the iron to manganese ratio of 3 : 1 in metal-rich sediments of the ridge is representative of the average exit value of hydrothermal systems (*6, 23-25*) then hot springs at greater than 100°C may be common, perhaps as short-lived phenomena following volcanic eruptions, involving interaction of seawater and flow interiors and dikes as proposed (*9*). These observations have implications for the formation of massive sulfide deposits by submarine hydrothermal systems.

The extent of inferred subsurface mixing and cooling is more than sufficient to preclude transport of sulfide-forming elements to the sea floor, but is not (except for the case of Dandelions) enough to produce an oxidizing system. In systems such as the Dandelions, since the solutions will be oxidizing and, because of lack of relative buoyancy, quite slow-moving, conditions around the exits should provide the ideal regime for formation of the very pure manganese oxide crusts discovered on the Mid-Atlantic Ridge (*48*) and elsewhere on the Galápagos spreading axis (*49*). Since the manganese is maintained in reduced form in the vent regions sampled and the ascent rates are very rapid, no such crusts were observed.

The helium isotopes have proved useful for understanding the global implications of our observations of these thermal springs. Jenkins *et al.* (*50*) showed that helium exhibits a regular increase proportional to silica and temperature. Combined with the best estimate for the total flux of helium-3 through the oceans to the atmosphere of Craig *et al.* (*51*), and assuming that this flux is all produced from hydrothermal systems at the spreading center with similar relations between helium and heat, we can estimate global hydrothermal flux of heat (*50*). This gives a flux of about 5×10^{19} calories per year, in close agreement with recent estimates based on heat flow (*20, 22, 52*).

The flux of the elements into or out of the ocean has been computed and compared to the river input. Using the same reasoning as for heat flux, we can use the flux of ^3He to calculate these values for elements we have measured. Of course, it is by no means clear that the existing data are representative of all ridge crest hydrothermal systems. However, the results are impressive (Table 1). The ridge crest hydrothermal system is the major sink for magnesium and the major source for sedimentary manganese and for lithium. It provides a substantial proportion of the silica and barium. It is a minor source of calcium and CO_2 and a significant sink for alkalinity. The systems we measured are also sinks for copper, nickel, and cadmium, but the dependence of the redox behavior of the systems on their physical characteristics—that is, flow rates and exit temperatures—makes it difficult to generalize from our data.

The heat to mass ratios for conservative elements also provide insight into the extent of interaction of seawater with the crust. Combined with data for the concentration of these elements in the magma and the heat content of the magma, they provide strong constraints on the depth of water penetration and the extraction of heat, gases, and other rock components (*53*).

Biological Observations in the Vent Areas

Large white objects identified as clamshells appeared in a few of the *Pleiades* deep tow photographs taken along the axial ridge (*54*). Observations from *Alvin* showed these to be evidence of highly productive communities of exceptionally large and densely spaced animals living within the hydrothermal vent areas. These animals provided a valuable indicator of thermal vents in our *Angus* photographs. We collected some of the animals amenable to sampling with the equipment available, primarily large clams and mussels, limpets, and tube worms, and also collected water samples and filtered suspended matter for biological studies (55).

The clams resemble an unusual family, the Vesicomyidae, which is probably allied to the common successful shallow-water veneroid clams such as quahogs. Vesicomyids have a considerable zoogeographic and bathymetric range (that is, greater than 3000 m), and their length may be only a couple of millimeters or up to 150 mm. They were found only in one active vent area, where they are of somewhat uniform large size, from about 150 to 300 mm, and cover large patches of the ocean floor filling the depressions between basalt pillows.

The mussels all appear to belong to the same new genus and species of the family Mytilidae. Their anatomy is distinctly different from shallow-water forms and from other known deepwater forms. The mussels were abundant in the Clambake where they averaged from 100 to 150 mm in length, and in the Oyster Bed where they ranged from 7 to 53 mm in length.

The limpets belong to a new gastropod family; they resemble members of the Calyptraeidae, a group in which abyssal species are unknown. However, there are some unusual anatomical features that indicate an affinity with more primitive groups of mollusks. In the new species, the gills are large and evidently used for filter feeding by ciliary currents, as in the Calyptraeidae. They are commonly attached to rock surfaces not occupied by the mussels in Oyster Bed and are abundant in the Garden of Eden where they line the interpillow fissures from which the warm water flows.

The tube worms belong to the class Vestimentifera of the phylum Pogonophora. The phylum is mostly restricted to the ocean depths. The Vestimentifera are the largest of the Pogonophora and two species of a single genus have been described. The hot-spring specimens (Figs. 4 and 5) represent a new genus of Vestimentifera and are the largest members of the class and phylum, with chitinous tubes ranging up to 338 mm in length and 25 mm in diameter. They possess a plume of fused tentacles that can be extended from the tube, presumably for respiration. They lack a gut or other visible digestive system, a characteristic of the Pogonophora, and live attached to rocks directly in the flow of warm waters from the vents.

In each vent area, we observed and photographed what appeared to be an apparently undescribed species of ophidioid fish swimming in the rising plumes of warm water. This fish was not observed outside the vent areas, and other fish seen outside the vents were not observed inside the active vent fields. Three benthic fish recovered on *Southtow* (4) from more than 80 dead benthic fish observed floating on the surface did not include an ophidioid, but did include species similar to those we photographed outside the vent areas.

Table 1. Hydrothermal flux of various components to the oceans. The global hydrothermal flux for each element is estimated by multiplying the mass to heat ratio in the fluids for the element by the estimated global hydrothermal heat flux. The hydrothermal flux of ^3He is assumed equal to the total oceanic flux of ^3He, and is used to estimate the global hydrothermal heat flux. Values for the global river flux (or global sedimentation rate) are given for each component for comparison (A_t = alkalinity). References are given in parentheses.*

Element	Mass to heat ratio (mole/cal)	Hydrothermal flux (mole year)	Global river flux (mole/year)	Global sedimentation rate (mole/year)
^3He	22×10^{-18}	1.08×10^{3} (*48*)		
Li			13.5×10^{9} (*63*)	
High	2.6×10^{-9}	130×10^{9}		
Low	1.5×10^{-9}	75×10^{9}		
Mg	-190×10^{-9}	-9.3×10^{12}	5.4×10^{12} (*64*)	
Ca			12×10^{12} (*64*)	
High	-85×10^{-9}	4.2×10^{12}		
Low	48×10^{-9}	2.4×10^{12}		
Ba			1.4×10^{10} (*65*)	
High	130×10^{-12}	6.4×10^{9}		
Low	49×10^{-12}	2.4×10^{9}		
Si	57×10^{-9}	2.9×10^{12}	7.1×10^{12} (*64*)	
A_t	-25×10^{-9}	-1.3×10^{12}	30×10^{12} (*64*)	
CO_2	25×10^{-9}	1.3×10^{12}		
Ni	-0.5×10^{-12}	-25×10^{-6}		2.4×10^{9} (*66*)
Mn				$52. \times 10^{9}$ (*67*)
High	3.9×10^{-9}	190×10^{9}		
Low	1.5×10^{-9}	74×10^{9}		

There are many questions to be answered about these animal communities. One concerns how they locate and colonize new vents. It is clear that an individual vent area has a finite lifetime. We discovered several dead vent areas along the axial ridge, recognizable by the abundant dead clamshells that were slowly dissolving away, and by the dark manganese staining around the base of the pillow basalts. Radiometric dating of the shells suggests that the shells are no more than 10 to 20 years old (56), indicating a minumum age for the thermal springs of the same order.

The animal population at each vent area is distinctly different from the populations at other vent areas. Clams and large mussels dominate Clambake (Fig. 2); small mussels and Pogonophora are the most abundant species at Oyster Bed (Fig. 5); Pogonophora and limpets are abundant at the Garden of Eden (Fig. 4). The only abundant attached animal at Dandelions has not been identified (see Fig. 3). In addition, there were a few small mussels at the Garden of Eden, and a few decayed-appearing Pogonophora tubes in Clambake. The fact that the mussels are all large in the Clambake and small in the Oyster Bed suggests a single separate colonization in each vent area. The differences between vents suggest perhaps a progression of successive populations culminating in the mussel- and clam-dominated communities such as Clambake 1.

Perhaps the most important question about these animal communities concerns their source of food. Particular features of the Galápagos hot-spring environment, such as the depth and the abundance and size of the animals, suggest that the supportive organic nutrients could not be substantially derived from primary productivity in the surface water. Moreover, because of the high concentrations of hydrogen sulfide in these waters, a significant proportion of the organic carbon utilized within these hot-spring regions could be produced by chemolithotrophic sulfur-oxidizing bacteria.

High concentrations of sulfur-oxidizing and heterotrophic bacteria (from 10^8 to 10^9 bacteria per milliliter) were observed by epiflourescence microscopy in glutaraldehyde-preserved samples of water collected as it flowed from the vents at two hot springs (57). If one assumes that the average bacterium weighs 10^{-12} gram, the concentration of bacteria in the vent waters could be 0.1 to 1 gram per liter. This flux of bacteria from the vents must be supported by the productivity of a large population of bacteria living within the rock mass, lining the walls of fissures through which the hydrogen sulfide-laden fluids ascend. They

96

presumably become incorporated into the fluids which ultimately flow from the rocks and past the filter-feeding mussels, clams, limpets, and *Pogonophora* which surround and fill the vents.

The generation time for these populations of sulfur-oxidizing bacteria in situ is not known, but, unlike phytoplankton blooms, their productivity is, presumably, continuous. Experiments with isolated cultures of these chemolithothrophs showed that they have generation times of approximately 100 hours when grown at 15°C on sodium thiosulfate. Our data do not prove that the hot-spring animals feed on bacteria. Many sulfur bacteria similar to the species isolated from water were isolated from the *Pogonophora* and from the guts of mussels. Very few heterotrophs were isolated, suggesting that the feeding habits of these thermal-spring animals are somewhat different from their counterparts in areas not affected by warm water, which contain in their guts high numbers of heterotrophs, particularly species that degrade refractory organic material.

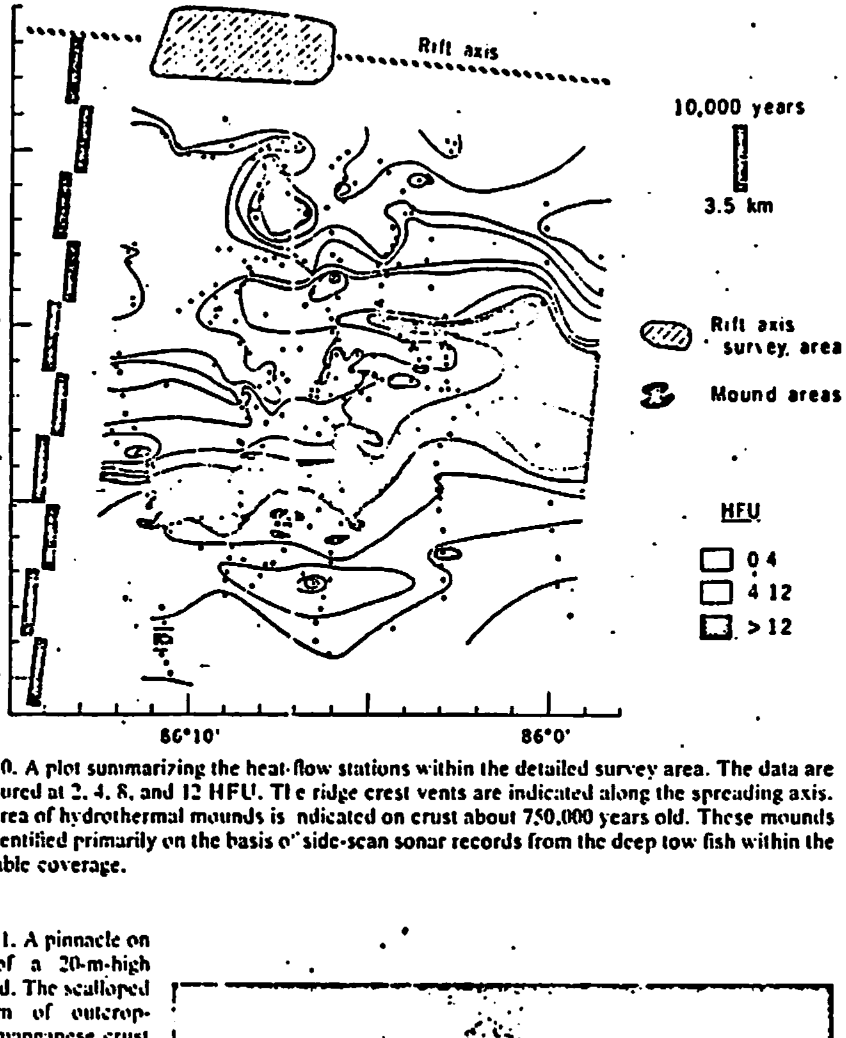


Fig. 10. A plot summarizing the heat-flow stations within the detailed survey area. The data are contoured at 2, 4, 8, and 12 HFU. The ridge crest vents are indicated along the spreading axis. The area of hydrothermal mounds is indicated on crust about 750,000 years old. These mounds are identified primarily on the basis of side-scan sonar records from the deep tow fish within the available coverage.

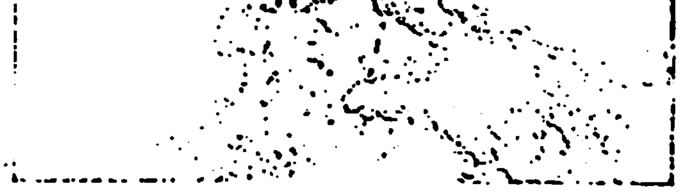


Fig. 11. A pinnacle on top of a 20-m-high mound. The scalloped pattern of outcropping manganese crust is typical of the undisturbed surfaces of the mounds. The interiors of such pinnacles were warm (to 6°C) and very soft, containing water-saturated nontronite and iron oxides covered by a more rigid manganese crust. Small "spikes" of iron oxides extended from the surface. These pinnacles appear to grow by deposition from the fluids flowing from the mound interiors. [Photo by D. Williams]

The high concentration of bacteria in water flowing from the vents suggests that they live to some depth in the rock mass, lining fractures and fissures, and that they may significantly influence the chemistry of the system. In systems in which there is extensive subsurface mixing of ambient water, the development of an abundant surface fauna is prevented, as in Dandelions, both by oxidation of all the sulfide and by dilution and destruction of the sulfide-oxidizing bacteria, presumably inhabiting the reducing zone at depth. This, coupled with the much smaller temperature anomaly, will make such vent fields very difficult to find.

Conductive Heat-Flow Patterns

The hydrothermal vents we have described are located at the rift axis, associated with newly erupted volcanic rocks. South of the spreading axis, we studied manifestations of hydrothermal circulation in older oceanic crust. These include a striking pattern of variations in conductive heat flow through the sediments, with maximum values associated with a band of spectacular mounds of hydrothermally deposited minerals precipitated from fluids percolating upward from the basaltic crust (Fig. 10).

More than 400 measurements of conductive heat flow have been made over this survey area on three different cruises (Fig. 10) (*Southtow*, *Pleiades*, *Knorr*-64). Because of the relatively high sedimentation rate (50 m per million years) and subdued topography, measurements were possible to within a few kilometers of the axis. The contoured heat-flow pattern is obviously two-dimensional, subparallel to the axis and topographic lineations. The break in two-dimensionality of the heat-flow pattern near the center of the map corresponds with a shift in the north-south ridge valley pattern of topography. In detail, higher values of heat flow are located over relative topographic ridges, and vice versa. A few measurements along a profile north of the ridge axis and along a line extending about 15 km south of the areal survey indicate that the oscillating patterns of heat flow extend to these regions also.

The oscillating pattern is clearly at variance with that of a simple conductively cooling plate, which would produce a monotonic decrease of heat flow with age. The observations support a mechanism of convective heat loss by hydrothermal circulation, most intense at the ridge axis.

The variation of surface heat flow (q) with age of the crust (t), based on a simple model of a one-dimensional conductively cooling plate, is given as $q = A/t^{1/2}$, where A is an empirically determined constant with a value of about 12 heat flow units (million year)$^{1/2}$, where 1 HFU equals 10^{-6} cal per second per square centimeter (*58*, *59*). The mean of the measured heat loss over the survey region (Fig. 10) is about 8 HFU, whereas the model predicts a mean heat loss of 24 HFU over this age range (0 to 1 million years), indicating a missing heat flux of 16 HFU on each side of the axis which could be accounted for by hydrothermal convection. Near the ridge axis, realistic geometries of the newly intruded material may modify some of the heat-loss calculations based on one-dimensional models, but not to the extent of removing such a major discrepancy.

If we consider both flanks of the ridge, this missing heat flux to 1 million years of age is equivalent to about 110 calories per second for each centimeter of ridge axis. Extrapolating this rate to the total global ridge system (55×10^3 km length) and correcting for the mean spreading rate, 2.7 cm/year, gives us about 1.5×10^{19} cal/year for the global heat loss by hydrothermal circulation. This is somewhat less than that calculated by Wolery and Sleep (*20*) ($4 = 0.4 \times 10^{19}$ cal/year) and from the estimate based on ^{3}He flux and the relation between helium and heat (5×10^{19} cal/year). This difference could result from the relatively rapid sealing of the Galápagos Rift by sediment, which would inhibit convective circulation (*60*), but the uncertainty in the calculations is such that such sealing of the crust is not required.

Hydrothermal Mound Deposits

Numerous rows of hydrothermally deposited mounds, subparallel to the spreading axis, are located within the band of high heat flow from 18 to 25 km south of the rift axis (Fig. 10) in crust from 500,000 to 700,000 years old. Although produced by the same basic process, convective flow of seawater through oceanic crustal rocks, the surface manifestation of the process in this sediment-covered area is distinctly different from that of rift axis systems. Results of deep tow studies of the mounds have been presented by Klitgord and Mudie (*3*), Williams *et al.* (*2*), and Lonsdale (*33*), and the chemistry of the mounds, based on samples dredged on *Pleiades*, leg 2, has been discussed by Corliss *et al.* (*37*).

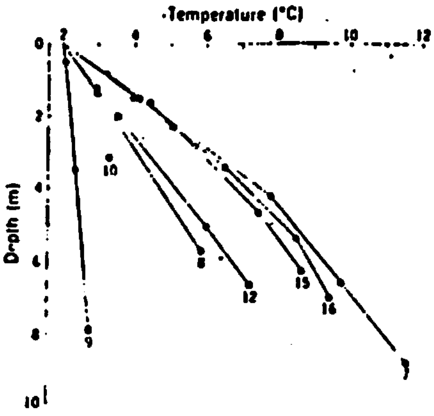


Fig. 12. Temperature gradients from piston-core measurements. All gradients have been adjusted for depth to make the bottom water temperature (2.04° to 2.06°C) consistent with extrapolation of upper gradients, except at 10 where only one temperature was successfully measured. Cores 7, 8, 15, and 16 contain mound deposits; 7, 15, and 16 show decreasing gradients with depth resulting from upward flow of interstitial water. Piston-core 9 was taken in the low-heat flow area north of the mounds (Fig. 10).

The mounds visited by *Alvin* varied in height from less than a meter to over 20 m. The small mounds typically had gentle slopes (< 30°) and were smooth or had a few small outcrops of hard ferromanganese material near their tops (Fig. 11). Away from outcrops these mounds were sediment-covered except for dark scalloped lines a few centimeters across which were found to be thin and friable ferromanganese crusts covered by a light coating of sediment. Larger mounds had correspondingly steeper slopes, sometimes vertical, with outcrops of ferromanganese material that varied from small pinnacles (Fig. 11) to cliffs a meter or two high to vertical spires and pinnacles 5 m high and 1 to 3 m across. Most mounds were roughly circular, although some mounds appeared to form essentially continuous ridges 100 to 200 m long. Generally, neighboring mounds in one row were quite similar, but an adjacent row could be quite different in size and morphology.

The geologic samples collected by *Alvin* from the mounds contain the same three hydrothermal phases found in the *Pleiades* dredge samples (*37*): black iron-free manganese crusts composed of mixtures of todorokite and birnessite, red-orange iron oxides containing manganese and silicon, and an aluminum-poor iron-nontronite that is normally green under reducing conditions within the mounds but can be oxidized to bright yellow when exposed to oxygenated bottom water. Iron and manganese are strongly fractionated into the nontronite and manganese oxides, respectively.

A common depth sequence of deposits, noted both visually and in cores and crust samples, is black manganese oxides overlying orange iron oxides which in turn overlie the green nontronite. The iron oxides often do not appear, and the exposed nontronite can be oxidized to a bright yellow. The manganese oxides appear to form a thin shell over the entire mound, often covered by a thin layer of sediment to form the scalloped outcrop pattern (Fig. 11). Drilling in the mounds area by the *Glomar Challenger* on leg 54 subsequent to our dives (*61*), has revealed thick beds of nontronite within and underlying the mounds, with manganese oxides nearer the surface.

In the mounds area extremely high temperature gradients, up to the equivalent of 30 HFU, were measured from the surface during *Southtow* and *Pleiades* with the heat-flow probe, and during the diving cruise on *Knorr* (*Knorr*-64) with thermistors on piston-core barrels (Fig. 12). Most of the piston-coring stations were targeted on mounds previously visited on *Alvin* dives. The descriptions of material recovered indicate that cores 7, 8, 15, and 16 sampled mound deposits, whereas cores 9 and 12 contain "normal" sediment.

All the highest thermal gradients from mounds appear to decrease significantly with depth, which suggests either nonconductive or nonsteady-state heat flux. One-dimensional modeling (*62*) of these gradients, assuming upward flow of pore waters gives velocities of the order of 10^{-4} cm per second. Such a velocity would be difficult to detect at the sediment and water interface. Temperatures of the source water at the base of the sediment layer are calculated to be about 12° to 20°C, which is consistent with temperatures deduced from extrapolation of the conductive gradients measured in the vicinity of the mounds to the sediment and basalt interface. Measurements with a short (0.5 m) thermal probe from *Alvin* on and near the mounds also showed wide variability in temperature gradients. The highest gradients or temperatures measured, at the crests of mounds, gave calculated conductive heat flows in the range 100 to 200 HFU or higher.

The *Southtow* and *Pleiades* cruises failed to locate any temperature anomalies in the water column above the mounds and no flowing hydrothermal vents were detected by *Alvin*. However, during four dives to the mounds, a visible flow of water with temperatures between 4° and 10°C was initiated by punching holes through the crusts on the mound surfaces. In addition, the manganese concentration in bottom water and water column samples collected over the mounds on *Pleiades*, leg 2, shows a strong systematic increase in total dissolvable manganese approaching the bottom, with highest values found just a few meters above the bottom (*36*, *37*).

Because of the low flow rates, the water collected with our sampling system from the holes was strongly diluted by ambient bottom water. The temperature anomaly of the sampled water was less than 0.5°C. Despite the strong dilution by normal bottom water these mound waters display very strong anomalies in their radon-222 contents. Total radon-222 values as high as 388,000 disintegrations per minute per 100 kilograms were measured. The values undoubtedly reflect the interaction of the thermal waters with the relatively uranium-rich sediments of the area. Other anomalies apparent in the shipboard data are limited to subtle increases in silicon in the mounds thermal water compared to the normal bottom water.

The high internal temperatures, the presence of hydrothermal minerals, and the discovery that warm waters will emanate from holes placed in the mounds all indicate that these mounds are deposited by circulating warm water. The fact that they form in linear rows strongly suggests that faults control the location of individual mounds (*33*). Their relation to the regional heat-flow pattern (Fig. 10) implies that they are located over the ascending limb of a large convective cell within the oceanic crust.

Conclusions

Venting of heated seawater was found in four vent areas along the axial ridge within the central rift of the Galápagos spreading center. The hydrothermal fluids flow from the rocks at maximum temperatures from 10° to 17°C, with measured flow rates of 2 to 10 cm/sec, and form large plumes that drift horizontally in the bottom currents. These were detected up to 150 m above the bottom where the temperature anomaly had dropped to 0.01°C.

Chemical data suggest that the interactions between the water and rocks may occur at temperatures of at least 300°C. Global hydrothermal fluxes of heat, and major and trace chemical constituents of seawater can be estimated by normalization to the ^{3}He flux. The chemical data indicate that submarine hydrothermal activity provides a major source of manganese and lithium; a minor source of calcium, barium, silicon, and carbon dioxide; and a significant sink for magnesium and alkalinity in the oceans. Nickel, copper, and cadmium are absent in the emerging vent water, because of the presence of hydrogen sulfide, which leads to the precipitation of these metals along with iron as sulfides within the rocks.

Convection of seawater in older sedimented crust south of the rift axis produces a distinctive pattern of heat flow and extensive hydrothermally deposited mounds. The mounds are formed by warm fluids percolating up from a convective cell in the underlying crust along fractures in the basalt, precipitating iron and silica in the reducing mound interiors, manganese at the bottom water interface, and carrying a significant flux of at least manganese into the bottom waters. The total hydrothermal heat loss is estimated from the heat-flow data to be 110 calories per centimeter per second, equivalent to 1.5 × 10^{18} cal/year globally, somewhat less than estimates based on global heat-flow patterns (≈ 4 × 10^{19} cal/year) and on normalization to ^{3}He flux in the rift axis vents (5 × 10^{19} cal/year).

The exploration of the vents revealed the existence of animal communities that appear to be totally dependent on energy derived from seawater-rock reactions and chemosynthesis by bacteria for their primary productivity. These fragile communities provide a unique opportunity for a wide range of zoological, bacteriological, ecological, and biochemical studies.

References and Notes

1. J. G. Sclater and K. D. Klitgord, *J. Geophys. Res.* 78, 6951 (1973).
2. D. L. Williams, R. P. von Herzen, J. G. Sclater, R. N. Anderson, *Geophys. J. R. Astron. Soc.* 38, 587 (1974).
3. K. D. Klitgord and J. D. Mudie, *ibid.*, p. 563.
4. K. C. MacDonald and J. D. Mudie, *ibid.* 36, 245 (1974).
5. J. W. Elder, *Am. Geophys. Union Monogr.* 8, 211 (1965).
6. K. Bostrom and M. N. A. Peterson, *Econ. Geol.* 61, 1258 (1966).
7. K. S. Deffeyes, in *Megatectonics of Continents and Oceans*, H. Johnson, Ed. (Rutgers Univ. Press, New Brunswick, N.J., 1970).
8. M. Talwani, C. C. Windish, M. G. Langseth, *J. Geophys. Res.* 76, 473 (1971).
9. J. B. Corliss, *ibid.*, p. 8128.
10. C. R. B. Lister, *Geophys. J. R. Astron. Soc.* 26, 515 (1972).
11. _____, *ibid.* 39, 465 (1974).
12. E. T. C. Spooner and W. S. Fyfe, *Contrib. Mineral. Petrol.* 42, 287 (1974).
13. J. L. Bischoff and F. W. Dickson, *Earth Planet. Sci. Lett.* 25, 385 (1975).
14. M. J. Mottl, thesis, Harvard University (1976).
15. A. Hajash, *Contrib. Mineral. Petrol.* 53, 205 (1975).
16. W. E. Seyfried and J. L. Bischoff, *Earth Planet. Sci. Lett.* 34, 71 (1977).
17. W. E. Seyfried and M. J. Mottl, in *Proceedings of the Second International Symposium on Water-Rock Interaction*, H. Paquet and Y. Tardy, Eds. (Université Louis Pasteur, Strasbourg, France, 1977), vol 4, pp. 173–180.
18. R. P. Lowell, *Geophys. J. R. Astron. Soc.* 40, 351 (1975).
19. C. R. B. Lister, *Tectonophysics* 37, 203 (1977).
20. T. J. Wolery and N. H. Sleep, *J. Geol.* 84, 249 (1976).
21. M. Lyle, *Geology* 4, 733 (1975).
22. D. L. Williams and R. P. von Herzen, *ibid.* 2, 327 (1974).
23. J. Dymond, J. B. Corliss, G. R. Heath, C. W. Field, E. J. Dasch, E. M. Veeh, *Geol. Soc. Am. Bull.* 84, 3355 (1973).
24. M. Bender, W. Broecker, V. Gornitz, V. Middel, R. Kay, S. S. Sun, P. Biscaye, *Earth Planet. Sci. Lett.* 12, 425 (1971).
25. G. R. Heath and J. Dymond, *Geol. Soc. Am. Bull.* 88, 723 (1975).
26. J. B. Corliss, J. L. Graf, Jr., B. J. Skinner, R. W. Hutchinson, *Geol. Soc. Am. (Abstracts with Programs)* 4, 476 (1972).
27. A. H. F. Robertson and J. D. Hudson, *Earth Planet. Sci. Lett.* 18, 93 (1973).
28. E. T. C. Spooner and C. J. Bray, *Nature (London)* 266, 808 (1977).
29. R. B. Scott, P. A. Rona, B. A. McGregor, M. R. Scott, *ibid.* 251, 301 (1974).
30. M. F. Glenn, *Int. Hydrogr. Rev.* 47, 35 (1970).
31. Cruise of the R/V *Melville*, June to July (1976).
32. K. Crane, *J. Geol.*, 86, 715 (1979).
33. P. Lonsdale, *Earth Planet. Sci. Lett.* 36, 92 (1977).
34. J. E. Lipton, R. F. Weiss, H. Craig, *Nature (London)* 267, 603 (1977).
35. R. F. Weiss, P. F. Lonsdale, J. E. Lipton, A. E. Bainbridge, H. Craig, *ibid.*, p. 600.
36. G. Klinkhammer, M. Bender, R. Weiss, *ibid.* 269, 319 (1977).
37. J. B. Corliss, J. Dymond, M. Lyle, K. Crane, *Earth Planet. Sci. Lett.* 40, 12 (1978).
38. Tj. H. van Andel and R. D. Ballard, *J. Geophys. Res.*, in press.
39. R. D. Ballard, R. H. Holcomb, Tj. H. van Andel, *ibid.*, in press.
40. R. D. Ballard and Tj. H. van Andel, *Geol. Soc. Am. Bull.* 88, 507 (1977).
41. G. W. Morey, R. O. Fournier, J. J. Rowe, *Geochim. Cosmochim. Acta* 26, 1029 (1962).
42. _____, *J. Geophys. Res.* 69, 1995 (1964).
43. R. O. Fournier and J. J. Rowe, *Am. J. Sci.* 264, 685 (1966).
44. W. G. Nelson and Tj. H. van Andel, *Mar. Geol.* 4, 365 (1966).
45. S. E. Humphris and G. Thompson, *Geochim. Cosmochim. Acta* 42, 107 (1974).
46. Manganese analyses were performed by G. Klinkhammer and M. Bender, University of Rhode Island.
47. W. C. Shanks III and J. L. Bischoff, *Geochim. Cosmochim. Acta* 41, 1507 (1977).
48. M. R. Scott, R. B. Scott, P. A. Rona, L. W. Butler, A. J. Nalwalk, *Geophys. Res. Lett.* 1, 355 (1974).
49. W. S. Moore and P. R. Vogt, *Earth Planet. Sci. Lett.* 29, 349 (1976).
50. W. J. Jenkins, J. M. Edmond, J. B. Corliss, *Nature (London)* 272, 156 (1978).
51. H. Craig, W. B. Clarke, M. A. Beg, *Earth Planet. Sci. Lett.* 26, 125 (1975).
52. J. B. Corliss, J. M. Edmond, L. I. Gordon, *Proceedings of the Second Annual EWING Symposium* (American Geophysical Union, Washington, D.C., in press).
53. R. N. Anderson, M. G. Langseth, J. G. Sclater, *J. Geophys. Res.* 82, 3391 (1977).
54. P. Lonsdale, *Deep-Sea Res.* 24, 857 (1977).
55. The clams were classified by K. Boss, Harvard University; the mussels by V. Kenk, San Jose State University; the limpets by J. McLean, Los Angeles County Museum of Natural History; the pogonophorans by M. Jones, National Museum of Natural History; and the fish photographs were examined by D. Cohen, National Museum of Natural History.
56. The analyses of ^{210}Pb were made by K. Turekian, Yale University.
57. Microbiological studies were conducted by J. Baross, Oregon State University.
58. C. R. B. Lister, *Tectonophysics* 41, 157 (1977).
59. B. Parsons and J. G. Sclater, *J. Geophys. Re* 82, 803 (1977).
60. R. N. Anderson and M. A. Hobart, *ibid.*, 8 2968 (1976).
61. Deep Sea Drilling Project, Leg 54, *Geotimes* 2 19 (1977).
62. E. A. Lubimova, R. P. Von Herzen, G. B. Udintsev, in *Terrestrial Heat Flow*, W. H. K. Lee, Ed. (American Geophysical Union Monograph No. 8, Washington, D.C., 1965), 78
63. T. J. Chow and E. D. Goldberg, *J. Mar. Res.* 2 163 (1962).

64. R. M. Garrels and F. T. MacKenzie, *Evolution of Sedimentary Rocks* (Norton, New York, 1971).

65. L. H. Chan, J. M. Edmond, R. F. Stallard, W. S. Broecker, Y. C. Chung, R. F. Weiss, T. L. Ku, *Earth Planet. Sci. Lett.* 32, 258 (1976).

66. F. R. Sclater, E. Boyle, J. M. Edmond, *ibid.* 31, 119 (1976).

67. M. L. Bender, G. T. Klinkhammer, D. W. Spencer, *Deep-Sea Res.* 24, 799 (1977).

68. The success of the expedition was made possible by the outstanding effort of the *Alvin* crew, under the direction of Larry Shumaker and Skip Marquet, the officers and crew of the R/V *Lulu*, Richard Flegenheimer, Master, and the R/V *Knorr*, Emerson Hiller, Master, and the members of the scientific party who carried out our heavy work schedule. Preliminary results from *Pleiades* leg I provided by P. Lonsdale, R. Weiss, J. Lupton and H. Craig are gratefully acknowledged. The expedition was funded by the Sea Bed Assessment program of the International Decade of Ocean Exploration Office, National Science Foundation; grant OCE 75-23352 to Oregon State University, and OCE 76-389 to Woods Hole Oceanographic Institution.

Mobilizing Technology for Developing Countries

Charles Weiss, Jr.

The 15 years since the United Nations Conference on Science, Technology, and Development in Geneva have taught us that what seemed at first to be technological obstacles to development frequently turn out on closer examination to have been policy failures.

If economic policies prevent com- etition, firms will pass up opportunities or the use of known technology to im- rove quality. If financial authorities vervalue local currency, firms will have n incentive to import equipment and aw materials in place of locally available upplies and will neglect local sources of chnology. If government marketing oards pay too little for higher quality, rocessors of agricultural commodities ill neglect their machines and will fail take advantage of improved technology. If banks are unwilling to extend redit to small farmers, the latter will continue to face severely restricted technological options.

If governments do not give high priority to the provision of safe water to poor people at a cost the country can afford, the agency charged with this responsibility will neither attract good managers, provide proper working conditions for its engineering staff, nor adopt appropriate design criteria, and hence will fail to employ the most suitable technology (*1*).

Summary. A new problem in technology policy—a discipline hitherto largely concerned with the modern industrial sector—is posed by the need for technology suited to creating productive jobs and providing minimum public services at a cost and level of sophistication within the reach of poor people in developing countries. Careful consideration must be given to overall and sectoral development objectives, economic and manpower resources, and the local institutional and sociocultural context. This may indicate the need for both hardware innovation, such as low-cost alternatives to waterborne sewerage, and social ("software") innovation, such as training large numbers of supervisors to implement improved technologies for labor-intensive civil works.

In each of these situations, the fact that a technology that seems suited to a local situation is, in fact, not used is a symptom of a deeper problem. In the language of experimental science, technology is a probe that reveals issues that might otherwise have escaped attention. But this does not mean that the problems thus revealed can necessarily be alleviated by the introduction of the "missing" technique. On the contrary, evidence is piling up that the impact of the introduction of any particular piece of equipment—whether tractors in southern Asia (*2*) or waterless toilets in Vietnam (*3*)—depends heavily on the social and institutional structures on which it is superimposed.

For this reason, there are many situations in which an intervention focused purely on technology—whether indigenous or foreign and whether new, adapted, or transferred—is likely to be doomed from the start. In such cases, the introduction of hardware must be accompanied by and integrated with a package of policy and institutional changes if a desired innovation is to be effected, so it is more illuminating to refer to the institutional change necessary to the solution of a social problem as the "software" of the technology.

An example may make this point clearer. One may imagine three approaches to the problem of providing water to small farmers from aquifers near the surface: (i) designing and testing a small hand- or pedal-powered pump to be used by one or two farmers and encouraging a system to market and maintain such pumps; (ii) encouraging the installation of diesel-powered tube wells serving 50 or so farmers and ensuring an equitable distribution of the water through cooperatives; and (iii) encouraging entrepreneurs to hire out truck-mounted pumping equipment by the hour to individual farmers.

Each of these overall approaches—hardware plus institutional support—constitutes an alternative technology. The choice among them should depend on careful overall assessment of local technoeconomic, geographic, ecological, and social factors, as well as the desired balance between growth and equity. Such a technology assessment, a key element in the choice of "appropriate" (locally suitable) technology for particular investment projects, should be built into procedures for project preparation and appraisal in governments and development assistance agencies (*4*).

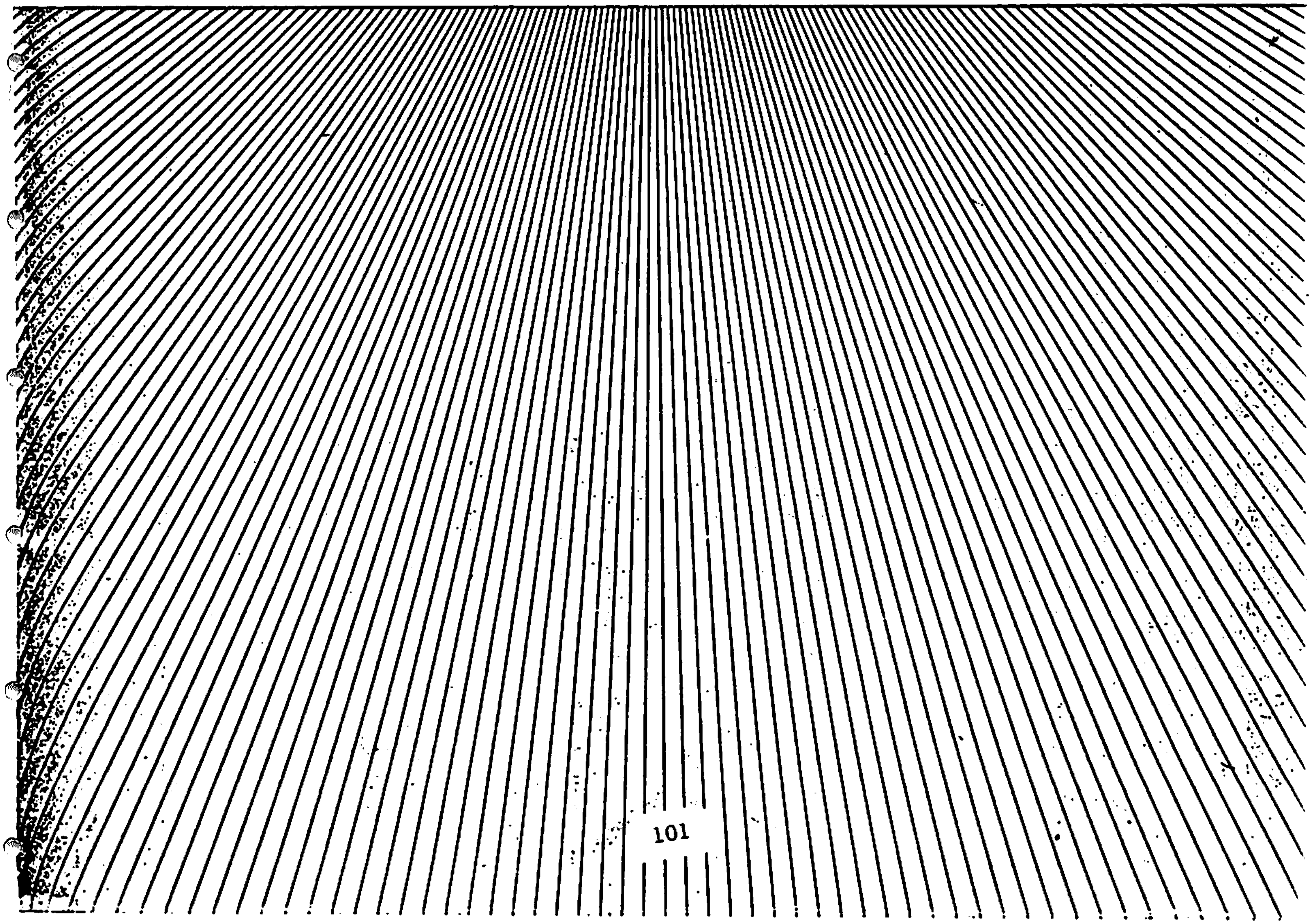
The crux of this approach is the focus on the problem rather than on the technology. "Technology is the answer—but what is the question?" (*5*). In the previous paragraphs, the objective is delivery of water, not development of pumps.

The author is Science and Technology Adviser at e World Bank, Washington, D.C. 20433.

1

Sagan, C.
The Solar System Chapter 1

THE SOLAR SYSTEM




PISCES
UARIUS
SYSTEMA TYCHONIS, vixit circa finem Sec. XVI.
Sic oculis,
SYSTEMA COPI
Sic ratio

The Solar System

CARL SAGAN

Presenting what is known about the sun and the bodies in orbit around it, with special reference to the knowledge gained in 18 years of exploration by space probes launched from the earth

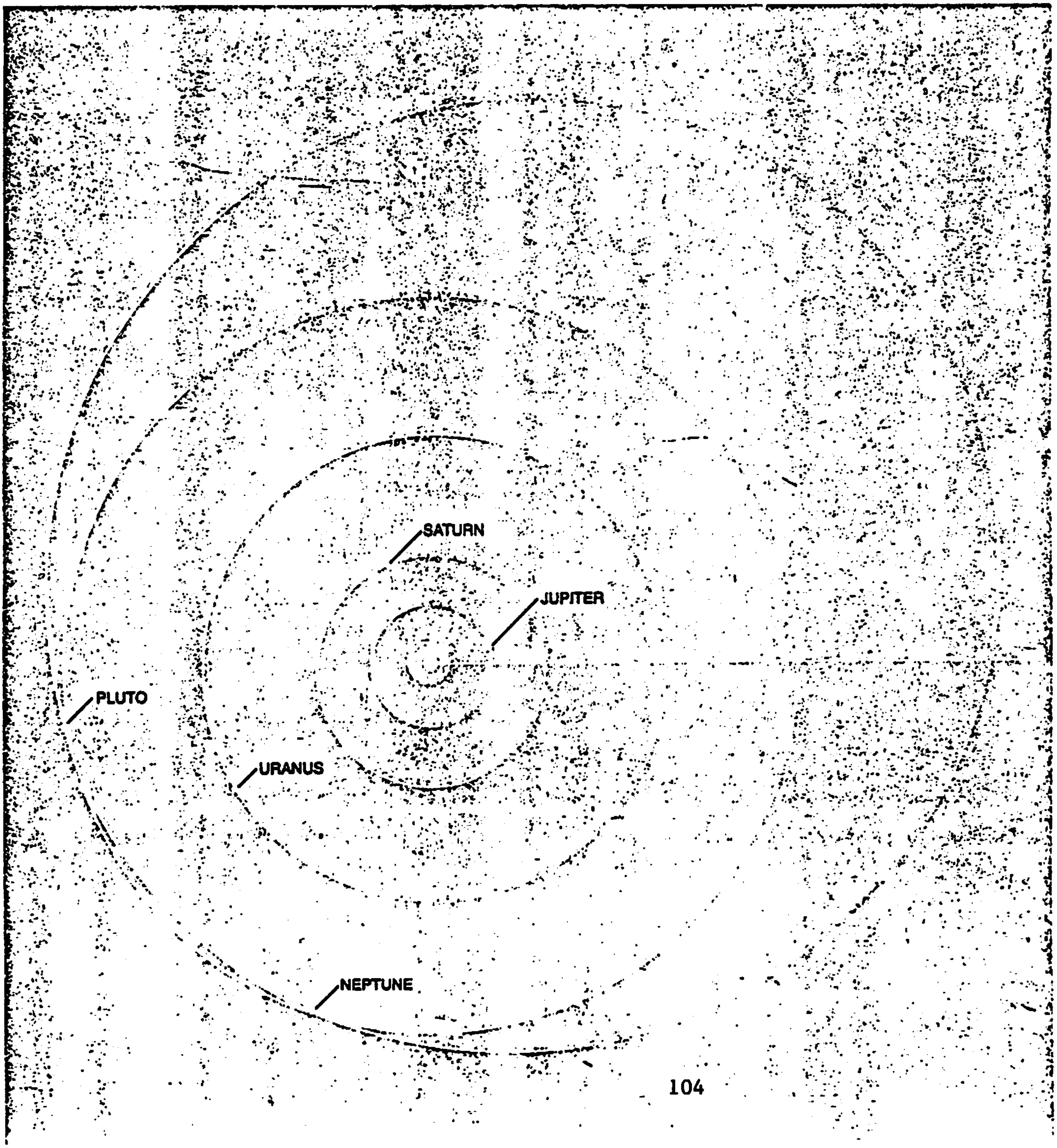
Imagine that the earth has been watched over the millenniums by a careful and extremely patient extraterrestrial observer. Some 4.6 billion years ago the planet completes its aggregation out of interstellar gas and dust. The last planetesimals that fall in to make the earth produce enormous impact craters; the planet heats up internally from the gravitational potential energy of its accretion and from the decay of its radioactive elements; the heavy liquid iron core separates from the lighter silicate mantle and crust; hydrogen-rich gases and condensable water are released from the interior to the surface, and a fairly straightforward organic chemistry yields complex molecules that combine into self-replicating molecular systems: the first terrestrial organisms. The rain of interplanetary boulders dwindles, and in time running water, wind, mountain building and other geological processes erase the scars of the earth's origin. A vast planetary convection engine is established that carries mantle material up through rifts in the ocean floor to form great crustal plates and then drives the material back into the mantle at the margins of the continents; collisions between plates push up chains of folded mountains, and the general configuration of land and sea, of icy and tropical regions shifts continuously. Meanwhile natural selection nominates from a wide range of alternative candidates those varieties of self-replicating molecular systems best suited to the latest change in the environment. Plants evolve that use visible light to break down water into hydrogen and oxygen, and the hydrogen escapes into space, changing the atmosphere from a reducing medium to an oxidizing one. Organisms of moderate complexity and modest intelligence eventually arise.

Throughout this sequence our imaginary observer is struck above all by the earth's isolation. Sunlight, starlight and cosmic rays, and occasionally some interplanetary debris, arrive at the earth's surface, but in all those aeons nothing save a little hydrogen and helium leaves the planet. And then, less than 20 years ago, the planet suddenly begins, like a dandelion gone to seed, to fire tiny capsules throughout the inner solar system. First they go into orbit around the earth and then to the planet's lifeless natural satellite, the moon. Six tiny capsules, larger than the rest, set down on the moon and from each two small organisms emerge, briefly explore their immediate surroundings and then sprint back to the earth, having tentatively extended a toe into the cosmic ocean. Five little spacecraft enter the hellhole of Venus' atmosphere and three of them survive some tens of minutes on the surface before being de-

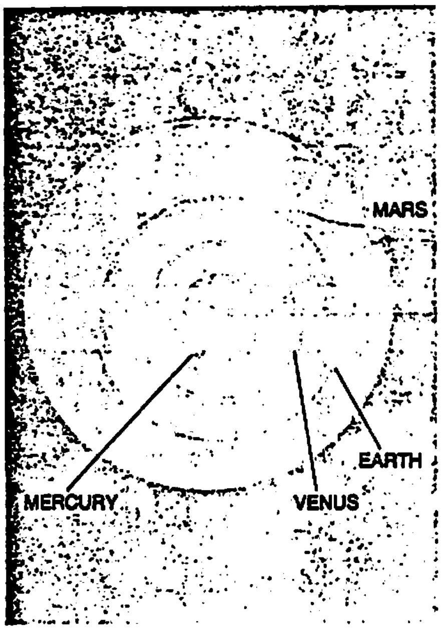
CRUCIAL DECISION in the evolution of the modern heliocentric model of planetary motion is represented symbolically in the illustration on the opposite page. The scene is a detail of a large hand-colored wood engraving of the solar system, one of a set of 30 such astronomical charts compiled by Johann Gabriel Doppelmayer of Nuremberg and published in 1742 under the title *Atlas novus coelestis*. The two circular diagrams depict the planetary system according to the two great 16th-century astronomers Tycho Brahe and Nicolaus Copernicus. The Latin inscriptions on the yellow ribbons outlining the two diagrams read in translation: "System of Tycho, who lived around the end of the 16th century" and "System of Copernicus, who lived around the beginning of the 16th century." The phrases in the italic letters just under each diagram can be translated as "Thus by eye" in the case of Tycho (who was noted primarily as an observer) and "Thus by reason" in the case of Copernicus (who was of course best known as the conceiver of the heliocentric theory). The female figure, presumably Urania, the muse of astronomy, appears to favor the Copernican system over the Tychonic system. (The Copernican system, as in the prevailing modern view, has the planets revolving around the sun at the center; the Tychonic system, following the medieval Ptolemaic tradition, continues to have the earth at the center but attempts to account for the observational data of the time by having the other planets revolve around the sun as the sun revolves around the earth.) The meaning of the symbols that appear in the chart is given in the illustration on page 6. The copy of the Doppelmayer atlas from which this [illegible] spacecraft are dispatched to Mars; one sends back information for a full year from its orbit around the planet. Another swings by Venus to encounter Mercury, on a trajectory that will cause it to pass close to the innermost planet many times. Two more successfully traverse the asteroid belt, fly close to Jupiter and are ejected by its gravity into interstellar space. It is clear, the observer might report, that something interesting is happening on the planet earth.

We have entered, almost without noticing it, an age of exploration and discovery unparalleled since the Renaissance, when in just 30 years European man moved across the Western ocean to bring the entire globe within his ken. Our new ocean is beyond that globe: it is the shallow disk of space occupied by the solar system. Our new worlds are the sun, the moon and the planets. In less than 20 years of space exploration we have learned more about those worlds than we have in all the preceding centuries of earthbound observation. We are beginning to assemble that information into a new picture of our solar system.

It is useful—and somewhat humbling—to start by placing our small solar neighborhood in its proper cosmic perspective. The earth is a tiny hunk of rock and metal that rides in a flood of sunlight through the innermost recess of the solar system. Other tiny spheres of rock and metal—Mercury, Venus and Mars—move in orbit around the sun



nearby. These inner planets and their satellites do not bulk very large in the solar system as a whole. Most of the mass, angular momentum and (from any extraterrestrial astronomer's viewpoint) ostensible interest of the solar system resides in the Jovian planets: four immense and rapidly rotating spheres. The inner two, Jupiter and Saturn, consist largely of hydrogen and helium; indeed, Jupiter is something like a star that failed. The outer two, Uranus and Neptune, are composed less of the lightest gases and more of such heavier gases as methane and ammonia. Jupiter takes almost 12 years to complete its trip around the sun at a mean distance of some five astronomical units. (An astronomical unit is the mean distance of the earth from the sun, about 93 million miles or 150 million kilometers). Beyond the Jovian planets Pluto, smaller and less familiar, orbits eccentrically at about 40 astronomical units. Much farther, at about 100,000 astronomical units, are some billions of tailless comets, kilometer-size snowballs slowly circling the distant sun.



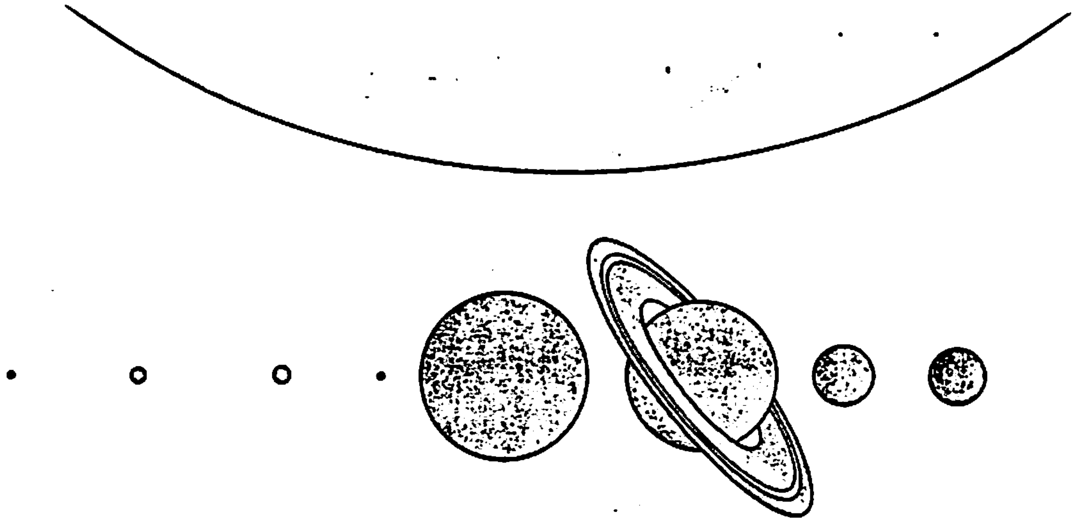
SOLAR SYSTEM FOR SEPTEMBER is represented in the large diagram at left with the planetary orbits all drawn to the correct scale; an enlargement of the inner portion of the diagram appears above. The white dots indicate only the positions of the planets; since the mean diameter of the earth's orbit is roughly 200 times the solar diameter, even the sun would be a barely perceptible speck at the scale of either diagram. By convention the heliocentric longitude of each planet is measured in degrees of arc from the vernal equinox (*straight line at right*). The broken curves denote the portion of each planet's orbit that lies below the earth's orbital plane (the ecliptic). Pluto's orbit is anomalous in several respects. It is the only planetary orbit whose eccentricity can be distinguished from a "zero ellipse" (a circle) at the scale of such diagrams. Moreover, it has by far the greatest inclination to the ecliptic of any planetary orbit: more than 17 degrees. These and other considerations have led some to regard Pluto as being not a planet but rather an escaped satellite of Neptune. Beginning in 1987 Pluto will lose even its questionable distinction of being the outermost planet in the solar system when it slips inside the orbit of Neptune on its way to perihelion (the point on its orbit closest to the sun).

From somewhat farther away, say a few hundred thousand astronomical units, the sun would appear to the unaided eye as a bright star with no hint of its retinue of planets. That would be a distance of a few light-years (a light-year is about 60,000 astronomical units), or the characteristic separation between stars in our galaxy. From a few dozen light-years away the sun would be quite undetectable to the unaided human eye—and a distance of a few dozen light-years is only about a thousandth of the distance from the sun to the center of our galaxy. The galaxy is a vast, ponderously rotating pinwheel of some 250 billion suns, and the dense central plane of the galaxy, seen edge on, is the diffuse band across the sky that we call the Milky Way. Our galaxy is one of at least billions, and perhaps hundreds of billions, of galaxies. Our particular sun and its companion planets constitute no more than one example of a phenomenon that must surely be repeated innumerable times in the vastness of space and time.

If the 4.6 billion years of earth history were compressed into a single year, the flurry of space exploration would have begun less than a tenth of a second ago. The fundamental changes in attitude and knowledge responsible for the remarkable transformation would have filled only the past few seconds, since the first widespread application of simple lenses and mirrors for astronomical purposes in the 17th century. Before that the planets had been recognized for millenniums as being different from the "fixed stars," which appeared not to move with respect to one another. The planets (the word comes from the Greek for "wanderer") were brighter than most stars, and they moved against the stellar background. Since the sun and the moon manifestly influenced the earth, astrological doctrine held that the planets must affect human life too, but in more subtle ways. Almost none of the ancients speculated that the planets were worlds in some sense like the earth. With the first astronomical telescope, however, Galileo was astonished and delighted to see Venus as a crescent lighted by the sun and to make out the mountains and craters of the moon. Johannes Kepler thought the craters were the constructions of intelligent beings inhabiting the moon, but Christiaan Huygens disagreed. He argued that the construction of such great circular depressions would require an unreasonably great effort—and he thought he could see natural explanations for them.

Huygens exemplified the marriage of advancing technology and experimental skills with a reasonable, skeptical mind and an openness to new ideas. He was the first to suggest that on Venus we are looking at an atmosphere and clouds, the first to understand something of the true nature of the rings of Saturn (which had seemed to Galileo like two "ears" enveloping the planet), the first to draw a picture of a recognizable marking on the surface of Mars (Syrtis Major) and the second (after Robert Hooke) to draw the Great Red Spot of Jupiter. The last two observations are of current significance because they establish the continuity of prominent planetary surface features over at least three centuries. (Huygens was, to be sure, not a thoroughly modern astronomer; he could not entirely escape the modes of belief of his time. Consider the curious argument by which he deduced the existence on Jupiter of hemp. Galileo had observed four moons traveling around Jupiter. Huygens asked a question of a kind few astronomers would ask today: Why is it that Jupiter has four moons? Well, why does the earth have one moon? Our moon's function, Huygens reasoned, apart from providing a little light at night and raising the tides, is to aid mariners in navigation. If Jupiter has four moons, there must be many mariners on that planet. Mariners imply boats; boats imply sails; sails imply ropes. And ropes imply hemp. I sometimes wonder how many of our own prized scientific arguments will appear equally foolish from the vantage of three centuries.)

A useful index of our knowledge about a planet is the number of bits of information necessary to characterize what we know of its surface—in effect the number of black and white dots in halftone photographic reproductions summarizing all existing imagery. Back in Huygens' day about 10 bits of information, all obtained by brief glimpses



	MERCURY	VENUS	EARTH	MARS	JUPITER	SATURN	URANUS	NEPTUNE	PLUTO
MAXIMUM DISTANCE FROM SUN (MILLIONS OF KILOMETERS)	69.7	109	152.1	249.1	815.7	1,507	3,004	4,537	7,375
MINIMUM DISTANCE FROM SUN (MILLIONS OF KILOMETERS)	45.9	107.4	147.1	206.7	740.9	1,347	2,735	4,456	4,425
MEAN DISTANCE FROM SUN (MILLIONS OF KILOMETERS)	57.9	108.2	149.6	227.9	778.3	1,427	2,869.6	4,496.6	5,900
MEAN DISTANCE FROM SUN (ASTRONOMICAL UNITS)	.387	.723	1	1.524	5.203	9.539	19.18	30.06	39.44
PERIOD OF REVOLUTION	88 DAYS	224.7 DAYS	365.26 DAYS	687 DAYS	11.86 YEARS	29.46 YEARS	84.01 YEARS	164.8 YEARS	247.7 YE
ROTATION PERIOD	59 DAYS	−243 DAYS RETROGRADE	23 HOURS 56 MINUTES 4 SECONDS	24 HOURS 37 MINUTES 23 SECONDS	9 HOURS 50 MINUTES 30 SECONDS	10 HOURS 14 MINUTES	−11 HOURS RETROGRADE	16 HOURS	6 DAYS 9 HOUR.
ORBITAL VELOCITY (KILOMETERS PER SECOND)	47.9	35	29.8	24.1	13.1	9.6	6.8	5.4	4.7
INCLINATION OF AXIS	<28°	3°	23°27′	23°59′	3°05′	26°44′	82°5′	28°48′	?.
INCLINATION OF ORBIT TO ECLIPTIC	7°	3.4°	0°	1.9°	1.3°	2.5°	.8°	1.8°	17.2°
ECCENTRICITY OF ORBIT	.206	.007	.017	.093	.048	.056	.047	.009	.25
EQUATORIAL DIAMETER (KILOMETERS)	4,880	12,104	12,756	6,787	142,800	120,000	51,800	49,500	6,000 (?
MASS (EARTH = 1)	.055	.815	1	.108	317.9	95.2	14.6	17.2	.1 (?)
VOLUME (EARTH = 1)	.06	.88	1	.15	1,316	755	67	57	.1 (?)
DENSITY (WATER = 1)	5.4	5.2	5.5	3.9	1.3	.7	1.2	1.7	?
OBLATENESS	0	0	.003	.009	.06	.1	.06	.02	?
ATMOSPHERE (MAIN COMPONENTS)	NONE	CARBON DIOXIDE	NITROGEN, OXYGEN	CARBON DIOXIDE, ARGON (?)	HYDROGEN, HELIUM	HYDROGEN, HELIUM	HYDROGEN, HELIUM, METHANE	HYDROGEN, HELIUM, METHANE	NONE DETECT
MEAN TEMPERATURE AT VISIBLE SURFACE (DEGREES CELSIUS) S = SOLID, C = CLOUDS	350(S) DAY −170(S) NIGHT	−33 (C) 480 (S)	22 (S)	−23 (S)	−150 (C)	−180 (C)	−210 (C)	−220 (C)	−230(?)
ATMOSPHERIC PRESSURE AT SURFACE (MILLIBARS)	10^{-9}	90,000	1,000	6	?	?	?	?	?
SURFACE GRAVITY (EARTH = 1)	.37	.88	1	.38	2.64	1.15	1.17	1.18	?
MEAN APPARENT DIAMETER OF SUN AS SEEN FROM PLANET	1°22′40″	44′15″	31′59″	21′	6′09″	3′22″	1′41″	1′04″	49″
KNOWN SATELLITES	0	0	1	2	[illegible]	10	5	2	0
SYMBOL	☿	♀	⊕	♂	♃	♄	♅	♆	♇

through telescopes, would have characterized man's knowledge of the surface of Mars. By the time of the close approach of Mars to the earth in 1877 that number had risen to perhaps a few thousand (if we exclude a large amount of erroneous information, such as the drawings of "canals" that we now know were entirely illusory). With further visual observations and the rise of astronomical photography the amount of information grew slowly until the advent of space-vehicle exploration of the planet provided a surge of new data. Just 22 photographs obtained in 1965 by the *Mariner 4* flyby mission represented five million bits of information, roughly comparable to all previous photographic knowledge of the planet, although they covered only a tiny fraction of the planet's area. The dual flyby mission of *Mariner 6* and *Mariner 7* in 1969 extended the coverage, increasing the bit total by a factor of 100, and in 1971 and 1972 the *Mariner 9* orbiter increased it by another factor of 100. The *Mariner 9* photographic results from Mars correspond roughly to 10,000 times the total previous photographic knowledge of Mars gathered over the history of mankind. The infrared and ultraviolet spectroscopic data and other information obtained by *Mariner 9* represent a similar enhancement.

The vast amount of new photographic information involves not only an advance in coverage, or quantity, but also a spectacular advance in resolution, or quality. Before the voyage of *Mariner 4* the smallest feature reliably detected on the surface of Mars was several hundred kilometers across. With the completion of the *Mariner 9* mission several percent of the planet's area has been observed at an effective resolution of 100 meters, an improvement in resolution by a factor of 1,000 in the past 10 years and by a factor of 10,000 since Huygens' time. It is only because of this improvement in resolution that we know of vast volcanoes, laminated polar formations, sinuous channels, great rift valleys, dune fields, crater-associated dust streaks and many other instructive and mysterious features of the Martian environment.

Both resolution and coverage are required in order to provide adequate information about a newly explored planet. For example, by an unlucky coincidence the *Mariner 4*, *Mariner 6* and *Mariner 7* spacecraft observed the old, cratered and comparatively uninteresting part of Mars and gave no hint of the young and geologically active third of the planet that was revealed by *Mariner 9*. Intelligent life on the earth would be entirely undetectable by photography in reflected sunlight unless about 100-meter resolution was achieved, at which point the urban and agricultural geometry of our technological civilization would become strikingly evident. This means that if there had been a civilization on Mars comparable in extent and level of development to our own, it would not have been detected photographically until the *Mariner 9* mission. There is no reason to expect such civilizations on other planets in our solar system; my point is that we are only now beginning an adequate reconnaissance of our neighboring worlds. There is no question that astonishments and delights await us as both resolution and coverage are dramatically improved in photography, and in spectroscopic and other methods, by future space-probe missions.

The vigor of the burgeoning planetary sciences and the volume and detail of recent findings will impress anyone who attends a meeting of the Division for Planetary Sciences of the American Astronomical Society. At the 1975 meeting in February there were reports on the discovery of water vapor in the atmosphere of Jupiter, of ethane on Saturn, of possible hydrocarbons on the asteroid Vesta, of an atmospheric pressure approaching that of the earth on Saturn's moon Titan and of radio bursts in the decameter-wavelength range from Saturn. Jupiter's moon Ganymede had been detected by radar, and the radio-emission spectrum of another Jovian moon, Callisto, had been elaborated. And spectacular new views of Jupiter and Mercury and their magnetospheres were presented by the *Pioneer 11* and *Mariner 10* experimenters.

Such discoveries are important and exciting in themselves, but it is their implications and interrelations that are most significant. Every new finding adds to the accumulation of evidence that is required before we can write an authentic history of the origin and evolution of the solar system. No complete version of that history has yet been accepted, but this field of study is now rich in provocative hints and ingenious surmises. Apart from an understanding of the solar system as a whole, it is becoming clear that information about any planet or satellite illuminates our knowledge of the others. In particular, if we are to understand the earth, we must have a comprehensive knowledge of the other planets. Let me give a few examples of what might be called comparative planetology.

There is now observational evidence to support an idea I first proposed in 1960: that the high temperatures on the surface of Venus are due to a runaway "greenhouse effect" in which water and carbon dioxide in the planetary atmosphere impede the emission of thermal radiation from the surface to space. The surface temperature rises to the point where there is an equilibrium between the visible sunlight arriving at the surface and the infrared radiation leaving it; this higher surface temperature results in a higher vapor pressure of the greenhouse gases, carbon dioxide and water, and the process continues until all the carbon dioxide and water is in the vapor phase, producing a planet with a high atmospheric pressure and a high surface temperature. The reason Venus has such an atmosphere and the earth does not seems to be that Venus receives a little more sunlight than the earth. If the sun were to become brighter or the earth's surface and clouds were to become darker, could the earth become a replica of this classical vision of hell? Venus may be a cautionary tale for our technical civilization, which has the capability to profoundly alter the environment of our small planet.

In spite of the expectations of almost all planetary scientists, Mars turns out to be covered with thousands of sinuous, tributaried channels that are probably one or two billion years old. Whether they were formed by running water or by running carbon dioxide, such channels could not be carved under present atmospheric conditions; they require

MAIN PROPERTIES of the planets are summarized in table on the opposite page. Drawings at top show the planets' sizes with respect to the sun. Minus sign in front of the rotation period of Venus and Uranus indicates that those planets rotate in a direction opposite to the direction in which the other planets rotate. Eccentricity of a planet's elliptical orbit is customarily expressed as the distance between the two foci divided by the length of the major axis. Oblateness, the amount by which a rotating body is flattened, is given as the difference between the equatorial and the polar diameters divided by the equatorial diameter of body. Data that are presented in this chart and the one on the next page were compiled from a number of sources with the assistance of Jay D. Goguen of Cornell University.

NAME OF SPACECRAFT	DATE OF LAUNCH	DESTINATION	DATE OF ENCOUNTER	NEAREST APPROACH (KILOMETERS)	STATUS OF MISSION
VENERA 1	2/12/61	VENUS	—	100,000	Radio contact lost 7.5 million kilometers from earth.
MARINER 1	7/22/62	VENUS	—	—	Booster rocket deviated from course and was destroyed by range safety officer.
MARINER 2	8/26/62	VENUS	12/14/62	35,000	First flyby of another planet; found high temperature (400 degrees Celsius) arises from surface, not atmosphere; no evidence of magnetic field.
MARS 1	11/1/62	MARS	—	190,000	Radio contact lost 106 million kilometers from earth.
ZOND 1	4/2/64	VENUS (?)	—	—	Radio contact lost within month after launch.
MARINER 3	11/5/64	MARS	—	—	Shroud failed to jettison; radio contact lost soon after launch.
MARINER 4	11/28/64	MARS	7/14/65	10,000	First flyby of Mars; returned 22 television pictures of Martian surface, other data.
ZOND 2	11/30/64	MARS	—	—	Radio contact lost 5/2/65.
VENERA 2	11/12/65	VENUS	2/27/66	24,000	Passed Venus but failed to return data.
VENERA 3	11/16/65	VENUS	3/1/66	LANDED	First spacecraft to land on another planet; failed to return data.
VENERA 4	6/12/67	VENUS	10/18/67	LANDED	First on-site measurements of temperature, pressure and composition of Venusian atmosphere; probe transmitted data during 94-minute parachute descent.
MARINER 5	6/14/67	VENUS	10/19/67	4,000	Measured structure of upper atmosphere of Venus during flyby.
VENERA 5	1/5/69	VENUS	5/16/69	LANDED	Probes transmitted data on pressure, temperature and composition of atmosphere during parachute descent; missions similar to that of *Venera 4*. First successful landing on another planet.
VENERA 6	1/10/69	VENUS	5/17/69	LANDED	
MARINER 6	2/25/69	MARS	7/31/69	3,390	Flyby obtained infrared and ultraviolet spectra of atmosphere; returned 76 pictures of surface, other data.
MARINER 7	3/27/69	MARS	8/5/69	3,500	Mission identical with that of *Mariner 6*; returned 126 pictures of surface, 33 of south-polar region.
VENERA 7	8/17/70	VENUS	12/15/70	LANDED	Mission similar to those of *Venera 4*, *Venera 5* and *Venera 6*.
MARINER 8	5/8/71	MARS	—	—	Malfunctioned during launch; crashed in Atlantic.
MARS 2	5/19/71	MARS	11/27/71	LANDED	Orbiter achieved Mars orbit; lander crashed to surface.
MARS 3	5/28/71	MARS	12/2/71	LANDED	Orbiter achieved Mars orbit and returned data; descent module soft-landed and transmitted 20 seconds of featureless television data before failing.
MARINER 9	5/30/71	MARS	11/13/71	1,395	First spacecraft to go into orbit around another planet. returned 7,329 pictures of surface, atmosphere, clouds and satellites, together with other data.
PIONEER 10	3/3/72	JUPITER	12/4/73	131,400	Successfully traversed asteroid belt; investigated interplanetary medium, Jovian magnetosphere and atmosphere; returned more than 300 pictures of Jovian clouds and satellites; first spacecraft to use gravity-assisted trajectory; first man-made object to escape solar system.
VENERA 8	3/26/72	VENUS	7/22/72	LANDED	Survived Venusian surface conditions for 50 minutes; determined radioactive content of surface; on entry measured winds and sunlight penetrating clouds.
PIONEER 11	4/6/73	JUPITER SATURN	12/3/74 (J) 9/79 (S)	46,400 (J)	Second Jupiter flyby; now en route to Saturn, then to leave solar system.
MARS 4	7/21/73	MARS	1/74	?	Went into orbit around Mars; returned photographs of surface and other data.
MARS 5	7/25/73	MARS	1/74	?	
MARS 6	8/5/73	MARS	2/74	LANDED	Descent module failed at touchdown, entry data suggest high argon content of atmosphere.
MARS 7	8/9/73	MARS	—	—	Radio contact lost 3/12/74.
MARINER 10	11/3/73	VENUS MERCURY	2/5/74 (V) 3/29/74 (M)	5,800 (V) 700 (M)	First probe of Mercury; returned more than 8,000 pictures and other data from Venus and Mercury; re-encountered Mercury 9/21/74 and 3/16/75.
VIKING 1	8/75	MARS	6/76	TO LAND	Orbiter to study atmosphere and photograph surface, lander to study atmosphere at surface, investigate surface geology and chemistry and test soil for signs of extraterrestrial life.
VIKING 2	9/75	MARS	8/76	TO LAND	
MARINER 11	8/77	JUPITER SATURN	1979 (J) 1981 (S)	?	To conduct comparative studies of two outer planets and their 23 satellites; to investigate nature of Saturn's rings; to measure interplanetary medium out to Saturn's orbit; 20,000 photographs planned.
MARINER 12	9/77	JUPITER SATURN	1979 (J) 1981 (S)		
PIONEER 12	5/78	VENUS	12/78	TO LAND	Orbiter to study interaction of atmosphere with solar wind over one 243-day period; "bus" to drop three small probes toward surface, then relay data to earth as they enter atmosphere
PIONEER 13	8 78	VENUS	12/78	TO LAND	

much higher pressures and probably higher polar temperatures. And so the channels bear witness to at least one epoch and perhaps many previous epochs of milder conditions on Mars, implying that there have been major climatic variations over the history of the planet. We do not know whether such variations are the result of internal causes or of external ones. If the causes are native to Mars, it becomes important to learn whether the earth might, perhaps even as a result of the activities of man, be subject to climatic excursions of Martian magnitude. If the Martian climatic variations were the result of external causes (perhaps variations in the luminosity of the sun), then a correlation of Martian paleoclimatology and terrestrial paleoclimatology would be extremely interesting.

Mariner 9 arrived at Mars in the midst of a great global dust storm, and its data make it possible to determine whether such storms heat a planetary surface or cool it. Any theory with pretensions to predicting the climatic consequences of an increase in the abundance of finely divided particles in the earth's atmosphere had better be able to provide the correct answer for that dust storm on Mars. In fact, drawing on our *Mariner 9* experience, James B. Pollack of the Ames Research Center of the National Aeronautics and Space Administration and Owen B. Toon and I at Cornell University have calculated the effects of single and multiple volcanic explosions on the earth's climate and have been able to reproduce, within the limits of experimental error, the climatic effects that were observed after actual volcanic explosions. The perspective of planetary astronomy, which alone enables us to view a planet as a whole, seems to be good training for studies of the earth. As another example of the contribution made by planetary studies to terrestrial problems, one of the main groups investigating the effect on the earth's ozone layer of the injection into the atmosphere of fluorocarbon propellants from aerosol cans is one headed by Michael B. McElroy of Harvard University—a group that cut its teeth on the physics and chemistry of the atmosphere of Venus.

INTERPLANETARY SPACECRAFT already launched by the U.S. and the U.S.S.R. are listed on the opposite page, along with several projected U.S. missions that are in the advanced planning stage. Excluded are earth-orbiting vehicles, lunar missions and the probes of the interplanetary medium.

We now know from space-vehicle observations something of the density of impact craters of different sizes on Mercury, the moon, Mars and its satellites, and radar studies are beginning to provide such information for Venus. Although the surface of the earth has been heavily altered by wind and water and by crustal folding and faulting, we also have some information about craters on the surface of the earth. If the population of objects that produced such impacts were the same for all these planets, it might be possible to work out both the absolute and the relative chronology of various cratered surfaces. The trouble is that we do not yet know whether the impacting objects are from a common source (for example the asteroid belt) or are of local origin (for example rings of debris swept up in the final stages of planetary accretion).

The heavily cratered lunar highlands speak to us of an early epoch in the history of the solar system, when the frequency of cratering was much higher than it is today; the present population of interplanetary debris fails by a large factor to account for the density of the highland craters. On the other hand, the lunar maria, or "seas," show a much lower crater density, which can be explained quite well by the present population of interplanetary debris: mostly asteroids and possibly dead comets. For planetary surfaces that are not so heavily cratered it is possible to determine something of the absolute age, a great deal about the relative age and in certain cases even something about the distribution of sizes in the population of objects that made the craters. On Mars, for example, we find that the flanks of the large volcanic mountains are almost free of impact craters, implying their comparative youth: they have not been around long enough to have accumulated much in the way of impact scars. That is the basis for the hypothesis of comparatively recent Martian volcanism.

The ultimate objective of comparative planetology, it might be said, is something like a vast computer program into which we insert a few input parameters (perhaps the initial mass, composition and angular momentum of a protoplanet and the population of neighboring objects that strike it) and then derive the complete evolution of the planet. We are far from having such a deep understanding of planetary evolution at present, but we are much closer than would have been thought possible only a few decades ago.

In addition every new set of discoveries raises a host of questions we were not until now even able to ask. I shall mention just a few of them.

The initial radar glimpse of the craters of Venus shows them as being extremely shallow. There is no liquid water to erode Venus' surface, and the lower atmosphere seems to move so slowly that its winds may not be strong enough to fill the craters with dust. Could the craters of Venus be filled by the slow collapse of very slightly molten walls, flowing like pitch?

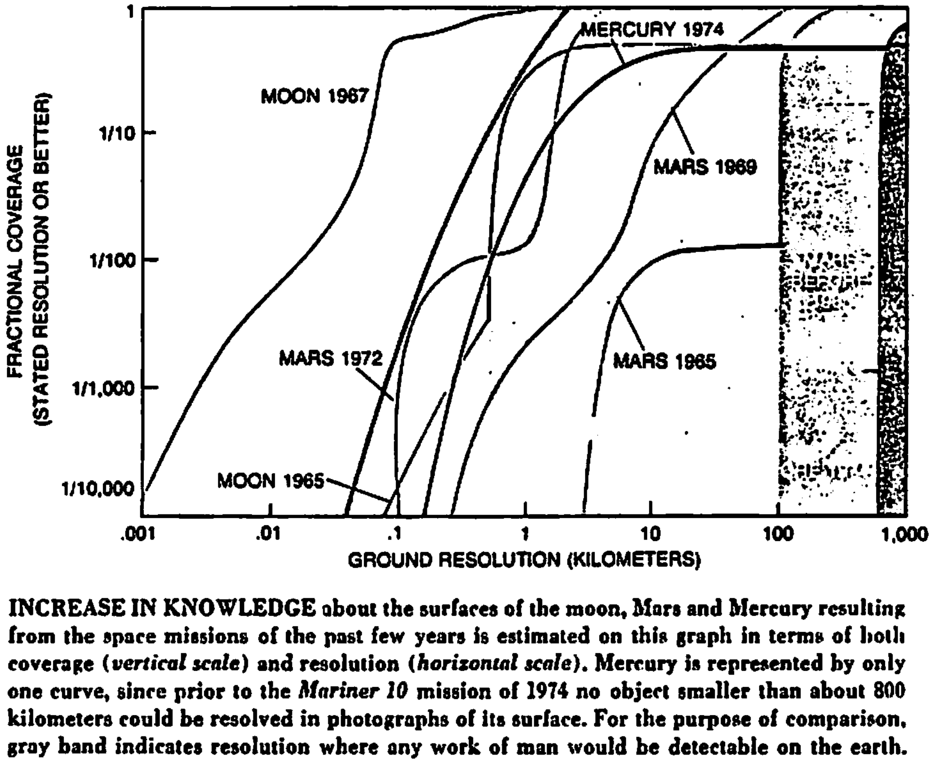
The most popular explanation for the generation of planetary magnetic fields invokes rotation-driven convection currents in an electrically conducting planetary core. Mercury, which rotates only once every 59 days, was expected to have no detectable magnetic field, but *Mariner 10* discovered one. Apparently a serious reappraisal of theories of planetary magnetism is in order.

Only Saturn has rings. Why?

There is an exquisite array of longitudinal sand dunes on Mars, nestling against the interior ramparts of the large eroded crater Procter. In Colorado, in the Great Sand Dunes National Monument, similar sand dunes nestle in a curve of the Sangre de Cristo Mountains. The Martian dunes and the terrestrial ones have the same total extent, the same dune-to-dune spacing and the same dune heights. Yet the Martian atmospheric pressure is only a two-hundredth of the pressure on the earth, so that the winds needed to push the sand grains around must be 10 times stronger than those on the earth; moreover, the distribution of particle sizes may be quite different on the two planets. How then can the dune fields produced by windblown sand be so similar?

Observations made from *Mariner 9* imply that the winds on Mars at least occasionally exceed half the local speed of sound. Are the winds ever much stronger? And if they are, what is the nature of a transonic meteorology?

There are pyramids on Mars that are about three kilometers across at the base and one kilometer high. They are not likely to have been constructed by Martian pharaohs. The rate of sandblasting by wind-transported grains on Mars is perhaps 10,000 times greater than the rate on the earth because of the greater speeds necessary to move particles in the thinner Martian atmosphere. Could the facets of the Martian pyramids have



INCREASE IN KNOWLEDGE about the surfaces of the moon, Mars and Mercury resulting from the space missions of the past few years is estimated on this graph in terms of both coverage (*vertical scale*) and resolution (*horizontal scale*). Mercury is represented by only one curve, since prior to the *Mariner 10* mission of 1974 no object smaller than about 800 kilometers could be resolved in photographs of its surface. For the purpose of comparison, gray band indicates resolution where any work of man would be detectable on the earth.

been eroded by millions of years of such sandblasting from more than one prevailing wind direction?

The moons of the outer solar system are almost certainly not replicas of our own rather dull satellite. Many of them have such a low density that they must consist largely of ices of methane, ammonia or water. What will their surfaces be like close up? How do impact craters erode on an icy surface? Might there be volcanoes of solid ammonia with lavas of liquid ammonia trickling down their sides? Why is Io, the innermost large satellite of Jupiter, enveloped in a cloud of gaseous sodium? Why is one side of Saturn's moon Iapetus six times brighter than the other? Is it because of a particle-size difference? A chemical difference? How did such differences become established, and why did they become established on Iapetus and nowhere else in the solar system in such a symmetrical way? The gravity of Saturn's largest moon, Titan, is low enough and the temperature of the upper atmosphere is high enough for the hydrogen in the atmosphere to escape rapidly into space. Yet the spectroscopic evidence suggests that a substantial quantity of hydrogen remains on Titan. Why?

110

Beyond Saturn the solar system is still almost literally clouded in ignorance. Our feeble telescopes have not even reliably determined the periods of rotation of Uranus, Neptune and Pluto, much

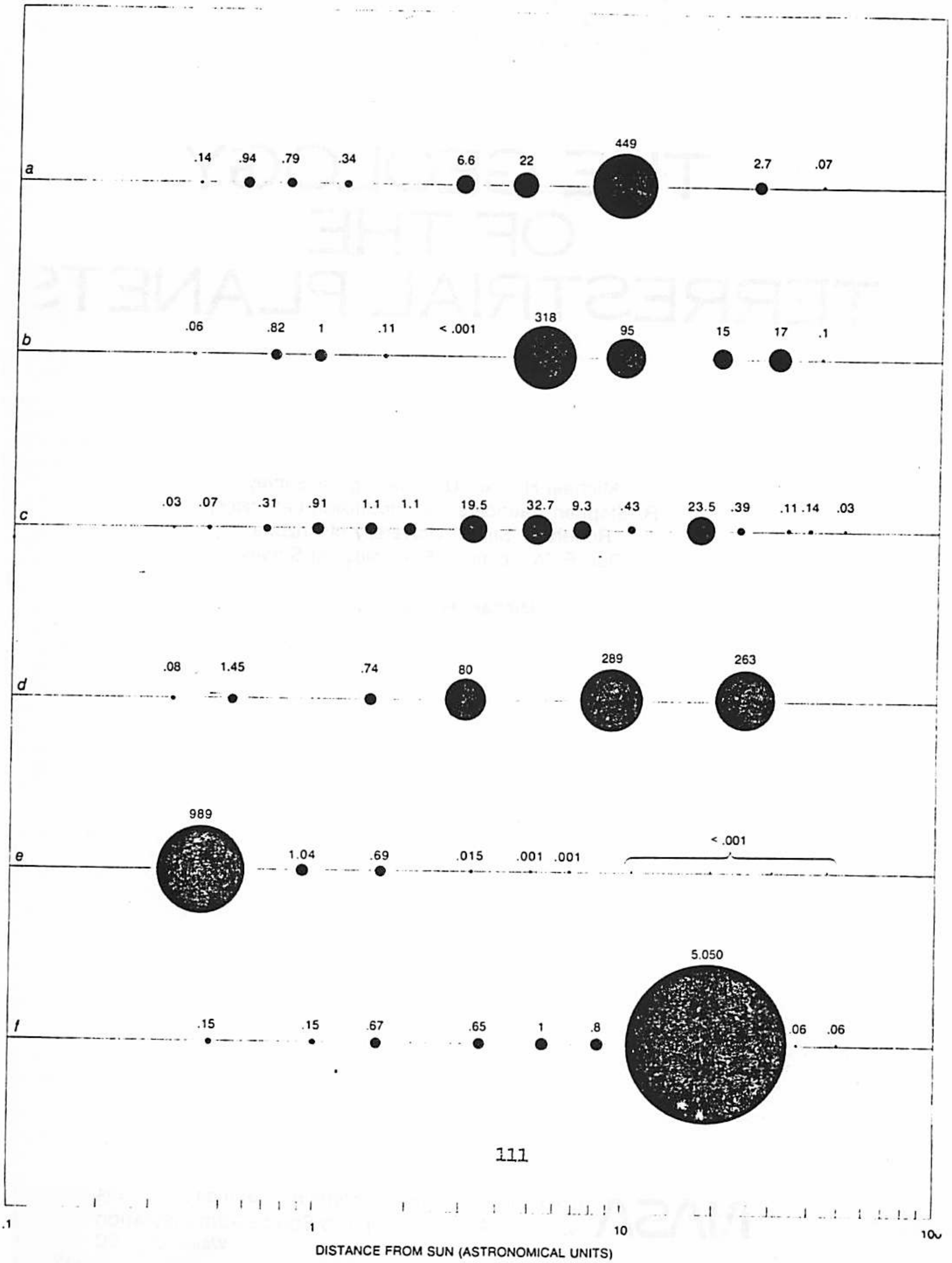
mospheres and the nature of their satellite systems.

One of the most tantalizing issues, and one that we are just beginning to approach seriously, is the question of organic chemistry and biology elsewhere in the solar system. The issue of whether there are organisms both large and small, on Mars in particular, is entirely open. The Martian environment is by no means so hostile as to exclude life, but we do not know enough about the origin and evolution of life to guarantee its presence there—or anywhere else. The three microbiology experiments, the organic chemistry experiment and the camera systems aboard the two Viking vehicles scheduled to land on Mars next summer may provide the first experimental evidence on the matter. The hydrogen-rich atmospheres of places such as Jupiter, Saturn, Uranus and Titan are in significant respects similar to the atmosphere of the earth at the time of the origin of life. From laboratory simulation experiments we know that organic molecules are synthesized in high yield under those conditions. (In the atmospheres of Jupiter and Saturn such molecules would be carried by convection to depths where they would be decomposed by heat, but even there the steady-state concentration of organic molecules may be significant.) In all simulation experiments the application of energy to such atmospheres produces a brownish polymeric material that in many respects in the clouds of Jupiter and Saturn. Titan may be completely covered with .. brownish organic material. It is possibl that the next few years will see majo and unexpected discoveries in the infa science of exobiology.

The principal means for the continue exploration of the solar system over th next decade or two will surely be unmanned planetary missions. Scientifi space vehicles have now been launche successfully to all the planets known t the ancients. If even a small fraction o the missions that are scheduled and hav been proposed are implemented, it i clear that the present golden age o planetary exploration will continue.

Yet even a preliminary reconnai sance of the entire solar system out t Pluto and the more detailed exploratio of a few planets (by, for example, vehicles that will traverse the surface of Mar or penetrate the atmosphere of Jupite will not solve the fundamental proble of solar-system origins. What we nee is to discover other solar systems, pe haps at various stages in their evolutio Advances in ground-based and spac borne instruments over the next two decades may make it possible to detect do ens of planetary systems around nearb single stars. Recent observational studie of multiple-star systems by Helmut Ab and Saul Levy of the Kitt Peak Nationa Observatory suggest that as many as third of all stars have planetary companions. We do not know whether su systems will be like ours or will be buil on very different principles. Richar Isaacman of Cornell and I have calculated a range of possible planetary systems based on a theoretical model originally devised by Stephen H. Dole of th Rand Corporation. The assumptions behind these models are so simple as t make us believe they are unrealistic and yet the range of systems to whicl they give rise is intriguing. The tim may not be far off when we shall hav observational information on the distribution in space of various types of planetary systems. We may then be able t echo Huygens: "What a wonderful an amazing Scheme we have here of th magnificent Vastness of the Universe So many Suns, so many Earths!"

Centuries hence, when current socia and political problems may seem as remote as the problems of the Thirt Years' War are to us, our age may b remembered chiefly for one fact: It wa the time when the inhabitants of th earth first made contact with the vas cosmos in which their small planet i



ALTERNATIVE PLANETARY SYSTEMS were calculated by the author and his colleague Richard Isaacman of Cornell on the basis of a theoretical model originally devised by Stephen H. Dole of the Rand Corporation. The assumptions behind such calculations are probably too simple at present, but the exercise is thought to be suggestive. Numbers above hypothetical planets denote mass in multiples of earth's mass. Horizontal scale indicates semimajor axes of planets' elliptical orbits. System labeled *b* is our solar system.

VENUS

Carr, M.H. (ed.)
The Geology of the Terrestrial
Planets "Venus" (Chapter 4)
Published by N.A.S.A
Used by permission

THE GEOLOGY OF THE TERRESTRIAL PLANETS

Michael H. Carr, U.S. Geological Survey
R. Stephen Saunders, Jet Propulsion Laboratory
Robert G. Strom, University of Arizona
Don E. Wilhelms, U.S. Geological Survey

Michael H. Carr, Editor

NASA Scientific and Technical Information Branch 1984
National Aeronautics and Space Administration
Washington, DC

4

VENUS

R. Stephen Saunders and Michael H. Carr

INTRODUCTION

Venus is of special interest to the geologist for the contrast it presents with Earth and for the clues it provides concerning Earth's origin and evolution. Both bodies are of similar size, composition, and distance from the Sun yet they have evolved to remarkably different states; one developed into a prolific haven for life, the other into a sterile inferno. Unfortunately, we are far from understanding why this happened. Geologic study of Venus is particularly difficult. The main obstacle is its thick atmosphere which prevents us from seeing the surface with either optical telescopes or conventional spacecraft imaging systems. Only in recent years, with the development of radar imaging techniques and deployment of landers on the surface, have we been able to get a glimmer of Venus' surface, and to begin its geologic study.

Venus is the brightest object in the heavens, after the Sun and the Moon. Because it lies inside Earth's orbit, its angular distance from the Sun is always small (<47°), so it is visible only at dawn and dusk. In ancient Egypt, Venus was seen as two objects—the evening star and the morning star. To the Phoenicians, Venus was Astarte; to the Chaldeans (Babylonians) it was Ishtar. The Chinese called it Tai-pe, which means Beautiful White One. The most ancient observations known were recorded by the Babylonians on the Venus Tablets around 1900 B.C. The correct reference to Venus is a matter of some debate. Moore (1961) used *Cytherean* as the adjectival form, a word derived from the old Sicilian name for Venus. As that term is probably unfamiliar to most, the more straightforward *Venusian* is used here.

Observations of Venus played a key role in testing the Copernican theory of the solar system against the hypothesis that Earth lies at the center. The test proposed by Copernicus was to observe the phases on Venus and thus demonstrate that it revolves around the Sun. Galileo observed the phases and revealed his discovery in a message to Kepler, which Galileo coded to allow more time to confirm the discovery and to establish his priority. The message to Kepler was "Haec immatura, a me, iam frustra, leguntur—o. y.," which translates to "These things not ripe are read by me.", ignoring the two letters "o. y." that do not fit into the anagram. The message can be rearranged to read "Cynthia figuras aemulatur Mater Amorum," the real message, or "The Mother of Love emulates the phases of Cynthia."

EARLY TELESCOPIC OBSERVATIONS

Astronomers have been observing Venus through the telescope since the instrument was invented in the early part of the seventeenth century. Among the best known of the early observers is Christiaan Huygens, who reported that Venus is featureless. However, many subsequent observers noted markings of various kinds, including bright patches and polar caps, although the best observers reported none that were reliably repeated. The watchers of the bright spots attempted to determine the length of the Venusian day, and a surprising number deduced rotation periods nearly the same as those of Earth and Mars. These included Giovanni Cassini (23 hr 21 min), J. J. Cassini (23 hr 28 min), Schroter (23 hr 21 min), and Trouvelot (24 hr).

The most extreme estimate of the Venusian day was by Schiaparelli, the discoverer of the Martian canals, who suggested in 1891 that Venus has a rotational period equal to its orbital period of 224 days 16 hr

48 min. A short time later Lowell built his observatory in Flagstaff, Arizona, primarily to study Schiaparelli's Martian canals. Lowell also saw canals on Venus and in 1897 published a map of them. Although Lowell's Venusian canals were never confirmed by others using large telescopes, Lowell appears to have been convinced that the markings were real and further, that the 225 day period was correct. Moore (1961) argued, in discussing Lowell's beliefs, that Schiaparelli's estimate of close to 225 days could not be correct since thermal observations of the dark side of Venus gave no indication that the dark side is colder than the sunlit side. In the early 1960s several astronomers documented retrograde atmospheric motions of 100 m/s at the equator, which gives a four day cloud circulation period (Young and Young, 1975), but still gave no indication as to the rotation rate.

The first published speculation that Venus has an atmosphere appears to have been by Schroter in 1796. The first conclusive indication of its thickness was through spectroscopic work in the 1930s, which revealed at least 300 times as much CO_2 as in Earth's atmosphere. In the early literature, the high albedo of the planet was variously attributed to a variety of causes including dust clouds, formaldehyde, or salts such as NaCl or MgCl from dried oceans.

Another frequently cited phenomenon was the Ashen Light. Apparently first described by Father Johannes Riccoli in 1643 (Moore, 1961), the Ashen Light is a faint phosphorescence of the night hemisphere, normally seen when Venus is a thin crescent. The light appears to vary in intensity and has been ascribed to extensive twilight or to electrical phenomena such as aurora or lightning activity.

Many of the early observers were concerned with predicting and accurately measuring the timing of transits—the passages of Venus across the Sun's disc. Transits of Venus are infrequent because of the relatively large inclination of its orbit (3.4°). Currently, they occur in pairs separated by eight years (1631, 1639; 1761, 1769; 1874, 1882; 2004, 2012). Kepler was the first to predict a transit, the one on December 7, 1631, which Gassendi unsuccessfully attempted to observe. Horrocks was successful, however, in predicting and observing the next transit in 1639.

By the mid-eighteenth century there was great interest in careful observation of the transits of Venus for the purpose of determining the astronomical unit, and the transits of 1761 and 1769 were widely observed for this purpose. The 1768 expedition of Captain James Cook, for example, was charged by the Royal Society with the responsibility of observing the transit of Venus from Tahiti. The observations were made on June 3, 1769.

The most poignant story among those of the early transit watchers is related by Moore (1961). The French astronomer Guillaume Legentil set out for Pondicherry, India, in 1760. Delayed by the Seven Years' War, he arrived too late for the 1761 transit and decided to wait eight years in India for the next one. Unfortunately, on the day of the transit, June 3, 1769, it was cloudy, and a presumably despondent Legentil departed India for home, but was shipwrecked twice en route. Eleven years after setting out for India, he reached Paris only to discover that he had been presumed dead and his property distributed to his heirs.

ORBITAL AND ROTATIONAL MOTIONS

Venus moves around the Sun in an orbital period of 224 days 16 hr 48 min, an eccentricity of 0.0068, and an inclination of 3° 24′. The eccentricity is less than any other planet, and the inclination is greater than that of any other planet's but Mercury and Pluto. There are about 584 days between inferior conjunctions, the times of closest approach of Venus and Earth. The exact time varies by about four days because of orbit eccentricities. At the time of closest approach, the Earth–Venus distance is 44 million km, and the Venus disc, although not visible, is 64″ across. At superior conjunction, when Venus and Earth are on opposite sides of the solar system, Venus has an angular diameter of 9.5″. As seen from Earth, the greatest angular distance of Venus from the Sun is at dichotomy when the angular separation is 47°. When seen in the evening, Venus is waning; in the morning it is waxing as it approaches superior conjunction. It is most brilliant at an angular distance of 40° from the Sun.

The semi-major axis of the orbit is 0.723 AU, so the planet receives about twice as much solar radiation as Earth. However, Venus has a high albedo (0.71), compared with Earth (0.39) and the Moon (0.07), because of its continuous cloud cover. As a consequence, despite its closer proximity to the Sun, Venus absorbs less radiation than Earth, giving it an effective radiation temperature of only 224 K as compared with Earth's 253 K.

The Venusian day has been only recently determined from radar observation. As noted above, early telescopic estimates ranged from 23 hours to 225 days, and this situation prevailed until the late 1950s. In the early 1960s, radar started to be used to observe Venus. W. B. Smith of the Massachusetts Institute of Technology looked for a Doppler shift in the radar echo as evidence of rotation. Although the period was not determined, there was an indication that the motion was retrograde. During the conjunction of 1962, several workers (Carpenter, 1964; Goldstein, 1964; Muhleman, 1964; Drake, 1964) confirmed that Venus' rotation was retrograde and estimated the period as close to 250 days. The estimate was later refined to 242.6 ± 0.9 days (Goldstein, 1965). The best current estimate is very close to 243 days, with the rotation axis inclined at 177°. Because of the similarity in the orbital and rotational periods, the sidereal day differs significantly from the solar day, which lasts 116.8 Earth-days. To an observer on the surface of Venus, the Sun would rise in the west, and daylight would persist for 58.4 Earth-days with a sky resembling that of a dark overcast day on Earth.

EARTH-BASED RADAR OBSERVATIONS OF THE SURFACE

Radar observations from Earth, besides providing us with a measure of the Venusian day, gave us our first information about the planet's surface. Radar studies of Venus have since been carried out in the U.S. mainly at three observatories: Arecibo in Puerto Rico, which has a 300 m diameter antenna using mostly 70 cm wavelength; Goldstone in the Mohave Desert of California, which has a 64 m antenna and operates primarily at 12.5 cm wavelength; and Haystack in Massachusetts, which has a 43 m antenna that is used at 3.8 cm wavelength. High resolution radar data can only be obtained around inferior conjunction, or about every 19 months. At inferior conjunction, the same hemisphere of Venus always faces the Earth.

The first Venus observations were ranging for ephemeris development. These range determinations had about one kilometer accuracy by the late 1960s. By 1970, images had been obtained by each of the three major observatories. These images had a resolution of about 50 km and showed only that Venus has fixed features and that the radar reflectivity varies from place to place. The technique used is to separate the echo into range bins representing circular regions on the planet equidistant from Earth. The range bins are subdivided by Doppler provided by the slow rotation of Venus. This still leaves a north-south ambiguity, in that each resolution element in the northern radar hemisphere has a counterpart with the same range and Doppler in the southern hemisphere. This ambiguity is now removed by an interferometry technique first developed by Rogers and Ingalls (1969, 1970) using the MIT Haystack telescope. The technique was subsequently used by Campbell et al. (1970) at Arecibo and by Goldstein and Rumsey (1970, 1972) at Goldstone. The theory behind the technique was described by Shapiro et al. (1972) and had been previously applied to the Moon (Zisk, 1972). The first images of Venus showed several circular features that resemble impact craters. Reflectivity and altimetry maps have since been made of small regions, showing numerous other surface features (Rumsey et al., 1974; Goldstein et al., 1976, 1978).

The highest resolution mode of the Goldstone images is about 10 km linear radar resolution and the images typically cover about 8° radius or regions of the planet about 1600 km in diameter (fig. 4.1). The theory and data processing operations of three station interferometry for obtaining the topography is described in Jurgens et al. (1980). The standard error of their altimetry is about 35 m near the sub-radar point increasing to about one kilometer at 6°. Recent Goldstone radar images (Jurgens et al., 1980) reveal small mountains 30–60 km in diameter, 1–2 km high, with mean slopes of 2.5° to 3.5°. A strong dependence of reflectivity on incidence angle is observed, suggesting that the surfaces are relatively smooth at the 12 cm-wavelength used.

The highest resolution images are those from Arecibo, which range in radar resolution down to 3 km (Campbell and Burns, 1980; Head and Campbell, 1982). These have substantially better resolution than the Pioneer Venus radar images, although obviously restricted to the hemisphere that faces Earth during opposition. Campbell and Burns used Arecibo images to construct a mosaic that covers about 25 percent of the planet's surface at a resolution of 10–20 km (fig. 4.2). The area covered stretches from 260° to 30°E between latitudes 70°N and 50°S.

The radar images show considerable surface detail reflecting mostly roughness variations. Freyja and Akna Montes, north of Lakshmi Planum, Maxwell Montes, and Theia and Rhea Montes are particularly prominent, implying rough surfaces at radar wavelengths. Two types of circular features are present: large, smooth, quasi-circular regions with diameters between 200 and 1300 km, and bright regions generally less than 300 km. Especially intriguing in the

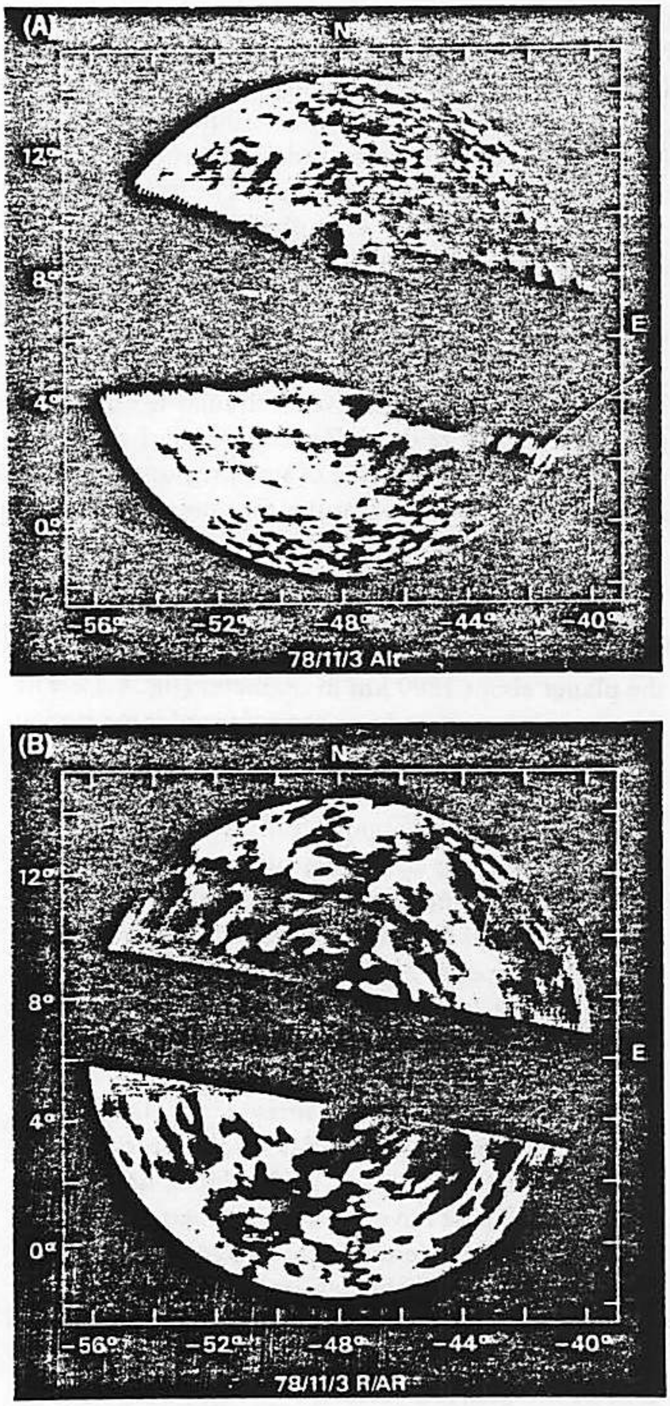


Figure 4.1. Goldstone radar images of a segment of the Venusian lowlands centered at 6°N, 312°E. (A) shows altimetry, the dark areas being low and the bright areas high. In the upper half is a crater 150 km in diameter and 750 m deep. Two bright spots in the lower half are isolated mountains, probably volcanoes. (B) shows variations in radar reflectivity for the same area. The variations are probably caused mainly by surface roughness, with bright areas rougher than dark areas.

Figure 4.2. Mosaic of Arecibo radar images which covers about 40 percent of the northern hemisphere about the 330° longitude. The image shows variations in reflectivity of the surface at radar wavelengths. The prominent dark feature at the top center is Lakshmi Planitia with the very bright Maxwell Montes to the east. The bright marking at the lower left edge is Beta Regio. (Refer to fig. 4.6 to correlate with relief.)

highest resolution pictures are some banded terrain, suggestive of folded or faulted rock sequences (Head and Campbell, 1982).

SPACECRAFT OBSERVATIONS

The first U.S. spacecraft to observe Venus was Mariner-Venus 2, which flew by the planet on December 14, 1962 (table 4.1). It carried two primary experiments: a microwave radiometer to measure the surface temperature and the atmospheric structure and an infrared radiometer to determine the cloud structure and temperatures. The principal scientific results were the determination of a surface temperature of around 400 K and an indication that the magnetic field, or at least the dipole moment, must be less than 10 percent that of Earth. Mariner 2 also provided an order of magnitude improvement in the mass estimate for Venus (e.g., Anderson et al 1964).

Table 4.1. Missions to Venus (from Colin, 1980)

Spacecraft	Launch	Encounter	Type	Encounter characteristics
Mariner 2	August 27, 1962	December 14, 1962	flyby	closest approach: 34 833 km
Venera 4	June 12, 1967	October 18, 1967	bus	burn-up
		entry probe	hard lander, nightside	
Mariner 5	June 14, 1967	October 19, 1967	flyby	closest approach: 4100 km
Venera 5	January 5, 1969	May 16, 1969	bus	burn-up
		entry probe	hard lander, nightside	
Venera 6	January 10, 1969	May 17, 1969	bus	burn-up
		entry probe	hard lander, nightside	
Venera 7	August 17, 1970	December 15, 1970	bus	burn-up
		entry probe	soft lander, nightside	
Venera 8	March 27, 1972	July 22, 1972	bus	burn-up
		entry probe	soft lander, dayside	
Mariner 10	November 3, 1973	February 5, 1974	flyby	closest approach: 5700 km
Venera 9	June 8, 1975	October 22, 1975	orbiter	periapsis: 1560 km; apoapsis: 112 200 km; period: 48 hr. 18 m; inclination 34°10'
		entry probe	soft lander, dayside	
Venera 10	June 14, 1975	October 25, 1975	orbiter	periapsis: 1620 km; apoapsis: 113 900 km; period: 49 hr, 23 m; inclination 29°30'
		entry probe	soft lander, dayside	
Pioneer Venus 1	May 20, 1978	December 4, 1978	orbiter	periapsis: <200 km; apoapsis: 66 000 km; period: 24 hr; inclination: 105°
Pioneer Venus 2	August 8, 1978	December 9, 1978	bus	burn-up, dayside
			entry probes	4 hard landers, dayside and nightside
Venera 11	September 9, 1978	December 25, 1978	flyby	closest approach: 25 000 km
		entry probe	soft lander, dayside	
Venera 12	September 14, 1978	December 21, 1978	flyby	closest approach: 25 000 km
		entry probe	soft lander, dayside	
Venera 13	October 30, 1981	March 1, 1982	entry probe	soft lander
Venera 14	November 4, 1981	March 5, 1982	entry probe	soft lander
Venera 15	June 2, 1983	October 10, 1983	orbiter	24 hr orbit, 1000 km periapsis
Venera 16	June 7, 1983	October 4, 1983	orbiter	24 hr orbit, 1000 km periapsis

Mariner 5, a spare Mariner-Mars 1964 spacecraft, was launched on June 14, 1967, and arrived at Venus on October 19, 1967. The main objective was to investigate the atmosphere, ionosphere, and magnetosphere of Venus; for the geosciences, an important result was an improved radius. Previous telescopic measurements had yielded a radius of 6120 ± 7 km (DeVaucouleurs, 1964). The radar determined radius was 6056 km (Ash et al., 1968). By combining the radio tracking of Mariner 5 with Earth-based radar reflections, a radius of 6054 ± 2 km was determined (Anderson et al., 1968). A high priority was to determine the atmospheric density and pressure at the surface. During an exciting five day period in October 1967, Mariner 5 arrived at Venus and the Soviet Venera 4 made a hard landing on the Venus surface. The Soviets put down a pressure sensor for a direct measurement of surface pressure, while the U.S. used radio occultations of the Mariner spacecraft. The surface temperature was about 700 K and the pressure about 100 bars. The cloud top was at 67 ± 10 km, and CO_2 was confirmed as the predominant atmospheric constituent. Measurements made by Venera 4 indicated that the magnetic field of Venus is <0.00001 that of Earth.

Over the next few years the Soviets sent several additional spacecraft to Venus and achieved the first soft landing on another planet when Venera 8 soft landed on the Venus surface on July 22, 1972. In addition to measuring various properties of the atmosphere during entry, the lander was able to measure the radioactivity of the surface materials with a gamma ray spectrometer (Vinogradov et al., 1973), and to obtain a measure of their density (1.5 g/cm^3).

The first spacecraft to orbit Venus were Veneras 9 and 10, placed in orbit in October 1975. They were in elliptical orbits with periapses of about 1500 km. The orbiters were not tracked continuously and were operated in a 3-axis stabilized mode only near periapsis. Elsewhere in the orbit they rotated slowly about the Venus–Earth direction. Veneras 9 and 10 also released landers to the surface and both successfully soft-landed. These landers carried an imaging system in addition to the gamma ray spectrometer/densitometer carried on Venera 8, and both spacecraft returned excellent pictures from the surface. In 1978 the Soviets sent two additional spacecraft, Veneras 11 and 12, to Venus, but neither returned useful data.

The U.S. also sent two spacecraft, Pioneer-Venus' 1 and 2, to Venus in 1978 (Colin, 1980). Pioneer 1 was an orbiter carrying a wide array of scientific instruments to examine the atmosphere, the surface, and the interaction of the planet with the solar wind. Of main geologic interest was a radar mapper designed to systematically map the elevation of the surface and its roughness. Pioneer 2 consisted of five separate spacecraft—a bus, a large atmospheric probe, and three small probes. The probes entered the atmosphere of Venus at different locations, two on the night side and two on the day side. Each carried several instruments designed primarily to determine the composition and structure of the atmosphere.

In 1981 the Soviets launched two more spacecraft to Venus, Veneras 13 and 14. In March 1982, both successfully soft-landed on the surface and photographed their surroundings. Both spacecraft carried an X-ray fluorescence instrument, which provided the first chemical analyses of the Venusian surface.

The most recent Soviet Venus missions are Venera 15 and 16. Identical spacecraft were placed in polar orbits to image the northern hemisphere using a radar system. Other experiments included a radiometer to measure surface brightness temperature, and altimeter to map surface topography, and an infrared spectrometer to obtain data on atmospheric composition.

The image resolution of Venera 15/16 radar is about 1.5 km. This is about the same as the best currently available Earth-based resolution from Arecibo (Campbell et al., in press). In comparable terms, the NASA Venus Radar Mapper will have a resolution of 180 to 460 m depending on position in orbit from periapsis.

Preliminary releases of the Soviet images show craters and linear tectonic features.

CONSTRAINTS ON THE COMPOSITION OF VENUS

The terrestrial planets are characterized as much by their differences as by their similarities and each must have started out on its unique evolutionary path as a result of conditions that prevailed very early in their history. Before examining the Venusian surface in detail, and trying to reconstruct the geologic history from the surface record, we will briefly examine what might be deduced about the planet's early history from models of formation of the planets. General models for compositional and density trends in the solar system have been developed by Lewis (1972, 1974) and Cameron (1963, 1973). In these models solid dust condensed in equilibrium with gas of the solar nebula at temperatures and pressures that varied with heliocentric distance (see Introduction). The dust grains accreted to form larger objects and ultimately planets. The more volatile the elements, or compounds, the farther out they condensed.

Condensation within a solar nebula that is cooling outward and with time, satisfactorily explains the gross compositional trends in the solar system, but the rare gases appear to follow an inverse trend. Pioneer-Venus discovered a large excess of nonradiogenic rare gases in the Venusian atmosphere. Neon and the nonradiogenic argon isotopes ^{36}Ar and ^{38}Ar are one hundred times more abundant on Venus than on Earth. In contrast, the mixing ratios for nitrogen and carbon dioxide are remarkably similar for the two planets. The volatile rare gases decrease in abundance outward rather than increase as the simple model predicts. The cause must be connected with the origin of planetary atmospheres. Pollack and Black (1979) outlined the basic theories of origin.

(1) Condensation/capture. Planets acquired their atmospheres from the primordial nebula or solar wind. Either source must have had a composition different from today's Sun.

(2) Cometary impact. Volatile-rich objects entered the inner solar system from the more distant parts where they had condensed. The objects accreted onto the inner planets, providing them with atmospheres. This model does not, however, explain the observed rare gas gradients.

(3) Grain-accretion. Pollack and Black (1979) proposed a grain-accretion hypothesis in which the rare gases became adsorbed on grains which were later accreted by the planets. The solar nebula, they postulated, had a relatively uniform temperature gradient but a steep pressure gradient, decreasing outward. Adsorption of volatiles is pressure dependent so that the grains that accreted nearest the center where the pressures were highest would have more adsorbed volatiles. Most of the non-rare gas volatiles such as nitrogen and water were chemically bound within the grains, and so their abundances were not as sensitive to the pressure gradient as the rare gases. The rare gases were thus fractionated with respect to the other volatiles.

Wetherill (1981) alternatively proposed that the solar wind accretion model best accounts for the observed trends. During the early stages of development of the solar system, material condensed from the nebula gas to form particles which accumulated into larger bodies (planetesimals), which in turn accumulated to form the planets. Wetherill suggested that the excess inert gases were implanted by an enhanced solar wind bombardment during planetesimal growth, and that the decreasing outward rare gas trends were produced by a geometric shading effect, causing implantation to be less efficient with increasing distance from the Sun because of shielding by particles closer in. Each planet preferentially incorporated material that had condensed in its part of the solar system, so that the planets nearer the Sun incorporated relatively more rare gases.

The origin of a planet's atmosphere is strongly coupled to its outgassing history. In general, the atmosphere does not represent the composition of the total inventory of degassed volatiles because most gases react with the surface materials or are lost by exospheric processes. Only the rare gases, except for He, are conserved. The isotopic abundance of rare gases, therefore, provides the best clues concerning atmospheric history but even these do not provide a unique solution. The outgassing history of Venus remains obscure, and will remain so until we obtain much more compositional information about the surface and the interior.

Presently, the composition of the interior of Venus is constrained only by the bulk density and inferences drawn from compositional determination made by the Soviet landers. After making corrections for the effects of compression, the density of Venus appears to be about 2 percent less than if it were compositionally identical to Earth (Ringwood and Anderson, 1977). Ringwood and Anderson argued that even if the high surface temperatures had persisted since the formation of the planet, the thermal wave could have penetrated only about 500 km. Temperature deeper than 500 km would be about the same as on Earth so that the density differences cannot be explained simply on the basis of the high surface temperatures. This argument assumes that heat is lost mainly by conduction, but the conclusion would be the same if heat transport was mainly by convection since the temperature at which mantle materials flow should be similar on both planets. Removing the effect of thermal expansion due to the higher near-surface temperatures on Venus lowers the bulk density somewhat, but still leaves an uncompressed density 1.7 percent smaller than Earth's.

The Lewis (1972, 1974) equilibrium condensation model qualitatively explains why Venus has a lower uncompressed density than Earth. Earth incorporated more sulfur than Venus, so has a higher sulfur to silicate ratio. Since sulfur has a greater atomic weight than the mean atomic weight of the silicates, Venus is less dense. However, Ringwood and Anderson suggested that there are problems with the equilibrium condensation model because they calculated the decrease in density resulting from removing all the sulfur from Earth's core and still concluded that Venus remains less dense. On the other hand, support for equilibrium condensation is provided by a model of sulfur chemistry in the atmosphere-lithosphere system based on Pioneer-Venus (Lewis and Kriemendahl, 1980) which suggested that the Venus lithosphere must have a far lower FeO content than Earth's. This is expected from the equilibrium condensation model, which predicts that FeO would not condense at the Venus distance.

Goettel et al. (1981) showed that the density differences between Earth and Venus can be eliminated by assuming a different temperature structure and a deeper basalt to eclogite transition. In addition, the low FeO content expected for Venus would tend to result in a lower observed (uncorrected) density for

Venus, since the depth to phase transitions involving olivine to spinel and perovskite structures is increased with decreasing FeO/(FeO + MgO) (Phillips and Malin, 1982). A low FeO content could also result in higher interior temperatures because iron-poor olivines have higher melting temperatures, and probably a higher viscosity at a given temperature, than iron-rich olivines. Temperature profiles may therefore stabilize at higher values in the iron-poor Venus mantle as compared with the iron-rich Earth mantle.

VENERA LANDER RESULTS

The Venera spacecraft have provided close-up views at four locations. Venera 9 landed on a rocky, sloping surface at 32°N, 291°E, on the eastern flanks of Beta Regio. The rocks are mostly flat and slab-like, ranging in size up to 70 cm in diameter and 20 cm high (Florensky et al., 1977). Many have fractures and grooves suggestive of layering. Between the rocks is a seemingly fine-grained, low-albedo material. Florensky et al. suggested that the slabs were in the process of slowly moving downslope by mass wasting. Three days later, on October 25, 1975, a second spacecraft was set down at 16°N, 291°E on a flat plain to the south of Beta Regio. The plain appears to be composed of scattered flat outcrops with darker, fine-grained material between. The outcrops are typically 1 to 3 m across and constitute about half the surface. The fine-grained material at this site also had a very low albedo (<3 percent); that of the rocks was close to 5 percent. After both landings, gamma-ray detectors were deployed to determine the radioactivity and density of the surface materials. At the Venera 10 site the material on which the detector rested has an estimated density of 2.8 ± 1 g/cm^3, which is typical of coherent silicate rocks (Surkov et al., 1977). Light levels decreased shortly after landing, probably as a result of the spacecraft engines raising dust. Some of the local materials are thus fine-grained and loosely consolidated. Both sites are at an elevation close to 6053 km. Surface temperatures ranged from 730 to 740 K and the pressure ranged from 88 to 94 atm.

Venera 13 landed at 7°S, 303°E on a flat plain at a similar elevation to the Veneras 9 and 10 sites; the Venera 14 site at 13°S, 311°E is at a somewhat lower elevation (fig. 4.3). Both sites resemble the Venera 10 site except that at the Venera 13 site there was a large area of dark regolith close to the spacecraft. Regolith is almost absent at the Venera 14 site and layering in the outcrops is more obvious (fig. 4.4).

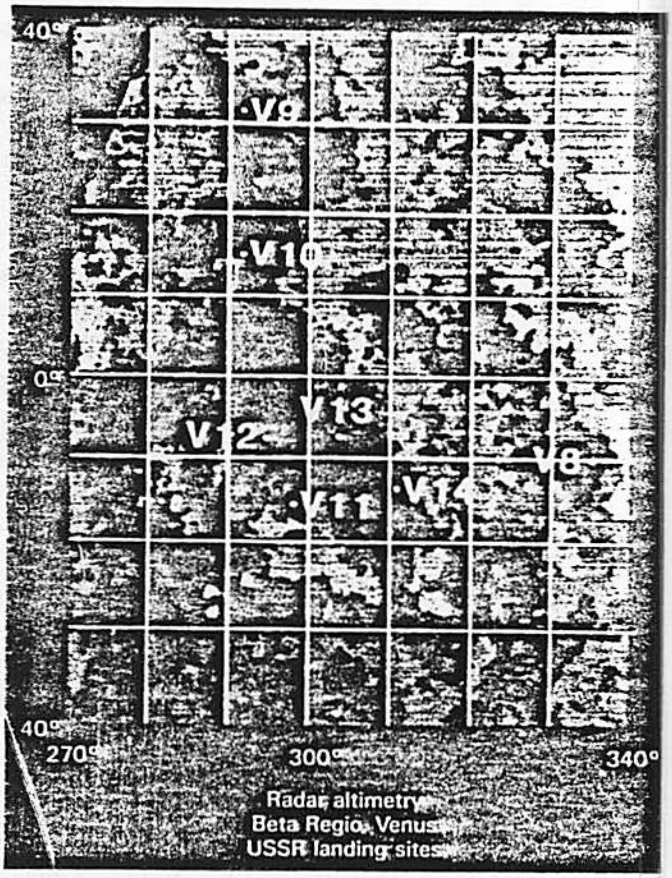


Figure 4.3. Location of the different Venera landing sites. (For key to the color coding of relief, see fig. 4.6.)

The abundances of U, Th, and K suggested by the gamma-ray results of Veneras 8, 9, and 10 (table 4.2) are more similar to those of terrestrial rocks than to those of lunar rocks and meteorites. The Venera 8 abundances are typical of terrestrial continental rocks, whereas the Veneras 9 and 10 abundances more closely approach the values for the terrestrial oceanic crust (see Earth chapter). However, these comparisons should be viewed with caution, for the precisions indicated in table 4.2 may not be a true reflection of the accuracies of the analyses.

More complete chemical analyses were obtained from the X-ray fluorescence instruments on Veneras 13 and 14 (table 4.3). The analyses from both sites are similar except for the K values. They resemble those of terrestrial ocean floor basalts, although the Venera 13 rocks have significantly more K and less Si.

EOLIAN EROSION AND TRANSPORT

At the Veneras 9 and 10 sites surface winds were measured directly by anemometers. They ranged from

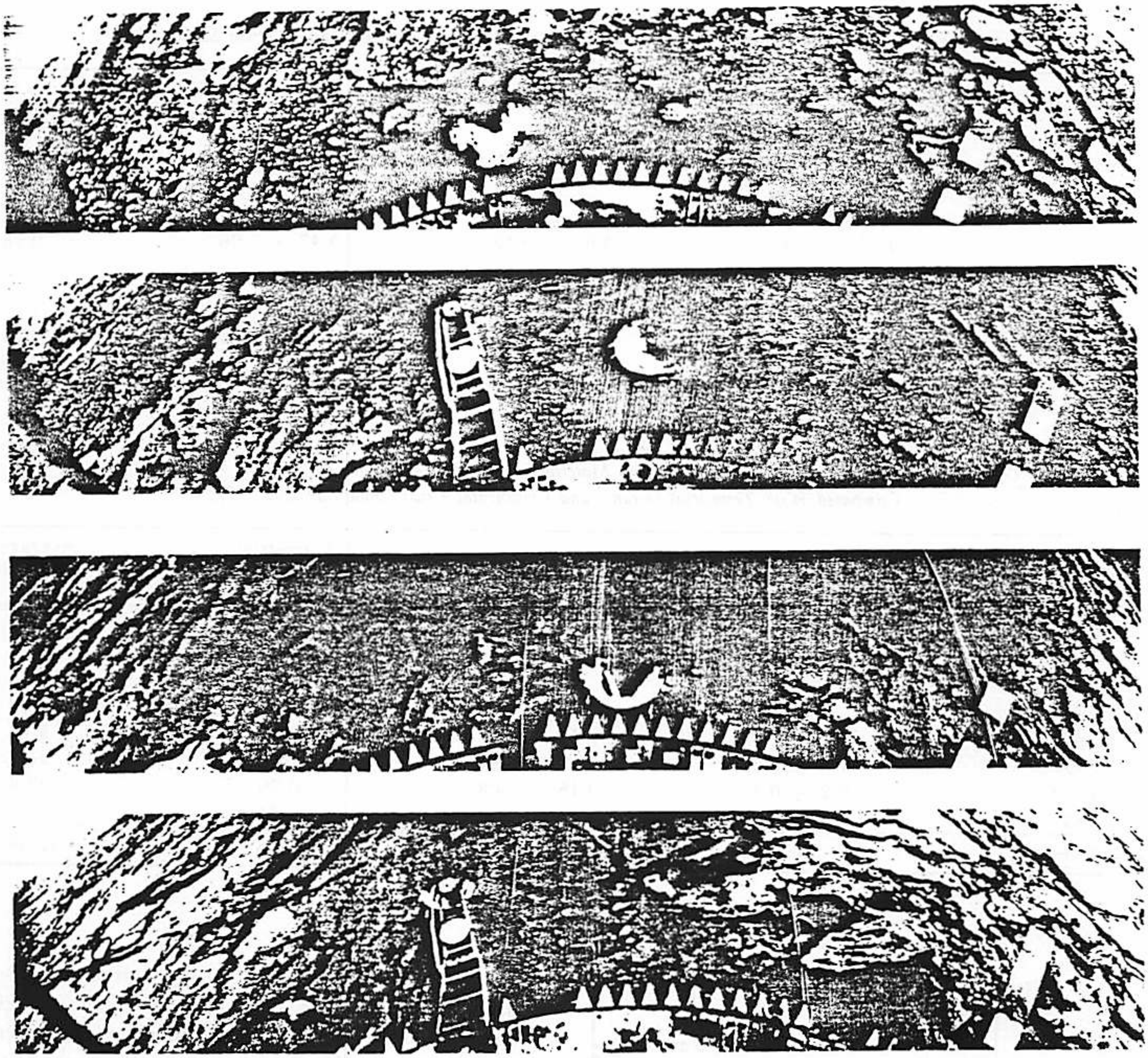


Figure 4.4. Views from the Venera 13 (top) and Venera 14 (bottom) spacecraft. Both spacecraft landed on a level plain with mostly flat, slabby rocks at the surface. The Venera 13 site is also partly covered with a fine-grained regolith.

0.4 to 0.7 m/sec^{-1} at the Venera 9 site and from 0.8 to 1.3 m/sec^{-1} at the Venera 10 site (Keldysh, 1977). In addition, surface winds could be estimated from the tracking of the Venera and Pioneer spacecraft as they descended through the atmosphere. Counselman et al. (1980) showed that the wind profile was similar at all four locations sampled by the Pioneer probes. They ranged from about 100 m/sec at an altitude of 65 km, down to 5 m/sec at an altitude of 10 km, and 1 m/sec at the surface. The dominant motion in the lower atmosphere is retrograde zonal (E–W) rotation. Meridional (N–S) velocities are small throughout the profile and appear to be due mostly to eddies.

These surface winds, though small, are sufficient to raise dust from the surface and to dislodge and move debris. Wind tunnel experiments that simulate Venusian conditions (Williams and Greeley, 1982; Greeley et al., 1982) show that the optimum particle size for wind transport on Venus is 70 μm. The minimum friction wind speed capable of moving particles is about 2 cm/sec (fig. 4.5), close to an order of magnitude less than on the Earth and two orders of magnitude less

Table 4.2. Venera Measurements of Uranium, Thorium, and Potassium Abundances (from Surkov, 1977)

Venera	U($\times 10^{-4}$ wt.%)	Th($\times 10^{-4}$ wt.%)	K(wt.%)	K/U($\times 10^{4}$)
8	2.2 ± 0.7	6.5 ± 0.2	4.0 ± 1.2	+1.65 1.82 −0.85
9	0.60 ± 0.16	3.65 ± 0.42	0.47 ± 0.08	+0.47 0.78 −0.27
10	0.46 ± 0.26	0.70 ± 0.34	0.30 ± 0.16	+1.65 0.65 −0.46

Table 4.3. Composition of Surface Materials at the Venera 13 and 14 Sites on Venus Compared With Terrestrial Oceanic and Continental Crust (from Barsukov, 1982)

Constituent	Venera 13	Venera 14	Terrestrial oceanic basalt	Average continental crust
MgO	10 ± 6	8 ± 4	7.56	2.2
Al_2O_3	16 ± 4	18 ± 4	16.5	16.0
SiO_2	45 ± 3	49 ± 4	51.4	63.3
K_2O	4 ± 0.8	0.2 ± 0.1	1.0	2.9
CaO	7 ± 1.5	10 ± 1.5	9.4	4.1
TiO_2	1.5 ± 0.6	1.2 ± 0.4	1.5	0.6
MnO	0.2 ± 0.1	0.16 ± 0.08	0.26	0.08
FeO	9 ± 3	9 ± 2	12.24	3.5
Total (%)	92.7	95.56	99.86	92.68

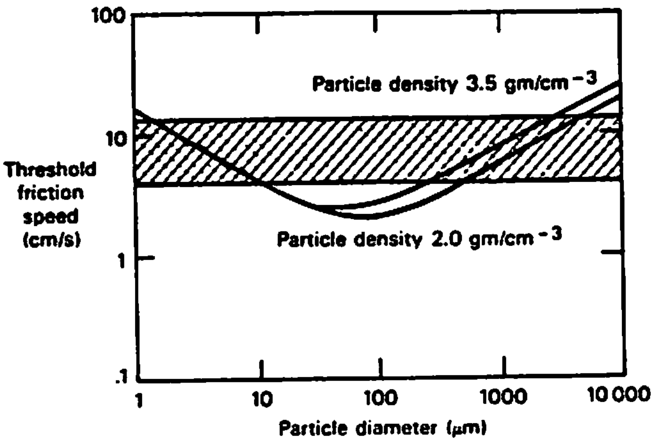


Figure 4.5. The threshold frictional wind speeds required to move particles of different diameters on the Venusian surface. Curves are given for particle densities of 3.5 and 2.0 gm/cm^{-3}, which spans the expected range. The shaded area shows the range of frictional wind velocities expected from the Venera wind data.

than on Mars, a reflection of the high atmospheric density at the Venusian surface. The Venera velocity measurements were made at a height of 1 m above the surface; the frictional wind velocity right at the surface should be about an order of magnitude lower. Thus the Venera measurements imply wind speeds in the 4 to 13 cm/sec range, well above the experimentally derived threshold speed of 2 cm/sec. Theoretical extrapolation of threshold friction speeds as a function of particle diameter to Venusian conditions, based on the work of White et al. (1976) and White (1979), give similar results.

One way in which eolian transport under Venusian conditions appears to differ from Earth and Mars is in the action of the particles after the threshold speeds for saltation are achieved. Particles tend to roll across the surface rather than saltate (Greeley et al., 1982). In other words, their saltation path lengths are very short. This effect which had been predicted on theoretical grounds (White, 1979) must affect the geometry of eolian features such as dunes and ripples. Ripples should, for example, have higher amplitudes and shorter wavelengths than on Earth.

While winds can readily move particulate debris across the Venusian surface, unconsolidated windblown debris appears to constitute only a small part of the surface. The dielectric constant of the Venusian surface has been estimated as 4.7 ± 0.8 (Kuzmin and Marov, 1974) which is more consistent with dry rock than poorly consolidated regolith. The dielectric constant of the Moon, for example, ranges from 2.6 to 2.8, a reflection of the presence of a fine-grained regolith. Dielectric constants of rock materials roughly correlate with density, and the range of dielectric constant for the Venusian surface implies densities that range from 1.3 to 3 g/cm^3 with a mean value of 2.2 g/cm^3, which is more consistent with porous rocks than unconsolidated sediments. Although unconsolidated eolian sediments appear to cover only a small fraction of the surface, many of the exposed rocks may be of eolian origin. The densities and dielectric constants are consistent with cemented fragmental debris. The layering observed in the rocks of the Venera sites could be sedimentary. Warner (1980, 1983) suggested that the low-lying dark plains of Venus are eolian sinks, with mostly cemented eolian deposits at the surface.

CHEMICAL WEATHERING

Several attempts have been made to deduce weathering processes at the Venusian surface by searching for plausible reactions that might occur between gases in the atmosphere and normal rock-forming minerals under Venusian conditions. Urey (1952) suggested that the Venus atmosphere might be buffered by the following reaction

$$\underset{\text{wollastonite}}{CaSiO_3} + CO_2 = CaCO_3 + SiO_2$$

The equilibrium pressure for this reaction at 740 K is 90 bars, which is remarkably close to the conditions on the surface However, as pointed out by Nozette and Lewis (1982), wollastonite is not a common rock-forming mineral and the efficacy of the reaction in buffering the atmosphere is questionable.

The question of how (or if) the Venus atmosphere is buffered is a complex one and intimately coupled with weathering reactions. Rates of reaction between the atmosphere and the surface under the high temperature conditions at the Venusian surface are likely to be high, despite the generally dry conditions. As pointed out by Warner (1983), although the mixing ratio of water vapor in the lower atmosphere, 10^{-3} to 10^{-4}, is up to a factor of 100 lower than in the Earth's atmosphere, the pressure is a factor of 90 higher, so that the partial pressure of water at the surface is about the same on the two planets. Since the temperature is about 450° C higher on Venus, the activity of water is considerably greater. If reaction rates are high, then the rate of fixation of carbon dioxide by reaction with silicates will depend on the rate at which new silicate materials are exposed at the surface by processes such as volcanism, mechanical weathering, and eolian transport. Venus maintains a thick carbon dioxide atmosphere (table 4.4). This could simply be a reflection of extremely low reaction rates of carbon dioxide with silicates at the surface. However, if reaction rates are fast, as has been proposed, then the carbon dioxide must be efficiently recycled back into the atmosphere either by deep-seated processes such as burial followed by volcanism, or by processes at the surface which result in carbon-dioxide-fixing reactions under some conditions, and carbon-dioxide-yielding reactions under other conditions. Several authors have suggested that the lower atmosphere of Venus is complexly buffered and that several gases (H_2O, CO, HF, HCl, SO_2, COS), in addition to CO_2 are buffered by reactions with each other and with the surface. The specific reactions involved are uncertain partly because of the lack of knowledge of the chemistry of the lower atmosphere and the surface rocks. Any equilibrium achieved is likely to be dynamic, depending to some

Table 4.4. Composition of the Atmosphere of Venus (from Nozette and Lewis, 1982)

Species	Mole fraction	Source (6, 7)
CO_2	0.97	Pioneer and Venera
CO	2×10^{-4}	Earth-based
H_2O	2×10^{-5} to 5×10^{-4}	Pioneer and Venera
HCl	$> 10^{-6}$	Earth-based
HF	$\geq 10^{-8}$	Earth-based
N_2	~ 0.03	Pioneer and Venera
$SO_2 + S_2$	$< 3 \times 10^{-4}$	Pioneer and Venera
H_2S	10^{-6}	Pioneer
COS	$> 3 \times 10^{-6}$	Pioneer

extent on the rate of turnover of the surface rocks, and to some extent on reaction rates.

Nozette and Lewis (1982) identified several reactions, involving common rock forming minerals and constituents identified in the lower atmosphere, that could take place at the surface (table 4.5). Some of the reaction could go in opposite directions at different elevations. At high elevations, for example, forsterite reacts with CO_2 to produce magnesite and enstatite but at elevations below 6052 km, where temperatures are higher, magnesite and enstatite recombine, giving off CO_2. Nozette and Lewis suggested that such reversible reactions could contribute to buffering of CO_2 in the atmosphere. In high areas, the surface reacts with the atmosphere to produce weathered products which are removed by creep or by the wind and deposited in the lowlands where the reactions are reversed.

Despite considerable attention given to weathering processes on the Venusian surface, large uncertainties remain, and have little chance of being resolved until better analyses are obtained of the lower atmosphere, and until we have good mineralogical analyses of the surface at various elevations.

GLOBAL TOPOGRAPHY AND SURFACE ROUGHNESS

Because of the near resonance between the rotation period of Venus and Earth's orbital period, the same side of Venus always faces Earth at closest approach. As a result, Earth-based elevation measurements and radar backscatter images are restricted to a limited region—between latitudes 50°S and 75°N, and 130° of longitude between 260° and 30°. Such restrictions do not, of course, apply to spacecraft data, and much of our recent increase in knowledge of planetwide variations in Venus surface properties results from the Pioneer mission. The Pioneer-Venus radar mapper operated at a wavelength of 17 cm and in two modes, altimetry and imaging (Pettengill et al., 1979a). In the altimetry mode, radar reflections from near the subspacecraft point were observed. Signal delay gives a

Table 4.5. Possible weathering reactions on Venus. Reactions marked with an asterisk proceed to the right at high altitudes and to the left in the hotter lowlands. Tremolite may, for example, form at high altitude but break down at low altitude (from Nozette and Lewis, 1982)

$Mg_2SiO_3 + 2CO_2$ forsterite	$2MgCO_3 + SiO_2$ magnesite quartz	(1)
$Mg_2SiO_4 + CO_2$	$MgCO_3 + MgSiO_3$ enstatite	(2)*
$Fe_2SiO_4 + 4\ COS$ fayalite	$2FeS_2 + SiO_2 + 2CO + 2CO_2$ pyrite	(3)*
$MgSiO_3 + 2HF$	$H_2O + MgF_2 + SiO_2$ sellaite	(4)
$CaCO_3 + MgSiO_3 + CO_2$	$CaMg(CO_3)_2 + SiO_2$ dolomite	(5)
$2CaAl_2Si_2O_8 + 5MgSiO_3 + SiO_2 + H_2O$	$Ca_2Mg_5Si_8O_{22}(OH)_2 + 2Al_2SiO_5$ tremolite	(6)*
$2CaMgSi_2O_6 + 3MgSiO_3 + SiO_2 + H_2O$ diopside	$Ca_2Mg_5Si_8O_{22}(OH)_2$	(7)*
$KAlSi_3O_8 + 3MgSiO_3 + 2HF$ orthoclase	$KMg_3AlSi_3O_{10}F_2 + 3SiO_2 + H_2O$ fluorophlogopite	(8)*
$Mg_2SiO_4 + 2CO_2 + SO_2$	$2MgSO_4 + SiO_2 + 2CO$	(9)
$CaAl_2Si_2O_8 + SO_2 + CO_2$	$CaSO_4 + Al_2SiO_5 + SiO_2 + CO$ anhydrite	(10)*
$CaMgSi_2Si_2O_6 + SO_2 + CO_2$	$CaSO_4 + MgSiO_3 + SiO_2 + CO$	(11)*
$Mg_2SiO_4 + 1/2CaMgSi_2O_6 + CO_2$	$1/2CaMg(CO_3)_2 + 2MgSiO_3$	(12)*

measure of surface elevation; backscatter efficiency gives an indication of slopes in the meter- to decameter-scale. Elevations were measured to an accuracy of about 200 m. The areal resolution cell ranged from 23 × 7 km to 101 × 101 km, depending on the altitude of the spacecraft during the observations (Pettengill et al., 1979b, 1980). In the imaging mode, the antenna was rotated to observe the surface on either side of the ground track, thereby enabling scattering efficiencies at relatively high angles (30°–58°) to be measured. At these angles scattering by small-scale (cm) roughness elements dominates.

The orbiter started mapping in December 1978 and continued, with some interruptions, until July 1980 (Colin, 1980). For most of the mission, the spacecraft was in a 24 hr orbit with an inclination of 140°; periapsis was maintained at altitudes between 140 and 190 km. As the planet rotated and moved in its orbit, the ground track of the orbiter was offset in longitude by 150 km at the equator. The orbit periapsis, about which most of the measurements were made, completed two and one-half passes around the planet during the course of the mission. On the second or third periapsis passes over the same area, gaps in the previous coverage were filled or new observations interleafed between the previous tracks. The mission data have now been integrated into relief maps (figs. 4.6 and 4.7) and meter-decameter-scale roughness maps, which cover over 90 percent of the planet's surface at a spatial resolution of 100–200 km. Maps have also been compiled of the cm-scale roughness of the area between 10°S and 50°N, at resolutions as low as 30 km (Pettengill et al., 1980, 1982; Masursky et al., 1980).

Much of the Venusian surface is a rolling plain of relatively uniform elevation. As a result, 60 percent of the planet's surface is within 500 m of the modal radius of 6051.1 km; 20 percent is within 125 m (fig. 4.8). Deviations from the modal value are biased heavily toward high elevation; about 20 percent of the surface has an elevation more than 1 km above the modal value, whereas less than 1 percent has elevation more than 1 km below it. The higher areas form a few continental size masses and smaller islands that stand above the global plain. The highest point so far measured (radius of 6062.1 km) is in the Maxwell Montes at 64°N, 2°E. The lowest point measured is in Diana Chasma at 14°S, 156°E, where the planet's radius is 6049.0 km.

The single mode of the elevation histogram for Venus is in sharp contrast to the strongly bimodal distribution of Earth's topography (fig. 4.8, table 4.6). The distributions for all the terrestrial planets have a high elevation tail, a phenomenon that appears to occur at all scales on natural topographic surfaces. On Mars, Earth, and the Moon, the elevation distributions are largely controlled by the various crustal provinces—continents and ocean basins on Earth; uplands and maria on the Moon; lowland plains, cratered uplands, and Tharsis plateau on Mars. The simple elevation distribution for Venus, and the similarity in its skewness to terrestrial continents, suggest a single global crustal province that is topographically similar to Earth's continents. Venus appears to have the same proportion of rugged, youthful mountains as Earth, although the mechanisms that created and modified them may be very different.

Masursky et al. (1980) divided the Venus surface into three major components on the basis of elevation. Their lowland province includes all areas below the reference radius of 6051.0 km and constitutes 27 percent of the surface. The most extensive lowland basin, Atalanta Planitia, centered at 65°N, 165°E is about the size of the Gulf of Mexico. Its surface averages 1.4 km below the datum, and like the surface of most other low areas of the planet, it is sparsely cratered and radar dark. Masursky et al. suggested that these low areas are volcanic plains analagous to the lunar maria, although as already noted, eolian deposits may preferentially accumulate in these low areas and be widely exposed at the surface. Their presence could in part account for scarcity of crater-like dark rings on this unit as compared with the rolling plains.

The second major component of the surface is the rolling plains, which include all those areas with radii between 6051.0 and 6053.0 km, about 65 percent of the planet. The unit is characterized by Root Mean Square slopes of 1°–3° in the meter- to decameter-scale, and probably gently rolling topography at the kilometer scale. Circular features 20–300 km in diameter, with shallow flat floors and radar-bright rims, are more common on this unit than on any other part of the surface. In addition to these crater-like features are several circular radar bright areas, 200–300 km across and up to 1.5 km high. Some have a central dark region of lower elevation and are tentatively interpreted as volcanoes.

The third major component of the surface is the highlands. These occur in three continent-size areas. The first is Ishtar Terra, centered at 70°, 340°E. Its highest part consists of the Maxwell Montes, which

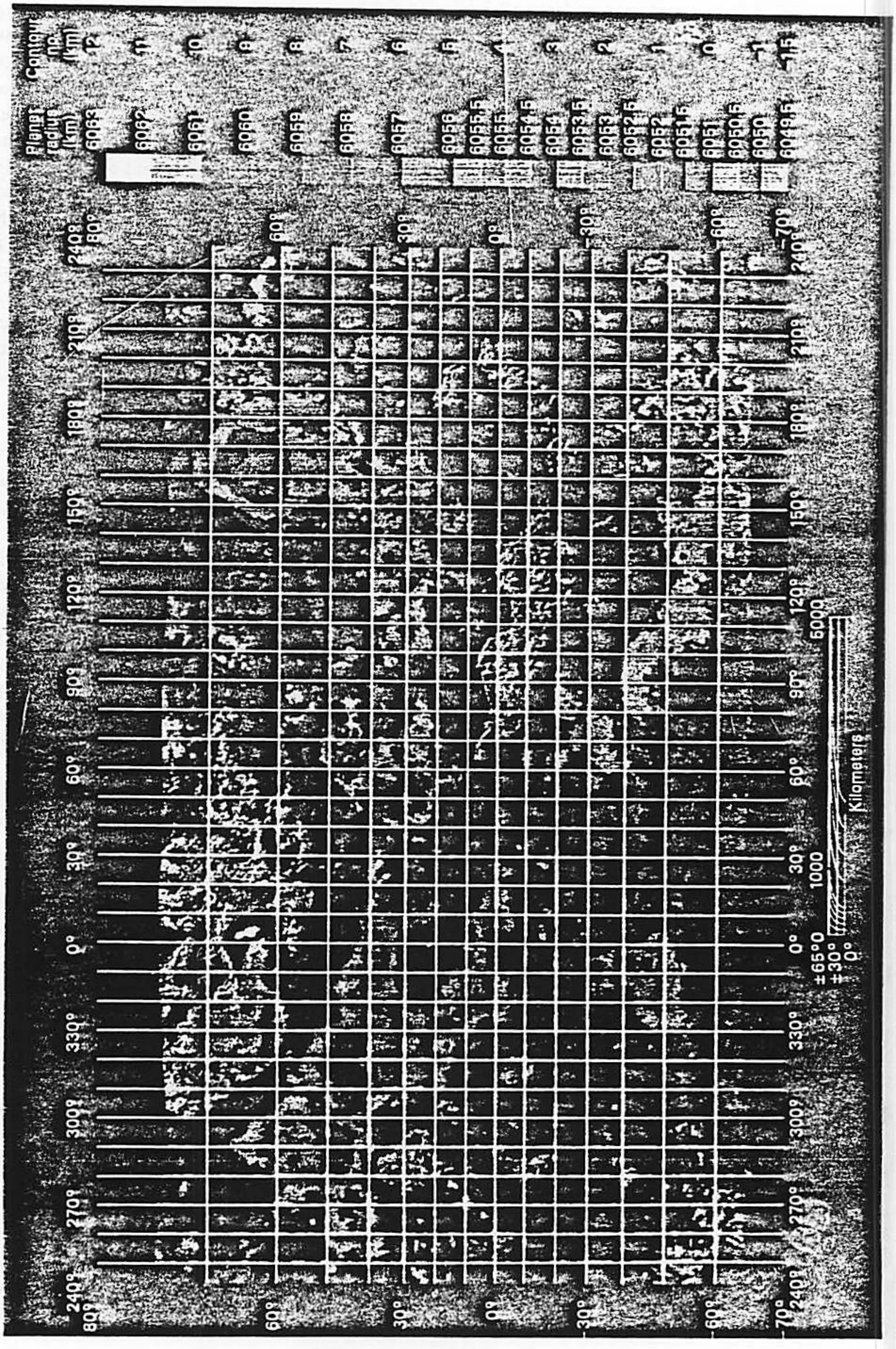


Figure 4.6. A color-coded contour map of elevations of the Venusian surface. High areas are shown in reds and yellows, low areas in blues.

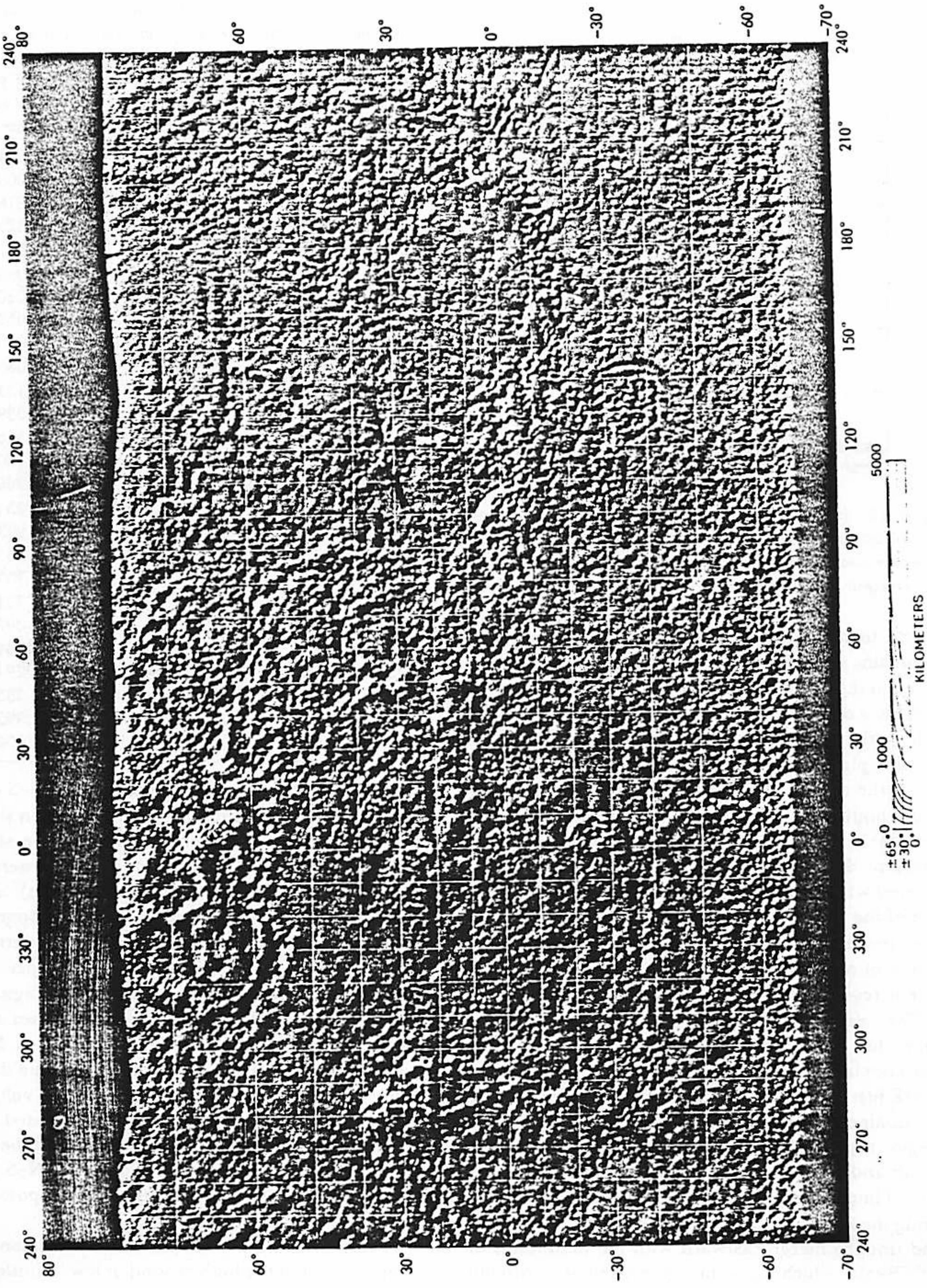


Figure 4.7. The same data as in figure 4.6 but depicted as shaded relief with illumination from the northeast.

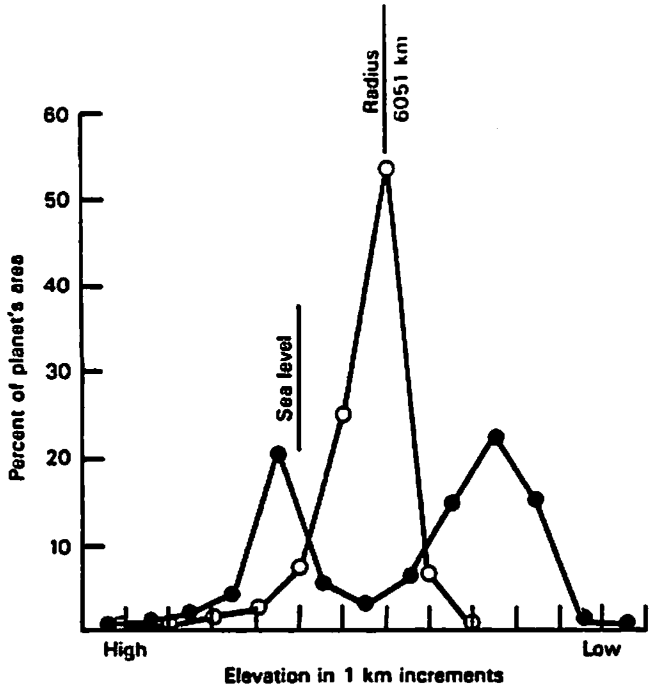


Figure 4.8. Hypsometric curves for Earth and Venus. The curve for the Earth is bimodal with peaks near sea level and at the depths of mature ocean floors. The Venus data are unimodal as skewed toward higher elevations.

Table 4.6. Percentage of Mapped Area of Venus Within Topographic Intervals (from Masursky et al., 1980)

Planetary radii class interval (km)	Percent area in interval	Cumulative percent area in and above interval
6061.9-6061.5	00.004	00.004
6061.5-6061.0	00.005	00.009
6061.0-6060.5	00.007	00.016
6060.5-6060.0	00.006	00.022
6060.0-6059.5	00.010	00.032
6059.5-6059.0	00.008	00.040
6059.0-6058.5	00.010	00.050
6058.5-6058.0	00.007	00.057
6058.0-6057.5	00.013	00.070
6057.5-6057.0	00.016	00.086
6057.0-6056.5	00.047	00.133
6056.5-6056.0	00.106	00.239
6056.0-6055.5	00.327	00.566
6055.5-6055.0	00.594	01.160
6055.0-6054.5	01.106	02.266
6054.5-6054.0	00.965	03.231
6054.0-6053.5	01.692	04.923
6053.5-6053.0	02.440	07.363
6053.0-6052.5	05.376	12.739
6052.5-6052.0	08.982	21.721
6052.0-6051.5	17.376	39.097
6051.5-6051.0	33.664	72.761
6051.0-6050.5	20.533	93.294
6050.5-6050.0	06.091	99.385
6050.0-6049.5	00.607	99.992
6049.5-6049.0	00.008	100.000

extend to heights of 11 km above the datum. The mountains are radar bright and have meter-decameter slopes in the range of 4°-10°; a dark circular feature, possibly a caldera, is situated 2-4 km below the summit. To the west of Maxwell Montes is a dark, roughly circular plain, Lakshmi Planum, standing 4-5 km above the datum. The plain appears to be cratered and is bounded on the south and west by an abrupt, arcuate scarp. Masursky et al. (1980) interpret Lakshmi Planum as an ancient uplifted surface covered with a thin veneer of younger flows. To the east of the Maxwell Montes is some topographically complex terrain, 2-3 km above the datum and consisting of numerous closed depressions with no consistent regional trends.

The second and largest major upland area is Aphrodite Terra, which is about the size of Africa and elongates in a roughly E-W direction between 70° and 210°E just south of the equator. Western and central mountainous regions named Ovda Regio and Thetis Regio, respectively, are separated by a low saddle. South and east of the central mountains are a complex of linear troughs and ridges, the most prominent being the semi-circular Artemis Chasma. The ridges and troughs merge eastward with the mountains of Atla Regio, which mark the eastern end of Aphrodite Terra. Here the terrain has a dominantly N-S trend, in contrast to the mainly E-W lineaments in the rest of Aphrodite. The eastern mountains are slightly higher (5.7 km above datum) than the western and central mountains (5.5 km above datum). All of Aphrodite exhibits complex patterns of topography, RMS slopes, and backscatter efficiency—patterns that suggest an abundance of steep blocky slopes.

The third major upland region is Beta Regio, centered at 30°N, 285°E. It is composed of two shield-shaped mountains, Theia Mons and Rhea Mons, which reach elevations of 4-5 km above the datum. Both have been interpreted as possible volcanoes (Malin and Saunders, 1977). Smaller elevated areas, Phoebe Regio and Themis Regio, occur to the south of Beta and appear connected to it by a N-S linear disruption zone upon which the two supposed volcanoes lie (McGill et al., 1981).

A vast series of canyons and linear disruption zones connect most of the high ground at low latitudes (fig.

4.9). The largest troughs, the Diana and Dali Chasmata, are 3000–3500 km long and 75–100 km wide (Schaber, 1982). They are part of a broad (1000–1500 km) linear zone of disruption that extends from the west end of Aphrodite through Atla Regio to Beta Regio over 20 000 km to the east. The fracture system thus extends almost three-quarters of the way around the planet in a roughly E–W direction. A second disruption zone, with a prominent rift, Devana Chasma, extends south from Beta Regio at the west end of Aphrodite. The most prominent of all the rifts, the semi-circular Artemis Chasma, does not lie on any of the disruption zones just outlined, but rather lies just to the south of the main E–W system.

SURFACE ROUGHNESS

To first order, roughness correlates strikingly with elevation; elevated regions are much rougher than low regions. This is true, both of the Pioneer-Venus data measured at a wavelength of 17 cm and Earth-based data. Venus is smoother at the centimeter- to meter-scale than the Moon or the rough regions of Mars. Most of the planet has meter-scale RMS slopes between 1° and 3° (Masursky et al., 1980). According to the Apollo bistatic radar experiment, the Moon has slopes of 3° to 4° (Moore et al., 1980; Tyler, 1979). Meter-scale roughness on the Moon is dominated by impact debris. The youngest craters, mostly of Copernican and Eratosthenian age, appear as bright rings in the radar images, but older craters do not, probably because continual break-up of blocks in the floor and rim ejecta reduces the surface roughness. Mars has regions that are much rougher than is typical for Venus, and regions that are perhaps smoother. The volcanic plains in the Tharsis region of Mars, for example, have 10° to 15° slopes, whereas the plains of Elysium and Syrtis Major have slopes of 1° to 2° at 12 cm wavelength (Downs et al., 1973, 1975; Simpson et al., 1978). An aging process that reduces roughness appears to occur both on Mars and the Moon. Older volcanic plains may be smoothed by sand blasting, the removal of weathered products by the wind, and

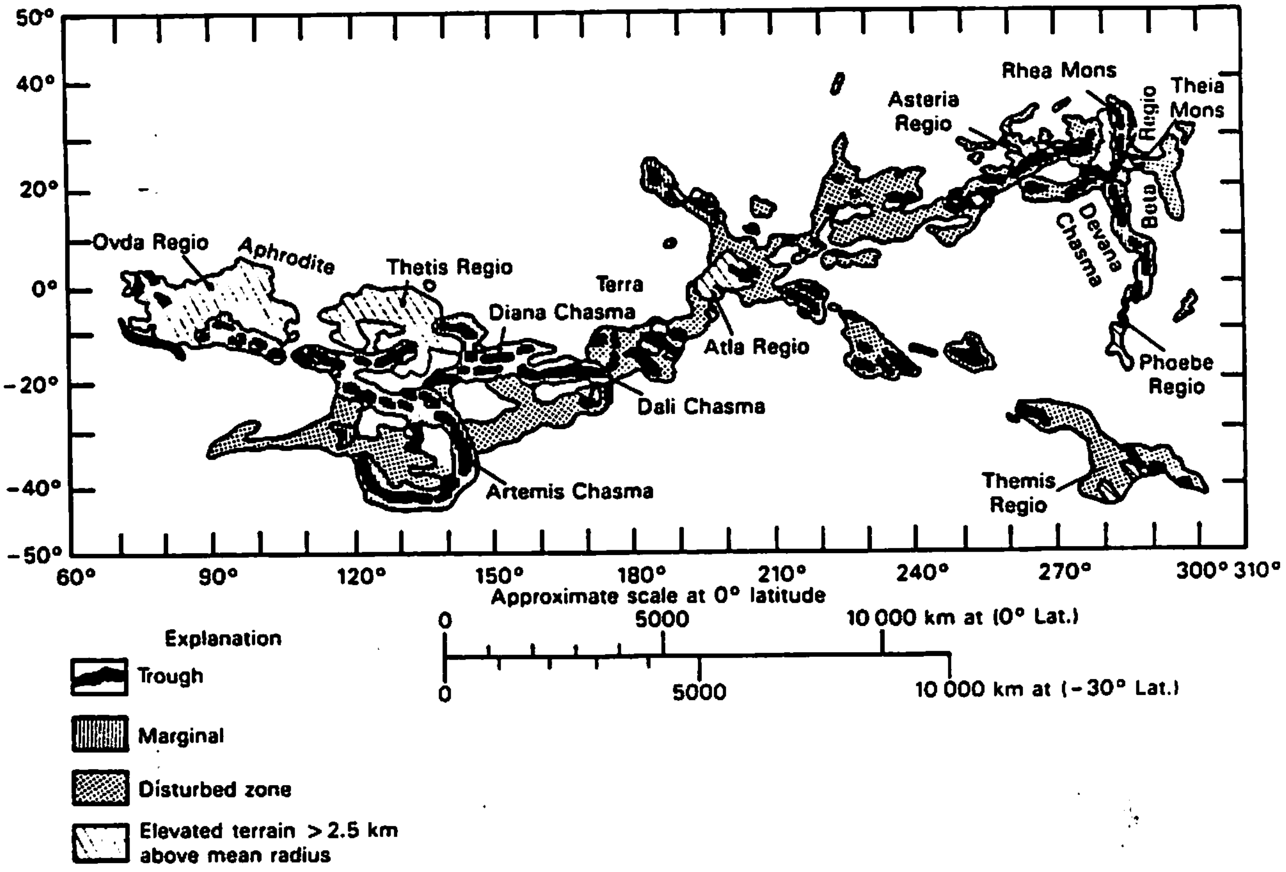


Figure 4.9. Structural sketch map showing interconnected troughs, marginal ridges, disturbed zones, and elevated terrain between Beta Regio and Themis Regio to the east and Aphrodite Terra to the west. The interconnected system stretches in a roughly E–W direction around roughly two-thirds of the planet near the equator (from Schaber, 1982).

through mantling by the products of eolian and chemical processes.

Venus has a wide range of roughness. The highest regions are probably rough, primary volcanic surfaces, but most of the surface has probably been smoothed by weathering and sedimentary mantling as on Mars. Numerous bright circular features have been interpreted as impact craters. If their present size frequency distribution (Campbell and Burns, 1980) represents an equilibrium between crater production and destruction by erosion and sedimentation, then crater obliteration rates of 10^{-5} to 10^{-7} cm/yr are implied, which are more comparable to Mars rates than terrestrial rates.

BRIGHT RADAR RINGS

Bright rings are prominent on many radar images of Venus (Rumsey et al., 1974; Campbell et al., 1979; Goldstein et al., 1976, 1978). They are common on the rolling plains unit but generally absent in the elevated regions. If they are impact craters, then they provide a means of assessing the age of the surface, and the rates of erosion and viscous relaxation of the lithosphere, so that their origin is of crucial importance for understanding the evolution of the surface.

A number of roughly circular bright rings can be seen in Goldstone images of the equatorial region (Rumsey et al., 1974; Goldstein et al., 1976, 1978). The incidence angles are generally less than 6°, and at these angles the backscatter is strongly slope dependent. The ratio of rim width to crater diameter is typical of that for impact craters on other planets, and generally different from that for volcanic features. The bright rings thus appear to be impact craters with raised rims (Saunders and Malin, 1977), but an impact origin is far from proven (McGill et al., 1982). Campbell and Burns (1980) identified 33 bright rings on Arecibo images which generally have higher incidence angles than the Goldstone images, and tend to emphasize variations in backscatter rather than slopes. Several large (>300 km) radar-dark circular features have also been interpreted as impact basins (Masursky et al., 1980).

Thompson et al. (1980, 1981) and Cutts et al. (1981) investigated radar images of lunar craters, and these may provide a basis for interpreting the Venus images. The radar signature of lunar craters is initially a continuous radar-bright region up to 20 times the diameter of the associated crater. Cutts et al. reported that most lunar crater signatures shrink with age rather rapidly until only the crater rim is bright and then this gradually fades. The floor signature has a much greater lifetime than that of the ejecta zone. Radar bright rings are rare on the Moon, and Cutts et al. suggested that they occur only where a crater has been embayed by mare material. This observation might lead to the conclusion that bright-ring features on Venus are volcanically filled craters. However, Cutts et al. suggested a similar result would follow from mantling of rough, radar-bright areas with fine impact debris produced during basin formation.

Saunders et al. (1982) examined the bright ring radar features in 3.8 cm lunar images (Zisk et al., 1974; Thompson, 1974). About 30 radar-bright rings were identified that met the selection criteria that they be complete rings encompassing darker floors and not be at extreme ranges of radar incidence angles. The features found in this study do not occur preferentially on maria, plains, or terra. Somewhat surprisingly, only three of the features have mare-flooded floors. With one exception, an Imbrian crater, all the rings are associated with Copernican and Eratosthenian craters. The diameter distribution of these craters falls along a production curve. Comparison with topography suggests that the inner diameter corresponds to the crater floor diameter, as suggested by Cutts et al. (1981). The outer diameter corresponds to the diameter of the raised crater lip. The implication is that, on the Moon, relatively level areas such as crater floors and the surrounding ejecta deposits quickly become dark to the 3.8 cm radar as a result of impact comminution. Slope appears to be the major contributor to radar brightness.

Saunders et al. (1982) proposed that the evolution of the radar signature of lunar craters is largely slope controlled. Impact craters are initially an extensive radar-bright region. The ejecta and floor become relatively smooth in a short time, but the rim lip and crater walls remain rough until their slopes have been so reduced by erosion that downslope movement of erosion products no longer continuously exposes blocky material.

The processes of crater erosion on Venus are not known. However, it might be expected that, as on the Moon, only a fraction of the total number of circular features is detected, and that their radar signature changes with time as the craters are eroded. The strong correlation between elevation and reflectivity apparent in the Pioneer-Venus data (Pettengill et al.,

1980) suggests that slope is a major factor in controlling backscatter. The radar-bright rings are therefore likely to be caused by steep slopes surrounding a relatively flat circular feature, and impact craters are the most plausible candidates.

If the radar-bright rings are impact craters, then estimates can be made of resurfacing rates on Venus and the ages of different Venusian surfaces. The flux of objects in the size range that would have produced the observed craters has been essentially identical for Earth and Venus, and affected little by the atmosphere (Tauber and Kirk, 1976). Assuming lunar fluxes, Campbell and Burns (1980) estimated from the density of radar-bright rings that the rolling plains have an impact age of 600–800 million years. However, Saunders and Malin (1976) found that the densities varied greatly, and while the plains overall have a crater density that is less by a factor of ten than that of the lunar highlands, densities in small areas may approach those of the lunar highlands. If the rings are of impact origin, the Campbell and Burns "age" is probably a minimum age, being more indicative of the lifetime of the radar-bright signatures than of the surface itself. Presumably older craters that have lost their radar signature could still be present, but remain undetected.

The impact crater hypothesis therefore leads to the supposition that the rolling plains are older than the highland regions and that some of the plains may date back to the decline in impact rates 3.8 billion years ago. Clearly, if a significant fraction of the rings are volcanic, then none of these conclusions are valid. The relative ages of highlands and lowlands remain unknown as well as the absolute ages.

The case for an impact origin for the large (>200 km) dark, quasi-circular features is far less convincing than that for the bright rings. Solomon et al. (1982) recognized two types: (1) circular areas of low radar backscatter and little topographic relief, and (2) large, roughly circular depressions, such as Atalanta Planitia, which is 4000 km across and over 2 km deep. Masursky et al. (1980) likened these dark features to lunar impact basins, and from their size frequently concluded that they were at least 3.8 billion years old. However, Solomon et al., demonstrated that if surface temperatures throughout much of Venus' history were comparable to those that presently prevail, then viscous relaxation would eliminate almost all topographic relief in basins larger than a few hundred kilometers across within 3 billion years. They conclude, therefore, that the large topographic basins, such as Atalanta Planitia, are geologically young and have formed by some process other than impact. One possibility is that they are analagous to terrestrial platform basins, produced by lithospheric extension and thermal subsidence.

GRAVITY

The gravity field of Venus shows strong correlation with topography, large gravity highs coinciding with large topographic highs (Ananda et al., 1980; Sjogren et al., 1980; Phillips et al., 1981; Reasenberg et al., 1981; Esposito et al., 1982). The situation thus differs from that on Earth where gravity is only poorly correlated with the regional-scale topography. The Venusian anomalies are generally smaller than those on the Moon and Mars where 100 mgal anomalies are common. The largest, that associated with Beta Regio, is 135 mgal at a reference altitude of 200 km (Esposito et al., 1982). For comparison, a 500–600 mgal anomaly, referenced to the global mean, occurs over Olympus Mons on Mars.

Simulation of the anomalies that would be produced by the observed topography implies that significant compensation of the topography has taken place. Phillips et al. (1981) attempted to place limits on lithosphere thickness and rigidity by comparing topography and gravity at different wavelengths. They showed that depths of compensation of 100 km or more are required if the topography is supported by density variations alone. If compensation is at shallower depths, then a combination of density differences and flexural rigidity of the lithosphere is required to support the topography. They concluded that, because of creep within the lithosphere, the topography can be supported passively only if it is very young, on the order of 10^7 years, or if the heat flow is considerably less than that on Earth, which they consider unlikely. Another possibility is that the lithosphere is supported dynamically from below, such as by convection beneath the lithosphere (McGill et al., 1981). These conclusions are reinforced by the more recent work of Esposito et al. (1982) who demonstrated that compensation depths of 300–400 km are required to explain the large anomaly over Beta Regio. Weertman (1979), however, questioned the use of terrestrial creep rates in estimating the strength of the Venusian crust. He suggested that, despite the higher temperatures, creep rates in the Venusian lithosphere may be substantially less than in Earth's because of the lower water content.

PLATE TECTONICS ON VENUS

Much of the speculation about the evolution of the Venusian surface has focused on whether plate tectonics have ever occurred on the surface (McGill, 1979; McGill et al., 1982; Phillips et al., 1981; Solomon and Head, 1982). The speculation is triggered by several considerations. First, the similarity in size of Venus and Earth should lead to similar surface heat flows if compositions are comparable. Second, the vast rift systems on Venus could be analogous to the divergent rift zones of Earth. Third, strong correlation of gravity and topography, coupled with some estimates of young crater ages, suggests the topography might be young. Fourth, early chemical analysis of the surface materials, from Venera 8, pointed to the possibility of granitic materials.

Head et al. (1981) and Arvidson and Davies (1981) assessed the likelihood of detecting present-day plate tectonics on Venus solely from the topography. They reconstructed images of Earth using digital terrain data with elements 100 km on a side in order to simulate the Pioneer-Venus resolution. In their simulations many of the diagnostic features of plate tectonics, such as mountain chains and island arcs, disappear. The continents look level and featureless but ocean floor ridges and some trenches are resolved. Arvidson and Davies asserted that a ridge system comparable to the ocean ridges of Earth should be discernible in the existing data even after correcting for the temperature of the Venusian surface and the lack of loading by ocean water. On the other hand, Head et al. (1981) and Solomon and Head (1982) believed several factors in addition to surface temperature and water loading could affect the configuration of the ridges; maintaining that a ridge system could be easily masked on the rolling plains.

Some care must be taken in the application of these analog studies. It is important to understand that the altimeter used by Pioneer-Venus has a horizontal resolution that is determined by the effective length of the pulse. In general, the range that is detected will be that of the nearest surface in the beam. Elevated features tend to be smeared out and depressions such as trenches appear narrower because of the geometric effects of the altimeter system. Experience with radar suggest that the smallest features that can be resolved have dimensions of many radar resolution elements, although linear features may require only three or four depending on how irregular they are.

Other reasons for exercising care in the application of Earth analogs to the interpretation of Venus data is that we do not fully understand the effects of the hydrosphere, atmosphere, and biosphere on Earth's topography. The major fold mountain systems would not appear as they do without the influence of any one of these factors. Of the highest peaks on Earth. the top one hundred or more consist primarily of marine sedimentary rocks.

Phillips et al. (1981) cited differences in profile between the oceanic ridges of Earth and possible counterparts in the Venusian highlands as evidence of their being of different origins. The flanks of the oceanic ridges are concave upward, reflecting contraction by cooling, whereas the elevated areas of Venus, such as Aphrodite and Beta Regio, are convex upward. The ridges of Venus are also convex upward. Kaula (1981) added that the ridges on Venus do not have a narrow distribution about a mode in crest height as do terrestrial oceanic ridges. But both these arguments are relevant only if the Venusian highlands are believed to be the analogs of the terrestrial oceanic ridges, which is doubtful (Solomon and Head, 1982). Furthermore, Brass and Harrison (1982) asserted that the general form and detectability of tectonically created features depend on the balance between erosion rates and rates of plate motion. If erosion rates on Venus are relatively high, linear trenches and rift valleys will fill and not be recognizable.

Theoretical arguments have also been raised against plate tectonics. Anderson (1981) argued that the high surface temperatures will prevent plate tectonics, because the lithosphere can never cool to a low enough temperature to become negatively buoyant. Thus, the currently favored mechanism for driving plate tectonics, drag of negatively buoyant lithosphere into subduction zones, cannot work. Weertman (1979) showed that plate tectonics could occur on Venus only if the rocks were extremely dry, for only then would the lithosphere have the required rigidity at the relatively high Venusian lithosphere temperatures. Finally, Kaula (1981) showed, from the number of possible spreading centers and the dimensions of the observed ridges, that if the ridges formed by spreading, which he doubts, the amounts of internal heat lost through plate tectonics can be no more than 15 percent of the total lost by the planet. On Earth, 70 percent of the internal heat is lost through plate motion.

While acknowledging that plate tectonics are unlikely at present, Phillips et al. (1981) speculated that plate tectonics may have occurred in the past. As we saw above, counts of large craters suggest the rolling plains are ancient, whereas the highlands are relatively young. Phillips et al. suggested that equatorial highlands may have been former zones of divergence, and that at some relatively ancient time in Venus' history, plate tectonics ceased and basalt crusts tended to accumulate over the former divergent zones to form the relatively young equatorial highlands. Possible causes for cessation of plate motion are crustal thickening, loss of water from the interior, destabilization of water at the surface, and rise in surface temperatures as the extreme greenhouse developed.

Kaula (1981) compared the evolution of Venus with that of other terrestrial planets. He noted that on the basis of size and composition Venus might be expected to have evolved along a path similar to Earth. Yet, Venus is more Mars-like than Earth-like. He suggested that the contrast between Earth and Venus is largely due to the thicker atmosphere and consequent higher surface temperatures. These result in a greater depth to the basalt-eclogite transition, which in turn inhibits recycling of the crust. Venus has therefore a thicker crust than Earth, which coupled with the higher surface temperatures, increases the buoyancy of the lithosphere and prevents subduction.

ACKNOWLEDGMENTS

G. McGill of the University of Massachusetts and S. Solomon of the Massachusetts Institute of Technology made several suggestions for improving the chapter. Part of the work was performed by the Jet Propulsion Laboratory, California Institute of Technology and sponsored by NASA.

The World into Which Darwin Led Us

IT has often been said that Darwin changed the world. It has less often been made clear just what the change has been. Darwin did not—to his credit he did not—make any of the discoveries that have led to our present overwhelming physical peril. Most, although not quite all, of our technology would be the same if Darwin's work had not been done, by him or anyone else. Doubtless we would in that case still have our same traffic jams, horror movies, bubble gum, and other evidences of high civilization. The paraphernalia of civilization are, however, superficial. The influence of Darwin, or more broadly of the concept of evolution, has had effects more truly profound. It has literally led us into a different world.

How can that be? If evolution is true, it was as true before Darwin as it is today. The physical universe has not changed. But our human universes, the ones in which we really have our beings, depend at least as much on our inner perceptions as on the external, physical facts. That can be made evident by an elementary example. Suppose a stone is seen by a small boy, an artist, and a petrologist. The small boy may perceive it as something to throw, the artist as something to carve into sculpture, the petrologist as a mixture of minerals formed under certain conditions. The stone is three quite different things to the three people, and yet they are seeing exactly the same thing. The stone has identical properties whatever anyone thinks about it.

In that trivial example all three conceptions of the stone, although profoundly different, are equally true. The stone can indeed be thrown, be sculptured, or be analyzed petrologically by procedures suitable to each of the three perceptions. But there are differing perceptions of objects and of our whole world that are not equally true in the same sense, which is the scientific sense of material testability. Perceptions that are not materially testable or that have been contradicted by adequate tests are not rationally valid. As they petrify into tradition and dogma they become superstitions. Perception of the truth of evolution was an enormous stride from superstition to a rational universe.

Years ago I lived for a time with a group of uncivilized Indians in South America. Their world is very different from ours: in space, a saucer a few miles across; in time, from a few years to a few generations back into a misty past; in essence, lawless, unpredictable, and haunted. Anything might happen. The Kamarakoto Indians quite believe that animals become men and men become stones; for them there is neither limitation nor reason in the flux of nature. There is also a brooding evil in their world, a sense of wrongness and fatality that they call *kanaima* and see manifested in every unusual event and object.

That level of invalid perceptions might be called the lower superstition. It is nevertheless superior in some respects to the higher superstitions celebrated weekly in every hamlet of the United States. The legendary metamorphoses of my Indian friends are grossly naive, but they do postulate a kinship through all of nature. Above all, they are not guilty of teleology. It would never occur to the Indians that the universe, so largely hostile, might have been created for their benefit.

It is quite wrong to think that uncivilized Indians are, by that token, primitive. Nevertheless, I suppose that the conceptual world of the Kamarakotos is more or less similar to that of ancient, truly primitive men. Indeed, even at the dawn of written history in the

cradles of civilization, the accepted world pictures do not seem very different from that of those Indians.

The world in which modern, civilized men live has changed profoundly with increasingly rational, which is to say eventually scientific, consideration of the universe. The essential changes came first of all from the physical sciences and their forerunners. In space, the small saucer of the savage became a large disk, a globe, a planet in a solar system, which became one of many in our galaxy, which in turn became only one nebula in a cosmos containing uncounted billions of them. The astronomers have finally located us on an insignificant mote in an incomprehensible vastness—surely a world awesomely different from that in which our ancestors lived not many generations ago.

As astronomy made the universe immense, physics itself and related physical sciences made it lawful. Physical effects have physical causes, and the relationship is such that when causes are adequately known effects can be reliably predicted. We no longer live in a capricious world. We may expect the universe to deal consistently, even if not fairly, with us. If the unusual happens, we need no longer blame *kanaima* (or a whimsical god or devil) but may look confidently for an unusual or hitherto unknown physical cause. That is, perhaps, an act of faith, but it is not superstition. Unlike recourse to the supernatural, it is validated by thousands of successful searches for verifiable causes. This view depersonalizes the universe and makes it more austere, but it also makes it dependable.

To those discoveries and principles, which so greatly modified concepts of the cosmos, geology added two more of fundamental, world-changing importance: vast extension of the universe in time, and the idea of constantly lawful progression in time. Estimates of geological time have varied greatly, but even in the eighteenth century it became clear to a few that the age of the earth must be in millions of years rather than the thousands then popularly accepted from biblical exegesis. Now some geological dates are firmly established, within narrowing limits, and no competent geologist considers the earth less than 3 billion years old. (Upper estimates for the solar system range from 5 to 10 billion.) That is still only a moment in eternity, but it characterizes a world very different from one conceived as less than 6000 years old.

With dawning realization that the earth is really extremely old, in human terms of age, came the knowledge that it has changed progressively and radically but usually gradually and always in an orderly, a natural, way. The fact of change had not earlier been denied in Western science or theology—after all, the Noachian Deluge was considered a radical change. But the Deluge was believed to have had supernatural causes or concomitants that were not operative throughout the earth's history. The doctrine of geological uniformitarianism, finally established early in the nineteenth century, widened the recognized reign of natural law. The earth has changed throughout its history under the action of material forces, only, and of the *same* forces as those now visible to us and still acting on it.

The steps that I have so briefly traced reduced the sway of superstition in the conceptual world of human lives. The change was slow, it was unsteady, and it was not accepted by everyone. Even now there are nominally civilized people whose world was created in 4004 B.C. Nevertheless, by early Victorian times the physical world of a literate consensus was geologically ancient and materially lawful in its history and its current operations. Not so, however, the world of life; here the higher (or at least later) superstition was still almost unshaken. Pendulums might swing with mathematical regularity and mountains might rise and fall through millennia, but living things belonged outside the realm of material principles and secular history. If life obeyed any laws, they were supernal and not bound to the physics of inert substance. Beyond its original, divine creation, life's history was trivial. Its kinds were each as created in the beginning, changeless except for minor and obvious variations.

Perhaps the most crucial element in man's world is his conception of himself. It is here that the higher superstition offered little real advance over the lower. According to the higher superstition, man is something quite distinct from nature. He stands apart from all other creatures; his kinship is supernatural, not natural. It may, at first sight, seem anomalous that those scientists who held this view did classify man as an animal. Linnaeus, an orthodox upholder of the higher superstition, even classified *Homo* with the apes and monkeys. No blood relationship was implied. The system of nature was the pattern of creation, and it included all created things, without any mutual affinities beyond the separate placing of each in one divine plan.

Another subtler and even more deeply warping concept of the higher superstition was that the world was created for man. Other organisms had no separate purpose in the scheme of creation. Whether noxious or useful, they were to be seriously considered only in their relationship to the supreme creation, the image of God. It required considerable ingenuity to determine why a louse, for example, was created to be a companion for man, but the ingenuity was not lacking. A world made for man is no longer the inherently hostile and evil world of *kanaima,* but that again is offset in some versions of the higher superstition by the belief that man himself is inherently evil or, at least, sinful.

Those elements of the higher superstition dominated European thought before publication of *The Origin of Species,* but various studies have exhaustively demonstrated that evolutionary ideas existed and were slowly spreading among a minority of *cognoscenti* long before Darwin. Some believed that a species, although divinely and separately created, might change, and in particular might degenerate from its form in the original plan of creation. That is not a truly evolutionary view, since it does not really involve the origin of one species from another, but it does deserve to be called proevolutionary in that it recognized the fact that each separate species may change. In the eighteenth century Buffon went that far, but hardly further, in spite of some apologists who now hail him as an evolutionist.

Some eighteenth-century worthies—among them Linnaeus in his later years—did go one step further. They conceived that each of the separately created "kinds" of Genesis might later have become considerably diversified, so that the unit of separate creation might be what we now call a genus or even a family or higher group, and the species or subgroups might have arisen, or indeed evolved, since the creation. Just as the many breeds of domesticated dogs are all dogs and of common origin, so the wolves, coyotes, foxes, jackals, and other wild species might all descend from a single creation of the dog-kind. That would still admit no relationship between the dog-kind and the now likewise diversified but singly and separately created cat-kind, for example. (It is an intellectual curiosity that precisely that variation of creationist superstition has recently been seriously revived by an American who had been exposed, at least, to excellent training in zoology.)

By the end of the eighteenth century there were a few true and thoroughgoing evolutionists—Charles Darwin's grandfather Erasmus was one, as has so often been pointed out. Their number increased during the first half of the nineteenth century. Some of them even had glimmerings of Darwin's great discovery, natural selection, although (contrary to some recent historians whose aim seems to be to denigrate Darwin) none of them elucidated that principle clearly and fully.

Practically all of the ideas in *The Origin of Species* had been dimly glimpsed, at least, by someone or other before 1859. The only surprising thing about that is that so many authors have thought it worthy of special emphasis. Organization, understanding, and conviction are the main contributions of theorists like Darwin, and obviously none ever succeeded until there already existed something to organize and to understand. It is, however, less obvious why Darwin was the first evolutionist ever to carry conviction to a majority of his fellow scientists. The whole answer is more complex, but its essentials are evident in a statement later made by Thomas Henry Huxley to explain why he was an antievolutionist until he read *The Origin of Species:*

> I took my stand upon two grounds: firstly that up to that time, the evidence in favor of transmutation [evolution] was wholly insufficient; and, secondly, that no suggestion respecting the causes of the transmutation assumed, which had been made, was in any way adequate to explain the phenomena. Looking back at the state of knowledge at that time, I really do not see that any other conclusion was justifiable.

The reason why *The Origin of Species* carried conviction was that it did supply sufficient evidence of evolution and also provided

an explanation of the phenomena of evolution. That twofold nature of Darwin's accomplishment has certainly been pointed out often enough, but the statement has also been criticized, and perhaps some small notice should here be given to some of the criticisms. It has, for one thing, been maintained that previous evidence *was* sufficient. It had persuaded Erasmus Darwin, Lamarck, Chambers (author of the anonymous *Vestiges of Creation*), and others, so (some critics say) it should have persuaded anyone without Charles Darwin's needing to recompile it. That conclusion is simply ridiculous. What anyone thinks *should* have happened has nothing to do with the question of historical fact. Previous evidence *did not* convince a majority of interested scientists; therefore it was insufficient for that purpose. Darwin's evidence *did* in fact convince them; therefore it was sufficient. (It may of course be recognized, as Darwin himself implied, that the way had been prepared by a changing climate of opinion and that even his evidence might have been insufficient if adduced at an earlier date.)

It has further been suggested that evolution could have been, perhaps should have been, established as a fact without requiring an explanation, and also that Darwin's explanation was not really adequate. The first proposition is debatable, certainly, and examples can be produced to support both sides. The inheritance of acquired characters was accepted by practically everyone, down to and including Darwin, even though no one had adequately explained it. Darwin himself did not like to deal with unexplained facts, and he did belatedly attempt to explain the inheritance of acquired characters. Since in this case the "facts" were not true, that particular Darwinian theory is now charitably forgotten. (Fortunately it was not really essential to his broader theory explanatory of evolution as a whole.) In any case, belief in the inheritance of acquired characters did not depend on any explanation of the supposed phenomena. (Is there perhaps a warning in the fact that the unexplained phenomena did not in truth occur?) On the other side of the argument is the modern example of extrasensory perception. A great mass of facts is claimed to demonstrate the reality of that unexplained phenomenon, and yet it is not generally accepted. It seems quite clear that it will not carry conviction unless some credible explanation is produced.

It does seem to me highly improbable that the fact of evolution would have been accepted so widely and quickly if it had been unaccompanied by an explanatory theory. Again, to question whether it *should* have been would be childish arguing with history.

The adequacy of Darwin's original explanation of evolution is also decidedly subject to debate. It was certainly an incomplete explanation, as Darwin was keenly aware. We now have much more extensive explanations, built in large part on Darwin's. Parts of Darwin's complex theory are also now known beyond serious doubt to have been wrong, although the more essential parts, those most stressed by Darwin, have been largely substantiated. Darwin's theory was adequate at the time in the sense of being convincing. The conviction did not depend entirely on the truth or falsity of different parts of his explanation, which was not wholly accepted by students who nevertheless were immediately persuaded of the truth of evolution. The essential point was demonstration that material causes of evolution are possible and can be investigated scientifically.

The fact—not theory—that evolution has occurred and the Darwinian theory as to how it has occurred have become so confused in popular opinion that the distinction must be stressed. The distinction is also particularly important for the present subject, because the effects on the world in which we live have been distinct. The greatest impact no doubt has come from the fact of evolution. It must color the whole of our attitude toward life and toward ourselves, and hence our whole perceptual world. That is, however, a single step, essentially taken a hundred years ago and now a matter of simple rational acceptance or superstitious rejection. How evolution occurs is much more intricate, still incompletely known, debated in detail, and the subject of most active investigation at present. Decision here has decidedly practical aspects and also affects our worlds even more intimately, and in even more ways, than the fact of evolution. The two will be separately considered.

The import of the fact of evolution depends on how far evolution extends, and here there are two crucial points: does it extend from the inorganic into the organic, and does it extend from the lower animals to man? In *The Origin of Species* Darwin implies that life did not arise naturally from nonliving matter, for in the very last sentence he wrote, ". . . life . . . having been originally breathed by the Creator into a few forms or into one. . . ." (The words *by the Creator* were inserted in the second edition and are one of many gradual concessions made to critics of that book.) Later, however, Darwin conjectured (he did not consider this scientific) that life will be found to be a "consequence of some general law"—that is, to be a result of natural processes rather than divine intervention. He referred to this at least three times in letters unpublished until after his death, the one from which I have quoted being the last letter he ever wrote (28 March 1882 to G. C. Wallich; Darwin died three weeks later).

Until comparatively recently, many—probably most—biologists agreed with Darwin that the problem of the origin of life was not yet amenable to scientific study. Now, however, almost all biologists agree that the problem can be attacked scientifically. The consensus is that life did arise naturally from the nonliving and that even the first living things were not specially created. The conclusion has, indeed, really become inescapable, for the first steps in that process have already been repeated in several laboratories. There is concerted study from geochemical, biochemical, and microbiological approaches. At a meeting in Chicago in 1959, a highly distinguished international panel of experts was polled. All considered the experimental production of life in the laboratory imminent, and one maintained that this had already been done—his opinion was not based on a disagreement about the facts, but depended on the definition of just where, in a continuous sequence, life can be said to begin.

At the other end of the story, it was evident to evolutionists from the start that man cannot be an exception. In *The Origin of Species* Darwin deliberately avoided the issue, saying only in closing, "Light will be thrown on the origin of man and his history." Yet his adherents made no secret of the matter and at once embroiled Darwin, with themselves, in arguments about man's origin from monkeys. Twelve years later (in 1871) Darwin published *The*

Descent of Man, which makes it clear that he was indeed of that opinion. No evolutionist has since seriously questioned that man did originate by evolution. Some, notably the Wallace who shared with Darwin the discovery of natural selection, have maintained that special principles, not elsewhere operative, were involved in human origins, but that is decidedly a minority opinion about the causes or explanations, not the fact, of evolution.

It is of course also true that the precise ancestry of man is not identified in full detail and so is subject to some disagreement. That is a minor matter of no real importance for man's image of himself. No one doubts that man is a member of the order Primates along with the lemurs, tarsiers, monkeys, and apes. Few doubt that his closest living relatives are the apes. On this subject, by the way, there has been too much pussyfooting. Apologists emphasize that man cannot be a descendant of any living ape—a statement that is obvious to the verge of imbecility—and go on to state or imply that man is not really descended from an ape or monkey at all, but from an earlier common ancestor. In fact, that common ancestor would certainly be called an ape or monkey in popular speech by anyone who saw it. Since the terms *ape* and *monkey* are defined by popular usage, man's ancestors *were* apes or monkeys (or successively both). It is pusillanimous if not dishonest for an informed investigator to say otherwise.

Evolution is, then, a completely general principle of life. (I refer here, and throughout, to organic evolution. Inorganic evolution, as of the stars or the elements, is quite different in process and principle, a part of the same grand history of the universe but not an extension of evolution as here understood.) Evolution is a fully natural process, inherent in the physical properties of the universe, by which life arose in the first place and by which all living things, past or present, have since developed, divergently and progressively.

This world into which Darwin led us is certainly very different from the world of the higher superstition. In the world of Darwin man has no special status other than his definition as a distinct species of animal. He is in the fullest sense a part of nature and not apart from it. He is akin, not figuratively but literally, to every living thing, be it an ameba, a tapeworm, a flea, a seaweed, an oak tree, or a monkey—even though the degrees of relationship are different and we may feel less empathy for forty-second cousins like the tapeworms than for, comparatively speaking, brothers like the monkeys. This is togetherness and brotherhood with a vengeance, beyond the wildest dreams of copy writers or of theologians.

Moreover, since man is one of many millions of species all produced by the same grand process, it is in the highest degree improbable that anything in the world exists specifically for his benefit or ill. It is no more true that fruits, for instance, evolved for the delectation of men than that men evolved for the delectation of tigers. Every species, including our own, evolved for its own sake, so to speak. Different species are intricately interdependent, and also some are more successful than others, but there is no divine favoritism. The rational world is not teleological in the old sense. It certainly has purpose, but the purposes are not imposed from without or anticipatory of the future. They are internal to each species separately, relevant only to its functions and usually only to its present condition. Every species is unique, and it is true that man is unique in new and very special ways. Among these peculiarities, parts of the definition of *Homo sapiens,* is the fact that man does have his own purposes that relate to the future—but of man's peculiarities I have more to say below.

The heart of Darwin's explanation of how evolution occurs was natural selection. He always considered this his most important contribution, and posterity agrees with that judgment. It is true that Wallace independently but later reached almost identical views on natural selection and that several others had anticipated both Darwin and Wallace on some points. It is further true that the concept of natural selection has changed through the years since 1859 and that its major importance has occasionally been questioned. Nevertheless, natural selection was primarily Darwin's discovery, later understanding of it has developed from his, and by overwhelming consensus it is now considered the main controlling factor in most evolutionary events.

From the first edition of *The Origin of Species* Darwin expressed the opinion "that natural selection has been the main but not the exclusive means of modification." Yet in the first edition he stressed it almost to the exclusion of other factors. Summing up in the last chapter, he wrote: "Species have changed, and are still slowly changing by the preservation and accumulation of successive slight favorable variations."

That is ambiguous as to what preserves and accumulates the variations, although in context it was obvious that natural selection was supposed to do so. The ambiguity was removed by rewording in the second edition: "Species have been modified, during a long course of descent, by the preservation or the natural selection of many successive slight favorable variations."

There was considerable criticism that Darwin imputed everything, or at any rate too much, to natural selection, and he tended to retreat from so strong a stand. In the fifth edition he changed his previously flat statement by saying that modification of species occurred only "chiefly" through natural selection. In the sixth edition, 1872, the last to be fully revised, Darwin complained that he had been misrepresented, and that he had never thought modification of species due exclusively to natural selection. He made this clear, and unfortunately retreated from a stronger position, by expanding the summary of factors believed to modify species: "This has been effected chiefly through the natural selection of numerous successive, slight, favourable variations; aided in an important manner by the inherited effects of the use and disuse of parts; and in an unimportant manner, that is in relation to adaptive structures, whether past or present, by the direct action of external conditions, and by variations which seem to us in our ignorance to arise spontaneously."

That summarizes the full and final Darwinian theory, which thus recognizes four factors or causes of evolution, in sequence of importance in Darwin's opinion: (1) natural selection; (2) inheritance of acquired characters due to use or disuse of organs; (3) inheritance of acquired characters due to direct effects of the environment; (4) what we now call mutations in the broadest sense.

Darwin rejected, without even mentioning them, various dualistic, vitalistic, or otherwise nonmaterialistic theories of evolution already proposed by 1872. He accepted only factors that were believed to be strictly materialistic or naturalistic, but among those he played safe. He accepted them all, although he considered the last two unimportant as explanations of adaptation. Later in the nineteenth century there was an interesting parceling out of Darwin's four factors into three distinct theories, each emphasizing one or two of those factors at the expense of the others.

One school took the attitude of which Darwin had, as he felt, been falsely accused. They emphasized Darwin's first factor, natural selection, and flatly rejected almost any others, explicitly the inheritance of acquired characters, whether acquired from habit or from environmental influence. Their theory, more Darwinian than Darwin's, came to be called flatly Darwinism or, more specifically, Neo-Darwinism.

A second school of theory accepted and emphasized the inheritance of acquired characters, Darwin's second and third factors, and minimized without necessarily wholly rejecting the influence of natural selection and of mutation. That theory is now usually called Lamarckian or Neo-Lamarckian, but the designations are misleading. Neo-Lamarckism rejects the very heart and basis of Lamarck's personal theory, which was an idealistic and vitalistic view of continuous climbing of a "ladder of nature," from simple to complex beings. Neo-Lamarckism also stresses a factor that Lamarck rejected: inheritance of direct effects of the environment. Neo-Lamarckism is more Darwinian than Lamarckian and is, indeed, about as Darwinian as Neo-Darwinism. It emphasizes Darwin's second and third factors rather than his first one, but it does not wholly reject any Darwinian factor, and it includes nothing that was not explicitly accepted by Darwin. Lamarck's own theory, so different from Neo-Lamarckism, is discussed in Chapter 3.

The third theory here in question emphasized Darwin's fourth factor, his "variations which seem to us in our ignorance to arise spontaneously," now called mutations. This was not, however, a development of parts of Darwinian and to some extent pre-Darwinian theory, as both Neo-Darwinism and Neo-Lamarckism were. To

Darwin, mutation (not yet under that name) was only one way, and the least important way, in which materials for evolution arose. The mutationists were striking out along quite new lines, developing modern genetics and rediscovering Mendelism. The extreme mutationists, notably De Vries, held that mutations were the *only* way in which significant evolutionary change occurs. They reduced natural selection to the minor and negative role of eliminating mutants so grossly malformed as to be unviable. They agreed with the Neo-Darwinians in denying the reality of the inheritance of acquired characters.

Those three theories, tagged as Neo-Darwinism, Neo-Lamarckism, and mutationism, seemed in the early twentieth century to be the principal if not the only alternatives as naturalistic explanations of evolution. They have in common the fact that they are naturalistic. That is, they hold that evolution is a consequence of the material, physical properties of the universe and that it is explicable without postulating any immediate nonphysical, non-natural influences. Beyond that quite basic philosophical point, the three theories do lead to three different views of the world and of man's nature and potentialities.

In the Neo-Darwinian view, the crucial point in evolutionary change is the comparative success of genetic variants in producing offspring. Given a store of varying genetic materials within a population, natural selection usually tends to produce and to increase genetic combinations that are likely to ensure survival and continued reproductive success for the group as a whole. Genetic variation in itself is not considered adaptive in origin, and it is not *directly* influenced by any needs, desires, or activities of individuals in the population. Yet genetic change through the generations is decidedly nonrandom, as a rule, and tends to be adaptive for the population. To that extent, the Neo-Darwinian theory is still accepted by a majority of biologists today. It has not been rejected but only modified by being integrated into a synthesis that is both broader and deeper. The Neo-Darwinian world view originally stressed individual survival, especially competitive success. The later synthesis has involved considerable modification of that emphasis.

The Neo-Lamarckians give prime importance to exactly those factors that were minimized by the Neo-Darwinians: the needs, desires, and activities of individuals. Those factors, together with the modifying influences of soil, climate, food, and other environmental features, are supposed to lead directly to evolutionary change. Thus, genetical modification is supposed to be adaptive in its very essence. The entire process is oriented by the direct, unmediated reactions of individual organisms to their environments. The simplicity of this view is appealing, and it also has a special emotional attraction. It suggests that personal accomplishment counts not only in one's own lifetime but also in posterity and in the eventual evolution of the human species. Improvement in physique by exercise, diet, and so on, may lead to stronger descendants, and education may lead to more intelligent ones. A world in which that was true would on the whole be a pleasant one, and also one in which human progress would be comparatively easy to control. Undoubtedly it is that appeal and its political implications that have made a form of Neo-Lamarckism popular among the rulers of the Soviet Union. As I have already pointed out, there is justification for not labeling this theory with the name of Lamarck. The Russians variously call it "Soviet creative Darwinism" (as opposed to capitalistic and unacceptable Neo-Darwinism), "Michurinism," or "Lysenkoism." Although it is improbable that any of the really able Russian biologists fully accepts that theory in private, it is publicly approved Communist dogma.

The only trouble with Neo-Lamarckism in any of its various seductive guises is that it is not true. Now that we understand the mechanism of inheritance, which Darwin could not know, it is certain that acquired characters cannot possibly be inherited in the way demanded by this theory, and that is that.

The extreme mutationist world view is very different from either the Neo-Darwinian or the Neo-Lamarckian. In it evolution is dominated by chance. Change within species or from one species to another is believed to be not only initiated but also carried through by a mutation or, eventually, a sequence of mutations. Mutations certainly have definite physical causes, even though these are unknown in most specific instances, and they have determinate effects. They arise, however, by chance, and their effects are random in the sense that the cause of a mutation has no evident relationship to the nature of the result and that the effects are unoriented with respect to usefulness or adaptation in the organism. The same cause, such as radiation, may result in any and all kinds of mutations, none producing changes adaptively related to the original radiation. Furthermore, if, for instance, animals are in a situation where increase in size would be adaptive, mutations for larger size do not thereby become either more or less frequent. Mutants are in these senses random or accidental. If there just happens to be some niche into which they can fit, they survive, and a step in evolution has occurred. A recent form of the theory calls such lucky mutants "hopeful monsters." If the mutant does not happen to fit anywhere, it dies, and that is all. Evolution in the mutationist world is not merely aimless but also directionless.

That mutations occur and are random in the stated sense of that word are facts established by innumerable observations. Mutationism, unlike Neo-Lamarckism, rests on a basis of real phenomena. Nevertheless, the further deductions drawn by the original and the extreme mutationists are flatly contradicted by other phenomena, notably those of adaptation. The origin of such an organ as an eye, for example, entirely at random seems almost infinitely improbable. Added to such considerations are many paleontological examples showing evolution occurring through millions of years not fitfully and haphazardly but in a perfectly definite and manifestly adaptive way. The theory that the direction of evolution is fully controlled by mutation simply cannot be true.

Adaptation and the apparent purposefulness of evolution are basic problems that a successful theory *must* solve. The rising science of genetics early in this century not only failed to solve the problem but also made it appear insuperably difficult. That explains why almost no students of other disciplines were inclined to accept mutationism, and why Neo-Lamarckism, an elegant but as we now know incorrect solution, hung on for so long. It also was one of several

reasons for the continued popularity of non-naturalistic theories, to which I allude below.

The way out of the dilemma seems simple now that it has been found. Mutationism is not an alternative to Neo-Darwinism but a supplement to it. If mutation is the source of new variation and yet is substantially nonadaptive, and if the actual course of evolution is to a large extent adaptive, then some additional factor or process must frequently intervene between the occurrence of mutations and the incorporation of some of them into evolving populations. The intervening process must be literally selective, because it must tend (not necessarily with full efficiency) to weed out disadvantageous mutations and genetical combinations and to multiply those that are advantageous in existing circumstances. Natural selection is just such a process, and the principal modern theory of evolution, although it contains much besides, is in large part a synthesis of selection theory and mutation theory.

Evolution is an extremely complex process, and we are here interested mainly in the effects of the concept on our world rather than in the process for its own sake. For that purpose I must, however, briefly note the main elements of the process now known. (Further details will be given in later chapters.) Genetic systems, governing heredity in each individual case, are composed of genes and chromosomes, discrete but complexly interacting units at different levels of size and complexity. The genes themselves, their organized associations in chromosomes, and whole sets of chromosomes have a large degree of stability as units, but all the kinds of units are shuffled and combined in various ways by the sexual processes of reproduction in most organisms. Thus, a considerable amount of variation is maintained, and, so to speak, genetic experimentation occurs in all natural populations. Mutations, in the broadest sense, affecting individual genes, chromosomes, or sets of chromosomes, introduce wholly new variation, which is fed into the processes of recombination.

Populations of similar animals, usually interbreeding among themselves and definable as species, have genetic pools, characterized by the total of genetic units in the included individuals and the distribution of combinations of those units through the population. Evolutionary change involves changes in the genetic pool, in kinds of included units, in frequencies of them, and in kinds and frequencies of combinations of them. Recombination alone does not tend to change the genetic pool. Only four processes are known to do so: mutation, fluctuation in genetic frequencies (what are known statistically as "sampling errors"), inflow of genes from other populations, and differential reproduction. The first three of those processes are not oriented toward adaptation. They are in that sense essentially random, and are usually inadaptive, although they may rarely and coincidentally be adaptive. By "differential reproduction" is meant the consistent production of more offspring, on an average, by individuals with certain genetic characteristics than by those without those particular characteristics. That is the modern understanding of natural selection, including but broader than the Darwinian or Neo-Darwinian concept, which emphasized mortality and survival more than reproduction. Natural selection in the Darwinian sense and still more in this expanded sense is nonrandom, and its trend is adaptive. It also tends, not always with complete success, to counteract the random effects of mutation and sampling error.

Evolutionary processes are tremendously more complicated in detail than this bald outline suggests. The point of the outline is that here is a mechanism, involving only materials and processes known beyond a doubt to occur in nature, capable (as one of its proponents has said) of generating just the degree of improbability evident in the phenomena of evolution.

Further information pertinent to our theme is provided by paleontology, the actual record of events in the history of life. Observation and experimentation with living organisms can extend over a few years, at most. There is always a possibility that processes there evident worked out differently over spans of millions of years, or that the actual history involved principles undetectable in shorter periods of time. There is admittedly some difference of opinion, but I think it fair to say that there is now a consensus for the view that

the fossil record is fully consistent with the modern synthetic theory of evolution and that it neither requires nor suggests any alternative explanation.

There is one thing demonstrated by the fossil record that is decidedly pertinent here and that probably would never have been inferred from study of living organisms. Throughout the whole history of life most species have become extinct, without issue. The statistically usual outcome of evolution is not, then, the progressive appearance of higher forms but simply obliteration. There has, indeed, been progression and even (still more rarely) progress, but this has been in the comparatively few, exceptional lines of descent. The adaptive mechanism of natural selection has guaranteed that some lineages would win, that the world would indeed be filled and kept filled with adapted organisms, but just as inexorably it has insured that most lineages would lose. It has, moreover, had the result that even the winners, the lineages that have survived so far, have not necessarily been progressive, from a human point of view at least. The primitive ameba has remained adapted, hence has survived, while the lordly dinosaurs lost adaptation and therefore life. The degenerate tapeworm is to all appearances as well adapted as the—we like to think—progressive man.

The theory just outlined obviously does not yet answer all questions or plumb all mysteries, even when the details here omitted are taken into consideration. It casts no light on the ultimate mystery—the origin of the universe and the source of the laws or physical properties of matter, energy, space, and time. Nevertheless, once those properties are given, the theory demonstrates that the whole evolution of life could well have ensued, and probably did ensue, automatically, as a natural consequence of the immanent laws and successive configurations of the material cosmos. There is no need, at least, to postulate any non-natural or metaphysical intervention in the course of evolution.

That conclusion has been questioned or opposed not only by many philosophers and theologians but also by a comparatively small number of scientists. The alternatives occasionally supported by scientists or scientific philosophers, and therefore pertinent here, comprise many shadings and variations of opinion, but most of them can be placed in the rubrics of vitalism and finalism.

The vitalists maintain that life is an essence or principle in itself, absent in nonliving matter and not reducible to the interaction of fully material factors. They usually point to a directedness or apparent purposefulness in the development and activities of living things and conclude that the vital, nonmaterial essence within them is a controlling influence in evolution. The finalists maintain that the evolutionary history of life has a preordained over-all pattern which, at least until the appearance of man, was purposefully directed toward a future goal or end. There is no absolute logical necessity that vitalism and finalism should go together, but the ideas are related if only because both are to some degree non-naturalistic and, in that sense, nonmaterialistic. More often than not, vitalists are finalists and finalists are vitalists.

Darwin's legacy in this respect was somewhat but not altogether negative. He did not discuss these issues explicitly and in plain terms. From the whole body of his work, and perhaps more particularly from notes and letters not written for publication, it is clear enough that he felt an antipathy for these philosophical approaches. The very fact that he did not specifically go into these problems amounts to a tacit but positive stand that metaphysical postulates are not necessary for a scientific explanation of evolution.

To that extent it is quite true, as has been so often said, especially by his enemies, that Darwin was a materialist. "Materialist" has become a highly ambiguous word and in some circles a dirty one. It is better here to use the word "naturalist," in the proper philosophical sense of a scientific inquirer who eschews recourse to the supernatural. Such an inquirer does not deny the possible existence of the supernatural but only excludes it from attempts at scientific explanation. Almost all scientists agree that such exclusion is pragmatically justified and indeed necessary. Appeal to the unknown or to the scientifically untestable always stultifies the progress of science, because it stops the search for material explanations that *are* scientifically testable—which, as a matter of experience, have generally been forthcoming when the search has been continued.

Most scientific evolutionists since Darwin have followed his lead in this matter and have continued to seek material, natural explanations of evolution without necessarily taking any overt stand on vitalism or finalism. To the extent that vitalism and finalism are nontestable, that attitude is justified, and the scientist, as scientist, has no right to go further than to repeat the classic remark that he has no need of that hypothesis. However, I do not see how the matter can in all candor be dropped at that point even by the least philosophical of evolutionists, for there are repeated claims by vitalists and finalists that their views *are* testable and that there *is* need for that hypothesis.

Those matters are further discussed later in this book. Here I shall only rather flatly state conclusions. These conclusions are not accepted by all evolutionists, but I think it safe to say that they are by most. The sort of testable evidence that would suggest vitalism or finalism would be the steady progression of life, and of each of its evolving lineages, toward a final and transcendentally worthy goal. That is not, in fact, what the known record of life's history shows. There is no clear over-all progression. Organisms diversify into literally millions of species, then the vast majority of those species perish and other millions take their places for an eon until they, too, are replaced. If that is a foreordained plan, it is an oddly ineffective one. Single lineages, when they can be followed for long, often do show rather steady change, but not indefinitely. They become extinct, or, if they survive, the directions and rates of their evolution change. They evolve exactly as if they were adapting as best they could to a changing world, and not at all as if they were moving toward a set goal. As for the directedness that does indeed characterize vital processes, it is amply explicable by natural selection without requiring any less mundane cause.

That sort of evidence, with much else in detail, convinces me, at least, that the hypotheses of vitalism and finalism are not necessary. Everything proceeds as if they were nonexistent. That does not prove that they are untrue, but it makes their positive adoption unjustified.

Vitalism and finalism have one other aspect that has no particular scientific bearing but that does require mention. They are sometimes advanced with the avowed hope of retaining something from the world of superstition. Vitalism then pretends to find a place in nature for the supernatural. Finalism tries to bring in by the back door the teleology that Darwin swept out the front door. (See Chapter 10.)

Let me summarize and conclude as to this world into which Darwin led us. In it man and all other living things have evolved, ultimately from the nonliving, in accordance with entirely natural, material processes. In part that evolution has been random in the sense of lacking adaptive orientation. As a rule, however, it has been oriented or directed toward achieving and maintaining adaptive relationships between populations of organisms and their whole environments. Nevertheless, this blind, amoral process has not guaranteed indefinite maintenance of adaptation for any given lineage of populations. On the contrary, it usually leads to eventual extinction and a repeopling of the world by the newly divergent offspring of a minority of earlier successful lineages. The mechanism of orientation, the nonrandom element in this extraordinarily complex history, has been natural selection, which is now understood as differential reproduction.

Man is one of the millions of results of this material process. He is another species of animal, but not just another animal. He is unique in peculiar and extraordinarily significant ways. He is probably the most self-conscious of organisms, and quite surely the only one that is aware of his own origins, of his own biological nature. He has developed symbolization to a unique degree and is the only organism with true language. This makes him also the only animal who can store knowledge beyond individual capacity and pass it on beyond individual memory. He is by far the most adaptable of all organisms because he has developed culture as a biological adaptation. Now his culture evolves not distinct from and not in replacement of but in addition to biological evolution, which also continues.

Concomitant with these developments is the fact that man has unique moral qualities. The evolutionary process is not moral—the word is simply irrelevant in that connection—but it has finally produced a moral animal. Conspicuous among his moral attributes is a sense of responsibility, which is probably felt in some way and to some degree by every normal human being. There has been disagreement and indeed confusion through the ages regarding to whom and for what man is responsible. The lower and the higher superstitions have produced their several answers. In the post-Darwinian world another answer seems fairly clear: man is responsible to himself and for himself. "Himself" here means the whole human species, not only the individual and certainly not just those of a certain color of hair or cast of features.

The fact that man knows that he evolves entails the possibility that he can do something to influence his own biological destiny. The fact that uncontrolled evolution often leads to degeneration and usually to extinction makes it highly advisable that man take a hand in determining his own future evolution. If man proceeds on the wrong evolutionary assumptions—for instance, on those of Neo-Lamarckism or Michurinism—whatever he does is sure to be wrong. If he proceeds on the right assumptions, what he does may still be wrong, but at least it has a chance of being right.

A world in which man must rely on himself, in which he is not the darling of the gods but only another, albeit extraordinary, aspect of nature, is by no means congenial to the immature or the wishful thinkers. That is plainly a major reason why even now, a hundred years after *The Origin of Species*, most people have not really entered the world into which Darwin led—alas!—only a minority of us. Life may conceivably be happier for some people in the older worlds of superstition. It is possible that some children are made happy by a belief in Santa Claus, but adults should prefer to live in a world of reality and reason.

Perhaps I should end this chapter on that note of mere preference, but it is impossible to do so. It is a characteristic of this world to which Darwin opened the door that unless *most* of us do enter it and live maturely and rationally in it, the future of mankind is dim indeed—if there is any future.

JOHN DEWEY

Most people, when they look at a spectrum, see a series of distinct colors side by side. When John Dewey looked at a spectrum he saw a continuum—a shading of one color into another, with no boundaries to indicate precisely where one color ends and another begins. Mind fades into matter, subject into object, means into ends. The individual merges with the social, liberal education mixes with vocational; science itself is part of a spectrum of things people do, like plowing the earth and sailing ships. There are no eternal essences with fixed outlines. The species move. "Truth" is simply that plastic, growing body of knowledge which serves as a tool in man's struggle to perpetuate his species.

No philosopher wasted less time brooding in metaphysical towers than this absent-minded, carelessly dressed pedagogue with the rimless glasses and Vermont drawl. He was active in hundreds of liberal organizations and causes. His influence on political thought, spelling out the meaning of such terms as "freedom" and "democracy," has been immense. He was never afraid to take partisan positions even when they were unfashionable; for instance, his vigorous condemnation of Stalin's purge trials at a time when most liberals tried to look the other way. Perhaps his greatest influence was in the field of elementary education. The old-fashioned bolted-down desk symbolized for him the old restraints. He wanted to unbolt them. He wanted to unbolt the mind. It could be done, he believed, only by extending the scientific attitude into every phase of human activity.

By a pleasant coincidence, John Dewey (1859-1952) was born the same year that *The Origin of Species* appeared. For Dewey, evolution was the great dissolver of fusty absolutisms,

---

and in the select
ing so. Originall
one of his best
very heart of h
technical writin
which led Max
a quotable sent
pile of his 36 l
can recover su
them," Dewey
questions of hi
alternatives, "w

m, see a series of
ewey looked at a
of one color into
sely where one
into matter, sub-
dual merges with
cational; science
do, like plowing
eternal essences
h" is simply that
serves as a tool

: in metaphysical
essed pedagogue
l. He was active
es. His influence
ng of such terms
nmense. He was
when they were
ndemnation of
rals tried to look
was in the field
olted-down desk
anted to unbolt
uld be done, he
rude into every

1859-1952) was
s appeared. For
sty absolutisms,

and in the selection chosen here he gives his reasons for thinking so. Originally a lecture delivered in 1909, it has become one of his best known, most influential essays, going to the very heart of his pragmatic philosophy. In later and more technical writings his style was often involved and dull, a fact which led Max Eastman to observe that if Dewey ever wrote a quotable sentence it had become permanently lost in the pile of his 36 books and 815 magazine articles. Perhaps we can recover such a sentence in this essay. "We do not solve them," Dewey concludes, concerning those great, burning questions of history that seem to demand exclusive either/or alternatives, "we get over them."

JOHN DEWEY

The Influence of Darwinism on Philosophy

I

THAT THE publication of the "Origin of Species" marked an epoch in the development of the natural sciences is well known to the layman. That the combination of the very words origin and species embodied an intellectual revolt and introduced a new intellectual temper is easily overlooked by the expert. The conceptions that had reigned in the philosophy of nature and knowledge for two thousand years, the conceptions that had become the familiar furniture of the mind, rested on the assumption of the superiority of the fixed and final; they rested upon treating change and origin as signs of defect and unreality. In laying hands upon the sacred ark of absolute permanency, in treating the forms that had been regarded as types of fixity and perfection as originating and passing away, the "Origin of Species" introduced a mode of thinking that in the end was bound to transform the logic of knowledge, and hence the treatment of morals, politics, and religion.

No wonder, then, that the publication of Darwin's book, a half century ago, precipitated a crisis. The true nature of the controversy is easily concealed from us, however, by the theological clamor that attended it. The vivid and popular features of the anti-Darwinian row tended to leave the impression that the issue was between science on one side and theology on the other. Such was not the case—the issue lay inism

ecies" marked an
sciences is well
of the very words
revolt and intro-
verlooked by the
n the philosophy
l years, the con-
ure of the mind,
of the fixed and
origin as signs
n the sacred ark
ns that had been
s originating and
luced a mode of
form the logic of
als, politics, and

Darwin's book,
e true nature of
however, by the
vid and popular
to leave the im-
on one side and
se—the issue lay

---

primarily within science itself, as Darwin himself early recognized. The theological outcry he discounted from the start, hardly noticing it save as it bore upon the "feelings of his female relatives." But for two decades before final publication he contemplated the possibility of being put down by his scientific peers as a fool or as crazy; and he set, as the measure of his success, the degree in which he should affect three men of science: Lyell in geology, Hooker in botany, and Huxley in zoology.

Religious considerations lent fervor to the controversy, but they did not provoke it. Intellectually, religious emotions are not creative but conservative. They attach themselves readily to the current view of the world and consecrate it. They steep and dye intellectual fabrics in the seething vat of emotions; they do not form their warp and woof. There is not, I think, an instance of any large idea about the world being independently generated by religion. Although the ideas that rose up like armed men against Darwinism owed their intensity to religious associations, their origin and meaning are to be sought in science and philosophy, not in religion.

II

Few words in our language foreshorten intellectual history as much as does the word species. The Greeks, in initiating the intellectual life of Europe, were impressed by characteristic traits of the life of plants and animals; so impressed indeed that they made these traits the key to defining nature and to explaining mind and society. And truly, life is so wonderful that a seemingly successful reading of its mystery might well lead men to believe that the key to the secrets of heaven and earth was in their hands. The Greek rendering of this mystery, the Greek formulation of the aim and standard of knowledge, was in the course of time embodied in the word species, and it controlled philosophy for two thousand years. To understand the intellectual face-about expressed in the phrase "Origin of Species" we must, then, understand the long dominant idea against which it is a protest.

Consider how men were impressed by the facts of life. Their eyes fell upon certain things slight in bulk, and frail in structure. To every appearance, these perceived things were inert and passive. Suddenly, under certain circumstances, these things—henceforth known as seeds or eggs or germs—begin to change, to change rapidly in size, form, and qualities. Rapid and extensive changes occur, however, in many things—as when wood is touched by fire. But the changes in the living thing are orderly; they are cumulative; they tend constantly in one direction; they do not, like other changes, destroy or consume, or pass fruitless into wandering flux; they realize and fulfil. Each successive stage, no matter how unlike its predecessor, preserves its net effect and also prepares the way for a fuller activity on the part of its successor. In living beings, changes do not happen as they seem to happen elsewhere, any which way; the earlier changes are regulated in view of later results. This progressive organization does not cease till there is achieved a true final term, a τέλος, a completed, perfected end. This final form exercises in turn a plenitude of functions, not the least noteworthy of which is production of germs like those from which it took its own origin, germs capable of the same cycle of self-fulfilling activity.

But the whole miraculous tale is not yet told. The same drama is enacted to the same destiny in countless myriads of individuals so sundered in time, so severed in space, that they have no opportunity for mutual consultation and no means of interaction. As an old writer quaintly said, "things of the same kind go through the same formalities"—celebrate, as it were, the same ceremonial rites.

This formal activity which operates throughout a series of changes and holds them to a single course; which subordinates their aimless flux to its own perfect manifestation; which, leaping the boundaries of space and time, keeps individuals distant in space and remote in time to a uniform type of structure and function: this principle seemed to give insight into the very nature of reality itself. To it Aristotle gave the name, εἶδος. This term the scholastics translated as *species*.

by the facts of life.
it in bulk, and frail in
perceived things were
ertain circumstances,
or eggs or germs—
e, form, and qualities.
vever, in many things
it the changes in the
mulative; they tend
t like other changes,
wandering flux; they
no matter how unlike
and also prepares the
ts successor. In living
seem to happen else-
nges are regulated
organization does not
term, a τέλος, a com-
exercises in turn a
eworthy of which is
ich it took its own
ycle of self-fulfilling

yet told. The same
countless myriads of
ered in space, that
consultation and no
uaintly said, "things
rmalities"—celebrate,

roughout a series of
which subordinates
anifestation; which,
e, keeps individuals
a uniform type of
med to give insight
Aristotle gave the
anslated as *species*.

The force of this term was deepened by its application to everything in the universe that observes order in flux and manifests constancy through change. From the casual drift of daily weather, through the uneven recurrence of seasons and unequal return of seed time and harvest, up to the majestic sweep of the heavens—the image of eternity in time—and from this to the unchanging pure and contemplative intelligence beyond nature lies one unbroken fulfilment of ends. Nature as a whole is a progressive realization of purpose strictly comparable to the realization of purpose in any single plant or animal.

The conception of εἶδος, species, a fixed form and final cause, was the central principle of knowledge as well as of nature. Upon it rested the logic of science. Change as change is mere flux and lapse; it insults intelligence. Genuinely to know is to grasp a permanent end that realizes itself through changes, holding them thereby within the metes and bounds of fixed truth. Completely to know is to relate all special forms to their one single end and good: pure contemplative intelligence. Since, however, the scene of nature which directly confronts us is in change, nature as directly and practically experienced does not satisfy the conditions of knowledge. Human experience is in flux, and hence the instrumentalities of sense-perception and of inference based upon observation are condemned in advance. Science is compelled to aim at realities lying behind and beyond the processes of nature, and to carry on its search for these realities by means of rational forms transcending ordinary modes of perception and inference.

There are, indeed, but two alternative courses. We must either find the appropriate objects and organs of knowledge in the mutual interactions of changing things; or else, to escape the infection of change, we *must* seek them in some transcendent and supernal region. The human mind, deliberately as it were, exhausted the logic of the changeless, the final, and the transcendent, before it essayed adventure on the pathless wastes of generation and transformation. We dispose all too easily of the efforts of the schoolmen to interpret nature and mind in terms of real essences, hidden forms, and occult faculties, forgetful of the seriousness and dignity of the ideas that lay behind. We dispose of them by laughing at the famous gentleman who accounted for the fact that opium put people to sleep on the ground it had a dormitive faculty. But the doctrine, held in our own day, that knowledge of the plant that yields the poppy consists in referring the peculiarities of an individual to a type, to a universal form, a doctrine so firmly established that any other method of knowing was conceived to be unphilosophical and unscientific, is a survival of precisely the same logic. This identity of conception in the scholastic and anti-Darwinian theory may well suggest greater sympathy for what has become unfamiliar as well as greater humility regarding the further unfamiliarities that history has in store.

Darwin was not, of course, the first to question the classic philosophy of nature and of knowledge. The beginnings of the revolution are in the physical science of the sixteenth and seventeenth centuries. When Galileo said: "It is my opinion that the earth is very noble and admirable by reason of so many and so different alterations and generations which are incessantly made therein," he expressed the changed temper that was coming over the world; the transfer of interest from the permanent to the changing. When Descartes said: "The nature of physical things is much more easily conceived when they are beheld coming gradually into existence, than when they are only considered as produced at once in a finished and perfect state," the modern world became self-conscious of the logic that was henceforth to control it, the logic of which Darwin's "Origin of Species" is the latest scientific achievement. Without the methods of Copernicus, Kepler, Galileo, and their successors in astronomy, physics, and chemistry, Darwin would have been helpless in the organic sciences. But prior to Darwin the impact of the new scientific method upon life, mind, and politics, had been arrested, because between these ideal or moral interests and the inorganic world intervened the kingdom of plants and animals. The gates of the garden of life were barred to the

---

new ideas; and only throu
to mind and politics. The in
resides in his having conque
principle of transition, and
application to mind and m
species what Galileo had s
he emancipated, once for
as an organon of asking q
tions.

The exact bearings upo
look are, of course, as yet,
the twilight of intellectual
ness of the prophet to th
venture a systematic expos
phy of the Darwinian m
as to its general bearing
and complexion, upon t
instinctive intellectual aver
mine, after all, our more d
this vague inquiry there
stone a problem of long
much discussed in Darwi
problem of design *versus*
causal explanation, first or

As we have already see
ried with it the idea of p
type is present directing
realization of its own perf
tive principle is not visible
be an ideal or rational forc
is gradually approximat
also follows that in and t
ideal force is working o
These inferences were ext
ing in vain; but all for an

hidden forms, and occult
's and dignity of the ideas
hem by laughing at the
or the fact that opium
t had a dormitive faculty.
ay, that knowledge of the
in referring the peculiar-
universal form, a doctrine
method of knowing was
unscientific, is a survival
dentity of conception in
theory may well suggest
come unfamiliar as well
rther unfamiliarities that

st to question the classic
edge. The beginnings of
ence of the sixteenth and
said: "It is my opinion
admirable by reason
is and generations which
expressed the changed
orld; the transfer of in-
anging. When Descartes
is much more easily
coming gradually into
nsidered as produced at
" the modern world be-
vas henceforth to control
"Origin of Species" is
ithout the methods of
successors in astronomy,
d have been helpless in
rwin the impact of the
, and politics, had been
or moral interests and
kingdom of plants and
life were barred to the

new ideas; and only through this garden was there access to mind and politics. The influence of Darwin upon philosophy resides in his having conquered the phenomena of life for the principle of transition, and thereby freed the new logic for application to mind and morals and life. When he said of species what Galileo had said of the earth, *e pur se muove*, he emancipated, once for all, genetic and experimental ideas as an organon of asking questions and looking for explanations.

III

The exact bearings upon philosophy of the new logical outlook are, of course, as yet, uncertain and inchoate. We live in the twilight of intellectual transition. One must add the rashness of the prophet to the stubbornness of the partizan to venture a systematic exposition of the influence upon philosophy of the Darwinian method. At best, we can but inquire as to its general bearing—the effect upon mental temper and complexion, upon that body of half-conscious, half-instinctive intellectual aversions and preferences which determine, after all, our more deliberate intellectual enterprises. In this vague inquiry there happens to exist as a kind of touchstone a problem of long historic currency that has also been much discussed in Darwinian literature. I refer to the old problem of design *versus* chance, mind *versus* matter, as the causal explanation, first or final, of things.

As we have already seen, the classic notion of species carried with it the idea of purpose. In all living forms, a specific type is present directing the earlier stages of growth to the realization of its own perfection. Since this purposive regulative principle is not visible to the senses, it follows that it must be an ideal or rational force. Since, however, the perfect form is gradually approximated through the sensible changes, it also follows that in and through a sensible realm a rational ideal force is working out its own ultimate manifestation. These inferences were extended to nature: (*a*) She does nothing in vain; but all for an ulterior purpose. (*b*) Within natural sensible events there is therefore contained a spiritual causal force, which as spiritual escapes perception, but is apprehended by an enlightened reason. (*c*) The manifestation of this principle brings about a subordination of matter and sense to its own realization, and this ultimate fulfilment is the goal of nature and of man. The design argument thus operated in two directions. Purposefulness accounted for the intelligibility of nature and the possibility of science, while the absolute or cosmic character of this purposefulness gave sanction and worth to the moral and religious endeavors of man. Science was underpinned and morals authorized by one and the same principle, and their mutual agreement was eternally guaranteed.

This philosophy remained, in spite of sceptical and polemic outbursts, the official and the regnant philosophy of Europe for over two thousand years. The expulsion of fixed first and final causes from astronomy, physics, and chemistry had indeed given the doctrine something of a shock. But, on the other hand, increased acquaintance with the details of plant and animal life operated as a counterbalance and perhaps even strengthened the argument from design. The marvelous adaptations of organisms to their environment, of organs to the organism, of unlike parts of a complex organ—like the eye—to the organ itself; the foreshadowing by lower forms of the higher; the preparation in earlier stages of growth for organs that only later had their functioning—these things were increasingly recognized with the progress of botany, zoology, paleontology, and embryology. Together, they added such prestige to the design argument that by the late eighteenth century it was, as approved by the sciences of organic life, the central point of theistic and idealistic philosophy.

The Darwinian principle of natural selection cut straight under this philosophy. If all organic adaptations are due simply to constant variation and the elimination of those variations which are harmful in the struggle for existence that is brought about by excessive reproduction, there is no call for a prior intelligent causal force to plan and preordain them.

Hostile critics charged D
making chance the cause
Some naturalists, like A
principle and attempted t
held to what may be callec
we conceive the "stream o
we may suppose that each
from the first to be selecte
and selection simply defi
causes" through which the
of design is none the wors
modus operandi.
Darwin could not accep
or rather he asserts that it
mense and wonderful unive
of looking far backwards a
blind chance or necessity.
since variations are in use
since the latter are sifted
conditions of struggle for
applied to living beings is
port there deprives it of s
in general. If the variatio
ficial selection give the p
for the sake of the breede
variations resulting in nat

So much for some of th
sion of design *versus* chanc
of life as a whole. We bro
as a crucial instance. Wh
to the bearing of Darwin
first place, the new logic o
will—one type of problems
Philosophy forswears inqu

ɔre contained a spiritual
escapes perception, but is
ason. (*c*) The manifesta-
a subordination of matter
d this ultimate fulfilment is
The design argument thus
efulness accounted for the
bility of science, while the
this purposefulness gave
nd religious endeavors of
nd morals authorized by
eir mutual agreement was

of sceptical and polemic
int philosophy of Europe
xpulsion of fixed first and
s, and chemistry had in-
of a shock. But, on the
vith the details of plant
interbalance and perhaps
n design. The marvelous
environment, of organs
complex organ—like the
wing by lower forms of
stages of growth for or-
ɔning—these things were
gress of botany, zoology,
ether, they added such
by the late eighteenth
ences of organic life, the
philosophy.
al selection cut straight
ic adaptations are due
elimination of those
uggle for existence that
tion, there is no call for
n and preordain them.

Hostile critics charged Darwin with materialism and with making chance the cause of the universe.

Some naturalists, like Asa Gray, favored the Darwinian principle and attempted to reconcile it with design. Gray held to what may be called design on the installment plan. If we conceive the "stream of variations" to be itself intended, we may suppose that each successive variation was designed from the first to be selected. In that case, variation, struggle, and selection simply define the mechanism of "secondary causes" through which the "first cause" acts; and the doctrine of design is none the worse off because we know more of its *modus operandi.*

Darwin could not accept this mediating proposal. He admits or rather he asserts that it is "impossible to conceive this immense and wonderful universe including man with his capacity of looking far backwards and far into futurity as the result of blind chance or necessity." But nevertheless he holds that since variations are in useless as well as useful directions, and since the latter are sifted out simply by the stress of the conditions of struggle for existence, the design argument as applied to living beings is unjustifiable; and its lack of support there deprives it of scientific value as applied to nature in general. If the variations of the pigeon, which under artificial selection give the pouter pigeon, are not preordained for the sake of the breeder, by what logic do we argue that variations resulting in natural species are pre-designed?

IV

So much for some of the more obvious facts of the discussion of design *versus* chance, as causal principles of nature and of life as a whole. We brought up this discussion, you recall, as a crucial instance. What does our touchstone indicate as to the bearing of Darwinian ideas upon philosophy? In the first place, the new logic outlaws, flanks, dismisses—what you will—one type of problems and substitutes for it another type. Philosophy forswears inquiry after absolute origins and absolute finalities in order to explore specific values and the specific conditions that generate them.

Darwin concluded that the impossibility of assigning the world to chance as a whole and to design in its parts indicated the insolubility of the question. Two radically different reasons, however, may be given as to why a problem is insoluble. One reason is that the problem is too high for intelligence; the other is that the question in its very asking makes assumptions that render the question meaningless. The latter alternative is unerringly pointed to in the celebrated case of design *versus* chance. Once admit that the sole verifiable or fruitful object of knowledge is the particular set of changes that generate the object of study together with the consequences that then flow from it, and no intelligible question can be asked about what, by assumption, lies outside. To assert—as is often asserted—that specific values of particular truth, social bonds and forms of beauty, if they can be shown to be generated by concretely knowable conditions, are meaningless and in vain; to assert that they are justified only when they and their particular causes and effects have all at once been gathered up into some inclusive first cause and some exhaustive final goal, is intellectual atavism. Such argumentation is reversion to the logic that explained the extinction of fire by water through the formal essence of aqueousness and the quenching of thirst by water through the final cause of aqueousness. Whether used in the case of the special event or that of life as a whole, such logic only abstracts some aspect of the existing course of events in order to reduplicate it as a petrified eternal principle by which to explain the very changes of which it is the formalization.

When Henry Sidgwick casually remarked in a letter that as he grew older his interest in what or who made the world was altered into interest in what kind of a world it is anyway, his voicing of a common experience of our own day illustrates also the nature of that intellectual transformation effected by the Darwinian logic. Interest shifts from the wholesale essence back of special changes to the question of how special changes serve and defeat concrete purposes; shifts from an intelli-

specific values and the
n.

ssibility of assigning the
gn in its parts indicated
o radically different rea-
ly a problem is insoluble.
high for intelligence; the
sking makes assumptions
s. The latter alternative
ted case of design *versus*
rifiable or fruitful object
hanges that generate the
consequences that then
tion can be asked about
To assert—as is often
cular truth, social bonds
shown to be generated
are meaningless and in
nly when they and their
at once been gathered
d some exhaustive final
gumentation is reversion
inction of fire by water
sness and the quenching
cause of aqueousness.
pecial event or that of
racts some aspect of the
to reduplicate it as a
h to explain the very
on.

narked in a letter that
or who made the world
of a world it is anyway,
f our own day illustrates
ansformation effected by
n the wholesale essence
n of how special changes
; shifts from an intelligence that shaped things once for all to the particular intelligences which things are even now shaping; shifts from an ultimate goal of good to the direct increments of justice and happiness that intelligent administration of existent conditions may beget and that present carelessness or stupidity will destroy or forego.

In the second place, the classic type of logic inevitably set philosophy upon proving that life *must* have certain qualities and values—no matter how experience presents the matter—because of some remote cause and eventual goal. The duty of wholesale justification inevitably accompanies all thinking that makes the meaning of special occurrences depend upon something that once and for all lies behind them. The habit of derogating from present meanings and uses prevents our looking the facts of experience in the face; it prevents serious acknowledgment of the evils they present and serious concern with the goods they promise but do not as yet fulfil. It turns thought to the business of finding a wholesale transcendent remedy for the one and guarantee for the other. One is reminded of the way many moralists and theologians greeted Herbert Spencer's recognition of an unknowable energy from which welled up the phenomenal physical processes without and the conscious operations within. Merely because Spencer labeled his unknowable energy "God," this faded piece of metaphysical goods was greeted as an important and grateful concession to the reality of the spiritual realm. Were it not for the deep hold of the habit of seeking justification for ideal values in the remote and transcendent, surely this reference of them to an unknowable absolute would be despised in comparison with the demonstrations of experience that knowable energies are daily generating about us precious values.

The displacing of this wholesale type of philosophy will doubtless not arrive by sheer logical disproof, but rather by growing recognition of its futility. Were it a thousand times true that opium produces sleep because of its dormitive energy, yet the inducing of sleep in the tired, and the recovery to waking life of the poisoned, would not be thereby one least step forwarded. And were it a thousand times dialectically demonstrated that life as a whole is regulated by a transcendent principle to a final inclusive goal, none the less truth and error, health and disease, good and evil, hope and fear in the concrete, would remain just what and where they now are. To improve our education, to ameliorate our manners, to advance our politics, we must have recourse to specific conditions of generation.

Finally, the new logic introduces responsibility into the intellectual life. To idealize and rationalize the universe at large is after all a confession of inability to master the courses of things that specifically concern us. As long as mankind suffered from this impotency, it naturally shifted a burden of responsibility that it could not carry over to the more competent shoulders of the transcendent cause. But if insight into specific conditions of value and into specific consequences of ideas is possible, philosophy must in time become a method of locating and interpreting the more serious of the conflicts that occur in life, and a method of projecting ways for dealing with them: a method of moral and political diagnosis and prognosis.

The claim to formulate *a priori* the legislative constitution of the universe is by its nature a claim that may lead to elaborate dialectic developments. But it is also one that removes these very conclusions from subjection to experimental test, for, by definition, these results make no differences in the detailed course of events. But a philosophy that humbles its pretensions to the work of projecting hypotheses for the education and conduct of mind, individual and social, is thereby subjected to test by the way in which the ideas it propounds work out in practice. In having modesty forced upon it, philosophy also acquires responsibility.

Doubtless I seem to have violated the implied promise of my earlier remarks and to have turned both prophet and partizan. But in anticipating the direction of the transformations in philosophy to be wrought by the Darwinian genetic and experimental logic, I do not profess to speak for any save those who yield themselves consciously or unconsciously to this logic. No one can fairly deny that at present there are

---

two effects of the Darw
hand, they are making
revise our traditional ph
with its demands. On th
recrudescence of absolu
type of philosophic kno
one which opens to us
which the sciences give
to something that essenti
action affects popular cr
as technical philosophies
sciences by the new ide
plicit and rigid separatio

Old ideas give way sl
logical forms and catego
deeply engrained attitud
over, the conviction pe
a hallucination—that all
has asked are questions
alternatives that the q
fact intellectual progress
donment of questions
they assume—an aband
creasing vitality and a c
solve them: we get ove
disappearing, evaporatin
ing to the changed atti
their place. Doubtless th
thought of old questio
methods, new intentions
by the scientific revolutic
of Species."

two effects of the Darwinian mode of thinking. On the one hand, they are making many sincere and vital efforts to revise our traditional philosophic conceptions in accordance with its demands. On the other hand, there is as definitely a recrudescence of absolutistic philosophies; an assertion of a type of philosophic knowing distinct from that of the sciences, one which opens to us another kind of reality from that to which the sciences give access; an appeal through experience to something that essentially goes beyond experience. This reaction affects popular creeds and religious movements as well as technical philosophies. The very conquest of the biological sciences by the new ideas has led many to proclaim an explicit and rigid separation of philosophy from science.

Old ideas give way slowly; for they are more than abstract logical forms and categories. They are habits, predispositions, deeply engrained attitudes of aversion and preference. Moreover, the conviction persists—though history shows it to be a hallucination—that all the questions that the human mind has asked are questions that can be answered in terms of the alternatives that the questions themselves present. But in fact intellectual progress usually occurs through sheer abandonment of questions together with both of the alternatives they assume—an abandonment that results from their decreasing vitality and a change of urgent interest. We do not solve them: we get over them. Old questions are solved by disappearing, evaporating, while new questions corresponding to the changed attitude of endeavor and preference take their place. Doubtless the greatest dissolvent in contemporary thought of old questions, the greatest precipitant of new methods, new intentions, new problems, is the one effected by the scientific revolution that found its climax in the "Origin of Species."

reptile into mammal is believed to have required several tens of millions of years.

The interpretation of the evolutionary process as very slow and gradual change due to small, or micromutations, in combination with the sorting out that accompanies sexual reproduction, all influenced by natural selection in accordance with the environment, is termed the neo-Darwinian interpretation of evolution. The essence of Darwinism is retained but Darwin's theories have been modified to accommodate them to the discoveries made since his time in genetics, molecular biology, etc. With very few exceptions, all evolutionists are neo-Darwinists.

PREDICTIONS BASED ON THE CREATION MODEL AND THE EVOLUTION MODEL

In the preceding discussion we have defined what is meant by creation and evolution. We have described the evolutionary mechanism that is accepted by almost all evolutionists, and we have described all that human knowledge will permit about the creation process. We are now ready to predict the evidence that must be found in the fossil record based on the creation model on the one hand and the evolution model on the other hand.

Creation Model. On the basis of the creation model, we would predict an explosive appearance in the fossil record of highly complex forms of life without evidence of ancestral forms. We would predict that all of the major types of life, that is, the basic plant and animal forms, would appear abruptly in the fossil record without evidence of transitional forms linking one basic kind to another.

We would thus expect to find the fossilized remains, for example, of cats, dogs, bears, elephants, cows, horses, bats, dinosaurs, crocodiles, monkeys, apes, and men without evidence of common ancestors. Each major kind at its earliest appearance in the fossil record would possess, fully developed, all the characteristics that are used to define that particular kind.

Evolution Model. On the basis of the evolution model, we would predict that the most ancient strata in which fossils are found would contain the most primitive forms of life capable of leaving a fossil record. As successively younger strata were searched, we would expect to see the gradual transition of these relatively simple forms of life into more and more complex forms of life. As living forms diverged into the millions of species which have existed in the past and which exist today, we would expect to find a slow and gradual transition of one form into another.

We would predict that new basic types would *not* appear suddenly in the fossil record possessing all of the characteristics that are used to define its kind. The earliest forms in each group would be expected to possess in incipient form some of the characteristics which are used to define that group while retaining characteristics used to define the ancestral group.

If fish evolved into amphibia, as evolutionists believe, then we would predict that we would find transitional forms showing the gradual transition of fins into feet and legs. Of course, many other alterations in the anatomy and physiology of fishes would have to occur to change an animal adapted to living its entire life span in water to one which

Gish, D.L.
Evolution, The Fossils say no!
"Predictions of evolutionary
theory and creationism"
Creation-Life Publisher
Used by permission

spends most of its life outside of water. The fin-to-feet transition would be an easily traceable transition, however.

If reptiles gave rise to birds, then we would expect to find transitional forms in the fossil record showing the gradual transition of the forelimbs of the ancestral reptile into the wings of a bird, and the gradual transition of some structure on the reptile into the feathers of a bird. These again are obvious transitions that could be easily traced in the fossil record. Of course, many other changes would have been taking place at the same time, such as the conversion of the hindfeet of the reptile into the perching feet of the bird, reptilian skull into bird-like skull, etc.

In the pterosaurs, or flying reptiles, the wing membrane was supported by an enormously lengthened fourth finger. If the pterosaurs actually evolved from a nonflying reptile, then we would predict that the fossil record would produce transitional forms showing a gradual increase in length of the fourth finger, along with the origin of other unique structures.

The fossil record ought to produce thousands upon thousands of transitional, or in-between forms. It is true that according to evolutionary geology only a tiny fraction of all plants and animals that have ever existed would have been preserved as fossils. It is also true that we have as yet uncovered only a small fraction of the fossils that are entombed in the rocks. We have, nevertheless, recovered a good representative number of the fossils that exist.

Sampling of the fossil record has now been so thorough that appeals to the imperfections in the record are no longer valid. George has stated:

> There is no need to apologize any longer for the poverty of the fossil record. In some ways it has become almost unmanageably rich and discovery is outpacing integration.[10]

It seems clear, then, that after 150 years of intense searching a large number of obvious transitional forms would have been discovered if the predictions of evolution theory are valid.

We have, for example, discovered literally billions of fossils of ancient invertebrates and many fossils of ancient fishes. The transition of invertebrate into vertebrate is believed to have required many millions of years. Populations are supposed to constitute the units of evolution and, of course, only successful populations survive. It seems obvious, then, that if we find fossils of the invertebrates which were supposed to have been ancestral to fishes, and if we find fossils of the fishes, we surely ought to find the fossils of the transitional forms.

We find fossils of crossopterygian fishes which are alleged to have given rise to the amphibia. We find fossils of the so-called "primitive" amphibia. Since the transition from fish to amphibia would have required many millions of years, during which many hundreds of millions, even billions, of the transitional forms must have lived and died, many of these transitional forms should have been discovered in the fossil record even though only a minute fraction of these animals have been recovered as fossils. As a matter of fact, the discovery of only five or six of the transitional forms scattered through time would be sufficient to document evolution.

So it would be throughout the entire fossil record.

There should not be the slightest difficulty in finding transitional forms. Hundreds of transitional forms should fill museum collections. If we find fossils at all, we ought to find transitional forms. As a matter of fact, difficulty in placing a fossil within a distinct category should be the rule rather than the exception.

SUMMARY

The contrast between the two models and the predictions based on each model may be summarized as follows:

Creation Model	Evolution Model
By acts of a Creator.	By naturalistic mechanistic processes due to properties inherent in inanimate matter.
Creation of basic plant and animal kinds with ordinal characteristics complete in first representatives.	Origin of all living things from a single living source which itself arose from inanimate matter. Origin of each kind from an ancestral form by slow gradual change.
Variation and speciation limited within each kind.	Unlimited variation. All forms genetically related.

These two models would permit the following predictions to be made concerning the fossil record:

Creation Model	Evolution Model
Sudden appearance in great variety of highly complex forms.	Gradual change of simplest forms into more and more complex forms.
Sudden appearance of each created kind with ordinal characteristics complete. Sharp boundaries separating major taxonomic groups. No transitional forms between higher categories.	Transitional series linking all categories. No systematic gaps.

Creationism in Schools: The Decision in McLean versus the Arkansas Board of Education

On 5 January 1982 U.S. District Court Judge William R. Overton enjoined the Arkansas Board of Education from implementing the "Balanced Treatment for Creation-Science and Evolution-Science Act" of the state legislature. This is the complete text of his judgment, injunction, and opinion in the case.

Judgment

Pursuant to the Court's Memorandum Opinion filed this date, judgment is hereby entered in favor of the plaintiffs and against the defendants. The relief prayed for is granted.

Dated this January 5, 1982.

Injunction

Pursuant to the Court's Memorandum Opinion filed this date, the defendants and each of them and all their servants and employees are hereby permanently enjoined from implementing in any manner Act 590 of the Acts of Arkansas of 1981.

It is so ordered this January 5, 1982.

Memorandum Opinion

Introduction

On March 19, 1981, the Governor of Arkansas signed into law Act 590 of 1981, entitled the "Balanced Treatment for Creation-Science and Evolution-Science Act." The Act is codified as Ark. Stat. Ann. §80-1663, *et seq.*, (1981 Supp.). Its essential mandate is stated in its first sentence: "Public schools within this State shall give balanced treatment to creation-science and to evolution-science." On May 27, 1981, this suit was filed (*1*) challenging the constitutional validity of Act 590 on three distinct grounds.

First, it is contended that Act 590 constitutes an establishment of religion prohibited by the First Amendment to the Constitution, which is made applicable to the states by the Fourteenth Amendment. Second, the plaintiffs argue the Act violates a right to academic freedom which they say is guaranteed to students and teachers by the Free Speech Clause of the First Amendment. Third, plaintiffs allege the Act is impermissibly vague and thereby violates the Due Process Clause of the Fourteenth Amendment.

The individual plaintiffs include the resident Arkansas Bishops of the United Methodist, Episcopal, Roman Catholic and African Methodist Episcopal Churches, the principal official of the Presbyterian Churches in Arkansas, other United Methodist, Southern Baptist and Presbyterian clergy, as well as several persons who sue as parents and next friends of minor children attending Arkansas public schools. One plaintiff is a high school biology teacher. All are also Arkansas taxpayers. Among the organizational plaintiffs are the American Jewish Congress, the Union of American Hebrew Congregations, the American Jewish Committee, the Arkansas Education Association, the National Association of Biology Teachers and the National Coalition for Public Education and Religious Liberty, all of which sue on behalf of members living in Arkansas (*2*).

The defendants include the Arkansas Board of Education and its members, the Director of the Department of Education, and the State Textbooks and Instructional Materials Selecting Committee (*3*). The Pulaski County Special School District and its Directors and Superintendent were voluntarily dismissed by the plaintiffs at the pre-trial conference held October 1, 1981.

The trial commenced December 7, 1981, and continued through December 17, 1981. This Memorandum Opinion constitutes the Court's findings of fact and conclusions of law. Further orders and judgment will be in conformity with this opinion.

I

There is no controversy over the legal standards under which the Establishment Clause portion of this case must be judged. The Supreme Court has on a number of occasions expounded on the meaning of the clause, and the pronouncements are clear. Often the issue has arisen in the context of public education, as it has here. In *Everson v. Board of Education*, 330 U.S. 1, 15–16 (1947), Justice Black stated:

> The "establishment of religion" clause of the First Amendment means at least this: Neither a state nor the Federal Government can set up a church. Neither can pass laws which aid one religion, aid all religions, or prefer one religion over another. Neither can force nor influence a person to go to or to remain away from church against his will or force him to profess a belief or disbelief in any religion. No person can be punished for entertaining or professing religious beliefs or disbeliefs, for church-attendance or non-attendance. No tax, large or small, can be levied to support any religious activities or institutions, whatever they may be called, or whatever form they may adopt to teach or practice religion. Neither a state nor the Federal Government can, openly or secretly, participate in the affairs of any religious organizations or groups and *vice versa*. In the words of Jefferson, the clause . . . was intended to erect "a wall of separation between church and State."

The Establishment Clause thus enshrines two central values: voluntarism and pluralism. And it is in the area of the public schools that these values must be guarded most vigilantly.

> Designed to serve as perhaps the most powerful agency for promoting cohesion among a heterogeneous democratic people, the public school must keep scrupulously free from entanglement in the strife of sects. The preservation of the community from divisive conflicts, of Government from irreconcilable pressures by religious groups, of religion from censorship and coercion however subtly exercised, requires strict confinement of the State to instruction other than religious, leaving to the individual's church and home, indoctrination in the faith of his choice. [*McCollum v. Board of Education*, 333 U.S. 203, 216–217 (1948). (Opinion of Frankfurter, J., joined by Jackson, Burton and Rutledge, J.J.)]

Overton,W.R.
"Judge Overton's Decision"
Science
Published by the A.A.A.S
Used by permission

The specific formulation of the establishment prohibition has been refined over the years, but its meaning has not varied from the principles articulated by Justice Black in *Everson*. In *Abbington School District v. Schempp*, 374 U.S. 203, 222 (1963), Justice Clark stated that "to withstand the strictures of the Establishment Clause there must be a secular legislative purpose and a primary effect that neither advances nor inhibits religion." The Court found it quite clear that the First Amendment does not permit a state to require the daily reading of the Bible in public schools, for "[s]urely the place of the Bible as an instrument of religion cannot be gainsaid." *Id.* at 224. Similarly, in *Engel v. Vitale*, 370 U.S. 421 (1962), the Court held that the First Amendment prohibited the New York Board of Regents from requiring the daily recitation of a certain prayer in the schools. With characteristic succinctness, Justice Black wrote, "Under [the First] Amendment's prohibition against governmental establishment of religion, as reinforced by the provisions of the Fourteenth Amendment, government in this country, be it state or federal, is without power to prescribe by law any particular form of prayer which is to be used as an official prayer in carrying on any program of governmentally sponsored religious activity." *Id.* at 430. Black also identified the objective at which the Establishment Clause was aimed: "Its first and most immediate purpose rested on the belief that a union of government and religion tends to destroy government and to degrade religion." *Id.* at 431.

Most recently, the Supreme Court has held that the clause prohibits a state from requiring the posting of the Ten Commandments in public school classrooms for the same reasons that officially imposed daily Bible reading is prohibited. *Stone v. Graham*, 449 U.S. 39 (1980). The opinion in *Stone* relies on the most recent formulation of the Establishment Clause test, that of *Lemon v. Kurtzman*, 403 U.S. 602, 612–613 (1971):

> First, the statute must have a secular legislative purpose; second, its principal or primary effect must be one that neither advances nor inhibits religion . . .; finally, the statute must not foster "an excessive government entanglement with religion." [*Stone v. Graham*, 449 U.S. at 40]

It is under this three part test that the evidence in this case must be judged. Failure on any of these grounds is fatal to the enactment.

II

The religious movement known as Fundamentalism began in nineteenth century America as part of evangelical Protestantism's response to social changes, new religious thought and Darwinism. Fundamentalists viewed these developments as attacks on the Bible and as responsible for a decline in traditional values.

The various manifestations of Fundamentalism have had a number of common characteristics (4), but a central premise has always been a literal interpretation of the Bible and a belief in the inerrancy of the Scriptures. Following World War I, there was again a perceived decline in traditional morality, and Fundamentalism focused on evolution as responsible for the decline. One aspect of their efforts, particularly in the South, was the promotion of statutes prohibiting the teaching of evolution in public schools. In Arkansas, this resulted in the adoption of Initiated Act 1 of 1929 (5).

Between the 1920's and early 1960's, anti-evolutionary sentiment had a subtle but pervasive influence on the teaching of biology in public schools. Generally, textbooks avoided the topic of evolution and did not mention the name of Darwin. Following the launch of the Sputnik satellite by the Soviet Union in 1957, the National Science Foundation funded several programs designed to modernize the teaching of science in the nation's schools. The Biological Sciences Curriculum Study (BSCS), a nonprofit organization, was among those receiving grants for curriculum study and revision. Working with scientists and teachers, BSCS developed a series of biology texts which, although emphasizing different aspects of biology, incorporated the theory of evolution as a major theme. The success of the BSCS effort is shown by the fact that fifty percent of American school children currently use BSCS books directly and the curriculum is incorporated indirectly in virtually all biology texts. (Testimony of Mayer; Nelkin, Px 1) (6).

In the early 1960's, there was again a resurgence of concern among Fundamentalists about the loss of traditional values and a fear of growing secularism in society. The Fundamentalist movement became more active and has steadily grown in numbers and political influence. There is an emphasis among current Fundamentalists on the literal interpretation of the Bible and the Book of Genesis as the sole source of knowledge about origins.

The term "scientific creationism" first gained currency around 1965 following publication of *The Genesis Flood* in 1961 by Whitcomb and Morris. There is undoubtedly some connection between the appearance of the BSCS texts emphasizing evolutionary thought and efforts by Fundamentalists to attack the theory. (Mayer)

In the 1960's and early 1970's, several Fundamentalist organizations were formed to promote the idea that the Book of Genesis was supported by scientific data. The terms "creation science" and "scientific creationism" have been adopted by these Fundamentalists as descriptive of their study of creation and the origins of man. Perhaps the leading creationist organization is the Institute for Creation Research (ICR), which is affiliated with the Christian Heritage College and supported by the Scott Memorial Baptist Church in San Diego, California. The ICR, through the Creation-Life Publishing Company, is the leading publisher of creation science material. Other creation science organizations include the Creation Science Research Center (CSRC) of San Diego and the Bible Science Association of Minneapolis, Minnesota. In 1963, the Creation Research Society (CRS) was formed from a schism in the American Scientific Affiliation (ASA). It is an organization of literal Fundamentalists (7) who have the equivalent of a master's degree in some recognized area of science. A purpose of the organization is "to reach all people with the vital message of the scientific and historic truth about creation." Nelkin, *The Science Textbook Controversies and the Politics of Equal Time*, 66. Similarly, the CSRC was formed in 1970 from a split in the CRS. Its aim has been "to reach the 63 million children of the United States with the scientific teaching of Biblical creationism." *Id.* at 69.

Among creationist writers who are recognized as authorities in the field by other creationists are Henry M. Morris, Duane Gish, G. E. Parker, Harold S. Slusher, Richard B. Bliss, John W. Moore, Martin E. Clark, W. L. Wysong, Robert E. Kofahl and Kelly L. Segraves. Morris is Director of ICR, Gish is Associate Director and Segraves is associated with CSRC.

Creationists view evolution as a source of society's ills, and the writings of Morris and Clark are typical expressions of that view.

> Evolution is thus not only anti-Biblical and anti-Christian, but it is utterly unscientific and impossible as well. But it has served effectively as the pseudo-scientific basis of atheism, agnosticism, socialism, fascism, and numerous other false and dangerous philosophies over the past century. [Morris and Clark, *The Bible Has The Answer*, (Px 31 and Pretrial Px 89) (8)]

Creationists have adopted the view of Fundamentalists generally that there are only two positions with respect to the origins of the earth and life: belief in the inerrancy of the Genesis story of creation and of a worldwide flood as fact, or belief in what they call evolution.

Henry Morris has stated, "It is impossible to devise a legitimate means of harmonizing the Bible with evolution." Morris, "Evolution and the Bible," *ICR Impact Series* Number 5 (undated, unpaged), quoted in Mayer, Px 8, at 3. This dualistic approach to the subject of origins permeates the creationist literature.

The creationist organizations consider the introduction of creation science into the public schools part of their ministry. The ICR has published at least two pamphlets (*9*) containing suggested methods for convincing school boards, administrators and teachers that creationism should be taught in public schools. The ICR has urged its proponents to encourage school officials to voluntarily add creationism to the curriculum (*10*).

Citizens For Fairness In Education is an organization based in Anderson, South Carolina, formed by Paul Ellwanger, a respiratory therapist who is trained in neither law nor science. Mr. Ellwanger is of the opinion that evolution is the forerunner of many social ills, including Nazism, racism and abortion. (Ellwanger Depo. at 32–34). About 1977, Ellwanger collected several proposed legislative acts with the idea of preparing a model state act requiring the teaching of creationism as science in opposition to evolution. One of the proposals he collected was prepared by Wendell Bird, who is now a staff attorney for ICR (*11*). From these various proposals, Ellwanger prepared a "model act" which calls for "balanced treatment" of "scientific creationism" and "evolution" in public schools. He circulated the proposed act to various people and organizations around the country.

Mr. Ellwanger's views on the nature of creation science are entitled to some weight since he personally drafted the model act which became Act 590. His evidentiary deposition with exhibits and unnumbered attachments (produced in response to a subpoena *duces tecum*) speaks to both the intent of the Act and the scientific merits of creation science. Mr. Ellwanger does not believe creation science is a science. In a letter to Pastor Robert E. Hays he states, "While neither evolution nor creation can qualify as a scientific theory, and since it is virtually impossible at this point to educate the whole world that evolution is not a true scientific theory, we have freely used these terms—the evolution theory and the theory of scientific creationism—in the bill's text." (Unnumbered attachment to Ellwanger Depo., at 2.) He further states in a letter to Mr. Tom Bethell, "As we examine evolution (remember, we're not making any scientific claims for creation, but we are challenging evolution's claim to be scientific) . . ." (Unnumbered attachment to Ellwanger Depo. at 1.)

Ellwanger's correspondence on the subject shows an awareness that Act 590 is a religious crusade, coupled with a desire to conceal this fact. In a letter to State Senator Bill Keith of Louisiana, he says, "I view this whole battle as one between God and anti-God forces, though I know there are a large number of evolutionists who believe in God." And further, ". . . it behooves Satan to do all he can to thwart our efforts and confuse the issue at every turn." Yet Ellwanger suggests to Senator Keith, "If you have a clear choice between having grassroots leaders of this statewide bill promotion effort to be ministerial or non-ministerial, be sure to opt for the non-ministerial. It does the bill effort no good to have ministers out there in the public forum and the adversary will surely pick at this point. . . . Ministerial persons can accomplish a tremendous amount of work from behind the scenes, encouraging their congregations to take the organizational and P.R. initiatives. And they can lead their churches in storming Heaven with prayers for help against so tenacious an adversary." (Unnumbered attachment to Ellwanger Depo. at 1.)

Ellwanger shows a remarkable degree of political candor, if not finesse, in a letter to State Senator Joseph Carlucci of Florida:

> 2. It would be very wise, if not actually essential, that all of us who are engaged in this legislative effort be careful not to present our position and our work in a religious framework. For example, in written communications that might somehow be shared with those other persons whom we may be trying to convince, it would be well to exclude our own personal testimony and/or witness for Christ, but rather, if we are so moved, to give that testimony on a separate attached note. (Unnumbered attachment to Ellwanger Depo. at 1.)

The same tenor is reflected in a letter by Ellwanger to Mary Ann Miller, a member of FLAG (Family, Life, America under God) who lobbied the Arkansas Legislature in favor of Act 590:

> . . . we'd like to suggest that you and your co-workers be very cautious about mixing creation-science with creation-religion . . . Please urge your co-workers not to allow themselves to get sucked into the "religion" trap of mixing the two together, for such mixing does incalculable harm to the legislative thrust. It could even bring public opinion to bear adversely upon the higher courts that will eventually have to pass judgment on the constitutionality of this new law. (Ex. 1 to Miller Depo.)

Perhaps most interesting, however, is Mr. Ellwanger's testimony in his deposition as to his strategy for having the model act implemented:

> Q. You're trying to play on other people's religious motives.
>
> A. I'm trying to play on their emotions, love, hate, their likes, dislikes, because I don't know any other way to involve, to get humans to become involved in human endeavors. I see emotions as being a healthy and legitimate means of getting people's feelings into action, and . . . I believe that the predominance of population in America that represents the greatest potential for taking some kind of action in this area is a Christian community. I see the Jewish community as far less potential in taking action . . . but I've seen a lot of interest among Christians and I feel, why not exploit that to get the bill going if that's what it takes. (Ellwanger Depo. at 146–147.)

Mr. Ellwanger's ultimate purpose is revealed in the closing of his letter to Mr. Tom Bethell: "Perhaps all this is old hat to you, Tom, and if so, I'd appreciate your telling me so and perhaps where you've heard it before—the idea of killing evolution instead of playing these debating games that we've been playing for nigh over a decade already." (Unnumbered attachment to Ellwanger Depo. at 3.)

It was out of this milieu that Act 590 emerged. The Reverend W. A. Blount, a Biblical literalist who is pastor of a church in the Little Rock area and was, in February, 1981, chairman of the Greater Little Rock Evangelical Fellowship, was among those who received a copy of the model act from Ellwanger (*12*).

At Reverend Blount's request, the Evangelical Fellowship unanimously adopted a resolution to seek introduction of Ellwanger's act in the Arkansas Legislature. A committee composed of two ministers, Curtis Thomas and W. A. Young, was appointed to implement the resolution. Thomas obtained from Ellwanger a revised copy of the model act which he transmitted to Carl Hunt, a business associate of Senator

James L. Holsted, with the request that Hunt prevail upon Holsted to introduce the act.

Holsted, a self-described "born again" Christian Fundamentalist, introduced the act in the Arkansas Senate. He did not consult the State Department of Education, scientists, science educators or the Arkansas Attorney General (*13*). The Act was not referred to any Senate committee for hearing and was passed after only a few minutes' discussion on the Senate floor. In the House of Representatives, the bill was referred to the Education Committee which conducted a perfunctory fifteen minute hearing. No scientist testified at the hearing, nor was any representative from the State Department of Education called to testify.

Ellwanger's model act was enacted into law in Arkansas as Act 590 without amendment or modification other than minor typographical changes. The legislative "findings of fact" in Ellwanger's act and Act 590 are identical, although no meaningful fact-finding process was employed by the General Assembly.

Ellwanger's efforts in preparation of the model act and campaign for its adoption in the states were motivated by his opposition to the theory of evolution and his desire to see the Biblical version of creation taught in the public schools. There is no evidence that the pastors, Blount, Thomas, Young or The Greater Little Rock Evangelical Fellowship were motivated by anything other than their religious convictions when proposing its adoption or during their lobbying efforts in its behalf. Senator Holsted's sponsorship and lobbying efforts in behalf of the Act were motivated solely by his religious beliefs and desire to see the Biblical version of creation taught in the public schools (*14*).

The State of Arkansas, like a number of states whose citizens have relatively homogeneous religious beliefs, has a long history of official opposition to evolution which is motivated by adherence to Fundamentalist beliefs in the inerrancy of the Book of Genesis. This history is documented in Justice Fortas' opinion in *Epperson v. Arkansas*, 393 U.S. 97 (1968), which struck down Initiated Act 1 of 1929, Ark. Stat. Ann. §§80-1627–1628, prohibiting the teaching of the theory of evolution. To this same tradition may be attributed Initiated Act 1 of 1930, Ark. Stat. Ann. §80-1606 (Repl. 1980), requiring "the reverent daily reading of a portion of the English Bible" in every public school classroom in the State (*15*).

It is true, as defendants argue, that courts should look to legislative statements of a statute's purpose in Establishment Clause cases and accord such pronouncements great deference. See, e.g., *Committee for Public Education & Religious Liberty v. Nyquist*, 413 U.S. 756, 773 (1973) and *McGowan v. Maryland*, 366 U.S. 420, 445 (1961). Defendants also correctly state the principle that remarks by the sponsor or author of a bill are not considered controlling in analyzing legislative intent. See, e.g., *United States v. Emmons*, 410 U.S. 396 (1973) and *Chrysler Corp. v. Brown*, 441 U.S. 281 (1979).

Courts are not bound, however, by legislative statements of purpose or legislative disclaimers. *Stone v. Graham*, 449 U.S. 39 (1980); *Abbington School Dist. v. Schempp*, 374 U.S. 203 (1963). In determining the legislative purpose of a statute, courts may consider evidence of the historical context of the Act, *Epperson v. Arkansas*, 393 U.S. 97 (1968), the specific sequence of events leading up to passage of the Act, departures from normal procedural sequences, substantive departures from the normal, *Village of Arlington Heights v. Metropolitan Housing Corp.*, 429 U.S. 252 (1977), and contemporaneous statements of the legislative sponsor, *Fed. Energy Admin. v. Algonquin SNG, Inc.*, 426 U.S. 548, 564 (1976).

The unusual circumstances surrounding the passage of Act 590, as well as the substantive law of the First Amendment, warrant an inquiry into the stated legislative purposes. The author of the Act had publicly proclaimed the sectarian purpose of the proposal. The Arkansas residents who sought legislative sponsorship of the bill did so for a purely sectarian purpose. These circumstances alone may not be particularly persuasive, but when considered with the publicly announced motives of the legislative sponsor made contemporaneously with the legislative process; the lack of any legislative investigation, debate or consultation with any educators or scientists; the unprecedented intrusion in school curriculum (*16*); and official history of the State of Arkansas on the subject, it is obvious that the statement of purposes has little, if any, support in fact. The State failed to produce any evidence which would warrant an inference or conclusion that at any point in the process anyone considered the legitimate educational value of the Act. It was simply and purely an effort to introduce the Biblical version of creation into the public school curricula. The only inference which can be drawn from these circumstances is that the Act was passed with the specific purpose by the General Assembly of advancing religion. The Act therefore fails the first prong of the three-pronged test, that of secular legislative purpose, as articulated in *Lemon v. Kurtzman*, *supra*, and *Stone v. Graham*, *supra*.

III

If the defendants are correct and the Court is limited to an examination of the language of the Act, the evidence is overwhelming that both the purpose and effect of Act 590 is the advancement of religion in the public schools.

Section 4 of the Act provides:

> Definitions, as used in this Act:
>
> (a) "Creation-science" means the scientific evidences for creation and inferences from those scientific evidences. Creation-science includes the scientific evidences and related inferences that indicate: (1) Sudden creation of the universe, energy, and life from nothing; (2) The insufficiency of mutation and natural selection in bringing about development of all living kinds from a single organism; (3) Changes only within fixed limits of originally created kinds of plants and animals; (4) Separate ancestry for man and apes; (5) Explanation of the earth's geology by catastrophism, including the occurrence of a worldwide flood; and (6) A relatively recent inception of the earth and living kinds.
>
> (b) "Evolution-science" means the scientific evidences for evolution and inferences from those scientific evidences. Evolution-science includes the scientific evidences and related inferences that indicate: (1) Emergence by naturalistic processes of the universe from disordered matter and emergence of life from nonlife; (2) The sufficiency of mutation and natural selection in bringing about development of present living kinds from simple earlier kinds; (3) Emergence by mutation and natural selection of present living kinds from simple earlier kinds; (4) Emergence of man from a common ancestor with apes; (5) Explanation of the earth's geology and the evolutionary sequence by uniformitarianism; and (6) An inception several billion years ago of the earth and somewhat later of life.
>
> (c) "Public schools" mean public secondary and elementary schools.

The evidence establishes that the definition of "creation science" contained in 4(a) has as its unmentioned reference the first 11 chapters of the Book of Genesis. Among the many creation epics in human history, the account of sudden creation from nothing, or *creatio ex nihilo*, and subsequent destruction of the world by flood is unique to Genesis. The concepts of 4(a) are the literal Fundamentalists' view of Genesis. Section 4(a) is unquestionably a statement of religion, with the exception of 4(a)(2) which is a negative thrust aimed at what the creationists understand to be the theory of evolution (*17*).

Both the concepts and wording of Section 4(a) convey an inescapable religiosity. Section 4(a)(1) describes "sudden creation of the universe, energy and life from nothing." Every theologian who testified, including defense witnesses, expressed the opinion that the statement referred to a supernatural creation which was performed by God.

Defendants argue that: (1) the fact that 4(a) conveys ideas similar to the literal interpretation of Genesis does not make it conclusively a statement of religion; (2) that reference to a creation from nothing is not necessarily a religious concept since the Act only suggests a creator who has power, intelligence and a sense of design and not necessarily the attributes of love, compassion and justice (*18*); and (3) that simply teaching about the concept of a creator is not a religious exercise unless the student is required to make a commitment to the concept of a creator.

The evidence fully answers these arguments. The ideas of 4(a)(1) are not merely similar to the literal interpretation of Genesis: they are identical and parallel to no other story of creation (*19*).

The argument that creation from nothing in 4(a)(1) does not involve a supernatural deity has no evidentiary or rational support. To the contrary, "creation out of nothing" is a concept unique to Western religions. In traditional Western religious thought, the conception of a creator of the world is a conception of God. Indeed, creation of the world "out of nothing" is the ultimate religious statement because God is the only actor. As Dr. Langdon Gilkey noted, the Act refers to one who has the power to bring all the universe into existence from nothing. The only "one" who has this power is God (*20*).

The leading creationist writers, Morris and Gish, acknowledge that the idea of creation described in 4(a)(1) is the concept of creation by God and make no pretense to the contrary (*21*). The idea of sudden creation from nothing, or *creatio ex nihilo*, is an inherently religious concept. (Vawter, Gilkey, Geisler, Ayala, Blount, Hicks.)

The argument advanced by defendants' witness, Dr. Norman Geisler, that teaching the existence of God is not religious unless the teaching seeks a commitment, is contrary to common understanding and contradicts settled case law. *Stone v. Graham*, 449 U.S. 39 (1980); *Abbington School District v. Schempp*, 374 U.S. 203 (1963).

The facts that creation science is inspired by the Book of Genesis and that Section 4(a) is consistent with a literal interpretation of Genesis leave no doubt that a major effect of the Act is the advancement of particular religious beliefs. The legal impact of this conclusion will be discussed further at the conclusion of the Court's evaluation of the scientific merit of creation science.

IV(A)

The approach to teaching "creation science" and "evolution science" found in Act 590 is identical to the two-model approach espoused by the Institute for Creation Research and is taken almost verbatim from ICR writings. It is an extension of Fundamentalists' view that one must either accept the literal interpretation of Genesis or else believe in the godless system of evolution.

The two model approach of the creationists is simply a contrived dualism (*22*) which has no scientific factual basis or legitimate educational purpose. It assumes only two explanations for the origins of life and existence of man, plants and animals: It was either the work of a creator or it was not. Application of these two models, according to creationists, and the defendants, dictates that all scientific evidence which fails to support the theory of evolution is necessarily scientific evidence in support of creationism and is, therefore, creation science "evidence" in support of Section 4(a).

IV(B)

The emphasis on origins as an aspect of the theory of evolution is peculiar to creationist literature. Although the subject of origins of life is within the province of biology, the scientific community does not consider origins of life a part of evolutionary theory. The theory of evolution assumes the existence of life and is directed to an explanation of *how* life evolved. Evolution does not presuppose the absence of a creator or God and the plain inference conveyed by Section 4 is erroneous (*23*).

As a statement of the theory of evolution, Section 4(b) is simply a hodgepodge of limited assertions, many of which are factually inaccurate.

For example, although 4(b)(2) asserts, as a tenet of evolutionary theory, "the sufficiency of mutation and natural selection in bringing about the existence of present living kinds from simple earlier kinds," Drs. Ayala and Gould both stated that biologists know that these two processes do not account for all significant evolutionary change. They testified to such phenomena as recombination, the founder effect, genetic drift and the theory of punctuated equilibrium, which are believed to play important evolutionary roles. Section 4(b) omits any reference to these. Moreover, 4(b) utilizes the term "kinds" which all scientists said is not a word of science and has no fixed meaning. Additionally, the Act presents both evolution and creation science as "package deals." Thus, evidence critical of some aspect of what the creationists define as evolution is taken as support for a theory which includes a worldwide flood and a relatively young earth (*24*).

IV(C)

In addition to the fallacious pedagogy of the two model approach, Section 4(a) lacks legitimate educational value because "creation science" as defined in that section is simply not science. Several witnesses suggested definitions of science. A descriptive definition was said to be that science is what is "accepted by the scientific community" and is "what scientists do." The obvious implication of this description is that, in a free society, knowledge does not require the imprimatur of legislation in order to become science.

More precisely, the essential characteristics of science are:

(1) It is guided by natural law;

(2) It has to be explanatory by reference to natural law;

(3) It is testable against the empirical world;

(4) Its conclusions are tentative, i.e., are not necessarily the final word; and

(5) It is falsifiable. (Ruse and other science witnesses).

Creation science as described in Section 4(a) fails to meet these essential characteristics. First, the section revolves around 4(a)(1) which asserts a sudden creation "from nothing." Such a concept is not science because it depends upon a supernatural intervention which is not guided by natural law.

It is not explanatory by reference to natural law, is not testable and is not falsifiable (*25*).

If the unifying idea of supernatural creation by God is removed from Section 4, the remaining parts of the section explain nothing and are meaningless assertions.

Section 4(a)(2), relating to the "insufficiency of mutation and natural selection in bringing about development of all living kinds from a single organism," is an incomplete negative generalization directed at the theory of evolution.

Section 4(a)(3) which describes "changes only within fixed limits of originally created kinds of plants and animals" fails to conform to the essential characteristics of science for several reasons. First, there is no scientific definition of "kinds" and none of the witnesses was able to point to any scientific authority which recognized the term or knew how many "kinds" existed. One defense witness suggested there may be 100 to 10,000 different "kinds." Another believes there were "about 10,000, give or take a few thousand." Second, the assertion appears to be an effort to establish outer limits of changes within species. There is no scientific explanation for these limits which is guided by natural law and the limitations, whatever they are, cannot be explained by natural law.

The statement in 4(a)(4) of "separate ancestry of man and apes" is a bald assertion. It explains nothing and refers to no scientific fact or theory (*26*).

Section 4(a)(5) refers to "explanation of the earth's geology by catastrophism, including the occurrence of a worldwide flood." This assertion completely fails as science. The Act is referring to the Noachian flood described in the Book of Genesis (*27*). The creationist writers concede that *any* kind of Genesis Flood depends upon supernatural intervention. A worldwide flood as an explanation of the world's geology is not the product of natural law, nor can its occurrence be explained by natural law.

Section 4(a)(6) equally fails to meet the standards of science. "Relatively recent inception" has no scientific meaning. It can only be given meaning by reference to creationist writings which place the age at between 6,000 and 20,000 years because of the genealogy of the Old Testament. See, e.g., Px 78, Gish (6,000 to 10,000); Px 87, Segraves (6,000 to 20,000). Such a reasoning process is not the product of natural law; not explainable by natural law; nor is it tentative.

Creation science, as defined in Section 4(a), not only fails to follow the canons defining scientific theory, it also fails to fit the more general descriptions of "what scientists think" and "what scientists do." The scientific community consists of individuals and groups, nationally and internationally, who work independently in such varied fields as biology, paleontology, geology and astronomy. Their work is published and subject to review and testing by their peers. The journals for publication are both numerous and varied. There is, however, not one recognized scientific journal which has published an article espousing the creation science theory described in Section 4(a). Some of the State's witnesses suggested that the scientific community was "close-minded" on the subject of creationism and that explained the lack of acceptance of the creation science arguments. Yet no witness produced a scientific article for which publication had been refused. Perhaps some members of the scientific community are resistant to new ideas. It is, however, inconceivable that such a loose knit group of independent thinkers in all the varied fields of science could, or would, so effectively censor new scientific thought.

The creationists have difficulty maintaining among their ranks consistency in the claim that creationism is science. The author of Act 590, Ellwanger, said that neither evolution nor creationism was science. He thinks both are religion. Duane Gish recently responded to an article in *Discover* critical of creationism by stating:

> Stephen Jay Gould states that creationists claim creation is a scientific theory. This is a false accusation. Creationists have repeatedly stated that neither creation nor evolution is a scientific theory (and each is equally religious). (Gish, letter to editor of *Discover*, July, 1981, App. 30 to Plaintiffs' Pretrial Brief)

The methodology employed by creationists is another factor which is indicative that their work is not science. A scientific theory must be tentative and always subject to revision or abandonment in light of facts that are inconsistent with, or falsify, the theory. A theory that is by its own terms dogmatic, absolutist and never subject to revision is not a scientific theory.

The creationists' methods do not take data, weigh it against the opposing scientific data, and thereafter reach the conclusions stated in Section 4(a). Instead, they take the literal wording of the Book of Genesis and attempt to find scientific support for it. The method is best explained in the language of Morris in his book (Px 31) *Studies in The Bible and Science* at page 114:

> . . . it is . . . quite impossible to determine anything about Creation through a study of present processes, because present processes are not creative in character. If man wishes to know anything about Creation (the time of Creation, the duration of Creation, the order of Creation, the methods of Creation, or anything else) his sole source of true information is that of divine revelation. God was there when it happened. We were not there. . . . Therefore, we are completely limited to what God has seen fit to tell us, and this information is in His written Word. This is our textbook on the science of Creation!

The Creation Research Society employs the same unscientific approach to the issue of creationism. Its applicants for membership must subscribe to the belief that the Book of Genesis is "historically and scientifically true in all of the original autographs" (*28*). The Court would never criticize or discredit any person's testimony based on his or her religious beliefs. While anybody is free to approach a scientific inquiry in any fashion they choose, they cannot properly describe the methodology used as scientific, if they start with a conclusion and refuse to change it regardless of the evidence developed during the course of the investigation.

IV(D)

In efforts to establish "evidence" in support of creation science, the defendants relied upon the same false premise as the two model approach contained in Section 4, i.e., all evidence which criticized evolutionary theory was proof in support of creation science. For example, the defendants established that the mathematical probability of a chance chemical combination resulting in life from non-life is so remote that such an occurrence is almost beyond imagination. Those mathematical facts, the defendants argue, are scientific evidences that life was the product of a creator. While the statistical figures may be impressive evidence against the theory of chance chemical combinations as an explanation of origins, it requires a leap of faith to interpret those figures so as to support a complex doctrine which includes a sudden creation from nothing, a worldwide flood, separate ancestry of man and apes, and a young earth.

The defendants' argument would be more persuasive if, in fact, there were only two theories or ideas about the origins of life and the world. That there are a number of theories was acknowledged by the State's witnesses, Dr. Wickramasinghe and Dr. Geisler. Dr. Wickramasinghe testified at length in support of a theory that life on earth was "seeded" by comets which delivered genetic material and perhaps organisms to the earth's surface from interstellar dust far outside the solar system. The "seeding" theory further hypothesizes that the earth remains under the continuing influence of genetic material from space which continues to affect life. While Wickramasinghe's theory (*29*) about the origins of life on earth has not received general acceptance within the scientific community, he has, at least, used scientific methodology to produce a theory of origins which meets the essential characteristics of science.

The Court is at a loss to understand why Dr. Wickramasinghe was called in behalf of the defendants. Perhaps it was because he was generally critical of the theory of evolution and the scientific community, a tactic consistent with the strategy of the defense. Unfortunately for the defense, he demonstrated that the simplistic approach of the two model analysis of the origins of life is false. Furthermore, he corroborated the plaintiffs' witnesses by concluding that "no rational scientist" would believe the earth's geology could be explained by reference to a worldwide flood or that the earth was less than one million years old.

The proof in support of creation science consisted almost entirely of efforts to discredit the theory of evolution through a rehash of data and theories which have been before the scientific community for decades. The arguments asserted by creationists are not based upon new scientific evidence or laboratory data which has been ignored by the scientific community.

Robert Gentry's discovery of radioactive polonium haloes in granite and coalified woods is, perhaps, the most recent scientific work which the creationists use as argument for a "relatively recent inception" of the earth and a "worldwide flood." The existence of polonium haloes in granite and coalified wood is thought to be inconsistent with radiometric dating methods based upon constant radioactive decay rates. Mr. Gentry's findings were published almost ten years ago and have been the subject of some discussion in the scientific community. The discoveries have not, however, led to the formulation of any scientific hypothesis or theory which would explain a relatively recent inception of the earth or a worldwide flood. Gentry's discovery has been treated as a minor mystery which will eventually be explained. It may deserve further investigation, but the National Science Foundation has not deemed it to be of sufficient import to support further funding.

The testimony of Marianne Wilson was persuasive evidence that creation science is not science. Ms. Wilson is in charge of the science curriculum for Pulaski County Special School District, the largest school district in the State of Arkansas. Prior to the passage of Act 590, Larry Fisher, a science teacher in the District, using materials from the ICR, convinced the School Board that it should voluntarily adopt creation science as part of its science curriculum. The District Superintendent assigned Ms. Wilson the job of producing a creation science curriculum guide. Ms. Wilson's testimony about the project was particularly convincing because she obviously approached the assignment with an open mind and no preconceived notions about the subject. She had not heard of creation science until about a year ago and did not know its meaning before she began her research.

Ms. Wilson worked with a committee of science teachers appointed from the District. They reviewed practically all of the creationist literature. Ms. Wilson and the committee members reached the unanimous conclusion that creationism is not science; it is religion. They so reported to the Board. The Board ignored the recommendation and insisted that a curriculum guide be prepared.

In researching the subject, Ms. Wilson sought the assistance of Mr. Fisher who initiated the Board action and asked professors in the science departments of the University of Arkansas at Little Rock and the University of Central Arkansas (*30*) for reference material and assistance, and attended a workshop conducted at Central Baptist College by Dr. Richard Bliss of the ICR staff. Act 590 became law during the course of her work so she used Section 4(a) as a format for her curriculum guide.

Ms. Wilson found all available creationists' materials unacceptable because they were permeated with religious references and reliance upon religious beliefs.

It is easy to understand why Ms. Wilson and other educators find the creationists' textbook material and teaching guides unacceptable. The materials misstate the theory of evolution in the same fashion as Section 4(b) of the Act, with emphasis on the alternative mutually exclusive nature of creationism and evolution. Students are constantly encouraged to compare and make a choice between the two models, and the material is not presented in an accurate manner.

A typical example is *Origins* (Px 76) by Richard B. Bliss, Director of Curriculum Development of the ICR. The presentation begins with a chart describing "preconceived ideas about origins" which suggests that some people believe that evolution is atheistic. Concepts of evolution, such as "adaptive radiation" are erroneously presented. At page 11, figure 1.6, of the text, a chart purports to illustrate this "very important" part of the evolution model. The chart conveys the idea that such diverse mammals as a whale, bear, bat and monkey all evolved from a shrew through the process of adaptive radiation. Such a suggestion is, of course, a totally erroneous and misleading application of the theory. Even more objectionable, especially when viewed in light of the emphasis on asking the student to elect one of the models, is the chart presentation at page 17, figure 1.6. That chart purports to illustrate the evolutionists' belief that man evolved from bacteria to fish to reptile to mammals and, thereafter, into man. The illustration indicates, however, that the mammal from which man evolved was *a rat*.

Biology, A Search For Order in Complexity (*31*) is a high school biology text typical of creationists' materials. The following quotations are illustrative:

> Flowers and roots do not have a mind to have purpose of their own; therefore, this planning must have been done for them by the Creator. (at page 12)

> The exquisite beauty of color and shape in flowers exceeds the skill of poet, artist, and king. Jesus said (from Matthew's gospel), "Consider the lilies of the field, how they grow; they toil not, neither do they spin . . ." (Px 129 at page 363)

The "public school edition" texts written by creationists simply omit Biblical references but the content and message remain the same. For example, *Evolution—The Fossils Say No!* (*32*) contains the following:

> Creation. By creation we mean the bringing into being by a supernatural Creator of the basic kinds of plants and animals by the process of sudden, or fiat, creation.
>
> We do not know how the Creator created, what processes He used, *for He used processes which are not now operating anywhere in the natural universe*. This is why we refer to creation as Special Creation. We cannot discover by scientific investigation anything about the creative processes used by the Creator. (page 40)

Gish's book also portrays the large majority of evolutionists as "materialistic atheists or agnostics."

Scientific Creationism (Public School Edition) by Morris, is another text reviewed by Ms. Wilson's committee and rejected as unacceptable. The following quotes illustrate the purpose and theme of the text:

> Foreword
>
> Parents and youth leaders today, and even many scientists and educators, have become concerned about the prevalence and influence of evolutionary philosophy in modern curriculum. Not only is this system inimical to orthodox Christianity and Judaism, but also, as many are convinced, to a healthy society and true science as well. (at page iii)
>
> The rationalist of course finds the concept of special creation insufferably naive, even "incredible". Such a judgment, however, is warranted only if one categorically dismisses the existence of an omnipotent God. (at page 17)

Without using creationist literature, Ms. Wilson was unable to locate one genuinely scientific article or work which supported Section 4(a). In order to comply with the mandate of the Board she used such materials as an article from *Readers Digest* about "atomic clocks" which inferentially suggested that the earth was less than 4½ billion years old. She was unable to locate any substantive teaching material for some parts of Section 4 such as the worldwide flood. The curriculum guide which she prepared cannot be taught and has no educational value as science. The defendants did not produce any text or writing in response to this evidence which they claimed was usable in the public school classroom (*33*).

The conclusion that creation science has no scientific merit or educational value as science has legal significance in light of the Court's previous conclusion that creation science has, as one major effect, the advancement of religion. The second part of the three-pronged test for establishment reaches only those statutes having as their *primary* effect the advancement of religion. Secondary effects which advance religion are not constitutionally fatal. Since creation science is not science, the conclusion is inescapable that the *only* real effect of Act 590 is the advancement of religion. The Act therefore fails both the first and second portions of the test in *Lemon v. Kurtzman*, 403 U.S. 602 (1971).

IV(E)

Act 590 mandates "balanced treatment" for creation science and evolution science. The Act prohibits instruction in any religious doctrine or references to religious writings. The Act is self-contradictory and compliance is impossible unless the public schools elect to forego significant portions of subjects such as biology, world history, geology, zoology, botany, psychology, anthropology, sociology, philosophy, physics and chemistry. Presently, the concepts of evolutionary theory as described in 4(b) permeate the public school textbooks. There is no way teachers can teach the Genesis account of creation in a secular manner.

The State Department of Education, through its textbook selection committee, school boards and school administrators will be required to constantly monitor materials to avoid using religious references. The school boards, administrators and teachers face an impossible task. How is the teacher to respond to questions about a creation suddenly and out of nothing? How will a teacher explain the occurrence of a worldwide flood? How will a teacher explain the concept of a relatively recent age of the earth? The answer is obvious because the only source of this information is ultimately contained in the Book of Genesis.

References to the pervasive nature of religious concepts in creation science texts amply demonstrate why State entanglement with religion is inevitable under Act 590. Involvement of the State in screening texts for impermissible religious references will require State officials to make delicate religious judgments. The need to monitor classroom discussion in order to uphold the Act's prohibition against religious instruction will necessarily involve administrators in questions concerning religion. These continuing involvements of State officials in questions and issues of religion create an excessive and prohibited entanglement with religion. *Brandon v. Board of Education*, 487 F.Supp 1219, 1230 (N.D.N.Y.), *aff'd.*, 635 F.2d 971 (2nd Cir. 1980).

V

These conclusions are dispositive of the case and there is no need to reach legal conclusions with respect to the remaining issues. The plaintiffs raised two other issues questioning the constitutionality of the Act and, insofar as the factual findings relevant to these issues are not covered in the preceding discussion, the Court will address these issues. Additionally, the defendants raised two other issues which warrant discussion.

V(A)

First, plaintiff teachers argue the Act is unconstitutionally vague to the extent that they cannot comply with its mandate of "balanced" treatment without jeopardizing their employment. The argument centers around the lack of a precise definition in the Act for the word "balanced." Several witnesses expressed opinions that the word has such meanings as equal time, equal weight, or equal legitimacy. Although the Act could have been more explicit, "balanced" is a word subject to ordinary understanding. The proof is not convincing that a teacher using a reasonably acceptable understanding of the word and making a good faith effort to comply with the Act will be in jeopardy of termination. Other portions of the Act are arguably vague, such as the "relatively recent" inception of the earth and life. The evidence establishes, however, that relatively recent means from 6,000 to 20,000 years, as commonly understood in creation science literature. The meaning of this phrase, like Section 4(a) generally, is, for purposes of the Establishment Clause, all too clear.

V(B)

The plaintiffs' other argument revolves around the alleged infringement by the defendants upon the academic freedom of teachers and students. It is contended this unprecedented intrusion in the curriculum by the State prohibits teachers from teaching what they believe should be taught or requires them to teach that which they do not believe is proper. The evidence reflects that traditionally the State Department of Education, local school boards and administration officials exercise little, if any, influence upon the subject matter taught by classroom teachers. Teachers have been given freedom to teach and emphasize those portions of subjects the individual teacher considered important. The limits to this discretion have generally been derived from the approval of textbooks by the State Department and preparation of curriculum guides by the school districts.

Several witnesses testified that academic freedom for the teacher means, in substance, that the individual teacher should be permitted unlimited discretion subject only to the bounds of professional ethics. The Court is not prepared to adopt such a broad view of academic freedom in the public schools.

In any event, if Act 590 is implemented, many teachers will be required to teach material in support of creation science which they do not consider academically sound. Many teachers will simply forego teaching subjects which might trigger the "balanced treatment" aspects of Act 590 even though they think the subjects are important to a proper presentation of a course.

Implementation of Act 590 will have serious and untoward consequences for students, particularly those planning to attend college. Evolution is the cornerstone of modern biology, and many courses in public schools contain subject matter relating to such varied topics as the age of the earth, geology and relationships among living things. Any student who is deprived of instruction as to the prevailing scientific thought on these topics will be denied a significant part of science education. Such a deprivation through the high school level would undoubtedly have an impact upon the quality of education in the State's colleges and universities, especially including the pre-professional and professional programs in the health sciences.

V(C)

The defendants argue in their brief that evolution is, in effect, a religion, and that by teaching a religion which is contrary to some students' religious views, the State is infringing upon the student's free exercise rights under the First Amendment. Mr. Ellwanger's legislative findings, which were adopted as a finding of fact by the Arkansas Legislature in Act 590, provides:

> Evolution-science is contrary to the religious convictions or moral values or philosophical beliefs of many students and parents, including individuals of many different religious faiths and with diverse moral and philosophical beliefs. Act 590, §7(d).

The defendants argue that the teaching of evolution alone presents both a free exercise problem and an establishment problem which can only be redressed by giving balanced treatment to creation science, which is admittedly consistent with some religious beliefs. This argument appears to have its genesis in a student note written by Mr. Wendell Bird, "Freedom of Religion and Science Instruction in Public Schools," 87 Yale L.J. 515 (1978). The argument has no legal merit.

If creation science is, in fact, science and not religion, as the defendants claim, it is difficult to see how the teaching of such a science could "neutralize" the religious nature of evolution.

Assuming for the purposes of argument, however, that evolution is a religion or religious tenet, the remedy is to stop the teaching of evolution, not establish another religion in opposition to it. Yet it is clearly established in the case law, and perhaps also in common sense, that evolution is not a religion and that teaching evolution does not violate the Establishment Clause, *Epperson v. Arkansas, supra, Willoughby v. Stever*, No. 15574-75 (D.D.C. May 18, 1973); *aff'd.* 504 F.2d 271 (D.C. Cir. 1974), *cert. denied*, 420 U.S. 924 (1975); *Wright v. Houston Indep. School Dist.*, 366 F.Supp. 1208 (S.D. Tex. 1978), *aff'd*. 486 F.2d 137 (5th Cir. 1973), *cert. denied* 417 U.S. 969 (1974).

V(D)

The defendants presented Dr. Larry Parker, a specialist in devising curricula for public schools. He testified that the public school's curriculum should reflect the subjects the public wants taught in schools. The witness said that polls indicated a significant majority of the American public thought creation science should be taught if evolution was taught. The point of this testimony was never placed in a legal context. No doubt a sizeable majority of Americans believe in the concept of a Creator or, at least, are not opposed to the concept and see nothing wrong with teaching school children about the idea.

The application and content of First Amendment principles are not determined by public opinion polls or by a majority vote. Whether the proponents of Act 590 constitute the majority or the minority is quite irrelevant under a constitutional system of government. No group, no matter how large or small, may use the organs of government, of which the public schools are the most conspicuous and influential, to foist its religious beliefs on others.

The Court closes this opinion with a thought expressed eloquently by the great Justice Frankfurter:

> We renew our conviction that "we have staked the very existence of our country on the faith that complete separation between the state and religion is best for the state and best for religion." *Everson v. Board of Education*, 330 U.S. at 59. If nowhere else, in the relation between Church and State, "good fences make good neighbors." [*McCollum v. Board of Education*, 333 U.S. 203, 232 (1948)]

An injunction will be entered permanently prohibiting enforcement of Act 590.

It is ordered this January 5, 1982.

—William R. Overton *in the U.S. District Court, Eastern District of Arkansas, Western Division*

Notes

1. The complaint is based on 42 U.S.C. §1983, which provides a remedy against any person who, acting under color of state law, deprives another of any right, privilege or immunity guaranteed by the United States Constitution or federal law. This Court's jurisdiction arises under 28 U.S.C. §§1331, 1343(3) and 1343(4). The power to issue declaratory judgments is expressed in 28 U.S.C. §§2201 and 2202.
2. The facts necessary to establish the plaintiffs' standing to sue are contained in the joint stipulation of facts, which is hereby adopted and incorporated herein by reference. There is no doubt that the case is ripe for adjudication.
3. The State of Arkansas was dismissed as a defendant because of its immunity from suit under the Eleventh Amendment. *Hans v. Louisiana*, 134 U.S. 1 (1890).
4. The authorities differ as to generalizations which may be made about Fundamentalism. For example, Dr. Geisler testified to the widely held view that there are five beliefs characteristic of all Fundamentalist movements, in addition, of course, to the inerrancy of Scripture: (1) belief in the virgin birth of Christ, (2) belief in the deity of Christ, (3) belief in the substitutional atonement of Christ, (4) belief in the second coming of Christ, and (5) belief in the physical resurrection of all departed souls. Dr. Marsden, however, testified that this generalization, which has been common in religious scholarship, is now thought to be historical error. There is no doubt, however, that all Fundamentalists take the Scriptures as inerrant and probably most take them as literally true.
5. Initiated Act 1 of 1929, Ark. Stat. Ann. §80-1627 *et seq.*, which prohibited the teaching of evolution in Arkansas schools, is discussed *infra* at text accompanying note 26.
6. Subsequent references to the testimony will be made by the last name of the witness only. References to documentary exhibits will be by the name of the author and the exhibit number.
7. Applicants for membership in the CRS must subscribe to the following statement of belief: "(1) The Bible is the written Word of God, and because we believe it to be inspired thruout (sic), all of its assertions are historically and scientifically true in all of the original autographs. To the student of nature, this means that the account of origins in Genesis is a factual presentation of simple historical truths. (2) All basic types of living things,

including man, were made by direct creative acts of God during Creation Week as described in Genesis. Whatever biological changes have occurred since Creation have accomplished only changes within the original created kinds. (3) The great Flood described in Genesis, commonly referred to as the Noachian Deluge, was an historical event, worldwide in its extent and effect. (4) Finally, we are an organization of Christian men of science, who accept Jesus Christ as our Lord and Savior. The account of the special creation of Adam and Eve as one man and one woman, and their subsequent Fall into sin, is the basis for our belief in the necessity of a Savior for all mankind. Therefore, salvation can come only thru (sic) accepting Jesus Christ as our Savior." (Px 115)

8. Because of the voluminous nature of the documentary exhibits, the parties were directed by pre-trial order to submit their proposed exhibits for the Court's convenience prior to trial. The numbers assigned to the pre-trial submissions do not correspond with those assigned to the same documents at trial and, in some instances, the pre-trial submissions are more complete.
9. Px 130, Morris, *Introducing Scientific Creationism Into the Public Schools* (1975), and Bird, "Resolution for Balanced Presentation of Evolution and Scientific Creationism," *ICR Impact Series* No. 71, App. 14 to Plaintiffs' Pretrial Brief.
10. The creationists often show candor in their proselytization. Henry Morris has stated: "Even if a favorable statute or court decision is obtained, it will probably be declared unconstitutional, especially if the legislation or injunction refers to the Bible account of creation." In the same vein he notes, "The only effective way to get creationism taught properly is to have it taught by teachers who are both willing and able to do it. Since most teachers now are neither willing nor able, they must first be both persuaded and instructed themselves." Px 130, Morris, *Introducing Scientific Creationism Into the Public Schools* (1975) (unpaged).
11. Mr. Bird sought to participate in this litigation by representing a number of individuals who wanted to intervene as defendants. The application for intervention was denied by this Court. *McLean v. Arkansas*, ——— F.Supp. ———, (E.D. Ark. 1981), aff'd. *per curiam*, Slip Op. No. 81-2023 (8th Cir. Oct. 16, 1981).
12. The model act had been revised to insert "creation science" in lieu of creationism because Ellwanger had the impression people thought creationism was too religious a term. (Ellwanger Depo. at 79.)
13. The original model act had been introduced in the South Carolina Legislature, but had died without action after the South Carolina Attorney General had opined that the act was unconstitutional.
14. Specifically, Senator Holsted testified that he holds to a literal interpretation of the Bible; that the bill was compatible with his religious beliefs; that the bill does favor the position of literalists; that his religious convictions were a factor in his sponsorship of the bill; and that he stated publicly to the *Arkansas Gazette* (although not on the floor of the Senate) contemporaneously with the legislative debate that the bill does presuppose the existence of a divine creator. There is no doubt that Senator Holsted knew he was sponsoring the teaching of a religious doctrine. His view was that the bill did not violate the First Amendment because, as he saw it, it did not favor one denomination over another.
15. This statute is, of course, clearly unconstitutional under the Supreme Court's decision in *Abbington School Dist. v. Schempp*, 374 U.S. 203 (1963).
16. The joint stipulation of facts establishes that the following areas are the only *information* specifically required by statute to be taught in all Arkansas schools: (1) the effects of alcohol and narcotics on the human body, (2) conservation of national resources, (3) Bird Week, (4) Fire Prevention, and (5) Flag etiquette. Additionally, certain specific courses, such as American history and Arkansas history, must be completed by each student before graduation from high school.
17. Paul Ellwanger stated in his deposition that he did not know why Section 4(a)(2) (insufficiency of mutation and natural selection) was included as an evidence supporting creation science. He indicated that he was not a scientist, "but these are the postulates that have been laid down by creation scientists." Ellwanger Depo. at 136.
18. Although defendants must make some effort to cast the concept of creation in non-religious terms, this effort surely causes discomfort to some of the Act's more theologically sophisticated supporters. The concept of a creator God distinct from the God of love and mercy is closely similar to the Marcion and Gnostic heresies, among the deadliest to threaten the early Christian church. These heresies had much to do with development and adoption of the Apostle's Creed as the official creedal statement of the Roman Catholic Church in the West. (Gilkey.)
19. The parallels between Section 4(a) and Genesis are quite specific: (1) "sudden creation from nothing" is taken from Genesis, 1:1–10 (Vawter, Gilkey); (2) destruction of the world by a flood of divine origin is a notion peculiar to Judeo-Christian tradition and is based on Chapters 7 and 8 of Genesis (Vawter); (3) the term "kinds" has no fixed scientific meaning, but appears repeatedly in Genesis (all scientific witnesses); (4) "relatively recent inception" means an age of the earth from 6,000 to 10,000 years and is based on the genealogy of the Old Testament using the rather astronomical ages assigned to the patriarchs (Gilkey and several of defendants' scientific witnesses); (5) Separate ancestry of man and ape focuses on the portion of the theory of evolution which Fundamentalists find most offensive, *Epperson v. Arkansas*, 393 U.S. 97 (1968).
20. "[C]oncepts concerning . . . a supreme being of some sort are manifestly religious . . . These concepts do not shed that religiosity merely because they are presented as philosophy or as a science . . ." *Malnak v. Yogi*, 440 F.Supp. 1284, 1322 (D.N.J. 1977); *aff'd per curiam*, 592 F.2d 197 (3d Cir. 1979).
21. See, e.g., Px 76, Morris, *et al.*, *Scientific Creationism*, 203 (1980) ("If creation really is a fact, this means there is a *Creator*, and the universe is His creation.") Numerous other examples of such admissions can be found in the many exhibits which represent creationist literature, but no useful purpose would be served here by a potentially endless listing.
22. Morris, the Director of ICR and one who first advocated the two model approach, insists that a true Christian cannot compromise with the theory of evolution and that the Genesis version of creation and the theory of evolution are mutually exclusive. Px 31, Morris, *Studies in the Bible & Science*, 102–103. The two model approach was the subject of Dr. Richard Bliss's doctoral dissertation. (Dx 35). It is presented in Bliss, *Origins: Two Models-Evolution, Creation* (1978). Moreover, the two model approach merely casts in educationalist language the dualism which appears in all creationist literature—creation (i.e. God) and evolution are presented as two alternative and mutually exclusive theories. See, e.g., Px 75, Morris, *Scientific Creationism* (1974) (public school edition); Px 59, Fox, *Fossils Hard Facts from the Earth*. Particularly illustrative is Px 61, Boardman, *et al.*, *Worlds Without End* (1971) a CSRC publication: "One group of scientists, known as creationists, believe that God, in a miraculous manner, created all matter and energy . . .

   "Scientists who insist that the universe just grew, by accident, from a mass of hot gases without the direction or help of a Creator are known as evolutionists."
23. The idea that belief in a creator and acceptance of the scientific theory of evolution are mutually exclusive is a false premise and offensive to the religious views of many. (Hicks) Dr. Francisco Ayala, a geneticist of considerable reknown and a former Catholic priest who has the equivalent of a Ph.D. in theology, pointed out that many working scientists who subscribed to the theory of evolution are devoutly religious.
24. This is so despite the fact that some of the defense witnesses do not subscribe to the young earth or flood hypotheses. Dr. Geisler stated his belief that the earth is several billion years old. Dr. Wickramasinghe stated that no rational scientist would believe the earth is less than one million years old or that all the world's geology could be explained by a worldwide flood.
25. "We do not know how God created, what processes He used, *for God used processes which are not now operating anywhere in the natural universe.* This is why we refer to divine creation as Special Creation. We cannot discover by scientific investigation anything about the creative processes used by God." Px 78, Gish, *Evolution? The Fossils Say No!*, 42 (3d ed. 1979) (emphasis in original).
26. The evolutionary notion that man and some modern apes have a common ancestor somewhere in the distant past has consistently been distorted by anti-evolutionists to say that man descended from modern monkeys. As such, this idea has long been most offensive to Fundamentalists. See, *Epperson v. Arkansas*, 393 U.S. 97 (1968).
27. Not only was this point acknowledged by virtually all the defense witnesses, it is patent in the creationist literature. See, e.g., Px 89, Kofahl & Segraves, *The Creation Explanation*, 40: "The Flood of Noah brought about vast changes in the earth's surface, including vulcanism, mountain building, and the deposition of the major part of sedimentary strata. This principle is called 'Biblical catastrophism.' "
28. See n. 7, *supra*, for the full text of the CRS creed.
29. The theory is detailed in Wickramasinghe's book with Sir Fred Hoyle, *Evolution From Space* (1981), which is Dx 79.
30. Ms. Wilson stated that some professors she spoke with sympathized with her plight and tried to help her find scientific materials to support Section 4(a). Others simply asked her to leave.
31. Px 129, published by Zunderman Publishing House (1974), states that it was "prepared by the Textbook Committee of the Creation Research Society." It has a disclaimer pasted inside the front cover stating that it is not suitable for use in public schools.
32. Px 77, by Duane Gish.
33. The passage of Act 590 apparently caught a number of its supporters off guard as much as it did the school district. The Act's author, Paul Ellwanger, stated in a letter to "Dick," (apparently Dr. Richard Bliss at ICR): "And finally, if you know of any textbooks at any level and for any subjects that you think are acceptable to you and also constitutionally admissible, these are things that would be of *enormous* to these bewildered folks who may be caught, as Arkansas now has been, by the sudden need to implement a whole new ball game with which they are quite unfamiliar." (sic) (Unnumbered attachment to Ellwanger depo.)

Selections from P.H. Gosse
Omphalos 1857

and, for the greater part of them, to the very existence of the infant.

Thus, once more, there is no condition of the animal, on which we may fix, as being so simple, as to have no retrospective history.

The umbilical cicatrix I have already alluded to; but I may be permitted to mention it again; because, in all the higher MAMMALIA, at least, it exists, throughout life, an eloquent witness to the organic connexion of the individual with a mother, and therefore to her pre-existence. If it were legitimate to suppose that the first individual of the species Man was created in the condition answering to that of a new-born infant, there would still be the need of maternal milk for its sustenance, and maternal care for its protection, for a considerable period; while, if we carry on the suggested stage to the period when this provision is no longer indispensable, the development of hair, nails, bones, &c., will have proceeded through many stages. And, in either condition, the navel cord or its cicatrix remains, to testify to something anterior to both.

XII.

THE CONCLUSION.

"We have no experience in the creation of worlds."
CHALMERS.

WE have passed in review before us the whole organic world: and the result is uniform; that no example can be selected from the vast vegetable kingdom, none from the vast animal kingdom, which did not at the instant of its creation present indubitable evidences of a previous history. This is not put forth as a *hypothesis*, but as a *necessity*; I do not say that it was *probably* so, but that it was *certainly* so; not that it *may have been thus*, but that it *could not have been otherwise*.

I do not touch the inorganic world: my acquaintance with chemistry is inadequate for this: perhaps the same law does not extend to the inorganic elements: perhaps their developments and combinations are not, like the economy of plants and animals, essentially and exclusively cyclical: perhaps carbon and oxygen and hydrogen could be created in conditions, which obviously did not depend on any previously existing conditions. This I do not know: I neither affirm nor deny it. But I think I have demonstrated in these pages, that such a cyclical character does attach to, and is inseparable from, the history of all organic essences; and that creation can be nothing else than a series of irruptions into circles: that, supposing the irruption to have been made at what part of the circle we please, and varying this condition indefinitely at will,—we cannot avoid the conclusion that each organism was from the first marked with the records of a previous being. But since creation and previous history are inconsistent with each other; as the very idea of the creation of an organism excludes the idea of pre-existence of that organism, or of any part of it; it follows, that such records are *false*, so far as they testify to time; that the developments and processes thus recorded have been produced without time, or are what I have called *prochronic*.

Nor is this conclusion in the least degree affected by the actual chronology of creation. The phenomena were equally eloquent, and equally false, whether any individual organism were created six

thousand years ago, or innumerable ages; whether primitively, or after the successive creations and annihilations of former organisms.

The law of creation supersedes the law of nature; so far, at least, as the organic world is concerned. The law of nature, established by universal experience, is, that its phenomena depend upon certain natural antecedents: the law of creation is, that the same phenomena depend upon *no* antecedents. The philosopher who should infer the antecedents from the phenomena alone, without having considered the law of creation, would be liable to form totally false conclusions. In order to be secure from error, he must first assure himself that creation is eliminated from the category of facts which he is investigating; and this he could do only when the facts come within the sphere of personal observation, or of historic testimony. Up to such a period of antiquity as is covered by credible history, and within such a field of observation as history may be considered fairly cognisant of,—the inference of physical antecedents from physical phenomena, in the animal or vegetable world, is legitimate and true. But, beyond that period, I cannot safely deduce the same conclusion; because I cannot tell but that at any given moment included in my inquiry, creation may have occurred, and have been the absolute beginning of the circular series.

The question of the actual age of any species, whether plant or animal, is one which cannot be answered, except on historic testimony. The sequence of cause and effect is not adequate to answer it; for a legitimate use of this principle, supposing it the only element of the inquiry, would inevitably lead us to the eternity of all existing organic life.

One of the familiar street-exhibitions in the metropolis is a tiny coach and horses of glittering metal; which, by means of simple machinery, course round and round the margin of a circular table. Let us suppose two youths of philosophical turn to come up during the process. They gaze for a while, and one asks his companion the following question.

"How long do you suppose that coach has been running round?"

"How long! for an indefinite period, for aught I know. I have counted twenty-two turns, and can see no change: nor can I suggest any point where the course could have begun."

Here a shrewd lad, carrying a grocer's basket, breaks in.

"Oh no; there have been only six-and-twenty turns altogether. Four turns had been made when you came up. The whole began by the man taking the carriage out of a box; then he set it down out there, just opposite to us, and gave it a little push with his finger, and it has been running ever since. I saw him do it."

Now perhaps you will say that a glance at the machinery beneath the table would show in a moment how many turns had been made, and how many could be made. Very true: but what if the tramp had locked up his clock-work, and would not let you look at it?

The only evidence worth a rush is that of the lad who saw the whirligig set a-going.

I wish it to be distinctly understood that I am not proving the exact or approximate antiquity of the globe we inhabit. I am not attempting to show that it has existed for no more than six thousand years. I wish this to be distinctly stated, because I am sure I shall meet with many opponents unfair enough, or illogical enough, to misrepresent or misunderstand my argument, and sound the trumpet of victory, because I cannot demonstrate *that.* *All* I set myself to do, is to invalidate the testimony of the witness relied on for the indefinitely remote antiquity; to show that in a very large and important field of nature, evidence exactly analogous to that relied on would inevitably lead to a false conclusion, and must, therefore, be rejected, or received only contingently; received only as indicative of probability, and that only in the absence of any positive witness to the contrary.

Perhaps it may be objected, that there is no sufficient analogy between the phenomena from which the past history of a single organism is inferred, and those from which the past history of a world is inferred. Is there not?

Permit me to repeat an illustration I have already used. The geologist finds a fossil skeleton. His acquaintance with anatomy enables him to pronounce that the objects found are bones. He sees cylinders, condyles, cavities for the marrow, scars of attachment of muscles and tendons, foramina for the passage of nerves and blood-vessels; he finds the internal structure, no less than the form and surface, such as to leave not a doubt that these are real *bones*. Now universal experience has taught him that bones imply the existence of flesh; that flesh implies blood; that blood implies life; that life implies time. He therefore concludes unhesitatingly, that this skeleton was once alive, and that time passed over it in that living condition.

Is not this process of reasoning exactly parallel to that which he would have pursued if he had examined an animal the moment after its creation, (supposing this fact to be unknown to him,) and by which he would in like manner have inferred past time? And where is the vital difference between the two cases, which would operate to make a conclusion which is manifestly false in the one case necessarily true in the other?

One of the most eminent of living botanists has set forth in striking terms the parallelism which I am suggesting. Speaking of the *shoot* as the vegetable individual, and the woody trunk as a kind of ever-accumulating ground, which supports successive generations of shoots, he uses the following comparison.

"The history of the grand development of nature on the surface of our globe presents an analogy, which may perhaps serve to set this relation in a clear light. The successive geological formations superposed during the course of countless ages, present, buried in their depths, the traces of as many formations of the organic world, each of which carpeted the then superior stratum of the earth with a new life, until it found its own grave in the succeeding formation, when a new uprising of organic life took its place. In the same way, the stem of a tree is a multistratified ground, in whose layers the history of earlier growths is legibly preserved. The number of the woody layers indicates the number of the generations which have perished, *i. e.* the age of the whole tree: a distinct annual ring is the monument of a vigorous season, an indistinct one of a bad season, a sickly one (which is often found among healthy ones) indicates the unhealthiness of the foliage of that particular year. The practised woodman can decipher many facts of the past in the layers of the trunk; *e. g.* a good season for foliage or for seed, damage by frost or by insects, &c."*

In order to perfect the analogy between an organism and the world, so as to show that the law which prevails in the one obtains also in the other, it would be necessary to prove that the development of the physical history of the world is circular, like that already shown to characterise

---

* Dr. Alex. Braun, "On the Veget. Individual." (Ann. N. H. Nov. 1855.)

the course of organic nature. And this I cannot prove. But neither, as I think, can the contrary be proved.

The life of *the individual* consists of a series of processes which are cyclical. In the tree this is shown by the successive growths and deaths of series of leaves: in the animal by the development and exuviation of nails, hair, epidermis, &c.

The life of *the species* consists of a series of processes which are cyclical. This has been sufficiently illustrated in the preceding pages, in the successive developments and deaths of generations of individuals.

We have reason to believe that species die out, and are replaced by other species, like the individuals which belong to the species, and the organs which belong to the individual. But is the life of *the species* a circle returning into itself? In other words, if we could take a sufficiently large view of the whole plan of nature, should we discern that the existence of species δ necessarily involved the pre-existence of species γ, and must inevitably be followed by species ϵ? Should we be able to trace the same sort of relation between the tiger of Bengal and the fossil tiger of the Yorkshire caves, between *Elephas Indicus* and

Elephas primigenius, as subsists between the leaves of 1857 and the leaves of 1856; or between the oak now flourishing in Sherwood Forest and that of Robin Hood's day, from whose acorn it sprang?[*]

I dare not say, we should; though I think it highly probable. But I think you will not dare to say, we should *not*.[†]

It is certain that, when the Omnipotent God proposed to create a given organism, the course of that

---

[*] It may be objected that *Elephas primigenius* is absolutely distinct from *E. Indicus*. I answer, Yes, *specifically* distinct; and so am I distinct from my father,—*individually* distinct. But as individual distinctness does not preclude the individual from being the exponent of a circular revolution in the life-history of the species, so specific distinctness may not preclude the species from being the exponent of a circular revolution in some higher, unnamed, life-history.

[†] "We may assert of the individual, as well as of the species, that it completes the cycle of its existence in a succession of subordinate generations; while, on the other hand, we may affirm of the species, that, like the individual, it exhibits a determinate cycle of development." "The species itself may be regarded as an inferior 'momentum' of a still more comprehensive cycle of development."—*Dr. A. Braun*, "*On the Vegetable Individual.*"

"The species is an individual of a higher rank."—*Link: Elements of Botanical Science*, vi. 11.

"Species, like individuals, have a certain limited term of existence. It is the fact, that, *according to some general law*, species of animals are introduced, last for a limited period, and are then succeeded by others performing the same office."—*Ansted's Ancient World*, 52, 54.

organism was present to his idea, as an ever revolving circle, without beginning and without end. He created it at some point in the circle, and gave it thus an arbitrary beginning; but one which involved all previous rotations of the circle, though only as ideal, or, in other phrase, prochronic. Is it not possible—I do not ask for more—that, in like manner, the natural course of the world was projected in his idea as a perfect whole, and that He determined to create it at some point of that course, which act, however, should involve previous stages, though only ideal or prochronic?

All naturalists have speculated upon the great plan of Nature; a grand array of organic essences, in which every species should be related in like ratio to its fellow species, by certain affinities, without gaps and without redundancies; the whole constituting a beautiful and perfect unity, a harmonious scheme, worthy of the infinite Mind that conceived it. Such a perfect plan has never been presented by any existing fauna or flora; nor is it made up by uniting the fossil faunas and floras to the recent ones; *yet the discovery of the fossil world has made a very signal approach to the filling up of the great outline;* and the more minutely this has been investigated, the more have hiatuses been bridged over, which before yawned between species and species, and links of connexion have been supplied which before were lacking.*

It is not necessary,—at least it does not seem so to me,—that all the members of this mighty commonwealth should have an actual, a diachronic existence; any more than that, in the creation of a man, his fœtal, infantile, and adolescent stages should have an actual, diachronic existence, though these are essential to his normal life-history. Nor would their diachronism be more certainly inferrible from the physical traces of them, in the one case than in the other. In the newly-created Man, the proofs of successive processes requiring time, in the skin, hairs, nails, bones, &c. could in no respect be distinguished from the like proofs in a Man of to-day;

---

* "The unity of the plan of organization, and the regular succession of animal forms, point out a *beginning* of this great kingdom on the surface of our globe, although the earliest stages of its development may now be effaced: and the continuity of the series though all geological epochs, and the *gradual transitions* which *connect* the species of one formation with those of the next in succession, distinctly indicate that they form *the parts of one creation*, and not the heterogeneous remnants of successive kingdoms begun and destroyed: so that, while they present the best records of the changes which the surface of the globe has undergone, they likewise afford the best testimony of the recent origin of the present crust of our planet, and of all its organic inhabitants."—*Dr. Grant, in Br. Sci. Annual for* 1839.

yet the developments to which they respectively testify are widely different from each other, so far as regards the element of time. Who will say that the suggestion, *that the strata of the surface of the earth, with their fossil floras and faunas, may possibly belong to a prochronic development of the mighty plan of the life-history of this world*,—who will dare to say that such a suggestion is a self-evident absurdity? If we had no example of such a procedure, we might be justified in dealing cavalierly with the hypothesis; but it has been shown that, without a solitary exception, the whole of the vast vegetable and animal kingdoms were created,—mark! I do not say *may* have been, but MUST have been created—on this principle of a prochronic development, with distinctly traceable records. It was *the law of organic creation.*

It may be objected, that, to assume the world to have been created with fossil skeletons in its crust,—skeletons of animals that never really existed,—is to charge the Creator with forming objects whose sole purpose was to deceive us. The reply is obvious. Were the concentric timber-rings of a created tree formed merely to deceive? Were the growth lines of a created shell intended to deceive? Was the navel of the created Man intended to deceive him into the persuasion that he had had a parent?*

These peculiarities of structure were inseparable from the adult stage of these creatures respectively, without which they would not have been what they were. The Locust-tree could not have been an adult *Hymenæa*, without concentric rings;—nay, it could not have been an exogenous tree at all. The *Dione* could not have been a *Dione* without those foliations and spines that form its generic character. The Man would not have been a Man without a navel.

To the physiologist this is obvious; but some unscientific reader may say, Could not God have created plants and animals without these retrospective marks? I distinctly reply, No! not so as to preserve their specific identity with those with which we are familiar. A Tree-fern without scars on the trunk! A Palm without leaf-bases! A Bean without a hilum! A Tortoise without laminæ on its plates! A Carp without concentric lines on its scales! A Bird without feathers! A Mammal without hairs, or claws, or teeth, or bones, or blood! A Fœtus without a placenta! I have indeed written the preceding pages in vain, if I have not demonstrated, in a multitude of examples, the absolute necessity of retrospective phenomena in newly-created organisms. But if it can be undeniably shown in one single example, our

---

* Dr. Harris has the following observations:—

"Why might not God have created the crust of the earth, just as it is, with all its numberless stratifications, and diversified formations, complete? And the analogy for such an exercise of creative power is supposed to be found in the creation of Adam, not as an infant, but as *an adult;* and in the production of the *full-sized* trees of Eden. To which the reply is direct: the maturity of the first man, and of the objects around him, could not deceive him by implying that they had slowly grown to that state. His first knowledge was the knowledge of the contrary. He lived, partly, in order to proclaim the fact of his creation. And, could his own body, or any of the objects created at the same time, have been subjected to a physiological examination, they would, no doubt, have been found to indicate their miraculous production in their very destitution of all the traces of an early growth; whereas the shell of the earth is a crowded storehouse of evidence of its gradual formation. So that the question, expressed in other language, amounts to this: Might not the God of infinite truth have enclosed in the earth, at its creation, evidence of its having existed ages before its actual production? Of course, the objector would disavow such a sentiment. But such appears to be the real import of the objection; and, as such, it involves its own refutation."—*Pre-Adamite Earth*, p. 83.

Now this reasoning appeared, doubtless, very triumphant to the worthy Doctor: and yet a very little acquaintance with physiology would have taught him that he was enunciating an absurdity. The very supposition which he considers as self-refuting, is an indubitable physiological fact. I have abundantly shown, in the text, that the *cells which compose* the tree or the animal are as undeniable evidences of past processes as the concentric cylinders of timber, or the superposed layers of bone and scale.

failure to perceive it in ninety-nine other instances would in nowise invalidate the deduction from that one. Granted that you can triumphantly convict me of a *non-sequitur*, in ninety-nine out of every hundred of the cases in which I have attempted to show this connexion; still, if I have conclusively proved that in one solitary instance an animal or a plant was created with but one solitary evidence of pre-development, the principle for which I contend is established.

I trust, however, it does not rest on one example, nor on twenty, nor on a hundred. It may be thought that I have multiplied my illustrations needlessly: ten times as many might have been given. I wished to show that the proof is of a cumulative character: a single good example would, indeed, have established the principle; but I wished to show how widely applicable it is; that it is, indeed, of universal application in the organic kingdoms.

If, then, the existence of retrospective marks, visible and tangible proofs of processes which were prochronic, was so necessary to organic essences, that they could not have been created without them,—is it absurd to suggest the *possibility* (I do no more) that the world itself was created under

the influence of the same law, with visible tangible proofs of developments and processes, which yet were only prochronic?

Admit for a moment, as a hypothesis, that the Creator had before his mind a projection of the whole life-history of the globe, commencing with any point which the geologist may imagine to have been a fit commencing point, and ending with some unimaginable acme in the indefinitely distant future. He determines to call this idea into actual existence, not at the supposed commencing point, but at some stage or other of its course.* It is clear, then, that at the selected stage it appears, exactly as it would have appeared at that moment of its history, if all the preceding eras of its history had been real. Just as the new-created Man was, at the first moment of his existence, a man of twenty, or five-and-twenty, or

---

* I here assume the life-history of the globe to be represented by a straight line, because I cannot *prove* it to be a circle. I cannot even *imagine* its circularity. I do not mean the possibility;—I can imagine *that*: but the *mode* I cannot conceive. This, however, does not disprove the possibility. If man's science extended not beyond the accumulated observations of his own life, he would probably be quite incompetent to conceive how the life-history of such a tree as the Oak could be a circle; if he had never seen more than one individual, which was a tree when he was born, and continued to flourish till his death.

thirty years old; physically, palpably, visibly, so old, though not really, not diachronically. He appeared precisely what he would have appeared, had he lived so many years.

Let us suppose that this present year 1857 had been the particular epoch in the projected life-history of the world, which the Creator selected as the era of its actual beginning. At his fiat it appears; but in what condition? Its actual condition at this moment:—whatever is now existent would appear, precisely as it does appear. There would be cities filled with swarms of men; there would be houses half-built; castles fallen into ruins; pictures on artists' easels just sketched in; wardrobes filled with half-worn garments; ships sailing over the sea; marks of birds' footsteps on the mud; skeletons whitening the desert sands; human bodies in every stage of decay in the burial-grounds. These and millions of other traces of the past would be found, *because they are found in the world now;* they belong to the present age of the world; and if it had pleased God to call into existence this globe at *this* epoch of its life-history, the whole of which lay like a map before his infinite mind, it would certainly have presented all these phenomena; not to puzzle the philosopher,

but because they are inseparable from the condition of the world at the selected moment of irruption into its history; because they constitute its condition; they make it what it is.

Hence the minuteness and undeniableness of the proofs of life which geologists rely on so confidently, and present with such justifiable triumph, do not in the least militate against my principle. The marks of Hyænas' teeth on the bones of Kirkdale cave; the infant skeletons associated with adult skeletons of the same species; the abundance of coprolites; the foot-tracks of Birds and Reptiles; the glacier-scratches on rocks; and hundreds of other beautiful and most irresistible evidences of pre-existence, I do not wish to undervalue, nor to explain away. On the hypothesis that the actual commencing point of the world's history was subsequent to the occurrence of such things in the perfect ideal whole, these phenomena would appear precisely as if the facts themselves had been diachronic, instread of prochronic, as was really the case.*

---

* The existence of Coprolites—the fossilized excrement of animals--has been considered a more than ordinarily triumphant proof of real pre-existence. Would it not be closely parallel with the presence of fæces in the intestines of an animal at the moment of creation! Yet this appears to me demonstrable. It

Perhaps some one will say, "All this might be tenable, supposing the world were an organism. Your argument goes to show that organic essences in every stage of their existence present proofs of pre-existence; but what analogy is there between the lifeless inorganic globe (in which evidences of past processes are apparent, independent of the fossil organisms), and a living organic being,—plant or animal?"

I answer, The point in the economy of the organic creatures, on which their prochronism rests, is not the organic, but the circular condition of their being. The problem, then, to be solved, before we can certainly determine the question of analogy between the globe and the organism, is this:—Is the life-history of the globe a cycle? If it is (and there are many reasons why this is

probable), then I am sure prochronism must have been evident at its creation, since there is no point in a circle which does not imply previous points. At all events, geologists cannot prove that it is not.

Wherever we can discern a cyclical condition, there the law of which I am treating must hold good; and it certainly obtains in other things beside organisms. When the inorganic crust of the globe was first cleft to contain rivers, whence came the water that flowed through the fissures? A river is the produce of rivulets, which issue from mountain springs; these originate in the water that percolates through the soil; and this is derived from the rains, and snows, and dews, that are deposited from the atmosphere. But there would be no deposition from the atmosphere if the water had not first been carried up by evaporation; and the vaporable fluid is obtained from the moistened soil; from the lakes and rivers; and from the seas and oceans, whose loss is perpetually recruited from the flowing rivers. Here, then, we get a circle closely analogous to that of organic being. Was a given drop of water created as a component particle of a running stream? Its position and condition looked back to the mountain spring whence it must naturally have issued. Was

---

may seem at first sight ridiculous, and will probably be represented so; but truth is truth. I have already proved that blood must have been in the arteries and veins of the newly-created Man (*vide* p. 276, *supra*), and that blood presupposes chyle and chyme; but what became of the indigestible residuum of the chyme, when the chyle was separated from it? Would it not, as a matter of course, be found in the intestines? If the principle is true, that the created organism was exactly what it would have been had it reached that condition by the ordinary course of nature, then fæcal residua must have been in the intestines, as certainly as chyle in the lacteals, or blood in the capillaries.

it called into being in the spring? It looked up to the surface, whence it must have oozed. Was it formed on the surface? It looked to the clouds, whence it must have dropped. Was it created in the cloud? It looked down to the surface of the lake or sea, whence it must have been raised. Was it created in the lake? It looked to the river, whence it must have flowed.

The chief pelagic currents, which have hitherto so often been the destruction of the navigator, but which may yet become his able and subject servants, flow in circular systems. There is such an one in the southern part of the Indian Ocean, known as the Hurricane Region; another immense one ever running round and round the North Pacific; and, above all, that wondrous river of hot water—a river whose well-marked banks are not solid earth, but cold water—the Gulf Stream.

"The fruit of trees belonging to the torrid zone of America is annually cast ashore on the western coasts of Ireland and Norway. Pennant observes that the seeds of plants which grow in Jamaica, Cuba, and the adjacent countries, are collected on the shores of the Hebrides. Thither also barrels of French wine, the remains of vessels wrecked in the West Indian seas, have been carried. In 1809

His Majesty's ship *Little Belt* was dismasted at Halifax, Nova Scotia, and her bowsprit was found, eighteen months after, in the Basque Roads. The mainmast of the *Tilbury*, buried off Hispaniola in the Seven Years' war, was brought to our shores."*

These facts are dependent on the eastward set of this majestic current; and so is another great physical fact of immeasurable importance to us;—the superiority in temperature of the western shores of Europe over the eastern shores of North America. The harbour of St. John's, Newfoundland, is frequently fast closed by ice in the month of June; yet the latitude of St. John's is considerably south of that of the port of Brest, in France.

Impelled by the rotatory motion of the earth, and by the trade-wind,† the equatorial waters of the Atlantic are ever urged, a broad and rapid river, into the Caribbean sea, and the Gulf of Mexico, the narrowing shores of which compress the stream as in a funnel. The Andes here present a slender but impregnable barrier to its further progress westward; and the trend of the Isthmus

---

* *Blackwood*; April, 1849; p. 412.

† Strictly speaking, the current is a lagging behind of the water, which cannot keep pace with the speed communicated to the solid crust of the globe at its equatorial regions. The trade-wind is owing to the same cause.

turns it to the northward. Still finding no outlet, the impatient current, like a wild-beast pacing round its cage, courses the Gulf of Mexico, doubles the peninsula of Florida, and pursues its way first to the north-east, and then to the east, crossing the Atlantic in a retrograde direction, until it laves with its warm billows the coasts of Europe. Here it turns to the southward, and after embracing the "Fortunate" isles that lie off the African shores,—the Azores, the Madeiras, and the Canaries,—it joins the great equatorial set beneath the trade-wind, and returns on its westward course.

This mighty circulation of water must have been going on from the instant that the earth commenced rotating on its axis, or (granting this to have been chronologically subsequent) from the instant the Atlantic occupied its present bed. Whether sooner or later, it commenced at *some* instant; but at that instant all the previous elements of the circle were presupposed, and a boundless succession of former circles. An intelligent stranger, looking on the movement immediately after its commencement, but ignorant of its origin, would not be able to assign any limit to its past duration. From his observation of the velocity of the current in different parts of the circle, he would say with

confidence,—"These identical particles of water, which I see now urged on their ceaseless course towards the middle of the North Atlantic, were, yesterday morning at this hour, in the latitude of the mouth of the Chesapeake; on the morning before, off Cape Hatteras; on the morning before that, off Cape Lookout;" and so backwards interminably.

Whether the economy of the globe is circular, or not, I am not in a position to show. But its movements certainly are; and so are the movements of all the myriad worlds with which astronomy is conversant. Asteroids, planets, satellites, comets, suns, —nay, even the stellar universe itself—obey *in their motions*, the grand universal law of circularity. Take any one of these;—our Moon. When its orbital motion commenced, it commenced at some point or other of the circle which it describes in its course around the earth. The pre-existence, or at least the co-existence, of the Earth, and also that of the Sun, are necessary to its motion. Supposing it possible for a spectator, furnished with modern astronomical knowledge, to have looked at that instant on the newly-spun orb, would he not confidently have inferred, from its position at that moment, its position a week before? Would he not have felt able to indicate with unhesitating certainty the solar and lunar eclipses of a century or a chiliad before, just as he now calculates the time of the eclipse that marked the death of Herod the Great? Undoubtedly he would; for he would assume the constancy of those movements which modern science has deduced from the observations of many centuries; and, granting him the fact of their constancy, we could not invalidate his conclusions. Yet *what* would he have shown? The conditions and phenomena of bodies before they had begun to exist. The conditions are legitimately deducible; but they are prochronic conditions.

The mention of the celestial orbs suggests to remembrance the famous argument for the vast antiquity of the material universe, founded on the time which is required for the propulsion of light. I believe it owes its origin to Sir William Herschel.

Speaking of the known velocity of light in connexion with the immense distance of certain nebulæ, that eminent astronomer made these remarks:—

"Hence it follows, that, when we . . . see an object of the calculated distance at which one of these very remote nebulæ may still be perceived. . . . . the rays of light which convey its image to the eye must have been more than nineteen hundred and ten thousand, that is, almost *two millions*, of years on their way; and that, consequently, so many years ago, this object must already have had an existence in the sidereal heavens, in order to send out those rays by which we now perceive it." *

try 10^{7}

The notion has been amplified, with some interesting details, by a writer in the *Scottish Congregational Magazine* for *January* 1847; who thus throws the statements into a tabular form, and comments on them.

"From	the Moon, light comes to the earth in			1½	second
„	the Sun	„	„ in	8	minutes
„	Jupiter	„	„ in	52	„
„	Uranus	„	„ in	2	hours
„	a fixed Star of	1st	magnitude —	3 to 12	years
„	„	2d	„	20	„
„	„	3d	„	30	„
„	„	4th	„	45	„
„	„	5th	„	66	„
„	„	6th	„	96	„
„	„	7th	„	180	„
„	„	12th	„	4000	„

"Now, as we see objects by the rays of light passing from those objects to our eye, it follows that we do not perceive the heavenly bodies, *as they*

---

* Philos. Trans. for 1802; p. 498.

187

are at the moment of our seeing them, but *as they were* at the time the rays of light by which we see them left those bodies. Thus when we look at the moon, we see her, not as she is at the moment of our beholding her disc, but as she was a second and a quarter before; for instance, we see her not at the moment of her rising above the horizon, but 1¼ second after she has risen. The sun also when he appears to us to have just passed the meridian, has already passed it by 8 minutes. So, in like manner, of the planets and the fixed stars. We see Jupiter, not as he is at the moment of our catching a sight of him, but as he was 52 minutes before. Uranus appears to us, not as he is at the moment of our discovering him, but as he was 2 hours previously. And a star of the 12th magnitude presents itself to our eye as it was 4,000 years ago: so that, suppose such a star to have been annihilated 3,000 years back, it would still be visible on the earth's surface for 1,000 years to come: or, suppose a star of the same magnitude had been created at the time the Israelites left Egypt, it will not be perceptible on the earth for nearly 700 years from this date."

Beautiful, and at first sight unanswerable as this argument is, it falls to the ground before the spear-

touch of our Ithuriel, the doctrine of prochronism. There is nothing more improbable in the notion that the sensible undulation was created at the observer's eye, with all the pre-requisite undulations prochronic, than in the notion that blood was created in the capillaries of the first human body. The latter we have seen to be a fact: is the former an impossibility?

It may perhaps be said:—"The traces of prochronism you have adduced in created organisms may be granted, because they are inseparable from the presumed condition of those organisms respectively. The blood in the vessels, the hair, the teeth, the nails, may afford evidences of past processes; but then those are only past stages of what yet exists. The case, however, is not parallel with the fossil skeletons, many of which have no connexion with anything now existing. The concentric rings of a timber-tree are essential to its adult state; but how is the existence of the *Pterodactyle* or the *Megatherium* essential to that of the recent *Draco volans*, or the South American Sloth? Can you show in the new-formed creature any trace of some organ which does not come into its present condition of being,—of something which has quite passed away?"

Perhaps I can. The very concentric rings of the tree are considered by botanists as, in some sense, dead. The paradoxical dictum of Schleiden, —"No tree has leaves,"[*]—is grounded on this circumstance, that the woody portion of the mass is the inert result of former generations, and that the present race of leaves is growing, not out of the woody portion of the tree, but out of its herbaceous extremities, "which grow upon the woody stem *as upon a ground*, formed by the process of vegetation. This common ground, namely, the woody stem, *which is almost lifeless* in comparison with the herbaceous parts engaged in active growth, is annually covered with a vigorous sheath under the protecting bark; and this sheath is the ground of the nourishment of all the vegetating herbaceous extremities."[†]

The polygonal plates into which the bark of the *Testudinaria* divides, not only show many superposed laminæ, at any given moment of its adult condition, but also bear witness, in the broad existent surface of each one, to former laminæ, yet older than the oldest now present, which have disintegrated and dropped off.

The Palm and the Tree-fern show, in their trunk-

---

[*] Beitrage, p. 152.

[†] Dr. A. Braun, On the Veg. Indiv.

scars, evidences of organs which have completely died away and disappeared; while, between these scars and the generation of living fronds, there is, at any given moment of the tree's history, a series of fronds which are quite dead and dry, but which have not yet disappeared.

The *Nerita*, a genus of beautiful shells from the tropical seas, dissolves away and removes, in the progress of growth, the spiral column, which originally formed the axis of development; so that, in adult age, the spiral direction of the whole testifies to the past existence of a column which has quite disappeared.

In that species of *Murex*,[*] which, on account of the long and slender rostellum, and the spines with which it is covered, is known to collectors as the Thorny Woodcock (*M. tenuispina*), the shelly spines of the earlier whorls would interfere with such as came, in process of development, to be superposed on them; for they cross the area which is to be the cavity enclosed by the advancing lip. They are, however, removed by absorption; but not so completely but that traces may still be discovered where they formerly existed: evidences of the quondam existence of what exists no longer.

---

[*] See *ante*, p. 233.

Towards one side of the upper surface of the pretty Star-fish, *Cribella rosea*, (as in many other species of Star-fishes,) there is a curious little mark, known as the *madreporic plate*, the use of which has greatly puzzled naturalists. Sars, the Norwegian zoologist, has unveiled the mystery.* The young larva, before it assumes the stellar form, is furnished with a sort of thick column, divided into four diverging clubbed arms, which are adhering organs, ancillary to locomotion. In the process of development, however, new locomotive organs are formed; and this four-fold column, being no longer needed, sloughs away; and that so completely, that not a trace of its existence remains, *except this scar*, or "*madreporic plate;*" which is therefore a permanent record of something that has quite passed away.

But the closest parallel to the relation borne by the skeleton of an extinct species to an extant one, is presented by that of the hilum to a seed, or of the umbilicus to a mammal. Each of these is a legible and undeniable record of a being, whose individuality was totally distinct from that of the being by which it is presented, and of which all vestiges have disappeared, *save this record.* Nor

---

* Fauna Littor. Norveg.; i. 47.

is the parallel founded on obscure or rare examples; both the umbilicus and the hilum are generally conspicuous; and both are extensively found in their respective kingdoms, the former pervading the viviparous Vertebrata, the latter characterising the whole of the cotyledonous types of vegetation.

Once more. An objection may arise to the reception of the prochronic principle, on the ground that the examples I have adduced are not to be compared, in point of grandeur, with the mighty revolutions, which are presumed to have written their records in the crust of the globe; and that hence no analogy can be fairly drawn from one to the other. To the philosopher, however, there is no great or small, as there is none in the works of God. We have every reason to believe that He has wrought by the same laws in all portions of his universe: the principle on which an apple falls from the branch to the ground, is the same as that which keeps the planet Neptune in the solar system. I have shown that the principle of prochronic development obtains wherever we are able to test it; that is, wherever another principle, that of *cyclicism*, exists; whether the cycle be that of a gnat's metamorphosis, or of a planet's orbit. The distinction of great or small, grand or mean, does not apply to it. If it cannot be proved to be universal, it is only because we are not sufficiently acquainted with some of the economies of nature to be able to pronounce with certainty whether they are cyclical or not. I am not aware of any natural process, in which its existence can be absolutely denied.

And this makes all the difference in the world between my position and that of the old simple-minded observers, with which a superficial reader might think it to possess a good deal in common. A century ago, people used to talk of *lusus naturæ;* of a certain *plastic power* in nature; of abortive or initiative attempts at making things which were never perfected; of imitations, in one kingdom, of the proper subjects of another, (as plants were supposed to be imitated by the frost on a window-pane, and by the dendritic forms of metals). Still later, many persons have been inclined to take refuge from the conclusions of geology in the absolute sovereignty of God, asking,—"Could not the Omnipotent Creator make the fossils in the strata, just as they now appear?"

It has always been felt to be a sufficient answer to such a demand, that no reason could be adduced for such an exercise of mere power; and that it would be unworthy of the Allwise God.

But this is a totally different thing from that for which I am contending. I am endeavouring to show that a grand LAW exists, by which, in two great departments of nature at least, the analogues of the fossil skeletons were formed without pre-existence. An arbitrary acting, and an acting on fixed and general laws, have nothing in common with each other.

Finally, the acceptance of the principles presented in this volume, even in their fullest extent, would not, in the least degree, affect the study of scientific geology. The character and order of the strata; their disruptions and displacements and injections; the successive floras and faunas; and all the other phenomena, would be *facts* still. They would still be, as now, legitimate subjects of examination and inquiry. I do not know that a single conclusion, now accepted, would need to be given up, except that of actual chronology. And even in respect of this, it would be rather a modification than a relinquishment of what is at present held; we might still speak of the inconceivably long duration of the processes in question, provided we understand *ideal* instead of *actual* time;—that the duration was projected in the mind of God, and not really existent.

The zoologist would still use the fossil forms of non-existing animals, to illustrate the mutual analogies of species and groups. His recognition of their prochronism would in nowise interfere with his endeavours to assign to each its position in the scale of organic being. He would still legitimately treat it as an entity; an essential constituent of the great Plan of Nature; because he would recognise the Plan itself as an entity, though only an ideal entity, existing only in the Divine Conception. He would still use the stony skeletons for the inculcation of lessons on the skill and power of God in creation; and would find them a rich mine of instruction, affording some examples of the adaptation of structure to function, which are not yielded by any extant species. Such are the elongation of the little finger in *Pterodactylus*, for the extension of the alar membrane; and the deflexion of the inferior incisors in *Dinotherium*, for the purposes of digging or anchorage. And still would he find, in the fossil forms, evidences of that complacency in beauty, which has prompted the Adorable Workmaster to paint the rose in blushing hues, and to weave the fine lace of the dragonfly's wing. The whorls of the *Gyroceras*, the foliaceous or zigzag sutures of the *Ammonites*, and the radiating pattern of *Smithia*, are not less elegant than anything of the kind in existing creation, in which, however, they have no parallels. In short, the readings of the "stone book" will be found not less worthy of God who wrote them, not less worthy of man who deciphers them, if we consider them as prochronically, than if we judge them diachronically, produced.

GYROCERAS.

Here I close my labours. How far I have succeeded in accomplishing the task to which I bent myself, it is not for me to judge. Others will determine that; and I am quite sure it will be determined fairly, on the whole. To prevent misapprehension, however, it may be as well to enunciate what the task was, which I prescribed, especially because other (collateral, hypothetical) points have been mooted in these pages.

All, then, that I consider myself responsible for is summed up in these sentences:—

I. The conclusions hitherto received have been but inferences deduced from certain premises: the witness who reveals the premises does *not* testify to the inferences.

II. The process of deducing the inferences has been liable to a vast incoming of error, arising from the operation of a *Law*, proved to exist, but hitherto unrecognised.

III. The amount of the error thus produced we have no means of knowing; much less of eliminating it.

IV. The whole of the facts deposed to by this witness are irrelevant to the question; and the witness is, therefore, out of court.

V. The field is left clear and undisputed for the one Witness on the opposite side, whose testimony is as follows:—

"IN SIX DAYS JEHOVAH MADE HEAVEN AND EARTH, THE SEA, AND ALL THAT IN THEM IS."

INDEX.

First signs of life on Earth

Geologists now consider that life originated very soon after the formation of the Earth and evidence is building up that living forms developed and diversified much earlier than we have hitherto believed

Chris Peat
Will Diver

One feature of the fossil record that puzzled 19th century geologists was that rocks below a certain level seemed to be devoid of fossils of animal or plant remains. The lowest obviously fossil-bearing deposits, which the geologists first identified in Wales, they assigned to the Cambrian period. (The name Cambrian is derived from Cambria—the Roman name for Wales.) For a long time geologists termed rocks from the period between the formation of the Earth and the emergence of obvious forms of life as "Precambrian". With the advent in the 20th century of radioisotope methods to date rocks it was soon found that animals with fossilised hard parts first appeared about 600 million years (Myr) ago, and this, then, by definition, is when the Cambrian began. But given that the Earth is about 4600 Myr old, this means that the fossil record with which geologists are most familiar represents only the latest 13 per cent of Earth history (Figure 1).

Columnar stromatolites, "living fossils," in the intertidal zone at Shark Bay, Western Australia

Yet most of the significant developments of the Earth as a planet occurred in the preceding 87 per cent of geological time; in the Precambrian. As living things were well established by the beginning of the Cambrian, then it is clear that physical conditions became conducive to the emergence of life some time during the Precambrian. What those conditions were is unknown. We do know that most of the oxygen in our atmosphere is produced by photosynthesis, but we do not know when the Precambrian atmosphere became oxygenic or how quickly this occurred. We know that different styles of mountain-building evolved and died out during the Precambrian but we do not know how significantly the disposition or area of the Precambrian continents differed from that of today, nor how this could have affected the environments in which life developed. This means that although we may use physical and chemical principles to reconstruct processes on the ancient Earth, in Precambrian times our planet exhibited features totally different from those we know today. It is important to consider the evolution of life in this context, not as a separate aspect, but as an integral part of planetary evolution.

Evidence for the first life forms that must have emerged during the Precambrian has not been easy to come by, however; and the search for evidence of earlier life forms becomes progressively more difficult as we examine older rocks. Very old rocks are rarely exposed at the surface, and often they have been greatly changed (metamorphosed) by the high temperatures and pressures that they experienced during burial and mountain-building episodes. In addition, the most likely candidates for early life forms would be simple, very small organisms (measured in micrometres) that would not be robust chemically or physically. Only in exceptional circumstances can we expect to find such microorganisms preserved in any clearly recognisable form in sedimentary rocks layed down *before* the start of the Cambrian period.

The oldest rocks found on Earth are in south-western Greenland and are about 3800 Myr old. Although the rocks are highly metamorphosed, there are signs that erosion and sedimentation were occurring when these rocks were being deposited, which indicates that the Earth was sufficiently cool for water to exist in liquid form. Geologists have found no conclusive evidence of life in these Greenland rocks; however, there is no reason to think that the early sedimentary environment was inhospitable to simple organisms.

Essentially there are four lines of evidence for life in the Precambrian, before 600 million years ago. The first is provided by chemical compounds, albeit now degraded into complex hydrocarbons and asphaltic residues. The second consists of tracks and physical impressions of soft-bodied organisms roving on the primeval substrate. Then there are the strange, mound-like structures called stromatolites that were formed through the interaction of microbes and sedimentary processes. Lastly, there are the organic remains of individual microbes themselves and the impressions of larger algae.

Chemical fossils

Complex organic compounds can be isolated from most sedimentary rocks, especially from dark laminated, fine-grained ones (shales), where such compounds are abundant and may be a source of oil and gas. Many of these compounds are so distinctive that they were almost certainly formed by living organisms. Unfortunately, all rocks contain minute pores and fractures through which recent organic residues can move. Consequently, much of the early work on chemical fossils has now been invalidated because contamination cannot be ruled out.

Even when chemical fossils can be shown to be indigenous, their molecular structures are likely to be highly modified and analytical results thus require careful evaluation. Only now are we beginning to study the degradation of microbial remains within recent sediments, but already a complex series of chemical transformations of organic and inorganic material is emerging. In many instances little is known of the microorganisms concerned and we have only a rudimentary understanding of the influence of their metabolic activity on sedimentary processes. One thing is certain: microbes can influence isotopic ratios and compositions and they can affect mineral transformations within sediments at depths of up to 100 metres. In geological terms, such influences would have affected sediments in their early history; subsequently the effects of

---

Chris Peat is Guy Newton Junior Research Fellow in Biology at Wolfson College, Oxford. Will Diver is a senior lecturer in the Geology Division of the Department of Environmental Sciences at Plymouth Polytechnic.

Peat, C. and Diver, W.
"First Signs of Life on Earth"
New Scientist

increased temperature and pressure will be superimposed if burial continues with the accumulation of kilometres of sediment.

Sophisticated techniques now enable us to investigate microbial processes in recent sediments. These techniques include heating the complex organic materials to release volatile components that can then be separated by gas chromatography and identified by mass spectrometry. At first such analysis was concerned mainly with identifying specific molecules. Now there is increasing emphasis on the distribution of compounds in suites of samples. Such work is very relevant to Precambrian material and may eventually help us in the tortuous task of evaluating the influence of ancient microorganisms on sedimentary processes.

Trace fossils

The idea that multicellular animals (or metazoa) with their cells organised in tissues instantaneously appeared at the end of the Precambrian period is incorrect, according to geologists working on a refined definition of the beginning of the Cambrian period. In support of their views is much new evidence from trace fossils—the preserved tracks, trails and burrows made in sediments by organisms. Although precise dating by radiometric techniques is not available, a pattern is emerging of the events as the metazoa evolved before organisms appeared with a hard outer shell or skeleton . . . the first fossils with obvious hard parts (Figure 1).

The earliest generally accepted trace fossils are about 750 Myr old (that is, are Precambrian) and are simple burrows and trails made by worm-like creatures. These must have been made by quite complex metazoa capable of moving about in and on the sediment, so the earliest *very* simple animals must have evolved well before this. The simple traces are succeeded by impression fossils of soft-bodied metazoa, such as jellyfish, and bilaterally symmetrical worms, all of which lived between 680 and 600 Myr ago (still Precambrian, but only just). At this time more complex tracks and trails are found, some of which resemble those associated with Cambrian skeletal faunas. This strongly suggests that many of the skeletonised animal groups evolved from soft-bodied forms with a similar body plan.

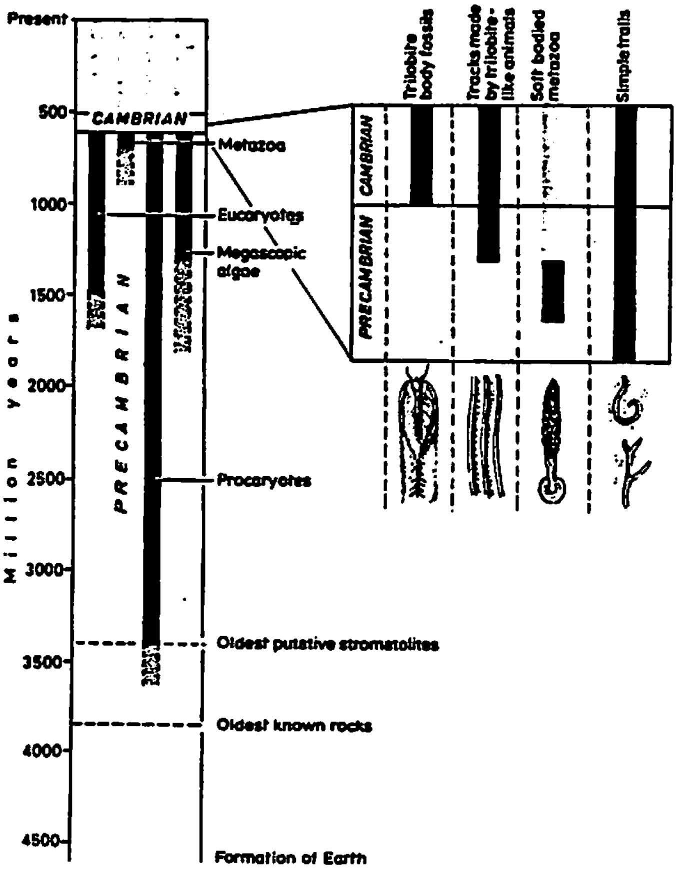


Figure 1 *Drawing the geological column to a linear time scale emphasises that the Precambrian period accounts for 87 per cent of elapsed geological time. For most of this period there is evidence of life, although fossils older than 2000 Myr are difficult to interpret. The evolution of the metazoa is recorded by a distinct sequence of trace fossil types across the Precambrian/Cambrian boundary (above right)*

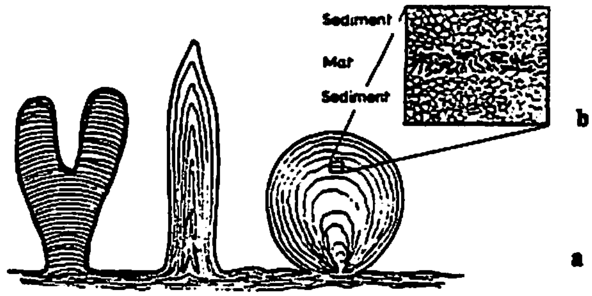


Figure 2 *Stromatolite forms include flat mats, columns and domes* (a). *All consist of alternating layers of sediment and mat bound together by the mat-forming microorganisms* (b)

Stromatolites

Stromatolites are not fossils in the strict sense but the products of a complex interaction between organisms and sedimentary processes (Figure 2). They vary greatly in shape and size, particularly those of the Precambrian, but all have a characteristic lamination.

Stromatolites are still forming today in several parts of the world (Figure 3), generally in arid and ecologically restricted environments such as the coastal plains along the shores of the Persian Gulf. Here saline flats, or *sabkhas*, are produced on the shore just above the tide-mark by a combination of low rainfall, a high exposure to the Sun, and a high rate of evaporation. In this environment the brines are of such extreme salinity that minerals such as anhydrite ($CaSO_4$) and gypsum ($CaSO_4.2H_2O$), are formed within the sediment. Most fossil stromatolites are found in limestones, many of which contain relics of these evaporite minerals.

On the tidal flats and *sabkhas* of the Persian Gulf today the surface of the sediment is covered with felted mats of microbes dominated by cyanobacteria (=blue-green algae). Many of these organisms secrete copious slimy (mucilaginous) sheaths around their cells. These protect the organisms from desiccation and intense solar radiation. Some of the filamentous forms move by a puzzling gliding motion, and when a layer of sediment is deposited on top of the mat they glide upwards between the particles of sediment to establish a new mat, abandoning their former sheaths in the buried mat beneath. As this process continues fine layers, organic and sediment-rich, gradually build up (Figure 2b). These layers are the characteristic laminations of the stromatolite.

The organic components of stromatolites are usually destroyed by decay and oxidation, but during the solidification and compaction of the sediment (lithification) subtle textural fabrics develop that often enable the original influence of microbial activity to be recognised. Many stromatolites, particularly those in the Precambrian, have distinctive overall shapes such as domes, columns and cones that provide an invaluable record of the behaviour, distribution and abundance of ancient communities of microbes ("microbiota").

The oldest structures interpreted as stromatolites come from the Pilbara region of Western Australia where they

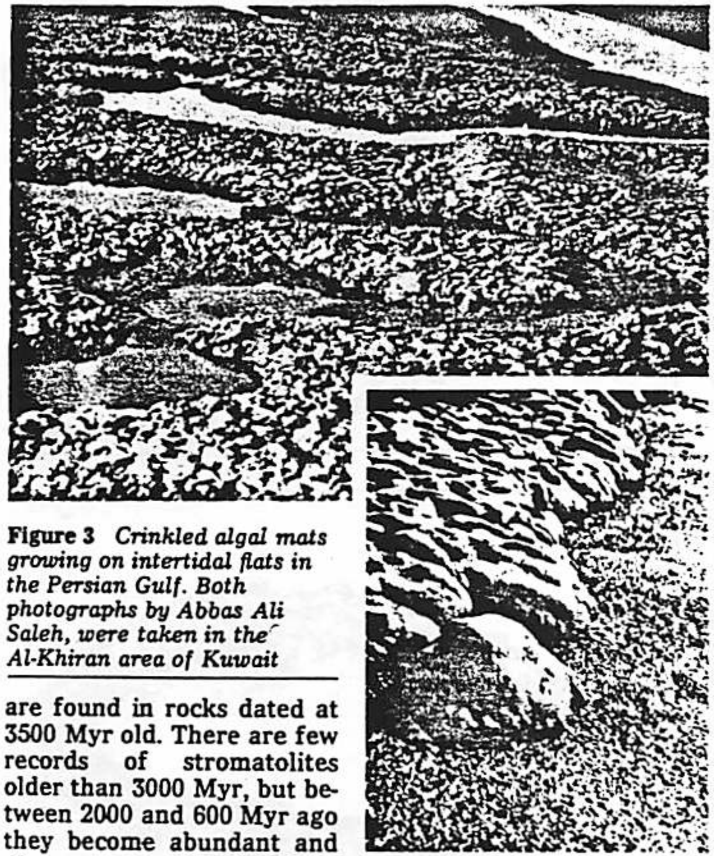


Figure 3 *Crinkled algal mats growing on intertidal flats in the Persian Gulf. Both photographs by Abbas Ali Saleh, were taken in the Al-Khiran area of Kuwait*

are found in rocks dated at 3500 Myr old. There are few records of stromatolites older than 3000 Myr, but between 2000 and 600 Myr ago they become abundant and diverse. But suddenly, 600 Myr ago, these ancient stromatolites began to be grazed by the newly-evolved molluscs and almost disappeared.

In addition to providing a record of the activity of ancient microbes, stromatolites are an invaluable source of Precambrian microfossils. In some ancient stromatolites, lithification began very early and embedded the microbes in mineral material. The microfossils are excellently preserved where the original limestone minerals have been replaced by silica to form a very hard flinty rock called chert. The remains of the microorganisms are thus embedded in flinty, fine-grained rock, and are often preserved in three dimensions. By grinding a thin section of the rock, the microfossils can be seen still embedded in the mineral matrix. Although the preservation in stromatolitic cherts may be striking, the preserved microfossils rarely represent more than a partial and fragmentary record of the original living mat. The microfossils can be dissolved out of the chert with hydrofluoric acid; but study in thin sections avoids chemical procedures which could introduce contaminants.

Precambrian microorganisms

One of the best-known sources of Precambrian microfossils is a rock called the Gunflint Chert that is exposed along the north-west shores of Lake Superior in North America and is about 2000 Myr old. (The Gunflint Chert was so named by early settlers who used it in their flintlock guns.) The organic-rich layers (laminae) of stromatolites in these cherts contain abundant spheres, filaments and rare other forms. In 1954, the field geologist Stanley Tyler, and Elso Barghoorn, a palaeobotanist, described the chert remains, and this was the first convincing evidence of Precambrian microbial fossils. At last, it had been shown unequivocally that some parts of delicate microorganisms could be preserved and recognised in ancient sedimentary rocks, and that it was feasible to look for direct evidence of the early evolution of life on Earth. The description of the Gunflint "microbiota" triggered a global search for other assemblages. This surge of research, which concentrated on stromatolites, has led to the discovery of at least 30 microbiotas of which those from the Bitter Springs Formation in central Australia (800 Myr), Amelia Dolomite in the Northern Territory of Australia (1500 Myr), Beck Springs Dolomite in California (1600 Myr) and Belcher Islands in Canada (1800 Myr) are the best known. These fossil assemblages are all comparable with living stromatolitic microbiotas. Some of the commonest fossils are hollow organic tubes, which we interpret as the sheaths abandoned by their occupying cyanobacteria. The Bitter Springs Formation contains well-preserved cellular filaments, and also spheroidal forms, some of which contain a small, round, opaque inclusion. It has been suggested that these inclusions represent the degraded remnants of nuclei and that the microfossils containing them were organisms akin to the red or green algae. In the Belcher Islands assemblages, layers of small spheroidal cells occur in thick mucilage which is discoloured like its living counterpart. Such close resemblances prompted many palaeontologists to compare Precambrian microfossils directly with living forms, and the fossils have often been given similar names. One of the difficulties of doing this is that fossil assemblages consist of microbes in varying states of decomposition and that a single living species can be degraded to produce several different fossil shapes. The converse is also true, and several different living species could form morphologically indistinguishable microfossils.

Research on living microorganisms reveals that far more fundamental than the subdivision of life into plants and animals is the schism between cells without a well-defined nucleus (prokaryotes) and those with a distinct nucleus (eukaryotes). Eukaryotic cells are characterised by their relatively elaborate internal organisation and membrane-bounded organelles such as nuclei, mitochondria and chloroplasts. Prokaryotes do not have such organelles. In the living microbiota, all bacteria and cyanobacteria are prokaryotic while all other organisms are eukaryotic.

It is possible that some eukaryotes originated by symbiosis between different kinds of prokaryotes, or by the ingestion (but not digestion) of one prokaryote by another. In any event, it is likely that the earliest eukaryotes were similar in size and shape to their contemporary pro-

Figure 4 *Scanning electron microscopy of a microfossil extracted from shales of the McMinn Formation of the Roper Group, Northern Territory, Australia*

karyotes. This makes the task of recognising the earliest eukaryotes in the fossil record extremely difficult because even in the best preserved fossil material only mucilaginous sheaths, cell walls and degraded contents are preserved. The delicate cellular fine structure is usually destroyed rapidly when a cell dies, often by the cell's own enzyme systems. Under some conditions the contents of a dying cell contract within the cell wall to form a small, round, dense body. In modern cyanobacteria the resulting structures are indistinguishable from the inclusions in the Bitter Springs microfossils. The controversy surrounding the interpretation of these fossil cells with inclusions emphasises the need for definitive morphological criteria with which to recognise simple eukaryotic microorganisms. Dorothy Oehler, the American biologist, has carefully examined these microfossils using a transmission electron microscope and has described differences in ultrastructure between the walls of fossils and their different inclusions. Although the development of ultrastructural techniques is an important advance, we still do not know what the spots represent, nor whether the cells were eukaryotic.

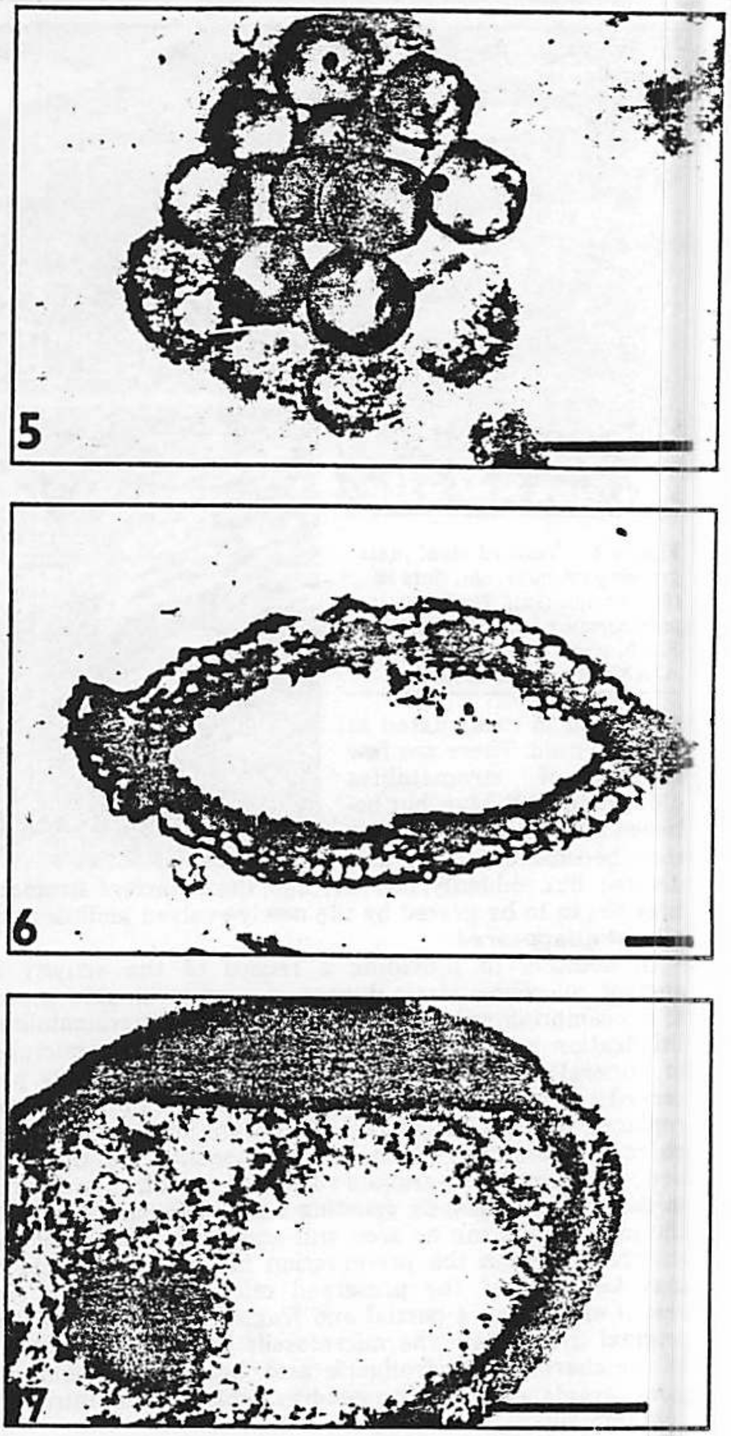
Irrespective of details of interpretation, it is clear that the associations of microbes that produced Precambrian stromatolites were both widespread and successful, and that microbiotas in stromatolitic chert provide a valuable catalogue of microbial life in one environment through much of the Precambrian. We have much to learn about stromatolites and the organisms that built them, but it is possible to recognise some broad trends. For example, there seems to have been a marked increase in the mean diameter of single cells in stromatolites about 1500 million years ago. Whether this reflects a major step in the evolution of prokaryotes, the origin of eukaryotes or some other evolutionary or environmental development remains to be seen.

Despite these trends, stromatolitic microbiotas have changed little since the middle Precambrian, and although they are valuable indicators of environmental conditions they are consequently of limited value in evolutionary studies. The greater part of our current sample of Precambrian life is dominated by these stable, perhaps ecologically specialised or restricted associations of microorganisms. We need to evaluate the extent to which Precambrian stromatolites represent extreme or ecologically restricted environments, but their microbiotas provide a biased sample of the Precambrian biosphere.

Microfossils from shales

Fortunately there are other sediments in which organic-walled microfossils are preserved. Dark grey or black shales represent ancient muds that were deposited in a wide range of sedimentary environments. They are a rich source of organic-walled microfossils that can be examined by the same methods as those preserved in chert, although greater care is needed in collecting and preparing the samples. Many of the microfossils from shales and cherts are morphologically identical, but assemblages from shales usually show a greater diversity of forms. Complex microfossils are rare in Precambrian rocks but, as studies continue, more and more are being discovered in shales.

The interpretation of shale microbiotas is less straightforward than with stromatolitic assemblages where the majority of the organisms are preserved where they lived. In shales we are usually dealing with assemblages of microorganisms that have been moved, mixed and sorted to an unknown extent by the water currents responsible for the transport and deposition of the host sediment. The microfossils in a single lamina of shale may have lived in markedly different habitats and it is not possible to compare them directly with living microbial assemblages from any particular recent environment. This means that a far wider range of microorganisms must be considered as possible living analogues of microfossils from a shale than of those from a stromatolite. Living microorganisms cannot be identified by their overall shape, partly because the cysts, or spores, of many species look much the same, and partly because one species can take very different forms in different conditions or during various stages of its life cycle. Because all fossil microorganisms are considerably altered and only cell walls, mucilage and cysts are preserved with any fidelity, the biological affinities of most simple microfossils may remain obscure for some time.



Figures 5-7 *(5) Colonial vesicles in a thin section of shale. (6) Partially flattened spheroidal microfossil with a vesicular wall. (7) Detail of a spheroidal microfossil with a decorated wall. All from the Torridon Group, NW Scotland. (Scale bar represent 20 µm in Figures 5-11.)*

All the fossil microbes illustrated in Figures 4-10 could represent prokaryotic microorganisms or extinct microbial forms, but comparisons with living microorganisms suggest

Figure 8 *Spheroidal microfossil with reticulate wall decoration. Locality as for Figure 4*

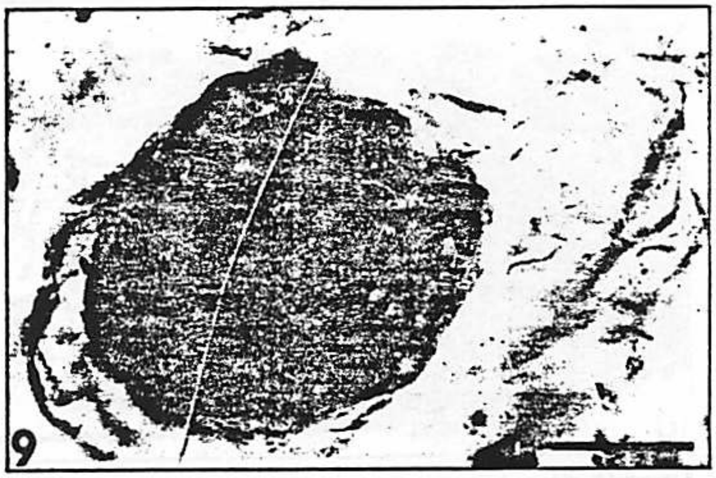
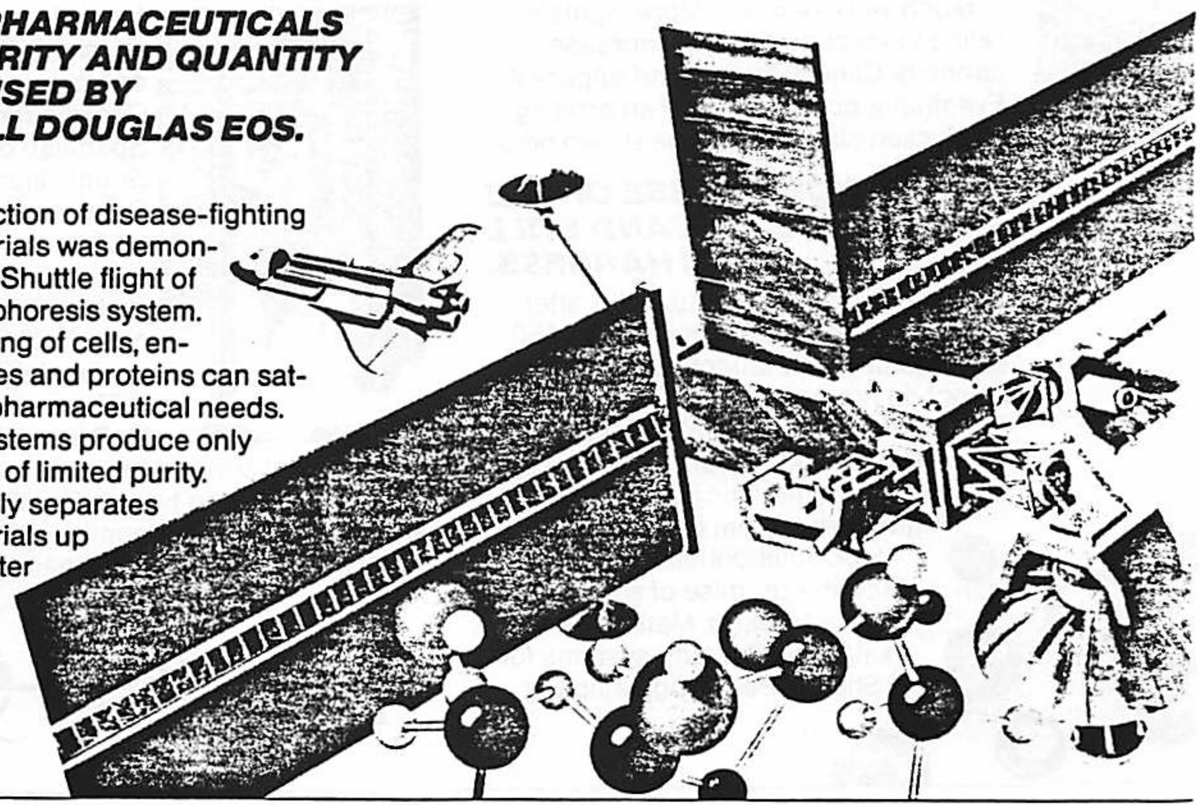


Figure 9 *Spheroidal microfossil being released by the splitting of an outer membrane. Locality as for Figure 4*

that the more complex forms are likely to be the remains of eukaryotes. Present-day prokaryotes do not produce large structures with complex or highly decorated walls, nor large thick-walled vesicles within a membrane or mucilage. The microfossils from the Torridon Group and Roper Group illustrated here provide some of the best evidence for Precambrian microbial eukaryotes.

Despite the difficulties of interpreting the microfossils, careful studies can provide a basis that will enable them to be used as evolutionary markers, for correlating sequences of rocks, and for analysis of palaeo-environments. Other kinds of fossils have all been classified into genera, species and higher taxonomic categories on the basis of preserved structure. We have recently recommended a similar approach to the relatively simple organic-walled microfossils. We call these fossil microbes *cryptarchs*, from the Greek *cryptos* (unkown) and *arche* (origin), because of the difficulties of establishing their biological affinities. The taxonomy and nomenclature of cryptarchs should be as objective as possible and kept free from biological inferences and preconceptions. This will make it much easier for geologists to use these fossils in stratigraphical studies and in mapping, and will not hinder research and speculation on their biological affinities.

One difficulty in this work is the very fine control of cryptarch assemblages by the environments in which the



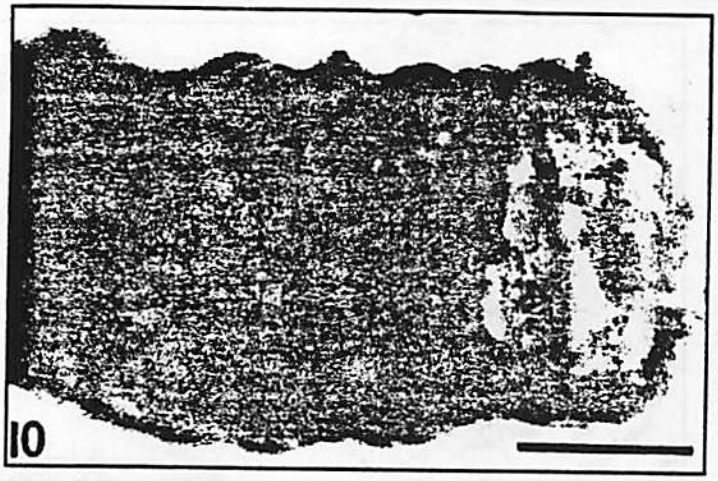


Figure 10 *The open end of a transversely striated tubular microfossil. Locality as for Figure 4*

organisms lived and in which their remains were deposited and preserved. Detailed microscope work on thin sections of Torridon Group shales has shown that successive laminae in a finely laminated shale can contain markedly different assemblages of cryptarchs. Work at Plymouth Polytechnic using both fossil and living material is now being aimed at distinguishing between indigenous and transported fossils, and also at interpreting the different assemblages and processes of sediment accumulation. Current work at Oxford is concentrating on the classification, nomenclature and fine structure of cryptarchs.

Near the beginning of the Cambrian period there appeared a group of microfossils that have organic walls and are relatively complex. These are called acritarchs (Figure 11), and are important index fossils in sediments of the lower Palaeozoic (between 600 and 430 Myr ago). Many cryptarchs are evolutionary forerunners of the acritarchs and are evidence of a gradual structural elaboration that began in the middle Precambian. There seems to be a gradual increase in morphological complexity as we move from older to younger shales. We have yet to establish the precise timing of the appearance of microfossils with complex walls and their environmental significance because we have been able to examine so few shales properly. There is every reason to believe that cryptarchs will eventually be used as index fossils for the later Precambrian, but the study of shallow-water shale assemblages in the 1200-1800 Myr age bracket is an urgent priority, and can be expected to result in major advances in our understanding of Precambrian evolution. □

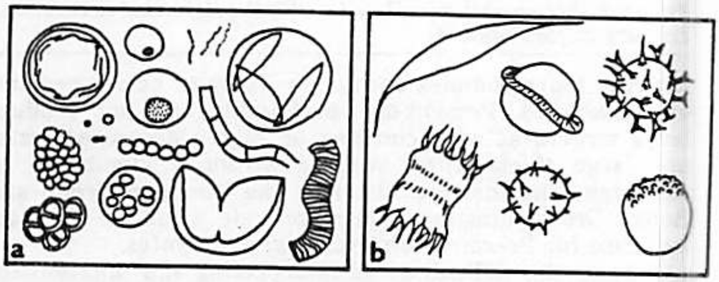


Figure 11 *Cryptarchs are the only organic-walled microfossils found in Precambrian sediments and include simple spheres, colonies and filaments* (a). *The relatively complex acritarchs with their odd shapes and varied processes* (b) *do not appear until after the beginning of the Cambrian period*

Gould, S.J.
"Living with connections"
Natural History
American Museum of Nat. Hist.
Used by permission

Living with Connections

Are Siamese twins one person or two?

by Stephen Jay Gould

La Grand Galerie of the Muséum d'Histoire Naturelle in Paris has been closed for fifteen years. This great space, supported by iron and roofed in glass, is no longer structurally sound. Like the capacious nineteenth-century railroad stations that served as its model, La Grande Galerie has passed into history. Its exhibits, too, reflect the thoughts and concerns of another age, the expansive and aggressive Victorian era that took so seriously, as a guide for collection and display, the words of Genesis (1:22): "Be fruitful and multiply, and fill the waters in the seas, and let fowl multiply in the earth." If modern museums emphasize intimacy, good lighting, tasteful display, and well-chosen words, their Victorian predecessors judged quality by quantity and crammed as many large animals as possible into their vast open spaces. At Lord Rothschild's museum in Tring, the stuffed zebras are supine, so that four tiers, rather than two or three, may be fitted from floor to ceiling.

La Grande Galerie is the granddaddy of them all. Built in 1889, and unchanged since, its skeletons and stuffed animals occupy every available inch. The great central pyramid almost reaches the high glass ceiling. One side is all zebras, another all antelopes; six giraffes crown the summit. The dust is thick, the hall dark and empty; eerie silence marks a dingy majesty.

The companion hall, La Galerie d'Anatomie Comparée, is smaller, well lit, and still open. Its style is identical—row upon endless row, tier upon tier of blanched skeletons. I wandered up and down the aisles, marveling at a long row of walruses and five superposed tiers of monkey skulls. Then I passed by cabinet 106 and stopped short. It contains a sideshow to offset the neighboring forest of sleek lions, and to remind complacent Victorians that nature can be capricious and cruel, as well as bountiful. Cabinet 106 holds a collection of teratological specimens, skeletons of deformed and abnormal births. Most are human and most represent that puzzling and frightening phenomenon of joined birth, or "Siamese" twinning. Skeleton A8597 has two heads, three arms, and two legs; A8613 has four arms, two legs, and two heads projecting from the ends of a joined vertebral column; A8572 is almost normal, but a tiny, headless brother with arms and legs projects from his chest. All are small and clearly died at birth or soon thereafter.

One skeleton stands out for its considerably larger size. A8599 is (or are)—and this is the issue we will soon discuss—twin girls with two well-formed heads and upper bodies with four full arms. Two distinct vertebral columns nearly join at their base, and only two well-formed legs extend below. The label reads *monstre humain dicéphale*, or "two-headed human monster." But A8599 was born live and survived several months. The twins were baptized and given names. The label records this poignant detail and includes, under the number and description, the identification "Ritta-Christina."

I mused much over Ritta and Christina, wondering about their life and death. Yet I would not have made the transition from troubled thought to column had I not discovered, quite by accident, a dusty old tome in a bookstore two days later—volume 11, for 1833, of the *Memoirs of the Royal Academy of Sciences*. It contained a long monograph by the great medical anatomist Etienne Serres: *Théorie des formations et des déformations organiques, appliquée à l'anatomie de Ritta Christina, et de la duplicité monstrueuse* ("Theory of organic development and deformation, applied to the anatomy of Ritta Christina, and to duplicate monsters in general").

Anyone who does not grasp the close juxtaposition of the vulgar and the scholarly has either too refined or too compartmentalized a view of life. Abstract and visceral fascination are equally valid and not so far apart. Two days before, I had seen young schoolchildren standing before Ritta-Christina in open-mouthed awe or horror, soon masked by forced humor. Now I learned that France's finest medical anatomist had dissected Ritta-Christina and used her to support a general theory of organic (not only human) embryology. Both themes seemed equally compelling to me; indeed, I had wallowed in both myself for two days. The children might not have generalized, but I have no doubt that M. Serres once gulped, as well as thought. I bought the book.

Ritta and Christina were born on March 23, 1829, to poor parents in Sardinia. Times were hard and social mobility scarcely possible in ordinary circumstances. Parents today would receive pity and experience only sorrow; in 1829, any realistic person, whatever his private feelings, must have recognized that such a child represented potential and substantial revenue, otherwise quite unobtainable. Thus, the parents of Ritta-Christina scraped together some funds and brought her to Paris, hoping to display her at fancy prices. The Hottentot Venus had provoked enough protest fifteen years earlier (see last month's column); but she was whole, however exotic. Public sensibilities had limits, and the authorities forbade any open display of Ritta and Christina. But she was shown privately, many times too often—for she died, in part from overexposure, after five months of life.

I have consciously switched back and forth from singular to plural in describing Ritta-Christina. When the vulgar and scholarly meet, a common question often underlies our joint fascination. One question has always overridden all others in this case—individuality. Was Ritta-Christina one person or two? This issue motivated the feeble jokes of my terrified schoolchildren. It also served as a focus for Serres's scientific investigation. And it underlay public fascination in 1829. When she died, a Parisian newspaper wrote: "Already it is a matter of grave consideration with the spiritualists, whether they had two souls or one."

One or two? Through all scholarly excursions and sideshow huckstering, this single question has been the focus of our fascination since Siamese twinning received its name. The originals, Eng and Chang, were born of Chinese parents in a small village near Bangkok in 1811 (Thailand was then called Siam). During the late 1820s and 1830s, they exhibited themselves in Europe and America and became quite wealthy. They decided to live in North Carolina where, at age 44, they married two sisters of English birth and settled down in two neighboring households to a comfortable life as successful (and, yes, even slaveholding) farmers. They switched houses at three-day intervals, traveling the one and a half mile distance by carriage. By the customs of the day, Chang was unquestioned master in his domicile, while Eng gave the orders *chez lui*. The unions were undeniably productive, for Chang had 10 children and Eng 12.

Chang and Eng were physically complete human beings connected by a thin band of tissue, three and a quarter inches at its widest and only one and five-eighths inches at its thickest. Each had a full set of parts down to the last toenail. They carried on independent conversations with visitors and had distinct personalities. Chang tended to be moody and melancholy and finally took to drink; Eng was quiet, contemplative, and more cheerful. Yet even they, history's most independent Siamese twins, apparently harbored private doubts about their individuality. They signed all legal documents "Chang Eng" and often spoke about their ambiguous feelings of autonomy.

But what of Ritta and Christina, whose bodily independence did not extend below the navel? They seemed, at first glance, to be two people above and but one below. One might have taken the easy way out and opted for the old cultural criterion of head and brain—two heads, two people. But as a scientist, Serres could not accept this solution, for he had also studied Siamese twins with one head, two arms, and four legs. He reasoned that a uniformity of process must underlie both types of twinning and could not accept the facile and simplistic resolution—one person if zipped halfway but starting from the top; two if zipped from the bottom.

Serres struggled with this momentous issue for 300 pages and finally concluded that Ritta *and* Christina were two people. His arguments and basic style of science belong to another era in the history of biology. They are worth recounting if only because few intellectual exercises are more rewarding than an examination of how radically different systems of thought treat a common subject of mutual interest. I also believe that Serres was at least half wrong.

Serres embodied the great early nineteenth-century tradition of romantic biology, called *Naturphilosophie* ("nature philosophy") in Germany and transcendental morphology in his native France. If modern morphologists study form either to determine evolutionary relationships and reconstruct the tree of life or to discover adaptive significance by examining function and behavior, Serres and his colleagues had markedly different goals. They were obsessed with the idea that some overarching, transcendental law must underlie and regulate all the apparent diversity of life.

These laws, in the Platonic tradition, must exist before any organism arises to obey their regularities. Organisms are accidental incarnations of the moment; the simple, regulating laws reflect timeless principles of universal order. Biology, as its primary task, must search for underlying patterns amidst the apparently confusing diversity of life. It must seek, in short, the "laws of form."

Serres followed his mentor, the great biologist Etienne Geoffroy Saint-Hilaire, in arguing that all animals share a common plan of construction, based on the vertebra as a building block. Linked vertebrae make a segmented animal; our skulls are the coalesced product of four primordial vertebrae. Insects and mammals share the same universal motif. Our soft parts are arranged around our vertebral axis. Insects have a segmented external skeleton; therefore, they dwell *within* their own vertebrae. (We now recognize that the segmentation of arthropods and vertebrates represents separate evolutionary events, and that no overarching law of form need underlie a superficial similarity.)

Original Siamese twins Chang and Eng

The Granger Collection

Serres made his most important contribution to the transcendental tradition in embryology. His colleagues had concentrated upon static adult form, searching for underlying patterns in final products alone. But organisms grow their own complexity from egg to adult. If laws of form regulate morphology, then we must discover principles for dynamic construction, not merely for relationships among finished creatures.

Serres's monograph on Ritta-Christina begins with an abstruse 200-page dissertation on the principles of morphology and their application to embryology. Unless we sneak a peek at the alluring plates in the back, we hear nothing of the famous Sardinian twins until our senses are numbed by generality. This organization, in itself, reflects a style of science strikingly different from our own. We maintain an empirical perspective and like to argue that generalities arise from the careful study and collation of particulars. Any modern embryologist would discuss Ritta-Christina first and only venture some short and cautious conclusions at the end. But Serres, as a transcendentalist, believed that laws of form existed before the animals that obeyed them. If abstraction preceded actuality in nature, why not in human creativity as well? Thought and theory first, application later. (Neither extreme well represents the intricate interplay of fact and theory in our actual practice of science, a theme of so many columns. Still, I suspect that Serres's "inverted" order is no worse a distortion of complex reality than our modern stylistic preferences.)

In the first pages of his monograph, Serres tries to reduce the embryology of all animals to three basic laws of "organology." First, by the law of *eccentric development*, otherwise known as the law of circumference to center, organs form initially at the edge of the developing embryo and then migrate toward the center. Second, by the law of *symmetry*, organs that become single and central in adults begin as double symmetrical rudiments on opposite edges of the developing embryo. Third, by the law of *affinity*, these symmetrical rudiments are drawn to each other until they fuse in the center to form a single adult organ. (Let me be charitable and simply state that these laws are unwarranted extensions of patterns that operate occasionally in development. Serres was writing before the establishment of cell theory and just a few years after Karl Ernst von Baer's discovery of the human ovum. His formal approach to morphology, so foreign to a world that can assess cellular and even molecular causes, fit the knowledge and mores of his own era.)

Two hundred pages later, when Serres finally discusses his dissection of Ritta-Christina, we understand why he devoted so much preceding space to the three great laws of organology—for they provide his solution to the great dilemma of individuality. Ritta and Christina are two people, albeit imperfect, and the laws of form proclaim *their* status.

No one quarreled with the double verdict on Ritta and Christina from the waist up; the dilemma had always rested upon their well-formed, but clearly single, lower half—one anus, one vaginal opening, two legs. If she were two people all the way from stem to stern, how could her lower half form so well in the shape of one? How could the incomplete parts of two separate creatures fuse and blend into a form indistinguishable from the lower half of such unambiguous singletons as you or me?

Serres used his laws of organology to render Ritta and Christina's lower half as the conjoined product of two people. After all, the harmonious, well-formed single organs of ordinary individuals arise (by the law of symmetry) as separate and double parts at the embryo's edge, and then move inward (by the law of circumference to center), eventually meeting and fusing (by the law of affinity) into one integral organ. If our single heart, stomach, and liver begin as two symmetrical rudiments, then why should we view the presence of a single, well-formed organ in Ritta and Christina's lower half as any argument against its construction from the mingled and melded parts of two embryonic individuals? If the twins have but one uterus, then the right half came from Ritta, the left from Christina. The two rudiments formed at the embryonic edges, in regions unambiguously assigned to Ritta or to Christina (law of symmetry). They moved toward the midline (circumference to center) and joined there (law of affinity) to form a single organ.

Serres announced proudly that his laws of form had resolved the great dilemma in favor of duality: "How could we possibly have conceived that each child furnished half of an organ common to both, if the law of eccentric development had not taught us that single organs are, in their normal state, originally double."

Nor did Serres shrink from the decidedly peculiar logical implications of his solution. He noted that the large uterus had proper connections with the ovaries and vaginal canal and saw no reason why Ritta and Christina might not have borne children had they lived to maturity. (Serres also found a second, rudimentary uterus that would not have worked.) He concluded that the large uterus had formed half from Ritta and half from Christina, and admitted that any children would have two natural mothers:

> This disposition of Ritta and Christina's genital organs evidently shows . . . that while nature had taken measures to assure the lives of these children, she had not forgotten the possibility of their reproduction. Now, for this reproduction, nature had combined everything, so that all the pleasures and pains would be shared. . . . Supposing that conception occurred in the large uterus, a single child would have had two distinct mothers, a singular result of this associated life.

Serres then discussed a pair of conjoined males with four legs and a single head, and opted for consistency and duality: the single well-formed brain shared the combined thoughts of two.

> There is a perfect unity produced by two distinct individualities. There are sense organs and cerebral hemispheres for a single individual, adapted to the service of two, since it is evident that there are two *me*'s in this single head [*deux* moi *dans cette tête unique*].

Thus Serres made a valiant and consistent attempt to resolve a question that seemed hopelessly ambiguous. We may appreciate the effort and enjoy an excursion into the different view of biology that supported it. But we must reject the conclusion.

Fertilized human eggs usually develop into single individuals. Rarely, the dividing cells separate into discrete groups, and two embryos develop. These one-egg (or identical) twins are genetic carbon copies. In some ultimate, biological sense, they are the same iterated individual—and the psychological literature contains ample testimony to feelings of imperfect separation shared by many so-called identical twins. Yet, at least for definition's sake, we have no difficulty in identifying one-egg human twins as undeniably separate personalities for two excellent reasons: first, physical separation is the essence of our vernacular definition of individuality; second, human personalities are so subtly and pervasively shaped by complex environments of life (whatever the quirky similarities between one-egg twins reared apart) that each person follows a unique path.

With vastly greater rarity, the dividing cells of a fertilized egg begin to separate into two groups, but do not complete the process—and conjoined (or Siamese) twins develop. Conjoined twins span the entire conceivable range from a single individual bearing a few rudimentary parts of an imperfect twin, to superficially joined, complete individuals like Chang and Eng. Ritta and Christina fall squarely in the middle of this continuum. With our modern knowledge of their biological formation, I fear that we must reject Serres's solution, and admit instead that his dilemma cannot be answered.

We live in a complex world. Some boundaries are sharp and permit clean, either/or definitions. But nature is replete with continua that cannot be nicely parceled into two piles of unambiguous yeses and noes. Biologists have heaped almost universal scorn upon attempts by anti-abortionists to define an unambiguous "beginning of life," because we know so well that the sequence from ovulation or spermatogenesis to birth is an unbreakable continuum—and surely no one will define masturbation as murder. Our congressmen may create a legal fiction for statutory effect but they may not seek support from biology. Ritta and Christina lay in the middle of another unbreakable continuum. They are in part two and in part one. And that, I am sorry to say, is the biological nonanswer to the question of the centuries.

If this claim leaves you with an empty feeling after so much verbiage, I can only contend that the most liberating response to an old mystery is often, The question has no answer because you asked the wrong question. The old question of individuality is rooted in the assumption that events can be pigeonholed into discrete categories. If we recognize that our world is full of irreducible continua, we will not be troubled by the intermediate status of Ritta and Christina.

Dante punished schismatics by dismembering them in hell to exact a physical punishment worthy of their ideological crime: "Lo, how is Mohammed mangled. . . . Whom here thou seest, while they lived, did sow scandal and schism, and therefore thus are rent."

Let us value connections. As Dante analogized physical with ideological separation, perhaps we can learn from the indissoluble union of Ritta and Christina that our intellectual world has its continuity as well.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

Gould, S.J.
"A most ingenious paradox"
Natural History
American Museum of Nat. Hist.
Used by permission

A Most Ingenious Paradox

Stephen Jay Gould

Museum of Comparative Zoology
Harvard University
Cambridge, Mass. 02138

Abstinence has its virtuous side, but enough is enough. I have always felt especially sorry for poor Mabel, betrothed to Frederic the pirate apprentice. On the very threshold of married happiness, she discovers that she must wait another 63 years to claim her beloved at age 84 -- and, as could only happen in Gilbert and Sullivan, she actually promises to wait.

The Pirate King and Ruth, Frederic's old nurse maid and jilted paramour, present the reason for this extraordinary delay. Frederic, wrongfully apprenticed to the pirate band, has reached his 21st year and longs for freedom, respectability and Mabel. But he was formally bound until his 21st birthday, and he was born on February 29. "You are a little boy of five," the Pirate King informs him with glee and expectation of prolonged service. The three principals of the Pirates of Penzance then analyze the complexities of this predicament in the famous paradox song:

> How quaint the ways of paradox
> At common sense she gaily mocks.

The classic paradox presents us with two contradictory interpretations, each quite correct in its own context. Consider our western prototypes, the so-called paradoxes of Zeno: the arrow that can never reach its destination because, at any instant, it must occupy a fixed position; and Achilles who will never catch the tortoise because he must first traverse half the remaining distance and any gap, no matter how small, can still be halved. We delight in paradox because it appeals both to the sublime and whimsical aspects of our psyche. We laugh with Frederic, but also feel that something deep about the nature of logic and life lies hidden in Zeno's conundrums.

Biology too has its classical paradox. It flared as a major issue in the 19th century, probably because scientists then felt that it might be resolved. All the best naturalists struggled with it: Huxley and Agassiz lined up on opposite sides; Haeckel tried to mediate. The 20th century has largely bypassed it, probably because we now realize that no simple answer can be given. Yet, if our fascination with paradox be justified, the question can still enlighten us by virtue of its stubborn intractability.

Physalia, the Portuguese man of war, embodies all this fuss. It is a siphonophore, a relative of corals and jellyfish. The old paradox addresses an issue that could not be more fundamental -- the definition of an organism, and the general question of boundaries in nature. Specifically: are siphonophores organisms or colonies?

Siphonophores belong to the phylum Cnidaria (or Coelenterata). Two aspects of cnidarian biology set the context of our paradox. First, many cnidarians live as colonies of connected individuals -- our massive coral reefs are gigantic congeries made of many million tiny, conjoined polyps. Second, the cnidarian life cycle features a so-called "alternation of generations." The sessile polyp, a fixed cylinder with a fringe of tentacles, is asexual and generates by budding the free-swimming medusa, or "jellyfish." The medusa produces sexual cells that unite and grow into a polyp. And so it goes.

Different kinds of cnidarians may emphasize one of these generations and suppress the other. Of the three major cnidarian groups, the Scyphozoa (or true jellyfish) have abandoned polyps and emphasized medusae, while the Anthozoa (or true corals) have dispensed with medusae and constructed their reefs of polyps and their skeletons. In the third group, the Hydrozoa, many members retain the full cycle, with prominent polyp and medusa. Siphonophores are hydrozoans. The technical literature, not usually noted either for charm or directness, has transcended its usual limitations in this case: amidst a forest of formidable jargon for other parts of cnidarian anatomy, it refers to the polyp and medusa stages of a single life cycle as "persons."

The Portuguese man of war, with its float above and tentacles below, looks superficially like a jellyfish (that is, a single medusa). When studied more carefully, we find that this floating weapon is a colony of many persons, both polyps and medusae. The pneumatophore, or float, is probably a greatly modified medusa (though some scientists think that it may be an even more altered polyp). The "tentacles," though varied and specialized for different roles of capturing food, digestion and reproduction, are not simple parts of a "jellyfish," but modified polyps -- that is, each "tentacle" arises as a discrete person. (Another common siphonophore, <u>Velella</u>, literally the "little sail," but popularly given the lovely name of "by-the-wind sailor," provokes even more confusion. Its persons are few enough and so well coordinated that the colony looks like a simple float surrounded by tentacles -- in other words, like a simple jellyfish. But the float is a medusa person and each tentacle a polyp person.)

If this degree of division of labor among persons impresses you, nature has much more to offer. <u>Physalia</u> and <u>Velella</u> are simple siphonophores, with relatively few types of modified persons. The more complex siphonophores are, by far, nature's most integrated colonies. Their parts are so differentiated and specialized, so subordinate to the entire colony, that they function more as organs of a body than persons of a colony.

Most siphonophores are small, transparent creatures of the open sea. They float at the surface among the plankton or swim actively, usually at shallow depths. As carnivores, they capture small planktonic animals in their net of tentacles. Larger siphonophores, <u>Physalia</u> among them, can ensnare and devour fish of substantial size; as many of us know to our sorrow, they can also inflict painful stings upon human bathers.

Complex siphonophores include an imposing array of well differentiated structures. Their bodies may be roughly divided into two parts, an upper set of bulbs and pumps for locomotion and a lower set of tubes and filaments for feeding and reproduction. Each part contains a series of different polyps and medusae.

Consider first the range of forms and activities assumed by polyp persons. We find three basic types, and a myriad of modifications. The feeding organs, or siphons (hence the group's name — siphonophore means "siphon bearer"), are tubular structures with stomach and trumpet shaped mouth, usually hanging in profusion below the floats and swimming persons. The siphons are minimally modified polyp persons, and we can easily comprehend their origin as complete organisms. All other types of polyps (and medusae) are more highly altered and specialized, and therefore more difficult to link with their original personality. The second order of polyp persons, the so-called dactylozooids ("finger" or touching animals), capture and transport food to the siphons. These are the extended thin tentacles, sometimes more than 50 feet long in Physalia, that carry the painful nematocysts, or stinging cells, and form a transparent web to ensnare prey. They have retained neither mouth nor digestive apparatus, and might easily be taken for parts rather than persons if we could not trace their origin as discrete buds in growth.

These capturing parts often display a remarkable complexity of form and function. The stinging cells may be concentrated into swellings, or "batteries," sometimes protected by a hood. In Stephanophyes, each battery ends in a delicate terminal filament, and contains about 1700 stinging cells of four different types. The terminal filament lassoes the prey and discharges its few stinging cells. If these fail to dispatch the victim, the filament contracts and carries the prey to the far end of the battery itself, where another volley of larger stinging cells transfixes it. If the prey continues to struggle, another contraction moves it up the battery to the near end, where the largest and most powerful stinging cells finally end its torment before passing it along to the siphon for ingestion.

Jennifer E. Purcell has recently presented further evidence that feeding and capturing persons do not form a simple passive network, like the web of a spider, but play an active role in obtaining food (*Science*, vol. 209, 1980, pp. 1045-1047). She found that the stinging cell batteries of two species function as lures by resembling, in both form and motion, small zooplankton that serve as prey for animals eaten by siphonophores. The batteries of *Agalma okeni* look like a copepod with two long antennae; each contracts independently at varying intervals of 5 to 30 seconds, creating a web of motion that simulates the darting and swimming of a copepod school (or whatever you call an aggregation of these tiny planktonic arthropods). To seal the story, Purcell opened the stomachs of *Agalma*, and found the remains of three creatures, all predators of copepods. The batteries of another species, *Athorybia rosacea*, resemble the planktonic larvae of fish and contract rapidly, mimicking the swimming and feeding motions of their models.

Gonozooids, the third type of polyp persons, are reproductive structures. They are usually short, simple tubes, without mouth or motion. But they bud off the medusa persons that make reproductive cells to produce the next generation of siphonophores.

The medusa persons of a complex siphonophore include four basic types: swimming, floating, protection, and reproduction. The swimming organs, or nectophores, are minimally modified medusae -- the upper swimming bells without the lower tentacles. Some siphonophores carry several orderly rows of nectophores; their rhythmic muscular contractions propel the creature, often in elaborate looping trajectories. The passive floats, or pneumatophores, are filled with gas (of a composition near ordinary air) and maintain the siphonophore at the surface or at some intermediate depth. Their origin is a matter of controversy; long interpreted as modified medusa persons, some biologists now regard them as even more elaborately transformed polyps. The two most familiar siphonophores, Velella and Physalia, build large floats but have no nectophores. They therefore move passively on winds and currents, often drifting into bays and beaches in vast accumulations.

The covering organs, or bracts, are the most curiously modified structures of all. They are usually flat, shaped like a prism or a leaf, and so different in form and function from a medusa person that we would scarcely suspect their origin if we could not follow their growth and budding.

The reproductive medusae, or gonophores, are budded off from polyp persons, the gonozooids discussed earlier. In a few species, the gonophores are detached and float in the ocean as independent objects. But they cannot feed and die soon after releasing their sex cells. In most siphonophores, however, the gonophores never separate from the parental colony and remain attached as a kind of sexual organ.

The paradox of the Siphonophora expresses an issue that I have been avoiding, or rather skirting about, in presenting this taxonomy of persons or parts. I have described the various swimming, floating, protecting, feeding, capturing, and reproducing structures as persons -- that is, as individual polyp or medusa organisms. Using evolutionary history as a criterion, this designation is almost surely correct, and accepted now by nearly all biologists. By history, siphonophores are colonies; they evolved from simpler aggregations of discrete organisms, each reasonably complete and able to perform a nearly full set of functions (as in modern coral colonies). But the colony has become so integrated, and the different persons so specialized in form and subordinate to the whole, that the entire aggregation now functions as a single individual, or superorganism.

The persons of a siphonophore no longer maintain individuality in a functional sense. They are specialized for a single task, and perform as organs of the whole. They do not look like organisms and could not survive as separate creatures. The entire colony works as a single entity, and its parts (or persons) move in a coordinated way. Although each nectophore (or swimming bell) has its own nervous system, all are connected by a common nerve tract. Impulses along this pathway regulate the rows of nectophores in an integrated manner that permits the whole colony (or animal) to move with precision and grace. Touch the float of Nanomia at one end and nectophores at the other extremity contract to remove the animal (or colony if you will) from danger. Siphons pump their digested food along the common stem to the rest of the colony, but empty siphons also join in the general peristalsis and food, as a result, reaches the entire colony (or organism) more effectively.

My studied parentheticals of the last paragraph underscore the fundamental paradox. Shall we call the entire siphonophore a colony or an organism -- for it is a colony by evolutionary history, but more an organism by current function. And what of the parts or persons? By history, they are modified individuals; by current function, they are organs of a larger entity. What is to be done?

This issue fueled the great siphonophore debate of 19th century natural history. T.H. Huxley studied siphonophores during his long apprenticeship at sea on H.M.S. Rattlesnake (less celebrated than Darwin's adventure on the Beagle, but another example of the same extended, exemplary and largely extinct style of training in natural history). He interpreted siphonophores as conventional organisms, their parts as true organs and not modified persons. They served as his primary example in a famous essay on the nature of individuality in biology.

Louis Agassiz studied the Portuguese man of war on the shores of his adopted America (I have included his beautiful lithograph of Physalia with this essay), and decided that siphonophores are colonies, their integration a sign of divine handiwork. Ernst Haeckel, artist and naturalist extraordinaire, described the siphonophores collected on that most celebrated of scientific expeditions in oceanography, the voyage of H.M.S. Challenger, 1873-1876. He published with his report a series of plates (including all other illustrations to this essey), unmatched ever since for beauty (though a bit short on accuracy, since Haeckel often added a touch of heightened symmetry for artistic effect). Haeckel also included several plates of siphonophores in his Kunstformen der Natur (Art Forms in Nature) of 1904, the great series of 100 lithographs, with plants and animals arranged in weirdly distorted form and swirling symmetry, in the best tradition of reigning art nouveau, so well embodied in contemporary kiosks of the Paris Métro.

Haeckel's theory of siphonophores would require an essay in itself to explain and explore, but it tried to mediate between Huxley and Agassiz by viewing these creatures in part as colonies (the poly-person theory in his words), in part as organisms (the poly-organ theory). Haeckel also used siphonophores, as Huxley had, to illustrate by dubious analogy his views on the proper organization of human societies. In his Über Arbeitstheilung in Natur und Menschenleben (On the Division of Labor in Nature and Human Life), he compared the simple colonies of other cnidarians with the life styles of "primitive" humans and their limited division of labor for repetitive tasks performed by all: "The wild people of nature, who have remained on the lowest level right to our own day, lack both culture and division of labor -- or they limit division of labor, as do most animals, to the different tasks of the two sexes." He then compared complex colonies of siphonophores with the "advances" that division of labor permits in "higher" human societies -- including modern warfare, whose instruments of destruction "require hundreds of human hands, working in different ways and manners."

Is there any resolution to this ancient debate, any possible mediation between two legitimate criteria that seem to yield opposite results -- the criterion of history supporting the poly-person theory (siphonophores are colonies and their parts are persons), and the criterion of current function upholding the poly-organ theory (siphonophores are organisms and their parts are organs). Could we tip the balance in favor of one view or the other by invoking the third major criterion of natural history -- growth and form?

Growth and form provides us with an *embarras de richesses* by presenting evidence for and against both theories. As strong support for the poly-organ theory, siphonophores develop from a single fertilized egg cell. A siphonophore begins life as an unambiguous person -- should we not regard any later development as an elaboration of this one individual? Moreover, the adult siphonophore acts as a discrete object. Many species have definite and complex symmetry governing all parts considered together. Some Portuguese men of war, for example, come in right and left handed versions.

We may, however, cite equally good arguments for the poly-person theory. Admittedly, each colony begins life as a single ovum, but it then develops a series of entities -- full persons in this view -- by budding from a common stem. This is a familiar mode of growth for many aggregations conventionally regarded as colonies. A stand of bamboo or a field of dandelions may trace its origin to a single seed, yet we usually view each budded stem or flower as an individual.

Second, highly specialized structures sometimes bear vestigial parts that testify to their status as persons. In the poly-person theory, for example, nectophores are medusae that have lost all feeding and digestive parts, retaining only the jellyfish bell. But some nectophores grow rudimentary tentacles; in one species, the tentacles even retain eye spots. Protective bracts are the most modified and specialized of all siphonophore parts, but the covering organs of two species retain a vestigial mouth -- an indication that bracts arose as full medusa persons.

It looks like a toss up again. We might resolve our paradox if growth occurred in either of two ways -- but nature doesn't oblige. If all structures began growth as complete persons with a full set of parts, and then lost unneeded parts as they specialized for swimming, protecting or eating, then the poly-person theory would gain a big boost. If buds from the main stem began as complete persons and then disarticulated -- the bell parts becoming nectophores and the tentacle parts siphons, for example -- then the poly-organ theory would be affirmed. But most specialized parts simply grow as we find them. Nectophores differentiate as nectophores, bracts as bracts. We are immersed in an unresolvable conflict among equally legitimate criteria: discrete buds grow like a person with a specialized set of parts like an organ. What shall we make of a gonophore, for example, the degenerate reproductive medusa budded from a polyp? If it separates from the colony, we may choose to regard it as an organism. But it has no mouth and cannot feed, and must therefore die after releasing the sexual cells -- do we call such a limited breeding machine an individual? And if it remains attached to the colony, as it usually does, should we regard it as any more than a sexual organ?

When an inquiry becomes so convoluted, we must suspect that we are proceeding in the wrong way. We must return to go, change gears, and reformulate the problem, not pursue every new iota of information or nuance of argument in the old style, hoping all the time that our elusive solution simply awaits a crucial item, yet undiscovered.

Nature, in some respects, comes to us as continua, not discrete objects with clear boundaries. One of nature's many continua runs from colonies at one end to organisms at the other. Even the words -- organism and colony -- have no precise and unambiguous separation. We may, however, use the two criteria of our vernacular as a guide. We tend to call a biological object an organism if it is spatially separated from others and if its parts are so well integrated that they work only in coordination with others, and for the proper function of the whole.

Most creatures lie near one or the other end of this continuum, and we have no trouble defining them as organisms or colonies. People are organisms -- even though all multicellular creatures probably arose as colonies about a billion years ago. This origin is so distant, and so much has happened since that we detect no signs of coloniality in our current operation. Thus, we are organisms by any reasonable use of language. Reef-building corals are colonies because each polyp is a complete creature, fully functional in its own right, though attached to its fellows.

But, since nature has built a continuum from colony to organism, we must encounter some ambiguity at the center. Some cases will be impossible to call -- as a property of nature, not an imperfection of knowledge. Consider a progression from evident organisms toward the undefinable center. Human societies are made of organisms; each person is genetically distinct and spatially separate. What about ants? We still opt for organisms even though ants may so submerge their individuality in tightly knit societies that some naturalists refer to an ant hill as a superorganism.

What about aphids? We begin to lose clarity. An aphid clone is all female; each mother grows her young within her own body, without benefit of fertilization. All members of the clone are genetically identical. Are they all separate individuals, or one gigantic evolutionary body with many thousand separate parts, all identical.(one prominent evolutionary biologist has recently urged this second view)?

What about dandelions? Harder still. All stems and flowers are members of a clone; they are genetically identical, and attached to a common underground stem. Is each plant above ground a person or a part? We still usually opt for persons (though many biologists demur) because each plant looks much the same and has a full set of structures.

Finally, then, what about siphonophores? We are now squarely in the middle of a continuum, and we cannot provide a clear answer. The parts of siphonophores are persons by history, organs by current function and a bit of both by growth. Our criteria of separation and independent operation have failed, but we cannot deny a history that still stares us in the face.

Siphonophores do not convey the message -- a favorite theme of unthinking romanticism -- that nature is but one gigantic whole, all its parts intimately connected and interacting in some higher, ineffable harmony. Nature is full of boundaries and distinctions; we inhabit a universe of structure. But since our universe has evolved historically, it must present us with fuzzy boundaries, where one kind of thing grades into another. Objects at these boundaries will continue to confuse and frustrate us as long as we follow old habits of thought and insist that all parts of nature can be pigeonholed unambiguously to assuage our poor and overburdened intellects.

The siphonophore paradox does have an answer of sorts, and a profound one at that. The answer is that we asked the wrong question -- a question that has no meaning because its assumptions violate the ways of nature. Are siphonophores organisms or colonies? Both and neither; they lie in the middle of a continuum where one grades to the other.

The siphonophore paradox is illuminating, not discouraging. It cannot be resolved, but when we understand why, we grasp a great truth about nature's structure. It embodies the same message as that old one about the lady who visits her butcher one Friday morning, seeking a large chicken for the Sabbath meal. The butcher looks in his bin and is chagrined to find that he has but one scrawny animal left. He takes it out with great fanfare and puts it on the scale. Two pounds. "Not big enough," the lady says. He puts it back in the bin, pretends to rummage amidst a large pile of non-existent alternatives, finally pulls out the same chicken, puts it on the scale, and puts his thumb on the scale. Three pounds. "Fine," says the lady. "I'll take them both." Things that seem separate are often the different sides of a unity.

Knoll, A.H. "The Precambrian
Evolution of Terrestrial Life"
In Papagiannes, M.D., The Search
for Extraterrestrial Life:
Recent Developments pp. 201-211
Reidel Publishing Company
Used by permission

THE PRECAMBRIAN EVOLUTION OF TERRESTRIAL LIFE

A. H. Knoll
Department of Organismic and Evolutionary Biology
Harvard University
Cambridge, Massachusetts 02138
U.S.A.

ABSTRACT. Paleontological evidence indicates that terrestrial life existed at least 3500 Ma ago, and it is quite possible that the earliest cells arose well before that time. The early appearance of life on Earth suggests that under appropriate environmental conditions the probability of chemical evolution proceeding to the point of biogenesis may be reasonably high. Most of biological history has been the history of microorganisms, with tissue-grade plants and animals characterizing only the most recent 15% or so of the fossil record. Intelligent life has occupied only the latest instant in geological time. The time table of terrestrial evolution is governed more by the particulars of our planet's physical and biological history than by some universal tempo of evolutionary change. One aspect of terrestrial life that is likely to be universal is the organization of populations into efficient biogeochemical systems.

1. INTRODUCTION

In the 1960's, enthusiasm over the explanatory powers of the newly emergent theory of plate tectonics led to the publication of numerous pronouncements on what is and is not possible in orogenesis. Exasperated by the failure of many authors to reconcile their models with known geological facts observable in the field, the late Marshall Kay coined what he termed the Zeroeth Law of Thermodynamics. Briefly stated, this law holds that anything that *did* happen, *can* happen.

Kay's tongue-in-cheek canon aptly captures the spirit in which the exobiologist must regard the evolution of life on Earth. Earth is the only known example of a planet containing life, so its history is immediately relevant to any discussion of life in the universe. Yet, we have no way of knowing to what extent Earth history may serve as an appropriate model for biological evolution in other solar systems. In the present context, it may in fact be most fruitful to examine Earth history for evidence of evolutionary roads not taken or for developments that need not have eventuated elsewhere in order to gain a greater appreciation of what, if any, elements of terrestrial evolution may potentially be universal.

In this paper, I will examine briefly the geological evidence for early biological evolution, concentrating in particular on the timing of major evolutionary events. It is of some importance to the search for extraterrestrial life to ask whether this time scale is governed by some characteristic tempo of evolution that might be expected to obtain on other planets or whether it reflects a history of biological and environmental events particular to the Earth.

2. THE PALEONTOLOGICAL RECORD

2.1. The Emergence of Life

Of the many evolutionary events whose timing one would like to be able to document in the geological record, none is more central than the origin of life itself. Unfortunately, paleontology can provide only a minimum date for the emergence of biological systems because the oldest negligibly metamorphosed sedimentary rocks known at present already contain compelling evidence for the existence of life at the time of their deposition.

The rocks in question are volcanic-sedimentary sequences preserved in ancient cratonic areas of southern Africa and Western Australia dated at 3400-3500 million years (Ma) old. Silicified chemical sediments in both successions contain large populations of small (2-4 μm diameter) organic walled spheroids, some preserved in various stages of binary division (Muir and Grant, 1976; Knoll and Barghoorn, 1977; Awramik et al., 1983). These microstructures are simple in organization and, consequently, are difficult to identify unequivocally as fossils; however, the regularity of their size frequency distributions, the spatial distribution of populations along bedding planes, surface and internal diagenetic features that compare closely with those characteristic of known microfossils in younger rocks, the absence of mineral grains that could have formed templates for the organic spheroids, and the evidence for binary division collectively suggest that at least some of these Archean microstructures are most parsimoniously interpreted as the remains of once-living microorganisms. This interpretation is strengthened by the recent discovery in both Africa and Australia of filamentous microfossil populations (Awramik et al., 1983; Walsh and Lowe, 1983). The few microfossils unearthed to date suggest a prokaryotic biota of limited morphological diversity, but the available sample is so small--and thus liable to sampling bias--that no strong conclusions can be drawn from these data alone.

Stromatolites can be considered as the trace fossils of microbial mat-building communities, and their presence in the two ancient sedimentary sequences provides further evidence for the early emergence of life. The microorganisms responsible for stromatolite accretion were presumably phototactic and, most likely, were photosynthetic (Walter, 1983). Independent evidence for the presence of photosynthesis is provided by stable isotope analyses of organic carbon preserved in the sequences (Schidlowski et al., 1983). Other geochemical evidence indicates that very little free oxygen was present in the contemporary biosphere (Holland, 1984), from which one can conclude that early biotas were anaerobic.

Combining geological evidence with newly available data on microbial phylogeny obtained by means of comparative molecular and physiological analyses (e.g., Fox et al., 1980; Gest, 1980), one can state with reasonable confidence that life originated early in our planet's history, that the initial diversification of anaerobic Eubacteria and Archaebacteria was rapid, and that by 3400-3500 Ma ago, the Earth contained complex anaerobic ecosystems fueled by photosynthesis (see Knoll, 1985). In the Isua region of southwestern Greenland, still older (approximately 3800 Ma) sedimentary rocks are found. The distribution of reduced carbon in these rocks and, particularly, the evidence for isotopic fractionation in the carbon are consistent with an even earlier diversification of life (Schidlowski, 1984); however, the Greenland rocks are severely metamorphosed and, hence, their unequivocal biogeochemical interpretation is impossible.

Although no record of prebiotic chemical evolution is known to be preserved in Archean sedimentary successions, ancient rocks do fix some broad environmental constraints on the conditions under which life originated. The geological record also supplies one other very important fact about the origin of life--it occurred early. The various syntheses, polymerizations, and other reactions that comprise chemical evolution culminated in the emergence of living systems within a relatively short time, a few hundred million years at most and, conceivably, much more quickly. This suggests that given a physical environment similar to that which characterized the early Earth, the genesis of living systems from simple precursor molecules may not be the extremely low probability event that some have assumed.

2.2. The Emergence of Aerobic Metabolism

The evolution of cyanobacterial photosynthesis initiated profound changes in the Precambrian biosphere and made possible a second radiation of prokaryotic microorganisms, that of the aerobes. The time of origin of "green plant" photosynthesis is not known with any certainty, but John Hayes (1983) has interpreted temporal patterns in the stable carbon isotope ratios of Precambrian organic matter as indicating the presence of oxygen evolving photoautotrophs 2800 to 2900 Ma ago. This may well be a minimum estimate.

Although cyanobacterial photosynthesis may have originated early, it appears that the transition to an aerobic biosphere was a protracted process (e.g., Knoll, 1979; Schopf, 1983). Early oxygenic environments were probably local and short-lived, and organisms capable of exploiting these habitats may have been facultatively aerobic, or amphiaerobic (Schopf, 1983). Mineralogical, gaseous, and organic sinks for photosynthetically produced oxygen were sufficient to keep O_2 levels extremely low over much of the planet's surface. Some mineralogical and isotopic evidence suggests that oxygen grew in biogeochemical importance in the earliest Proterozoic Eon (Figure 1), concomitant with the appearance of large, stable continents (Knoll, 1979; Cameron, 1982, 1983). I have elsewhere discussed the ecological reasons for expecting cyanobacterial productivity to rise as large continental blocks stabilized (Knoll, 1979, 1984). Despite this increase, fully aerobic environmental conditions may not have been established on a stable basis throughout the world's oceans until the mid-Proterozoic (Schopf, 1983; Holland, 1984). Far better geochemical data on the rise of atmospheric oxygen are needed, and until such data become available, one can really only discuss this transition in broad terms. It is clear, however, that aerobic metabolism evolved during this period of environmental change and that the coupled aerobic/anaerobic biogeochemical pathways characteristic of the modern Earth's surface were established at this time.

2.3. Increasing Intra- and Intercellular Complexity

It is now widely accepted that the complex intracellular organization characteristic of eukaryotic organisms originated in part by means of microbial endosymbiosis. In particular, the principal energy-yielding organelles of eukaryotic cells, the mitochondria and plastids, are descended from purple non-sulfur bacteria and cyanobacteria, respectively, that lived symbiotically within a host cell whose genetic organization had differentiated from that of the Eubacteria extremely early in biological history (Margulis, 1981; Gray and Doolittle, 1982; Woese, 1983). Although the geological record can contribute little to the debate over how eukaryotic cells originated, it does provide an estimate of when nucleated cells evolved (e.g., Knoll, 1983). Once again, this is a minimum estimate. The oldest fossils that can plausibly be interpreted as the remains of eukaryotes are large (40-100 μm), thick-walled organic vesicles found in sedimentary sequences up to approximately 1400 Ma old (e.g., Horodyski, 1980). These microfossils are similar in construction and distribution to the cysts and walls of planktonic algae that are commonly preserved in younger rocks, and they have been interpreted in comparable terms (e.g., Tappan, 1980; Vidal and Knoll, 1983). Such an interpretation is reasonable, but it is not entirely iron-clad, because the robust spheroidal envelopes of colonial cyanobacteria can exhibit somewhat comparable morphologic features. Perhaps the strongest evidence that can be marshalled in support of the interpretation of these large 1400 Ma old microfossils as eukaryotes is the observation that demonstrably eukaryotic microfossils become increasingly abundant and diverse in succeeding Late Proterozoic sequences (Vidal and Knoll, 1983). If this interpretation is correct, then the first appearance of eukaryotic microfossils in the fossil record represents not the initial evolution of eukaryotic cellular organization, nor even the origin of algae, but the emergence of algae capable of producing degradationally resistant walls and cysts. Thus, the first eukaryotic microfossils that are recognizable in the geological record may significantly post-date the origin of the group.

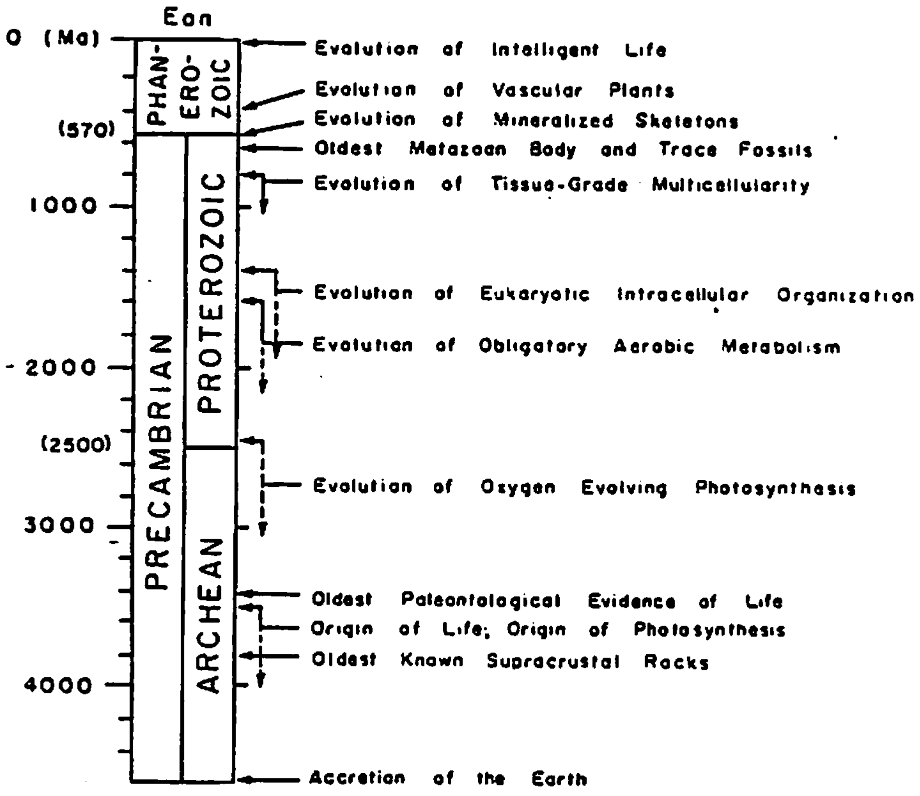
With time, some populations of nucleated cells became organized into multicellular clonal aggregates, and from these evolved tissue-grade seaweeds and animals. Macroscopic carbonaceous compressions are found throughout the Proterozoic geologic record, but many of these may be ripped up fragments of microbial mats. Among the oldest fossils that are incontestably the remains of multicellular organisms are large populations of morphologically regular carbonaceous ribbons preserved in siltstones and shales from northwestern Canada, China, and Svalbard estimated to be approximately 850 to 900 Ma old (Hofmann and Aitken, 1979). The phylogenetic affinities of these earliest megafossils are not clear, but the remains of demonstrable metazoans are found in latest Precambrian (650-600 Ma) sedimentary rocks from several continents (Glaessner, 1979). The oldest animal fossils are casts and molds of large, flat, soft-bodied invertebrates, but preserved tracks, trails, and burrows found in the same sequences indicate that a diversity of small bottom dwelling animals also inhabited latest Precambrian coastal environments.

The evolution of mineralized skeletons, which occurred independently in a number of algal and animal phyla, marks the beginning of the Phanerozoic Eon (literally, the age of visible animals). It has been during this latest approximately 570 Ma interval of biological history that many of the events traditionally discussed as the "history of life" occurred--the diversifaction of animal structure, the invasion of the land surface by plants and animals, the dramatic increase in global biomass and primary productivity that accompanied the radiation of woody vascular plants, and the recent appearance of a species capable of changing the course of evolution by means of technology.

3. THE TIMING OF TERRESTRIAL EVOLUTION

Figure 1 summarizes the timing of major evolutionary events in Earth history. Microorganisms of prokaryotic ultrastructural organization were the sole inhabitants of our planet for the first 50% or more of biological history, and tissue-grade metazoans have been present for only the last 15 to 20%. In Figure 1, hominid history corresponds to the last 130 μm of the time axis, *Homo sapiens* the last 2 μm, and the Industrial Revolution a mere 60 Å at the top of the figure.

On Earth, intelligent life emerged 3500 to 4000 Ma after the initial radiation of living systems. Can one justifiably infer from this that terrestrial history reflects some characteristic time scale for the evolution of intelligent life? Both tacit and explicit support for this view can be found in the literature. For example, more than two decades ago, S. S. Huang (1959) wrote, "One may question the wisdom of using 3 x 10^9 years for t_b [the duration of biological evolution necessary to produce intelligent life] in general, because this value is based on a single case--that of the earth. However, we can argue that the time scale, t_b cannot be greatly shortened because, according to current ideas in biology, the natural selection and evolution of organisms is the result of mutations which are of a random nature and are therefore slow." More recently, M. Cohen (1981) has reiterated this view, stating flatly that "there seems to be a consensus that the



THE TIME SCALE OF TERRESTRIAL EVOLUTION

Figure 1. Chart showing the chronology of biological evolution on Earth. For events whose timing is not known with certainty, the horizontal arrow points to the youngest time of origin consistent with the geological record while the vertical arrow indicates that an earlier origin is probable. For a discussion of dates, see text and references cited therein.

emergence of intelligence on a planet requires about three billion years of relative constancy of stellar output."

Despite these expressions of faith, there is really little basis for the supposition that evolution on Earth has proceeded at some universal rate. In fact, there is little evidence to support the underlying assumption that rates of macroevolutionary change correspond closely to rates of mutation. Clearly, the time scale for the emergence of intelligent life on Earth is governed not so much by the relatively rapid changes in multicellular organisms that have occurred during the Phanerozoic Eon as by the evolutionary events of the long Precambrian era, events such as the evolution of aerobic metabolism, the radiation of eukaryotic cells, and the evolution of regulatory systems to control the expression of different genes in daughter cells within a single colony--a major step in the evolution of tissue-grade multicellular organisms. Of course, one must bear in mind a point argued forcefully by G. G. Simpson (1964), namely that there is no good reason to assume that intelligence is the inevitable product of biological evolution (see also Cloud, 1983).

Controls on the timing of major evolutionary innovations are complex. Each event reflects the influence of previous biological history, as well as the constraints and opportunities presented by the physical environment and other organisms. The relative importance of different factors may vary from one event to another, making generalizations about large scale evolutionary tempo impossible. As a single example, one can consider the emergence of aerobic metabolism. An obvious precondition for the evolution of aerobic physiological pathways was the appearance of environments containing metabolically useful concentrations of oxygen. The required oxygen was produced by cyanobacteria, but it is apparent from the geological record that the establishment of stable oxygenic environments long post-dates the physiological evolution of cyanobacterial photosynthesis (see section 2.2 and references cited therein). The growth of atmospheric and oceanic oxygen levels occurred when levels of O_2 production increased to a point where they exceeded those of oxygen consumption. This, in turn, may well have been a consequence of the physical evolution of the Earth, as large, stable continents flanked by broad shallow shelves and zones of upwelling appeared and, concomitantly, rates of production of reduced gases and minerals by volcanoes abated. Laboratory experiments have repeatedly demonstrated the ability of bacteria to adapt rapidly to environments containing novel substrates and altered environmental conditions, and so it is likely that when oxygenic conditions appeared, microbial populations quickly evolved the capacity to exploit them metabolically. Thus, following the evolution of cyanobacterial photosynthesis, the timing of evolution of aerobic pathways was probably controlled most strongly by the tectonic development of the planet. On Earth, surface environments became oxygenic about 1500 to 2000 Ma after the origin of life. All other factors being equal, under different tectonic conditions, the same event might have happened 1000 Ma earlier, 2000 Ma later, or not at all.

This single example shows that, were all other features of terrestrial evolution left unchanged, a more rapid development of the crust-mantle system could have almost halved the time required for biology to proceed from biogenesis to intelligence. A slower development could have doubled the time, and under some tectonic regimes, oxygen production might never have outstripped consumption. Generally similar arguments could be made for most major evolutionary events of the Precambrian era. The point is that the time scale of terrestrial evolution appears to have been governed less by some intrinsic evolutionary metronome than by the particulars of our planet's physical and biological history, including such extrinsic agencies as directional environmental change and mass extinction.

In a stimulating paper, Carter (1983) recently argued on the basis of what he terms the anthropic principle that the intrinsically expected average time for the evolution of intelligence must be far longer than the period actually required on Earth and, therefore, far longer than the main sequence lifetimes of life supporting stars. If this argument is correct, then the universe probably contains few intelligent civilizations. However, in light of geological observations, it is not at all clear that one can reject the alternative hypothesis that because of constraints related to the particular course of Earth history, evolution on this planet has proceeded slowly relative to rates that might be possible elsewhere (see Maddox, 1984). Indeed, the history of our own planet leads one to entertain the idea that the timescale determining early events of biological evolution are significantly constrained by the physical evolution of planets. If this is true, and if, as seems reasonable, planetary development is related to the larger scale evolution of solar systems, then the observed similarity in the time required for the evolution of intelligence on Earth and the main sequence lifetime of the Sun may have a physical basis and, therefore, not be a coincidence that can be dismissed as being *a priori* improbable. If this possibility can be admitted, then according to the probability equations of Carter (1983), intelligent life may be widespread in the universe. The fact is, of course, that without any knowledge of evolutionary rates and trends on even a single other planet, one cannot determine whether the appearance of intelligent life midway through our Sun's main sequence lifetime should be regarded as unusually early, unusually late, or unusual at all.

4. ARE THERE UNIVERSAL FEATURES OF TERRESTRIAL EVOLUTION?

Having sided with those who argue that both the tempo and products of evolution may vary on different planets, I close by asking whether any features of terrestrial life are likely to be universal. I would argue that the answer is "yes", but that the universal features do not necessarily include biochemical pathways, morphological designs, or specific taxa. More likely the universal aspects of life on Earth include the organization of metabolically disparate populations into biogeochemical systems capable of transferring energy and cycling biologically important elements. It is only by means of these systems that life can be maintained for evolutionarily long periods of time. Therefore, it is likely that any planet that sustains life will contain series of taxa having biogeochemically complementary metabolisms for the rapid biological cycling of materials, as well as a tectonically active planetary surface capable of maintaining the longer term biogeochemical balance of biospheres.

5. ACKNOWLEDGMENTS

My research on problems of Precambrian biological evolution has been supported by NSF Grant BSR 82-13682. S. Goldberg drafted the evolutionary timescale figure.

6. REFERENCES

Awramik, S. M., Schopf, J. W., and M. R. Walter. 1983. 'Filamentous fossil bacteria from the Archean of Western Australia' *Precambrian Res.* **20**: 357-374.

Cameron, E. M. 1982. 'Sulphate and sulphate reduction in early Precambrian oceans' *Nature* **296**: 145-148.

Cameron, E. M. 1983. 'The start of sulfur oxidation in continental environments about 2.2 x 10^9 years ago' *Science* **221**: 549-551.

Carter, B. 1983. 'The anthropic principle and its implications for biological evolution' *Phil. Trans. R. Soc. Lond.* A**310**: 347-363.

Cloud, P. E. 1983. 'Cosmobiology' *Quart. Rev. Biol.* **58**: 57-60.

Cohen, M. 1981. 'Stellar influences on the emergence of intelligent life' Pp. 115-118. In: Billingham, J., ed., *Life in the Universe*. MIT Press; Cambridge, Massachusetts.

Fox, G. E., *et. al.* 1980. 'The phylogeny of prokaryotes' *Science* **209**: 457-463.

Gest, H. 1980. 'The evolution of biological energy-transducing systems' *FEMS Microbiol. Lett.* **7**: 73-77.

Glaessner, M. F. 1979. 'Precambrian' Pp. A79-A118. In: Robison, R. A., and C. Teichert, eds., *Treatise on Invertebrate Paleontology, Part A: Introduction*. Geol. Soc. Amer. and Univ. Kansas Press; Boulder, Colorado, and Lawrence, Kansas.

Gray, M. W., and W. F. Doolittle. 1982. 'Has the endosymbiont hypothesis been proven?' *Microbiol. Rev.* **46**: 1-42.

Hayes, J. M. 1983. 'Geochemical evidence bearing on the origin of aerobiosis, a speculative hypothesis. Pp. 292-301. In: Schopf, J. W., ed., *Earth's Earliest Biosphere, Its Origin and Evolution*. Princeton Univ. Press; Princeton, New Jersey.

Hofmann, H. J., and J. D. Aitken. 1979. 'Precambrian biota from the Little Dal Group, Mackenzie Mountains, northwestern Canada' *Can. J. Earth Sci.* **16**: 150-166.

Holland, H. D., 1984. *The Chemical Evolution of the Atmosphere and Oceans*. Princeton Univ. Press; Princeton, New Jersey.

Horodyski, R. J. 1980. 'Middle Proterozoic shale-facies microbiota from the lower Belt Supergroup, Little Belt Mountains, Montana' *J. Paleontol.* **54**: 649-663.

Huang, S. S. 1959. 'The occurrence of life in the universe' *Amer. Sci.* **47**: 397-402.

Knoll, A. H. 1979. 'Archean photoautotrophy: some alternatives and limits' *Origins of Life* **9**: 313-327.

Knoll, A. H. 1983. 'Biological interactions and Precambrian eukaryotes' Pp. 251-283. In: Tevesz, M. J. S., and P. L. McCall, eds., *Biotic Interactions in Recent and Fossil Benthic Communities*. Plenum Publ. Corp.; New York.

Knoll, A. H. 1984. 'The Archean/Proterozoic transition: a sedimentary and paleobiological perspective' Pp. 221-242. In: Holland, H. D., A. F. Trendall, eds., *Patterns of Change in Earth Evolution*. Dahlem Konferenzen. Springer Verlag; Berlin.

Knoll, A. H. 1985. 'Patterns of Evolution in the Precambrian Era' *Paleobiology*. In Press.

Knoll, A. H., and E. S. Barghoorn. 1977. 'Archean microfossils showing cell division from the Swaziland System of South Africa' *Science* **198**: 396-398.

Maddox, J. 1984. 'New twist for anthropic principle' *Nature* **307**: 409.

Margulis, L. 1981. *Symbiosis and Cell Evolution*. W. H. Freeman and Co.; San Francisco.

Muir, M. D., and P. R. Grant. 1976. 'Micropaleontological evidence from the Onverwacht Group, South Africa' Pp. 595-604. In: Windley, B. F., ed., *The Early History of the Earth*. Wiley; London.

Schidlowski, M. 1983. 'Evolution of photoautotrophy and early atmospheric oxygen levels' *Precambrian Res.* **20**: 319-335.

Schidlowski, M., Hayes, J. M., and I. R. Kaplan. 1983. 'Isotopic inferences of ancient biochemistries: carbon, sulfur, hydrogen, and nitrogen' Pp. 149-186. In: Schopf, J. W., ed., *Earth's Earliest Biosphere, Its Origin and Evolution*. Princeton Univ. Press; Princeton, New Jersey.

Schopf, J. W., ed. 1983. *Earth's Earliest Biosphere, Its Origin and Evolution*. Princeton Univ. Press; Princeton, New Jersey.

Simpson, G. G. 1964. 'The non-prevalence of humanoids' *Science* **143**: 769-775.

Tappan, H. 1980. *The Paleobiology of Plant Protists*. W. H. Freeman and Co.; San Francisco.

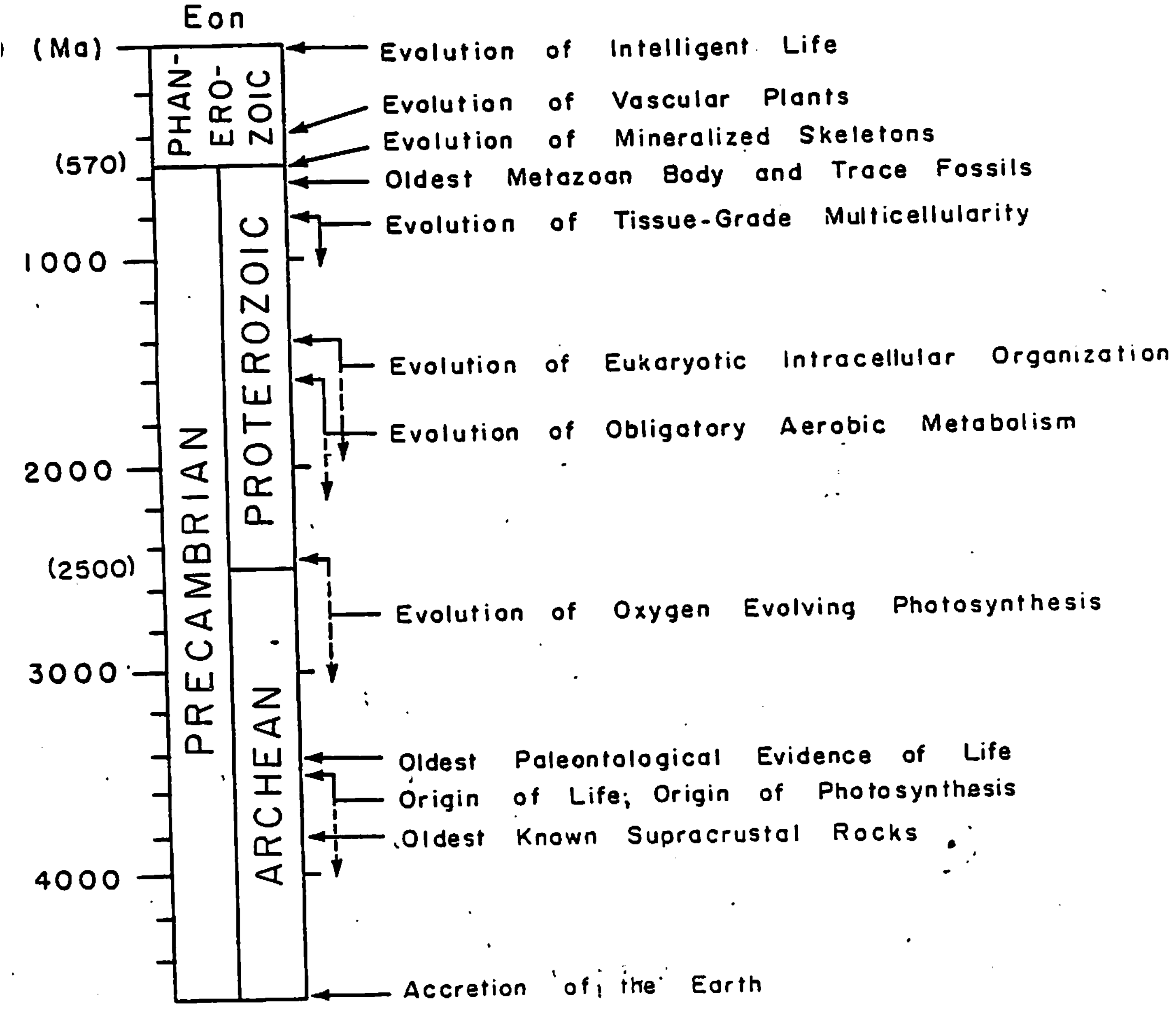
Vidal, G., and A. H. Knoll. 1983. 'Proterozoic plankton' *Geol. Soc. Amer. Mem.* **161**: 265-277.

Walsh, M. M., and D. R. Lowe. 1983. 'Filamentous microfossils from the 3.1-3.5 billion-year-old Swaziland Supergroup, Barberton Mountain Land, South Africa' *Lunar and Planetary Sci. XIV*, Abstr. Pp. 814-815.

Walter, M. R., 1983. 'Archean stromatolites: evidence of the Earth's earliest benthos' Pp. 187-213. In: Schopf, J. W., ed., *Earth's Earliest Biosphere, Its Origin and Evolution*. Princeton Univ. Press; Princeton, New Jersey.

Woese, C. R. 1983. 'The primary lines of descent and the universal ancestor' Pp. 209-233. In: Bendall, D. S., ed., *Evolution from Molecules to Men*. Cambridge Univ. Press; Cambridge, U.K.

THE TIME SCALE OF TERRESTRIAL EVOLUTION



Glaessner, M.F.
The Dawn of Animal Life pp. 1-42
Cambridge University Press
Used by permission

1

Precambrian life and its environment: a review

1.1 Theories about the origin of life and its environment

In the eighteenth century, James Hutton, one of the founders of the science of earth history, saw 'no vestige of a beginning, no prospect of an end'. Much has been learned in the 200 years which have elapsed since these words were written. We still see them as containing rather more than a grain of truth but we give them a different meaning from that which was in Hutton's eighteenth-century mind. Our aim now is to understand ourselves and our environment as part of an ongoing process of change. The model, still frequently if unconsciously applied because it is so close to the individual's experience, of birth, youth, maturity, old age and death, is not helpful when we consider superindividual processes in the universe. It does not imply a tidy and tedious calendar of a journey from point A, labelled 'beginning' to point B, 'the end', of which we wish to compile a partial diary. The aim of our science today is to compile a record of interactions which help us to understand the complexity of ourselves and our environment as revealed by ongoing research at ever increasing rates. The part of history of the universe which is our subject is animal life and its increasing diversity. To be sure, it had a beginning in the sense that there was a time when there were no animals, but, as Hutton recognized, the record contains no vestiges of a beginning and it would be a futile strategy to search for them. The history of life on earth may come to an end: we know that our technology has given us the means to end its existence but no end is in sight, at least at this point in time. James Hutton wrote that 'time is to nature endless and as nothing'. It has taken two centuries of thought and experiment to give content and definition to the concept of geological time which Hutton intuitively felt as 'endless' compared with experienced human life time and therefore as 'nothing' within our experience of time. The study of a process implies the study of rates of change in time and therefore the Huttonian endlessness must be measured and the supposed nothingness must be filled with evidence of events. We operate now with a geological time scale (Van Eysinga 1975, Cohee, Glaessner & Hedberg 1978). The order of magnitude of divisions of this scale, required for the study of the emergence and diversification of animals, is from 1 to 1000 million years (m.y.). The major division of the scale with which we are concerned is labelled Precambrian, for reasons which will be examined in detail in Chapter 4.

The origin of life is the result of circumstances and events close to the beginning of Precambrian time. In the absence of material vestiges of the first organisms we must briefly consider current theories about the beginning and about the distinguishing characters of what is called life. The question of the origin of life as a subject for scientific investigation was first raised in theoretical discussions independently by Haldane (1929) and Oparin (1924). The basis for a unified theory of the abiogenic origin of life is its basic chemical uniformity, together with well-founded assumptions about the possible composition of the atmosphere, hydrosphere and lithosphere at the relevant time, 3500–4500 m.y. ago. Haldane invoked from the then available experimental evidence the concept of an abundance of organic molecules ('sugars and some of the molecules from which proteins are built up') which are necessary for the formation of the elementary building materials of all organisms. He postulated that 'they must have accumulated till the primitive oceans reached the consistency of hot dilute soup' (Haldane in Bernal 1967, p. 246). This came to be referred to as the primitive soup or 'primordial broth'. Oparin considered the behaviour of colloidal solutions, pointing to the importance of cell walls and heterotrophic nutrition (from pre-existing organic compounds) in early organisms. Several years later, the experiments of Miller (1953) (see also Miller & Urey 1959) showed that energy discharge (electric sparks) in oxygen-free gas mixtures produced amino acids, the building stones of proteins, and similar experiments with other energy sources and gas mixtures produced further essential biological molecules.

We can now summarize the theoretical requirements for the origin of life as follows.

(1) The formation of organic molecules from which the basic components of living systems (proteins, lipids, carbohydrates, nucleic acids) can be built up. Some of their precursors occur in interstellar space. They have been identified by radioastronomy (see Chang 1981). Others occur in meteorites, particularly those belonging to the class of carbonaceous chondrites, or they could exist in comets. Some can be and have been produced experimentally from what would have been common materials in the primitive earth crust and a reconstituted early atmosphere, in the absence of free oxygen.

(2) The atmosphere had to be reducing but the arguments continue about the need for a strongly or minimally reducing one, and accordingly about its composition. It may have contained hydrogen, water, ammonia, methane; or water, nitrogen and carbon dioxide; or carbon monoxide, hydrogen sulfide, ammonia and methane.

(3) A source of energy was required for the reaction of synthesis of prebiotic molecules. This could have been lightning which is essentially an electric spark, as in the Miller experiment, or ultraviolet radiation which was then unimpeded by the subsequently formed ozone screen in the upper atmosphere, or volcanic heat, as in hot springs, or solar heat, as in a desert where dew falls or where tide pools exist on an arid shore. Ionizing radiations have also been considered.

(4) 'The continued supply of external energy by radiation, etc., led to a build-up of substances of high free energy content on the early earth. The products must have remained dissolved or suspended in the water... The aqueous solution and suspension was the medium in which according to Oparin and Haldane life developed – Haldane's "dilute soup"' (Broda 1978, p. 27). No quantitative estimates of the possible dilution (or concentration) of this 'soup' are available. The need for an additional mechanism for localization of synthetic processes was felt by Bernal as early as 1951 when he suggested that some of the ubiquitous clay minerals, or alternatively quartz crystals, may have acted as catalysts for polymerizations. The difficulty of having an ocean turned into a soup bowl so as to bring organic molecules close enough together to make them react in quantity is only slightly mitigated by modern geological thought about the existence of many 'small' oceans in the early stages of crustal evolution. The unease about the organic-filled prebiotic ocean led to speculation about hot springs or lagoons as the geographical environment for the origin of life. However, with this model intercommunication of possible sites of reaction would be lost and the biochemical unity of the nascent life would be harder to understand.

(5) An unsolved problem is the beginning of the mechanisms confining the molecules required to form individual organisms. All living organisms form cells, except viruses which 'are not relevant to the origin of life' (Broda 1978, p. 30), being essentially adaptations of biochemical systems to parasitism. Cells are enclosed in membranes consisting of proteins and lipids. The cell membrane not only keeps together and separates from the environment the basic materials for life processes (metabolism, self-reproduction, bioenergetics) 'but is also the place of the mechanism for

transport into the cell and out of the cell, including the pumps for active transport' (Broda 1978, pp. 43–4). Protective cell walls consisting of sugars and peptides or of cellulose or chitin may surround the cell membrane but do not occur in primitive animal cells, with grave consequences for the possibility of their preservation in rocks. The recognition of the problem of separation of individual protobionts from the primeval broth led Oparin to fruitful experimentation with coacervates, self-congregating units in colloidal systems. It also led to preparation of 'microspheres' by Fox (see Fox & Dose 1972) and his followers. They are natural aggregations of proteinoids with interesting cell-like properties and appearance.

(6) The final problem is which – if any – of the processes of prebiotic chemical evolution, demonstrated to be possible by experiment, have actually led to the origin of life. There is some agreement about the actual occurrence of chemical evolution (Calvin 1969). It suggests a natural selection among chemical processes and products, leading to the survival of those most suited to continued self-reproduction through adaptation to the constraints and resources of the environment. This formulation points to two aspects of the problem which are under active investigation but still unsolved. Self-reproduction of all organisms is based on automatic transmission of information. This is contained in the genetic code, incorporated in the structure of the DNA molecule and uniform throughout the existing organic world. Its origin is still obscure. The second problem is the demonstration that what could have happened in the earliest stages of the history of the biosphere has actually happened. This would be possible if organic remains claimed to be produced by early organisms could be reliably distinguished as formerly living organisms from prebiotic chemical fossils.

The discovery of apparent fossil remains of single-celled organisms resembling bacteria and algae in rocks more than 3000 m.y. old (Barghoorn & Schopf 1966, Pflug 1966, Schopf & Barghoorn 1967) is significant. Critical reviews by Schopf (1975*a*) have thrown doubt on the status of these fossils as organisms and indicate their resemblance to shapes assumed by abiotic, in this instance probably prebiotic, organic matter. They have been referred to, with other similar finds, as 'fossil-like objects' (Schopf 1978). However, fossil sedimentary structures of organic origin known as stromatolites occur in rocks 3500–2500 m.y. old in Australia, southern Africa and Canada (Walter 1983). It is now well known that they are formed by the life activities of Prokaryota (bacteria and blue-green algae) (see Walter 1976, 1977, 1978, 1983). Undisputed fossil remains of microorganisms which produced stromatolites have been observed. There is evidence (Walter 1978, 1983) that some Archaean stromatolites were built by photosynthetic Cyanophyta and by purple or green bacteria which assimilate carbon dioxide but do not produce free oxygen (Broda 1978). Many formerly existing forms of life, including microbes, must have been supplanted sooner or later by more efficient ones. This will become increasingly important as our story progresses but it causes the greatest difficulty in reconstructing microbial life of the past, a problem of classifying organisms whose observable life processes are vastly more significant than their morphological configurations.

Before we proceed to the next stage in the history of life it is necessary to consider the placing of the documented events in precise time scales and also to understand fossilization processes which convert the results of life processes into components of sedimentary rocks.

1.2 The Precambrian, its subdivisions, and the dating of biohistoric events

The construction and application of a geological time scale is almost as old as the science of geology. It implies a division of the uninterrupted flow of time during the entire history of the earth by a sequence of distinctive events. The model for it was, either consciously or unconsciously, human history with its division into antiquity, the Middle Ages and modern times, and its subdivisions such as Egyptian or Chinese dynasties, or the reigns of European monarchs, or socio-economic changes. There are also unique marker events like the Renaissance or the French Revolution, and the divisions of prehistory according to technological advances such as stone age, bronze age, iron age. The scales define primarily only *sequence* (the placing of other events as earlier or later than the key event), not *chronology* (dating of events and consequently duration of intervals). The geological chronology is being established with ever increasing accuracy by relating geological events to clock-like, fixed-rate processes such as radioactive decay. This makes it possible to calibrate the sequential stratigraphic scale. This scale was established early, by the application of the law of superposition to stratigraphic sequences. Except for later tectonic disturbances, lower strata must have been laid down before higher ones. The study of fossils contained in the rocks showed more than 150 years ago that they facilitated distinction between rocks occurring in sequence, and comparison of such sequences. These observations led soon to the development of biostratigraphy which eventually found in irreversible evolution and concomitant extinction of forms of life a guide for the recognition of increasingly finer divisions of the geological time scale. Biostratigraphy could be applied as far back in history as abundant fossils could be found, that is, to the beginning of the Cambrian

Period, the earliest part of the Palaeozoic Era. This and the two succeeding eras (Mesozoic and Cenozoic) were grouped together as the Phanerozoic Eon whose name was taken to mean the time of visible (or obvious) life.

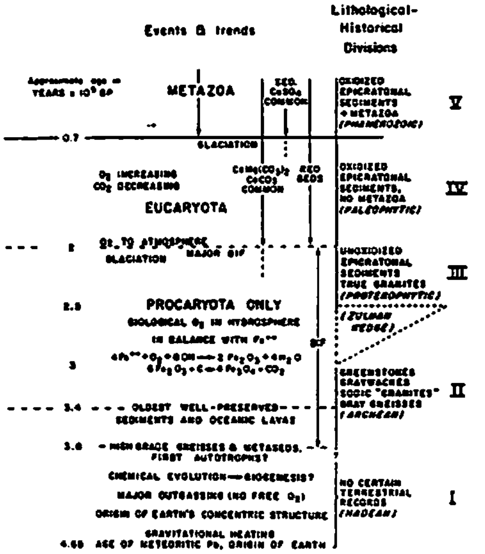
When these names were proposed, no obvious traces of life were known from rocks older than those of Cambrian age, hence they were lumped together as Precambrian, of unknown duration. When calibration indicated that Precambrian time comprised about seven-eighths of geological time and when interest in Precambrian rocks increased for various other reasons, an extension of the more detailed time scale became necessary. The Subcommission on Precambrian Stratigraphy (a unit of the International Union of Geological Sciences) is preparing proposals which will eventually be submitted to an International Geological Congress for ratification. A proposal for a conventional division of Precambrian time into eons, the Archaean and the Proterozoic, has been approved by the subcommission. The boundary is defined by its chronometric (isotopic, commonly but loosely called radiometric) age, at 2500 m.y. ago. Possible subdivisions of the Archaean are still under discussion. The oldest presently known rocks are about 3800 m.y. old. Nothing is known from geological data about events between that age and the formation of the earth about 4600 m.y. ago. The dating of this event is based on cosmological evidence from meteorites and moon rocks. There is increasing information about the Archaean lithosphere, its composition and dynamics (Windley 1976, Tarling 1978). By the end of that eon, the earth's crust appears to have evolved significantly, to approximate the present lithosphere more closely, and so the concept of an eon boundary at about 2500 m.y. (James 1978), which is also significant for the history of life, is likely to find wide acceptance. A subdivision of the following eon, the Proterozoic, into three eras has also been proposed. The terms Early, Middle and Late Proterozoic will be used here for these divisions. Their proposed boundaries at 1600 and 900 m.y. (Sims 1980) are times of significant geological events documented in many continents.

The end of the Precambrian is necessarily the beginning of the Cambrian. Its precise definition is being considered by an international working group, which intends to base it on biostratigraphic changes in a stratotype section which has yet to be designated. This means that it will be based on the same principles as the Phanerozoic scale (Hedberg 1974) and that its age will be established by the dating of that concrete boundary event rather than by being based on an abstract chronological definition alone. The age of the base of the Cambrian is generally considered as approximately 550–570 m.y. The uncertainty may be greater than this range and further isotopic dating is urgently required. This boundary and its implications will be considered further in Chapter 4.

From the viewpoint of biostratigraphy, the Proterozoic is the time when stromatolites became abundant, while animal fossils from at least the greater part of this era are absent or rare. The middle of Precambrian time, between the age of the oldest rocks and 570 m.y. would be at about 1600 m.y. The informal concept of a mid-Precambrian time to which reference is made occasionally would encompass some early and some middle Proterozoic time from, say, about 1800 to 1400 m.y. ago. As we shall see, this is thought to be the time of the emergence of eukaryotic cells, with nuclei and organelles, from their prokaryote, non-nucleated predecessors. This important evolutionary step either rapidly led to or actually coincided with the first appearance of animal protists (Protozoa). The refinement of meiosis in cell division, sexual reproduction with its concomitant increase in evolutionary potential, colonial organisation and the origin of Metazoa, multicellular animals developing distinct tissues, followed sooner or later. The order of these events is based firmly on biological theory (see Margulis 1970, Patterson 1978) and will be further discussed in section 1.4. The palaeontological evidence and its geochronological dating are uncertain. This uncertainty encompasses nearly the entire mid-Precambrian interval, if not the entire Early and Middle Proterozoic, up to 1000 m.y. The evolutionary rates which determine the timing of these early stages of metazoan prehistory are unknown and cannot be known as long as remains of fossil animals, body and trace fossils, from sediments of Early and Middle Proterozoic age are unknown. Cloud (1976*a*) believes that the first Metazoa appeared about 700 m.y. ago which is the approximate age of the first unquestioned remains of metazoan body fossils. Older traces of animals exist but regrettably little is known about them. The first manifest metazoan fossil representatives are not what we would expect on theoretical grounds, i.e. small, planuloid or at the most advanced possible level, primitive coelenterates. It is more than likely that such animals could not have left traces in the sediments but a time span should and could be allotted to them without undue appeal to speculation. This has been attempted. The curve depicting oxygen accession to the atmosphere – oxygen being essential for oxidative respiration – is occasionally invoked as a means of dating biological events. However, the physical theory of oxygen (and ozone) accession, developed by Berkner & Marshall (1965) and brought to the attention of geologists by Cloud (1968), Fischer (1965, 1972) and Rutten (1962, 1966) who have modified it, does not result in a time scale. Berkner & Marshall placed the 'Pasteur point' (1% of the present atmospheric level (PAL) of oxygen) at the beginning of the Cambrian, which was thought to be the time of the oldest fossil animals. We know now that this is incorrect. Not enough is known quantitatively to calibrate precisely in terms of geological time the Berkner–Marshall curve of oxygen accession.

Cloud (1976*a*,*b*) divided the post-Archaean (Proterozoic) time, on geochemical grounds of change from unoxidized to oxidized epicratonal sediments and on grounds of biological change from Prokaryota to Eukaryota, into Proterophytic and Palaeophytic Eras, with a boundary at about 2000 m.y. (Figs. 1.1, 1.2). There is at present no indication of general acceptance of this biostratigraphic scale and nomenclature. Even soundly based and euphonious names are not necessarily generally accepted. Time scale divisions are conventions depending on consensus rather than on strictly scientific generalizations of a multitude of empirical data. International discussions on the time scale of the Precambrian and the naming of its subdivisions are planned for the 1980s. After a review of the Late Proterozoic Metazoa, the time framework for that period will be re-examined (Chapter 4).

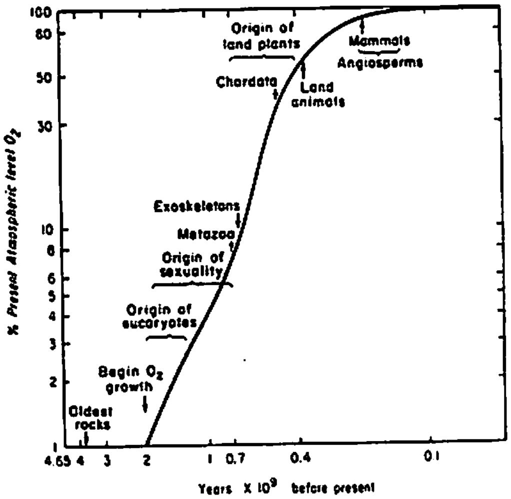
Fig. 1.1. Major divisions of earth history and related aspects of biological and geochemical evolution. (BIF – banded iron formations; BP – before present.) Adapted from Cloud (1976*b*).



1.3 The preservation and recognition of organic remains in rocks: chemofossils, fossil organisms, pseudofossils

Before proceeding with the study of the history of life to the point in time when animal remains become recognizable, we have to consider the nature of existing documentation of the life of the past and the interpretation of these documents. There is an analogy between them and the material evidence of human history. The analogy is valid in the sense that as we go back to the oldest times, the record of both kinds of history becomes more difficult to decipher and to interpret. It also becomes inherently more incomplete, full of gaps and false clues which are apt to mislead. Much has been written about the inherent incompleteness of the Phanerozoic fossil record. Estimates have been made of the percentage of living species and higher taxa which could not have been preserved. Much can be learned from the well-known Phanerozoic record about fossilization

Fig. 1.2. Apparent timing of events in biospheric evolution compared with hypothetical levels of atmospheric oxygen. Both scales logarithmic; curve highly generalized, excursions from the mean likely. From Cloud (1976*b*).



potential, but as will be seen in Chapter 4, evolution of mineralized skeletons has made a quantum jump at the Precambrian–Cambrian transition. Their absence has not rendered the Precambrian entirely unfossiliferous as had been expected earlier, but it has necessitated much effort in building up a specific methodology of discovery and interpretation of Precambrian fossils which is still incomplete. The availability of fossils for the study of life history is a function not only of the fossilization potential of the biota of different periods but also of the exploration effort. In Darwin's time so few Phanerozoic fossils had been discovered that he was unable to make any significant use of them for his theories of evolution by natural selection. One hundred years later, no theoretical discussion of the course of evolution is complete without reference to palaeontological evidence. The rapid increase in the exploration and study of Precambrian geology, stimulated by the abundance of mineral resources formed during that time span, and the recognition of its great length, has revealed an abundance of Precambrian fossils. In general terms they are quantitatively and qualitatively inferior to those of later Phanerozoic age but they are highly significant for several reasons. It is true that with greater age of rocks there are greater chances of their alteration by repeated tectonic and magmatic processes leading to the destruction of organic remains but this is not the general fate of all Precambrian rocks, contrary to what had been expected. Exploration has revealed that many of them have remained not too strongly altered to contain organic remains. There are thick sequences of sediments previously considered as unfossiliferous (and often in the absence of clear evidence dated as Palaeozoic) which, with the application of new biostratigraphic methods, have been shown to be of Precambrian age and which have since been dated chronometrically. New methods of the physical sciences, electron microscopy and modern analytical chemistry, are leading to discoveries of great significance which are still being evaluated. The occurrence of organic matter in ancient rocks has been known for a long time and has proved puzzling, in the absence of other traces of fossil organisms and in the face of authoritative views of their presumed absence. Molecular biology, a dominant branch or viewpoint in biological sciences, has led to suggestions of a 'molecular palaeontology' (Zuckerkandl & Pauling 1962, Zuckerkandl 1965, Florkin 1966, McLaughlin & Dayhoff 1973, Broda 1978). They are based on evolutionary studies of proteins (haemoglobins, cytochromes) from mainly microbes and vertebrates, and evaluation of their genetically controlled alterations with respect to phylogeny and time. Biochemists have considered chemical evolution (Calvin 1969). The evolution of bioenergetic processes (Broda 1978) is not a supplement but an essential substrate for these studies, with far-reaching consequences for the understanding of the first appearance of animals whose metabolic activities and bioenergetic requirements are different from those of their predecessors.

The palaeobiological significance of the study of stable organic compounds isolated from ancient rocks as 'chemical fossils' has been recognized in organic geochemistry (e.g. McKirdy 1974). The main problems are the possibility of abiotic or prebiotic origin of some organic compounds, the danger of contamination when the analyst has to deal with minute quantities, the possibility of migration of fluids through even weakly porous or fractured rocks during long periods of time, and the alteration of chemical structures by diagenesis and incipient metamorphism. In recent years there has been much progress in the understanding of the alteration of insoluble organic matter in sedimentary rocks (kerogen) which must be syngenetic and coeval with the enclosing rock. The possible linking of chemical marker fossils with evolutionary events is particularly significant. The early occurrence of photosynthesis is partly deduced from such studies but the dating by chemofossils of the change from prokaryote to eukaryote cells or from fermentation to oxidative respiration is still essentially in the future.

The recognition of fossil organic remains on the basis of their configuration rather than their chemical composition is particularly difficult in ancient rocks. These difficulties have been discussed by Hofmann (1971, 1972) who proposed an elaborate classification which distinguishes 'pseudofossils' and 'problematica' or 'dubiofossils' from definite remains of organisms ('fossils'). Here we shall distinguish between difficulties of recognition caused by (1) alteration of rocks, (2) similarities of organic and inorganic configurations and (3) problems of evolutionary morphology affecting fossilization.

(1) Alteration. It is generally true that Precambrian rocks are more highly altered, i.e. more affected by diagenetic, metamorphic and tectonic processes than younger rocks. Fossils have been found in younger metamorphic rocks but their preservation under conditions of elevated temperatures and pressures or metasomatic processes is dependent on a degree of robustness of embedded organic remains which was not generally attained, at least by animal bodies, until Phanerozoic time. The search for remains of fossil organisms is generally confined to special classes of altered sedimentary rocks. Early diagenesis can enclose organic remains so as to seal them hermetically against further alteration. This can occur in a matrix of chert (cryptocrystalline silica) or in films of organic matter (kerogen). It may also be possible to recognize traces of organic activity (burrowing, coprolitic pellets) in the fabric of mildly altered sediments.

Few if any of them will be unquestionably biogenic and most of them will be therefore classed as 'problematica'. The odd shapes of the results of some tectonic deformations (boudinage) and of concretions will generally make them recognizable as pseudofossils.

(2) Similarities of organic and inorganic configurations. Geometric regularity of configurations which could not be immediately identified as crystals was taken as an indication of organic origin, at least in the early stages of the search for Precambrian fossils. Clay pellets often resemble fossil shells, infillings of drying cracks can resemble worm burrows or bodies, the outlines of angular shards of dried clay layers have been combined to produce misleading reconstructions of imaginary fossil arthropods. Fractured or sectioned glass bubbles in volcanic tuffs may strikingly resemble sponge spicules or chambered Foraminifera or Radiolaria. Configurations such as vertical stalks ending in lobate structures have turned out to be sand volcanoes. Jellyfish-like concentric structures on bedding planes may be found, on closer examination, to have vertical feeder channels and to be formed by escaping gas or fluids. Others, without stalks, have turned out to be pressure marks or moulds of crystal rosettes. The possibility of mechanical origin such as current effects or bedding lineations must be rigorously excluded or at least made highly improbable by observation of specific characteristics before similar configurations are admitted as genuine fossils. An interesting example of the need for caution in evaluating the biogenicity of reputed Precambrian fossils arises from a comparison of an early distribution table of groups of fossils in the Precambrian (Glaessner 1966, Fig. 2) with present knowledge. Ten stratigraphic units from various countries were included as containing either trace fossils or (animal) megafossils. Two of them were considered as questionably fossiliferous. Three supposed trace fossil occurrences and the two questionable ones are now known to contain only configurations of mechanical origin, one contains megascopic algae mistaken for animal remains. Some trace fossils from the Precambrian–Cambrian Vindhyan rocks of India may be Cambrian and one figured trace was apparently caused by the rasping of the radula of a living snail on the surface of the rock. The origins of other trace fossils are still being questioned. Cloud (1968, 1973) has acted consistently as an invaluable sceptic in the evaluation of Precambrian configurations considered as organic remains. They require continued strenuous efforts to verify their organic origin, more than do younger fossils.

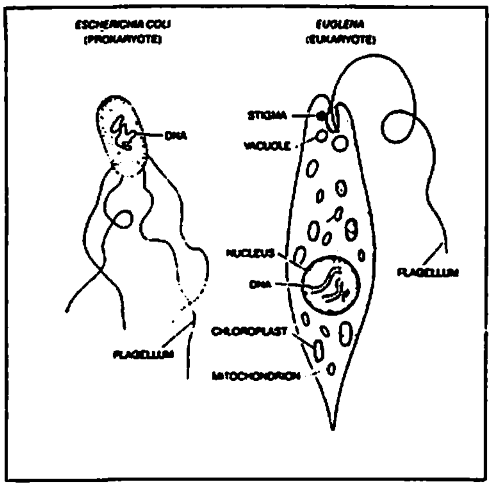
(3) Evolutionary problems. Not long ago the course of evolution before the Cambrian was undocumented and even the fauna of the Early Cambrian was poorly known. This explains why the occurrence in the

Precambrian of large eurypterids ('reconstructed' from the outlines of large clay flakes) or of articulate brachiopods (which subsequently proved to come from inliers of Cambrian rocks in Precambrian terrains) was accepted. True shells or skeletons originated about the time of the transition from Precambrian to Cambrian (see Chapter 4). The real problem in finding and interpreting Precambrian fossils is the fact that these animals were without shells or skeletons (soft-bodied). The preservation of soft-bodied organisms poses many problems of the reconstruction of their original form and of the taxonomic assessment of their variability. These problems and the methods assisting in their solution are rarely discussed in the literature. A recent authoritative statement on fossilization (A. H. Müller 1979) devotes one paragraph to the preservation of soft parts in some 70 pages of text. Meanwhile the wasteful labour of description and publication of pseudofossils mistaken for fossils continues.

1.4 The cell: organization and life processes. The 'kingdoms' of organisms

All organisms, except viruses, are built up from cells or organized as cells. They are separated from their environment and able to communicate with it through the structure of their surrounding membranes. To fulfil their tasks, including the basic one of self-reproduction, the cells cannot be simple pieces of bioengineering. The simplest kind, a bacterial cell, has been compared in complexity with a chemical factory, with an information content, at the atomic level, of 10^{12} 'bits' or 'yes/no decisions' built into it. There are among living organisms two basic kinds or levels of cell organization, the prokaryote cells of the bacteria and the blue-green algae (which are now frequently referred to as cyanobacteria), and the eukaryote cells which make up all other organisms. The prokaryote cells are mostly small (1–10 μm), without a nuclear membrane or endoplasmic reticulum (cytoplasmic membrane system), no mitochondria (enzyme-containing bodies), no chloroplasts for photosynthesis; they divide by binary fission (Fig. 1.3). The eukaryote cells are generally larger (10–100 μm, exceptionally larger by up to three orders of magnitude) with a nuclear membrane and with the organelles which are absent in the prokaryotes; cell division occurs by mitosis followed by meiosis, and the cells have complexly but uniformly structured flagella or cilia. They are primarily aerobic and produce oxygen in the process of photosynthesis. 'The numerous and fundamental differences between eukaryotic and prokaryotic organisms...have been fully recognized only in the past few years. In fact, this basic divergence in cellular structure which separates the bacteria and the blue-green algae from all other cellular organisms, probably represents the greatest single

Fig. 1.3. Comparison of prokaryote and eukaryote cells. The drawings represent a bacterium compared with a single-celled green alga. Its cell contents include ribosomes and food storage bodies (unlabelled). The chloroplasts are also known as plastids and the stigma as the eye spot (Tappan 1980, pp. 895–6). From Schopf (1978). Copyright © 1978 by Scientific American Inc. All rights reserved.



	PROKARYOTES	EUKARYOTES
ORGANISMS REPRESENTED	BACTERIA AND CYANOBACTERIA	PROTISTS, FUNGI, PLANTS AND ANIMALS
CELL SIZE	SMALL, GENERALLY 1 TO 10 MICROMETRES	LARGE, GENERALLY 10 TO 100 MICROMETRES
METABOLISM AND PHOTOSYNTHESIS	ANAEROBIC OR AEROBIC	AEROBIC
MOTILITY	NONMOTILE OR WITH FLAGELLA MADE OF THE PROTEIN FLAGELLIN	USUALLY MOTILE, CILIA OR FLAGELLA CONSTRUCTED OF MICROTUBULES
CELL WALLS	OF CHARACTERISTIC SUGARS AND PEPTIDES	OF CELLULOSE OR CHITIN, BUT LACKING IN ANIMALS
ORGANELLES	NO MEMBRANE-BOUNDED ORGANELLES	MITOCHONDRIA AND CHLOROPLASTS
GENETIC ORGANIZATION	LOOP OF DNA IN CYTOPLASM	DNA ORGANIZED IN CHROMOSOMES AND BOUNDED BY NUCLEAR MEMBRANE
REPRODUCTION	BY BINARY FISSION	BY MITOSIS OR MEIOSIS
CELLULAR ORGANIZATION	MAINLY UNICELLULAR	MAINLY MULTICELLULAR, WITH DIFFERENTIATION OF CELLS

evolutionary discontinuity to be found in the present-day living world' (Stanier, Adelberg & Doudoroff 1963). The most unexpected fact observed in Precambrian palaeontology, the dominance of remains of bacteria and blue-green algae in rocks dated as representing the first half of Precambrian time, is an expression of this first and highly significant 'evolutionary discontinuity'. It raises three questions: (1) What evolutionary step or steps could have bridged this discontinuity? (2) How are these evolutionary events related to changes in the environment? (3) What evolution, if any, took place at prokaryote level, during Early Precambrian time, through more than 2000 m.y. since deposition of the oldest fossiliferous rocks?

(1) One perhaps surprising answer to the first question was given by the symbiotic theory of the origin of the eukaryotes (Margulis 1970). Margulis considers the classical view that the more highly organized forms of life originated from bacteria and blue-green algae 'by the accumulation of selectively advantageous mutations [as] inconsistent with many facts'. The symbiotic theory asserts that prokaryotes acquired such organelles as mitochondria in the form of symbiotic aerobic bacteria; flagella and cilia were similarly pre-existent as spirochaetes and incorporated in prokaryotes where they evolved eventually to participate not only in locomotion but also in advanced reproductive processes ('mitotic apparatus'). Some blue-green algae were originally endosymbiotic in protozoan-like heterotrophs to form plastids and finally chloroplasts as sites of photosynthesis. The details of this theory and the factual evidence need not concern us here. They were lucidly presented by Margulis, with due consideration of available palaeontological and geohistorical data documenting as far as possible what could have happened at the relevant time in the history of life. G. E. Hutchinson remarked in his foreword to Margulis' work (1970, p. xvii) that 'not every interested biologist may accept all her suggestions'. This has proved correct (see Broda 1978, pp. 132, 180 for contrasting views; Taylor 1974, with reply by Margulis 1975; Cavalier-Smith 1975, discussed in detail by Cloud 1976*a*). A review of relevant new observations and interpretations appeared too late to be considered here (L. Margulis, *Symbiosis in cell evolution*, W. H. Freeman & Co., San Francisco, 1981). It strengthens and widens the basis for the endosymbiosis theory. At least some of its postulates have proved widely acceptable, fruitful for further research, and significant for the origin of animals, an event not likely to have occurred much later than the first appearance of eukaryotic cells in mid-Precambrian time. The various hypotheses about their evolution and its timing were clearly and fairly reviewed by Tappan (1980, pp. 84–96).

(2) The answer to the second question is that these evolutionary events at cell level do not presuppose any rapid and fundamental changes in the

environment. It may be said that they have gradually laid the foundations necessary for an expansion of the biosphere in mass, diversity and in its influence on the history of the earth. The lithosphere and hydrosphere of at least Early Proterozoic time were, according to geochemical and tectonic data, probably not very different from their present composition and dynamic state, though not as rich in diverse ecological sites (niches) for various life forms. The most distinctive development in the physical environment in Early Precambrian time was the accumulation in the atmosphere of free oxygen. According to Cloud (1976*a*) it amounted to about 3% PAL at the time of the origin of the eukaryotes. There is biogeochemical evidence of photosynthesis occurring at the time of the formation of the oldest sedimentary rocks. It is still practised by a few bacteria. They never produce free oxygen but the blue-green algae and the plants do. The oxygen in the present atmosphere is the product of this life activity, with only minor contributions from photodissociation of water. The process is efficient enough: 'It can be calculated, for example, that the offspring of one gram of algae (assuming unlimited space and nutrients and a reasonable rate of cell division) could photosynthetically produce an amount of oxygen equal to that of the present atmosphere in less than 40 days' (Schopf 1975*b*, p. 55). But of course space and nutrients were not unlimited and so it took longer. How much longer is still a controversial question; estimates vary from the assumption of a fully oxygenic atmosphere in Early Precambrian time, which conflicts with the fact that unoxidized minerals of that age were found where they would have been exposed to the atmosphere when deposited, to Cloud's estimate (1976*a*) of only about 50% PAL at about the end of Cambrian time. Cloud has repeatedly drawn attention to a relevant geological fact. Banded iron formations, presently the major source of industrial iron production, had their worldwide maximum development about 2000 m.y. ago and then they disappeared from the record almost completely. They were formed under conditions different from those which the fully oxidized redbeds required for their deposition. Typically their colour is due mainly to quartz grains coated with iron oxides. Cloud considers this change as the first indication of conditions which permitted a change of life processes from fermentation to oxidative respiration, a change which now occurs in facultatively aerobic organisms at about 1% PAL in the organisms' environment (the so-called Pasteur point). This change in the organisms' energy acquisition occurred certainly under considerable selection pressure, as oxidative respiration is more than twice as efficient than fermentation, in terms of free energy produced by the basic reaction.

(3) The fact that Prokaryota (bacteria and blue-green algae) dominate

the Precambrian fossil record does not necessarily indicate very low rates of evolution during this long time. Studies of detailed stratigraphic sequences are still too few and too new for final conclusions to be drawn. Not all difficulties of interpretation of fossilization processes and of their influence on original micromorphology have been overcome. There are indications of evolutionary changes in size and complexity (Schopf 1977). Schopf has repeatedly stressed the importance of the development of mitotic and meiotic reproductive mechanisms of the eukaryote cell for significant increases in diversification rates. This is very likely but the palaeontological evidence for their occurrence is by no means unequivocal, because of problems in interpreting detailed observations on fossil remains of single or possibly dividing cells. What is probably more important is the need for the evolution of complex and sophisticated new biochemical pathways within the structural framework of the prokaryote cell to enable it to make efficient use of the changed environment for its metabolic and bioenergetic needs. The time required for this biochemical evolution under conditions of a significant increase in the oxygen content of the atmosphere and the hydrosphere cannot be estimated but it was probably long in geological terms. In other words, the morphological simplicity of the fossil and existing prokaryotes may conceal a lengthy evolution of their basic life activities and adaptations and possibly also extinctions of less well adapted life forms.

Energy in the form of light is a precondition of life which (to use a profound thought expressed variously at different times by such eminent physical scientists as Boltzmann, Schrödinger or Broda) interposed itself into the flux of radiant energy from the hot sun to the warm earth and on to cold space. Oxygen is the basis for oxidative respiration, the efficient transformation of radiant energy to the energy of the chemical bond in the adenosine triphosphate (ATP) molecule of living organisms. The accession of oxygen to the atmosphere as the result of photosynthetic use of radiant energy is favourable for the further evolution of life under two conditions. Some of it had to be available to protect living organisms from harmful ultraviolet radiation, by forming an ozone screen in the atmosphere. The ability of some prokaryote cells to repair radiation damage has been proved experimentally but could hardly have been effective enough to assure their dominance during 2000 m.y. of early history of the biosphere. Other protective mechanisms have been invoked, such as the shielding effect of 10 m of water or of the calcareous deposits on algal mats which are preserved throughout the geological record as stromatolites. There have been objections to the presumed efficacy of these mechanisms. On the other hand, there have been objections to the suggestion that the ozone screen began to operate as soon as there was free oxygen in the atmosphere. If it appeared late, the apparent delay in the occurrence of abundant life on dry land could be explained. Abundant plant or animal fossils are not found in non-marine sediments older than 400–500 m.y. However, there may have been other reasons for the late appearance of abundant terrestrial life. The conclusion must be that the dating of the screening out of lethal ultraviolet radiation is still controversial: 600–650 m.y. ago according to Cloud (1968), or the time when as little as 0·1% PAL of oxygen was present (Ratner & Walker 1972).

The second precondition for the use of free oxygen in life processes is avoidance of its toxicity. While some biologists believe that the necessary protective enzyme systems could have evolved in a geologically insignificant time span, concomitantly with the respiratory pathways, Cloud (1976*a*) assumes a slow start of respiration in oxygen-poor environments and designates the onset of redbed sedimentation about 2000 m.y. ago as the time 'when enzymatic mediation of that particular atmospheric pollutant became efficient enough to tolerate O_2 levels above ~ 1% P.A.L.'

We can now turn to the question of the time of the first appearance of eukaryote organisms. Cloud (1976*a*), who has made the most comprehensive studies of the beginnings of biospheric evolution, concluded that the origin of eukaryotes may have occurred at any time from 2000 m.y. ago onward but that the oldest really persuasive morphological evidence for that development is not found until about 1300 m.y. ago. Before discussing the evidence for the earliest occurrence of animal remains in the geological record, the distinction between animals and other forms of life has to be considered. All animals and plants (all organisms other than bacteria and blue-green algae) have eukaryotic cells. Some exist as single cells or as colonial cell aggregates. They are known as Protista, or if they are known to live as animals do, as Protozoa. The multicellular animals with cell systems differentiated and functioning as tissues are known as Metazoa. It is the almost generally held view that the Metazoa evolved from the Protista. Within the living Protista the distinction between plants and animals is not significant but to some extent only a matter of terminology or semantics. There are single-cell organisms that sometimes nourish themselves by photosynthesis, and at other times swim about digesting food particles (Broda 1978, p. 138, with reference to Margulis 1970). Apart from those Protista which can function either as plants or as animals, photosynthesis in plants and phagotrophy in animals, which consequently develop locomotion, are generally valid distinctions. They are not important enough to exclude, for example, Phytomonadina (plant flagellates) from discussion in a standard textbook on protozoology (Grell 1973) and they could not generally be applied to extinct Protista but only to living organisms whose feeding habits can be observed. The old taxonomic term Protophyta has become obsolete; yet it does not make sense biologically to remove the photosynthetic flagellates from their systematic grouping with other algae as long as even the prokaryote Cyanophyta are commonly referred to as blue-green algae (Hanson 1977). The problem concerning us is the tracing of possible evolutionary pathways from early eukaryotes to Metazoa. The extent to which fossil remains can be used to test the historical reality of the theoretical conclusions concerning animal ancestors will depend on our ability to recognize fossil evidence of the distinctive animal *functions*: locomotion, and ingestion of food particles, with its corollary, the excretion of structured fecal matter. This is possible only at the grade of Metazoa. The recognition of fossil Protozoa as their possible ancestors is excluded in most instances by the difficulties of their preservation. While plant microbes have cell walls of cellulose, animal cells have less resistant membranes consisting of proteins and lipids, or chitin. Hardening of walls by agglutination of foreign particles, mineralization with silica or carbonates, or sclerotization of chitin are secondary phenomena which did not occur before Late Precambrian time.

The decay of the cell content during fossilization can leave a residue looking like intracellular membranes or organelles and a dark spot near its centre can resemble a nucleus. This problem has been investigated by experimental 'fossilization' of known organisms (Oehler 1976, Francis, Margulis & Barghoorn 1978), with different results according to the methods used. Similar difficulties cloud the interpretation of fossil evidence of cell division in eukaryotes as distinct from binary fission of prokaryotes, and much other palaeocytological detail (Knoll & Barghoorn 1975, Oehler, Oehler & Muir 1976). A significant conclusion from experimental testing of earlier observations states: 'Although it is a logical conclusion that the eukaryotic level of cell organization arose before 680 m.y. ago, there is no definite cytological evidence for an earlier date of appearance of eukaryotes in the fossil record'. (Francis *et al.* 1978, p. 97). On this negative evidence we cannot disregard 'logical conclusions'. Work on morphology, palaeobiochemistry and fossilization of Early and Middle Proterozoic microfossils is progressing and the latest conclusions about the timing of evolutionary and environmental events may be subject to further revision in the light of new data and advances in methods of study and experimentation. The richest and best preserved assemblages of such microfossils are from the Gunflint and Bitter Springs stromatolitic and non-stromatolitic cherts of North America and Australia, respectively. The first named is about 2000 and the second about 850 m.y. old and there are other similarly significant microfossils of intermediate age which have been discussed by Cloud, Schopf and others in numerous review articles. A relevant and cautiously expressed conclusion seems to be that the time of deposition of the Gunflint chert 'corresponds approximately to the transition from an oxygen-poor atmosphere to one in which there were significant levels of free oxygen', about 1800 to 2000 m.y. ago; a critical period in the history of life during which 'enhanced morphological experimentation produced the problematical microorganisms now found in the Gunflint' (Awramik & Barghoorn 1977, p. 128). While according to these authors the Gunflint microbiota is wholly prokaryotic, the Bitter Springs microbiota is likely to contain eukaryotes which may have first appeared 1400–1500 m.y. ago (Schopf & Oehler 1976). Comparison of early eukaryotes suggests affinities to green algae to most observers, but others have different opinions (see Knoll & Golubic 1979) or refrain from definite taxonomic assignments. The earliest microfossils resembling possible Protozoa are from the Chuar Group of the Grand Canyon, Arizona, dated at more than 650 and less than 850 m.y. old (Bloeser *et al.* 1977), from southwest Brazil (Fairchild, Barbour & Haralyi 1978), from Saudi Arabia (Binda & Bokhari 1980) and from the uppermost Riphean of Greenland (Vidal 1979) and Sweden (Knoll & Vidal 1980). The finds have been correlated with rock sequences which are about 700–800 m.y. old. These microfossils are described as Chitinozoa ('heterotrophic protists or primitive metazoans') by Bloeser *et al.*, but as 'Chitinozoan-like' by Vidal and Binda & Bokhari.

We have seen that the simplest eukaryotes are the unicellular (or colonial unicellular) Protista and that among them the first manifestations of animal life must have occurred. They are unlikely to have left any recognizable structural marks on fossils or in rocks. The proposers of the multikingdom concepts of classification including a kingdom Protista have recognized this as a confederation of those eukaryotes which are neither Metaphyta nor Metazoa (Whittaker 1969, Whittaker & Margulis 1978). Living Protozoa which could serve as models for the first animals must be among the zooflagellates and ciliates. They are not likely to be found among the Rhizopoda of which those with preservable skeletons (Radiolaria, Foraminifera) appear late in the geological record, with clearly 'primitive' representatives initiating a comparatively well documented, complexly radiating evolution not earlier than in Phanerozoic time. Hanson (1977) considers the Protozoa as polyphyletic, i.e. as derived from various colourless algal protists. His conclusions on the origin of the Metazoa will be discussed together with others in Chapter 3. Here we state only that it is not unlikely that groups of Protista which could have included ancestors of Metazoa have survived to the present. If not, if these

ancestors have become extinct, it is unlikely that we shall obtain any information about them from the geological record. In any case, what occurred among the Precambrian Protista was probably not so much a morphological change as one towards dominant ingestive nutrition. This form of nutrition combined with motility is found in phytoflagellates which are also equipped for photosynthesis. Hence evolutionary transition to animal protists could have occurred through loss of photosynthetic organelles (plastids).

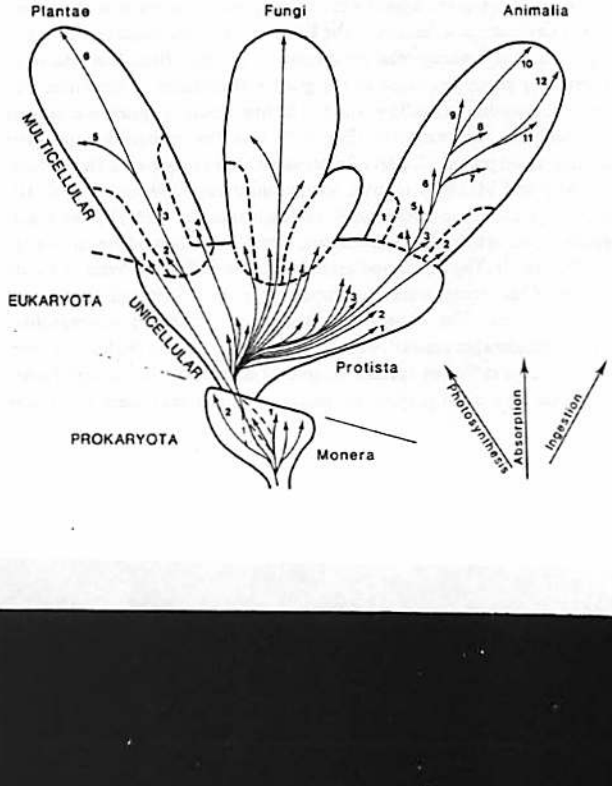


Fig. 1.4. The five kingdoms of organisms. After Whittaker (1969, Fig. 3, simplified). (Copyright 1969 by the American Association for the Advancement of Science.) Numbered subdivisions, more or less relevant to discussions in the text, are as follows. Monera (or Prokaryota) – 1, bacteria; 2, Cyanophyta (or Cyanobacteria). Protista – 1, Ciliata (or Ciliophora); 2, Sarcodina (Foraminiferida, Radiolaria, amoebae); 3, Zoomastigina (animal flagellates). Plantae – 1, Rhodophyta; 2, Chlorophyta; 3, Charophyta; 4, Phaeophyta; 5, Bryophyta; 6, Tracheophyta. (Alternative classifications (Whittaker & Margulis 1978) include 1–4 and the lower Fungi with the Protista, as 'protoctista'.) Animalia – 1, Mesozoa; 2, Porifera (or Parazoa); 3, Coelenterata (Cnidaria and Ctenophora); 4, Platyhelminthes; 5, Aschelminthes; 6, Tentaculata (or Lophophorata); 7, Chaetognatha; 8, Annelida; 9, Mollusca; 10, Arthropoda; 11, Echinodermata; 12, Chordata. (In the later classification (Whittaker & Margulis 1978) the Pogonophora and Hemichordata were grouped with the Chaetognatha between 11 and 12.)

The desirability of excluding in a formal classification scheme the Protista from both the plant and animal kingdoms is obvious, despite its inconvenience. This, as Whittaker (1969, p. 154) remarks, 'is not so much the fault of the systematist as faults of the living world as a subject of classification'. The stream of evolution widens and divides, diversifying to fill available niches (Fig. 1.4). The level of organization above that of Protista is characterized by multicellularity and tissue development and the basic divisions are defined on the basis of nutrition because of its far-reaching consequences on structure and function. Nutrition is absorptive in Fungi. Therefore, and because of structural adaptations to their mode of life, they deserve separation from the other kingdoms. In the kingdom Plantae *sensu stricto* it is mainly photosynthetic, with nutrients obtained from water, 'soil' and air. The organisms are basically autotrophic and sessile and should include the multicellular 'higher' algae. The fact that nutrition is by ingestion of captured food in the kingdom Animalia raises the multifarious problems of locomotion. It has been suggested that with the Metazoa this kingdom should also include the small groups of Mesozoa and Parazoa, with absent or limited tissue differentiation. The Mesozoa are without fossil representatives and will not be considered here. The Parazoa (sponges) are here included with the Metazoa as animals. For the rest, widely accepted formal classifications of the kingdom Animalia will be used in this study. It is not its aim to criticize or develop them.

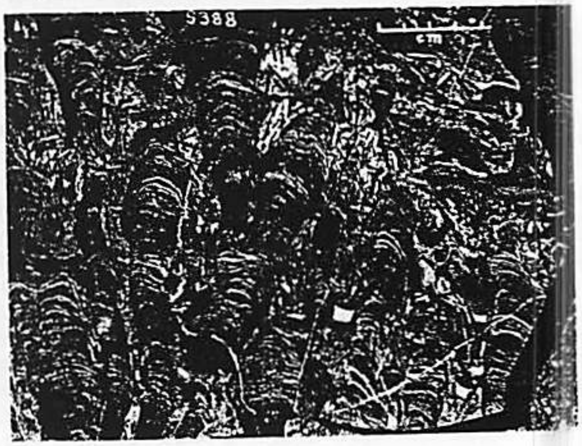
1.5 Proterozoic fossils and environments

At the beginning of the Proterozoic Eon, autotrophic prokaryotes (bacteria and blue-green algae) were in existence. The most striking and generally characteristic feature of the Early Proterozoic fossil record is the first abundant and widespread occurrence of stromatolites. It is related to and probably partly the cause of the greater abundance of limestone and dolomite deposition in shallow water. The first result of systematic studies of Precambrian stromatolites (Fig. 1.5) was the recognition of their biostratigraphic potential. Early Proterozoic (Aphebian), Middle Proterozoic (Early and Middle Riphean, stromatolite assemblages I and II–III, respectively) and Late Proterozoic (assemblages III and IV) were distinguished (see articles by Semikhatov, Preiss and Donaldson in Walter 1976, Chapter 7). The numbered assemblages (specified by Preiss in a note on p. 368 of his article) were described first in the USSR and later found in other regions. The subsequent studies of Aphebian stromatolites revealed considerable similarities in their configuration on the level of form genera but also sufficient specific structural differences to confirm rather than negate their stratigraphic potential, provided that their study was carried out in necessary detail. Due consideration had to be given to environmental, preservational and diagenetic factors. At the same time, considerable efforts were made to find and describe the fossil remains of the algal communities which had contributed to the construction of stromatolites (Schopf *et al.* 1977, Schopf & Prasad 1978) as well as non-stromatolitic occurrences of Proterozoic fossil algae. It was reported that about 20 distinctive assemblages of microfossils occur in biostratigraphically useful sequences, about 10 of them in strata which are more than 1000 m.y. old.

The abundance and diversity of stromatolites declined sharply in Late Precambrian time, from about 800 m.y. ago onwards (Gebelein in Walter 1976). The fact that they had flourished until that time points to the absence or insignificance of Metazoa and Metaphyta in their environment (photic, subtidal to intertidal) in which shelf-type limestones have been deposited. In subsequent times the predation by Metazoa together with competition by red and green algae and other marine plants confined the stromatolite builders mostly to hypersaline and freshwater environments. This occurred rapidly during Early Palaeozoic time and continued to the present when they survive as relicts.

The major primary producers of food resources for heterotrophic

Fig. 1.5. Stromatolites. *Gymnosolen ramsayi* Steinmann, from a boulder in the Late Proterozoic Tapley Hill Formation, Flinders Ranges, South Australia. Photo by courtesy of Dr W. V. Preiss.



organisms in the sea are the photoautotrophic phytoplankton organisms. It is difficult to establish their fossil record through the great time spans of the Proterozoic. It has been suggested that their quantitative fluctuations in Phanerozoic times had a decisive influence on the evolution of the marine fauna (Tappan & Loeblich 1971). The present level of our knowledge does not permit reliable extrapolation into Precambrian time, for a number of reasons. Planktonic algae belonging to extant taxa make their first appearance in Phanerozoic strata: the chrysophycean coccolithophorids probably in the Jurassic, the diatoms in the Cretaceous, the dinoflagellates possibly in the Silurian or Permian and definitely in the Triassic. This makes the identification of Precambrian microfossils as members of extant taxa of Protista of the marine plankton impossible, in contrast to some extant Cyanophyta which can be recognized as belonging to it, with considerable probability, in Precambrian time. On the other hand, large numbers of organic walled microfossils are known which are placed in the 'Group Acritarcha' (Evitt 1963). This name indicates only that their affinities are unknown. Precambrian acritarchs are rarely older than Late Proterozoic. They are mostly simple spherical bodies with thin, smooth, organic envelopes. They are classed in a subgroup Sphaeromorphitae (Downie, Evitt & Sarjeant 1963). Their biological functions are unclear. Often assigned to this group, the genus *Chuaria* Walcott, 1899 has a remarkable geographic range throughout North America, northern Europe, the USSR, Iran, China and India. It occurs generally in considerable abundance in certain horizons of the Late Proterozoic and Vendian (Ford & Breed 1973, Hofmann 1977, Hofmann & Aitken 1979, Vidal 1981). It has a remarkable size range, from 0.5 mm to 5 mm diameter, according to Ford & Breed. A closely similar Late Proterozoic (mid-Vendian) fossil *Beltanelloides* Sokolov (Fig. 1.6) measures from 5 to 45 mm (Sokolov 1974). Its size makes it even more difficult to understand its growth and function as a phytoplankton organism. Sokolov considered it originally to be a 'medusa' but as it is a simple, originally spheroidal or more likely discoidal envelope without openings or appendages it is certainly not a coelenterate nor any other kind of animal, as Sokolov came to note later.

Fig. 1.6. *Beltanelloides sorichevae* Sokolov (= *Beltanelliformis brunsae* Menner), from a bore at Leino, eastern Russian Platform, Vendian, Redkino Series. From Sokolov (1972*b*). Magnification: × 3.

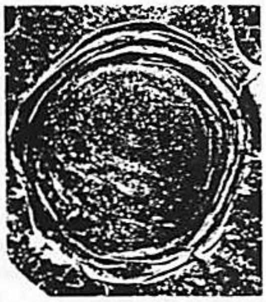
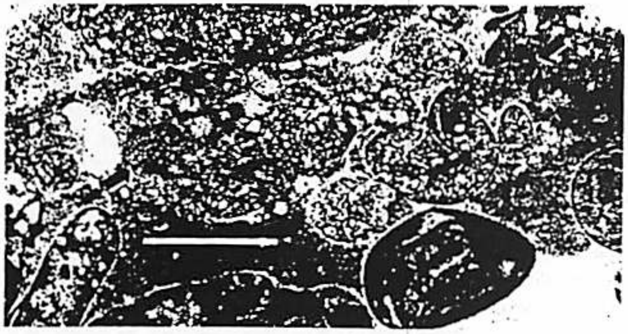


Fig. 1.7. Convoluted fecal string (arrowed) of a sediment-feeding metazoan. Upper Riphean, Ural Mountains. From Sabrodin (1971). Scale bar, 1 mm.



The earliest fossils which could be considered as animal remains (see review in Glaessner 1983) are unfortunately poorly known and incompletely described. Microfossils resembling Chitinozoa which may represent Protozoa or Metazoa have been mentioned (p. 20). Other microfossils occurring in the Lakhandin 'Series' of Siberia (950–1000 m.y. old) appear to be small metazoan organic structures. They were extracted by B. Timofeev and T. N. German from laminae of organic matter. They are still undescribed. All other metazoan fossils of that age are traces of organic activities rather than body fossils. A sigmoidal string of what appear to be fecal pellets about 0·37 mm in diameter (Fig. 1.7) from the middle Zilmerdak 'Series' (lower part of Upper Riphean of the Ural Mountains) was figured by Sabrodin (1971, 1972), together with a possible burrow. Clemmey (1976) described what he believes to be burrows from the Middle Roan formation of Zambia (about 1000 m.y. old) but Cloud (1978*a*; Cloud, Gustafson & Watson 1980) considers them as Recent termite burrows. Other questionable traces of bioturbation were recorded by Squire

(1973) from the Brioverian of the Channel Islands, considered to be about 750 m.y. old. Beer (1919) described a spiral measuring 13 × 11 mm preserved in relief on the lower surface of a slab of Rhotas Limestone of the uppermost Vindhyan of northern India which should be re-examined as it appears to be a trace fossil of metazoan origin. However, a Cambrian age of the Upper Vindhyan is possible. Supposed trace fossils from the Belt Supergroup in Montana (USA) have been redescribed as megascopic algae, aged 1300 m.y. (Walter, Oehler & Oehler 1976). A possible trace fossil from the Lower or Middle Riphean equivalents in the Ukraine, from a formation resting unconformably on a weathering horizon on volcanics which are 1000–1400 m.y. old, was described as *Rugoinfractus ovruchensis* by Palij (1974). Durham (1978, p. 37) considered it to 'merit special attention'. It is, however, also comparable with infillings of desiccation (syneresis) cracks which have been described repeatedly from Precambrian and younger rocks. The interpretation of a presumed medusoid fossil *Brooksella canyonensis* Bassler was questioned by Cloud (1968). Its subsequent assignment to the trace fossils (Glaessner 1969*a*) on the basis of its resemblance to *Asterosoma* was made more likely by Cloud's discovery of a second specimen. This interpretation as a possible 'infaunal deposit feeder' (Brasier 1979) was apparently confirmed by Kauffman (in Kauffman & Steidtmann 1981, p. 925) who gives the age as 1100–1300 m.y. although it is often stated as 900–1000 m.y.

The record of metazoan life from Early and Middle (?) Proterozoic rocks is apparently non-existent and that from 1000 to about 680 m.y. ago (Upper Riphean and equivalents) exists but is extremely poorly documented. The coincidence of the absence of metazoan remains in the early Middle Proterozoic and of definite eukaryotes favours the view that eukaryotes originated in mid-Proterozoic time or about the beginning of the Late Proterozoic, not in Early Proterozoic nor in Vendian time. We cannot yet assert as a fact that deposit feeders existed about 1000 m.y. ago but there are indications which make it more likely than the assertion of their absence. A fuller understanding of the distribution of structured fecal pellets in sediments should reveal the presence of sediment feeders more clearly than bioturbation, which may require more efficient structural adaptations of its originators to enable them to leave recognizable traces in ancient sediments. On the other hand it is probable that the earliest Metazoa were small detritus feeders. They may have developed cuticles of structural proteins which could be preserved together with other organic debris and discovered by palynological methods (acid treatment and bleaching) in unmetamorphosed sediments.

The Vendian faunas. The latest Precambrian (Vendian, about 680–560 m.y. ago) presents a clear picture of distinctive palaeontological documentation of a stage of metazoan evolution. Metazoans are widespread and locally abundant in rocks of this age (Chapter 2). Extended and thorough investigations have shown that a considerable number of sequences of Precambrian strata of that age contain distinctive fossils. Most of them became known only in the last 20 or 30 years. They have very little in common with the succeeding Cambrian faunas. The most noteworthy facts about the Metazoa of the uppermost Precambrian are the general absence of mineralized tissues such as calcareous or siliceous shells and the presence of representatives of several phyla, most of them extant in the living fauna. These are the Cnidaria, Annelida, Echiura, Arthropoda, Pogonophora (?) and possibly one or two others (Glaessner 1979*a*,*b*). The absence or probable absence of some phyla which are well represented in the Early Cambrian fauna, such as Porifera, Archaeocyatha, Mollusca, Brachiopoda and Echinodermata, and of the Platyhelminthes which on theoretical grounds are believed to have existed in Late Precambrian time, is noteworthy. The absence of Bryozoa and Chordata which make their first appearance in later Cambrian time should also be noted.

Late Precambrian metazoan body fossils are at present known from about 20 different regions in Africa, Asia, Australia, Europe and North America (Fig. 1.8). In some of these regions (southwestern Africa, South Australia, northern Russia, Ukraine) there are numerous localities where such fossils have been found, while from others we know so far only single finds. At the present time some of these fossils are regrettably still undescribed and some descriptions are still unpublished. Most areas have some distinctive kinds of Precambrian fossils in common but from South Carolina we know at present of an occurrence of a worm-like organism not found elsewhere (Cloud, Wright & Glover 1976). It is remarkable that even a few species are common to areas which are now as far distant geographically as South Australia, southwest Africa and northern Russia. A possible explanation for this fact will be discussed in Chapter 3. However, there are marked regional differences and we are not dealing with a cosmopolitan fauna in a biogeographic sense. There is evidence of considerable palaeoecological differences between fossil assemblages. What they have in common, however, is their distinctive level of diversification, with possible indications of increasing diversity within the time span of the fossiliferous Late Precambrian strata at the different localities. To give an order of magnitude to the possible time span, we anticipate here

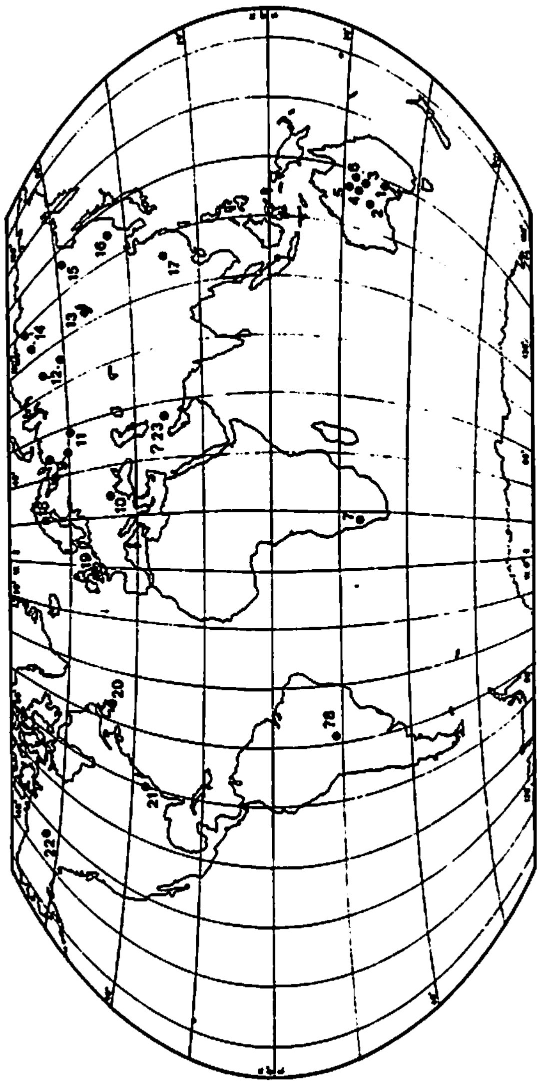


Fig. 1.8. Worldwide occurrences of Ediacarian fossil Metazoa. *Australia*: 1 – Ediacara, Flinders Ranges; 2 – Punkerri Hills, Officer Basin; 3 – Deep Well, southeast of Alice Springs; 4 – Laura Creek, southwest of Alice Springs; 5 – Mt Skinner; 6 – Jervois Ranges, southwest Georgina Basin. *Africa*: 7 – southwest Africa/Namibia. *South America*: 8 – Mato Grosso, southwest Brazil. USSR: 9 – northern Russia; 10 – southwest Ukraine (Podolia); 11 – western Urals; 12 – Yenisey River (Igarka–Turukhansk); 13 – Irkutsk (Lake Baikal); 14 – Anabar and Olenek (northern Siberia); 15 – River Maya; *China*: 16 – northeast China (Longshan); 17 – Yangtze Gorge. *Northern Europe*: 18 – Lake Torneträsk (Sweden); 19 – English Midlands (Leicester). *North America*: 20 – southeast Newfoundland; 21 – North Carolina; 22 – northwest Canada (Mackenzie Mountains); 23 – central Iran. (Several localities without body fossils were omitted. Sequence of numbers follows text. Material from all localities except 8, 13, 15, 18, 22 has been available to me for study.)

the results of an analysis of the datings of individual finds which will be discussed later. All these finds represent a time range of no more than 100 m.y., from about 580 to 680 m.y. before the present. It may be shorter but it is unlikely to be longer. It is no more than one fifth of subsequent Phanerozoic time, equal to the time which elapsed since the mid-Cretaceous. The fossiliferous rocks from this interval of time represent a variety of marine biotopes, but some of those which exist now, particularly those created by later differentiation of the biota, e.g. reefs, are absent, as are dark shales like the uncommonly richly fossiliferous Middle Cambrian Burgess Shale. The absence of most of the predators occupying the higher trophic levels of the modern marine faunas is of fundamental significance for the evolutionary level reached in Late Precambrian time. The definition of Metazoa as multicellular heterotrophs feeding by ingestion presupposes the presence of primary producers, presumably planktonic Protista, of which we have little direct knowledge. The abundance of life which can be documented at least at some sites must have led to an abundance of organic detritus and in turn to an abundance of detritus feeders, some of which can be documented by trace fossils of Late Precambrian age.

Between the very fragmentary record of traces of metazoan life about 900–1000 m.y. ago and the relatively abundant record of the latest Precambrian (680–580 m.y. ago) there is a significant time gap. It may be partly filled when more attention is given to a detailed search for traces of animal life activities in sediments 700–900 m.y. old. Pending further discoveries, we have to consider the possible effects of environmental changes during that time. In this connection we note that it is the time of the first of a number of Late Precambrian glaciations.

The sedimentary record of mid-Proterozoic time (1700–1000 m.y. ago) does not show any significant deviation from that of later times. It gives no grounds for assuming that the evolutionary origin of Eukaryota and the subsequent emergence of animal life, either as single-celled Protista or later as Metazoa, was due to some fundamental change in the environment. Once the oxygen level permitting oxidative respiration was reached, the adaptation to this efficient energy source was controlled not so much by changes of the external environment as by advances in the biochemistry and physiology of the organisms. We can speak of a phase of cytological evolution without contradicting any basic assumptions of modern selectionist (neo-Darwinian) evolutionary theory. Selection would favour physiologically more efficient organisms functioning in an unchanging environment. The significance of cytological changes in cell division leading to meiosis and redistribution of hereditary material by sexual processes in early eukaryotes has been discussed (Schopf *et al.* 1973,

Glaessner, M.F.
The Dawn of Animal Life pp. 1-42
Cambridge University Press
Used by permission

1

Precambrian life and its environment: a review

1.1 Theories about the origin of life and its environment

In the eighteenth century, James Hutton, one of the founders of the science of earth history, saw 'no vestige of a beginning, no prospect of an end'. Much has been learned in the 200 years which have elapsed since these words were written. We still see them as containing rather more than a grain of truth but we give them a different meaning from that which was in Hutton's eighteenth-century mind. Our aim now is to understand ourselves and our environment as part of an ongoing process of change. The model, still frequently if unconsciously applied because it is so close to the individual's experience, of birth, youth, maturity, old age and death, is not helpful when we consider superindividual processes in the universe. It does not imply a tidy and tedious calendar of a journey from point A, labelled 'beginning' to point B, 'the end', of which we wish to compile a partial diary. The aim of our science today is to compile a record of interactions which help us to understand the complexity of ourselves and our environment as revealed by ongoing research at ever increasing rates. The part of history of the universe which is our subject is animal life and its increasing diversity. To be sure, it had a beginning in the sense that there was a time when there were no animals, but, as Hutton recognized, the record contains no vestiges of a beginning and it would be a futile strategy to search for them. The history of life on earth may come to an end: we know that our technology has given us the means to end its existence but no end is in sight, at least at this point in time. James Hutton wrote that 'time is to nature endless and as nothing'. It has taken two centuries of thought and experiment to give content and definition to the concept of geological time which Hutton intuitively felt as 'endless' compared with experienced human life time and therefore as 'nothing' within our experience of time. The study of a process implies the study of rates of change in time and therefore the Huttonian endlessness must be measured and the supposed nothingness must be filled with evidence of events. We operate now with a geological time scale (Van Eysinga 1975, Cohee, Glaessner & Hedberg 1978). The order of magnitude of divisions of this scale, required for the study of the emergence and diversification of animals, is from 1 to 1000 million years (m.y.). The major division of the scale with which we are concerned is labelled Precambrian, for reasons which will be examined in detail in Chapter 4.

The origin of life is the result of circumstances and events close to the beginning of Precambrian time. In the absence of material vestiges of the first organisms we must briefly consider current theories about the beginning and about the distinguishing characters of what is called life. The question of the origin of life as a subject for scientific investigation was first raised in theoretical discussions independently by Haldane (1929) and Oparin (1924). The basis for a unified theory of the abiogenic origin of life is its basic chemical uniformity, together with well-founded assumptions about the possible composition of the atmosphere, hydrosphere and lithosphere at the relevant time, 3500–4500 m.y. ago. Haldane invoked from the then available experimental evidence the concept of an abundance of organic molecules ('sugars and some of the molecules from which proteins are built up') which are necessary for the formation of the elementary building materials of all organisms. He postulated that 'they must have accumulated till the primitive oceans reached the consistency of hot dilute soup' (Haldane in Bernal 1967, p. 246). This came to be referred to as the primitive soup or 'primordial broth'. Oparin considered the behaviour of colloidal solutions, pointing to the importance of cell walls and heterotrophic nutrition (from pre-existing organic compounds) in early organisms. Several years later, the experiments of Miller (1953) (see also Miller & Urey 1959) showed that energy discharge (electric sparks) in oxygen-free gas mixtures produced amino acids, the building stones of proteins, and similar experiments with other energy sources and gas mixtures produced further essential biological molecules.

We can now summarize the theoretical requirements for the origin of life as follows.

(1) The formation of organic molecules from which the basic components of living systems (proteins, lipids, carbohydrates, nucleic acids) can be built up. Some of their precursors occur in interstellar space. They have been identified by radioastronomy (see Chang 1981). Others occur in meteorites, particularly those belonging to the class of carbonaceous chondrites, or they could exist in comets. Some can be and have been produced experimentally from what would have been common materials in the primitive earth crust and a reconstituted early atmosphere, in the absence of free oxygen.

(2) The atmosphere had to be reducing but the arguments continue about the need for a strongly or minimally reducing one, and accordingly about its composition. It may have contained hydrogen, water, ammonia, methane; or water, nitrogen and carbon dioxide; or carbon monoxide, hydrogen sulfide, ammonia and methane.

(3) A source of energy was required for the reaction of synthesis of prebiotic molecules. This could have been lightning which is essentially an electric spark, as in the Miller experiment, or ultraviolet radiation which was then unimpeded by the subsequently formed ozone screen in the upper atmosphere, or volcanic heat, as in hot springs, or solar heat, as in a desert where dew falls or where tide pools exist on an arid shore. Ionizing radiations have also been considered.

(4) 'The continued supply of external energy by radiation, etc., led to a build-up of substances of high free energy content on the early earth. The products must have remained dissolved or suspended in the water... The aqueous solution and suspension was the medium in which according to Oparin and Haldane life developed – Haldane's "dilute soup"' (Broda 1978, p. 27). No quantitative estimates of the possible dilution (or concentration) of this 'soup' are available. The need for an additional mechanism for localization of synthetic processes was felt by Bernal as early as 1951 when he suggested that some of the ubiquitous clay minerals, or alternatively quartz crystals, may have acted as catalysts for polymerizations. The difficulty of having an ocean turned into a soup bowl so as to bring organic molecules close enough together to make them react in quantity is only slightly mitigated by modern geological thought about the existence of many 'small' oceans in the early stages of crustal evolution. The unease about the organic-filled prebiotic ocean led to speculation about hot springs or lagoons as the geographical environment for the origin of life. However, with this model intercommunication of possible sites of reaction would be lost and the biochemical unity of the nascent life would be harder to understand.

(5) An unsolved problem is the beginning of the mechanisms confining the molecules required to form individual organisms. All living organisms form cells, except viruses which 'are not relevant to the origin of life' (Broda 1978, p. 30), being essentially adaptations of biochemical systems to parasitism. Cells are enclosed in membranes consisting of proteins and lipids. The cell membrane not only keeps together and separates from the environment the basic materials for life processes (metabolism, self-reproduction, bioenergetics) 'but is also the place of the mechanism for

transport into the cell and out of the cell, including the pumps for active transport' (Broda 1978, pp. 43–4). Protective cell walls consisting of sugars and peptides or of cellulose or chitin may surround the cell membrane but do not occur in primitive animal cells, with grave consequences for the possibility of their preservation in rocks. The recognition of the problem of separation of individual protobionts from the primeval broth led Oparin to fruitful experimentation with coacervates, self-congregating units in colloidal systems. It also led to preparation of 'microspheres' by Fox (see Fox & Dose 1972) and his followers. They are natural aggregations of proteinoids with interesting cell-like properties and appearance.

(6) The final problem is which – if any – of the processes of prebiotic chemical evolution, demonstrated to be possible by experiment, have actually led to the origin of life. There is some agreement about the actual occurrence of chemical evolution (Calvin 1969). It suggests a natural selection among chemical processes and products, leading to the survival of those most suited to continued self-reproduction through adaptation to the constraints and resources of the environment. This formulation points to two aspects of the problem which are under active investigation but still unsolved. Self-reproduction of all organisms is based on automatic transmission of information. This is contained in the genetic code, incorporated in the structure of the DNA molecule and uniform throughout the existing organic world. Its origin is still obscure. The second problem is the demonstration that what could have happened in the earliest stages of the history of the biosphere has actually happened. This would be possible if organic remains claimed to be produced by early organisms could be reliably distinguished as formerly living organisms from prebiotic chemical fossils.

The discovery of apparent fossil remains of single-celled organisms resembling bacteria and algae in rocks more than 3000 m.y. old (Barghoorn & Schopf 1966, Pflug 1966, Schopf & Barghoorn 1967) is significant. Critical reviews by Schopf (1975*a*) have thrown doubt on the status of these fossils as organisms and indicate their resemblance to shapes assumed by abiotic, in this instance probably prebiotic, organic matter. They have been referred to, with other similar finds, as 'fossil-like objects' (Schopf 1978). However, fossil sedimentary structures of organic origin known as stromatolites occur in rocks 3500–2500 m.y. old in Australia, southern Africa and Canada (Walter 1983). It is now well known that they are formed by the life activities of Prokaryota (bacteria and blue-green algae) (see Walter 1976, 1977, 1978, 1983). Undisputed fossil remains of microorganisms which produced stromatolites have been observed. There is evidence (Walter 1978, 1983) that some Archaean stromatolites were built by photosynthetic Cyanophyta and by purple or green bacteria which assimilate carbon dioxide but do not produce free oxygen (Broda 1978). Many formerly existing forms of life, including microbes, must have been supplanted sooner or later by more efficient ones. This will become increasingly important as our story progresses but it causes the greatest difficulty in reconstructing microbial life of the past, a problem of classifying organisms whose observable life processes are vastly more significant than their morphological configurations.

Before we proceed to the next stage in the history of life it is necessary to consider the placing of the documented events in precise time scales and also to understand fossilization processes which convert the results of life processes into components of sedimentary rocks.

1.2 The Precambrian, its subdivisions, and the dating of biohistoric events

The construction and application of a geological time scale is almost as old as the science of geology. It implies a division of the uninterrupted flow of time during the entire history of the earth by a sequence of distinctive events. The model for it was, either consciously or unconsciously, human history with its division into antiquity, the Middle Ages and modern times, and its subdivisions such as Egyptian or Chinese dynasties, or the reigns of European monarchs, or socio-economic changes. There are also unique marker events like the Renaissance or the French Revolution, and the divisions of prehistory according to technological advances such as stone age, bronze age, iron age. The scales define primarily only *sequence* (the placing of other events as earlier or later than the key event), not *chronology* (dating of events and consequently duration of intervals). The geological chronology is being established with ever increasing accuracy by relating geological events to clock-like, fixed-rate processes such as radioactive decay. This makes it possible to calibrate the sequential stratigraphic scale. This scale was established early, by the application of the law of superposition to stratigraphic sequences. Except for later tectonic disturbances, lower strata must have been laid down before higher ones. The study of fossils contained in the rocks showed more than 150 years ago that they facilitated distinction between rocks occurring in sequence, and comparison of such sequences. These observations led soon to the development of biostratigraphy which eventually found in irreversible evolution and concomitant extinction of forms of life a guide for the recognition of increasingly finer divisions of the geological time scale. Biostratigraphy could be applied as far back in history as abundant fossils could be found, that is, to the beginning of the Cambrian

Period, the earliest part of the Palaeozoic Era. This and the two succeeding eras (Mesozoic and Cenozoic) were grouped together as the Phanerozoic Eon whose name was taken to mean the time of visible (or obvious) life.

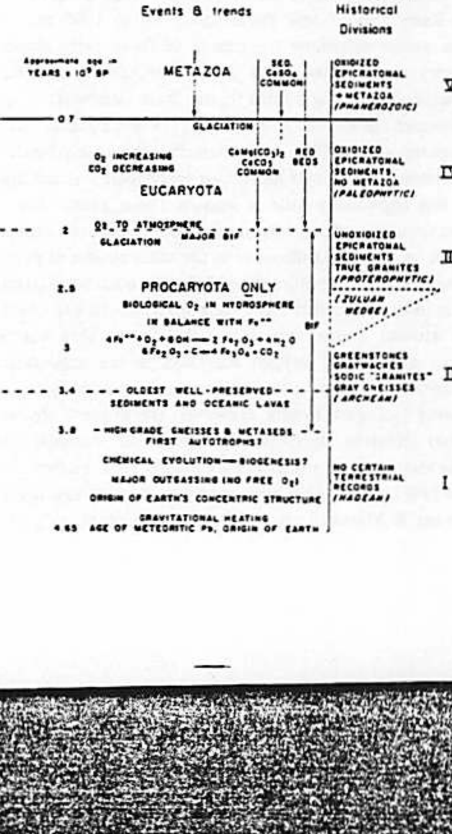
When these names were proposed, no obvious traces of life were known from rocks older than those of Cambrian age, hence they were lumped together as Precambrian, of unknown duration. When calibration indicated that Precambrian time comprised about seven-eighths of geological time and when interest in Precambrian rocks increased for various other reasons, an extension of the more detailed time scale became necessary. The Subcommission on Precambrian Stratigraphy (a unit of the International Union of Geological Sciences) is preparing proposals which will eventually be submitted to an International Geological Congress for ratification. A proposal for a conventional division of Precambrian time into eons, the Archaean and the Proterozoic, has been approved by the subcommission. The boundary is defined by its chronometric (isotopic, commonly but loosely called radiometric) age, at 2500 m.y. ago. Possible subdivisions of the Archaean are still under discussion. The oldest presently known rocks are about 3800 m.y. old. Nothing is known from geological data about events between that age and the formation of the earth about 4600 m.y. ago. The dating of this event is based on cosmological evidence from meteorites and moon rocks. There is increasing information about the Archaean lithosphere, its composition and dynamics (Windley 1976, Tarling 1978). By the end of that eon, the earth's crust appears to have evolved significantly, to approximate the present lithosphere more closely, and so the concept of an eon boundary at about 2500 m.y. (James 1978), which is also significant for the history of life, is likely to find wide acceptance. A subdivision of the following eon, the Proterozoic, into three eras has also been proposed. The terms Early, Middle and Late Proterozoic will be used here for these divisions. Their proposed boundaries at 1600 and 900 m.y. (Sims 1980) are times of significant geological events documented in many continents.

The end of the Precambrian is necessarily the beginning of the Cambrian. Its precise definition is being considered by an international working group, which intends to base it on biostratigraphic changes in a stratotype section which has yet to be designated. This means that it will be based on the same principles as the Phanerozoic scale (Hedberg 1974) and that its age will be established by the dating of that concrete boundary event rather than by being based on an abstract chronological definition alone. The age of the base of the Cambrian is generally considered as approximately 550–570 m.y. The uncertainty may be greater than this range and further isotopic dating is urgently required. This boundary and its implications will be considered further in Chapter 4.

From the viewpoint of biostratigraphy, the Proterozoic is the time when stromatolites became abundant, while animal fossils from at least the greater part of this era are absent or rare. The middle of Precambrian time, between the age of the oldest rocks and 570 m.y. would be at about 1600 m.y. The informal concept of a mid-Precambrian time to which reference is made occasionally would encompass some early and some middle Proterozoic time from, say, about 1800 to 1400 m.y. ago. As we shall see, this is thought to be the time of the emergence of eukaryotic cells, with nuclei and organelles, from their prokaryote, non-nucleated predecessors. This important evolutionary step either rapidly led to or actually coincided with the first appearance of animal protists (Protozoa). The refinement of meiosis in cell division, sexual reproduction with its concomitant increase in evolutionary potential, colonial organisation and the origin of Metazoa, multicellular animals developing distinct tissues, followed sooner or later. The order of these events is based firmly on biological theory (see Margulis 1970, Patterson 1978) and will be further discussed in section 1.4. The palaeontological evidence and its geochronological dating are uncertain. This uncertainty encompasses nearly the entire mid-Precambrian interval, if not the entire Early and Middle Proterozoic, up to 1000 m.y. The evolutionary rates which determine the timing of these early stages of metazoan prehistory are unknown and cannot be known as long as remains of fossil animals, body and trace fossils, from sediments of Early and Middle Proterozoic age are unknown. Cloud (1976*a*) believes that the first Metazoa appeared about 700 m.y. ago which is the approximate age of the first unquestioned remains of metazoan body fossils. Older traces of animals exist but regrettably little is known about them. The first manifest metazoan fossil representatives are not what we would expect on theoretical grounds, i.e. small, planuloid or at the most advanced possible level, primitive coelenterates. It is more than likely that such animals could not have left traces in the sediments but a time span should and could be allotted to them without undue appeal to speculation. This has been attempted. The curve depicting oxygen accession to the atmosphere – oxygen being essential for oxidative respiration – is occasionally invoked as a means of dating biological events. However, the physical theory of oxygen (and ozone) accession, developed by Berkner & Marshall (1965) and brought to the attention of geologists by Cloud (1968), Fischer (1965, 1972) and Rutten (1962, 1966) who have modified it, does not result in a time scale. Berkner & Marshall placed the 'Pasteur point' (1% of the present atmospheric level (PAL) of oxygen) at the beginning of the Cambrian, which was thought to be the time of the oldest fossil animals. We know now that this is incorrect. Not enough is known quantitatively to calibrate precisely in terms of geological time the Berkner–Marshall curve of oxygen accession.

Cloud (1976*a,b*) divided the post-Archaean (Proterozoic) time, on geochemical grounds of change from unoxidized to oxidized epicratonal sediments and on grounds of biological change from Prokaryota to Eukaryota, into Proterophytic and Palaeophytic Eras, with a boundary at about 2000 m.y. (Figs. 1.1, 1.2). There is at present no indication of general acceptance of this biostratigraphic scale and nomenclature. Even soundly based and euphonious names are not necessarily generally accepted. Time scale divisions are conventions depending on consensus rather than on strictly scientific generalizations of a multitude of empirical data. International discussions on the time scale of the Precambrian and the naming of its subdivisions are planned for the 1980s. After a review of the Late Proterozoic Metazoa, the time framework for that period will be re-examined (Chapter 4).

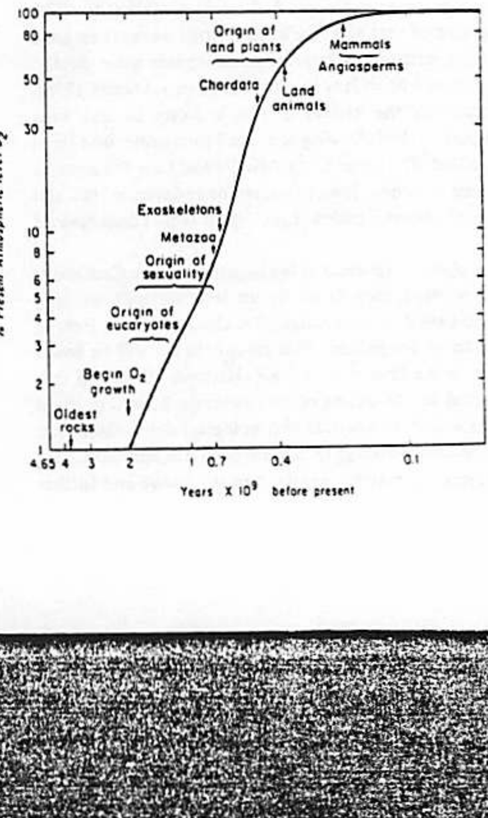
Fig. 1.1. Major divisions of earth history and related aspects of biological and geochemical evolution. (BIF – banded iron formations; BP – before present.) Adapted from Cloud (1976*b*).



1.3 The preservation and recognition of organic remains in rocks: chemofossils, fossil organisms, pseudofossils

Before proceeding with the study of the history of life to the point in time when animal remains become recognizable, we have to consider the nature of existing documentation of the life of the past and the interpretation of these documents. There is an analogy between them and the material evidence of human history. The analogy is valid in the sense that as we go back to the oldest times, the record of both kinds of history becomes more difficult to decipher and to interpret. It also becomes inherently more incomplete, full of gaps and false clues which are apt to mislead. Much has been written about the inherent incompleteness of the Phanerozoic fossil record. Estimates have been made of the percentage of living species and higher taxa which could not have been preserved. Much can be learned from the well-known Phanerozoic record about fossilization

Fig. 1.2. Apparent timing of events in biospheric evolution compared with hypothetical levels of atmospheric oxygen. Both scales logarithmic; curve highly generalized, excursions from the mean likely. From Cloud (1976*b*).



potential, but as will be seen in Chapter 4, evolution of mineralized skeletons has made a quantum jump at the Precambrian–Cambrian transition. Their absence has not rendered the Precambrian entirely unfossiliferous as had been expected earlier, but it has necessitated much effort in building up a specific methodology of discovery and interpretation of Precambrian fossils which is still incomplete. The availability of fossils for the study of life history is a function not only of the fossilization potential of the biota of different periods but also of the exploration effort. In Darwin's time so few Phanerozoic fossils had been discovered that he was unable to make any significant use of them for his theories of evolution by natural selection. One hundred years later, no theoretical discussion of the course of evolution is complete without reference to palaeontological evidence. The rapid increase in the exploration and study of Precambrian geology, stimulated by the abundance of mineral resources formed during that time span, and the recognition of its great length, has revealed an abundance of Precambrian fossils. In general terms they are quantitatively and qualitatively inferior to those of later Phanerozoic age but they are highly significant for several reasons. It is true that with greater age of rocks there are greater chances of their alteration by repeated tectonic and magmatic processes leading to the destruction of organic remains but this is not the general fate of all Precambrian rocks, contrary to what had been expected. Exploration has revealed that many of them have remained not too strongly altered to contain organic remains. There are thick sequences of sediments previously considered as unfossiliferous (and often in the absence of clear evidence dated as Palaeozoic) which, with the application of new biostratigraphic methods, have been shown to be of Precambrian age and which have since been dated chronometrically. New methods of the physical sciences, electron microscopy and modern analytical chemistry, are leading to discoveries of great significance which are still being evaluated. The occurrence of organic matter in ancient rocks has been known for a long time and has proved puzzling, in the absence of other traces of fossil organisms and in the face of authoritative views of their presumed absence. Molecular biology, a dominant branch or viewpoint in biological sciences, has led to suggestions of a 'molecular palaeontology' (Zuckerkandl & Pauling 1962, Zuckerkandl 1965, Florkin 1966, McLaughlin & Dayhoff 1973, Broda 1978). They are based on evolutionary studies of proteins (haemoglobins, cytochromes) from mainly microbes and vertebrates, and evaluation of their genetically controlled alterations with respect to phylogeny and time. Biochemists have considered chemical evolution (Calvin 1969). The evolution of bioenergetic processes (Broda 1978) is not a supplement but an essential substrate for these studies, with far-reaching consequences for the understanding of the first appearance of animals whose metabolic activities and bioenergetic requirements are different from those of their predecessors.

The palaeobiological significance of the study of stable organic compounds isolated from ancient rocks as 'chemical fossils' has been recognized in organic geochemistry (e.g. McKirdy 1974). The main problems are the possibility of abiotic or prebiotic origin of some organic compounds, the danger of contamination when the analyst has to deal with minute quantities, the possibility of migration of fluids through even weakly porous or fractured rocks during long periods of time, and the alteration of chemical structures by diagenesis and incipient metamorphism. In recent years there has been much progress in the understanding of the alteration of insoluble organic matter in sedimentary rocks (kerogen) which must be syngenetic and coeval with the enclosing rock. The possible linking of chemical marker fossils with evolutionary events is particularly significant. The early occurrence of photosynthesis is partly deduced from such studies but the dating by chemofossils of the change from prokaryote to eukaryote cells or from fermentation to oxidative respiration is still essentially in the future.

The recognition of fossil organic remains on the basis of their configuration rather than their chemical composition is particularly difficult in ancient rocks. These difficulties have been discussed by Hofmann (1971, 1972) who proposed an elaborate classification which distinguishes 'pseudofossils' and 'problematica' or 'dubiofossils' from definite remains of organisms ('fossils'). Here we shall distinguish between difficulties of recognition caused by (1) alteration of rocks, (2) similarities of organic and inorganic configurations and (3) problems of evolutionary morphology affecting fossilization.

(i (1) Alteration. It is generally true that Precambrian rocks are more highly altered, i.e. more affected by diagenetic, metamorphic and tectonic processes than younger rocks. Fossils have been found in younger metamorphic rocks but their preservation under conditions of elevated temperatures and pressures or metasomatic processes is dependent on a degree of robustness of embedded organic remains which was not generally attained, at least by animal bodies, until Phanerozoic time. The search for remains of fossil organisms is generally confined to special classes of altered sedimentary rocks. Early diagenesis can enclose organic remains so as to seal them hermetically against further alteration. This can occur in a matrix of chert (cryptocrystalline silica) or in films of organic matter (kerogen). It may also be possible to recognize traces of organic activity (burrowing, coprolitic pellets) in the fabric of mildly altered sediments.

Few if any of them will be unquestionably biogenic and most of them will be therefore classed as 'problematica'. The odd shapes of the results of some tectonic deformations (boudinage) and of concretions will generally make them recognizable as pseudofossils.

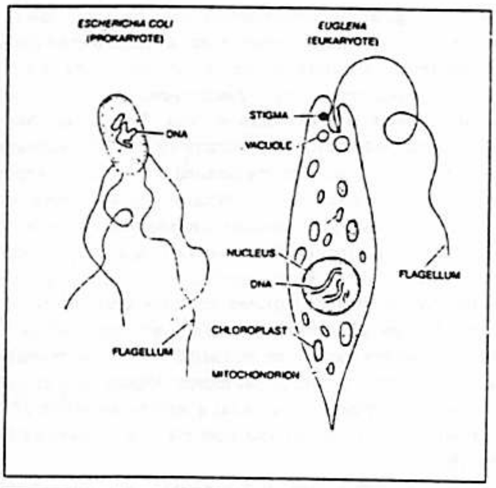
(2) Similarities of organic and inorganic configurations. Geometric regularity of configurations which could not be immediately identified as crystals was taken as an indication of organic origin, at least in the early stages of the search for Precambrian fossils. Clay pellets often resemble fossil shells, infillings of drying cracks can resemble worm burrows or bodies, the outlines of angular shards of dried clay layers have been combined to produce misleading reconstructions of imaginary fossil arthropods. Fractured or sectioned glass bubbles in volcanic tuffs may strikingly resemble sponge spicules or chambered Foraminifera or Radiolaria. Configurations such as vertical stalks ending in lobate structures have turned out to be sand volcanoes. Jellyfish-like concentric structures on bedding planes may be found, on closer examination, to have vertical feeder channels and to be formed by escaping gas or fluids. Others, without stalks, have turned out to be pressure marks or moulds of crystal rosettes. The possibility of mechanical origin such as current effects or bedding lineations must be rigorously excluded or at least made highly improbable by observation of specific characteristics before similar configurations are admitted as genuine fossils. An interesting example of the need for caution in evaluating the biogenicity of reputed Precambrian fossils arises from a comparison of an early distribution table of groups of fossils in the Precambrian (Glaessner 1966, Fig. 2) with present knowledge. Ten stratigraphic units from various countries were included as containing either trace fossils or (animal) megafossils. Two of them were considered as questionably fossiliferous. Three supposed trace fossil occurrences and the two questionable ones are now known to contain only configurations of mechanical origin, one contains megascopic algae mistaken for animal remains. Some trace fossils from the Precambrian–Cambrian Vindhyan rocks of India may be Cambrian and one figured trace was apparently caused by the rasping of the radula of a living snail on the surface of the rock. The origins of other trace fossils are still being questioned. Cloud (1968, 1973) has acted consistently as an invaluable sceptic in the evaluation of Precambrian configurations considered as organic remains. They require continued strenuous efforts to verify their organic origin, more than do younger fossils.

(3) Evolutionary problems. Not long ago the course of evolution before the Cambrian was undocumented and even the fauna of the Early Cambrian was poorly known. This explains why the occurrence in the Precambrian of large eurypterids ('reconstructed' from the outlines of large clay flakes) or of articulate brachiopods (which subsequently proved to come from inliers of Cambrian rocks in Precambrian terrains) was accepted. True shells or skeletons originated about the time of the transition from Precambrian to Cambrian (see Chapter 4). The real problem in finding and interpreting Precambrian fossils is the fact that these animals were without shells or skeletons (soft-bodied). The preservation of soft-bodied organisms poses many problems of the reconstruction of their original form and of the taxonomic assessment of their variability. These problems and the methods assisting in their solution are rarely discussed in the literature. A recent authoritative statement on fossilization (A. H. Müller 1979) devotes one paragraph to the preservation of soft parts in some 70 pages of text. Meanwhile the wasteful labour of description and publication of pseudofossils mistaken for fossils continues.

1.4 The cell: organization and life processes. The 'kingdoms' of organisms

All organisms, except viruses, are built up from cells or organized as cells. They are separated from their environment and able to communicate with it through the structure of their surrounding membranes. To fulfil their tasks, including the basic one of self-reproduction, the cells cannot be simple pieces of bioengineering. The simplest kind, a bacterial cell, has been compared in complexity with a chemical factory, with an information content, at the atomic level, of 10^{12} 'bits' or 'yes/no decisions' built into it. There are among living organisms two basic kinds or levels of cell organization, the prokaryote cells of the bacteria and the blue-green algae (which are now frequently referred to as cyanobacteria), and the eukaryote cells which make up all other organisms. The prokaryote cells are mostly small (1–10 μm), without a nuclear membrane or endoplasmic reticulum (cytoplasmic membrane system), no mitochondria (enzyme-containing bodies), no chloroplasts for photosynthesis; they divide by binary fission (Fig. 1.3). The eukaryote cells are generally larger (10–100 μm, exceptionally larger by up to three orders of magnitude) with a nuclear membrane and with the organelles which are absent in the prokaryotes; cell division occurs by mitosis followed by meiosis, and the cells have complexly but uniformly structured flagella or cilia. They are primarily aerobic and produce oxygen in the process of photosynthesis. 'The numerous and fundamental differences between eukaryotic and prokaryotic organisms... have been fully recognized only in the past few years. In fact, this basic divergence in cellular structure which separates the bacteria and the blue-green algae from all other cellular organisms, probably represents the greatest single

Fig. 1.3. Comparison of prokaryote and eukaryote cells. The drawings represent a bacterium compared with a single-celled green alga. Its cell contents include ribosomes and food storage bodies (unlabelled). The chloroplasts are also known as plastids and the stigma as the eye spot (Tappan 1980, pp. 895–6). From Schopf (1978). Copyright © 1978 by Scientific American Inc. All rights reserved.



	PROKARYOTES	EUKARYOTES
ORGANISMS REPRESENTED	BACTERIA AND CYANOBACTERIA	PROTISTS, FUNGI, PLANTS AND ANIMALS
CELL SIZE	SMALL GENERALLY 1 TO 10 MICROMETRES	LARGE, GENERALLY 10 TO 100 MICROMETRES
METABOLISM AND PHOTOSYNTHESIS	ANAEROBIC OR AEROBIC	AEROBIC
MOTILITY	NONMOTILE OR WITH FLAGELLA MADE OF THE PROTEIN FLAGELLIN	USUALLY MOTILE, CILIA OR FLAGELLA CONSTRUCTED OF MICROTUBULES
CELL WALLS	OF CHARACTERISTIC SUGARS AND PEPTIDES	OF CELLULOSE OR CHITIN, BUT LACKING IN ANIMALS
ORGANELLES	NO MEMBRANE-BOUNDED ORGANELLES	MITOCHONDRIA AND CHLOROPLASTS
GENETIC ORGANIZATION	LOOP OF DNA IN CYTOPLASM	DNA ORGANIZED IN CHROMOSOMES AND BOUNDED BY NUCLEAR MEMBRANE
REPRODUCTION	BY BINARY FISSION	BY MITOSIS OR MEIOSIS
CELLULAR ORGANIZATION	MAINLY UNICELLULAR	MAINLY MULTICELLULAR WITH DIFFERENTIATION OF CELLS

evolutionary discontinuity to be found in the present-day living world' (Stanier, Adelberg & Doudoroff 1963). The most unexpected fact observed in Precambrian palaeontology, the dominance of remains of bacteria and blue-green algae in rocks dated as representing the first half of Precambrian time, is an expression of this first and highly significant 'evolutionary discontinuity'. It raises three questions: (1) What evolutionary step or steps could have bridged this discontinuity? (2) How are these evolutionary events related to changes in the environment? (3) What evolution, if any, took place at prokaryote level, during Early Precambrian time, through more than 2000 m.y. since deposition of the oldest fossiliferous rocks?

(1) One perhaps surprising answer to the first question was given by the symbiotic theory of the origin of the eukaryotes (Margulis 1970). Margulis considers the classical view that the more highly organized forms of life originated from bacteria and blue-green algae 'by the accumulation of selectively advantageous mutations [as] inconsistent with many facts'. The symbiotic theory asserts that prokaryotes acquired such organelles as mitochondria in the form of symbiotic aerobic bacteria; flagella and cilia were similarly pre-existent as spirochaetes and incorporated in prokaryotes where they evolved eventually to participate not only in locomotion but also in advanced reproductive processes ('mitotic apparatus'). Some blue-green algae were originally endosymbiotic in protozoan-like heterotrophs to form plastids and finally chloroplasts as sites of photosynthesis. The details of this theory and the factual evidence need not concern us here. They were lucidly presented by Margulis, with due consideration of available palaeontological and geohistorical data documenting as far as possible what could have happened at the relevant time in the history of life. G. E. Hutchinson remarked in his foreword to Margulis' work (1970, p. xvii) that 'not every interested biologist may accept all her suggestions'. This has proved correct (see Broda 1978, pp. 132, 180 for contrasting views; Taylor 1974, with reply by Margulis 1975; Cavalier-Smith 1975, discussed in detail by Cloud 1976*a*). A review of relevant new observations and interpretations appeared too late to be considered here (L. Margulis, *Symbiosis in cell evolution*, W. H. Freeman & Co., San Francisco, 1981). It strengthens and widens the basis for the endosymbiosis theory. At least some of its postulates have proved widely acceptable, fruitful for further research, and significant for the origin of animals, an event not likely to have occurred much later than the first appearance of eukaryotic cells in mid-Precambrian time. The various hypotheses about their evolution and its timing were clearly and fairly reviewed by Tappan (1980, pp. 84–96).

(2) The answer to the second question is that these evolutionary events at cell level do not presuppose any rapid and fundamental changes in the

environment. It may be said that they have gradually laid the foundations necessary for an expansion of the biosphere in mass, diversity and in its influence on the history of the earth. The lithosphere and hydrosphere of at least Early Proterozoic time were, according to geochemical and tectonic data, probably not very different from their present composition and dynamic state, though not as rich in diverse ecological sites (niches) for various life forms. The most distinctive development in the physical environment in Early Precambrian time was the accumulation in the atmosphere of free oxygen. According to Cloud (1976*a*) it amounted to about 3% PAL at the time of the origin of the eukaryotes. There is biogeochemical evidence of photosynthesis occurring at the time of the formation of the oldest sedimentary rocks. It is still practised by a few bacteria. They never produce free oxygen but the blue-green algae and the plants do. The oxygen in the present atmosphere is the product of this life activity, with only minor contributions from photodissociation of water. The process is efficient enough: 'It can be calculated, for example, that the offspring of one gram of algae (assuming unlimited space and nutrients and a reasonable rate of cell division) could photosynthetically produce an amount of oxygen equal to that of the present atmosphere in less than 40 days' (Schopf 1975*b*, p. 55). But of course space and nutrients were not unlimited and so it took longer. How much longer is still a controversial question; estimates vary from the assumption of a fully oxygenic atmosphere in Early Precambrian time, which conflicts with the fact that unoxidized minerals of that age were found where they would have been exposed to the atmosphere when deposited, to Cloud's estimate (1976*a*) of only about 50% PAL at about the end of Cambrian time. Cloud has repeatedly drawn attention to a relevant geological fact. Banded iron formations, presently the major source of industrial iron production, had their worldwide maximum development about 2000 m.y. ago and then they disappeared from the record almost completely. They were formed under conditions different from those which the fully oxidized redbeds required for their deposition. Typically their colour is due mainly to quartz grains coated with iron oxides. Cloud considers this change as the first indication of conditions which permitted a change of life processes from fermentation to oxidative respiration, a change which now occurs in facultatively aerobic organisms at about 1% PAL in the organisms' environment (the so-called Pasteur point). This change in the organisms' energy acquisition occurred certainly under considerable selection pressure, as oxidative respiration is more than twice as efficient than fermentation, in terms of free energy produced by the basic reaction.

(3) The fact that Prokaryota (bacteria and blue-green algae) dominate the Precambrian fossil record does not necessarily indicate very low rates of evolution during this long time. Studies of detailed stratigraphic sequences are still too few and too new for final conclusions to be drawn. Not all difficulties of interpretation of fossilization processes and of their influence on original micromorphology have been overcome. There are indications of evolutionary changes in size and complexity (Schopf 1977). Schopf has repeatedly stressed the importance of the development of mitotic and meiotic reproductive mechanisms of the eukaryote cell for significant increases in diversification rates. This is very likely but the palaeontological evidence for their occurrence is by no means unequivocal, because of problems in interpreting detailed observations on fossil remains of single or possibly dividing cells. What is probably more important is the need for the evolution of complex and sophisticated new biochemical pathways within the structural framework of the prokaryote cell to enable it to make efficient use of the changed environment for its metabolic and bioenergetic needs. The time required for this biochemical evolution under conditions of a significant increase in the oxygen content of the atmosphere and the hydrosphere cannot be estimated but it was probably long in geological terms. In other words, the morphological simplicity of the fossil and existing prokaryotes may conceal a lengthy evolution of their basic life activities and adaptations and possibly also extinctions of less well adapted life forms.

Energy in the form of light is a precondition of life which (to use a profound thought expressed variously at different times by such eminent physical scientists as Boltzmann, Schrödinger or Broda) interposed itself into the flux of radiant energy from the hot sun to the warm earth and on to cold space. Oxygen is the basis for oxidative respiration, the efficient transformation of radiant energy to the energy of the chemical bond in the adenosine triphosphate (ATP) molecule of living organisms. The accession of oxygen to the atmosphere as the result of photosynthetic use of radiant energy is favourable for the further evolution of life under two conditions. Some of it had to be available to protect living organisms from harmful ultraviolet radiation, by forming an ozone screen in the atmosphere. The ability of some prokaryote cells to repair radiation damage has been proved experimentally but could hardly have been effective enough to assure their dominance during 2000 m.y. of early history of the biosphere. Other protective mechanisms have been invoked, such as the shielding effect of 10 m of water or of the calcareous deposits on algal mats which are preserved throughout the geological record as stromatolites. There have been objections to the presumed efficacy of these mechanisms. On the other hand, there have been objections to the suggestion that the ozone screen began to operate as soon as there was free oxygen in the atmosphere. If it appeared late, the apparent delay in the occurrence of abundant life on dry land could be explained. Abundant plant or animal fossils are not found in non-marine sediments older than 400–500 m.y. However, there may have been other reasons for the late appearance of abundant terrestrial life. The conclusion must be that the dating of the screening out of lethal ultraviolet radiation is still controversial: 600–650 m.y. ago according to Cloud (1968), or the time when as little as 0·1% PAL of oxygen was present (Ratner & Walker 1972).

The second precondition for the use of free oxygen in life processes is avoidance of its toxicity. While some biologists believe that the necessary protective enzyme systems could have evolved in a geologically insignificant time span, concomitantly with the respiratory pathways, Cloud (1976*a*) assumes a slow start of respiration in oxygen-poor environments and designates the onset of redbed sedimentation about 2000 m.y. ago as the time 'when enzymatic mediation of that particular atmospheric pollutant became efficient enough to tolerate O_2 levels above $\sim 1\%$ P.A.L.'

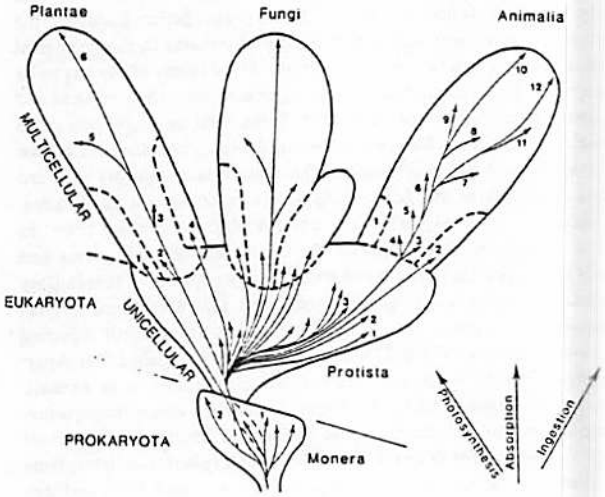
We can now turn to the question of the time of the first appearance of eukaryote organisms. Cloud (1976*a*), who has made the most comprehensive studies of the beginnings of biospheric evolution, concluded that the origin of eukaryotes may have occurred at any time from 2000 m.y. ago onward but that the oldest really persuasive morphological evidence for that development is not found until about 1300 m.y. ago. Before discussing the evidence for the earliest occurrence of animal remains in the geological record, the distinction between animals and other forms of life has to be considered. All animals and plants (all organisms other than bacteria and blue-green algae) have eukaryotic cells. Some exist as single cells or as colonial cell aggregates. They are known as Protista, or if they are known to live as animals do, as Protozoa. The multicellular animals with cell systems differentiated and functioning as tissues are known as Metazoa. It is the almost generally held view that the Metazoa evolved from the Protista. Within the living Protista the distinction between plants and animals is not significant but to some extent only a matter of terminology or semantics. There are single-cell organisms that sometimes nourish themselves by photosynthesis, and at other times swim about digesting food particles (Broda 1978, p. 138, with reference to Margulis 1970). Apart from those Protista which can function either as plants or as animals, photosynthesis in plants and phagotrophy in animals, which consequently develop locomotion, are generally valid distinctions. They are not important enough to exclude, for example, Phytomonadina (plant flagellates) from discussion in a standard textbook on protozoology (Grell 1973) and they could not generally be applied to extinct Protista but only to living organisms whose feeding habits can be observed. The old taxonomic term Protophyta has become obsolete; yet it does not make sense biologically to remove the photosynthetic flagellates from their systematic grouping with other algae as long as even the prokaryote Cyanophyta are commonly referred to as blue-green algae (Hanson 1977). The problem concerning us is the tracing of possible evolutionary pathways from early eukaryotes to Metazoa. The extent to which fossil remains can be used to test the historical reality of the theoretical conclusions concerning animal ancestors will depend on our ability to recognize fossil evidence of the distinctive animal *functions*: locomotion, and ingestion of food particles, with its corollary, the excretion of structured fecal matter. This is possible only at the grade of Metazoa. The recognition of fossil Protozoa as their possible ancestors is excluded in most instances by the difficulties of their preservation. While plant microbes have cell walls of cellulose, animal cells have less resistant membranes consisting of proteins and lipids, or chitin. Hardening of walls by agglutination of foreign particles, mineralization with silica or carbonates, or sclerotization of chitin are secondary phenomena which did not occur before Late Precambrian time.

The decay of the cell content during fossilization can leave a residue looking like intracellular membranes or organelles and a dark spot near its centre can resemble a nucleus. This problem has been investigated by experimental 'fossilization' of known organisms (Oehler 1976, Francis, Margulis & Barghoorn 1978), with different results according to the methods used. Similar difficulties cloud the interpretation of fossil evidence of cell division in eukaryotes as distinct from binary fission of prokaryotes, and much other palaeocytological detail (Knoll & Barghoorn 1975, Oehler, Oehler & Muir 1976). A significant conclusion from experimental testing of earlier observations states: 'Although it is a logical conclusion that the eukaryotic level of cell organization arose before 680 m.y. ago, there is no definite cytological evidence for an earlier date of appearance of eukaryotes in the fossil record'. (Francis *et al.* 1978, p. 97). On this negative evidence we cannot disregard 'logical conclusions'. Work on morphology, palaeobiochemistry and fossilization of Early and Middle Proterozoic microfossils is progressing and the latest conclusions about the timing of evolutionary and environmental events may be subject to further revision in the light of new data and advances in methods of study and experimentation. The richest and best preserved assemblages of such microfossils are from the Gunflint and Bitter Springs stromatolitic and non-stromatolitic cherts of North America and Australia, respectively. The first named is about 2000 and the second about 850 m.y. old and there are other similarly significant microfossils of intermediate age which have been discussed by Cloud, Schopf and others in numerous review articles. A relevant and cautiously expressed conclusion seems to be that the time of deposition of the Gunflint chert 'corresponds approximately to the transition from an oxygen-poor atmosphere to one in which there were significant levels of free oxygen', about 1800 to 2000 m.y. ago; a critical period in the history of life during which 'enhanced morphological experimentation produced the problematical microorganisms now found in the Gunflint' (Awramik & Barghoorn 1977, p. 128). While according to these authors the Gunflint microbiota is wholly prokaryotic, the Bitter Springs microbiota is likely to contain eukaryotes which may have first appeared 1400–1500 m.y. ago (Schopf & Oehler 1976). Comparison of early eukaryotes suggests affinities to green algae to most observers, but others have different opinions (see Knoll & Golubic 1979) or refrain from definite taxonomic assignments. The earliest microfossils resembling possible Protozoa are from the Chuar Group of the Grand Canyon, Arizona, dated at more than 650 and less than 850 m.y. old (Bloeser *et al.* 1977), from southwest Brazil (Fairchild, Barbour & Haralyi 1978), from Saudi Arabia (Binda & Bokhari 1980) and from the uppermost Riphean of Greenland (Vidal 1979) and Sweden (Knoll & Vidal 1980). The finds have been correlated with rock sequences which are about 700–800 m.y. old. These microfossils are described as Chitinozoa ('heterotrophic protists or primitive metazoans') by Bloeser *et al.*, but as 'Chitinozoan-like' by Vidal and Binda & Bokhari.

We have seen that the simplest eukaryotes are the unicellular (or colonial unicellular) Protista and that among them the first manifestations of animal life must have occurred. They are unlikely to have left any recognizable structural marks on fossils or in rocks. The proposers of the multikingdom concepts of classification including a kingdom Protista have recognized this as a confederation of those eukaryotes which are neither Metaphyta nor Metazoa (Whittaker 1969, Whittaker & Margulis 1978). Living Protozoa which could serve as models for the first animals must be among the zooflagellates and ciliates. They are not likely to be found among the Rhizopoda of which those with preservable skeletons (Radiolaria, Foraminifera) appear late in the geological record, with clearly 'primitive' representatives initiating a comparatively well documented, complexly radiating evolution not earlier than in Phanerozoic time. Hanson (1977) considers the Protozoa as polyphyletic, i.e. as derived from various colourless algal protists. His conclusions on the origin of the Metazoa will be discussed together with others in Chapter 3. Here we state only that it is not unlikely that groups of Protista which could have included ancestors of Metazoa have survived to the present. If not, if these

ancestors have become extinct, it is unlikely that we shall obtain any information about them from the geological record. In any case, what occurred among the Precambrian Protista was probably not so much a morphological change as one towards dominant ingestive nutrition. This form of nutrition combined with motility is found in phytoflagellates which are also equipped for photosynthesis. Hence evolutionary transition to animal protists could have occurred through loss of photosynthetic organelles (plastids).

Fig. 1.4. The five kingdoms of organisms. After Whittaker (1969, Fig. 3, simplified). (Copyright 1969 by the American Association for the Advancement of Science.) Numbered subdivisions, more or less relevant to discussions in the text, are as follows. Monera (or Prokaryota) – 1. bacteria; 2. Cyanophyta (or Cyanobacteria). Protista – 1. Ciliata (or Ciliophora); 2. Sarcodina (Foraminiferida, Radiolaria, amoebae); 3. Zoomastigina (animal flagellates). Plantae – 1. Rhodophyta; 2. Chlorophyta; 3. Charophyta; 4. Phaeophyta; 5. Bryophyta; 6. Tracheophyta. (Alternative classifications (Whittaker & Margulis 1978) include 1–4 and the lower Fungi with the Protista, as 'protoctista'.) Animalia – 1. Mesozoa; 2. Porifera (or Parazoa); 3. Coelenterata (Cnidaria and Ctenophora); 4. Platyhelminthes; 5. Aschelminthes; 6. Tentaculata (or Lophophorata); 7. Chaetognatha; 8. Annelida; 9. Mollusca; 10. Arthropoda; 11. Echinodermata; 12. Chordata. (In the later classification (Whittaker & Margulis 1978) the Pogonophora and Hemichordata were grouped with the Chaetognatha between 11 and 12.)



The desirability of excluding in a formal classification scheme the Protista from both the plant and animal kingdoms is obvious, despite its inconvenience. This, as Whittaker (1969, p. 154) remarks, 'is not so much the fault of the systematist as faults of the living world as a subject of classification'. The stream of evolution widens and divides, diversifying to fill available niches (Fig. 1.4). The level of organization above that of Protista is characterized by multicellularity and tissue development and the basic divisions are defined on the basis of nutrition because of its far-reaching consequences on structure and function. Nutrition is absorptive in Fungi. Therefore, and because of structural adaptations to their mode of life, they deserve separation from the other kingdoms. In the kingdom Plantae *sensu stricto* it is mainly photosynthetic, with nutrients obtained from water, 'soil' and air. The organisms are basically autotrophic and sessile and should include the multicellular 'higher' algae. The fact that nutrition is by ingestion of captured food in the kingdom Animalia raises the multifarious problems of locomotion. It has been suggested that with the Metazoa this kingdom should also include the small groups of Mesozoa and Parazoa, with absent or limited tissue differentiation. The Mesozoa are without fossil representatives and will not be considered here. The Parazoa (sponges) are here included with the Metazoa as animals. For the rest, widely accepted formal classifications of the kingdom Animalia will be used in this study. It is not its aim to criticize or develop them.

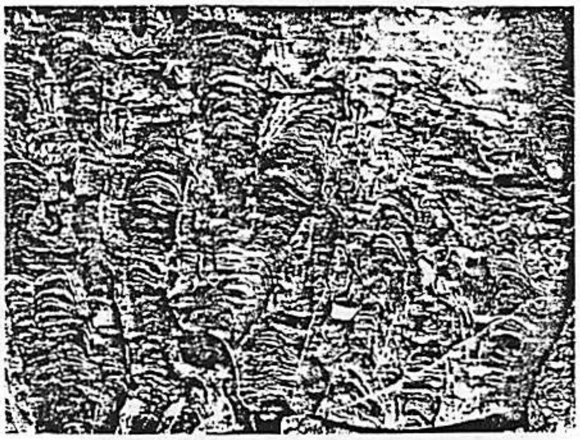
1.5 Proterozoic fossils and environments

At the beginning of the Proterozoic Eon, autotrophic prokaryotes (bacteria and blue-green algae) were in existence. The most striking and generally characteristic feature of the Early Proterozoic fossil record is the first abundant and widespread occurrence of stromatolites. It is related to and probably partly the cause of the greater abundance of limestone and dolomite deposition in shallow water. The first result of systematic studies of Precambrian stromatolites (Fig. 1.5) was the recognition of their biostratigraphic potential. Early Proterozoic (Aphebian), Middle Proterozoic (Early and Middle Riphean, stromatolite assemblages I and II–III, respectively) and Late Proterozoic (assemblages III and IV) were distinguished (see articles by Semikhatov, Preiss and Donaldson in Walter 1976, Chapter 7). The numbered assemblages (specified by Preiss in a note on p. 368 of his article) were described first in the USSR and later found in other regions. The subsequent studies of Aphebian stromatolites revealed considerable similarities in their configuration on the level of form genera but also sufficient specific structural differences to confirm rather than negate their stratigraphic potential, provided that their study was carried out in necessary detail. Due consideration had to be given to environmental, preservational and diagenetic factors. At the same time, considerable efforts were made to find and describe the fossil remains of the algal communities which had contributed to the construction of stromatolites (Schopf *et al.* 1977, Schopf & Prasad 1978) as well as non-stromatolitic occurrences of Proterozoic fossil algae. It was reported that about 20 distinctive assemblages of microfossils occur in biostratigraphically useful sequences, about 10 of them in strata which are more than 1000 m.y. old.

The abundance and diversity of stromatolites declined sharply in Late Precambrian time, from about 800 m.y. ago onwards (Gebelein in Walter 1976). The fact that they had flourished until that time points to the absence or insignificance of Metazoa and Metaphyta in their environment (photic, subtidal to intertidal) in which shelf-type limestones have been deposited. In subsequent times the predation by Metazoa together with competition by red and green algae and other marine plants confined the stromatolite builders mostly to hypersaline and freshwater environments. This occurred rapidly during Early Palaeozoic time and continued to the present when they survive as relicts.

The major primary producers of food resources for heterotrophic

Fig. 1.5. Stromatolites. *Gymnosolen ramsayi* Steinmann, from a boulder in the Late Proterozoic Tapley Hill Formation, Flinders Ranges, South Australia. Photo by courtesy of Dr W. V. Preiss.



organisms in the sea are the photoautotrophic phytoplankton organisms. It is difficult to establish their fossil record through the great time spans of the Proterozoic. It has been suggested that their quantitative fluctuations in Phanerozoic times had a decisive influence on the evolution of the marine fauna (Tappan & Loeblich 1971). The present level of our knowledge does not permit reliable extrapolation into Precambrian time, for a number of reasons. Planktonic algae belonging to extant taxa make their first appearance in Phanerozoic strata: the chrysophycean coccolithophorids probably in the Jurassic, the diatoms in the Cretaceous, the dinoflagellates possibly in the Silurian or Permian and definitely in the Triassic. This makes the identification of Precambrian microfossils as members of extant taxa of Protista of the marine plankton impossible, in contrast to some extant Cyanophyta which can be recognized as belonging to it, with considerable probability, in Precambrian time. On the other hand, large numbers of organic walled microfossils are known which are placed in the 'Group Acritarcha' (Evitt 1963). This name indicates only that their affinities are unknown. Precambrian acritarchs are rarely older than Late Proterozoic. They are mostly simple spherical bodies with thin, smooth, organic envelopes. They are classed in a subgroup Sphaeromorphitae (Downie, Evitt & Sarjeant 1963). Their biological functions are unclear. Often assigned to this group, the genus *Chuaria* Walcott, 1899 has a remarkable geographic range throughout North America, northern Europe, the USSR, Iran, China and India. It occurs generally in considerable abundance in certain horizons of the Late Proterozoic and Vendian (Ford & Breed 1973, Hofmann 1977, Hofmann & Aitken 1979, Vidal 1981). It has a remarkable size range, from 0.5 mm to 5 mm diameter, according to Ford & Breed. A closely similar Late Proterozoic (mid-Vendian) fossil *Beltanelloides* Sokolov (Fig. 1.6) measures from 5 to 45 mm (Sokolov 1974). Its size makes it even more difficult to understand its growth and function as a phytoplankton organism. Sokolov considered it originally to be a 'medusa' but as it is a simple, originally spheroidal or more likely discoidal envelope without openings or appendages it is certainly not a coelenterate nor any other kind of animal, as Sokolov came to note later.

Fig. 1.6. *Beltanelloides sorichevae* Sokolov (= *Beltanelliformis brunsae* Menner), from a bore at Leino, eastern Russian Platform, Vendian, Redkino Series. From Sokolov (1972*b*). Magnification: × 3.

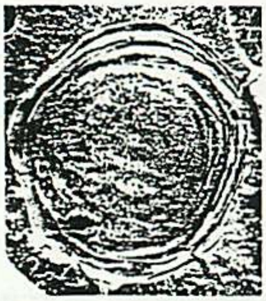
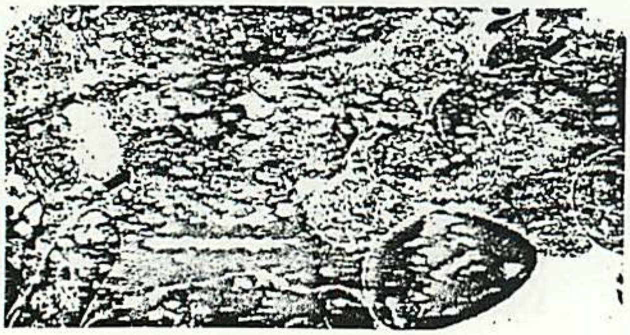


Fig. 1.7. Convoluted fecal string (arrowed) of a sediment-feeding metazoan. Upper Riphean, Ural Mountains. From Sabrodin (1971). Scale bar, 1 mm.



The earliest fossils which could be considered as animal remains (see review in Glaessner 1983) are unfortunately poorly known and incompletely described. Microfossils resembling Chitinozoa which may represent Protozoa or Metazoa have been mentioned (p. 20). Other microfossils occurring in the Lakhandin 'Series' of Siberia (950–1000 m.y. old) appear to be small metazoan organic structures. They were extracted by B. Timofeev and T. N. German from laminae of organic matter. They are still undescribed. All other metazoan fossils of that age are traces of organic activities rather than body fossils. A sigmoidal string of what appear to be fecal pellets about 0·37 mm in diameter (Fig. 1.7) from the middle Zilmerdak 'Series' (lower part of Upper Riphean of the Ural Mountains) was figured by Sabrodin (1971, 1972), together with a possible burrow. Clemmey (1976) described what he believes to be burrows from the Middle Roan formation of Zambia (about 1000 m.y. old) but Cloud (1978*a*; Cloud, Gustafson & Watson 1980) considers them as Recent termite burrows. Other questionable traces of bioturbation were recorded by Squire

(1973) from the Brioverian of the Channel Islands, considered to be about 750 m.y. old. Beer (1919) described a spiral measuring 13 × 11 mm preserved in relief on the lower surface of a slab of Rhotas Limestone of the uppermost Vindhyan of northern India which should be re-examined as it appears to be a trace fossil of metazoan origin. However, a Cambrian age of the Upper Vindhyan is possible. Supposed trace fossils from the Belt Supergroup in Montana (USA) have been redescribed as megascopic algae, aged 1300 m.y. (Walter, Oehler & Oehler 1976). A possible trace fossil from the Lower or Middle Riphean equivalents in the Ukraine, from a formation resting unconformably on a weathering horizon on volcanics which are 1000–1400 m.y. old, was described as *Rugoinfractus ovruchensis* by Palij (1974). Durham (1978, p. 37) considered it to 'merit special attention'. It is, however, also comparable with infillings of desiccation (syneresis) cracks which have been described repeatedly from Precambrian and younger rocks. The interpretation of a presumed medusoid fossil *Brooksella canyonensis* Bassler was questioned by Cloud (1968). Its subsequent assignment to the trace fossils (Glaessner 1969*a*) on the basis of its resemblance to *Asterosoma* was made more likely by Cloud's discovery of a second specimen. This interpretation as a possible 'infaunal deposit feeder' (Brasier 1979) was apparently confirmed by Kauffman (in Kauffman & Steidtmann 1981, p. 925) who gives the age as 1100–1300 m.y. although it is often stated as 900–1000 m.y.

The record of metazoan life from Early and Middle (?) Proterozoic rocks is apparently non-existent and that from 1000 to about 680 m.y. ago (Upper Riphean and equivalents) exists but is extremely poorly documented. The coincidence of the absence of metazoan remains in the early Middle Proterozoic and of definite eukaryotes favours the view that eukaryotes originated in mid-Proterozoic time or about the beginning of the Late Proterozoic, not in Early Proterozoic nor in Vendian time. We cannot yet assert as a fact that deposit feeders existed about 1000 m.y. ago but there are indications which make it more likely than the assertion of their absence. A fuller understanding of the distribution of structured fecal pellets in sediments should reveal the presence of sediment feeders more clearly than bioturbation, which may require more efficient structural adaptations of its originators to enable them to leave recognizable traces in ancient sediments. On the other hand it is probable that the earliest Metazoa were small detritus feeders. They may have developed cuticles of structural proteins which could be preserved together with other organic debris and discovered by palynological methods (acid treatment and bleaching) in unmetamorphosed sediments.

The Vendian faunas. The latest Precambrian (Vendian, about 680–560 m.y. ago) presents a clear picture of distinctive palaeontological documentation of a stage of metazoan evolution. Metazoans are widespread and locally abundant in rocks of this age (Chapter 2). Extended and thorough investigations have shown that a considerable number of sequences of Precambrian strata of that age contain distinctive fossils. Most of them became known only in the last 20 or 30 years. They have very little in common with the succeeding Cambrian faunas. The most noteworthy facts about the Metazoa of the uppermost Precambrian are the general absence of mineralized tissues such as calcareous or siliceous shells and the presence of representatives of several phyla, most of them extant in the living fauna. These are the Cnidaria, Annelida, Echiura, Arthropoda, Pogonophora (?) and possibly one or two others (Glaessner 1979*a*, *b*). The absence or probable absence of some phyla which are well represented in the Early Cambrian fauna, such as Porifera, Archaeocyatha, Mollusca, Brachiopoda and Echinodermata, and of the Platyhelminthes which on theoretical grounds are believed to have existed in Late Precambrian time, is noteworthy. The absence of Bryozoa and Chordata which make their first appearance in later Cambrian time should also be noted.

Late Precambrian metazoan body fossils are at present known from about 20 different regions in Africa, Asia, Australia, Europe and North America (Fig. 1.8). In some of these regions (southwestern Africa, South Australia, northern Russia, Ukraine) there are numerous localities where such fossils have been found, while from others we know so far only single finds. At the present time some of these fossils are regrettably still undescribed and some descriptions are still unpublished. Most areas have some distinctive kinds of Precambrian fossils in common but from South Carolina we know at present of an occurrence of a worm-like organism not found elsewhere (Cloud, Wright & Glover 1976). It is remarkable that even a few species are common to areas which are now as far distant geographically as South Australia, southwest Africa and northern Russia. A possible explanation for this fact will be discussed in Chapter 3. However, there are marked regional differences and we are not dealing with a cosmopolitan fauna in a biogeographic sense. There is evidence of considerable palaeoecological differences between fossil assemblages. What they have in common, however, is their distinctive level of diversification, with possible indications of increasing diversity within the time span of the fossiliferous Late Precambrian strata at the different localities. To give an order of magnitude to the possible time span, we anticipate here

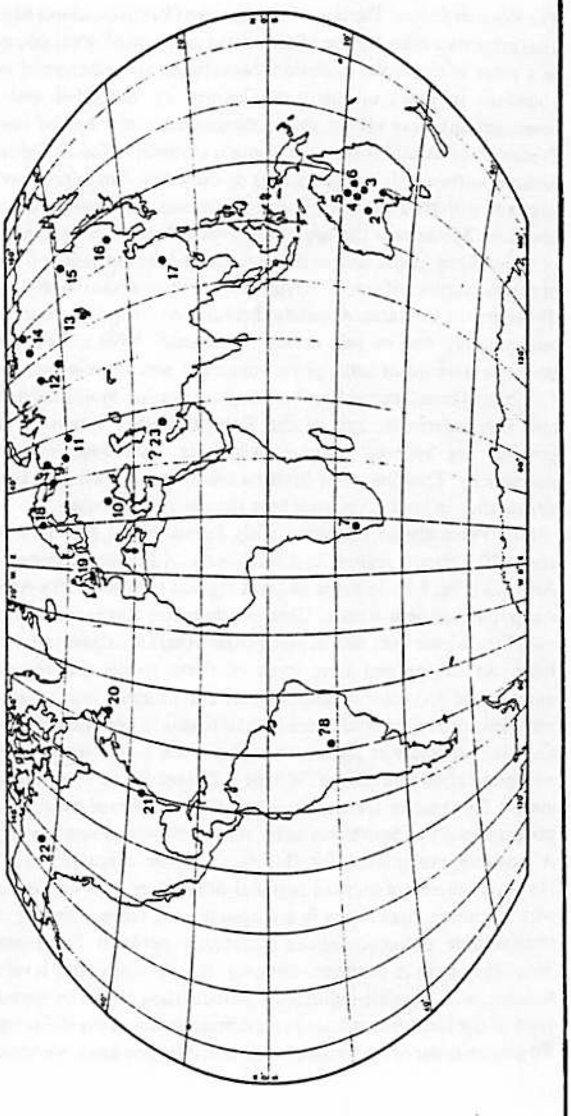


Fig. 1.8. Worldwide occurrences of Ediacarian fossil Metazoa. *Australia*: 1 – Ediacara, Flinders Ranges; 2 – Punkerri Hills, Officer Basin; 3 – Deep Well, southeast of Alice Springs; 4 – Laura Creek, southwest of Alice Springs; 5 – Mt Skinner; 6 – Jervois Ranges, southwest Georgina Basin. *Africa*: 7 – southwest Africa/Namibia. *South America*: 8 – Mato Grosso, southwest Brazil. USSR: 9 – northern Russia; 10 – southwest Ukraine (Podolia); 11 – western Urals; 12 – Yenisey River (Igarka-Turukhansk); 13 – Irkutsk (Lake Baikal); 14 – Anabar and Olenek (northern Siberia); 15 – River Maya [illegible] 16 – northeast China (Longshan); 17 – Yangtze Gorge. *Northern Europe*: 18 – Lake Torneträsk (Sweden); 19 – English Midlands (Leicester). *North America*: 20 – southeast Newfoundland; 21 – North Carolina; 22 – northwest Canada (Mackenzie Mountains); 23 – central Iran. (Several localities without body fossils were omitted. Sequence of numbers follows text. Material from all localities except 8, 13, 15, 18, 22 has been available to me for study.)

the results of an analysis of the datings of individual finds which will be discussed later. All these finds represent a time range of no more than 100 m.y., from about 580 to 680 m.y. before the present. It may be shorter but it is unlikely to be longer. It is no more than one fifth of subsequent Phanerozoic time, equal to the time which elapsed since the mid-Cretaceous. The fossiliferous rocks from this interval of time represent a variety of marine biotopes, but some of those which exist now, particularly those created by later differentiation of the biota, e.g. reefs, are absent, as are dark shales like the uncommonly richly fossiliferous Middle Cambrian Burgess Shale. The absence of most of the predators occupying the higher trophic levels of the modern marine faunas is of fundamental significance for the evolutionary level reached in Late Precambrian time. The definition of Metazoa as multicellular heterotrophs feeding by ingestion presupposes the presence of primary producers, presumably planktonic Protista, of which we have little direct knowledge. The abundance of life which can be documented at least at some sites must have led to an abundance of organic detritus and in turn to an abundance of detritus feeders, some of which can be documented by trace fossils of Late Precambrian age.

Between the very fragmentary record of traces of metazoan life about 900–1000 m.y. ago and the relatively abundant record of the latest Precambrian (680–580 m.y. ago) there is a significant time gap. It may be partly filled when more attention is given to a detailed search for traces of animal life activities in sediments 700–900 m.y. old. Pending further discoveries, we have to consider the possible effects of environmental changes during that time. In this connection we note that it is the time of the first of a number of Late Precambrian glaciations.

The sedimentary record of mid-Proterozoic time (1700–1000 m.y. ago) does not show any significant deviation from that of later times. It gives no grounds for assuming that the evolutionary origin of Eukaryota and the subsequent emergence of animal life, either as single-celled Protista or later as Metazoa, was due to some fundamental change in the environment. Once the oxygen level permitting oxidative respiration was reached, the adaptation to this efficient energy source was controlled not so much by changes of the external environment as by advances in the biochemistry and physiology of the organisms. We can speak of a phase of cytological evolution without contradicting any basic assumptions of modern selectionist (neo-Darwinian) evolutionary theory. Selection would favour physiologically more efficient organisms functioning in an unchanging environment. The significance of cytological changes in cell division leading to meiosis and redistribution of hereditary material by sexual processes in early eukaryotes has been discussed (Schopf *et al.* 1973, Stanley 1976*b*). While palaeontological evidence is not as unambiguous as was once thought, it is clear that these steps in cytological evolution were determined by selective processes favouring more efficient reproduction and increased genetic variability in populations, without changes in their environment.

The picture changes when we approach the latest Proterozoic. Worldwide glaciations of Late Riphean and Early Vendian age (Chumakov 1978, 1981) are dated, at various localities, from not less than about 800 to about 650 m.y. before the present. Abundant evidence of metazoan body fossils is found in sediments which are *younger* than those containing evidence of the latest glaciation, in North America, northern Europe, China, southwestern Africa (where there may have been later cold periods) and Australia (Glaessner 1977) (see p. 102). This observation is not only of stratigraphic significance but is also likely to reflect influences on the history of the biosphere. The existence of rare traces of animal activities locomotion and feeding) and of still unidentified microscopic fossils of probable animal origin recorded from preglacial Late Proterozoic (Upper Riphean) sediments 1000–800 m.y. old supports the view that the Ediacarian fauna does not represent the earliest Metazoa. Even without these data, this is obvious from the relatively high level of metazoan differentiation. A considerable interval of time was required for metazoan 'prehistory'.

1.6 Origins and early differentiation of the Metazoa

Some zoologists have expressed the view that the study of fossils is useless for investigation of phylogeny because of the incompleteness of the fossil record. They may have overlooked the fact that zoological knowledge of the great majority of a million living species is incomplete in respect of vital facts of their genetics, molecular biology, ethology and even morphology. That does not necessarily limit the value of their research or make it essentially speculative. The existence of certain kinds of animals, whether expected or unexpected, whether well or incompletely known, at certain times in the past must be taken into consideration. Without historical background, without palaeobiology, the pursuit of biological science must remain incomplete. The study of the fine structure of mineralized tissues of living organisms was aided and stimulated by palaeontological work, and the systematics of many major groups of animals was significantly modified and refined by the integration of palaeo- and neobiological knowledge. What has been discussed so far in this chapter was based on integration of theories and observations in many fields with a fossil record that was until recently thought to be non-existent. It was shown to exist although it is less complete the further back we go into the past. It is appropriate to speak of a prehistory of the biosphere, as we speak of the prehistory of mankind when we go back beyond the periods for which we have written documentary records. In order to understand the earliest preserved abundant assemblages of Precambrian Metazoa in relation to the history of life at earlier Precambrian times, we shall briefly consider some of the many biological theories of origins of Metazoa and their early differentiation. These theories are essentially phylogenetic since we are trying to trace the course of evolution, the evolutionary pathways from one group or groups, extinct or still living, to others. Phylogenetic theories and the theory of phylogeny on a supraspecific level are once again becoming respectable subjects of scientific enquiry and debate (Dougherty 1963, Brien 1969, Eldredge & Gould 1972, Jägersten 1972, Gould 1977, Hallam 1977, Hanson 1977, Heberer 1967–1971, Riedl 1978, Stanley 1979, and many others), but this is not the place for a critical review of evolutionary or phylogenetic theory. The varieties of applied methodologies and of the results presented, often in the graphic form of taxa linked by presumed lines of descent, is bewildering. Approaches and suggestions which appear to be less controversial than others are here considered with the specific purpose of providing a framework for the linking of prehistorical and historical data on the early evolution and differentiation of animals. The outline of our knowledge of Precambrian fossils presented above makes it clear that phylogenetic discussions relevant to them must be kept on the level of higher taxa. No significant evidence for evolution at the species level in Precambrian time has been discovered. Omission of discussion at this level does not necessarily imply reliance on special mechanisms of 'macroevolution' as distinct from evolution of species.

Hanson (1977) has laid down three 'procedural guidelines' for phylogenetic analysis. Taking for granted, without questioning it, because it is considered as outside the scope of our discussion, the first rule 'Species define the basis for interorganismic comparisons', we turn to the second one. It says: 'Phylogenetic comparisons must examine the exploitive, homeostatic, and reproductive functions and/or their anatomical correlates whenever possible'. The third rule states that 'Every innovative step must be selectively advantageous'. These guidelines (a better term than rules) are necessary constraints which can lift phylogenetic investigations above the level of speculation. They are here rephrased in order to make them self-explanatory. Hanson's third rule means that every innovative step implied in the proposed phylogeny should be seen as a possible result of a selective pressure. Rule two means that phylogenetic analysis is based on functional morphology, the primary functions being (*a*) bioenergetically and environmentally conditioned (*b*) reproductively advantageous in their result and (*c*) constrained by homeostatic principles. Homeostatic in this context means preserving constancy of internal environment generally, and of conditions for development and growth specifically, during evolutionary changes in the form–function complex. The significance of the basic functions of the animals and their expression in morphology are discussed in most textbooks of biology and need no further explanation here. The homeostatic principles guiding phylogeny concern, firstly, the existence of homologous structures in different organisms. The concept of homology cannot be put aside as allegedly based on circular reasoning. It is an essential operational tool of phylogenetic analysis (Boyden 1947, Remane 1956, Simpson 1961, Peters 1972, Hanson 1977, Riedl 1978, Rieppel 1980). Secondly, homeostatic principles underlie the channelling of development discussed extensively by Waddington in many of his works, and ultimately the relations between ontogeny and phylogeny (de Beer 1951, Gould 1977). Again, this has been brushed aside mistakenly, as based on the supposedly long-discredited biogenetic 'law' of Haeckel. Thirdly, they underlie the morphotype or better groundplan or structural plan of organisms, a concept which is more often invoked in German than in Anglo-Saxon literature. It has no necessary epistemological connection with idealistic morphology but can be based on genetics and information theory as well as on considerations of biophysics and materials science as applied to organic form and function (Pantin 1951, Gould 1970, Peters & Gutmann 1971). The necessity for every part of an organism to function in accordance with structural and mechanical principles, within the constraints of strength of available materials, and to preserve ability to function throughout all stages of growth and in response to all tolerable environmental stresses and the exigencies of competition is generally admitted. All this can be expressed briefly as structural plan ('Bauplan') or its synonyms but care must be taken to avoid the literal, typological, theological or teleological implications of the words 'plan' or 'type'. When reference to engineering principles is made, it should be remembered that engineers order materials according to specifications, and mathematical analysis precedes design. Evolution proceeds with inherited material, by trial and error. It has often been said that it is 'opportunistic' in the choice of environments for its products. Careful consideration of functional plans in the sense here indicated is particularly important when the deformable soft bodies of extinct Precambrian Metazoa have to be interpreted as formerly functioning organisms.

The gap in fossil evidence between Metazoa and their presumed protozoan ancestors can be filled at the present time only by biological theory. Hanson (1977) is mainly concerned with the interrelations of the Protozoa and their possible origins from other Protista, most of which remain unidentified. Concerning metazoan origins, he derives the Porifera from choanoflagellates, which is a widely accepted view, and the turbellarian platyhelminths from primitive ciliates, considered as descended from unidentified zooflagellates. The Cnidaria are questionably derived from amoeboid zooflagellates. The very large amount of factual data from the living Protozoa and lower Metazoa contained in Hanson's voluminous and thoughtful work cannot be reviewed here. While it throws as much light on the world of animals at dawn as can be derived from present knowledge of their living representatives, it stresses repeatedly, as various groups are reviewed, the scantiness or total absence of their palaeontological record. On present knowledge and with present methods, the distinction between Protozoa and other Protista on which Hanson insists (for other views on systematics see Whittaker & Margulis 1978) cannot be made for those microfossils which have no living representatives. Precambrian microfossils not assignable to blue-green algae, eukaryotic algae and fungi may have no living representatives. However, considered views on the possible origin of Metazoa from Protozoa can be used as a guide to further search which may reveal modes of fossilization as unexpected as that of the rich Gunflint assemblage of Prokaryota.

Comparative embryological studies of phylogeny of the lower invertebrates were critically reviewed by Ivanov (1968). His short but important book is concerned with the origin of the multicellular animals. A critical history of phylogenetic theories is followed by an exposition of the author's own views (Fig. 1.9). They evolved from the work of the famous Russian school of embryologists of the last century (Kovalevsky, Metchnikoff and others) and their modern successors, including the comparative morphologist Beklemishev (1964). Work in other countries, from Haeckel to Hanson, up to the 1960s, is given due attention. Ivanov's proposed stages in the phylogeny of the Metazoa can be considered as a model bridging the gap in metazoan prehistory to which reference was made above (p. 30). It can be tested, to some extent, at least against biochronological and palaeobiological facts, if not against palaeomorphological data. One of the reasons for the absence of such data becomes obvious when the (unstated) scale of size of the organisms shown in Fig. 1.9 is considered. In dealing with the problem of the derivation of small multicellular from single-celled organisms, Ivanov argues convincingly against theories of cellularization (i.e. the separation within multinucleate protistan cells of compartments destined to become cells of Metazoa). This course of evolution is still

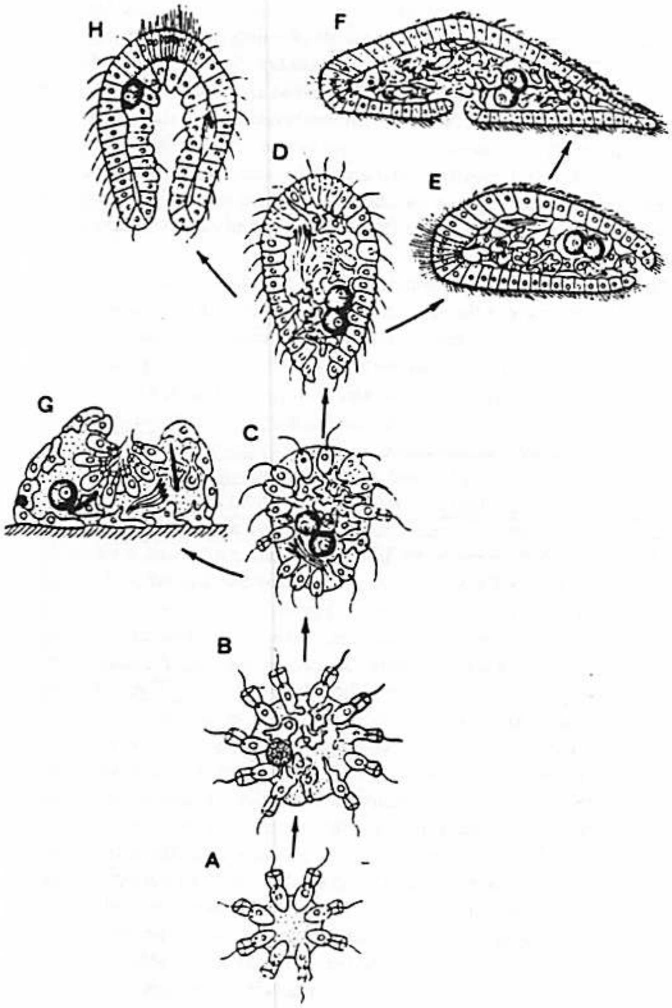


Fig. 1.9. Hypothetical stages in the phylogeny of the lower Metazoa. (After Ivanov 1968, Fig. 118; 1970, Fig. 8.) A. Colonial choanoflagellate (*Sphaeroeca*-type); B, the same, *Proterospongia*-type; C, early Phagocytella; D, later Phagocytella; E, ancestral benthic turbellarian flatworm with mouth and bilateral

accepted by Hanson. Ivanov starts his phylogenetic construction with a free-swimming spherical colony of undifferentiated animal flagellates, resembling the living *Sphaeroeca* belonging to the class Craspedomonadida (or Choanoflagellida). Reproduction was asexual, by separation and division of individual cells and formation of new colonies. This stage in evolution is followed by a colonial stage with increased integration of the cells. Somatic and reproductive cells are differentiated, sexual reproduction leads through cleavage of the zygote to the formation of a blastula-like organism with multiplication of somatic cells by mitotic division and development of radial symmetry. *Proterospongia haeckeli* Kent is said to show some approximation to this stage. It is a rare organism assigned to the choanoflagellates and considered to be close to the ancestry of Porifera. It requires further study (Ivanov 1968, p. 223, Hanson 1977, p. 452), particularly concerning its reproduction. The next stage is frankly hypothetical as an adult form. It is based on Metchnikoff's Phagocytella which he had earlier named Parenchymella and which is similar to the planuloid stage in the ontogeny of various lower Metazoa. In Metchnikoff's and Hyman's (1940) view it is considered as a free-swimming organism, a possible ancestor of the Eumetazoa. During its embryogeny a blastula-like larva was supposedly formed which grew through cell division. Differentiation of ecto- and endoderm occurred through multipolar immigration of specialized cells into the interior of the organism. Among them were phagocytes moving in a parenchyma. Thus there is at that stage a differentiation into an external, locomotory *kinetoblast* and an internal, digestive *phagocytoblast*. It is noteworthy that Ivanov accepts the notion of a solid Phagocytella originally without a mouth while objecting firmly to Haeckel's generalized gastraea theory and describing Lankester's assumption of a digestive cavity in a 'Planula' without a mouth as 'obvious nonsense' (Ivanov 1968, p. 133). However, according to Ivanov a mouth is necessarily formed subsequently in this Phagocytella at the posterior end of this free-swimming organism where food particles in the water driven by the flagella of the kinetoblast congregate in the 'backwash'. It follows from this view that the Porifera originated prior to this stage, becoming totally sessile, with the kinetoblast transferring its hydrokinetic

Caption for Fig. 1.9 (*cont.*)
symmetry; F, ancestral acoeloid turbellarian, with enhanced cell differentiation and ventral mouth. Sidelines from stage B leading to G, a primitive sponge, sessile, locomotion replaced by hydrokinetic function; and from stage D leading to H, a primitive coelenterate (Gastraea). (No scale, note small number of cells in each section, hence microscopic size of organisms.)

function into interior cavities of the organism. The Turbellaria Acoela are also close to the 'Phagocytella' but remained motile by ciliary locomotion and originated after it had evolved a mouth. Their internal organization, in the absence of an intestine, corresponds according to Ivanov (1968, p. 263) to the phagocytoblast of Metchnikoff's Phagocytella. This hypothetical but potentially functioning organism also gave rise to the swimming Ctenophora and, through adaptation to a sessile mode of life, to the Cnidaria. Ivanov considers single hydroid polyps as their ancestral forms. All 'coelenterates' have acquired diploblastic organization. Only their ancestral forms are described by Ivanov as 'gastraea-like'.

Jägersten (1972) expanded his earlier 'Bilaterogastraea' theory into a complete phylogeny of the Metazoa based on an untestable hypothesis. He states that two phases of their life cycle arose when 'the adult of the primeval ancestor of the metazoans, viz. the holopelagic, radially symmetrical Blastaea, descended to life on the bottom' and became bilateral as a Bilaterogastraea with coelomic pouches derived from gastric pouches of coelenterates. The 'juvenile stage remained in the pelagic zone' (p. 216) in most metazoans, not by adaptation to dispersal but because of recapitulation of the postulated primitive ontogeny. Jägersten's simplistic and formalistic theory on which Hyman (1959, p. 759) made cogent critical remarks but which has found some supporters, implies incipient formation of a coelom and metamerism in coelenterates. This was unacceptable to Clark (1964, 1969, 1979) whose views are based mainly on functional considerations, relating primarily to locomotion as a distinctive basic problem in metazoan diversification. Clark's position on the earliest history of the Metazoa is somewhat agnostic (1979, p. 63) as was Hyman's earlier conclusion (1959). Hanson (1977) does not accept the Turbellaria Acoela as direct ancestors of all other Metazoa including the coelenterates, considering with the great majority of zoologists Hadži's (1963) theory as 'discredited', but he places a question mark at the origin of the Cnidaria, where others (Hand 1959, Hyman 1959, Ivanov 1968) had placed a planuloid organism. Clark (1979) concerns himself with the radiations of the acoelomates and coelomates which we shall discuss (in Chapter 3) after a review of palaeontological data and their chronological implications (Chapter 2). However, several theoretical viewpoints must be briefly considered before proceeding to these matters.

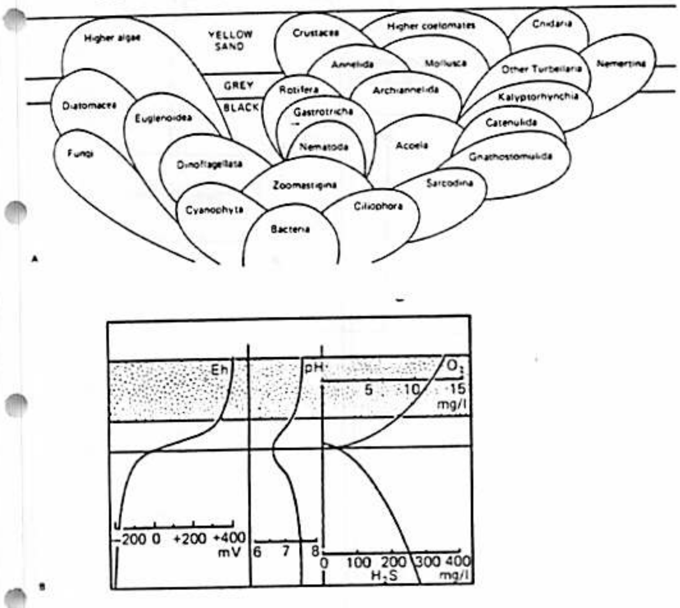
(1) Phylogenetic theories. A generalized historical background to the still controversial questions of invertebrate phylogeny was given by Clark (1964) who later reviewed (1979), in the light of his studies on coelom formation and metamerism, arguments and constructions presented in the course of lively debates in the 1950s and 1960s from the viewpoints of comparative functional morphology and embryology (or developmental anatomy). This was followed by a fierce debate, mainly in the German literature, between classical morphologists (Reisinger 1970, Siewing 1972, Ulrich 1972, Remane 1973) and the 'Frankfurt School' (see Gutmann 1977, 1981, and references therein) whose views are based dominantly and almost exclusively on functional–biomechanical analysis and model construction. This continuing debate will not be reviewed here (although it is significant and mostly constructive) because it was so far largely concerned with the origin of the chordates. This is essentially outside the scope of our examination of the Precambrian–Cambrian evolution of the Metazoa.

Phylogenetic hypotheses concerning the transition from Protozoa to early Metazoa have advanced from early speculations or subsequently refuted observations to more sophisticated models proposed in recent years. Combined evidence from histology, embryology, comparative and functional morphology, physiology and ecology of living organisms is used extensively. *The construction of hypothetical ancestral organisms on a purely morphological or embryological basis without full exploration and explanation of their functioning* 'ab initio' *and of the selection pressures leading to their origination and transformation into descendant taxa is being discouraged.*

(2) Palaeoecological considerations. There is at present hardly any evidence for the nature of the Late Precambrian metazoans prior to the radiations which the Ediacarian faunas reflect and which will be discussed later (Chapter 2). There are, however, hypotheses about how and where the early metazoans lived. Palaeoecological theorizing based on living taxa of the lower Metazoa is notoriously hazardous. Their survival may be due to their limitation to environments hostile to predators when the latter appeared later. An example among Prokaryota is the 'survival' of columnar stromatolites in hypersaline environments which tells us much about stromatolites and their cyanophyte builders but little about the habitat of their Precambrian precursors. This example does not necessarily lead to a refutation of the hypothesis that some of the adaptive radiations of the lower Metazoa occurred in the reducing environments existing within the substratrum of marine sediments. The planuloid hypothesis concerns the origination of the early Metazoa but does not explain their differentiation. Some of them would have contributed to the zooplankton while others could have descended to the sea floor and developed the ciliary creeping habits of Platyhelminthes. The organic remains of microplankton are incorporated in the sediments. In the absence of mixing with oxygenated water, or when the oxygen is exhausted by their decomposition, an anoxic environment is produced below the sediment/water interface. This is now

known to be occupied by a varied biota (Fenchel & Riedl 1970, Rhoads & Morse 1971, Boaden 1975, Clark 1979). This infaunal biota (Fig. 1.10) includes Platyhelminthes (Turbellaria Acoela and Catenulida) and Gnathostomulida in the sulfide biome or thiobios of near-shore and marginal marine zones. Representatives of additional, more advanced taxa occur when and where more oxygenated sediments are deposited above reducing black muds. Considering the subsequent colonization of the hostile sulfide-enriched environment by secondarily and convergently adapted members of other phyla as a subsequent modification of the biota, Fenchel & Riedl, supported by Boaden, conclude that the occurrence there of many primitive Metazoa is a relict of a biosystem which preceded the fully aerobic biosphere. The probable attainment of an atmospheric level of at least 1% PAL of oxygen about the time of the origin of the Eukaryota makes it feasible to link the acoelomate radiation with infaunal habitat and to consider some of the present sulfide biota as hypothetical relicts of Late Precambrian (pre-Vendian) faunas. The absence of Cnidaria from such environments indicates not a later origination of coelenterates but other, concurrently occurring adaptive radiations in co-existent, more oxygenic environments. At any time during the development of the oxygen content above 1% PAL there were opportunities for the capture of nutrients from primary producers, i.e. the phytoplankton anywhere in the water column from the surface to the bottom. Sediments formed at the relevant time indicate the existence of normal conditions of oxygenation of sea bottom water: black shales which formed under reducing conditions are not prevalent among the terrigenous formations of the Upper Proterozoic, hence the sulfide biome was not an exclusive or particularly likely environment for evolutionary developments. An hypothesis requiring this lacks geological or palaeoecological verification.

Fig. 1.10. Depth distribution of benthic marine taxa in relation to deoxygenation of sediments. A: The names are plotted according to maximum depth of occurrence relative to the 'redox potential discontinuity layer', here labelled 'grey'. Lines around names suggest 'relationship between depth and evolutionary origin' according to the author of the figure (Boaden 1975). B: Schematic representation of Eh and pH profiles in sediment from near-shore and restricted marine environments. Part of Fig. 4 of Fenchel & Riedl (1970). The two horizontal lines across the middle of the figure delimit the 'grey' zone of A. (Rhoads (1974) has reversed the terms 'black' and 'grey'. Note that these sediment layers do not normally represent rock sequences but presently observed stages of syndiagenesis of sediments. Under unchanging conditions the zones and their biota will move upward with continuing sedimentation, producing dark, organic-rich clastics.)

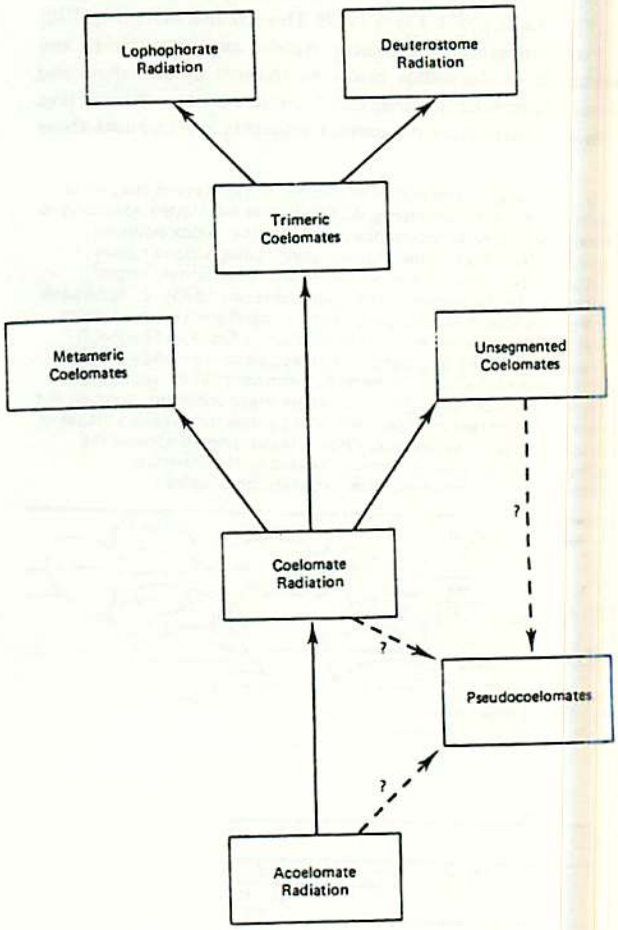


Summarizing the present evidence from biological observations for metazoan prehistory we reach the following result. Boaden's (1975) conclusion that the Gnathostomulida, Platyhelminthes and Aschelminthes originally evolved under dysaerobic conditions can be accepted, assuming that dysaerobic means atmospheric oxygen levels of a few percent of the present level. Boaden's further conclusion that the early Metazoa were benthic anaerobes does not necessarily follow from either his own or geological evidence. The rate of oxygen accession to the atmosphere and to bottom waters and the distribution of oxygenation in benthic habitats at the relevant time are not sufficiently well known to support conclusions based on observations in the present near-shore thiobiome alone. There is no need to confine early metazoan evolution to this type of habitat. The later origin of the coelenterates compared with that of the Bilateria is also unacceptable. The basically sessile benthic coelenterates were not the likely ancestors of the locomotory, bilateral metazoans. It should be remembered that for good reasons Hyman (1959, pp. 750 ff) considered the derivation of the coelom from the coelenteric pockets of the Anthozoa (rugose corals), which was updated by Remane from nineteenth-century notions and accepted by some of his followers, as 'fantastic nonsense' and Jägersten's bilaterogastraea theory as 'absurd'. For these reasons *Ivanov's hypothetical phylogeny of the earliest Metazoa is here considered as acceptable and Clark's discussion of subsequent radiations will be the basis of further comparisons with documented events of Late Precambrian time* (Chapter 3).

(3) Relative timing of metazoan evolution in the Late Proterozoic. A common assertion in many phylogenetic theories is the independent origin of the Porifera from choanoflagellates. Lacking organized tissues and possessing the potential of regenerating the entire organism from the smallest fragments, they are often separated from the Eumetazoa as belonging to a subkingdom Parazoa (Margulis 1974). The implication is that their origination would be an early event in the history of animals. The fact that most of them possess mineralized spicules would lead to the expectation that their remains would be found in Late Precambrian sediments. This expectation has not been fulfilled. One or two reports of Precambrian sponge spicules have not yet been disproved but they are questionable. Three explanations of these observations are possible. (i) The sponges could have evolved early with bodies strengthened by organic fibres without significant fossilization potential. Abundant living sponges possess only spongin fibres. (ii) The sponges with spicular skeletons could have evolved early but remained undetected in the fossil record because of small size and lack of distinctive characters. (iii) They could have evolved at the end of Precambrian time when the diversification of many animals with mineralized tissues took place. Assumption (iii) is supported by analogy with a similar late development in the Foraminifera and Radiolaria which evolved rapidly from few structurally primitive early Palaeozoic representatives. This hypothesis would not necessarily exclude (i) if in fact the Porifera without mineralized spicules can be considered as ancestral to spicule-bearing major taxa.

The second postulate is the independent early origination of the coelenterates which are also known as Radialia or as diploblastic Metazoa. No fossil Ctenophora are known and therefore only the phylum Cnidaria will be considered here. The high degree of taxonomic diversity attained by the cnidarians in the Ediacarian faunas (Chapter 2) indicates a long preceding period of existence. The Phanerozoic record of soft-bodied Cnidaria is remarkably patchy. Suitable conditions of preservation could have been similarly rare in Precambrian time. The evolutionary rate of diversification being unknown, the 'long' period of cryptic evolution is to be seen only as a measure of time elapsed in relation to the time span required for the diversification of other Late Precambrian Metazoa. Those definitely known include one representative of the phylum Echiura, a number of annelids and a few primitive arthropods. The annelids are

Fig. 1.11. Outline of metazoan phylogeny assuming acoeloid/planuloid origins and polyphyletic origin of the coelom. From Clark (1979, Fig. 9).



represented by at least six genera belonging to at least four very different families. None of them could possibly be considered as primitive by comparison with living members of this phylum. This contrasts sharply with the representation of the arthropods. Only two kinds are known, one representing primitive trilobitomorpha or chelicerates, the other primitive crustaceans. *It must be concluded that the origination of the arthropods was closer to the age of the Ediacarian faunas than that of the annelids which in turn was closer than that of the cnidarians.*

Most studies of the evolution of the living fauna make it likely that the diversification of acoelomate metazoans (mainly Platyhelminthes, Nemertina and Gnathostomulida) preceded that of the coelomates (Fig. 1.11), though Clark (1979, pp. 95–7) leaves the other alternative open. The acoelomates and pseudocoelomates ('Aschelminthes': Nematoda, Nematomorpha, etc.) are entirely or almost totally absent from the fossil record including that of the Precambrian. Omitting the living non-marine and parasitic forms from consideration, we note that these lower free-living Metazoa are small, many of them millimetres in size, and their tissues are extremely soft. These are sufficient reasons for their absence from most kinds of sedimentary rocks. Future lucky finds of some remains of some of these animals are not impossible but it cannot be concluded that they did not exist prior to the appearance of the coelomates because of their absence from the Precambrian record of life. Tubes of worm-like animals (Sabelliditida) may occur in pre-Vendian sediments (see p. 89). The Mollusca are believed to be non-metameric descendants of Platyhelminthes (Salvini-Plawen 1969, Clark 1979). No molluscan body fossils are known from the Precambrian but their locomotion traces could and probably do occur. *The presence of metameric coelomates in the Late Precambrian (Ediacarian) faunas is obvious* (Chapter 2).

2

The Ediacarian faunal assemblages: discovery, composition, significance

2.1 Discoveries at Ediacara

In 1946 R. C. Sprigg, who was then the Assistant Government Geologist of South Australia, examined old lead mines in the Ediacara Hills, a desolate, low range in arid country some 600 km north of Adelaide. The main objective of his work was to review the possibilities of re-opening old silver–lead mines, in the course of a re-assessment of the state's mineral resources. Having acquired the habit of fossil collecting during his student days, Sprigg kept his eyes to the ground even when traversing unpromising quartzite outcrops a short distance southwest of the old mines. To his surprise he found numerous casts and impressions of 'jellyfishes'. He described some of them briefly in the following year, noting correctly that they are 'among the oldest direct records of animal life in the world' and that 'they all appear to lack hard parts and to represent animals of very varied affinities' (Sprigg 1947, p. 212). More fossils were collected soon afterwards by the discoverer and by Professor Sir Douglas Mawson of the University of Adelaide and his students, at the same locality and its northern extensions. They were described in some detail by Sprigg (1949). He stated (p. 73) that there can be no hesitation in placing many of the forms with either the Hydrozoa or Scyphozoa and added: 'Some forms are referable to algae but these will not be described in this paper'. Sprigg's interest in his remarkable finds continued to the present time but his extensive subsequent work in regional geological and mineral exploration and numerous other varied and fruitful activities did not leave him time for further palaeontological research.

What happened some 10 years later was recorded by Hans Mincham (1958), then a school teacher and later a staff member of the South Australian Museum and a well-known author and naturalist. 'It was just after Christmas in 1956 that Mr. Ben Flounders and I set out to

27 August 1982, Volume 217, Number 4562 **SCIENCE**

Cloud, P. & Glaessner, M.F.
"The Ediacaran System: Meta-
zoa inherit the Earth"
Science
Published by A.A.A.S
Used by permission

The Ediacarian Period and System: Metazoa Inherit the Earth

Preston Cloud and Martin F. Glaessner

Ever since Darwin, the geologically abrupt appearance and rapid diversification of early animal life have fascinated biologists and students of Earth history alike. Yet, until recently, convention and lack of data limited serious discussion of planetary evolution to its last eighth, with emphasis on details of metazoan and tracheophyte evolution, sedimentology, and paleogeography, and with but passing reference to earlier Earth history.

Summary. The Ediacarian, here defined as the initial period and system of the Phanerozoic Eon, is characterized by the oldest known multicellular animal life. The distinctive biotal assemblage comprises naked Metazoa, represented in the type region by 26 species in 18 genera and 4 or more phyla, plus simple metazoan surface tracks. Elements of this unique biota appeared worldwide at low paleolatitudes, following terminal Proterozoic glaciation. Ediacarian history lasted from about 670 million to 550 million years ago. This interval, plus Early Cambrian, was the time during which metazoan life diversified into nearly all of the major phyla and most of the invertebrate classes and orders subsequently known.

This constraint loosened with the growing perception that some seven-eighths of geologic time had already elapsed before the Cambrian Period began. Increased knowledge, both of the extent of time and of the older rocks, stimulated reconsideration of the nature of biological processes during those long eons that preceded the Cambrian and accelerated the search for evidence concerning them. Well-preserved sequences of sedimentary rocks, at places extending far beneath the Cambrian without apparent major interruption, showed that there was neither a great historical discontinuity nor a universally high degree of metamorphism beneath the Cambrian, as once supposed. The common presence in these rocks of glauconite and other sedimentary peculiarities indicative of a marine origin bolstered the expectation that, if there were antecedents to the Cambrian biotas, they would be found here.

That prognosis was confirmed with the discovery by Sprigg in 1947 (*1*) of what has come to be called the Ediacara fauna, together with its description by Glaessner and others (*2–5*), and with the recognition that Ediacara fossils are older than and different from those of the conventional Cambrian (*4*, *6*). Further support came with the discovery of similar faunas in ancient rocks of southwestern Africa (*7*) and new finds in the English Midlands (*8*). It is the gist of this article that this Ediacara fauna characterizes a distinctive episode of pre-Cambrian history but of Phanerozoic age and that, among names available, it is most appropriately called the Ediacarian Period. Rocks formed within this period of history then constitute the Ediacarian System, and the appearance of this fauna marks the geological transition from the preceding Proterozoic Eon to the following Phanerozoic Eon, comprising the rest of geologic time.

Semantic impediments have made unambiguous discussion difficult. Until 1930 no term existed for all of Earth history characterized by metazoan faunas and evolution, while pre-Cambrian was the only inclusive term available for the long preceding interval which, heretofore, had yielded no convincing records of visible animal life. Chadwick (*9*) then proposed a solution: the geological record characterized by conspicuous animal life would become the Phanerozoic (Greek for visible plus animal life), whereas antecedent history and rocks would be called Cryptozoic.

Phanerozoic is now widely accepted, but Cryptozoic has had only limited use. By definition, the term Phanerozoic must be extended downward to include older discoveries of manifest animal life. In the opinion of one of us (P.C.) this should take the Paleozoic with it, adding a basal extension to previous additions at the younger end of the original Paleozoic Era. Chadwick's definition, geologic consistency, and etymological congruity all equate the base of the Phanerozoic Eon with that of the Paleozoic Era. Inasmuch as the initial Phanerozoic rocks and history are also pre-Cambrian, still older divisions of rocks and history can no longer be unambiguously designated as simply pre-Cambrian. Clarity demands that the term Cryptozoic be accepted or pre-Phanerozoic employed where the intent is to designate by one word the long sequence of rocks and history that preceded the appearance of metazoan body fossils, imprints, and tracks as conspicuous components of the geologic record of life. The other writer (M.F.G.) considers it unlikely that the great majority of geologists will refrain from using the entrenched formal term Precambrian.

Discussion relevant to the base of the Cambrian (and, by implication, the Paleozoic) became active from the late 1940's through the 1960's (*10–12*). It grew in scale and scope with research and discussion within the framework of

---

The authors are, respectively, Senior Queens Fellow at the Baas Becking Geobiological Laboratory, Canberra City, ACT 2601, Australia, and Professor Emeritus, Department of Geology, University of Adelaide, Adelaide, SA 5001, Australia. Dr. Cloud's current address is Department of Geological Sciences, University of California, Santa Barbara 93106.

0036-8075/82/0827-0783$01.00/0

the International Geological Correlation Program (IGCP), established by the International Union of Geological Sciences and UNESCO in 1971 (*13–16*). As a result, the conventional base of the Cambrian has moved stepwise downward from being defined by the first appearance of the trilobite *Olenellus* to nearly coinciding with the massive incoming of shelly fauna—mainly small primitive mollusks and tubular to conical fossils of uncertain systematic position.

A broad consensus has also emerged concerning the grounds on which useful historical-stratigraphic subdivisions are best made (*11, 14, 17*). The agreed upon criteria are biological, with emphasis on the ranges in time of marine invertebrate animals where dealing with the Phanerozoic. More attention has focused on boundaries than on the modal characteristics or distinctive contents of the divisions bounded, perhaps because the regional variations are so great. Whether for boundaries or for modal characterization, however, it is clear that paleobotanical, paleomicrobiological, and other criteria must also become important as we undertake to define the lower limits of Phanerozoic and Paleozoic rocks and history and to delineate older divisions of the historical and sedimentary record.

Granted that the purpose of nomenclature is to facilitate unambiguous discussion, we must somehow specify the semantic content of terms utilized and seek to eliminate or clarify words that now mean different things to different people.

Ediacarian Period and System

In this spirit, we turn to the matter of how best to treat the historical episode represented by the Ediacara fauna and its equivalents. It turns out that this fauna not only has a circumglobal distribution, but it is clearly antecedent to conventional Cambrian. In fact, the term Ediacarian has already been applied to this part of Earth history—as the Ediacarian Stage of Termier and Termier (*12*) and the Ediacarian Period of Cloud (*16*). In the present article we seek more clearly to define and document the historical episode, to consider its appropriate position in the geologic time scale, and to review alternative nomenclatures. Other names that have included some or all of the rocks and history we here call Ediacarian either have not been clearly defined, have meant something different from the type Ediacarian, or have fluctuated too much in meaning to be useful (*18*).

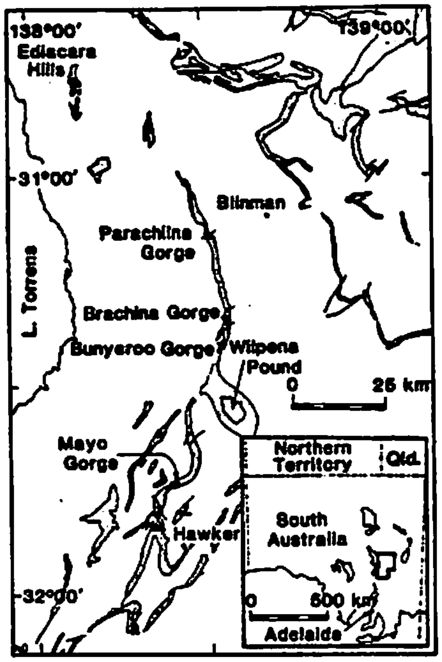


Fig. 1. Index map for key localities in Flinders Ranges. Stippling indicates outcrop of Pound Subgroup. Stars mark Ediacarian fossil localities. [Adapted from Wade (*19*) and Jenkins and Gehling (*48*)]

In brief, the Ediacarian Period and System—named from and characterized by fossil evidence now known from some dozen South Australian localities along almost continuously exposed lines of outcrop 140-kilometers long (Fig. 1), and with equivalent strata worldwide—is not synonymous either with the officially defined Vendian of the Soviet Union or the Vend of general usage. The Ediacarian System is the oldest unit definable by metazoan fossil content. In its stratotype area it follows conformably above tillite-containing late Proterozoic sequences, but it is succeeded disconformably or unconformably by fossiliferous Lower Cambrian deposits.

We then define the Ediacarian Period as that interval of geologic history characterized by the soft-bodied, macroscopic, marine invertebrates of the Ediacara fauna and allied forms. Rocks of the Ediacarian System range from immediately above the last preceding episode of glacial sedimentation in South Australia and, with important exceptions, the last of the fossil stromatolite *Conophyton* beneath equivalents elsewhere, to the massive onset of shelly fauna and distinctive trace fossils (for instance, *Skolithos*, *Rusophycus*, *Plagiogmus*, and *Phycodes pedum*) that marks the initial Cambrian at many places.

In the stratotype area of the Ediacarian System, described below, these rocks are represented by the fossiliferous strata of the Wilpena Group. The distinctive fauna is one of soft-bodied arthropods (*Praecambridium* and *Parvancorina*), segmented worms (*Spriggina* and *Dickinsonia*, Fig. 2, C and D), primitive colonial cnidarian coelenterates (*Charniodiscus*, Fig. 2A; *Glaessnerina*, *Pteridinium*, and *Phyllozoon*), a possible proechinoderm (*Tribrachidium*, Fig. 2E), and some dozen genera of early medusoids (Fig. 2B) and other simple coelenterates. It is widely distributed through an interval that reaches a thickness of 112 meters near the middle of the Pound Subgroup at Mayo Gorge 15 km north of Hawker (*19*) (Fig. 1). The fossiliferous portion of the Pound has been called the Ediacara Member (*20*).

On fossil evidence from the type sequence the base of the Ediacarian is at least as low in the Wilpena Group as the middle of its upper clastic division, the Pound Subgroup. A new discovery of simple metazoan tracks in the Elkera Formation of north central Australia implies a still lower position—at or beneath the base of Bunyeroo Formation (*21*). Combining paleontology with paleoclimatology, operational convenience, and the probability that the metazoan record will be extended downward to some extent, we find it more appropriate, at this time, to equate the base of the Ediacarian with the base of the first stratigraphic unit above the highest tillite-containing deposits beneath known Ediacarian fossils. In South Australia that is the conformable lower surface of the Nuccaleena Dolomite and equivalent rocks at the base of the local Wilpena Group, overlying the tillitic Elatina Formation of the Umberatana Group. From data now available, this could be of roughly the same age as the terminal surfaces of other ice-related deposits that occur shortly below faunas of Ediacarian aspect in other parts of the world (*22, 23*).

Biological support for placing the boundary so low is the structure called *Bunyerichnus* (Fig. 2F), found near the middle of the Brachina Formation some 1800 m below the main Ediacarian fossil zone (*3*). Although we agree with Jenkins (*24*) that this surface marking is not a true crawling track, we do not agree that it is "unlikely to be of metazoan origin" (*25*).

Thus the basal Ediacarian approximately coincides with the oldest record of manifest animal life. For those who may still harbor reservations about that conclusion, it should be noted that most (P.C. thinks all) so far reported pre-Ediacarian Metazoa are pseudofossils, are misdated, or are not Metazoa. Some are even modern organisms or their traces intruded into older rocks (*16, 26*). Among recently proposed examples, the structure called *Rugoinfractus*, suggested as evidence for a metazoan presence 1100 million years ago, appears to be the fillings of shrinkage cracks in underlying strata (*27*), a common form of pseudofossil. Other supposed ancient Metazoa cited by Durham (*27*) from the Huronian Ajibik Quartzite of Michigan and the early Proterozoic Medicine Peak Quartzite of Wyoming are also of inorganic origin, as at least one of their authors has now recognized (*27a*).

Concerning prior usage, Termier and Termier (*12*) proposed "L'Ediacarien, premier étage paléontologique . . . caracterise par ces fossiles et qui s'insère au sein du sous-système Eocambrien." Despite their assignment of Ediacarian tc Eocambrian and some erroneous correlations (for instance, with the much older Belt Supergroup of Montana), there is no doubt that the Termiers intended the Ediacarian to comprise the initial Phanerozoic strata, characterized paleontologically. Here, moreover, although we follow Cloud (*16*) in elevating this term from stage to period rank, we also comply with standard Phanerozoic nomenclature in adopting Ediacarian for both geohistory (period) and rocks (system). Although biostratigraphically defined, it is not a biostratigraphic unit (a zone) in terms of the current International Stratigraphic Guide (*28*). In those terms it is a chronostratigraphic (system) and geochronologic (period) unit.

Division of the Ediacarian into formal epochs of history and series of rocks is not proposed here, although it may be suggested by faunal variation. Cribricyathids of Early Cambrian aspect (*Cloudina*) occur with Ediacarian forms through much of the lower Nama Group in the southwest Africa, whereas typical Early Cambrian trace fossils (such as *Phycodes pedum*) mark the overlying Nomtsas Formation and the Fish River Subgroup (*29*). By contrast, strata of the Wilpena Group have so far revealed neither shelly fauna nor such advanced trace fossils. This, with its unconformable base, suggests (to P.C.) that the Kuibis fauna of the lower Nama Group may be younger than the Ediacara fauna proper and that, if this should be substantiated, subdivision of Ediacarian on faunal grounds into epochs and series may become possible. Until then, the divisions shown as Early (Lower) and Late (Upper) Ediacarian on our global correlation chart are provisional.

Type Site and Global Reference Section

The discovery site for the Ediacara fauna is at the southern end of the Ediacara [also Idracourza, Idyakra, or Etikaura (*29a*)] Hills near Randell Lookout, southwest corner of the 1/250,000 Copley map sheet (~ 30°50'S, 138°8'E). Most Ediacarian fossils so far described from Australia have come from this area, but only a small part of the sequence of interest is exposed there.

The thickest fossiliferous sequence, 112 m, is in Mayo Gorge at the southwest corner of the 1/250,000 Parachilna map sheet, whereas the type sections for the Wilpena Group itself and all its named subdivisions are in the middle third of the Parachilna sheet between Mayo Gorge and Brachina Gorge, along the western flank of the central Flinders Ranges (Fig. 1). The general stratigraphy, paleogeography, and correlation of the rocks of the Wilpena Group and related strata in this and adjacent areas are summarized by Preiss *et al.* (*30*). Descriptions of the individual units appear in papers by others (*19, 31*).

We designate the steeply northwest dipping (55° to 60°) sequence in Bunyeroo Gorge (~ 31°25'S, 138°32'–35'E), 380 km north of Adelaide, as the standard reference section for the Ediacarian System (Fig. 3) (*32*), with the sections in

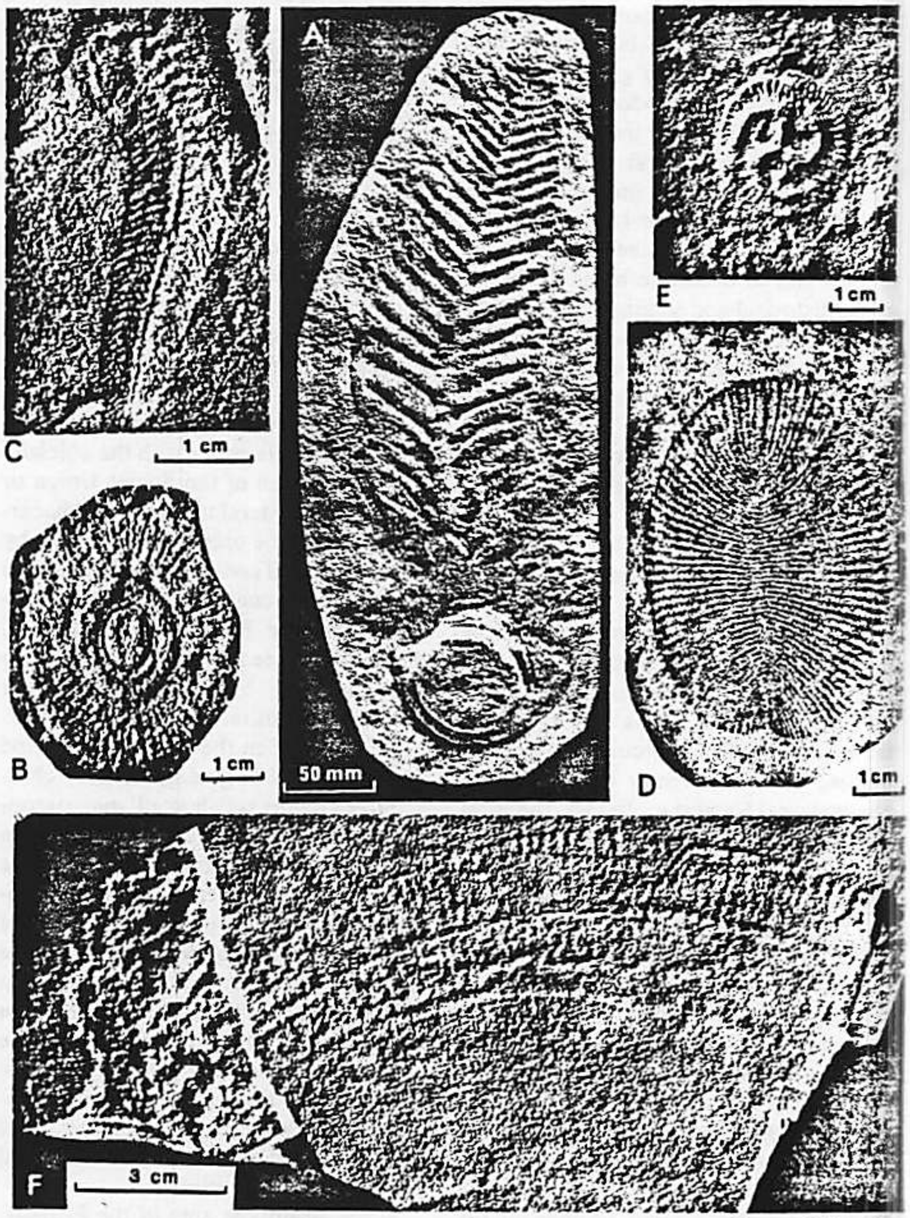


Fig. 2. Some distinctive Ediacarian Metazoa (A) to (E) are from Ediacara Hills. (A) *Charniodiscus arboreus* (Glaessner), plaster mold, photography by R. J. F. Jenkins; (B) *Cyclomedusa radiata* Sprigg; (C) *Spriggina*; (D) *Dickinsonia costata* Sprigg; (E) *Tribrachidium heraldicum* Glaessner. (F) *Bunyerichnus dalgarnoi* Glaessner, from near middle of Brachina Formation in Bunyeroo Gorge; compare with (B).

the Brachina Gorge and Mayo Gorge as supplementary reference sections. We choose Bunyeroo Gorge as the stratotype (*32*) because (i) an uninterrupted sequence of some 3000 m of Ediacarian strata is here compactly exposed in almost continuous outcrop within a horizontal distance of less than 5 km, (ii) the usual fossiliferous interval is well displayed, and (iii) the oldest known record of likely metazoan life (*Bunyerichnus*) occurs in this sequence. The Brachina Gorge section is more accessible by car, and the incomplete Mayo Gorge section adds information on the stratigraphic range of the most distinctive Ediacarian fossils.

Geochronology

The available geochronologic data for rocks of the Ediacarian System have limited usefulness. Some papers in which ages are cited do not state the method employed, the kind of rock or mineral dated, or the decay constant utilized. Where such data are available, the method used was often potassium-argon dating, which is susceptible to error in both directions. Argon loss with time and metamorphism results in minimal ages for materials dated. Other K-Ar numbers may be too old because of absorption of argon by thermally altered pyroxenes or because the mineral dated was glauconite, which may be detrital or formed by alteration of old biotite, giving exaggerated ages for enclosing sediments. Such a number, for example, led to the erroneous report of a Proterozoic age for the Cambrian trace fossil *Skolithos* in the Wessel Group of Northern Australia (*33*). Finally, international agreement on decay constants was only reached in 1976. Before then potassium-argon, rubidium-strontium, and even uranium-lead ages included a computational variance among laboratories of several percent.

Fig. 3. Aerial view of stratotype area. Abbreviations: *€h*, Cambrian, Hawker Group; *€hp*, Parachilna Formation (basal Hawker Group); *E*, Ediacarian System; *Ewpr*, Ediacarian, Wilpena Group, Pound Subgroup, Rawnsley Quartzite; *Ewpb*, Pound Subgroup, Bonney Sandstone; *Eww*, Wonoka Formation; *Ewbu*, Bunyeroo Formation; *Ewa*, ABC Range Quartzite; *Ewb*, Brachina Formation; *Ewn*, Nuccaleena Dolomite; *U*, Umberatana Group (Proterozoic). North is at the top. [Courtesy of the Department of Lands of South Australia]

Cowie and Cribb (*34*) sought to ameliorate such problems for the Cambrian by choosing 40 selected radiometric ages and recalculating them to the new decay constants. These were plotted on a calibrated chart to ascertain average numbers for Cambrian, Ediacarian, and related rock boundaries. The number so obtained for the base of the Cambrian was 560 million years if defined as the top of the Tommotian Stage, or 590 million years if placed at its base, a placing that we favor. We have reservations, however, about the validity of dating statistics that depend so heavily on K-Ar ages of glauconite.

A more credible age for the Ediacarian-Cambrian transition comes from assessment of recent data for the English Midlands: Nuneaton, Warwickshire; Charnwood Forest, Leicestershire; and the Wrekin area of Shropshire.

The Nuneaton ages are on distinctive diorites ("markfieldites") that intrude (and are thus younger than) the Caldecote Volcanics of supposed latest pre-Cambrian age. At Cliffe Hill in Charnwood Forest these "southern diorites" also intrude strata containing an Ediacarian type of medusoid fossil (*35*). This locality provided most of the samples from which a Rb-Sr isochron age of 552 ± 58 million years was obtained for the southern diorites by Cribb (*36*), recalculated to 540 ± 57 million years with the 1976 decay constant. That is close to the 546 ± 22 million years found for similar South Leicestershire diorites some 10 km distant. In the Nuneaton area the Caldecote Volcanics are unconformably overlain by the Hartshill Formation, containing fossils of the Baltic Lower Cambrian *Mobergella* zone 250 m above the volcanics (*37*). These Lower Cambrian fossils are thus younger than the age of 540 million to 546 million years found for the diorites that intrude rocks with Ediacarian fossils.

In the Wrekin area, about 100 km west of Charnwood Forest, Patchett *et al.* (*38*) obtained a good Rb-Sr whole-rock isochron age of 533 ± 13 million years on the Ercall granophyre, unconformably overlain by fossiliferous Lower Cambrian of uppermost Tommotian to Atdabanian age.

The numbers and stratigraphic relations in these three Midlands areas thus reveal an episode of igneous activity around 533 million to 546 million years ago, following deposition of most of the Charnian strata of Ediacarian age. Much but not all of the Lower Cambrian, however, is younger than that. In order to find a realistic age for the base of the Cambrian we must add some years to these numbers to compensate for missing Tommotian sediments. Estimating that missing time as 10 million to 20 million years and averaging, we find 550 million to 560 million years as a likely age for basal Tommotian and hence the Ediacarian-Cambrian boundary.

That number is not far from earlier accepted estimates of around 570 million years. It differs significantly from the estimate of 590 million years given by Cowie and Cribb (*34*), for basal Tommotian, based mainly on K-Ar ages for glauconite from the Soviet Union. These ages, however, are increasingly being questioned by Soviet authors.

Concerning numbers relevant to the base of the Ediacarian, Coates and Preiss (*39*) reviewed Rb-Sr data for the ages of glacial and overlying sedimentary rocks in Western Australia, converting them to the new decay constant $\lambda^{87}\text{Rb} = 1.42 \times 10^{-11}$ year^{-1}. They concluded that three shales of lower to middle Ediacarian age above the uppermost glacials gave reasonably good Rb-Sr whole rock isochrons. Two of the three, correlated with the lower Brachina Formation of the stratotype Ediacarian, give ages of 672 ± 70 million and 670 ± 84 million years, respectively, compared with a less well constrained date of 676 ± 204 million years on the Brachina itself. A younger shale equivalent to the Bunyeroo Formation midway of the stratotype sequence gives an age of 639 ± 47 million years.

Other dates of interest include those related to fossils of Ediacarian aspect in Scandinavia, Newfoundland, and North Carolina. Kulling (*40*) illustrated simple trace fossils and a medusoid (*Kullingia concentrica* Glaessner) from beds correlated with strata above the uppermost Varangerian tillites and below Early Cambrian fossils (*Volborthella*) in northeast Sweden. The age is younger than an adjusted Rb-Sr isochron age of 654 million years reported (*41*) for the Nyborg Formation below the uppermost tillites of nearby Finmark in northern Norway, a number close to those reported for Soviet diamictites of presumed Varangerian age and glacial origin (*42*). Ages around 650 million years are commonly suggested for the Varangerian elsewhere. Thus it is possible either that these mainly K-Ar ages are too young (because of Ar loss), that they date metamorphism rather than sedimentation, or that the north European tillites are in fact younger than the ones that give older Rb-Sr dates in Australia. In any event, the top of the terminal Proterozoic tillites cannot at present be dated more closely than between ~ 670 million and 650 million years.

Table 1. Taxonomic affiliations of metazoan fossils from the Ediacarian of South Australia. Numbers of additional taxa from central Australia are added in parentheses. Percentages refer to specimens collected at Ediacara.

Phylum and class	Genera	Species
Cnidaria (67 percent)		
Hydrozoa, Chondrophorina	3	3
Scyphozoa	4 (+2)	4 (+2)
Conulata	1	1
Others (medusoids)	3 (+1)	6 (+1)
Colonial Cnidaria	4	5
Annelida (25 percent)		
Polychaeta	3	7
Arthropoda (5 percent)	2	2
Phylum uncertain (3 percent)	1	1
	21	29
Trace fossils	6	7
	27	36

In Newfoundland, Hughes and Brückner (*43*) suggested that the intrusive Holyrood Granite, with an adjusted Rb-Sr age of 594 million years (*44*), is also of about the same age as the Harbour Main volcanics and the Conception Group of marine volcanogenic sediments. The latter contains in its upper part the Mistaken Point fauna of Ediacarian affinity (*45*), found well above tillites but below the Random Formation and equivalent or older rocks with Tommotian fossils (*46*). Similarities between the Mistaken Point fauna and the Ediacarian of Charnwood Forest combine with lithological resemblances to imply equivalence and predrift proximity.

The firmest number available for rocks that probably represent part of the Ediacarian System is a U-Pb concordia age of 620 ± 20 million years on zircons from little metamorphosed pyroclastic rocks in conformable stratigraphic sequence with fossiliferous volcanogenic sediments in North Carolina (*47*). No distinctive Ediacarian elements are present at this place, but ages and local succession strongly imply an Ediacarian equivalence, and imprints of soft-bodied wormlike metazoans occur in these sediments.

The numbers given above are believed to be the best so far available for Ediacarian rocks. They imply a range in time from perhaps 670 million years at the base to around 550 million years at the top.

Metazoa

The distinctive fauna of the Pound Quartzite at Ediacara (Fig. 2 and Table 1) has already been described (*1–5, 48*) and tabulated (*49*). It consists entirely of remains and imprints of soft-bodied Metazoa and trace fossils. Some show indications of chitinous or minor spicular strengthening elements such as are found in similar living soft-bodied invertebrates, but none had mineralized shells or solid skeletons. About two-thirds of the specimens known are cnidarian coelenterates. Such a dominance is rare at younger pre-Cenozoic fossiliferous localities and unknown in the Cenozoic. Unique though they be, however, fossils of the Ediacarian assemblage are clearly related to younger fossil and even living forms.

Hydrozoa are represented by the distinctive floats of the Chondrophorina. Among them is *Eoporpita*, belonging to the same family as the living *Porpita*, colonies of differentiated polyps that drift at the surface of the sea. The only other living genus, *Velella*, resembles the Ediacarian genera *Ovatoscutum* and *Chondroplon*, although these have bilaterally symmetrical floats. A number of other genera of similar colonial hydrozoans occur in the Paleozoic, but only two still live. It seems that either competition among planktotrophic Metazoa or predation on them increased with time.

The Scyphozoa are apparently represented by such genera as *Brachina*, *Ediacaria*, *Rugoconites*, and *Kimberella*, considered to be a possible ancestor of the Cubomedusae. The Conulata, an important group of Paleozoic fossils, are represented by *Conomedusites*, strikingly like the Ordovician *Conchopeltis*. A number of medusoid fossils, particularly the common *Cyclomedusa* and *Medusinites*, are not sufficiently distinctive to be placed in classes and orders of the Cnidaria. There is, however, little doubt that most of them lived like common modern jellyfish.

Sessile colonial Cnidaria have left

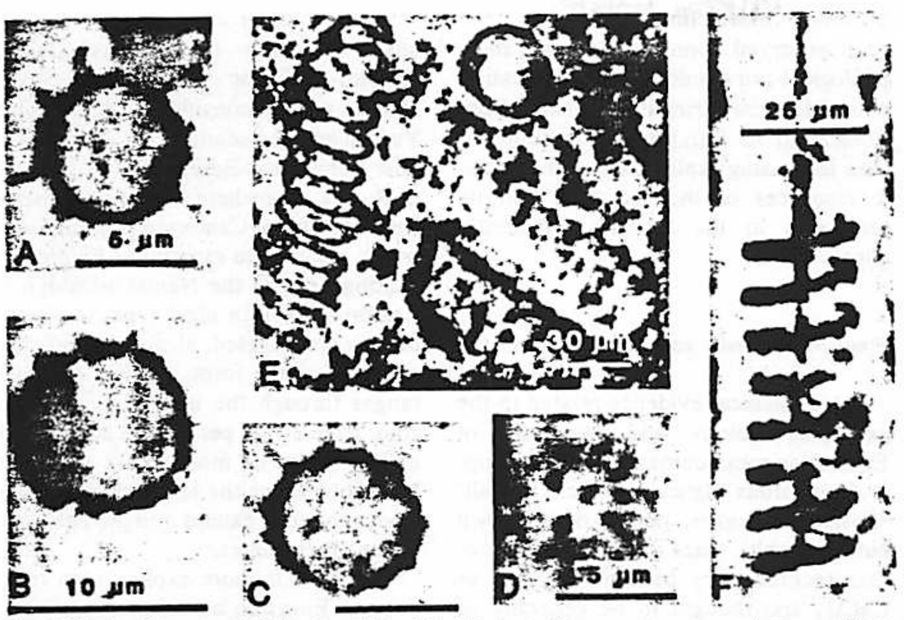


Fig. 4. Distinctive microbial fossils. (A) to (D) *Micrhystridium* sp. from lower Yudoma Group (Ediacarian), right bank Aldan River, upstream from Belaya (or Xanda) River, East Siberia. (E) and (F) *Obruchevella parva* Reitlinger from Muraykhah Formation, Jubaylah Group (Ediacarian or Cambrian), Jabal Umm al'Aisah, northeastern Saudi Arabia (*59*).

strikingly leaf-like fossils, up to 1 m long. Some of them (*Charniodiscus*, Fig. 2A) show distinctive characters of the living Pennatulacea (Anthozoa, Octocorallia), while others (*Pteridinium*, *Phyllozoon*), although pennatulacean in some respects, are more difficult to interpret. New specimens of *Rangea* (*50*) have shown it to be an endemic Namibian genus to which the first finds of this group of fossils from Ediacara were mistakenly assigned. The term "Petalonamae," proposed by Pflug (*50*) for these and other Namibian leaf- or cup-shaped fossils and for similar Ediacarian genera, has proved confusing and unnecessary. The significance of these fossils is the wide geographic distribution and the great age of ancient colonial cnidarians, some of which closely resemble living seapens (for instance, *Pennatula*), organisms that are rare or missing in most of the Phanerozoic fossil record.

About 25 percent of the specimens collected at Ediacara are annelids. The most common genus, *Dickinsonia*, may have survived into Paleozoic time (*51*). A similar form, *Spinther*, is still living as an ectoparasite on sponges. *Spriggina* is a very different kind of free-living, probably nectobenthic, polychaete worm. These fossils are still under study in connection with problems of annelid-arthropod relations. Two primitive arthropods, *Praecambridium* and *Parvancorina*, are rare. Together with the enigmatic triradiate *Tribrachidium* (resembling edrioasteroid echinoderms without calcareous plates) they amount to but a few percent of the fauna (Table 1).

The status of the Ediacarian as a valid Phanerozoic chronostratigraphic unit of system rank in the standard global stratigraphic scale is reflected in this fauna and reinforced by its position in the geochronologic scale. It is characterized by the global distribution of distinctive metazoan assemblages resembling those found at Ediacara and differing from those of the succeeding Cambrian System. The recognition of their potentiality for long-range correlation began with the discovery of Charniidae and medusoids comparable with those from Ediacara in the Charnian of the English Midlands (*2–5, 8, 48, 49*) and related forms in southwestern Africa (*7, 29*). Charnian types of fossils were later found, together with others, in the Mistaken Point fauna of Newfoundland (*45*) and in Siberia (*52*). A remarkable display of this fauna is that in the Valdai sediments of the White Sea Coast of the Soviet Union (*53*). It includes about ten species and ten genera (not all represented by previously known species) also found at Ediacara. A similar assemblage occurs on the southwestern margin of the East European platform. The Soviet faunas include numerous *Dickinsonia costata*, *Tribrachidium*, and the primitive arthropod *Vendia*, similar to *Praecambridium* from Ediacara. These assemblages, irrespective of sedimentary facies, share the numerical dominance of the phylum Cnidaria, reflecting evolutionary level at Ediacarian time. From some regions (central Australia, China, Scandinavia, and North Carolina), however, finds reported are limited or unique.

The wide distribution of Ediacarian faunas and even species supports paleogeographic reconstructions that place fossiliferous localities of the Proterozoic-Phanerozoic transition interval in low latitudes (*54*). Equatorial currents favored by such reconstructions would have facilitated the observed spread of marine faunas.

The Proterozoic to Early Cambrian sedimentary succession has been intensively studied in recent years in search of data on which a stratigraphic definition of the base of the Cambrian might be based. It is clear from this work (*15*) that the Ediacarian fauna of soft-bodied animals disappeared from or became very scarce in the preserved record after Ediacarian time. In its place appeared a fauna of small shelly fossils, many difficult to place in the zoological system of classification (*5*). They are not descended from members of the Ediacarian assemblage. Their ancestors may have been too small to be fossilized or may not have developed mineralized tissues. That this post-Ediacarian fauna with abundant and diverse small shelly fossils should be considered Cambrian, despite its lack of trilobites, is supported by the occurrence in Siberia of small, primitive archaeocyathids at the base of the Tommotian Stage. They soon developed complex calcareous skeletons which built reef-like structures in post-Ediacarian time. Their evolution through the Early Cambrian to their extinction in early Middle Cambrian time illustrates the successes of minor groups during the initial Phanerozoic adaptive radiation of the emerging Metazoa.

Like other transitions in the geologic record, however, the replacement of Ediacarian soft-bodied by younger skeletal Cambrian faunas was neither complete nor geologically instantaneous. The evolution of chitinous cuticles and of tubes formed from discrete grains cemented with mucus started during Ediacarian times, and we note reports of Sabelliditidae with organic tubes, possibly representing Pogonophora, of putative Riphean to Early Cambrian age. The characteristically Lower Cambrian cribricyathids are represented by the genus *Cloudina* in calcareous strata of the Ediacarian Nama Group of Namibia, mentioned above as implying a young Ediacarian age.

The poor showing in the Phanerozoic fossil record of soft-bodied Metazoa, abundant in the present marine fauna, is probably due to preservational or collecting bias, the evolution of macrophagous predators, an increase in saprophagous macro- and microbiota, or some combination of such factors. Biostratigraphic studies have shown that the "sudden" appearance of abundant small shells of the Tommotian Stage (basal Cambrian) was preceded in the latest Ediacarian (latest Yudomian of Siberia and probably the Sinian of the Yangtze Platform) by the occurrence of three or four genera of minute tubular and spicular fossils (*Anabarites trisulcatus*, probably a worm tube; *Protohertzina*, similar to chaetognath grasping spicules; and *Cambrotubulus* and/or *Hyolithellus*). The seemingly abrupt appearance of small shelly fossils in many places is enhanced by the common transgression of Early Cambrian on older sediments, leaving gaps in the record. Ecological effects of this transgression are well displayed in England (*55*). The onset of biomineralization was part of the rapid early Phanerozoic diversification of the Metazoa. No universally causal extrinsic environmental factor has been credibly invoked for it.

Metazoan diversification during the Ediacarian-Cambrian transition was paralleled by an increase in diversity of trace fossils (*3, 53, 56*). These are indicators of mainly metazoan life activities on and in sediments, made by live animals at the places where their markings are found. They may reveal much about the nature of these activities (feeding, locomotion, and so on), but only exceptionally do they indicate the place of the originator in the zoological classification. Different kinds of traces can be made by different activities of the same animal and similar traces by different animals. Thus the number of named form genera does not indicate taxonomic diversity but ethological differentiation.

Known Ediacarian trace fossils, probably of detritus feeders, are simple shallow burrows and relatively poorly oriented search trails on bedding surfaces. By contrast, early Cambrian assemblages include deep, vertical burrows (*Skolithos*), complex burrows (*Phycodes pedum*, *Diplocraterion*, *Chondrites*, *Treptichnus*, and *Plagiogmus*), excavations made by trilobite-like arthropodan appendages (*Rusophycus*), and a variety of trackways. Fedonkin (*53*) recognized Ediacarian trace fossils as dominantly two-dimensional and horizontal, in contrast to three-dimensional Early Cambrian forms. Some relatively complex locomotion trails that occur only rarely in the Ediacarian become common in the Cambrian, for instance, *Didymaulichnus*, which suggests the actions of a mollusk-like gliding foot rather than the peristaltic movement of a wormlike animal. Much more analytical work remains to be done in paleoichnology, but what has been observed confirms the rapid morphological and ethological diversification of the Metazoa during the transition from Ediacarian to Cambrian, a product of their increasing exploitation of the trophic resources of the sea floor and its sediments in the littoral and neritic zones.

Plant Microfossils and Stromatolites

Paleobotanical evidence related to the age, paleoecology, and correlation of Ediacarian rocks comes from microscopic filamentous algae, acritarchs (small, spheroidal, organic bodies of unknown but probably plant origin), stromatolites (accretionary buildups, mainly of $CaCO_3$ and thought to be generally of blue-green algal origin), and perhaps some ambiguous structures called microphytolites.

The microbial flora of the pre-Phanerozoic and early Phanerozoic sediments is only beginning to receive the attention it warrants. Available evidence, however, implies that such a flora has existed from Archean times to the present and that forms preserved increased markedly in size, abundance, and diversity during later Proterozoic and early Phanerozoic time. Although such material has not yet been systematically studied in the Ediacarian and immediately underlying strata of Australia, it is known to be biostratigraphically useful in correlative beds elsewhere and will become more so as our now limited knowledge of it increases (*57*).

The ranges of four distinctive, stratigraphically limited, organic-walled, microbial forms known to us are shown on the left of the correlation chart. Spiny planktonic acritarchs of the sort called *Micrhystridium* (*58*) and the spiral filaments of the microalga *Obruchevella* (*59*), illustrated in Fig. 4, apparently first appear in strata of Ediacarian age and range upward through the Cambrian. The spheroidal, multicomponent form genus *Bavlinella*, representing the endosporangia or clonal colonies of *Sphaerocongregus* (*57, 60*), ranges from upper Proterozoic through Ediacarian equivalents into the Cambrian of the Northern Hemisphere. Vidal (*57*) describes it as a characteristically Vendian form, but records it in the Cambrian and in and beneath the Varangerian tillites, as well as in other truly upper Proterozoic rocks such as those of the upper Sinian "Suberathem" of northern China. Bushlike colonies of the bifid branching microalga *Epiphyton* (commonly commensal on archaeocyathids or a bioherm builder) are unknown below the Nemakit-Daldyn and lower Baltic beds of the Soviet Union, widely considered to be upper Yudomian (Ediacarian) in Siberia but post-Vend (post-Ediacarian) in European Russia. Elsewhere we have not seen it below the basal Cambrian (Tommotian), which leads us to regard the *Epiphyton*-bearing beds of the Nemakit-Daldyn as Cambrian. A fifth algal type, the large, usually compressed, algal spheroids that comprise the form genus *Chuaria*, ranges through the uppermost Proterozoic, from strata perhaps as old as 1200 million years or more to its very top. Here it overlaps the lower range of *Bavlinella* and may extend into the Paleozoic under other names.

Vidal (*57*) is more explicit with reference to Eurasian acritarchs. Among 22 previously known acritarch taxa he studied, he finds that 18 percent first appear in the Vend, 10 percent are limited to the Vend, and 6 percent are limited to the upper Riphean (uppermost Proterozoic). We are not sufficiently confident about either the identification of acritarch species or the ages of the stratigraphic units discussed (for instance, the upper Sinian Suberathem of northern China) to accept these estimates as a basis for the definition of systemic boundaries. We are hopeful, however, that continued acritarch research will eventually lead to a better Proterozoic biostratigraphy and contribute to a more cogent definition than we can presently give for the base of the Ediacarian.

Stromatolites undergo a decline in abundance and diversity with the transition from Proterozoic to Phanerozoic, and with the appearance of metazoan browsers, eukaryotic precipitation and secretion of $CaCO_3$, and competitive exclusion. They are, however, still relatively abundant during Ediacarian time. Although few are uniquely Ediacarian, an association of certain distinctive columnar forms with strongly enveloping laminae or lateral wall-like structures would suggest a late Proterozoic to Ediacarian age (*61*). The simple peripherally enveloped columns of *Boxonia grumulosa*, combined with branching columnar "walled" forms like *Gymnosolen* and the lumpy *Linella ukka* and *Paniscollenia*, would be such an association. Where found above a sequence containing evidence of extensive late Proterozoic glaciation an Ediacarian age would seem likely.

Finally, in the Soviet Union and to some extent in China, subdivision and correlation of late Proterozoic and early Phanerozoic strata has, in some in-

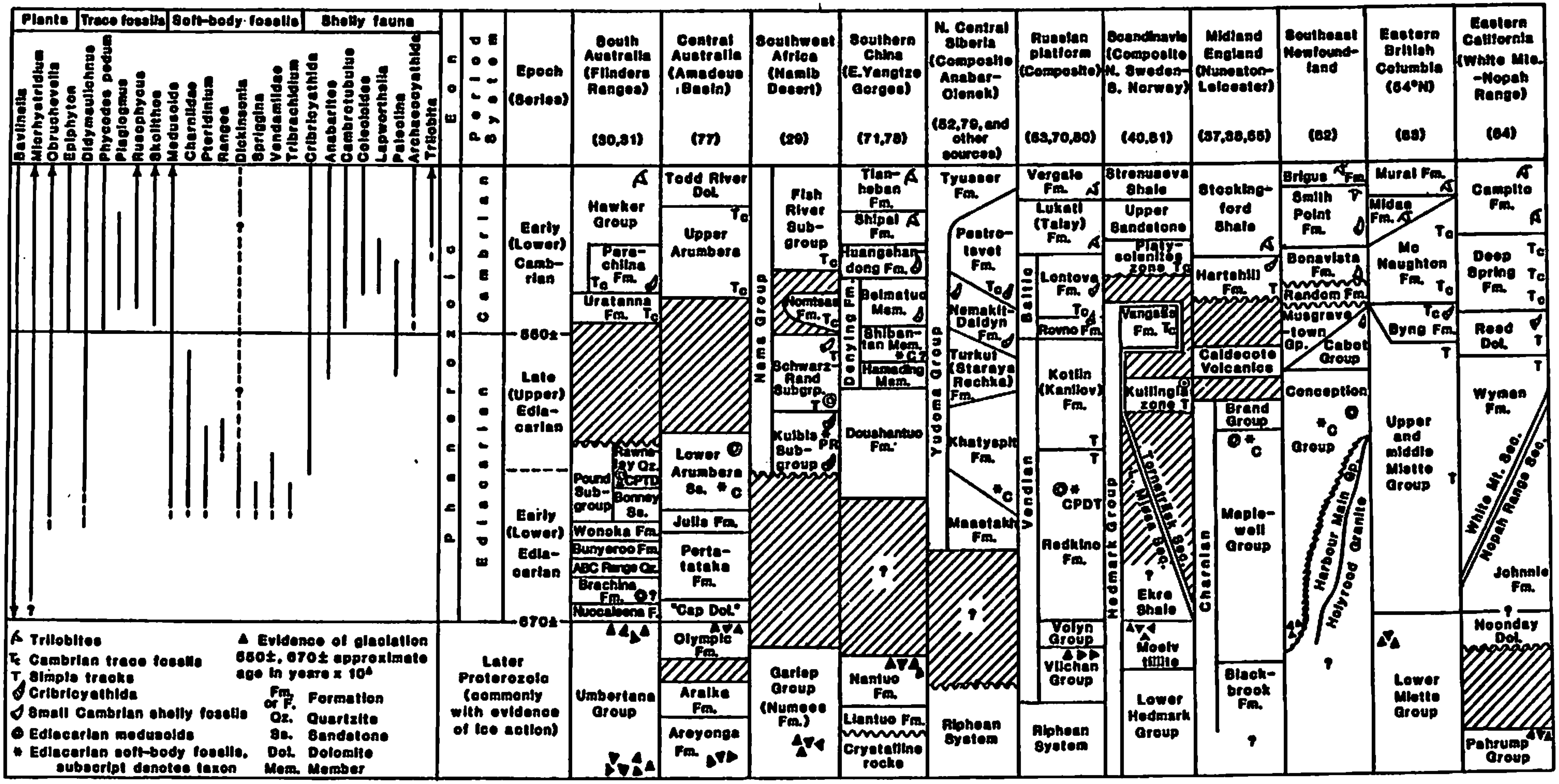


Fig. 5. Provisional global correlation of Ediacarian System. (Not to scale; omits many details and some whole regions.)

stances, been based on a heterogeneous group of microstructures referred to as microphytolites. Although we have doubts about the biological nature of most of these structures and the validity of subdivision and correlation based on them, several named entities have been reported from Ediacarian and adjacent strata in South Australia, and it is claimed that partial success in correlation was attained in tests based on samples of provenance unknown to the identifier (*62*).

Global Distribution

Beyond South Australia, Ediacarian fossils are known from the Amadeus Basin of central Australia and from some dozen other regions on four other continents (Africa, Asia, Europe, and North America). Figure 5 summarizes the main features of nine of these fossiliferous sequences plus two western North American sequences that are expected eventually to yield Ediacarian fossils. Space, regrettably, does not permit further discussion of these areas in this article. Their stratigraphy, biotas, and paleoecology, however, are detailed in references cited at the top of each regional column.

It should be stressed that Fig. 5 shows only relative positions in geologic sequence and does not conform to any scale of thickness or time, other than as ages are indicated for the top and base of the Ediacarian. The underlying upper Proterozoic is simply whatever occurs beneath the Ediacarian, with no special reference to age other than the supposed approximate time equivalence of the probably glacial rocks. The overlying Lower Cambrian here includes only the Tommotian and Atdabanian Stages of earliest Cambrian age.

In addition to sequences shown in Fig. 5, evidence available is *consistent* with an Ediacarian presence in North Carolina in the United States (*47*); the upper Brioverian of Brittany, Normandy, and the Channel Islands (*63*); Antarctica (*64*); South America (*65*); and the northern Arabian Shield (*59*). Also possible is an Ediacarian presence somewhere beneath or in the basal part of the Cándana Quartzite of northern Spain (*66*). In India, Svalbard, and Alaska also, well-preserved sedimentary rocks of the right age range and facies to yield an Ediacarian biota may be present. Indeed, there is reason to believe that future discovery will show the Ediacarian to have an areal extent worthy of its rich historical and evolutionary interest (*85*).

References and Notes

1. R. C. Sprigg, *Trans. R. Soc. South Aust.* 71, 212 (1947).
2. M. F. Glaessner, *ibid.* 81, 185 (1958); *Nature (London)* 183, 1472 (1959); *Sci. Am.* 204 (March), 72 (1961); *Biol. Rev. Cambridge Philos. Soc.* 37, 467 (1962).
3. ——, *Lethaia* 2, 369 (1969).
4. ——, *Geol. Rundsch.* 47, 522 (1958).
5. —— and B. Daily, *South Aust. Mus. Rec.* 13, 369 (1959); M. F. Glaessner and M. Wade, *Palaeontology* 9, 599 (1966).
6. L. C. Noakes, *20th Int. Geol. Congr. El Sistema Cambrico* (1956), part 2, pp. 213–238.
7. R. Richter, *Senckenbergiana Lethaea* 36, 243 (1955).
8. T. D. Ford, *Proc. Yorks. Geol. Soc.* 31, 211 (1958).
9. G. H. Chadwick, *Geol. Soc. Am. Bull.* 41, 47 (1930).
10. H. E. Wheeler, *J. Geol.* 55, 153 (1947); C. J. Stubblefield, *20th Int. Geol. Congr. El Sistema Cambrico* (1956), part 1, pp. 1–44; V. V. Missarzhevsky and A. Yu. Rozanov, *Izv. Akad. Nauk SSSR Ser. Geol.* 2, 60 (1963); P. Cloud, *Evolution* 2, 322 (1948).
11. M. F. Glaessner, *J. Geol. Soc. Aust.* 10, 223 (1963).
12. H. Termier and G. Termier, *Rev. Gen. Sci. Pures Appl. Bull. Assoc. Fr. Av. Sci.* 67, 79 (1960).
13. B. S. Sokolov, *23d Int. Geol. Congr. Contrib. Sov. Geol. Probl.* 9 (1968), pp. 5–15; *Geol. Geofiz.* 12 (No. 6), 13 (1971); *24th Int. Geol. Congr. Sec. 1* (1972), pp. 78–84; *23d Int. Geol. Congr. Rec. Int. Paleontol. Union* (1972), pp. 79–84; *Paleontol. Zh.* 1, 3 (1976).
14. B. S. Sokolov, *Geol. Geofiz.* 15 (No. 2), 3 (1974).
15. V. E. Savitsky, *Geol. Mag.* 115, 127 (1978); A. Yu. Rozanov, *ibid.* 104, 415 (1967); A. Yu. Rozanov *et al.*, *The Tommotian Stage and the Problem of the Cambrian Lower Boundary*, (Doklady Geologiya Instituta Akademii Nauk SSSR 206, Moscow, 1969); A. Martinsson, *Lethaia* 13, 46 (1980); M. F. Glaessner, *Bull. Geol. Soc. Am.* 82, 509 (1971); J. W. Cowie and M. F. Glaessner, *Earth Sci. Rev.* 11, 209 (1975); J. W. Cowie and A. Yu. Rozanov, *Geol. Mag.* 111, 237 (1974).
16. P. Cloud, in *Evolution and Environment*, E. T. Drake, Ed. (Yale Univ. Press, New Haven, Conn., 1968), p. 37; *Paleobiology* 2, 351 (1976).
17. B. S. Sokolov, *Geol. Geofiz.* 15 (No. 4), 18 (1974); *Paleontol. Zh.* 1, 3 (1976).
18. The terms "Eocambrian" (*67, 68*) and "Infracambrian" (*68, 69*) are ambiguous. They are not based on biological criteria, no type sequences have been designated, and they have been used loosely to include widely varying parts of the Precambrian sequence and locally, though unintentionally, rocks now known to be Cambrian (for instance, the Ringsaker Quartzite of Norway, upper member of the Vangsås Formation). They have been generally abandoned. Sinian, although once officially adopted in the Soviet Union for rocks of Vendian and Riphean age (*70*), suffers similar disadvantages. It is still widely used in China, but in two very different senses. The thin Sinian System of the Yangtze Gorges and neighboring parts of southern China, younger than about 680 to 720 million years (*71*) and with glacial sediments at its base, is apparently not even represented in the standard northern China reference section for the Sinian Suberathem. Only if the glacial sediments were excluded from its base and the local basal Cambrian (Tommotian) from its top would it approximately coincide in time with our Ediacarian System.

    The term most deserving consideration and comparison with the Ediacarian is Vend or Vendian. This name was proposed by Sokolov (*72*), who later (*73*) referred to the Ediacarian as "the Australian equivalent of the Vendian." This equivalence is less straightforward than it may appear. A new stratigraphic scale for the Proterozoic of the Soviet Union was accepted by a large number of specialists at a conference in Ufa in 1977 (*74*). This makes the Vend the uppermost division of the Proterozoic, including (from base upward) the Vilchan, Volyn, and Valdai Groups. Its stratotype is in the western part of the Moscow basin, where it was studied from numerous bore holes. It is inaccessible to direct observation.

    The Vend was originally equated with the Valdai only (*72*). It was later extended downward to include the Volyn, partly volcanogenic (*70*), and then the Vilchan, partly glaciogenic (together comprising the Drevlyany "series" of some). The Ediacarian fauna of soft-bodied Metazoa is found only in the Valdai, and, for this and other reasons, Keller (*74*), Yakobson and Krylov (*75*), and other Soviet geologists have deplored the acceptance of the "greater Vendian." If, as many believe, the late Proterozoic glaciogenic deposits of the East European Platform and the margins of the Baltic Shield are of approximately the same age as those of Australia, only the Valdai (including Redkino and Kotlin strata) would correspond to the Ediacarian as here defined, not the official Vendian *sensu lato* which Sokolov accepts at least since 1973 (*52*). Vendian has also been used in this broad sense by Harland and Herod (*76*) and by Cowie and Cribb (*34*), among others, to include both fossiliferous rocks of Ediacarian age and the Varangerian tillites beneath.
19. M. Wade, *Trans. R. Soc. South Aust.* 94, 87 (1970).
20. R. J. F. Jenkins, *Geol. Soc. Aust. Abstr. 1st Geol. Conv.* (1975), p. 21.
21. M. R. Walter, personal communication to P.C., who examined the fossils and agrees to their metazoan origin.
22. W. B. Harland and I. N. Herod, *Geol. J. Spec. Issue* 6, 189 (1975).
23. P. R. Dunn *et al.*, *Nature (London)* 231, 498 (1971).
24. R. J. F. Jenkins, *Geol. Soc. Aust. Abstr. 1st Aust. Geol. Conv.* (1975), p. 19; R. J. F. Jenkins, P. S. Plummer, K. C. Moriarty, *Trans. R. Soc. South Aust.* 105, 67 (1981).
25. The enigmatic structure called *Bunyerichnus* is so far known only from the fragment shown in Fig. 2F, its counterpart, and one smaller fragment. The view illustrated is of an upper bedding surface. The arcuate grooves and ridges of this specimen, from the scalloped outer margin to the faint groove at the lower edge (*x* in Fig. 2F) deviate only slightly from a set of concentric circles that would close on themselves, if continued, with a maximum diameter of 24 cm. They clearly represent either (i) the tracing of a longish (12 cm) stalklike, flexible structure of some complexity that swung by currents about a point of attachment or (ii) the imprint of a large (24 cm in diameter) discoidal structure, having enough plasticity to account for asymmetries observed. These characteristics imply that it was the product of a once living organism, very likely a metazoan. Although such a marking might be produced by the swinging about a central point of an elongate wormlike form [for instance, *Vermiforma* Cloud (in *47*)] or a pennatulacean (*53*), the positive epi-relief displayed is more consistent with it being an imprint than a drag mark; for example, compare Fig. 2F with Fig. 2B. The counterpart imprint also shows features resembling gonadal imprints of some recent medusae. We conclude that *Bunyerichnus* is certainly biologic, very probably metazoan, and most likely medusoid.
26. P. Cloud, L. B. Gustafson, J. A. L. Watson, *Science* 210, 1013 (1980).
27. J. W. Durham, *Annu. Rev. Earth Planet. Sci.* 6, 21 (1978); V. M. Palij, *Dopov. Akad. Nauk Ukr. RSR Ser. B* 36 (No. 1), 34 (1974).

27a. H. Faul, personal communication to P.C.

28. H. D. Hedberg, Ed., *International Stratigraphic Guide* (Wiley, New York, 1976), pp. 66–93.
29. G. J. B. Germs, *Bull. Precamb. Res. Unit Univ. Capetown* 12, 1 (1972); *Am. J. Sci.* 272, 752 (1972); A. Kroner, M. D. Williams, G. J. B. Germs, A. B. Reid, K. E. L. Schalk, *ibid.* 280, 942 (1980).

29a. R. Lacour-Gayet, *A Concise History of Australia*, J. Grieve, Transl. (Penguin, Blackburn, Victoria, Australia, 1976), p. 3.

30. W. V. Preiss, R. W. R. Rutland, B. Murrell, in *Precambrian of the Southern Hemisphere*, D. R. Hunter, Ed. (Elsevier, Amsterdam, 1981), pp. 327–354.
31. R. Goldring and C. N. Curnow, *J. Geol. Soc. Aust.* 14, 195 (1967); B. G. Forbes, *Trans. R. Soc. South Aust.* 95, 219 (1971); P. S. Plummer, *ibid.* 102, 25 (1978).
32. Briefly, the divisions of the Wilpena Group, comprising the type Ediacarian System in Bunyeroo Gorge, are (from bottom upward) as follows:

    1) Nuccaleena Formation; basal Wilpena Group, conformable on the underlying locally tillitic rocks of the Umberatana Group. Pinkish, laminated, flaggy dolomite going upward to purple shale with dolomite lenses. These were the initial shallow marine sediments of the postglacial warming and marine innundation that introduced the Ediacarian System. Thickness, 10 m.

    2) Brachina Formation; reddish-brown to olive-green, thin-bedded, micaceous siltites with shale and fine-grained sandstone. *Bunyerichnus* near the middle. Current marks, flute marks, local coarsening and grading, and ripple marks imply submarine conditions of varying depth.

with flat-topped ripples indicating local subaerial exposure. Thickness, ~ 1200 m.

3) ABC Range Quartzite: light colored, flaggy to massive, cross-bedded and ripple-marked, ridge-forming, locally feldspathic sandstone and quartzite. Flat-topped ripple marks and desiccation polygons indicate episodes of subaerial exposure within a shoaling marine environment. Thickness in Bunyeroo Gorge is 80 to 120 m, thickening westward.

4) Bunyeroo Formation: monotonous, reddish-brown and green, silty shales with thin cupriferous dolomite bands, local carbonaceous shale, and dolomitic sandstones showing local ball-and-pillow structure and sole markings indicative of a western source and gravity mass transport. Basally equivalent Elkera Formation of central Australia contains metazoan trace fossils (*21*). Thickness in Bunyeroo Gorge is ~ 400 m (reaching 700 m elsewhere).

5) Wonoka Formation; greenish-gray calcareous siltstones and fine-grained channel sandstones in the lower part, going upward to mainly cross-bedded, silty, commonly sole-marked, detrital limestones, locally stromatolitic. Thickness is 460 m in Bunyeroo Gorge, but the formation thickens dramatically, changes facies, and cuts downward across underlying sequences in a series of submarine canyon deposits dispersed south to north across the center of the Copley map sheet, next north of the Parachilna sheet [C. C. von der Borch, R. Smit, A. E. Grady, *Flinders Univ. Inst. Aust. Geodyn. 81/3* (1981), pp. 1–44].

6) Pound Subgroup; massive ridge- and bluff-forming sandstone and quartzite, including the lower Bonney Sandstone and upper Rawnsley Quartzite:

6A) Bonney Sandstone; mostly arkosic and micaceous lenticular siltstones showing ripple marks, desiccation polygons, local ball-and-pillow structure, and cross-bedding, including trough cross-beds indicative of fluviatile processes. A periodically emerged coastal environment is indicated. Thickness, ~ 300 m.

6B) Rawnsley Quartzite; whitish quartzite and sandstone with local siltite layers, locally ripple-marked and cross-bedded, gradational from Bonney Sandstone beneath. Contains typical Ediacarian body fossils and imprints [illegible] a massive, basal, bluff-forming unit. Represents final overfilling of Adelaide Geosyncline prior to late Ediacarian and Cambrian truncation and onlap of initial Cambrian seas. Thickness, ~ 500 m. Unconformably overlain by argillaceous sandstones of Parachilna Formation, Hawker Group (lower Cambrian), with vertical trace fossils *Skolithos* and *Diplocraterion*.

Total thickness, ~ 2970 m.

The description above applies to the stratotype section and nearby outcrops. The sequence is broadly similar over other parts of the Flinders Ranges but with the variations in thickness and facies to be expected in an evolving mobile shelf environment (*30*).

33. K. A. Plumb, J. H. Shergold, M. Z. Stefanski, *B.M.R. J. Aust. Geol. Geophys.* 1, 51 (1976).
34. J. W. Cowie and S. J. Cribb, in *The Geologic Time Scale*, G. V. Cohee, M. F. Glaessner, H. D. Hedberg, Eds. (American Association of Petroleum Geologists, Tulsa, Okla., 1978), pp. 355–362.
35. L. Rose, personal communication to M.F.G.
36. S. J. Cribb, *J. Geol. Soc. London* 131, 203 (1975).
37. M. D. Brasier and R. A. Hewitt, *U.S. Geol. Surv. Open-File Rep. 81-743* (1981), p. 29.
38. P. J. Patchett, N. H. Gale, R. Goodwin, M. J. Humm, *J. Geol. Soc. London* 137, 649 (1980).
39. R. P. Coates and W. V. Preiss, *Precamb. Res.* 13, 181 (1980).
40. O. Kulling, *Sver. Geol. Unders. Ser. Ba* 19, 28 (1964).
41. J. R. Pringle, *Geol. Mag.* 109, 465 (1973).
42. N. M. Chumakov, *Dokl. Akad. Nauk SSSR* 198 (No. 2), 419 (1971).
43. C. J. Hughes and W. D. Brückner, *Can. J. Earth Sci.* 8, 899 (1971).
44. H. W. Fairbain, S. Moorbath, A. O. Ramo, W. H. Pinson, P. M. Hurley, *Earth Planet. Sci. Lett.* 2, 321 (1967).
45. S. B. Misra, *Bull. Geol. Soc. Am.* 80, 2133 (1969).
46. S. Bengston and T. P. Fletcher, *U.S. Geol. Surv. Open-File Rep. 81-743* (1981), p. 18; B. Greene and H. Williams, *Can. J. Earth Sci.* 11, 319 (1973).
47. P. Cloud, J. Wright, L. Glover III, *Am. Sci.* 64, 396 (1976).
48. R. J. F. Jenkins and J. G. Gehling, *Rec. South Aust. Mus.* 17 (No. 23), 347 (1978).
49. M. F. Glaessner, in *Treatise on Invertebrate Paleontology* (Geological Society of America, Boulder, Colo., 1979), part A, p. 79; ______ and M. R. Walter, in *Precambrian of the Southern Hemisphere*, D. R. Hunter, Ed. (Elsevier, Amsterdam, 1981), pp. 361–396.
50. H. D. Pflug, *Palaeontogr. Abt. A.* 134 (Nos. 4–6), 226 (1970); G. J. B. Germs, *Lethaia* 6, 1 (1973).
51. L. I. Borovikov, *Dokl. Akad. Nauk SSSR* 231 (No. 5), 1182 (1976).
52. B. S. Sokolov, *Mem. Am. Assoc. Pet. Geol.* 19, 204 (1973).
53. M. A. Fedonkin, *Tr. Geol. Inst. Akad. Nauk SSSR* 342, 1 (1981).
54. C. R. Scotese *et al.*, *J. Geol.* 87, 217 (1979); P. A. Jell, *ibid.* 82, 319 (1974); L. P. Zonenshain and A. M. Gorodnitskiy, *Geotektonika* 2, 3 (1977).
55. M. D. Brasier, R. A. Hewitt, C. J. Brasier, *Geol. Mag.* 115, 21 (1978); M. D. Brasier and R. A. Hewitt, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 27, 35 (1979).
56. S. P. Alpert, in *Trace Fossils 2*, T. P. Crimes and J. C. Harper, Eds. (Seel House, Liverpool, 1977), pp. 1–8; N. L. Banks, in *Trace Fossils*, T. P. Crimes and J. C. Harper, Eds. (Seel House, Liverpool, 1970), pp. 19–34; J. W. Cowie and A. M. Spencer, in *ibid.*, pp. 91–100; B. Daily, *Univ. Adelaide Cent. Precamb. Res. Spec. Pap. 1* (1972), p. 13; B. D. Webby, *Lethaia* 3, 79 (1970).
57. G. Vidal, *Fossils Strata* 9, 1–57 (1976); *Nor. Geol. Unders. Bull.* 362, 1 (1981).
58. W. A. S. Sarjeant, *Rev. Micropaleontol.* 9, 201 (1967); S.-C. C. Lo, *Precamb. Res.* 13, 109 (1980).
59. P. Cloud, S. M. Awramik, K. Morrison, D. G. Hadley, *Precamb. Res.* 10, 73 (1979).
60. M. Moorman, *J. Paleontol.* 48, 524 (1974).
61. P. E. Cloud and M. A. Semikhatov, *Am. J. Sci.* 267, 1017 (1969); M. R. Walter, *Stromatolites and the Biostratigraphy of the Australian Precambrian and Cambrian* (special paper 11, Paleontological Association, London, 1972); M. R. Walter and W. V. Preiss, *24th Int. Geol. Congr.* Sec. 1, 85 (1972).
62. W. V. Preiss and I. N. Krylov, *Izv. Akad. Nauk SSSR Geol. Ser.* 7, 61 (1980).
63. A. D. Squire, *Geol. Mag.* 10, 223 (1973).
64. E. L. Yochelson and E. Stump, *J. Paleontol.* 51, 872 (1977).
65. C. A. L. Isotta, A. C. Rocha-Campos, R. Yoshida, *Nature (London)* 122, 466 (1969); E. L. Yochelson and H. E. Herrera, *Ameghiniana* 11, 283 (1974); T. R. Fairchild, *Soc. Bras. Geol. 30th Congr. Bras. Geol. Recife Bol.* 1, 181 (1978).
66. T. P. Crimes, J. Legg, A. Marcos, M. Arboleya, in *Trace Fossils 2*, T. P. Crimes and J. C. Harper, Eds. (Seel House, Liverpool, 1977), pp. 91–138.
67. W. C. Brøgger, *Norge i 19de Aarhundred* (Kristiana Central Torylekerret, 1900), vol. 1, pp. 1–32; B. Asklund, *Colloq. Int. CNRS* 76, 39 (1958).
68. R. B. Neuman and A. R. Palmer, *20th Int. Geol. Congr. El Sistema Cambrico* (1956), part 2, p. 427.
69. N. Menchikoff, *Bull. Soc. Geol. Fr.* 16, 451 (1949); P. Pruvost, *Bull. Geol. Soc. Belge* 60, 43 (1951).
70. B. S. Sokolov, *Colloq. Int. CNRS* 76, 104 (1958).
71. F.-T. Chung, *Sci. Sin.* 20 (No. 6), 818 (1977).
72. B. S. Sokolov, *Izv. Akad. Nauk SSSR, Ser. Geol.* 5, 21 (1952).
73. ______, *Siberian Branch Akad. Nauk SSSR Trans. Inst. Geol. Geofiz.* 232, 5 (1975).
74. B. M. Keller, *Geol. Mag.* 116, 419 (1979).
75. K. E. Yakobson and N. S. Krylov, *Int. Geol. Rev.* 20, 709 (1978).
76. W. B. Harland and I. N. Herod, *Geol. J.* 6, 189 (1975), special issue.
77. M. R. Walter, *Aust. Bur. Miner. Res. Geol. Geophys. Rep. 214* (1980), microform Mf92 with chart.
78. Y. Qian, M.-E. Chen, Y.-Y. Chen, *Acta Paleontol. Sin.* 18 (No. 3), 207 (1979); R-J. Cao and Y-Z. Liang, *Acad. Sin. Mem. Nanjing Inst. Geol. Paleontol.* 5, 1 (1974); P. Cloud, field notes, October 1979; M. F. Glaessner, field notes, May 1979.
79. S. N. Serebryakov, personal communication.
80. A. P. Brangulis, V. V. Kirjyjanov, K. A. Mens, A. Yu. Rozanov, *U.S. Geol. Surv. Open-File Rep. 81-743* (1981), p. 26; B. S. Sokolov, personal communications.
81. S. Føyn and M. F. Glaessner, *Nor. Geol. Tidsskr.* 59, 25 (1979); K. Bjørlykke, A. Elvsborg, T. Høy, *ibid.* 56, 233 (1976); K. Bjørlykke, *Geol. Surv. Can. Pap. 78-13* (1978), p. 49.
82. M. M. Anderson, *Can. J. Earth Sci.* 9, 1710 (1972); B. Greene and H. Williams, *ibid.* 11, 319 (1974); S. Bengston and T. P. Fletcher, *U.S. Geol. Surv. Open-File Rep. 81-743* (1981), p. 18.
83. F. G. Young, *Geol. Surv. Can. Bull.* 288, 1 (1979); H. Gabrielse, *Am. J. Sci.* 272, 521 (1972).
84. C. A. Nelson, *Geol. Mag.* 115, 121 (1978); P. Cloud, *Am. J. Sci.* 273, 193 (1973); S. P. Alpert, in *Trace Fossils 2*, T. P. Crimes and J. C. Harper, Eds. (Seel House, Liverpool, 1977), vol. 2, pp. 1–8.
85. While this article was in press a parallel endorsement of a new basal Phanerozoic System and Period was published by R. J. F. Jenkins [*Trans. R. Soc. South Aust.* 105 (No. 4), 179 (1981)].
86. We are grateful to so many individuals and organizations for help with aspects of the research reported here that space does not permit appropriate acknowledgment. We mention therefore only our special obligation to Dr. W. V. Preiss of the South Australian Department of Mines and Energy, to the several colleagues in the Soviet Union and China who furthered our researches there, and to the Queen's Fellowship Committee of the Australian Department of Science and the Baas Becking Geobiological Laboratory, who sponsored P.C.'s year in Australia.

Gould, S.J.
"The Ediacaran experiment"
Natural History
American Museum of Nat. Hist.
Used by permission

The Ediacaran Experiment

Life's first try didn't pan out. But from this fiasco, scientists are gaining insights that may lead to a general theory of extinction

by Stephen Jay Gould

To many outsiders, Indianapolis is nothing but one weekend a year and 500 miles of auto racing. In continuous reality, it is an attractive city filled with modern amenities and a liberal sprinkling of those older structures that unite our frenetic and uncertain present with a more comforting past. Last week, on a break from stated duties, I wandered along through the Murat Temple of the Shrine and the enormous cathedral of Scottish Rite Masonry. These lodges must once have dominated the social life of Indianapolis; they may yet, for all I know, be important. But their gigantic buildings look forlorn and abandoned—cavernous Victorian rooms in dark wood and stained glass, dimly lit by available light, filled with old, overstuffed chairs occupied rarely by a few elderly men in odd-shaped hats. Surely, the old order changeth.

I was in Indianapolis to attend the annual meeting of the Geological Society of America. There I watched, listened, and joined the debate as a group of my colleagues in paleontology began to dismantle an old order of thinking about old objects—and to construct a new and striking approach to a major feature of life's history on earth: mass extinctions.

Paleontologists have known about mass extinctions from the inception of our science as a modern discipline. We have used them to mark the major divisions of our geological time scale—the boundaries between eras. The Permian extinction that rang out the Paleozoic era eliminated half the families of marine invertebrates; the Cretaceous extinction, marking the transition from Mesozoic to Cenozoic eras, wiped out more than 25 percent of marine families, along with the most popular of all terrestrial creatures, the dinosaurs.

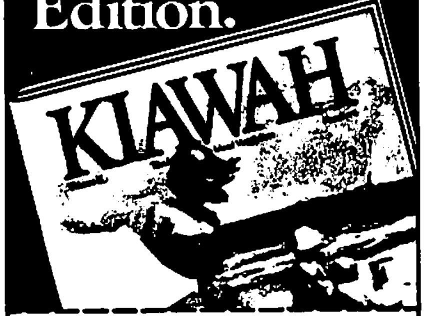
Nonetheless, though we have always acknowledged the reality of these great dyings, we have tried, in a curious way, to mitigate their effects, probably because our strong biases for gradual and continuous change force us to view mass extinctions as anomalous and threatening. We have, in short, attempted to depict mass extinction as a simple, quantitative extension of the slower disappearance, species by species, that characterizes normal times—larger and more abrupt to be sure, but basically just more of the same. We have pursued two principal strategies to temper mass extinctions and bring them into harmony with events of ordinary times. First, we have emphasized continuity across the boundaries by trying to find direct ancestors for new forms that appear after an extinction among species that flourished just before the event. Second, we have toted the numerical patterns of extinctions to argue that the peaks were neither high nor abrupt enough to please a person of catastrophic bent—that is, we have argued that pulses of extinction were preceded by gradual declines lasting for millions of years, and that the peaks themselves do not stand so noticeably above the "background" rates of normal times.

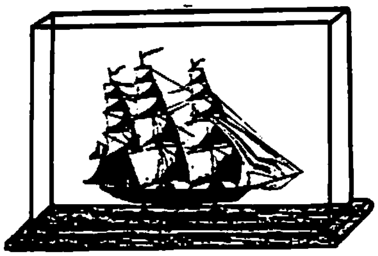
Both these traditions were strongly challenged in Indianapolis in a series of separate and ostensibly unconnected papers that point to a common conclusion: mass extinctions have been more frequent, more unusual, more intense (in numbers eliminated), and more different (in effect versus the patterns of normal times) than we had ever suspected. Any adequate theory of life's history will have to treat them as special controlling events in their own right. They will not be fully explained by the evolutionary theory we have constructed for interaction among organisms and populations of normal times—that is, by nearly all of conventional evolutionary theory as it now stands.

The centerpiece of this unplanned assault upon tradition lay in a paper presented by Adolf Seilacher, professor of geology at Tübingen in Germany. Dolf is the greatest observer I have ever had the privilege of knowing. He looks at common objects, scrutinized by generations of researchers, and invariably sees something new and unexpected. This time he turned his superior gaze upon the oldest of all metazoan (multicellular animal) assemblages—the Ediacaran fauna. His paper offered a fundamental reinterpretation of these fossils, complete with wide-ranging implications for the entire history of life—and I sat spellbound as wave after wave of expanded meaning cascaded over me.

About 570 million years ago, our modern fossil record began with the greatest of geological bangs—the Cambrian explosion. Within a few million years, nearly all major groups of invertebrates with hard parts made their first appearance in the fossil record. For fully three billion years before, life had been little more than a long sequence of bacteria and blue-green algae. But we do encounter one exception—first discovered in Australia but now known throughout the world—the Ediacaran fauna (named for the main Australian locality). In rocks just predating the Cambrian explosion, we find a moderately diverse assemblage of medium to large (up to a meter in length), soft-bodied, shallow-water marine invertebrates.

In the continuationist tradition that I identified above as a first strategy for softening the impact of mass extinctions, paleontologists have always tried to identify





the Ediacaran animals with modern groups. Thus, we have Ediacaran jellyfish, corals, and worms—a continuity of evolutionary relationship across the greatest of all geological boundaries. Yet, as I argued just a few months ago in my column on conodonts (July 1983), the traditional ploy of forcing old and problematical animal fossils into modern taxonomic categories often fails badly. We must recognize that the early history of life should be studded with failed experiments—small groups that never achieved much diversity and bear only distant relationship with any modern animal. We might expect that our oldest fauna should contain a large number of such curiosities—yet all Ediacaran animals have been shoehorned, often with considerable effort, into modern groups.

Dolf Seilacher now argues, turning the old view completely on its head, that the Ediacaran fauna contains not simply a few creatures with no modern analogues—but that every animal in it shares a basic mode of organization quite distinct from the architecture of living groups. The entire Ediacaran fauna, in other words, represents a unique and extinct experiment in the basic construction of living things. Our planet's first fauna was replaced after a mass extinction, not simply improved and expanded.

Dolf began by showing that the similarities of Ediacaran and modern animals are misleading and superficial, and that the Ediacaran forms could not work as their supposed living counterparts. Nearly all Ediacaran fossils have been falsely fit into three modern groups: jellyfish, corals, and segmented worms. Living jellyfish move by contracting a prominent ring of concentric muscles located at the outer edge of their bell; radial grooves for feeding lie within the concentric muscles, toward the center. But the so-called Ediacaran medusoids have a reversed arrangement that could not work in the same way: concentric structures surround the center, and radial grooves lie on the outside.

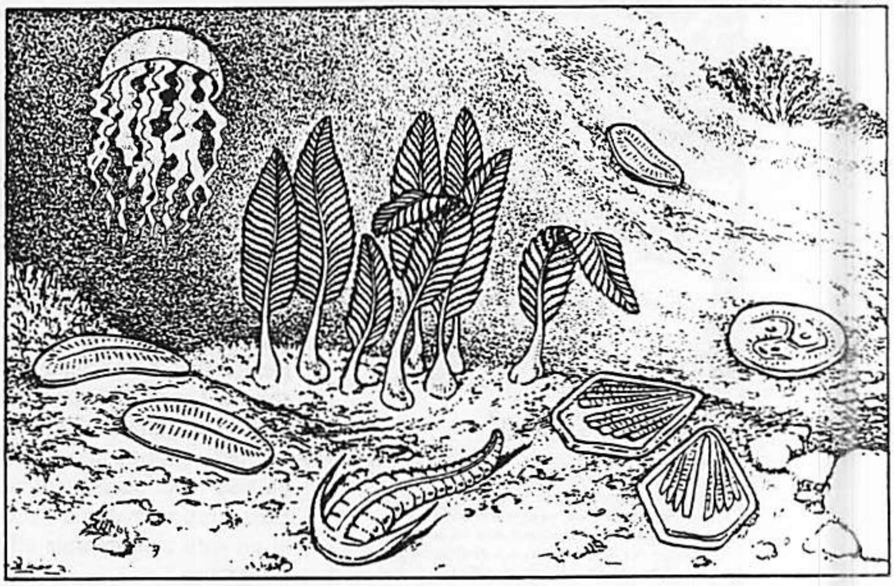
Modern alcyonarian corals ("soft" corals, or sea pens) invariably bear distinct branches, often springing from a common stem. The branches must be separated so that water, bearing oxygen and nutrients, can reach the individual polyps (members of the colony) growing on them. At first glance, the Ediacaran "sea pens" look superficially like their modern counterparts in general shape, but they form a continuous, quilted structure, not a set of separated branches—and could therefore not operate like a modern soft coral colony. The Ediacaran "worms" are segmented and bilaterally symmetrical like their supposed modern analogues, but so are many other creatures—and such a basic and repeatable architecture need not imply close relationship. In other respects, the Ediacaran creatures are most unwormlike. They may be up to a meter in length and flat as a pancake—more like films than the substantially thickened bodies of most modern segmented worms.

After tracing the differences between Ediacaran animals and their supposed modern counterparts, Seilacher examined the similarities that unite all Ediacaran forms. They seem to share an architecture only rarely utilized by modern animals—and not by any living creature ever linked to an Ediacaran fossil. They look like ribbons, pancakes, and films, sometimes slightly "blown up" as air mattresses with a foliate or quilted structure.

The Ediacaran animals evolved before any creature had invented mineralized skeletons or any external hard parts. Perhaps their unique *Bauplan* (to use the convenient German term for a basic scheme of organic architecture), or "building plan," records a pathway to large size that animals without supporting hard parts might follow—light and thin structures, woven together for added strength. In any case, and following a favorite theme of these columns for more than a decade, the Ediacaran fossils seem to represent one of two possible solutions—the one *not* followed by modern animals—to the basic structural problem of large size: the imposed decline of surfaces relative to volumes since surfaces (growing as length squared) must increase more slowly than volumes (growing as length cubed) as objects of similar shape get bigger. Since so many organic functions depend upon surfaces (respiration and feeding, to name just two) yet must serve the entire body's volume, this decline in relative surface cannot be tolerated for long.

Of the two possible solutions, most modern animals have retained their rounded or globular shapes but have evolved internal organs to increase surface areas—lungs for respiration and the complexly folded surface of the small intestine for absorption of food, for example. Another potential solution, followed rarely today but exploited by some large parasites, including tapeworms, permits large size without any internal complexity by changing the body's basic shape into something very thin—a ribbon or pancake—so that no internal space will be far from the external surface, the only locus of respiration and absorption of food in the absence of internal organs. The Ediacaran animals, as a group, have followed this second pathway to large size and therefore

Juan Barbens, AMNH



In conventional reconstructions of Ediacaran animals, they are depicted as the ancestors of modern forms—jellyfish, soft corals, and worms.

from *A View of Life* by S. E. Luria, S. J. Gould, and S. Singer

represent a coherent fauna strikingly different from any modern counterpart in basic design.

Following a progressivist bent, I might be gratified that life's first "try" used the simpler of two solutions—a change in body shape rather than an evolution of complex internal organs. Be that as it may, the important point remains that if Seilacher is right, the Ediacaran fauna represents a different, unique, and coherent experiment in organic architecture—not a set of precursors for modern animals. To emphasize this discontinuity, the first Paleozoic fauna with hard parts, the so-called Tommotian assemblage, is filled with tiny tubed, coiled, and cap-shaped creatures bearing precious little similarity to Ediacaran forms. The ancestry of these later creatures may reside in indirect evidence for other Precambrian animals not included among the Ediacaran fossils. We have abundant records, through "trace fossils" of feeding and burrowing tubes but, alas, no "body fossils" as yet, of animals with more conventional rounded shapes—a good source for later Tommotian descendants.

Seilacher ended his paper with a stunning argument. We have, he pointed out, been searching with no success, and little hope, for complex extraterrestrial creatures, primarily because we wonder so powerfully what an independent experiment in the development of life might produce. How similar would it be with life on earth? how strong a constraint does the physics and chemistry of objects impose? how different could life be elsewhere? It now appears, however, that an independent experiment occurred right here on earth, expressing itself as the Ediacaran fauna, our first assemblage of multicellular animals.

As for the theme of mass extinctions, we used to say that the first era boundary, between Precambrian and Paleozoic some 570 million years ago, was an anomaly marked by a profound radiation (the Cambrian explosion) but no previous extinction. But if the Ediacaran fauna, lying just below the base of the Paleozoic in strata throughout the world, represents a coherent and different experiment in life's architecture, then a major extinction marks this initial boundary as well. The first strategy for mitigating mass extinction fails, and we trace little continuity across the opening and most profound boundary of life's complex history.

Other papers at Indianapolis challenged the second strategy by arguing for a greater separation in effect and magnitude of mass extinctions and events in ordinary times. Some conclusions of previous years, already documented in these columns, have paved the way: (1) An asteroidal impact as the source, or at least the *coup de grâce*, of our terminal Cretaceous extinction (column of June 1980)—organisms, after all, can scarcely "prepare" for such a trigger. (2) David Raup's estimate (column of November 1980) that a 50 percent extirpation of families (the counted figure for the Permian extinction) might translate to as much as 96 percent of all species (half the families mean many more species since most species die without eliminating their families—a more inclusive category—while the death of a family must include all its species). For a removal so profound, we must seriously consider the possibility that entire groups will be lost for purely random reasons. (3) The calculation of Raup and Jack Sepkoski (column of August 1982) that major extinctions stand higher and more distinctly above the background level than previously recognized.

Some refinements and major surprises were added to this growing theme in papers presented at Indianapolis. Jack Sepkoski, a former student of mine now flourishing mightily at the University of Chicago, has spent years compiling the most consistent and complete data set on extinctions ever developed—a listing at the family level that includes everything from protozoans to mammals. With these data, we finally have a basis for the fine-scaled consideration of quantitative patterns in extinction that this second strategy demands. (Good science may require genius and imagination, as these columns so often emphasize, but never forget that new conclusions are the fruit of hard empirical work as well; without these labors, highfalutin thought is so much waffling.)

Using the Sepkoski data, Raup and Sepkoski have now identified a striking cyclicity in mass extinctions for 225 million years since the great Permian dying. Every 26 million years, with eight hits and just two apparent misses (a pattern too regular and striking to be accidental on statistical grounds), we find a peak of mass extinction; all previously identified major extirpations lie right on the highs of this 26-million-year cycle. What cause could yield a periodicity so regular, yet so widely spaced? If we understand geology aright, no purely internal process of climate, volcanism, or plate tectonics cycles so regularly with such a long period. Raup and Sepkoski therefore speculate that some astronomical cycle must be involved—a solar or galactic property, although for the moment, we have no idea what. If the cycles are so frequent and caused by events so utterly beyond an organism's control or anticipation (how can populations track a 26-million-year cycle?), and if the mass extinctions shape life's pattern so fundamentally, then mass dying is not ordinary death extrapolated.

David Jablonski, a paleobiologist from the University of Arizona at Tucson, then added two cogent points to emphasize the abruptness and the different character of mass extinctions. For abruptness, Jablonski noted that the raw data of mass extinctions often include a long period of apparently slow and steady decline among groups that crash more profoundly at the peak itself. These slow declines have long been taken as a sign of continuity between normal and mass extinction. But are they real or an artifact of our imperfect geological record?

For more than one hundred years, geologists have sought terrestrial agents to associate with mass extinction. The litany is long, yet all but one have failed—mountain building, volcanism, fluctuations in temperature, to name just a few old and unsuccessful favorites. Falling sea level represents the one good correlation (and the 26-million-year-cycle theorists had better take it into account). Nearly all mass extinctions are preceded by a marked regression of sea level.

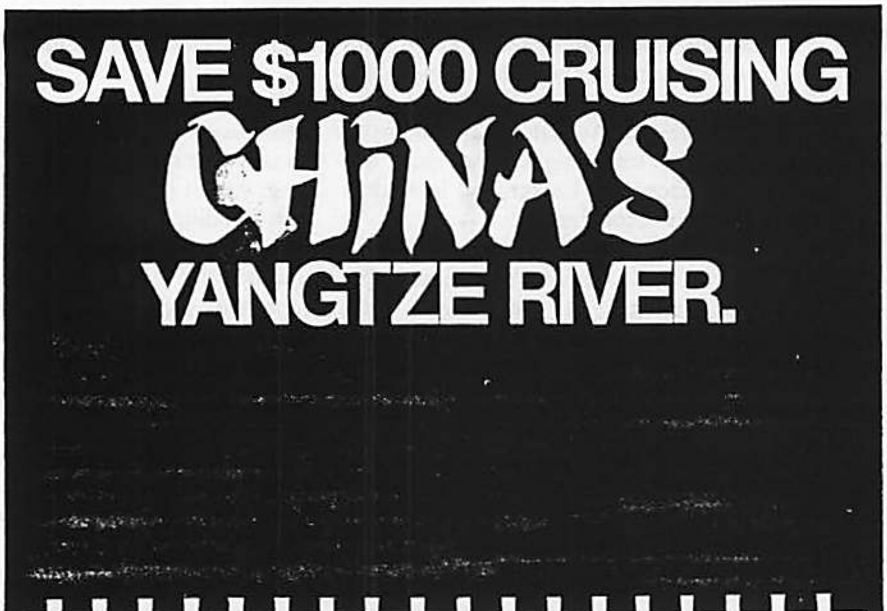
Falling sea level may well be a causal factor in extinctions (our fossil record is strongly biased toward shallow-water marine invertebrates), but it also imposes an obvious artifact upon the data. As sea level falls, fewer sedimentary rocks are deposited to hold the fossils of these regressive epochs. Perhaps the slow decline that precedes most mass extinctions only records the decreasing volume of available rock for finding fossils, not a true and gradual decrease presaging the later peak.

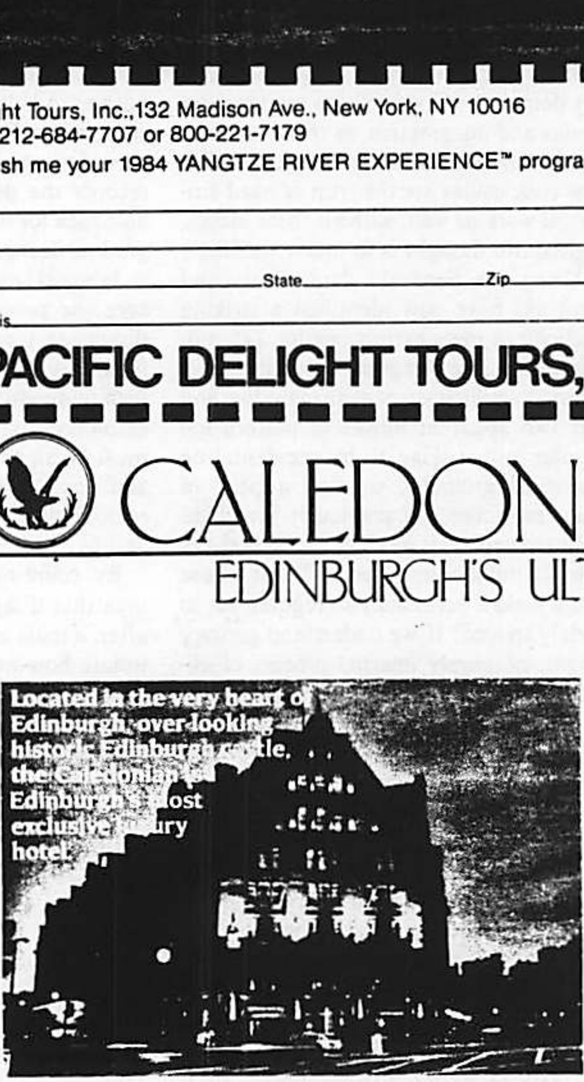
Jablonski used a clever method to measure the potential artifact. Some forms disappear from the record as sea level falls, only to be found again when seas return to deposit more rocks after the mass extinction itself. These temporary losses must be an artificial effect of falling seas and decreasing amounts of fossiliferous rock. Jablonski refers to these reappearing groups as "Lazarus taxa."

By counting the number of Lazarus taxa that disappear before, but reappear after, a mass extinction, Jablonski can estimate how much of a counted slow decline before a mass extinction might be the artificial result of less available rock for finding fossils, and how much must record a real and gradual event tying peaks of mass extinction with normal times before.

In some cases, subtraction of the Lazarus taxa still leaves a residue of slow disappearance, and the pattern must be real (decline of ammonites before the Cretaceous extinction, for example). But for many groups at the Cretaceous, measured slow decline can be explained entirely by the artifact of decreasing available rock. Thus, the Cretaceous extinction, and others as well, may be more abrupt than we have previously realized. The asteroid's candidacy is strengthened, and its dominant role affirmed. Mass extinction is something quick and special.

Jablonski then examined the behavior of groups during normal times and during episodes of mass extinction to see if he could detect consistent differences that might accentuate the special character of mass extinctions. He found some intrigu-





ing disparities. Some branches of the evolutionary tree contain many species either because new species form easily or because they are unusually resistant to extinction once they arise. Jablonski calls these "species-rich clades" as opposed to "species-poor clades," or branches that never contain many species.

During normal times, species-rich clades tend to increase their numbers of species continually—and to win increasing numerical advantage over species-poor clades. The environments of normal times must encourage either rapid speciation or persistence thereafter. But why, then, don't species-rich clades take over the biosphere entirely? Jablonski finds that these same species-rich clades fare worse than species-poor clades during mass extinctions. The individual species in species-poor clades have wider geographic ranges and broader ecological tolerances than the narrow-niched taxa of species-rich clades. This geographic and ecological breadth probably protects these species in extreme environments that mass extinction must generate. These same features of breadth may cut down their rate of speciation in normal times (fewer opportunities for isolation and exploitation of new environments), thus rendering their groups species-poor.

This contrary behavior of species-rich clades in normal and catastrophic times preserves a balance that permits both species-rich and species-poor clades to flourish throughout life's history. More important in our context, it emphasizes the qualitative difference between normal times and catastrophic zaps. Mass extinctions are not simply more of the same. They affect various elements of the biosphere in a distinctive manner, quite different from the patterns of normal times.

As we survey the history of life since the inception of multicellular complexity in Ediacaran times, one feature stands out as most puzzling—the lack of clear order and progress through time among marine invertebrate faunas. We can tell tales of improvement for some groups, but in honest moments we must admit that the history of complex life is more a story of multifarious variation about a set of basic designs than a saga of accumulating excellence. The eyes of early trilobites, for example, have never been exceeded for complexity or acuity by later arthropods. Why do we fail to find this expected order?

Perhaps the expectation itself is faulty, a product of a pervasive, progressivist bias in Western thought and never a prediction of evolutionary theory. Yet if natural selection rules the world of life, we should see some fitful accumulation of better and

278

more complex design through time—amidst all the fluctuations and backings and forthings that must characterize a process primarily devoted to constructing a better fit between organisms and changing local environments. Darwin certainly thought so when he wrote:

> The inhabitants of each successive period in the world's history have beaten their predecessors in the race for life, and are, insofar, higher in the scale of nature; and this may account for that vague yet ill-defined sentiment, felt by many paleontologists, that organization on the whole has progressed.

I regard the failure to find a clear "vector of progress" in life's history as the most puzzling fact of the fossil record. But I also believe that we are now on the verge of a solution, thanks to a better understanding of evolution in *both* normal and catastrophic times. We need a two-tiered explanation for patterns (or nonpatterns) in the history of life.

I have devoted the last ten years of my professional life in paleontology to constructing an unorthodox theory for lack of expected patterns during normal times—the theory of punctuated equilibrium. Niles Eldredge and I, who must admit responsibility for this particularly uneuphonious name, argue that the pattern of normal times is not one of continuous adaptive improvement within lineages. Rather, species form rapidly in geological perspective (thousands of years) and tend to be highly stable for millions of years thereafter. Evolutionary success must be assessed among species themselves, not at the traditional Darwinian level of struggling organisms within populations. The reasons that species succeed are many and varied—high rates of speciation and strong resistance to extinction, for example—and often involve no reference to traditional expectations for improvement in morphological design. If punctuated equilibrium dominates the pattern of normal times, then we have come a long way toward understanding the curiously fluctuating directions of life's history. Until recently, I suspected that punctuated equilibrium might resolve the dilemma all by itself.

I now realize that the fluctuating pattern must be constructed by a complex and fascinating interaction of two distinct tiers of explanation—punctuated equilibrium for normal times, and the different effects produced by separate processes of mass extinction. Whatever accumulates by punctuated equilibrium in normal times can be broken up, dismantled, reset, and dispersed by mass extinction. If punctuated equilibrium upset traditional expectations, mass extinction is even worse. Organisms cannot track or anticipate the environmental triggers of mass extinction. No matter how well they adapt themselves to environmental ranges of normal times, they must take their chances in catastrophic moments. And if extinctions can demolish more than 90 percent of all species, then we must be losing groups forever as a result of pure bad luck among a few clinging survivors designed for another world.

Heretofore, we have thrown up our hands in frustration at the lack of expected pattern in life's history—or we have sought to impose a pattern that we hoped to find on a world that does not really display it. Perhaps now we can navigate between a Scylla of despair and a Charybdis of comforting unreality. If we can develop a general theory of mass extinction, we may finally understand why life has thwarted our expectations—and we may even extract an unexpected kind of pattern from apparent chaos. The fast track of an extraordinary meeting in Indianapolis may be pointing the way.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

Seilacher, A. "Late Precambrian
and Early Cambrian Metazoa:
Preservational or Real Extinc-
tions?" In Patterns in Earth
Evolution
Springer-Verlag
Used by permission

Patterns of Change in Earth Evolution, eds. H.D. Holland and A.F. Trendall, pp. 159-168.
Dahlem Konferenzen 1984. Berlin, Heidelberg, New York, Tokyo: Springer-Verlag.

Late Precambrian and Early Cambrian Metazoa: Preservational or Real Extinctions?

A. Seilacher
Geologisches Institut der Universität Tübingen
7400 Tübingen, F.R. Germany

Abstract. The interpretation of Ediacara-type body fossils in terms of modern soft-bodied metazoans must be questioned. Their morphology rather suggests foliate, non-locomotory quasi-autotrophs. Their mode of preservation, which has no counterpart in comparable post-Vendian rocks, remains problematical, since associated trace fossils attest to oxic conditions and the presence of worm-like heterotrophic burrowers in the same environment. It seems that Vendian biota mark not simply a non-skeletal start of metazoan evolution, but a distinct episode in the history of life that was followed by a major extinction.

INTRODUCTION

Paleontologic research during the last decades has changed our views of early metazoan evolution tremendously. We recognize not only a sudden appearance of a diversified Cambrian megafauna, but also a pre-trilobite (Tommotian) shelly fauna, which was preceded by a phase (Vendian-Ediacarian), in which larger, presumably metazoan organisms were present that had not yet developed mineralized skeletons. We can only speculate about a still earlier phase of metazoan history. One could, for instance, interpret the presence of similar planktonic larvae in different phyla of modern marine organisms, not simply as a convergently evolved means of dispersal, but as an archaic heritage from a hypothetical planktonic and larviform stage of metazoan evolution. Metazoans of this stage would not have left a fossil record because of their microscopic size and non-skeletal nature. The Vendian radiation might have expressed the metazoan conquest of the benthic realm, in which size was no longer limiting. The resulting diversified extension of ontogenetic histories could therefore follow diverse pathways of adaptation to the new habitat: Wormlike infaunal organisms left a trace fossil record and epifaunal organisms left a record as body fossils. Cambrian radiations followed the introduction of mineralized skeletons that opened new possibilities of constructional design.

In this paper we do not want to emphasize the appearance, but rather the disappearance of these early metazoan faunas, and to discuss whether they are preservational in nature and hence reflect major changes in the necrolytic and diagenetic regimes, or whether they are due to major extinctions.

THE PROBLEM OF THE EDIACARAN FOSSILS

Several decades ago the discovery of impressions of distinctive soft-bodied organisms in sandy deposits of late Precambrian age in South Australia created a sensation. The Cambrian appearance of trilobites and other "shelly" organisms with mineralized skeletons no longer marked the beginning of metazoan radiation but could be regarded as a second major step in the evolution of multicellular animals. Since the discovery of the Ediacara fauna, similar or even identical impressions have been found in similar stratigraphic positions in more than a dozen localities - particularly in Australia, South Africa, the Soviet Union, England, and Newfoundland (1, 2, 7). Without a doubt we are dealing not with a local preservational "bonanza" such as the Burgess Shales, but with the distinctive fauna of an Ediacarian period that preceded the Cambrian radiation.

This development has raised a problem as vexing as the fauna itself: that of its preservation. Most authors agree that Ediacara-type organisms were soft-bodied animals, similar to the jellyfishes that are so common in most occurrences. Such organisms have existed ever since. Nevertheless, no remains comparable to the Ediacara soft-bodied fossils, at least like their more characteristic members, have been found in younger rocks. Individual "Fossil-Lagerstätten" with an extraordinary level of preservation, including that of soft-bodied organisms, have formed repeatedly since late Precambrian time. These deposits usually consist of very fine-grained sediments deposited in continuously or intermittently stagnant basins, or of coarser deposits in which whole organisms became smothered. In contrast, Ediacara fossils are found in sandstones and sand/shale sequences whose sedimentary structures and trace fossil content are characteristic of normal and well aerated shallow marine environments (6). A question that is particularly relevant in the context of this workshop is this: "Why did the Ediacaran mode of preservation become 'extinct' with the beginning of the Cambrian?" The recent discovery of non-skeletal metazoans in the Precambrian of the Oleniok Uplift (Fedonkin and Rozanov, Moscow, personal communication) will be of great interest in this connection, since they are preserved in thinly bedded dolomites, possibly of the lithographic type, i.e., in a facies with different preservational properties.

We must first test the reality of the "extinction" by answering the following questions:

1. Could the extinction be an artefact in the sense that we are dealing with pseudo-fossils of inorganic origin? This is certainly true of many so-called "Precambrian fossils." The supposed jellyfishes from the Hakatai Shales of the Grand Canyon (3, 14) are surely nothing more than compactional haloes around sand salt crystals. Another example of pseudo-fossils are the sinuous, wormlike ridges of "Manchuriophycus" (3, 13), which are only shrinkage patterns of thin mud lenses deposited in ripple troughs. The complex forms of Ediacaran impressions clearly exclude the possibility of a mechanical origin.

2. Do similar impressions occur in post-Vendian deposits, and have they so far escaped attention? Most Ediacaran fossils are faint but distinctive impressions on sole faces of sandstone beds; these are also optimal sites for the preservation of burrows. Students of trace fossils have probably searched square miles of such sole faces in rocks of all ages without ever coming across a truly Ediacaran type of preservation. It is therefore most unlikely that any similar impressions occur in post-Vendian deposits.

A full reevaluation of Ediacaran fossils is clearly needed; the following section of this paper is limited, however, to comments on published material which are immediately relevant to the final discussion and suggestions.

Ediacaran Medusoids

The most common elements of the Ediacara fauna are round impressions with radial and concentric structures that have been classified as medusoids. Most published pictures are small and lighted in a variety of different directions so that is is difficult to judge critical details. We therefore owe a great deal to M. Wade (18) for a critical analysis of preservational details. The mode of preservation of medusoids is consistently different from that of other groups of Ediacaran fossils. Unfortunately, Wade's reconstructions of preservational processes are based on the unproven assumption that these were really jellyfishes.

Stratinomy. Fossil-bearing sole faces record the change of pelitic to sandy sedimentation during an episodic rise in turbulence (commonly a storm event) in which coarser material was deposited in a graded fashion following an initial phase of mud erosion. Since medusoids are mainly found on these sole faces, they must have reached the bottom together with benthic types of Ediacaran organisms. This is not what we might expect from such light bodies. In the lithographic limestones (Upper Jurassic) of Pfalzpaint near Solnhofen - one of the few localities where undoubted jellyfishes did become fossilized - they are invariably found within the event-generated flinz beds and not at their base, where the heavier bodies of arthropods, echinoderms, squids, and fishes are located. If this is the case in a coccolithic mud, how much more should it apply to sand!

Postmortem deformation. Solnhofen specimens also show the characteristic alteration that occurs when a jellyfish shrinks by dehydration within sediments. Radial and concentric wrinkles develop in the peripheral zone of the umbrella, where the radial and circular muscle fibers must be located in medusoid swimmers, while the central part of the impression remains smooth or carries a regular number of radial furrows corresponding to the mesentheria within the stomach cavity. Ediacaran medusoids, in contrast, show strong concentric folds in the center and radial structures at the periphery.

Radial feeding burrows? Radial backfill burrows that reflect the probing of wormlike sediment feeders around a vertical burrow are among the structures that are commonly mistaken for fossil jellyfishes. Medusina, Palaeosemaeostoma, Kirklandia, and Gyrophyllites (8) are familiar examples from the Phanerozoic record. Brooksella from the late Precambrian of the Grand Canyon may be another example. Among the Ediacara medusoids, Mawsonites spriggi (see (5), Figs. 1-2 and (1), cover picture) is clearly such a burrow system. Like the Cretaceous Kirklandia, it shows several tiers of sharply separated radial lobes that extend from the central shaft with a downward inclination. Seleniform backfill lamallae can also be seen in the lobes as well as in the vertical shaft of published Mawsonites specimens.

Actinian (?) burrows? Circular burrows, possibly made by actinians that lived in sandy sediments, are another source of confusion. Paleozoic examples (Bergaueria) are usually preserved as smooth and dome-shaped casts, sometimes with a central depression; Mesozoic representations (Solicyclus) have regular radial grooves around the smooth central field. Without careful examination of the original material, it is difficult to assign Ediacaran medusoids to this type of trace fossils; however, the separation of specimens of Medusinites, Edicaria, and Cyclomedusa from the surrounding bedding surface by a sharp groove and the eccentrically arranged annular furrows in some specimens (18) may indicate that they belong to this kind of trace fossils.

Sandy skeletons of actinians? A last possibility is related to a new interpretation of Protolyella, a three-dimensional fossil of Cambrian to Ordovician age with a globular base and radial plus concentric ridges on the truncated upper surface. In the past, it has been interpreted as the sandy filling of a medusoid stomach cavity. New material from the Ordovician of Jordan (16), however, shows that the concentric grooves of the upper surface continue inside the body as concentric hemispherical laminae that cannot be explained by sedimentation within a cavity. Rather they reflect the laminar growth, possibly within the gastral cavity of an actinian, of a heavy sandy skeleton that stabilized the animal on the soft substrate. The occurrence of a central depression at the base of many Protolyella casts suggests that the organisms were similar to those responsible for Bergaueria burrows.

A three-dimensional character is also suggested by some Ediacara "medusoids." In one specimen ((18), Fig. 2) they are aligned in the troughs of oscillation ripples in the same manner as Protolyella in Cambrian and Ordovician sandstones. The animals may have accumulated on and burrowed in the sand during the event and were then smothered by the muddy tail of the same tempestite (see (15), Fig. 5b for a similar behavior in ophiuroids). The round basal disk of the Pennatula-like Charniodiscus ((1), Fig. 2A) could be a similar sandy weight belt.

Conclusion: None of the numerous "medusoids" in the Ediacaran fauna can be interpreted with certainty as a jellyfish. Rather we seem to be dealing with a heterogeneous group of trace fossils and remains of unidentified benthic organisms.

Non-medusoid Body Fossils

Other enigmatic fossils of the Ediacaran fauna have been variously assigned to pennatulid coelenterates, annelids, and arthropods, i.e., to phyla still living today (5). Only one author (11, 12) referred them to an extinct phylum Petalonamae. Affiliation with modern creatures, based on crude morphologic similarities, should, however, be considered tentative. Sea pens, for instance, are colonial organisms in which the branches must be separated so that the polyps can function properly. Nevertheless, the branches are never spread apart in Ediacaran "sea pens," not even in specimens that are preserved within the sandstone in a three-dimensionally twisted fashion. It is also strange that among the "annelids" only very unusual, large and foliate forms (Dickinsonia; Spriggina) should be preserved, while normal cylindrical worms, whose presence is well documented by associated trace fossils, have never been found as body fossils in Ediacara-type deposits. The only vaguely worm-like body fossil (4) is a three-dimensional internal cast and not the flattened impression that one would expect.

Nature of the cuticle. In our present discussion, questions of taxonomic affiliation are of secondary importance. It suffices to state that the non-"medusoid" Ediacaran impressions are neither pseudo-fossils nor trace fossils, but true body fossil impressions of highly organized organisms. Their mode of preservation is characteristically different from that of the medusoids (18): instead of convex casts they form concave external molds on sandstone sole surface. This is not what we would expect, because a cavity resulting from burrowing, or from the decay of a soft body, at a sand/mud interface should be filled with the noncohesive sand that casts the mold left in the more cohesive mud. The situation is reversed only after the sand has become somewhat immobilized by cementation. Since non-medusoid impressions are also preserved within the sandstone, we conclude that they reflect a cuticular skeleton resistant enough to survive until the sand had become diagenetically altered.

On the other hand, the way in which these fossils have become deformed in particular situations (Dickinsonia draping over Arborea, (5), plate 103, Fig. 1; twisted Pteridinium within the sandstone, (5), plate 101, Figs. 1-3) suggests that the cuticle was flexible but rigid enough not to become wrinkled. It must have consisted of a highly elastic material. In sandy marine sediments of later ages such biomaterials have become digested by microorganisms - except under anoxic conditions. The presence of burrowing organisms in the Ediacara sediments rules out this possibility.

Functional morphology. A feature that Ediacaran "pennatulids," "chondrophores," and Dickinsonia have in common is their quilt-like repetitive structure which has been interpreted as expressing metameric segmentation. It could, however, be simply a means to stiffen a hydraulic skeleton biomechanically, somewhat in the manner of an air mattress; it would also have provided an internal compartmentalization that facilitated metabolic processes (17). The air mattress model does not apply to Spriggina, because some of the specimens are laterally bent, nor to Tribrachidium, Parvancorina, and Praecambridium, in all of which the boundaries between radial sectors stick out as sharp ridges instead of forming reentrant grooves. Nevertheless, these forms were also quilted and frond-like.

The frond-like shape of most Ediacaran fossils could be explained in terms of metabolism, since it maximizes the external surface. This may have been necessary for respiration at a time when the atmosphere and seawater contained less oxygen than today. One might even speculate that dissolved food could have been absorbed directly through the enlarged body surface, since orifices and intestines have not been observed in Ediacaran fossils. Another possibility is an association with photosymbiotic algae, which would benefit from a frond-like shape and could have provided food and oxygen directly to the host.

A mode of nutrition unfamilar in modern Metazoa is also indicated by the non-locomotory character of most Ediacara-type organisms. Morphological distinction between front and rear end - a universal attribute of mobile benthic organisms except regular echinoids - is found only in Spriggina and Parvancorina. The rest, including the benthic "medusoids," were probably more or less sessile. We do not, however, have indications of the presence of lophophores or other filtering devices that are the base for sedentary life in modern marine organisms.

All these considerations are, of course, based on the assumption that we are really dealing with the Metazoa in the modern sense, and this is not at all certain.

Conclusion: Non-medusoid body fossils of the Ediacaran fauna seem to be the remains of benthic organisms that are not referable to extant phyla. They had a flexible cuticle whose survival suggests that it was composed of biomaterial indigestible for the contemporary microorganisms.

EVOLUTIONARY SCENARIO

The problem that appeared to be mainly a matter of preservation at the outset of this review has thus turned out to be of basic significance for the theme of this Dahlem workshop. If one views the Ediacaran fauna as the initial stage of metazoan evolution, it could be incorporated into the sigmoidal part of the Cambrian diversity explosion that was unique in the sense that it took place in an ecological vacuum. In terms of diversity, this curve is still valid, because the number of species and the degree of provincial differentiation was very low in the Vendian biota. But if the Ediacara-type fossils turn out to be basically different from Metazoa in the modern sense and if their disappearance is not only an artefact of preservation, then the Cambrian radiation was preceded by a major extinction event in the benthic realm.

One could go a step further and ask whether a similar relationship might not exist between the shelly faunas of the Tommotian and the trilobite-dominated stages of the Cambrian. After all, animals with a taxonomy based on that of modern organisms could have suppressed the basic constructional and ecological differences of this earliest shelly fauna. The problem of the Tommotian biota also has preservational overtones, because the shelly fossils of this stage are all very small and phosphatized. But this time-specific mode of preservation continues throughout the remainder of the Cambrian. It became "extinct," however, in later periods, when suitable micromolluscs were still present but are significantly absent in the phosphatic residues studied by conodont workers (O. Walliser, personal communication).

An unusually high phosphorous content in Cambrian seas was probably responsible for the easy phosphatization of mineralized as well as non-mineralized skeletons (9, 10), as well as for the high percentage of originally phosphatic shells in Tommotian faunas.

We conclude that metazoan evolution was no smoother in its initial than in its later phases. Extinctions caused by changes in the physical environment preceded radiations that increased the overall diversity in a step-like mode. Since the levels of diversity were generally low, these early extinctions had little in the way of quantitative effects. Nevertheless, their consequences played a major role in the course of metazoan evolution.

SUGGESTIONS FOR FUTURE RESEARCH

This review of the literature has been admittedly cursory. We have come to the conclusion that the nature of Ediacara-type body fossils of Vendian age may have been misinterpreted. Not only their affiliation with extant phyla, but also their truly metazoan nature should remain open to question. More definite answers regarding their affiliations can be given only after a very critical reexamination of the fossil material. Such a study should include associated trace fossils, which seem to be less different from post-Vendian forms and that may represent the stocks from which marine epibenthic phyla radiated at a later time. The study of trace fossils and associated sedimentary structures will also provide better information about the depositional environments and the distribution of food in sediments during those early times.

REFERENCES

(1) Cloud, P., and Glaessner, M.F. 1982. The Ediacarian period and system: Metazoa inherit the earth. Science 217: 783-792.

(2) Fedonkin, M.A. 1981. White sea biota of the Vendian. Precambrian non-skeletal fauna of the northern Russian platform. Trans. Akad. Nauk. SSR. 342: 1-100.

(3) Glaessner, M.F. 1969. Trace fossils from the Precambrian and basal Cambrian. Lethaia 2: 369-393.

(4) Glaessner, M.F. 1979. An echiurid worm from the Late Precambrian. Lethaia 12: 121-124.

(5) Glaessner, M.F., and Wade, M. 1966. The late Precambrian fossils from Ediacara, South Australia. Paleontology 9: 599-628.

(6) Goldring, R., and Curnow, C.N. 1967. The stratigraphy and facies of the late Precambrian at Ediacara, South Australia. J. Geol. Soc. Austral. 14: 195-214.

(7) Jenkins, R.J.F. 1981. The concept of an "Ediacaran Period" and its stratigraphic significance in Australia. Tran. Roy. Soc. S. Austral. 105: 179-194.

(8) Moore, R.C. 1956. Treatise on invertebrate paleontology. Part F. Coelenterata. Lawrence, KS: University of Kansas Press.

(9) Müller, K.J. 1979. Phosphatocopine ostracodes with preserved appendages from the Upper Cambrian of Sweden. Lethaia 12: 1-

(10) Müller, K.J. 1982. Weichteile von Fossilien aus dem Erdaltertum. Naturwissenschaften 69: 249-254.

(11) Pflug, H.D. 1972. Systematik der jung-präkambrischen Petalonamae. Paläont. Z. 46: 56-67.

(12) Pflug, H.D. 1974. Feinstruktur und Ontogenie der jung-präkambrischen Petalo-Organismen. Paläonet. Z. 48: 77-109.

(13) Schindewolf, O.H. 1956. Über präkambrische Fossilien. In Geotektonishes Symposium zu Ehren von Hans Stille, ed. F. Lotze, pp. 455-480. Stuttgart: F. Enke.

(14) Seilacher, A. 1956. Der Beginn des Kambriums als biologische Wende. N. Jb. Geol. Paläont. 103: 155-180.

(15) Seilacher, A. 1982. Distinctive features of sandy tempestites. In Cyclic and event stratification, eds. G. Einsele and A. Seilacher. Heidelberg: Springer-Verlag.

(16) Seilacher, A. 1982. Paleozoic sandstones in southern Jordan: trace fossils, depositional environments and biogeography. Contribution, 1st Jordanian Geological Congress, Amman.

(17) Sernetz, M.; Rufeger, H.; and Kindt, R. 1982. Interpretation of the reduction law of metabolism. Exp. Biol. Med. 7: 24.

(18) Wade, M. 1968. Preservation of soft-bodied animals in Precambrian sandstones at Ediacara, South Australia. Lethaia 1: 238-267.

to begin with: we must stipulate that the universe we are talking about is finite.

I do not have the slightest idea whether in real fact the universe is finite, or indeed whether the expression "real fact" makes any sense in this connection. Qualified cosmologists (I am not one of them) seem to disagree on that point. Apparently no one is sure, and probably no one ever will be. If, however, the universe were postulated to be infinite, some rather striking consequences would seem to be logically inescapable.

If the universe is not merely vast, which it plainly is, but literally infinite, then there are not a hundred million or a billion billion habitable planets or any enumerable amount of them but an infinite number. Now, in infinity everything that is physically possible, no matter how improbable, must exist and in fact it must exist an infinite number of times. What actually exists here on earth is obviously physically possible. Therefore, if the universe is infinite, somewhere there is another reader exactly like you reading exactly what I have written here. In fact this scene must be repeated an infinite number of times. There must also be an infinite number of readers just slightly different from you reading an essay slightly different from this one. And so on! In that case of course life like ours and men like us do exist on other worlds, on an endless number of them.

I am naive enough to be completely awed by that consequence of a possible infinity. Yet it has no particular bearing on the inquiry in the following chapters. The parts of the universe, if any, that we cannot observe and with which we could not communicate by any means do not "really" exist as far as we are concerned. The parts that we can observe and with which we could (in principle) communicate are by that very fact finite. It would make no sense for us to consider any but this finite universe or, mayhap, finite segment of an infinite universe. A finite segment of space, no matter how large, is infinitesimal in comparison with infinity. The conclusion of endless duplications of life does not at all apply to this infinitesimally finite universe of our discourse. In *it*, as concluded in the next chapter, we are not about to talk to humanoids on other planets.

Copyright (c) 1964 by George
Gaylord Simpson. Reprinted
from his volume THIS VIEW OF
LIFE by permission of Harcourt
Brace Jovanovich, Inc.

The Nonprevalence of Humanoids

THE possibility that life exists elsewhere than on earth has excited human imagination since antiquity. In our own days it has become the principal basis for a whole school of writing: science fiction, which remains mere entertainment even though some of its devotees do make an unjustified claim that it should be taken more seriously. There has also long been discussion that was scientific, at least in the sense that it was by professional scientists who did not intend to write fiction. Even in the nineteenth century there was serious, if not invariably sober, discussion of the view that life exists not only elsewhere but even everywhere in the cosmos.

There is, then, nothing new in the fact that this subject is being widely discussed and publicized. What is new is that the usual speculation and philosophizing are now accompanied by extensive (and incidentally expensive) research programs, by concrete plans for exploration, and by development of pertinent instrumentation. Although the interested scientists have by no means stopped talking, they are now, and for the first time in history, also acting. Our major space agency, NASA, has a "space bioscience" program. Biologists meeting under the auspices of the National Academy of Sciences have agreed that their "first and . . . foremost [task in space science] is the search for extraterrestrial life" (Hess *et al.*, 1962). The existence of this movement is as familiar to the reader of the newspapers as to those of technical publications. There is even increasing recognition of a new science of extraterrestrial life, sometimes called *exobiology*—a curious development in view of the fact that this "science" has yet to demonstrate that its subject matter exists!

Another curious fact is that a large proportion of those now discussing this biological subject are not biologists. Even when biochemists and biophysicists are involved, the accent is usually on chemistry and physics and not on biology, strictly speaking. It would seem obvious that organic evolution has a crucial bearing on the subject, which is essentially a problem in evolutionary systematics. Surely, then, it is odd that evolutionary biologists and systematists have rarely been consulted and have volunteered little to the discussion. A possible reason for this blatant omission was suggested long ago by an evolutionary systematist, W. D. Matthew, who wrote that, "[Physical scientists] are accustomed to hold a more receptive attitude . . . toward hypotheses that can not be definitely disproved . . . [while] the [evolutionary and systematic] biologist . . . is compelled . . . to leave out of consideration all factors that have not something in the way of positive evidence for their existence."

Matthew also remarked that, "To admit the probability of extra-mundane life opens the way to all sorts of fascinating speculation in which a man of imaginative temperament may revel free from the checks and barriers of earthly realities." Both of his points are illustrated delightfully and without conscious humor by a contemporary leader in exobiology who wrote in 1962, "We do not really know [what the atmosphere of Venus is like], and we are thus not severely limited in our conclusions"! (exclamation point mine).

As an evolutionary biologist and systematist, I believe that we should make ourselves heard in this field. Since part of our role must be to point out "the checks and barriers of earthly realities," we may at times seem merely to be spoilsports, but we do have other contributions as well.

Exobiology has three major questions: "'What kind of life?" "Where?" "How may it have evolved?" Each question in turn involves two complex, distinct fields of inquiry. Confusion of these fields frequently distorts judgment and confuses argument.

The alternative fields as to the kind of life are "life as we know it" and "life as we do not know it." Life as we know it obviously cannot be confined in this context to actual terrestrial species, but implies only a more general similarity. It must, at least, involve a carbon chemistry reacting in aqueous media and with such fundamental organic compounds as amino acids, carbohydrates, purine-pyrimidine bases, fatty acids, and others. It must almost certainly also involve the combination and polymerization of those or similar fundamental molecules into such larger molecules or macromolecules as proteins, polysaccharides, nucleic acids, and lipids. Life as we do not know it might be based on some multivalent element other than carbon, on some medium (perhaps even solid or gaseous) other than liquid water, and then necessarily on quite different kinds of compounds.

If we did encounter such systems or organisms, we might well fail to recognize them as living or might have to revise our conception of what life is. Here on earth, in spite of a border zone between, and enormous diversities within, each realm, we can recognize two kinds of configurations of matter, one living and one not. (Under "configuration" I mean to include not only chemical composition but also organization or anatomy in the fullest sense and energy states and transactions.) "Life as we do not know it," if recognized at all, might have to be recognized as a third fundamental kind of configuration and not, strictly speaking, as life. There has been considerable speculation along such lines, some of it diverting in a science-fictional sort of way. Yet there is not a scrap of evidence that "life as we do not know it" actually exists or even that it *could* exist—evidence, for example, in the form of detailed specifications for a natural system that might exhibit attributes of life without the basis of life as we do know it. (Computers and other artifacts that mimic some features of the life of their makers are not really pertinent to this question.) Here, at least, further consideration will be given only to life as we know it, to the minimal extent of depending on similar biophysical and biochemical substrates.

The dichotomy in discussing the "Where?" of possible extraterrestrial life is between our own solar system and presumed similar planetary systems anywhere else in the universe. Much has been learned over the years about the planets of our system by earth-based astronomical methods. Recently rocketry and telemetry have given us closer looks at the moon and at Venus and promise to give us many additional facts. Human visits to the moon and the closer planets, at least, make no evident further demands on our theoretical knowledge and require only a reasonable extrapolation of our technical potentialities into the near future. Here, then, we have actual observational data to work with, and the promise of many more.

Not so for any planetary systems that may exist outside our own. Statements in both the scientific and the popular literature that there are millions of such systems suitable for life and probably inhabited may give the impression that we know that they do exist. In fact we know no such thing in any way acceptable as sober science. There are no direct observational data whatever. It is inherent in any acceptable definition of science that statements that cannot be checked by observation are not really *about* anything—or at the very least they are not science. As long as we do not confuse what we are saying with reality, there is no reason why we should not discuss what we hope or expect to observe, but it is all too easy to take conjecture and extrapolation too seriously. It is not impossible that our descendants may some day make pertinent direct observations on other planetary systems, but that is far beyond our present capabilities or any reasonable extrapolation from them. With our present techniques, the only way we could obtain direct knowledge of life outside our solar system would be by receiving signals from someone or something out there. That point is involved in the third question, the directly evolutionary one, and its two major fields of enquiry: the origin of life and its subsequent history. Here is my main topic, to which I will return at length.

First it is necessary to refer briefly to the environmental conditions and possible evidence of life on the only planets for which we have any actual data, the planets of our own solar system. Apart from a few eccentrics, astronomers have long since agreed that life as we know it is now quite impossible on any extraterrestrial body in our solar system except Venus and Mars. (See, e.g., the book by Jackson and Moore cited in the notes to this chapter.) Opinion regarding Venus has been divided, but telemetry from the recent Venus probe seems to confirm beyond doubt the previous view that Venus is far too hot for life as we know it (Barath *et al.*). Although somewhat equivocal, such evidence as we have on the composition of the Venusian atmosphere also seems to be unfavorable on balance (see, e.g., Sagan). It would appear, then, that Venus can now be ruled out as a possible abode of recognizable life.

The evidence for Mars is also highly equivocal, but it does not at present entirely exclude the possibility of life there. Temperatures are rigorous and there is little or no free oxygen. Obviously neither man nor any of our familiar animals and plants could possibly live in the open on Mars. Simple microorganisms have, however, been grown in conditions possibly similar to those that just might exist on Mars (Hawrylewicz, Gowdy, and Ehrlich). This possibility depends in part on the usual belief that the so-called ice caps of Mars are indeed composed of water and that the atmosphere is mainly nitrogen with some carbon dioxide. Both beliefs have been authoritatively challenged by Kiess, Karrer, and Kiess, who maintain that the caps are N_2O_4. That and the accompanying concentrations of oxides of nitrogen in the atmosphere would make Mars lethal to life as we know it. In any case, there is increasing doubt that enough water exists on Mars to sustain any form of life.

Direct evidence for life on Mars has also been claimed. The old idea that the so-called canals of Mars were made by intelligent beings no longer merits sober consideration. It is, however, well known that there are dark areas on Mars that show seasonal changes in position and in apparent color. It has been claimed repeatedly that these areas must be covered with some form of plant life, and that idea received significant support when it was discovered that their infrared spectrum has a band similar to that of some organic compounds (Sinton). However, similar absorption can also be caused by oxides of nitrogen and by a variety of inorganic carbonates (partly unpublished work cited by Calvin). The question remains open, and plans to make direct observations by space probe are going forward (see, e.g., Levin *et al.*). These plans depend on the further doubtful proposition that there may be microorganisms on Mars that can be grown by the same methods used here to grow microorganisms in laboratories.

The only other direct evidence for extraterrestrial life worthy of serious consideration is derived from meteorites. It has been claimed that some of these contain hydrocarbons of organic origin and even actual fossils of microorganisms (see the articles by Nagy, Meinschein, and Hennessy; and by Nagy, Claus, and Hennessy). If confirmed, these observations would indicate that life (now extinct) had occurred on a planet of our system that has since been disrupted. However, further investigation strongly suggests that the materials observed are in part inorganic and in the remaining part terrestrial contaminants (Anders and Fitch). The most favorable possible verdict is "Not proven."

There is, then, no clear evidence of life anywhere else in our solar system. Wishful thinking, to which scientists are not immune, has obviously played a part here. The possibility is not excluded, but on what real evidence we have the chance of finding life on other planets of our system is slim.

It bears repeating that there are no observational data whatever on the existence, still less on the possible environmental conditions, of planets suitable for life outside our solar system. Any judgment on this subject depends on extrapolations from what we know of the earth and its life and from astronomical data that do not include direct observation. There are, indeed, considerable grounds for such extrapolations, but they still contain a large subjective element and have a strong tendency to go over into sheer fantasy.

There are four successive probabilities to be judged: the probability that suitable planets do exist; the probability that life has arisen on them; the probability that such life has evolved in a predictable way; and the probability that such evolution would lead eventually to humanoids (as defined in the next paragraph). The thesis I shall now develop, admittedly subjective and speculative but extrapolated from evidence, is that the first probability is fair, the second far lower but appreciable, the third exceedingly small, and the fourth almost negligible. Each of these probabilities depends on that preceding it, so that they must be multiplied together to obtain the over-all probability of the final event, the emergence of humanoids. The product of these probabilities, each a fraction, is probably not significantly greater than zero.

(Before proceeding, I should define "humanoid" for those not as addicted as I am to science fiction. A humanoid, in science-fiction terminology adaptable to the present also somewhat fanciful subject, is a natural, living organism with intelligence comparable to man's in quantity and quality, hence with the possibility of rational communication with us. Its anatomy and indeed its means of communication are not defined as identical with ours. An android, on the other hand, is a nonliving machine, servomechanism, or robot constructed in more or less human external shape and capable of performing some manlike actions.)

The first point, as to the existence of earthlike planets, need not detain us long. The astronomers seem to be in complete agreement that planets that are or have been similar to the earth when life arose here probably exist in large numbers (see the works by Hoyle; Shapley; and Jackson and Moore). Indeed the number of stars in the accessible universe (discernible by light or radio telescopy) is so incredibly enormous that even if the chances of any one of them having such a planet were exceedingly small, the probability that *some* of them do would be considerable. As a basis for further consideration, we may, then, reasonably postulate that conditions such as proved propitious to the origin of life on earth may have existed also outside our solar system.

The next question is: How did life arise on earth, and is it probable or perhaps inevitable that it would arise elsewhere under similar conditions? This is largely in the field of the biochemists, and they certainly have not neglected it. The literature is enormous.

Enough of it for our purposes is summarized or cited in the recent works of Oparin, Florkin, Calvin, and Ehrensvärd. There are wide differences of opinion as to the particular course followed, but here again there is near unanimity on the essential points. Virtually all biochemists agree that life on earth arose spontaneously from nonliving matter and that it would almost inevitably arise on sufficiently similar young planets elsewhere.

That confidence is based on chemical experience. If atoms of hydrogen and oxygen come together under certain simple and common conditions of energy, they always deterministically combine to form water. Formation of more complex molecules requires correspondingly more complex concatenations of circumstances but is still deterministic in what seems to be a comparatively simple way. That has, indeed, been demonstrated in the laboratory. If energy such as would be available on a young planet is put into a mixture of the simplest possible compounds of hydrogen, oxygen, carbon, and nitrogen, such as also could well occur on a young planet, amino acids and other building blocks of the essential complex organic molecules are formed. The crucial experiment was that of Miller. A large amount of later work, mostly noted in the books cited above, has confirmed and extended those results. The further synthesis of the building blocks into the macromolecules, especially nucleic acids and proteins, essential for life has not yet been accomplished under realistically primitive conditions. Nevertheless it is reasonable to assume that those steps, too, would occur deterministically, inevitably, if given enough time under conditions likely to hold on some primitive planets. It is also clear that there has indeed been enough time, for the earth is now definitely known to be more than three billion years old, and planets still older could well exist in this and other galaxies.

It is still a far cry from the essential preliminary formation of proteins, nucleic acids, and other large organic molecules to their organization into a system alive in the full sense of the word. This is the step, or rather the great series of steps, about which we now know the least even by inference and extrapolation. A fully living system must be capable of energy conversion in such a way as to accumulate negentropy, that is, it must produce a less probable, less random organization of matter and must cause the increase of available energy in the local system rather than the decrease demanded in open systems by the second law of thermodynamics. It must also be capable of storing and replicating information, and the replicated information must eventually enter into the development of a new individual system like that from which it came. The living system must further be enclosed in such a way as to prevent dispersal of the interacting molecular structures and to permit negentropy accumulation. At the same time selective transfer of materials and energy in both directions between organism and environment must be possible. Systems evolving toward life must become cellular individuals bounded by membranes.

The simplest true organisms have all those characteristics and more, but they are very far from being simple in microscopic and submicroscopic organization. Less organized associations of organic macromolecules, such as are seen today in the viruses, cannot perform all those feats on their own and therefore cannot be meaningfully viewed as primitive and true forms of life.

If evolution is to occur and organisms are to progress and diversify, still more is necessary. Living things must be capable of acquiring new information, of alteration in their stored information, and of its combination into new but still integrated genetic systems. Indeed it now seems that these processes, summed up as mutation, recombination, and selection, must already be invoked in order to get from the stage of loose macromolecules to that of true organisms, or cellular systems. There must be some kind of feedback and encoding leading to increased and diversified adaptation of the nascent organisms to the available environments. Basically such adaptation is the ability to reproduce and to maintain or increase continuous populations of individuals by acquiring, converting, and organizing materials and energy available from existing environments. These processes of adaptation in populations are decidedly different in degree from any involved in the prior inorganic synthesis of macromolecules. They also seem to be quite different in kind, but that is partly a matter of definition and is also obscured by the fact that they must have arisen gradually on the basis of properties already present in the inorganic precursors. In any case, something new has definitely been added in these stages of the origin of life. It requires an attitude of hope if not of faith to assume that the acquisition of organic adaptability was deterministic or inevitable to the same degree or even in the same sense in which that was probably true of the preceding, more simply chemical origin of the necessary macromolecules.

By that I do not mean to say that material causality has been left behind or that some mysterious vitalistic element has been breathed into the evolving systems. All must still be proceeding without violation of physical and chemical principles. Those principles must, however, now be acting in different ways because they are involved in holistic, organic, increasingly complex, multimolecular systems that far transcend simple chemical bonding. It is here that one must stop taking for granted the expectations and extrapolations of the chemist and can obtain further enlightenment only from the biologist as such, that is, the student of whole organisms as they exist in reproducing populations and in communities adapted to environments.

Given ample time and rather simple circumstances not likely to be unique in the universe, there does seem to be considerable probability, perhaps even inevitability, in the progression from dissociated atoms to macromolecules. The further organization of those molecules into cellular life would seem, on the face of it, to have a far different, very much lower order of probability. It is not impossible, because we know it did happen at least once. Nevertheless that event is so improbable that even if macromolecules have arisen many times in many places, it would seem that evolution must frequently or usually have ended at that preorganismal stage. Only the astronomical assurance that there may be many millions of earthlike planets permits us to assume that the origin of true, that is of cellular, life may have happened more than once. In the observable universe the lowest recent estimate for earthlike planets by a competent astronomer is, as far as I know, that of Shapley, who considers 100 million a highly conservative figure. On that basis it is reasonable to speculate that life has arisen repeatedly in the universe, even though we do not know and perhaps will never know whether that is a fact.

Here brief consideration may be given to the idea that once life had arisen somewhere, organisms in a state of cryptobiosis (Keilin) might have spread by "cosmozoan" transport from one planet to another. That possibility was especially urged by Arrhenius, following the still earlier, curious speculation of Richter and others that life may be coextensive with the whole cosmos both in space and in time. It now appears extremely improbable but not quite impossible that any organism, even encapsulated and in a cryptobiotic state of entirely suspended metabolism, could survive the radiation hazards in space without artificial shielding (Becquerel). Furthermore, passage from one solar system to another at any speed attainable by natural means (e.g., by the pressure of light) would require vastly more time than any established or probable duration of the cryptobiotic state, which is not known to have lasted longer than about fifty years in microorganisms or about a thousand years in any organisms (Keilin). A conservative conclusion would be that it is extremely improbable, almost to the point of impossibility, that any form of life has ever traveled by natural means from one planetary system to another. Such travel between earth and Mars, within the same planetary system, is still improbable, but the possibility is not absolutely ruled out.

We now turn to the subsequent evolution of postulated life forms once life has appeared on a planet, and we again move to a different order of probability. We have only a single sample on which to base judgment. Paleobiologists have shown us the general course followed by evolution on this planet. Neobiologists have shown in great, although still incomplete, detail the outcome of that process at one point in time, the present. Although these are far from being the only accomplishments of systematists, they are in themselves so important for current problems as to justify intensified research on this enormous subject.

The problem of extrapolating from this unique sample is to decide whether it is inevitable, probable, improbable, or impossible for life of independent origin to have followed a similar or identical course. Opinions have indeed varied from one end to the other of that scale. I believe that a reasonable choice among those opinions is possible, and furthermore that many, even most, of those who have recently considered the subject have made a wrong choice. Review of recent literature on exobiology, almost all of it by physical scientists and biochemists (or molecular biologists), shows that most of them have *assumed*, usually without even raising the question, that once life arose anywhere its subsequent course would be much as it has been on earth. Now, the only really sound basis for such an assumption would be the opinion that the course followed by evolution on earth is its only possible course, that life cannot evolve in any other way. In a review of two books in which that assumption is made, Blum has called this the "deterministic" point of view as contrasted with an "opportunistic" one. The choice of terms is not a happy one, if only because it is demonstrable that evolution fully deterministic in the philosophical sense would not necessarily, indeed would almost surely not, follow similar courses on different planets. Nevertheless, the two schools of thought do exist and what Blum calls the deterministic one is more commonly followed in current exobiological speculations.

There are here underlying problems of philosophy and indeed also theology. Those problems have been discussed in previous chapters (especially 9, 10, and 11) and need little further attention here. The pertinent *scientific* questions are: If the processes of evolution are the same everywhere as they are here on earth, will they elsewhere lead to the same material results, including men or humanoids? Just how inevitable is that outcome?

Those questions can be followed up in two different but related ways. First, we can examine the course of evolution here on earth to see whether in fact it has proceeded as if directed toward a goal or an inevitable outcome. Second, we can investigate the mechanisms or processes of evolution in order to judge whether and under what conditions their outcome was limited to a course eventuating in some kind of humanoid, that is, in ourselves in the terrestrial example. Those approaches have also been discussed in previous pages (especially Chapters 4, 8, and 12) and need only summaries at this point.

The fossil record shows very clearly that there is no central line leading steadily, in a goal-directed way, from a protozoan to man. Instead there has been continual and extremely intricate branching, and whatever course we follow through the branches there are repeated changes both in the rate and in the direction of evolution. Man is the end of one ultimate twig. The housefly, the dog flea, the apple tree, and millions of other kinds of organisms are similarly the ends of others. Moreover, we do not find that life has simply expanded, branching into increasing diversity, until the organisms now living had evolved. On the contrary, the vast majority of earlier forms of life have become extinct without issue. Usually their places in the economy of nature have then been taken by other organisms of quite different origin. In some cases, their places seem simply to have remained empty for shorter or longer periods.

Neither in its over-all pattern nor in its intricate detail can that record be interpreted in any simply finalistic way. If evolution is God's plan of creation—a proposition that a scientist as such should neither affirm nor deny—then God is not a finalist. But this still does not fully answer the particular question we are pursuing here. The whole nonfinalistic pattern *might* have been followed nearly enough on a planet of some other star to produce humanoids there also. We must turn then to the causal elements and limitations inherent in the process for further judgment of the probability of such an outcome.

Each new organism develops in accordance with a figurative message, coded information, received from its one or two parents. Evolution occurs only if there are changes in that information in the course of generations. Such changes in individuals occur for the most part in two ways, although each takes numerous and sometimes complicated forms: mutations, which introduce new elements into the message, and recombinations, which put these elements into new associations and sequences. In a stricter sense mutations are any changes within the code carried by a nucleic acid. Recombinations involve rearrangements of the various code units and particularly new associations of units from different sources. The latter sources of variation are sexual, and sexlike processes occur in even the most primitive living organisms although they have been secondarily lost in a relatively small number of both plants and animals.

In themselves, these processes are not adaptive; they have no direct relevance to fitting organisms into the economy of nature, permitting their survival and further evolution. Since most (but not all) evolutionary changes are adaptive and progressive evolution does occur, these processes alone cannot be the whole story. They are necessary for evolution, but something else must also be involved. There must be some interaction between organisms and environment and from this there must be some kind of feedback into the genetic code. The feedback is by natural selection and it occurs in populations through successive generations, not in individuals in their lifetimes. That is the whole point of natural selection: that it does feed back from environment to genetic code in such a way as to maintain or change the message in adaptive ways. It does this because, by and large, the better adapted organisms have more offspring. The more adaptive genetic messages thus tend to spread through the population in the course of generations. Also, in more complex ways that I need not go into here, new code combinations adaptive for the population as a whole are thus brought into being.

This feedback is basic for our present enquiry because it places definite limitations on the possible course of evolution. We can be quite sure that if the environments of their ancestors had been very different from what they were, the organisms of today would also be very different. It is also clear that evolution must be opportunistic in the sense that it can work only with what is there. Mutations can occur only in quite definite ways depending on the existing nature of the coded message. Recombination can recombine only the code elements that do exist in given organisms. Selection can work only on variations actually present in a population. The cause of evolution thus includes all the genetic, structural, physiological, and behavioral states of populations right back to the origin of life.

Even slight changes in earlier parts of the history would have profound cumulative effects on all descendant organisms through the succeeding millions of generations. In spite of the enormous diversity of life, with many millions of species through the years, it represents only a minute fraction of the possible forms of life. The existing species would surely have been different if the start had been different and if any stage of the histories of organisms and their environments had been different. Thus the existence of our present species depends on a very precise sequence of causative events through some two billion years or more. Man cannot be an exception to this rule. If the causal chain had been different, *Homo sapiens* would not exist. (These causal limitations were discussed in more detail in the preceding chapter.)

Both the course followed by evolution and its processes clearly show that evolution is not repeatable. No species or any larger group has ever evolved, or can ever evolve, twice. Dinosaurs are gone forever. Nothing very like them occurred before them or will occur after them. That is so not only because of the action of selection through long chains of nonrepetitive circumstances, as I have just briefly noted. It is also true because in addition to those adaptive circumstances there is a more or less random element in evolution involved in mutation and recombination, which are stochastic, technically speaking. Repetition is virtually impossible for nonrandom actions of selection on what is there in populations. It becomes still less probable when one considers that duplication of what are, in a manner of speaking, accidents is also required. This essential nonrepeatability of evolution on earth obviously has a decisive bearing on the chances that it has been repeated or closely paralleled on any other planet.

The assumption, so freely made by astronomers, physicists, and some biochemists, that once life gets started anywhere, humanoids will eventually and inevitably appear is plainly false. The chance of duplicating man on any other planet is the same as the chance that the planet and its organisms have had a history identical in all essentials with that of the earth through some billions of years. Let us grant the unsubstantiated claim of millions or billions of possible planetary abodes of life; the chances of such historical duplication are still vanishingly small.

Even if, as I believe, any close approximation of *Homo sapiens* elsewhere in the accessible universe is effectively ruled out, the question is not quite closed. Manlike intelligence is, after all, a marvelous adaptation, especially in its breadth. It has survival value in a wide range of environmental conditions, and therefore, if it became possible at all, might be favored by natural selection even under conditions different from those on earth. There is, to be sure, another serious hitch here. Man may be going to use one wild aspect of his intelligence to wipe himself out. I do not believe that will occur, but no realist can now deny it as a possibility. If it did happen, the adaptiveness of human intelligence would have been short-lived indeed, and the argument from its apparent broad adaptiveness would be negatived.

Apart from that point, is there not some play, so to speak, in the causations of history? Even in planetary histories different from ours might not some quite different and yet comparably intelligent beings—humanoids in a broader sense—have evolved? Obviously these are questions that cannot be answered categorically. I can only express an opinion. Evolution is indeed a deterministic process to a high degree. The factors that have determined the appearance of man have been so extremely special, so very long continued, so incredibly intricate that I have been able hardly to hint at them here. Indeed they are far from all being known, and everything we learn seems to make them even more appallingly unique. If human origins were indeed inevitable under the precise conditions of our actual history, that makes the more nearly impossible such an occurrence anywhere else. I therefore think it extremely unlikely that anything enough like us for real communication of thought exists anywhere in our accessible universe.

"Extremely unlikely" is not "impossible," and those who like to dream may still dream that mankind is not alone in the universe. But here another point comes up to trouble us. What is the nature and value of that dream? Unless we know or can seriously hope to learn in fact of other humanoids, the dream remains a dream, a fantasy, a science-fiction *divertissement*, a poetic consolation with no substance of reality. Suppose the near-impossible were to be true. What are the chances that we could in fact learn of the existence of extraterrestrial humanoids and eventually communicate with them? With a feeling almost of sorrow, I must conclude that the chances are vanishingly small.

In the present or any foreseeable state of our technology, the only way we could learn of other humanoids would be by their sending us a message or actually visiting us. That requires, in the first instance, that they must have developed manlike technology, which by no means follows automatically from the mere development of intelligence. (They *might* be intelligent enough to use their brains in better ways!) They must also have done so at just the right time, which involves another tricky point. Out of the billions of years of life on earth, there has been only an infinitesimal length of time, some sixty years, since man has been in a position either to send or to receive messages through outer space. How small the chance of coincidence that any other humanoid reached just this stage at just the right time!

Theoretically, the improbability of humanoids becomes a little less if we consider regions farther out in space and therefore include more stars in our calculations. If humanoids were on a planet a million light years away—and that is a very small distance in the vastness of the galaxies—a message to reach us now would have had to be sent precisely a million years ago. Improbability piled on improbability approaches impossibility. If again the apparently impossible happened, it would certainly be one of the most exciting events in history, but to what avail? The senders of the message would obviously be dead when we received it; their whole species might well be extinct. If, finally stretching the barest possibility to the utmost, we received a message from the relatively nearby stars, it would take years or more likely generations to send a message and receive a reply. Under those conditions the establishment of useful, intelligible intercommunication would still be impossible.

An actual visit to earth by extraterrestrial humanoids would require a technology extremely far advanced beyond ours. We do not, at present, even know that such a stage of technology is possible. All the difficulties previously noted, and more, here pile up. If such a feat is remotely possible and if humanoids are at all prevalent in the universe—the if's do tend to pile up, too, in this subject!—then one would think that we would have been visited by now. In spite of reports of flying saucers and little green men, which belong only in science fiction, the fact is that none have visited us. That would seem, indeed, a logical added reason to believe that humanoids are, to say the least, nonprevalent.

I cannot share the euphoria current among so many, even among certain biologists (some of them now ex-biologists converted to exobiologists). The reasons for my pessimism are given here only in barest suggestion. They will not, I know, convince all or indeed many. There are too many emotional factors and, to put it bluntly, selfish interests opposed to these conclusions. In fact I myself would like to be proved wrong, but a rational view of the evidence seems now to make the following conclusions logically inescapable:

1. There are certainly no humanoids elsewhere in our solar system.
2. There is probably no extraterrestrial life in our solar system, but the possibility is not wholly excluded as regards Mars.
3. There probably are forms of life on other planetary systems somewhere in the universe, but if so it is unlikely that we can learn anything whatever about them, even as to the bare fact of their real existence.
4. It is extremely improbable that such forms of life include humanoids, and apparently as near impossible as does not matter that we could ever communicate with them in a meaningful and useful way if they did exist.

I shall close this chapter with a plea. We are now spending billions of dollars a year and an enormously disproportionate part of our badly needed engineering and scientific manpower on space programs. The prospective discovery of extraterrestrial life is advanced as one of the major reasons, or excuses, for this. Let us face the fact that this is a gamble at the most adverse odds in history. Then if we want to go on gambling, we will at least recognize that what we are doing resembles a wild spree more than a sober scientific program.

To some it seems that the reward could be so great that facing any odds whatever is justified. The biological reward, if any, would be a little more knowledge of life. But we already have life, known, real, and present right here in ourselves and all around us. We are only beginning to understand it. We can learn more from it than from any number of hypothetical Martian microbes. We can, indeed, learn more about possible extraterrestrial life by studying the systematics and evolution of earthly organisms. Knowledge from enlarged programs in those fields is not a gamble because profit is sure.

My plea then is simply this: that we invest just a bit more of our money and manpower, say one-tenth of that now being gambled on the expanding space program, for this sure profit.

[redacted]

Gould, S.J.
"Seti & the wisdom of Casey
Stengel"
Discover
(c) Time, Inc.
Used by permission

A BIOLOGIST'S VIEW

THE WISDOM OF CASEY STENGEL

by STEPHEN JAY GOULD

Since the study of extraterrestrial life lacks any proven subject, opinions about the form and frequency of nonearthly beings record the hopes and fears of speculating scientists more than the constraints of evidence. Alfred Russel Wallace, for example, Darwin's partner in the discovery of natural selection and the first great evolutionist to consider exobiology in any detail, held firmly that man must be alone in the entire cosmos—for he could not bear the thought that human intelligence had not been the uniquely special gift of God, conferred upon an ideally suited planet. He wrote in 1903 that the existence of abundant and brainy extraterrestrials "would imply that man is an animal and nothing more, is of no importance in the universe, needed no great preparations for his advent, only, perhaps, a second-rate demon, and a third- or fourth-rate earth."

The endless debate about extraterrestrial life has focused upon the calculation of probabilities—how many stars, how many suitable planets, the chance that life will originate on appropriate earths, the probability that life will eventually generate intelligence. I must confess that I have always viewed this literature as dreary and inconclusive, too mixed up with hope and uncertainty to reach any respectable conclusions. But now, a different kind of direct and truly experimental search has been launched with the SETI program.

As an evolutionary biologist, I have no expert knowledge in most areas motivating the debate about SETI. I am moved to comment only because some opponents of SETI have used an argument from my field as one of their most powerful weapons. They state that all leading evolutionary biologists have proclaimed the existence of extraterrestrial life to be nearly inconceivable. The optimism of some physical scientists therefore resides in their failure to understand the distinctive character of evolutionary reasoning. I believe that opponents of SETI have misstated the biological argument, and I would like to explain why at least one evolutionary biologist thinks that SETI is a long shot well worth trying.

In a lengthy series of strongly worded articles, Frank Tipler has cited "two basic reasons for my disbelief in the existence of extraterrestrial-intelligent beings." The second reason, that any intelligent extraterrestrials, if they existed, would build von Neumann machines [see page 56] that would already be here, lies outside my field and I shall not dwell on it, though it must be mentioned. I don't mean to be a philistine, but I must confess that I simply don't know how to react to such arguments. I have enough trouble predicting the plans and reactions of those humans closest to me. I am usually baffled by

the thoughts and accomplishments of human beings in different cultures. I'll be damned if I can state with certainty what some extraterrestrial source of intelligence might do. Tipler's second argument follows the same speculative tradition that SETI, with its experimental approach, is designed to transcend.

But, as his first argument, Tipler features a different kind of claim, based on the methods and data of my field. He writes:

> First, all the great contemporary experts in the theory of evolution—Francisco Ayala, Theodosius Dobzhansky, Ernst Mayr, and George Simpson—are unanimous in claiming that the evolution of an intelligent species from simple one-celled organisms is so improbable that we are likely to be the only intelligent species ever to exist.

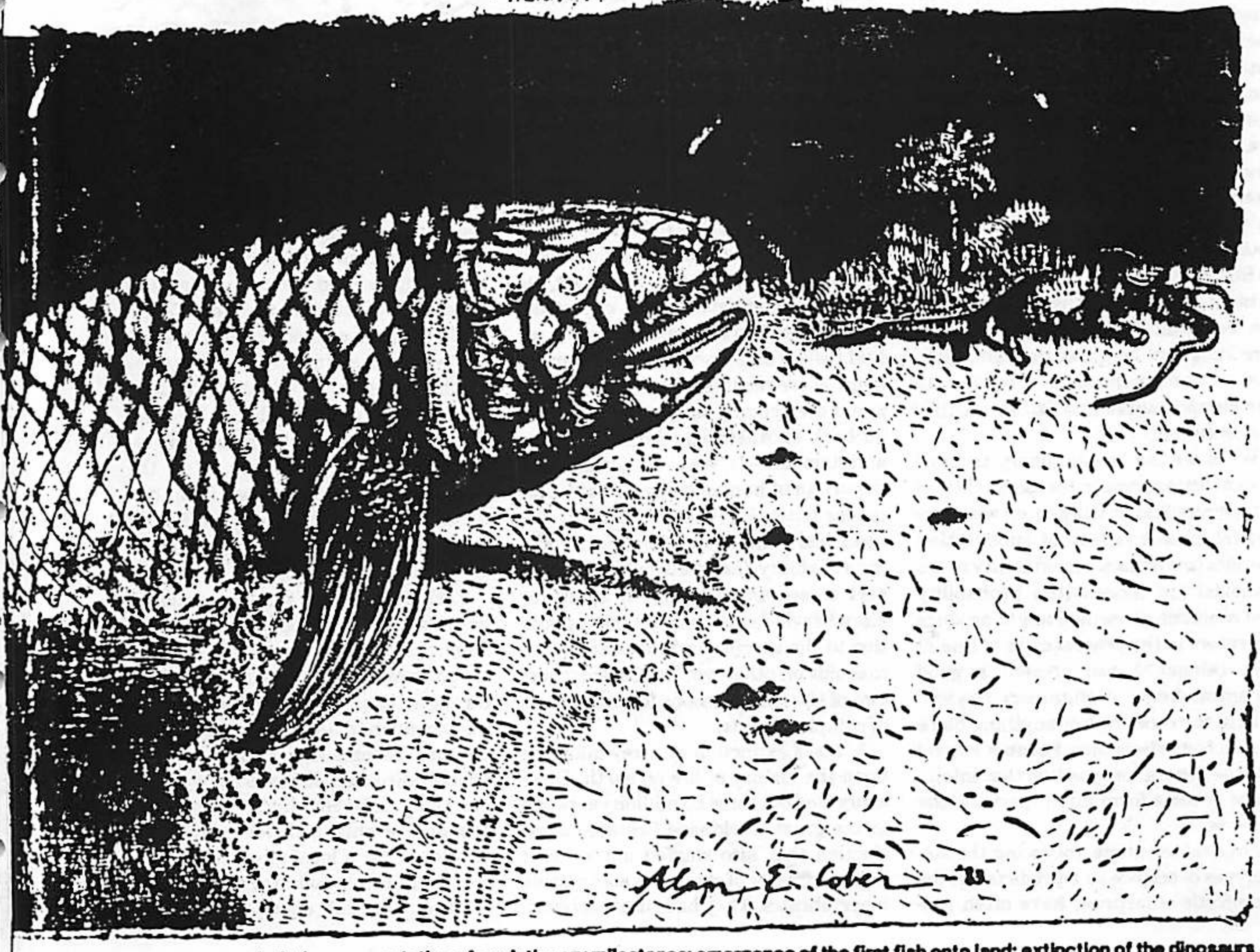
On the most mundane level, if I may play the nasty "expert game" for just one sentence, Tipler's statement is empirically false. I count at least four quite respectable evolutionists in the international pro-SETI petition released by Carl Sagan this October (Tom Eisner of Cornell, Dave Raup of the University of Chicago, Ed Wilson of Harvard, and, with apologies for arrogance, yours truly). Evolutionary biologists, in their usual consistency with nature's primary theme, maintain a *diversity* of views on this subject.

More important, I think that Tipler, by conflating two very different issues, has misunderstood what evolutionary biologists dismiss with such passion. All evolutionists who have discussed exobiology at length have clearly delineated two separate concerns—a specific claim and a general argument.

The *specific* issue treats the *detailed repeatability* of any particular evolutionary sequence—in this case, the evolution of creatures looking pretty much like us: bilaterally symmetrical with sense organs up front, two eyes, a nose in the middle, a mouth, and a brain. If we could start the earth's tape anew, would intelligent creatures again evolve in this form? If other worlds share our basic chemistry and conditions, would such "humanoids" evolve on them?

The *general* question asks whether attributes that we would identify as intelligence might arise in creatures of any conformation—blobs, films, spheres of pulsating energy, or diffuse and unimagined forms far beyond the limited visions of most science fiction writers.

All evolutionists have vociferously denied the specific claim, and I join them in all their vehemence. Many evolutionists have also gone on to doubt the general argument, but never with such



Artist's representation of evolutionary milestones: emergence of the first fish onto land; extinction of the dinosaurs

certainty—and always as a personal opinion, not as a proclamation bearing the indelible imprimatur of "evolutionary theory." I am among those evolutionists who deny the specific claim but feel that no strong opinion can be entertained about the general argument. SETI needs only the general argument to be worth doing.

Gregory Bateson, the recently deceased guru of sciences that deal with complex objects and interacting systems, continually emphasized that confusion of hierarchical categories may be the most common and serious fallacy of human reasoning. Bateson identified as primary among such "category confusions" the substitution of individuals for classes (or vice versa).

Casey Stengel, one of the greatest general gurus of our time, consciously committed Bateson's fallacy of categories to avoid the heat of scrutiny in a tough moment. He was roundly criticized for blowing the Mets' first pick in the expansion draft on a *particular* catcher of exceedingly modest ability (one Hobie Landrith by name). Casey answered by invoking the *class* of catchers in general: "You have to have a catcher, because if you don't, you're likely to have a lot of passed balls." Now Ol' Case, as usual, knew exactly what he was saying (never let the patter known as Stengelese fool you). He used humor to blunt criticism because he knew that we would all recognize the fallacy of reasoning and laugh at the conflation. But we commit the same error in subtler circumstances and fail to identify our confusion.

When we use "evolutionary theory" to deny categorically the possibility of extraterrestrial intelligence, we commit the classic fallacy of substituting specifics (*individual* repeatability of humanoids) for classes (the probability that evolution elsewhere might produce a creature in the *general* class of intelligent beings). I can present a good argument from "evolutionary theory" against the repetition of anything like a human body elsewhere; I cannot extend it to the general proposition that intelligence in some form might pervade the universe.

Physical scientists, following the stereotypes of science as a predictable, deterministic enterprise, have often reasoned that if human beings arose on earth, then we must infer (since cause leads inexorably to effect) that intelligent creatures of roughly human form would arise on any planet beginning with physical and chemical conditions similar to those that prevailed on the early earth. Perhaps this deterministic outlook is responsible for the paltry imagination of film makers and science fiction writers, with their endless creatures, all designed on a human model. This tendency could be forgiven when human actors had to play roles in our movies, but now that pieces of plastic can evoke our deepest emotions and move so subtly that E. T. becomes a national hero, the excuse no longer holds.

Styles of doing science are as diverse as its subject matter. Classical determinism and complete predictability may prevail for simple macroscopic objects subject to a few basic laws of motion (balls rolling down inclined planes in high school physics experiments), but living beings and other complex historical objects do not lend themselves to such easy treatment. In the history of life, all results are end products of long series of events, each so intricately dependent upon particular environments and previous histories that we cannot predict their future course. The historical sciences try to explain unique occurrences—immensely complex historical accidents. Evolutionary biologists, as historical scientists, do not expect detailed repetition and cannot use the actual results of history to establish probabilities for recurrence (would a Caesar again die brutally in Rome if we could go back to *Australopithecus* in Africa and start anew?). We evolutionists view the origin of human beings (or any particular butterfly, roach, or starfish) as a historical event of such complexity and improbability that we would never expect to see anything like it again (or elsewhere)—hence our strong opposition to the *specific* argument about humanoids on other worlds. Consider just two of the many reasons for uniqueness of complex events.

1. Mass extinction as a key influence upon the history of life on earth. Dinosaurs died out some 65 million years ago in the great worldwide Cretaceous extinction that also snuffed out at least half the species of shallow water marine invertebrates. They had ruled terrestrial environments for 100 million years and would probably reign still if they had survived the debacle. Mammals arose at about the same time and spent their first 100 million years as small creatures inhabiting the nooks and crannies of a dinosaur's world. If the death of dinosaurs had not provided their great opportunity, mammals would still be small and insignificant creatures. We would not be here, and, in all probability, no consciously intelligent life would grace our earth. Evidence gathered during the past two years indicates that the impact of an asteroid, or some other extraterrestrial body, engendered (or at least greatly accelerated) this extinction. What could be more unpredictable and unexpected than an asteroid striking the earth literally out of the blue? Yet without it, earth would probably lack consciously intelligent life. Five great extinctions (several larger than the Cretaceous event) have set basic patterns in the history of life, imparting an essential randomness to our evolutionary pageant.

Birds, insects, and bats took different

2. Each species as a concatenation of improbabilities. Any animal species, human, squid, or coral, is the end result of an evolutionary chain stretching through thousands of species back to the inception of life. If any of these species had become extinct or evolved in another direction, final results would be markedly different. Each chain of improbable events includes adaptations developed for a local environment and only fortuitously suited to support later changes. Our fish ancestors evolved a

fin with a sturdy and central bony axis. Without such a structure, landbound descendants could not have supported themselves in a nonbuoyant terrestrial environment. (Most lineages of fishes did not and could not evolve terrestrial descendants, because they lacked fins of this form.) Yet these fins did not evolve in anticipation of future terrestrial needs. They developed as adaptations to a local environment in water, and were luckily suited to permit a new, terrestrial direction later on. All evolutionary sequences include such a set of *sine quibus non*, a lucky series of accidents with respect to future evolutionary success. Human brains and bodies did not evolve up a direct and inevitable ladder but by a circuitous and tortuous route carved by a set of adaptations, evolved for different reasons, and fortunately suited to later changes.

evolutionary roads to the same end—flight

Thus the improbabilities of history proclaim that each species is unique and unrepeatable in detail. Evolutionary theory, as a science of history, does deny the specific argument for humanoids on other worlds. All leading evolutionists in their writings on exobiology have said so with gusto, and I agree. But all these evolutionists have also clearly distinguished this specific proposition about humanoids from the general argument that intelligence in some other form might arise elsewhere. On the general proposition, they have maintained a diversity of opinion—leading to the empirical conclusion that "evolutionary theory" has no clear pronouncement to make.

George Gaylord Simpson, for example, a leading detractor of the specific argument, also denies the general claim, but ever so much more gently, and as a personal opinion only:

> Even in planetary histories different from ours might not some quite different and yet comparably intelligent beings ... have evolved? Obviously these are questions that cannot be answered categorically. I can only express an opinion ... I think it extremely unlikely that anything enough like us for real communication of thought exists anywhere in our accessible universe.

Other evolutionists, including two of Tipler's experts who supposedly deny any possibility for SETI's success, also distinguish the specific from the general argument, but express more optimism for the generality. As Ayala and Dobzhansky wrote:

> Granting that the possibility of obtaining a man-like creature is vanishingly small even given an astronomical number of attempts ... there is still some small possibility that another intelligent species has arisen, one that is capable of achieving a technological civilization.

I am not convinced that the possibility is so small.

Does evolutionary theory offer any insight about the general argument? We gain some sense of probabilities for repetition of a general theme (but not of specific details) from the phenomenon known as convergence. Flight has evolved separately in insects, birds, pterosaurs (flying reptiles), and bats. Aerodynamic principles are similar, but morphologies differ widely (birds use feathers; bats and pterosaurs employ a membrane, but bats stretch it between several fingers, pterosaurs from only one). Marsupial "moles" and "wolves" evolved in Australia, a continent isolated from placental mammals elsewhere. Since adaptive themes are limited and animals so diverse, convergences of different evolutionary lineages to the same general solutions (but not to detailed repetition) are common. Highly adaptive forms that are easy to evolve arise again and again. More complex morphologies without adaptive drive offer little or no prospect for repetition. Conscious intelligence has evolved only once on earth, and has no real prospect for repetition should we choose to use our gift for destruction. But does intelligence lie within the class of phenomena too complex and historically conditioned for repetition? I do not think that its uniqueness on earth demands such a conclusion. Perhaps, in another form on another world, intelligence would be as easy to evolve as flight on ours.

Even if Tipler is right in claiming that von Neumann machines are the only way to go, he admits that we won't have the technology to build one for a century. But I'm an impatient and mortal fellow. As I think it cruel to ask disadvantaged minorities to "go slow" in demands for political change—thus guaranteeing that any practical benefits will fall only upon their children or children's children—so too do I selfishly wish to see some exobiological results (positive or negative) in my lifetime.

SETI is all we have for now. It is relatively cheap, and entirely sensible, it seems to me, from those perspectives that evolutionary theory can enlighten. Frankly, I think the chances of its success are somewhat lower than the probabilities envisioned by its more enthusiastic supporters among physical scientists. But we can't know until we try. Ultimately, however, I must justify the attempt at such a long shot simply by stating that a positive result would be the most cataclysmic event in our entire intellectual history. Curiosity impels, and makes us human. Might it impel others as well?

Goldsmith, D.
The Quest for Extraterrestrial Life pp. 74-75, 224-227, 241-242, 250-254
University Science Books
Used by permission

Our sun's family of nine planets contains just one on which life is known to exist, leaving eight candidates for speculation. Of these eight, by far the most intriguing to mankind has been Mars, the next planet out from the sun, whose rust-red color has suggested the god of war to many civilizations that admired it from Earth.

The possibility of life on Mars seemed to receive a tremendous increase, if not downright confirmation, from the announcement of *canali,* an Italian word meaning "channels" or "canals," made by Giovanni Schiaparelli more than a century ago. Schiaparelli made no statements about the nature of these apparent straight lines on the Martian surface, but the American astronomer Percival Lowell, possessed of a fortune large enough to build his own observatory in Flagstaff, Arizona, and a self-confidence to match, proclaimed that his observations showed an indisputable network of water-carrying channels, covering the planet except for the polar caps, from which they presumably carried irrigation water as the caps melted.

Lowell was wrong, and Schiaparelli was wrong: the Martian canals are optical illusions, straight lines produced by the human eye and brain, which links together small, faint markings in straight patterns. But Lowell's impact was tremendous, coming at a time when human interest in life on other planets was growing. Lowell's "social Darwinism,' his conclusions about how Martian society must have evolved under the harsh conditions of that planet's environment, fed straight into the mainstream of his era's political thought. Lowell's ability as a popular writer (he also wrote interesting books about his travels in China) helped his "discoveries" attain worldwide recognition. In England, H. G. Wells responded by writing one of his most famous science-fiction novels, *War of the Worlds* (published in 1897), which described an invasion of Earth by intelligent Martians, who proved unable in the end to survive terrestrial microbes.

Radio searches for signals from Mars were made as early as 1922, but in the absence of success, and with the onset of the Great Depression and the Second World War, support for such efforts faded. But the popular belief in life on other worlds showed itself vividly on Halloween night, 1938, when Orson Welles led a radio dramatization of H. G. Wells's novel, telling in a quasi-realistic fashion of a Martian invasion of New Jersey. Thousands of listeners throughout the East panicked: many of them fled their homes for supposedly safer regions of the country. Reassuring news soon greeted them, and many were later reluctant to admit the full extent of their participation in the evacuation.

By the time that the eminent British astronomer Sir James Jeans reviewed the search for extraterrestrial life in 1942, Lowell's influence had been replaced by a realization that the canals probably did not exist. Jeans concluded that none of the inner planets, Mercury, Venus, and Mars, offered a likely site for our sort of life, and he dismissed the possibility of life on the giant planets, Jupiter, Saturn, Uranus, and Neptune, because their outer layers were known to be tremendously cold.

By sending space probes to five of Earth's eight sister planets, and by studying them with better equipment from Earth's surface, we have found during the past decade that the generally negative outlook for life elsewhere in the solar system appears correct. However, we have discovered that the giant planets—Jupiter in particular—grow warmer as one descends into their gaseous envelopes, so that it would be premature to conclude, as Jeans did, that life cannot exist there simply because of the temperature problem. Temperature does indeed rule out life on Mercury (which alternately bakes and freezes), on the moon (for the same reason), and on Venus, which stifles under a blanket of carbon-dioxide atmosphere that keeps the planet's surface above 700° C. But the possibility of life on the giant planets seems more threatened—though by no means eliminated—through the lack of oxygen or a solid surface on which molecules can collect.

To search for life on Mars, we spent a billion dollars to send the two Viking orbiters and landers to the red planet in 1976. As Norman Horowitz describes, we found tantalizing indications, but on balance no clear evidence of life. When we manage to send probes that will enter the atmospheres of Jupiter and the other giant planets, we may find equally fascinating data, perhaps equally indecisive, on the outer members of the solar system.

41

Alfred Adler
Behold the Stars

In September, 1971, the first international conference on extraterrestrial civilizations was held in Soviet Armenia. Jointly organized by the US and USSR Academies of Sciences, it brought together eminent physicists (such as James Dyson), astronomers, biologists (Francis Crick), engineers, and some social scientists. Both the importance of the subject and the scientific credentials of the participants invited attention and respect, even a certain amount of hopeful anticipation. It is, after all, a reasonable hypothesis that life may exist in other parts of the universe, though nothing is yet understood of where it resides, or of how to approach it, or of what forms it might take. To those of us who believe that man's future resides in the whole universe rather than on earth alone, the conference held the promise of revelation and inspiration. Yet such confidence was totally misplaced. For, the mountain labored, and brought forth a

> CONFERENCE RESOLUTION
> *List of Possible Research Directions*
> It would be useful to concentrate efforts in two directions, both of which seem promising:
> I. Searches for civilizations at a technical level comparable with our own.
> II. Searches for civilizations at a technological level greatly surpassing our own.
> A wide circle of specialists, from astrophysicists to historians, should participate in the planning of this research.

But this was only the last of a collection of blows designed to make those who love science weep, and those who do not, laugh.

The full account is available in an instructive book called *Communication With Extraterrestrial Intelligence (CETI)* (MIT Press, $10.00). The primary occupation of the conference was explained by Carl Sagan, a young astronomer and an organizer of the event. Sagan proposed a formula for the number *N* of other civilizations within our galaxy:

$$N = R_* f_p n_e f_l f_i f_c L.$$

Since the formula appears scientific (mathematical), but is in fact totally nonscientific and even meaningless, and since the formula lies at the heart of almost all the work and interest and significance of the conference, I will have to try the reader's patience by identifying each of its terms. The definitions are all quotations from the book.

> R_* is the rate of star formation averaged over the lifetime of the Galaxy, in units of numbers of stars per year . . . the province of astrophysics.
> f_p is the fraction of stars which have planetary systems.
> n_e is the mean number of planets within such planetary systems which are ecologically suitable for life . . . n_e is determined at the boundary between astronomy and biology.
> f_l is the fraction of such planets on which life actually occurs.
> f_i is the fraction of such planets on which, after the origin of life, intelligence in some form arises.
> f_c is the fraction of such planets in which the intelligent beings develop to a communicative phase . . . a topic in anthropology, archaeology, and history.
> L is the mean lifetime of such technical civilizations . . . it involves psychology and psychopathology, history, politics, sociology, and many other fields.

And how are these numbers to be determined? The answers, fortunately, are provided by the distinguished conferees.

> [FRANCIS] CRICK: The point that I am making adds up to the following conclusion: It is not possible at the moment, with our knowledge of biochemistry, to make any reasonable estimate whatsoever of the factor f_l . . . until we have further information, we cannot really guess about the matter.
> [L.M.] MUKHIN: I do not quite understand how we can estimate f_i . . . when we cannot choose any rational approach for assessing f_i. I think it is correct to say that a reliable estimate cannot be given.
> [LESLIE] ORGEL: It is our opinion, as a result of experimental

---

Reprinted from the *Atlantic Monthly*, 234 (October 1974), 109.

> and theoretical work on the subject, that our science has not yet progressed to the point where a meaningful estimate [of f_i] can be made.
>
> [CARL] SAGAN: We are faced . . . with very difficult problems of extrapolating . . . in the case of L, from no examples at all.
>
> [R.B.] LEE: The first tool [for determining f_c] is the modern synthetic theory of evolution. . . . The second tool is the theory and method of historical materialism . . . pioneered by Marx and Engels. . . . The third tool is the commitment shared by most of us to search for the broadest, most comprehensive generalizations that can be drawn from available facts.

The "second tool," with its lunatic assertion that historical materialism has direct bearing on the nature of f_c (the fraction of planets in which intelligence develops to a communicative phase), causes one almost to forget that the conference was dealing with questions of extraterrestrial life. The "third tool" is, if possible, an even greater intellectual pollutant. And finally, the major substantive purpose of the conference, a determination of estimates for the number N, is quite clearly a total fraud.

And yet: *"Mon derrière est divisé en deux parts,"* said Gladstone of his own past, thus providing a fair description of the conference as well. A (lesser) part was indeed devoted to useful pursuits, to questions of message transmission, message reception, message decoding, and, finally, to the possible consequences of contact with other civilizations. On the whole these were technical exchanges of theory and experiment, with two expositions of particular general interest. One, by Philip Morrison, speculated on the possible impact on Earth of contact with extraterrestrial civilizations. The other, by Dyson, was a compassionate essay on the human dilemma, beginning with a discussion of Bernal's three enemies of man's rational nature (the cruelty of nature, the frailty of physical man, and the power of man's irrational components), and closing with an invocation which almost totally defied the spirit and content of the conference that preceded it:

> If we are wise, we shall preserve intact these qualities of the human species [toughness, courage, unselfishness, foresight, common sense, and good humor] through the centuries to come, and they will see us safely through the many crises of destiny that surely await us.

A moderately intelligent and humane note. Almost at the conference's end, yet precisely the point from which the conference should have proceeded. For the human qualities most displayed by the conferees were of another nature: those of cupidity, inanity, and triviality:

Cupidity: "Their [the components of the number N] *only* value is in assessing how much effort, time, and money we are willing to devote to the problem. . . . If it turns out that there is some rigorous argument to exclude extraterrestrial intelligence, a convincing demonstration of a small value of N, then a search would not be a useful allocation of resources." (Sagan)

We have already seen that a demonstration of any value of N, let alone of a small value of N by a rigorous argument of any kind, is out of the question. The disclaimer, in short, is an exercise in cynicism.

Inanity: "I believe that the state of society characteristic of Polynesia is one that we can look forward to in [all] the industrial nations. . . . And if their brains resembled ours . . . then all those advanced [extraterrestrial] societies would make up a galactic Polynesian archipelago." (Gunther Stent)

Trivia: "The strategy [in our quest for other civilizations] should be organized roughly as follows: First, the energy for transmitting one bit of information should be minimized. Second, interference in the vicinity of the sending side should be minimized. Third, the cost of the receiving apparatus should be minimized. Fourth, a signal-to-noise ratio greater than unity . . . is desirable. And lastly, the point I should particularly like to stress, is that the ultimate time must be minimized." (N.S. Kardashev).

Is there absolutely nothing that these pedants could not simply take for granted?

It is almost incredible that the truly distinguished scientists among the conferees (and there were indeed several of these) could be willing, almost eager, participants in a travesty of all that is taken seriously by men and women who love and value science and intellect. Possibly some of the scientists were not quite so willing as they appeared to be. Dyson in particular was terse and belligerent almost from beginning to end. But that does not explain the behavior of those who were indeed willing participants. Plato would have understood them well:

> In a state of democratic anarchy, the master fears and flatters his scholars . . . the old men condescend to the young and are full of pleasantry and gaiety; they are loath to be thought morose and authoritative, and therefore they adopt the manners of the young. . . . *(The Republic)*

So now, who are these young? Fortunately, they are not hard to find, nor shy or secretive. Carl Sagan, for one, has written a short, quite personal book of essays in which he has disclosed much more of himself and of his peers than perhaps even he or the reader would like. Called *The Cosmic Connection* (Anchor/Doubleday, $7.95), it is part autobiography, part personal philosophy and speculation, part self-aggrandizement, direct and indirect. (Sagan's wife, who provided an aesthetically offensive drawing for an intellectually offensive metal plaque sent off by NASA to greet extraterrestrial beings, is said to have created art "based on the classical models of Greek sculpture and the drawings of Leonardo da Vinci.") This book has much to teach about the nature of the technological mind.

The modern technologist is first of all a promoter. "It [the subject of extraterrestrial life] has now reached a practical stage where it can be pursued by rigorous scientific techniques where it has achieved scientific respectability," writes Sagan, fully aware of the total absurdity of every part of his assertion, and having in fact participated in a conference which exhibited the emperor's nakedness for even the most obtuse to see. The technologist, too, is gracefully immodest about his own accomplishments: "The greatest significance of the... plaque is not as a message to out there; it is as a message to back here." He is the vacant and specious authority on intellect: "The deflation of some of our more common conceits is one of the practical applications of astronomy." Whose conceits? What conceits? Today, in the second half of the twentieth century? His wit is arch and flat (chapter headings such as "Hello, Central Casting? Send Me Twenty Extraterrestrials," or "Some of My Best Friends Are Dolphins," or "The Cosmic Cheshire Cats," are a numskull's delight). He believes himself to be Renaissance Man, with a profound understanding of man's creative sources: "As the results of space exploration... permeate our society, they must, I believe, have consequences in literature and poetry, in the visual arts and music"—a platitude that in fact provides another perfect illustration of Sherlock Holmes's dog that did not bark. And he apparently regards himself as a master of politics and economics:

> Old economic assumptions, old methods of determining political leaders... may once have been valid... but today may no longer have survival value at all.... At the same time, there are vested interests opposed to change. These include individuals... who are unable in middle years to change the attitudes inculcated in their youth.

The modern technologist is a gifted, highly trained, opportunistic, humorless, and unimaginative ass. Not a barbarian, certainly, and not to be feared; but not to be flattered, pampered, or praised, either. None of his fatuous pseudo-science is science; all of it is empty of intellectual content, inflated with self-importance, and held accountable for nothing. He charges through subtleties and profundities where wise men hesitate to walk on tiptoe; he usurps domains about which he knows nothing and then proceeds to pre-empt them. He tells us that the way to roast a pig is first to find a cow, and expects to be rewarded, applauded, and honored for this genial advice.

The fundamental issues of a search for extraterrestrial civilizations are after all not so arcane or so inaccessible to the nonscientist. Reasonable hypotheses can be advanced and a few promising steps taken, even though at present, since we do not really know where we are going, almost any road will get us there. Already something is understood of how best to attempt the reception of messages from space, of how best to attempt to decode them, of where to seek them. These methods are highly technical, but not unacceptably expensive. Modest conjectures about what to expect of these efforts, and what some of their consequences might be, are also in order. Even though there is no a priori certainty that there exist any civilizations at all in the universe beyond Earth, it is reasonable to postulate that other civilizations exist, even that they exist in large numbers, and it does no harm, in any case, to proceed from such assumptions.

If, in fact there exist civilizations in the galaxy that have attained a level of technology far beyond ours, and if there also exist those whose technologies are approximately equivalent to ours, then it is also reasonable to conjecture that the second of these will contact us long before the first. Civilizations whose development stands to ours as does ours to the grasshopper's will either ignore us entirely or else observe us without bothering to communicate with us in any way. We would not understand most of what they had to tell us in any case. Thus the cultural shock of the first intragalactic messages to earth might well be far less severe than we now expect. We could not even be certain that any message from space would be a call from a living civilization. The speed of light (186,000 miles per second) is a universal upper limit for the speed of messages through space, so that signals would reach us years, or centuries, or more likely yet, millennia after their transmission, possibly arriving here long after the death of the civilization that had sent them.

Morrison, in one of the conference's few bright spots, proposed the thesis that contrary to all superficial expectations, contact with extraterrestrial civilizations would in time have a slow and soberly meditated impact on Earth, much like that of the Western world on nineteenth-century Japan, rather than an overpowering, even cataclysmic effect. Morrison's argument revolves around the observation that other civilizations would be so far removed from us, both in space and (because of the great distance any message must travel) in time, that there could be no question of military dominance or technical economic competition.

A similar argument suggests that the step from reception of the first message to a two-way communication would be a long and difficult one. Each of a series of questions and replies would take (at best) a prohibitive amount of time to travel from sender to receiver, the interval between successive messages spanning many human generations. A reasonable corollary to this conclusion is the suggestion that the matter of communication with other civilizations is not now, and will not be in the future, an urgent one, requiring vast expenditures and intensive exploratory programs. The whole subject, though of great interest and importance, can be approached in patient, measured fashion. It might well, in fact, be postponed entirely for another century (although there is no objective reason for such a postponement).

Human technology has only recently approached a level at which it can begin to liberate man from some of his terrestrial constraints; and even this accomplishment, if regarded from a slightly different perspective, indicates that human technology is still so much more primitive than that of civilizations capable today of contacting us that we could barely understand the other civilization's messages, or benefit from their advice.

The reception of a galactic message would be significant most of all because it would replace a human expectation with a certainty. It would prove to humanity that we are not alone in the universe, and in the proving might diminish man's self-consciousness and self-centeredness. The consequences could only be beneficial. And that is about all the profit that can be expected. Something tangible might indeed be learned from an extraterrestrial civilization, but this would have to be within the realm of our present knowledge and capabilities, or else we could not comprehend its meaning. Being in this realm, it would be something we would sooner or later have discovered for ourselves, without outside help. It would slightly accelerate our progress, no more. Any scientist, any rational person, would reach conclusions more or less like these if ever he decided to think seriously about the matter.

And yet most of the participants at the conference on extraterrestrial life, playing by the rules of their patrons, their new technologists, chose to avoid such rational discourse. Rational, civilized man appears to have become very tired, no longer able to withstand the onslaughts of the manic young masters who promote large grants of money and influence, leave him breathless at conferences, lavish the currency of vague new ideas upon all those around them, and exhaust their weakened elders with pure, assured, unself-conscious power. The behavior of scientists is but one small indication of this syndrome. The fact, for example, that our entire culture turns to M.D.'s for an understanding of the meanings and possibilities of conjugal and erotic love (M.D.'s! Is this to be believed?) suggests that our whole humanist and intellectual tradition has become tired too. It knows better, but goes along all the same, paying obeisance to its new masters in the spirit of the Arab proverb: If the King at noonday says it is night, behold the stars.

◆

45

John A. Ball

The Zoo Hypothesis

The most interesting scientific problem of our age involves the question of the existence of extraterrestrial intelligent life. Arguments summarized below make it likely that intelligence exists on many planets throughout our galaxy and that most of these civilizations are much older than our own. This problem has been the subject of considerable work both theoretical and experimental (see Oparin and Fesenkov, 1960; Cameron, 1963a; Shklovskii and Sagan, 1966; Sagan, 1973a; and other references therein) and our understanding of the subject has certainly progressed rapidly in the last decade or so. However, this problem has proved to be extremely difficult, in part because it involves understanding what a civilization much older than ours might be like. It is difficult enough to predict our own development for a few decades hence, but we need to know about other civilizations that may be older than ours not by decades but by eons.

Among currently popular ideas about extraterrestrial intelligence, the idea that "they" are trying to talk to us has many adherents (see, e.g., Drake, Reading 24). This idea seems to me to be unlikely to be correct and the zoo hypothesis is in fact the antithesis of this idea.

Starting Premises

Three working hypotheses or starting premises are used in most discussions of the problem of extraterrestrial life. These premises are stated below with a discussion of their origin and references to the literature. Although this discussion is brief, these premises are in fact crucial and if any of them proves to be incorrect, then the zoo hypothesis falls.

A. Whenever the conditions are such that life can exist and evolve, it will. Life is to be understood as a chemical reaction that occurs whenever the necessary reactants are present under the appropriate conditions for a sufficient time. This statement represents a considerable extrapolation of our present knowledge. In fact the opposite hypothesis, that life is statistically unlikely even in ideal conditions, has been expressed (e.g., by Townes, 1971). Discovery of primitive life on Mars or Venus would probably settle this question. Our current understanding of biochemistry seems to support premise A (Shklovskii and Sagan, 1966, Chapter 14; and Calvin, 1961).

B. There are many places where life can exist. Planets are probably quite common in the universe. As many as 20% of all stars may have planets and as many as 10% of these planets may have surfaces on which life can form. (However Oparin and Fesenkov, 1960, think that only one star in 10^5 or 10^6 has a planet with a surface suitable for life. See also von Hoerner, 1961). This statement also represents more than we know at present; no star other than our sun is definitely known to have planets comparable to the earth. Objects that may be planets have been detected around a few other stars (see Shklovskii and Sagan, 1966, Chapter 11; Huang, 1959b; and van de Kamp, 1969b), however these objects are much more massive than the earth. Planets comparable to the earth around almost any other star would go undetected with present techniques. The opposite hypothesis, that the solar system is unique, was believed by Jeans (1929, Chapter XVI), but is now discredited (see, e.g., Levin, 1964, for a summary of current thinking).

C. We are unaware of "them."

Who is Out There?

It is statistically unlikely that there exists anywhere in our whole galaxy any other civilization whose level of development is at all comparable to ours. We would expect to find either primitive life forms, perhaps comparable to those on the earth a few million years ago, or very advanced life forms, perhaps comparable to what will be on earth a few million years hence(!)

There are three general categories of possibilities defin-

Reprinted from *Icarus*, 19 (1973), 347. Copyright © 1973 by Academic Press, Inc.

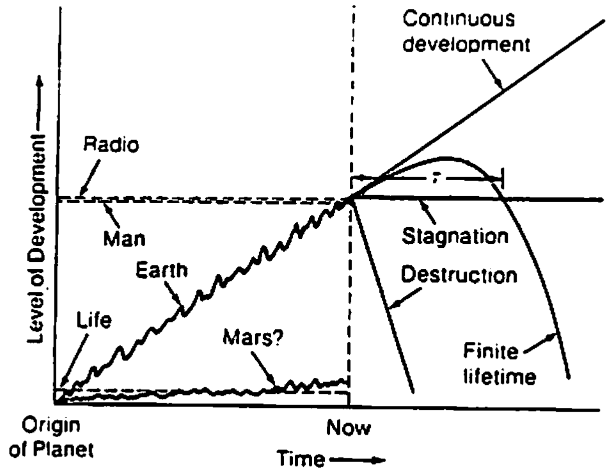


Figure 45.1. This is a sketch of the top level of development, defined in terms of complexity, versatility, and ability to control the environment, either of the organism itself or of the civilization to which it belongs. The various possible extrapolations for our future are discussed in the text.

ing the technological evolution of a civilization:

(1) Destruction (from within or without).
(2) Technological stagnation.
(3) Quasi-continuous technological progress.

Also there are many other mixed possibilities, such as partial destruction and rebuilding, and the surprisingly popular finite-lifetime idea. These possibilities are sketched diagramatically in Figure 45.1 with specific reference to our own extrapolated future. It is likely that some fraction of all civilizations follow each of these possibilities. However, analogy with civilizations on earth indicates that most of those civilizations that are behind in technological development would eventually be engulfed and destroyed, tamed, or perhaps assimilated. So, generally speaking, we need consider only the most technologically advanced civilizations because they will be, in some sense, in control of the universe.

Technological progress may be defined as increasing ability to control one's environment. Already at our level of technology we affect almost everything on earth from elephants to viruses. But we do not always exert the power we possess. Occasionally we set aside wilderness areas, wildlife sanctuaries, or zoos in which other species (or other civilizations) are allowed to develop naturally, i.e., interacting very little with man. The perfect zoo (or wilderness area or sanctuary) would be one in which the fauna inside do not interact with, and are unaware of, their zookeepers.

The Zoo Hypothesis

Premise C above now seems to me to be extremely significant. I believe that the only way we can understand the apparent non-interaction between "them" and us is to hypothesize that they are deliberately avoiding interaction and that they have set aside the area in which we live as a zoo.

The zoo hypothesis predicts that we shall never find them because they do not want to be found and they have the technological ability to insure this. Thus this hypothesis is falsifiable, but not, in principle, confirmable by future observations.

Conclusions

The zoo hypothesis as given here is probably flawed and incomplete. I hope that it can provide some sort of inspiration for further work. Among other hypotheses that one might consider, the laboratory hypothesis is one of the more morbid and grotesque. We may be in an artificial laboratory situation. However, this hypothesis is outside the purview of science because it leads nowhere, it immediately calls into question the premises on which it is based, and it makes no predictions. Or one might suppose that extraterrestrial civilizations have not yet found us or that they know we are here but they are uninterested in us. These latter two hypotheses are probably incompatible with the high level of technological sophistication they undoubtedly possess.

The zoo hypothesis seems to me to be pessimistic and psychologically unpleasant. It would be more pleasant to believe that they want to talk with us, or that they would want to talk with us if they knew that we are here. However the history of science contains numerous examples of psychologically unpleasant hypotheses that turned out to be correct.

Acknowledgments

Although the ideas in this paper are not in the mainstream of current scientific thought about the problem, they are also not new. Science-fiction authors, in particular have toyed with similar notions for many years. And at least a few previous writers have suggested such ideas as a serious possibility.

I thank Sebastian von Hoerner and Mrs. Lyle G. Boyd for pointing out relevant background material and for stimulating discussions. I am grateful to Prof. A. E. Lilley for his encouragement.

48

Sebastian von Hoerner

Where Is Everybody?

How frequently may intelligence and technology have developed on the planets of other far-away stars, and what are our chances of detecting any radio signals from these alien and mostly much older races? I seriously think that trying to establish contact with other beings in the universe is our next great task, and that success would mean the largest step in the evolution of mankind since our development of speech about a million years ago. And meanwhile, as a fringe benefit, meditating about life in space may give us a helpful distance and a better perspective for looking at our own terrestrial affairs and problems.

On a small scale, we have already started SETI, the *S*earch for *E*xtraterrestrial *I*ntelligence; and very large future SETI projects are being prepared and could be undertaken. But what chances do we have? Previous estimates have mostly given quite optimistic results, but during the last years some rather pessimistic arguments have turned up, which are quite puzzling and hard to beat. In the following, I will emphasize this pessimistic puzzle, hoping that someone else will come up with a convincing optimistic solution.

Assuming Us to be Typical

First, I will give a short summary of the previous estimates for SETI. The only thing we can go by, the only case of life we know, is our own life here on Earth. We should not assume to be anything special. Our Sun is just an average star, one out of $200 \cdot 10^9$ which make up our stellar system, the Galaxy. And from our theories of star formation (von Hoerner, 1975b) we think that planets are quite common, formed together with their central star from a contracting cloud of gas and dust. Sun and Earth are about $5 \cdot 10^9$ years old.

We know that our Sun is not an old star, and that most stars are about $5 \cdot 10^9$ years older because star formation was about 30 times more productive at the beginning than it is now. The very first stars will have had no solid Earth-like planets, because the original big-bang of the universe provided only hydrogen and helium, while all the heavy elements needed for solid planets are produced by stellar evolution and supernova explosions. But still we expect a very large number of habitable planets similar to Earth, most of them being about $4 \cdot 10^9$ years older than Earth.

Regarding life, we should assume similar developments under similar conditions. Which means that life and intelligence have developed wherever a planet similar enough to our Earth was going in the right orbit at the proper distance around a star, similar to our Sun and old enough. Many scientists have estimated how frequently it happens that all conditions for life are fulfilled (Cameron, 1963a; Sagan, 1973a; Bracewell, 1974; Dole, 1970). Within a factor of ten, say, the results mostly state that about 0.5% of all stars should have a planet where life similar to our own could be expected. Our Galaxy then would have about 10^9 of such habitable planets, and our nearest neighbors would be about 20 light years away. If our estimate of the percentage is wrong by a factor of ten, then the number of planets is wrong by a factor of ten, but the distance is only wrong by a factor of two ($10^{1/3} = 2.15$). This then would be the distance we had to travel in future spaceships, in search for habitable planets and for higher life.

If we want to communicate by some technical means, however, for example by radio waves, then our partner must be in a comparable state of mind, in order to use similar means and to understand each other. We should not assume that our present state of mind is the final goal of all evolution; it will be surpassed by completely different interests and activities which we cannot guess at our present state. Unfortunately, we have nothing to go by for estimating the duration of a "comparable technical state," because we are mere beginners. Using a wild guess of a 100,000 years, say, this would yield about 10,000 such

---

Reprinted from *Naturwissenschaften*, 65 (1978), 553. Copyright © 1978 by *Die Naturwissenschaften*.

civilizations in our Galaxy, and the nearest partners would be almost 1000 light years away from us. If our guess is wrong by a factor of ten, then the distance is only wrong by a factor of two.

What do these distances practically mean? Regarding radio contact, we could assume, for example, the same technical effort on both sides. If we as well as our partner would use equipment like our existing telescope at Arecibo in Puerto Rico, with its present best transmitter and receiver, we could already talk to a partner 40 light years away (provided we had agreed, before that, on the exact frequency and a narrow bandwidth, and knowing already at which star to point the telescope and when). And even a distance of 1000 light years could be bridged if we design a transmitter and receiver for this very purpose, which might take about three years and a million dollars.

Of course, guessing the right frequency channel, being lucky with the right time and finding a partner is a lot more difficult and time-consuming. Either we must assume that strong radio beacons are aimed at us, or we must build a large array of many big telescopes.

The distances also mean a very long waiting time for answers. No signal can go faster than light, and if our partner is 1000 light years away, then the answer to any question will take 2000 years. Not individuals, only whole civilizations can talk to each other. A two-way communication, with questions and answers, is only possible if the general time-scale of development is longer than the waiting time, so that there still is interest in the answer to a question asked so long ago. Our own hectic development is a lot faster than that, which would leave only a one-way communication (being given a wealth of general information without having asked for it). But this should never be underestimated: our whole western culture has been strongly influenced by the ancient Greeks from whom we got nothing but a few books and pieces of art in a one-way communication. Actually, all the traditions and cultural values we inherit from our ancestors are one-way communications. And our interstellar partners will be tremendously far ahead of us since we are mere beginners, and most stars are older than our Sun.

Regarding space travel, nothing can move faster than light (except in science fiction). From basic physical considerations one can show (von Hoerner, Reading 37) that space travel will probably be limited to $1/10$ the speed of light, $v \leq c/10$. Relativistic time effects then play no role. With our presently known technology and a good financial effort (Dyson, 1968) we could already achieve $v = c/100$ or maybe $c/30$. Thus, even a trip to the nearest habitable planets would take at least 200 years, with $v = c/10$ and a distance of 20 lightyears, much longer than the lifetime of a crew. Signals then look a lot more promising than travel, and radio waves between 1 cm and 30 cm wavelength seem most economical because there the background noise is smallest.

These promising estimates have led to several SETI searches for radio signals from outer space, beginning with Frank Drake's "Project Ozma" in 1960 at Green Bank (Cameron, 1963a). Meanwhile, about a dozen of searches have been done, with better equipment and new techniques, but all with existing equipment and no extra cost except the telescope time. Some hundred nearby solar-type stars have been occasionally searched, some star clusters and even a few other galaxies. So far, without any success.

The opposite SETI approach has been taken since 1970 by a group at Ames Research Center of NASA (and at JPL at Pasadena), called "Project Cyclops" (Oliver and Billingham, 1972): to find out what kind of equipment we should develop and build in order to have a really good chance of success within, say, 30 years of a dedicated search. The answer is a growing array of up to 1500 telescopes, 100 meter diameter each, with special receivers and correlators (doing normal radio astronomy about $1/10$ of the time). One would start with one such telescope and gradually build up more, searching all the time, and would stop building up whenever success is achieved. The total cost of all 1500 telescopes and receivers would be about 10^{10} dollars. This sounds ridiculously expensive; on the other side, this is just three months of what we actually did pay for the war in Vietnam.

If We Were Typical, We Should Not Exist

Second, I will describe a puzzling line of thought. The basic assumption of our previous estimates, that we on Earth are about average and nothing special, seems to lead to a serious contradiction. So far, we have considered and generalized only our previous and present activities. But what about our future ones? What will we probably be doing in the next few hundred years? And to what consequences does it lead if we generalize that again, assuming a similar development for other civilizations in space?

We have good reasons to believe that our space exploration will continue and expand. There is our great general curiosity, the drive to explore and to use our near and far surroundings. In addition, our planet Earth is getting unpleasantly crowded, and we are using up many of our limited resources at an alarming speed. Within the next few generations we must develop complete recycling systems for many metals and rare elements, and we must develop new energy sources. The next logical step then is to set up space colonies for mining (O'Neill, Reading 57; 1977), in large shelters or underground, on Moon and Mars and even on the many thousands of asteroids (little planets, big chunks of minerals and metals, orbiting around our Sun between Mars and Jupiter). These colonies will become self-supporting and multiplying, they will grow with their own babies and grandchildren. In

September I was at the 28th International Astronautical Congress at Prague, where many such plans were presented and discussed. A wealth of general problems, and even a lot of solid engineering tasks, have already been worked out in an amazing detail. We actually could start all this activity right now, if we wanted it and if we would be willing to spend the money. Sure, mining the asteroids is at present not economical, but that will change when our resources on Earth are getting used up.

Another future aspect must be mentioned. At present we live with a frightening arms race which has completely gotten out of hand, and which cannot continue for too long. Either we learn to be more reasonable, or we will blow each other to pieces. All the big nations spend $^1/_{10}$ of their gross national product on the development and manufacture of more and more powerful weapons, which was a worldwide total of $334 \cdot 10^9$ dollars per year in 1976. And nobody wants to be left behind: in 1976, the underdeveloped countries received $13 \cdot 10^9$ dollars for economical help, but bought weapons for $18 \cdot 10^9$ dollars, from the industrialized countries (Sampson, 1977). And the total destructive power of all nuclear bombs was, in 1972, about 40,000 megatons of TNT, which is the same as 10 tons of dynamite for each living person on Earth (von Hoerner, 1975a). Just try to imagine a solid round ball of dynamite with 2 meter diameter, one for each person.

Our chances for blowing up are really quite large. But in case we don't, if we somehow manage to stop the arms race, then a very large amount of money, a large work force of labor and engineering, must be redirected to other activities, and space engineering may come in quite handy and naturally. But especially if we do *not* stop the arms race, we should very soon build large space colonies: for the survival of the human race and culture in case of severe catastrophes on Earth.

Let us assume that after a hundred years or so, if we survive, we have many large self-supporting colonies in space, with people living there for generations, in shelters or underground; maintaining large-scale industries of mining, manufacturing, engineering, and exploration. After a few generations the cultural and emotional ties to the home planet will become less important. It seems to be the next logical step to send out large colonies (in mobile homes) on interstellar trips lasting many generations, in order to explore and colonize other planetary systems. If we do not destroy ourselves, we will probably do all this. Life, in whatever form we know it, has a strong tendency to expand and to fill out every possible niche, into all its corners. Plants, animals and man have settled in hot and dry deserts as well as close to the cold ice-covered poles.

Space travel becomes easy (even for our beginner's technology) if we drop the prejudice that it must be finished within an individual's lifetime. After we have colonized the nearby planetary systems, these settlements will grow, and after a while some of them will send out their own colonies in mobile homes to the next systems, and so on (just as hundreds of islands spread out over thousands of miles of the Pacific Ocean have been colonized long ago by Polynesians in their small primitive boats). We will have started a continuous wave of growing colonization, spreading out with about $^1/_{100}$ the speed of light. And in this way we can colonize our whole Galaxy, from one end to the other, in less than ten million years (Jones, 1976). This is only a very short time in astronomical terms. Even in biological terms it is not too long. The higher mammals are about 20 million years old; and 400 million years ago the first plants and animals moved out of the oceans and started conquering the continents of our Earth which took about 200 million years. Furthermore, ten million years would be needed for our present technology, and it might be reasonable to expect that a further-developed future technology could colonize the whole Galaxy within one million years, going with $^1/_{10}$ the speed of light.

Now let us generalize again; let us assume that we on Earth are about average and that similar developments are to be expected on similar planets, and that such planets are provided by 0.5% of all stars which means there are 10^9 such planets in our Galaxy (these are the basic assumptions for our previous and future SETI projects).

Assuming similar developments on these planets then leads to the following conclusion (Hart, Reading 42; Papagiannis, 1977): All what we just have described as our own probable future, large-scale exploration and colonization, all this should have happened long ago. It could have been started by any one out of 10^9 planetary civilizations $4 \cdot 10^9$ years ago, and it would have been finished only $1–10 \cdot 10^6$ years after it started, colonizing the whole Galaxy with everyone of its habitable planets. The whole Galaxy should be teeming with life, so obvious that there is no question about it. And, first of all: we humans should be the descendants of long-ago settlers from somewhere else. Which we certainly are not.

Furthermore it seems clear, from all we know, that all life on Earth had a common origin, and has developed here without any outside interference. All humans, animals and plants use the same basic organic chemistry, the same amino acids and even the same complicated genetic code. And the long chain of development from simple life forms to highly organized ones seems well enough described by mutations and natural selection. As to the origin of life (Sagan, 1977), we know from lab experiments that many large organic molecules, even amino acids and nucleotide bases and sugars (the most important building blocks of living matter), are formed abundantly from water and the gases of the original Earth's atmosphere, if energy is provided by ultraviolet radiation of the Sun or by electric discharges of lightning. We still do not know how the first self-reproducing organisms formed, or the genetic code, but the further development seems more or less clear. Also the development of human intelligence seems to follow quite naturally, since it provides such a wonderfully large and manyfold niche.

Not only our Earth, our whole planetary system appears uninhabited by any outside settlers. Otherwise, their large-scale mining industry or active radio communication would have been very obvious to us, the moment we invented optical and radio telescopes, since all this would be so nearby. There just are no extraterrestrials here.

We then are forced to the sad conclusion that our basic assumption was wrong, the assumption that we are not unique but are typical for the general development, and that life, intelligence and technology have developed in a similar way at a very large number of similar planets. But if all that is wrong, how can we ever expect to detect radio signals from other beings in space? If life is not abundant, and if we are not typical, then there is no chance of success for any SETI project.

We have shown that if our own development and attitude were typical, then the Earth should have been colonized long ago, there should be outside settlers here instead of us homegrown humans. With other words: "If we were typical, we should not exist." Which is such a revolting contradiction that we just cannot be typical. But then SETI has no chance.

The Large-Number Problem

How could we explain the absence of extraterrestrials in our system, or in general the absence of any obvious evidence of higher technology in our Galaxy? Because if even our beginner's technology is able to travel large distances and to send out strong radio signals, then the technology on old planets, after billions of years of further development, should surely be able to do all kinds of "astro-engineering" (Dyson, 1966), plainly visible and obvious, none of which we observe.

Where is everybody? This question has puzzled many scientists. Maybe the origin of self-reproducing life is so improbable that it almost never occurs. Maybe intelligence is so dangerous that it always ends up in self-destruction. Or, if this and other crises are to be mastered, any surviving civilization must be so highly regimented and stabilized that complete stagnation excludes any further technical progress (von Hoerner, 1975a). Maybe not a single one of 10^9 old civilizations felt the desire to colonize the Galaxy, or, first attempts of space colonization and astro-engineering have always met disastrous failures and have since been given up. Another nice thought is the "Zoo Hypothesis": that we live in a region of the Galaxy which has been declared a zoo or wildlife area, and a perfect zoo is one in which the animals don't see their waiters and spectators (Bova, 1963; Ball, Reading 45). Maybe the answer is "change of interest", meaning that science and technology are only a very short-lived phase of any long development to be surpassed by completely different future activities, which we cannot guess and which do not produce any obvious evidence (von Hoerner, 1975a). Well, most of these explanations would not give SETI much chance, and, anyway, none of them sounds very convincing and plausible.

The main difficulty is what I like to call the "Large-Number Problem". We should expect 10^9 habitable planets, with higher life $4 \cdot 10^9$ years old, and both are very large numbers. In order to explain the absence of extraterrestrial evidence, either habitable planets must occur only once in 10^9 cases, or life originates only once on 10^9 planets within $4 \cdot 10^9$ years; or intelligence destroys itself or stagnates without any exception in 10^9 cases, or poachers in wildlife areas are less frequent than one in 10^9 during $4 \cdot 10^9$ years, or not one in 10^9 governments granted funds for space mining, and so on. Any one of these explanations would sound much more acceptable if we had to consider only a small number of cases, but not so for 10^9 of them.

One way of avoiding the large-number problem would be to assume that during these billions of years, with lots of interstellar communications between the old civilizations, all these many original cultures have completely merged into one single galactic superculture which now acts as a whole (ant hill or bee hive), or at least with only a small number of subcultures. This would make several explanations possible, and it might give us some chances for SETI. But it works only if the galactic travel time of communication, 30,000 years from one half to the other for radio signals, were always much shorter than the timescale of individual planetary development, which does not sound plausible. It also leaves an unpleasant afterthought: what about all these newcomers, like us, about one per 100 years in the Galaxy? Those of the old establishment would have to investigate them, early enough and carefully, for finding out whether they can be educated into useful future members, or whether they are to be treated as dangerous bacteria. Hard luck for us, probably.

Conclusions

The absence of extraterrestrials in our solar system seems to show that there are none anywhere else in the Galaxy either (at least no "technical" ones). Even if we have no explanation for our amazing uniqueness, it still would follow that our searches for signals cannot have success.

If we want to continue our searches, and especially if we want to invest large sums, we should try to find a scenario which favors interstellar communication but prevents interstellar travel and colonization. Some suggestions have been made, but none so far looks plausible. In the absence of such a plausible scenario, it might be suggested to continue our searches with moderate means, in case the argument given above is wrong or incomplete, but to postpone any great expenditure until we can justify it again. This is the negative aspect I wanted to emphasize. However, different conclusions are also possible and should be discussed.

Even if one agrees that the absence of extraterrestrials seems to speak against SETI, one still could be in favor of spending larger sums for future SETI projects. First, on the general ground that one should never give up something so important before really having tried with all possible means; and especially so because our beginner's ignorance may have used assumptions and drawn conclusions which are all completely wrong. Second, not only would success be of tremendous importance for the human development, also the lack of success (after a dedicated effort) may have great impact, because it would be "quite a responsibility to know that we are the torch bearers of the flame of cosmic consciousness in our entire Galaxy" (Papagiannis, 1977). Third, in a discussion it was mentioned that Columbus started out for an impossible goal under wrong assumptions, but he still discovered America. Fourth, Phil Morrison suggested in a letter that interstellar colonization in its initial state may start as an organized process where the distance covered goes in proportion with time, but later on it would proceed (because of cultural and informational fragmentation) only as a diffusion or random walk, where the distance goes only with the square root of time, in which case there was still not enough time to cover the whole Galaxy and to reach us.

Finally, it seems to me that the Large-Number argument could be applied in the opposite direction, too; there are no extraterrestrials in our solar system because the colonizers have by chance overlooked, or by some reason neglected, at least one in 10^9 of habitable planets and useful mining sites. In this case there are no great odds against our own uniqueness, because our question, "Where is Everybody," *can* have been asked only at one of these few neglected places. Regarding SETI we then should not expect any strong contacting signals pointed at us, but we still may consider things like omnidirectional navigational beacons, or eavesdropping on someone else's local broadcast.

Any further suggestions?

6 June 1980, Volume 208, Number 4448 **SCIENCE**

Alvarez, L.W., Alvarez, W.,
Asaro, F., and Michel, H.V.
"Extra-terrestrial cause for
the Cretaceous-Tertiary Ex-
tinction"
Science
Published by the A.A.A.S
Used by permission

Extraterrestrial Cause for the Cretaceous-Tertiary Extinction

Experimental results and theoretical interpretation

Luis W. Alvarez, Walter Alvarez, Frank Asaro, Helen V. Michel

Summary. Platinum metals are depleted in the earth's crust relative to their cosmic abundance; concentrations of these elements in deep-sea sediments may thus indicate influxes of extraterrestrial material. Deep-sea limestones exposed in Italy, Denmark, and New Zealand show iridium increases of about 30, 160, and 20 times, respectively, above the background level at precisely the time of the Cretaceous-Tertiary extinctions, 65 million years ago. Reasons are given to indicate that this iridium is of extraterrestrial origin, but did not come from a nearby supernova. A hypothesis is suggested which accounts for the extinctions and the iridium observations. Impact of a large earth-crossing asteroid would inject about 60 times the object's mass into the atmosphere as pulverized rock; a fraction of this dust would stay in the stratosphere for several years and be distributed worldwide. The resulting darkness would suppress photosynthesis, and the expected biological consequences match quite closely the extinctions observed in the paleontological record. One prediction of this hypothesis has been verified: the chemical composition of the boundary clay, which is thought to come from the stratospheric dust, is markedly different from that of clay mixed with the Cretaceous and Tertiary limestones, which are chemically similar to each other. Four different independent estimates of the diameter of the asteroid give values that lie in the range 10 ± 4 kilometers.

In the 570-million-year period for which abundant fossil remains are available, there have been five great biological crises, during which many groups of organisms died out. The most recent of the great extinctions is used to define the boundary between the Cretaceous and Tertiary periods, about 65 million years ago. At this time, the marine reptiles, the flying reptiles, and both orders of dinosaurs died out (*1*), and extinctions occurred at various taxonomic levels among the marine invertebrates. Dramatic extinctions occurred among the microscopic floating animals and plants; both the calcareous planktonic foraminifera and the calcareous nannoplankton were nearly exterminated, with only a few species surviving the crisis. On the other hand, some groups were little affected, including the land plants, crocodiles, snakes, mammals, and many kinds of invertebrates. Russell (*2*) concludes that about half of the genera living at that time perished during the extinction event.

Many hypotheses have been proposed to explain the Cretaceous-Tertiary (C-T) extinctions (*3*, *4*), and two recent meetings on the topic (*5*, *6*) produced no sign of a consensus. Suggested causes include gradual or rapid changes in oceanographic, atmospheric, or climatic conditions (*7*) due to a random (*8*) or a cyclical (*9*) coincidence of causative factors; a magnetic reversal (*10*); a nearby supernova (*11*); and the flooding of the ocean surface by fresh water from a postulated arctic lake (*12*).

A major obstacle to determining the cause of the extinction is that virtually all the available information on events at the time of the crisis deals with biological changes seen in the paleontological record and is therefore inherently indirect. Little physical evidence is available, and it also is indirect. This includes variations in stable oxygen and carbon isotopic ratios across the boundary in pelagic sediments, which may reflect changes in temperature, salinity, oxygenation, and organic productivity of the ocean water, and which are not easy to interpret (*13*, *14*). These isotopic changes are not particularly striking and, taken by themselves, would not suggest a dramatic crisis. Small changes in minor and trace element levels at the C-T boundary have been noted from limestone sections in Denmark and Italy (*15*), but these data also present interpretational difficulties. It is noteworthy that in pelagic marine sequences, where nearly continuous deposition is to be expected, the C-T boundary is commonly marked by a hiatus (*3*, *16*).

In this article we present direct physical evidence for an unusual event at exactly the time of the extinctions in the planktonic realm. None of the current hypotheses adequately accounts for this evidence, but we have developed a hypothesis that appears to offer a satisfactory explanation for nearly all the available paleontological and physical evidence.

Luis Alvarez is professor emeritus of physics at Lawrence Berkeley Laboratory, University of California, Berkeley 94720. Walter Alvarez is an associate professor in the Department of Geology and Geophysics, University of California, Berkeley. Frank Asaro is a senior scientist and Helen Michel is a staff scientist in the Energy and Environment Division of Lawrence Berkeley Laboratory.

0036-8075/80/0606-1095$02.00/0

Identification of Extraterrestrial Platinum Metals in Deep-Sea Sediments

This study began with the realization that the platinum group elements (platinum, iridium, osmium, and rhodium) are much less abundant in the earth's crust and upper mantle than they are in chondritic meteorites and average solar system material. Depletion of the platinum group elements in the earth's crust and upper mantle is probably the result of concentration of these elements in the earth's core.

Pettersson and Rotschi (*17*) and Goldschmidt (*18*) suggested that the low concentrations of platinum group elements in sedimentary rocks might come largely from meteoritic dust formed by ablation when meteorites passed through the atmosphere. Barker and Anders (*19*) showed that there was a correlation between sedimentation rate and iridium concentration, confirming the earlier suggestions. Subsequently, the method was used by Ganapathy, Brownlee, and Hodge (*20*) to demonstrate an extraterrestrial origin for silicate spherules in deep-sea sediments. Sarna-Wojcicki *et al.* (*21*) suggested that meteoritic dust accumulation in soil layers might enhance the abundance of iridium sufficiently to permit its use as a dating tool. Recently, Crocket and Kuo (*22*) reported iridium abundances in deep-sea sediments and summarized other previous work.

Considerations of this type (*23*) prompted us to measure the iridium concentration in the 1-centimeter-thick clay layer that marks the C-T boundary in some sections in the Umbrian Apennines, in the hope of determining the length of time represented by that layer. Iridium can easily be determined at low levels by neutron activation analysis (NAA) (*24*) because of its large capture cross section for slow neutrons, and because some of the gamma rays given off during de-excitation of the decay product are not masked by other gamma rays. The other platinum group elements are more difficult to determine by NAA.

Italian Stratigraphic Sections

Many aspects of earth history are best recorded in pelagic sedimentary rocks, which gradually accumulate in the relatively quiet waters of the deep sea as individual grains settle to the bottom. In the Umbrian Apennines of northern peninsular Italy there are exposures of pelagic sedimentary rocks representing the time from Early Jurassic to Oligocene, around 185 to 30 million years ago (*25*). The C-T boundary occurs within a portion of the sequence formed by pink limestone containing a variable amount of clay. This limestone, the *Scaglia rossa*, has a matrix of coccoliths and coccolith fragments (calcite platelets, on the order of 1 micrometer in size, secreted by algae living in the surface waters) and a rich assemblage of foraminiferal tests (calcite shells, generally in the size range 0.1 to 2.0 millimeters, produced by single-celled animals that float in the surface waters).

Fig. 1. Photomicrographs of (a) the basal bed of the Tertiary, showing *Globigerina eugubina*, and (b) the top bed of the Cretaceous, in which the largest foraminifer is *Globotruncana contusa*. Both sections are from the Bottaccione section at Gubbio; they are shown at the same scale and the bar in (a) is 1 mm long.

In some Umbrian sections there is a hiatus in the sedimentary record across the C-T boundary, sometimes with signs of soft-sediment slumping. Where the sequence is apparently complete, foraminifera typical of the Upper Cretaceous (notably the genus *Globotruncana*) disappear abruptly and are replaced by the basal Tertiary foraminifer *Globigerina eugubina* (*16*, *26*). This change is easy to recognize because *G. eugubina*, unlike the globotruncanids, is too small to see with the naked eye or the hand lens (Fig. 1). The coccoliths also show an abrupt change, with disappearance of Cretaceous forms, at exactly the same level as the foraminiferal change, although this was not recognized until more recently (*27*).

In well-exposed, complete sections there is a bed of clay about 1 cm thick between the highest Cretaceous and the lowest Tertiary limestone beds (*28*). This bed is free of primary $CaCO_3$, so there is no record of the biological changes during the time interval represented by the clay. The boundary is further marked by a zone in the uppermost Cretaceous in which the normally pink limestone is white in color. This zone is 0.3 to 1.0 meter thick, varying from section to section. Its lower boundary is a gradational color change; its upper boundary is abrupt and coincides with the faunal and floral extinctions. In one section (Contessa) we can see that the lower 5 mm of the boundary clay is gray and the upper 5 mm is red, thus placing the upper boundary of the zone in the middle of the clay layer.

The best known of the Umbrian sections is in the Bottaccione Gorge near Gubbio. Here some of the first work on the identification of foraminifera in thin section was carried out (*29*); the oldest known Tertiary foraminifer, *G. eugubina*, was recognized, named, and used to define the basal Tertiary biozone (*16*, *26*); the geomagnetic reversal stratigraphy of the Upper Cretaceous and Paleocene was established, correlated to the marine magnetic anomaly sequence, and dated with foraminifera (*30*); and the extinction of most of the nannoplankton was shown to be synchronous with the disappearance of the genus *Globotruncana* (*27*).

Results from the Italian Sections

Our first experiments involved NAA of nine samples from the Bottaccione section (two limestone samples from immediately above and below the boundary plus seven limestone samples spaced over 325 m of the Cretaceous). This was supplemented by three samples from the nearby Contessa section (two from the boundary clay and one from the basal Tertiary bed). Stratigraphic positions of these samples are shown in Fig. 2.

Twenty-eight elements were selected for study because of their favorable nuclear properties, especially neutron capture cross sections, half-lives, and gamma-ray energies. The results of these analyses are presented in Fig. 3 on a logarithmic plot to facilitate comparison of the relative changes in elements over a wide range of concentrations. The only preparation given to these samples was removal of the $CaCO_3$ fraction by dissolution in dilute nitric acid. Figure 3 shows elemental abundances as gram of element per gram of insoluble clay residue. The limestones generally contain about 5 percent clay. The boundary clay layer contains about 50 percent $CaCO_3$, but this is coarse-grained calcite that probably crystallized during deformation long after deposition. Chemical yields of iridium in the acid-insoluble fraction averaged 44 percent for the red and gray Contessa boundary clays, and this value was assumed for all the other samples.

Twenty-seven of the 28 elements show very similar patterns of abundance variation, but iridium shows a grossly different behavior: it increases by a factor of about 30 in coincidence with the C-T boundary, whereas none of the other elements as much as doubles with respect to an "average behavior" shown in the lower right panel of Fig. 3. Figure 4 shows a typical gamma-ray spectrum used to measure the Ir abundance, 5.5 parts per billion (ppb).

In follow-up experiments we analyzed five more samples from the Bottaccione section, eight from Gorgo a Cerbara (28 kilometers north of Gubbio), and four large samples of the boundary clay from the two sections near Gubbio and two sections about 30 km to the north (*31*). The chemical yield of iridium in the acid-insoluble fraction was 95 ± 5 percent for the Contessa boundary clay, and a 100 percent yield was assumed for all the other samples.

Figure 5 shows the results of 29 Ir analyses completed on Italian samples. Note that the section is enlarged and that the scale is linear in the vicinity of the C-T boundary, where details are important, but changes to logarithmic to show results from 350 m below to 50 m above the boundary. It is also important to note that analyses from five stratigraphic sections are plotted on the same diagram on the basis of their stratigraphic position above or below the boundary. Because slight differences in sedimentation rate probably exist from one section to the next, the chronologic sequence of samples from different sections may not be exactly correct. Nevertheless, Fig. 5 gives a clear picture of the general trend of iridium concentrations as a function of stratigraphic level.

The pattern, based especially on the samples from the Bottaccione Gorge and Gorgo a Cerbara, shows a steady background level of ~ 0.3 ppb throughout the Upper Cretaceous, continuing into the uppermost bed of the Cretaceous. The background level in the acid-insoluble residues is roughly comparable to the iridium abundance measured by other

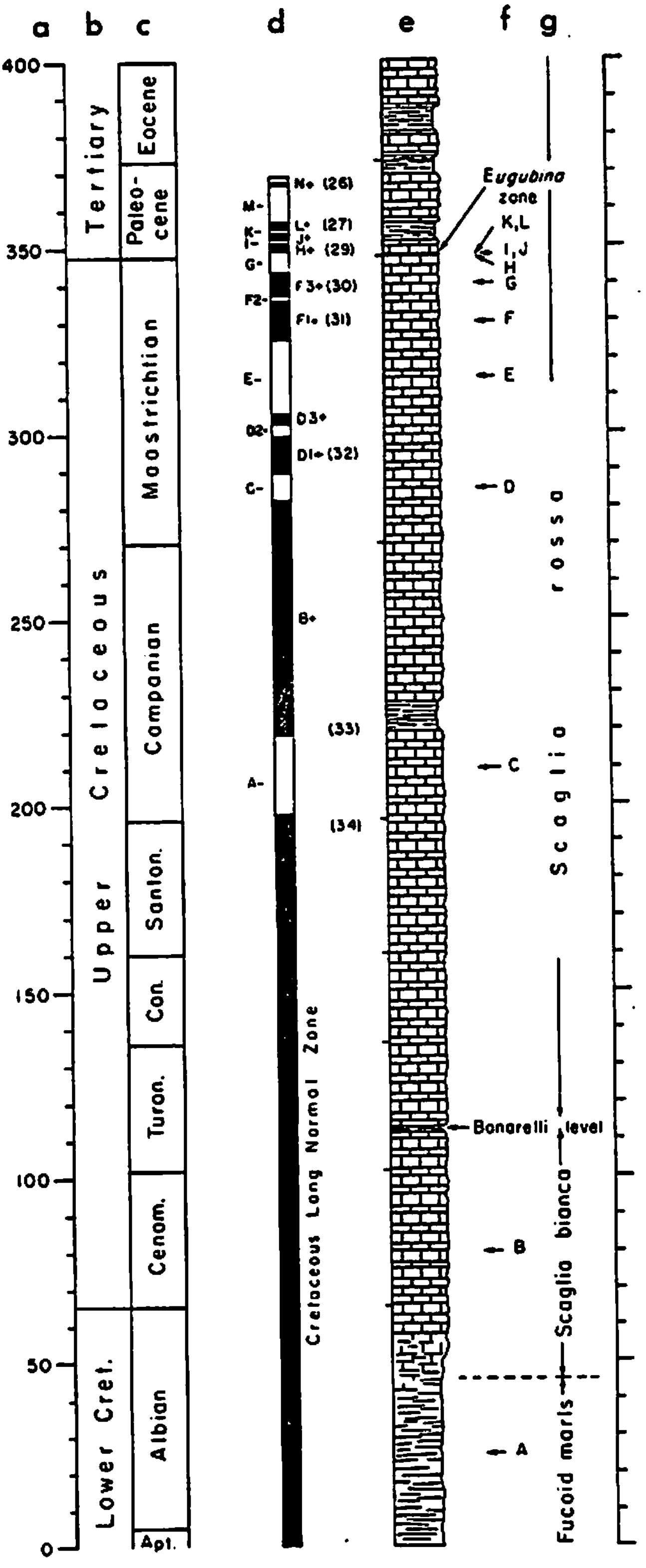


Fig. 2. Stratigraphic section at the Bottaccione Gorge, Gubbio (*30*). (a) Meter levels. (b) Systems. (c) Stages. (d) Magnetic polarity zones (black is normal, white is reversed polarity, letters give Gubbio polarity zonation, numbers are equivalent marine magnetic anomalies). (e) Lithology. (f) Samples used in first NAA study (samples *I*, *J*, and *L* are from equivalent positions in the Contessa section, 2 km to the northwest). (g) Formation names.

workers (*19*, *22*, *32*) in deep-sea clay sediments. This level increases abruptly, by a factor of more than 30, to 9.1 ppb, the Ir abundance in the red clay from the Contessa section. Iridium levels are high in clay residues from the first few beds of Tertiary limestone, but fall off to background levels by 1 m above the boundary. For comparison, the upper dashed line in Fig. 5 shows an exponential decay from the boundary clay Ir level with a half-height of 4.6 cm.

To test the possibility that iridium might somehow be concentrated in clay layers, we subsequently analyzed two red clay samples from a short distance below the C-T boundary in the Bottaccione section. One is from a distinctive clay layer 5 to 6 mm thick, 1.73 m below the boundary; the other is from a 1- to 2-mm bedding-plane clay seam 0.85 m below the boundary. The whole-rock analyses of these clays showed no detectable Ir with limits of 0.5 and 0.24 ppb, respectively. Thus neither clay layers from below the C-T boundary nor clay components in the limestone show evidence of Ir above the background level.

The Danish Section

To test whether the iridium anomaly is a local Italian feature, it was desirable to analyze sediments of similar age from another region. The sea cliff of Stevns Klint, about 50 km south of Copenhagen, is a classical area for the C-T boundary and for the Danian or basal stage of the Tertiary. A collection of up-to-date papers on this and nearby areas has recently been published, which includes a full bibliography of earlier works (*6*, vol. 1).

Our samples were taken at Højerup Church (*33*). At this locality the Maastrichtian, or uppermost Cretaceous, is represented by white chalk containing black chert nodules in undulating layers with amplitudes of a few meters and wavelengths of 10 to 50 m (*14*). These undulations are considered to represent bryozoan banks (*34*). The C-T boundary is marked by the *Fiskeler*, or fish clay, which is up to 35 cm thick in the deepest parts of the basins between bryozoan banks (*14*) but commonly only a few centimeters thick, thinning or disappearing, over the tops of the banks. The fish clay at Højerup Church was studied in detail by Christensen *et al.* (*14*), who subdivided it into four thin layers; we analyzed a sample mixing the two internal layers (units III and IV of Christensen *et al.*). These layers are black or dark gray, and the lower one contains pyrite concretions; the layers below and above (II and V) are light gray in color. Undisturbed lamination in bed IV suggests that no bottom fauna was present during its deposition (*14*). Above the fish clay, the Cerithium limestone is present to a thickness of about 50 cm in the small basins, disappearing over the banks. It is hard, yellowish in color, and cut by abundant burrows. Above this is a thick bryozoan limestone.

The presence of a thin clay layer at the C-T boundary in both the Italian and Danish sections is quite striking. However, there are notable differences as well. The Danish sequence was clearly deposited in shallower water (*35*), and the Danish limestones preserve an extensive bottom-dwelling fauna of bivalves (*36*), echinoderms (*37*), bryozoans (*38*), and corals (*39*).

Foraminiferal (*40*) and coccolith (*41*)

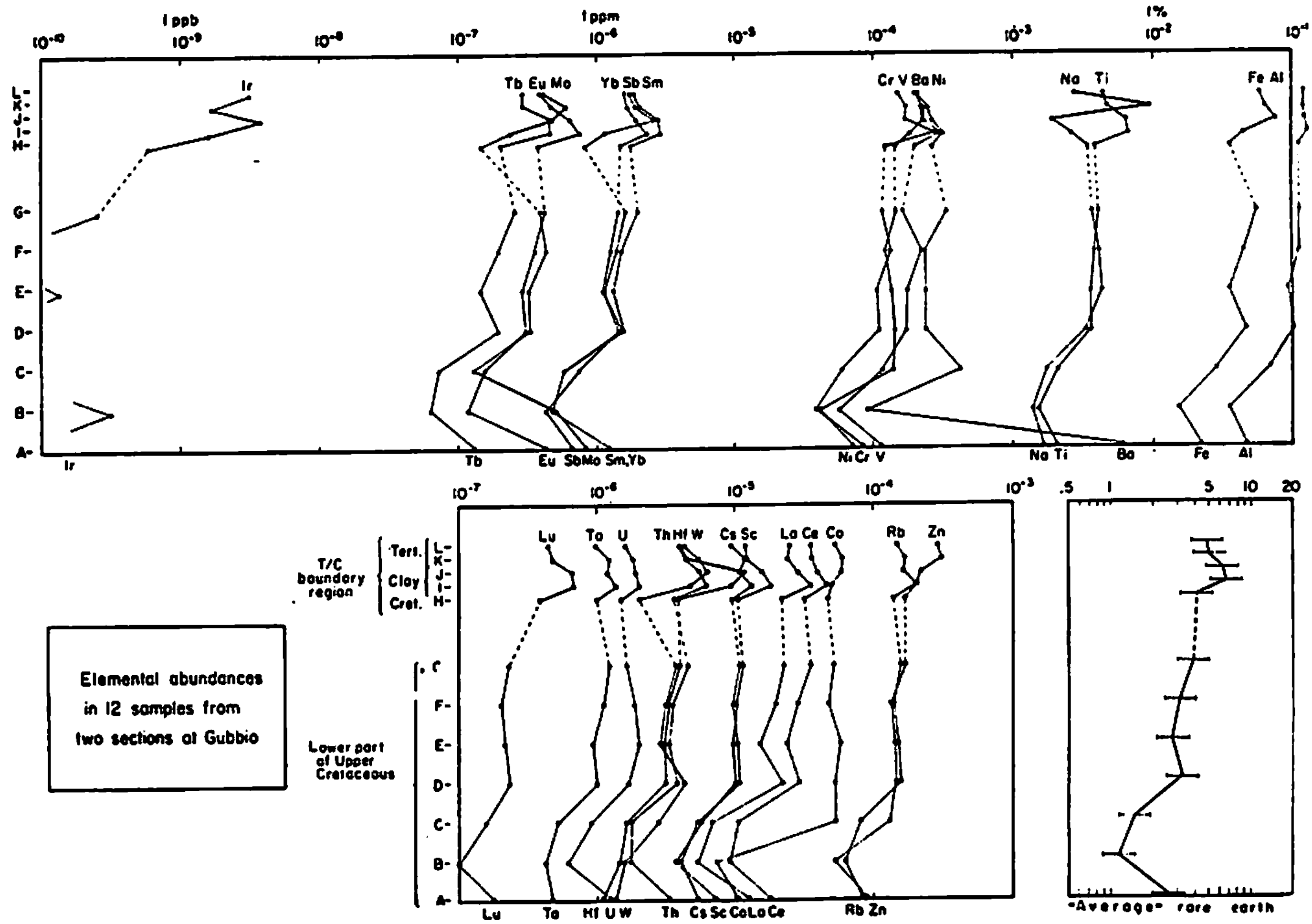


Fig. 3. Abundance variations of 28 elements in 12 samples from two Gubbio sections. Flags on "average rare earth" diagram are ± 30 percent and include all rare earth data.

zonation indicates that the C-T boundaries at Gubbio and Stevns Klint are at least approximately contemporaneous, and they may well be exactly synchronous. However, no paleomagnetic results are available from Stevns Klint, so synchroneity cannot be tested by reversal stratigraphy.

Results from the Danish Section

Seven samples were taken from near the C-T boundary (Fig. 6). Fractions of each sample were treated with dilute nitric acid, and the residues were filtered, washed, and heated to 800°C. The yield of acid-insoluble residue was 44.5 percent for the boundary fish clay and varied from 0.62 to 3.3 percent for the pelagic limestones (Table 1).

Neutron activation analysis (*24*) and x-ray fluorescence (XRF) (*42, 43*) measurements were made on all seven samples both before and after the acid treatment. This measurement regime was more sophisticated than that used for the Italian sections studied earlier, and 48 elements were determined.

The Cretaceous and Tertiary acid-insoluble residues were each rather homogeneous in all of the measured elements, and the two groups were only slightly different from each other. The residue from the clay boundary layer was much different in composition (Figs. 7 and 8 and Table 2), and this suggested a different source for the boundary clay.

As shown in Table 1, the Ir in the boundary layer residue rises by about a factor of 160 over the background level (~ 0.26 ppb). A 1-cm thickness of this layer would have about 72 × 10^{-9} gram of Ir per square centimeter. To test whether there is enough Ir in the seawater to contribute to this value, we made a measurement of the Ir in the ocean off the central California coast. In water passed through a 0.45-μm filter Ir was undetected, giving an upper limit of 4 × 10^{-13} g of Ir per gram of seawater. If the depth of the shallow ancient Danish sea is assumed to be less than 100 m and our limit for Ir in seawater is applicable, then the maximum Ir in the 100-m column of water should be 4 × 10^{-9} g/cm^2, almost a factor of 20 lower than the observed value. So there was probably not enough Ir stored in the seawater to explain the amount observed in the Danish boundary. Iridium has apparently not been detected in seawater. One tabulated result (*44*) contains a typographical error that places the value for indium in the atomic number position of iridium. Iridium has been detected (*45*) in a warm spring on Mount Hood in northern California at a level of 7 × 10^{-12} g per gram of water, and in two cold-water sources at levels of 3 × 10^{-13} to 4 × 10^{-13} g per gram of water. Many other cold-water sources in this area had Ir levels less than 1 × 10^{-13} g/g.

The Boundary Layers

The whole-rock composition of the Contessa boundary layer (a mixture of red and gray clay) is shown in Table 3. There are two recognizable sublayers, each about 0.5 cm thick, the upper being red in color and the lower gray. The elemental iron content, which may explain

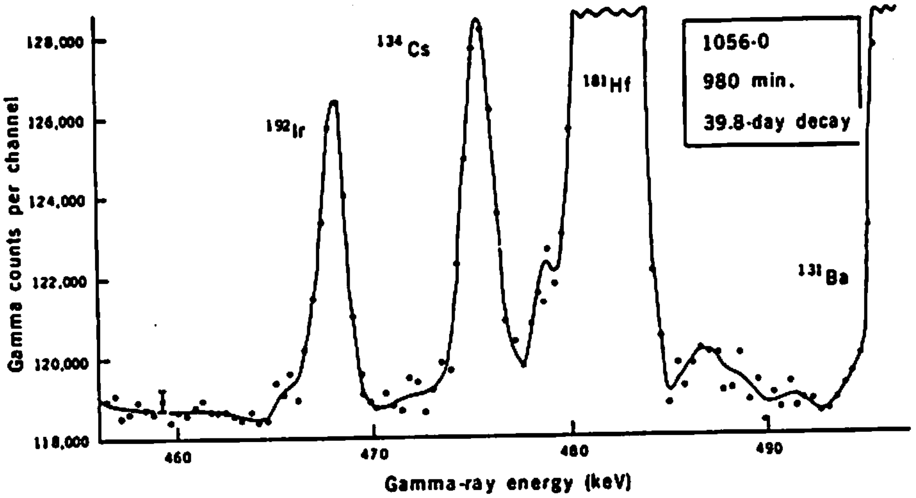
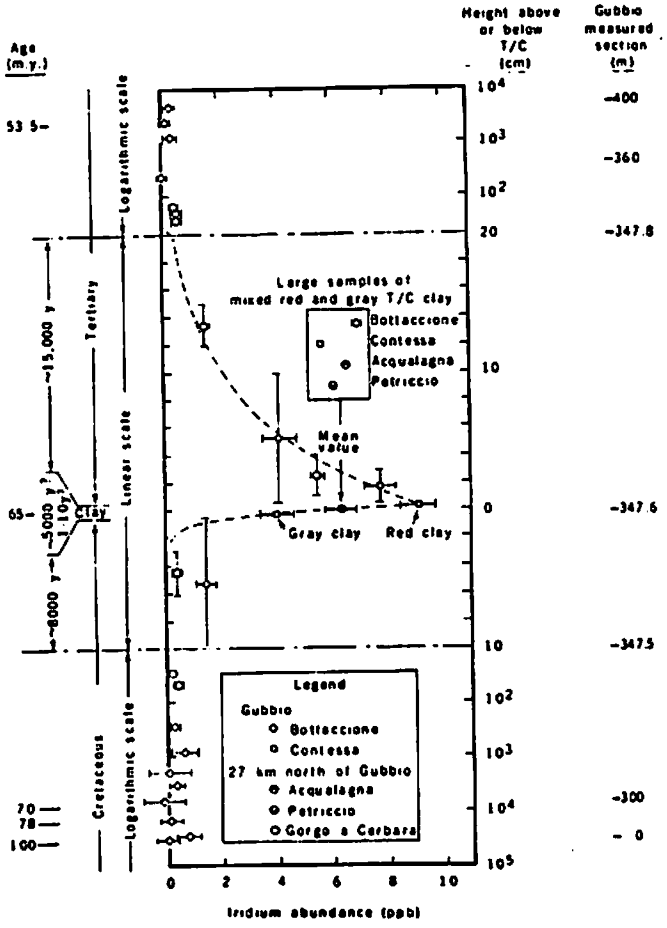


Fig. 4. Typical gamma-ray spectrum used to determine Ir abundance (5.5 ppb) in nitric acid-insoluble residues without further chemistry. Note that the entire spectrum rests on a background of 118,000 counts. Detector volume was 128 cm^3; length of count was 980 minutes. Count began 39.8 days after the end of the irradiation. Residue is from a Tertiary limestone sample taken 2.5 cm above the boundary at Gorgo a Cerbara (see Fig. 5).

Fig. 5. Iridium abundances per unit weight of 2*N* HNO_3 acid-insoluble residues from Italian limestones near the Tertiary - Cretaceous boundary. Error bars on abundances are the standard deviations in counting radioactivity. Error bars on stratigraphic position indicate the stratigraphic thickness of the sample. The dashed line above the boundary is an "eyeball fit" exponential with a half-height of 4.6 cm. The dashed line below the boundary is a best fit exponential (two points) with a half-height of 0.43 cm. The filled circle and error bar are the mean and standard deviation of Ir abundances in four large samples of boundary clay from different locations.



the color, is significantly higher in the residue of the red layer (7.7 versus 4.5 percent) than in the gray, and so is the Ir (9.1 ± 0.6 versus 4.0 ± 0.6 ppb). Boundary samples were analyzed from the Bottaccione Gorge nearby and two other areas about 30 km to the north.

In samples taken near the Italian boundary layer, the chemical compositions of all clay fractions were roughly the same except for the element Ir. However, there are discernible differences, as shown in Fig. 9, which suggest that at least part of the boundary layer clay had a different origin than the Cretaceous and Tertiary clays.

The Danish C-T boundary clay is somewhat thicker than 1 cm and is divided into four layers, as mentioned earlier. Only a single mixed sample from the two middle layers was measured, so no information is available on the chemical variations within the boundary. The average Ir abundance is 29 ppb in the whole rock or 65 ppb based on the weight of acid-insoluble residue.

The whole-rock abundances and mineral composition of the Danish boundary clay are shown in Table 3, and the abundances of pertinent trace elements are shown in Table 2. The major silicate minerals that must be present were not detected, so the other mineral abundances were normalized to give the amount of calcite expected from the calcium measurement. The boundary clay fraction is far different chemically from the limestone clay fractions above and below, which are similar to each other. Pyrite is present in the boundary clay, and elements that form water-insoluble sulfides are greatly enhanced in this layer. The trace elements that are depleted are those that often appear as clay components. The element magnesium is an exception. Its enhancement may be due to replacement of iron in the clay lattices in the sulfide environment or to a different, more mafic source for the boundary layer clay than for the Tertiary and Cretaceous clays.

Table 1. Abundance of iridium in acid-insoluble residues in the Danish section.

Sample*	Abundance of iridium (ppb)	Abundance† of acid-insoluble residues (%)
SK, +2.7 m	< 0.3	3.27
SK, +1.2 m	< 0.3	1.08
SK, +0.7 m	0.36 ± 0.06	0.836
Boundary	41.6 ± 1.8‡	44.5
SK, −0.5 m	0.73 ± 0.08	0.654
SK, −2.2 m	0.25 ± 0.08	0.621
SK, −5.4 m	0.30 ± 0.16	0.774

*Numerical values are the distances above (+) or below (−) the boundary layer; SK, Stevns Klint. †The boundary layer has a much higher proportion of clay than the pelagic limestones above and below. ‡Some iridium dissolved in the nitric acid. The whole-rock abundance was 28.6 ± 1.3 ppb.

Recent unpublished work by D. A. Russell of the National Museums of Canada and by the present authors has shown that the boundary layer whole-rock concentration of Ir in a section near Woodside Creek, New Zealand, is approximately 20 times the average concentration in the adjacent Cretaceous and Tertiary limestones.

A Sudden Influx of Extraterrestrial Material

To test whether the anomalous iridium at the C-T boundary in the Gubbio sections is of extraterrestrial origin, we considered the increases in 27 of the 28 elements measured by NAA that would be expected if the iridium in excess of the background level came from a source with the average composition of the earth's crust. The crustal Ir abundance, less than 0.1 ppb (*19*, *22*), is too small to be a worldwide source for material with an Ir abundance of 6.3 ppb, as found near Gubbio. Extraterrestrial sources with Ir levels of hundreds of parts per

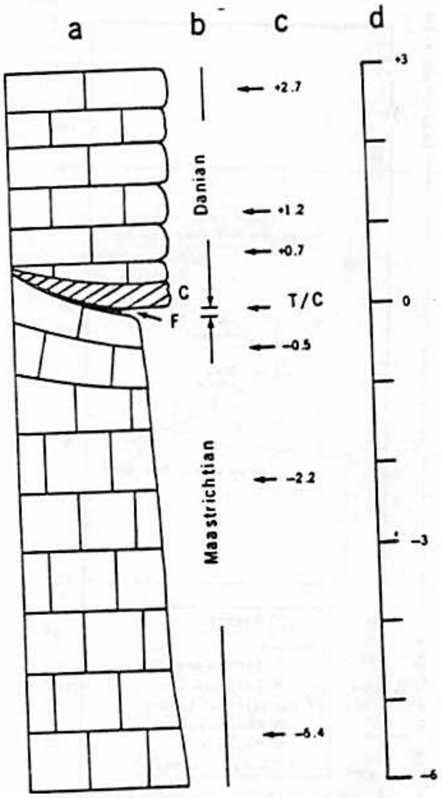


Fig. 6 (left). Stratigraphic section at Højerup Church, Stevns Klint, Denmark. (a) Lithology (*C*, Cerithium limestone; *F*, fish clay). (b) Stages. (c) Samples analyzed in this study. (d) Meter levels. Analytical results are given in Tables 1 to 3.

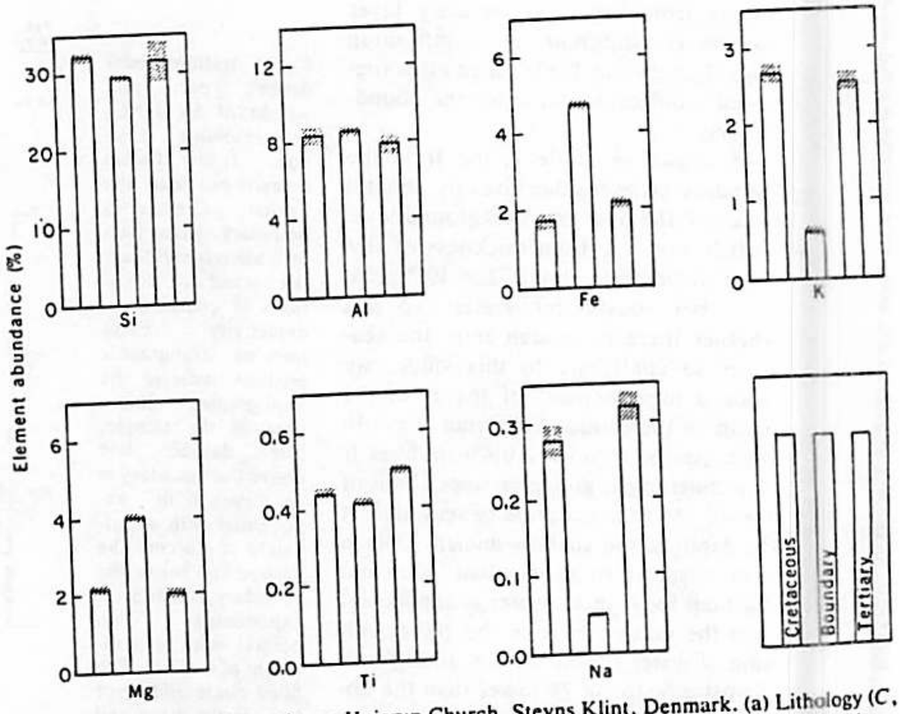


Fig. 7 (right). Major element abundances in acid-insoluble fractions from Danish rocks near the Tertiary-Cretaceous boundary. The crosshatched areas for the Cretaceous and Tertiary values each represent root-mean-square deviations for three samples. (Only two measurements of magnesium and silicon were included in the Cretaceous values.) For the boundary layer sample the crosshatched areas are the standard deviations associated with counting errors. Measurements of silicon and magnesium were done by XRF (*42*), all others were by NAA.

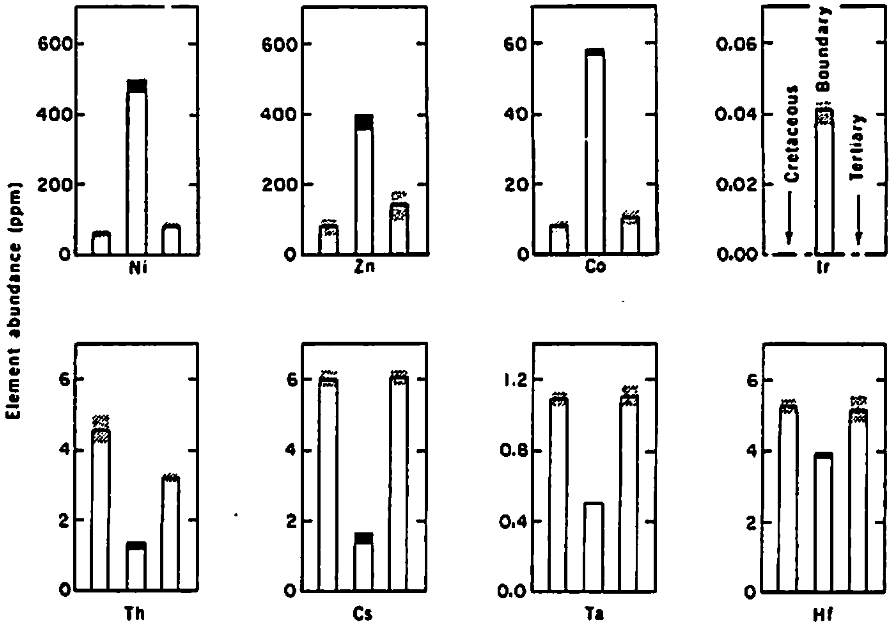


Fig. 8. Selected trace element abundances of Danish acid-insoluble residues. First bar is the mean value [root-mean-square deviation (RMSD) is shown by the crosshatched areas] for the given element in the three Cretaceous residues. Second bar is the abundance (counting error is shown by solid areas) for the given element in the boundary layer residue. Third bar is the mean value and RMSD for the given element in the three Tertiary residues. Measurements were by NAA, except for zinc, which was measured by XRF (*43*). Significant amounts of nickel, zinc, cobalt, iridium, and thorium in all samples dissolved in the 2*N* HNO_3. Very little cesium, tantalum, or hafnium in any of the samples dissolved in the acid.

billion or higher are more likely to have produced the Ir anomaly. Figure 10 shows that if the source had an average earth's crust composition (*46*), increases significantly above those observed would be expected in all 27 elements. However, for a source with average carbonaceous chondrite composition (*46*), only nickel should show an elemental increase greater than that observed. As shown in Fig. 11, such an increase in nickel was not observed, but the predicted effect is small and, given appropriate conditions, nickel oxide would dissolve in seawater (*47*). We conclude that the pattern of elemental abundances in the Gubbio sections is compatible with an extraterrestrial source for the anomalous iridium and incompatible with a crustal source.

The Danish boundary layer, which has much more Ir than the Italian C-T clay, is even less likely to have had a crustal origin. Rocks from the upper mantle (which has more Ir than the crust) have less than 20 ppb (*48*) and are therefore an unlikely worldwide source. There are, however, localized terrestrial sources with much higher Ir abundances; for example, nickel sulfide and chromite ores (*48*) have Ir levels of hundreds and thousands of parts per billion, respectively. The Danish boundary layer, however, does not have enough nickel [506 parts per million (ppm)] or chromium (165 ppm) to explain its Ir in this fashion, unless the marine chemistry concentrated Ir preferentially (even in the sulfide environment) and disposed of the other elements elsewhere. The probability of these effects occurring worldwide seems less likely than an extraterrestrial origin for the Ir.

We next consider whether the Ir anomaly is due to an abnormal influx of

Table 2. Abundance of trace elements in the Danish boundary layer (parts per million).

Element	(1) Abundance in whole rock/ abundance of residue*		(2) Abundance in residue*		Element	(1) Abundance in whole rock/ abundance of residue		(2) Abundance in residue	
	Enhanced elements†					*Depleted elements*			
V	391	± 27	330	± 31	Mn	102.0	± 1.3	21.3	± 0.5
Cr	371	± 13	358	± 9	Rb	27	± 7	35	± 4
Co	141.6	± 1.8	57.2	± 0.7	Y‡	79	± 6	6.3	± 1.8
Ni	1137	± 31	479	± 14	Zr‡	144	± 11	125	± 6
Cu‡	167	± 14	93	± 6	Nb‡	8	± 4	6.1	± 1.8
Zn‡	1027	± 49	378	± 18	Cs	1.87	± 0.19	1.51	± 0.14
As	96	± 8	68	± 4	La	61.1	± 1.6	6.8	± 0.4
Se§	46.5	± 0.6	12.1	± 0.3	Ce	57.0	± 1.2	9.7	± 0.6
Mo	29.0	± 2.5	20.3	± 1.4	Nd	63.4	± 2.7	5.4	± 0.6
Ag§	2.6	± 0.9	3.5	± 0.7	Sm	11.93	± 0.08	0.781	± 0.008
In§	0.245	± 0.022	0.086	± 0.019	Eu	2.76	± 0.11	0.121	± 0.010
Sb	8.0	± 0.4	6.7	± 0.4	Tb	1.84	± 0.04	0.148	± 0.014
Ba	1175	± 16	747	± 11	Dy	11.24	± 0.12	0.908	± 0.033
Ir	0.0643	± 0.0029	0.0416	± 0.0018	Yb	5.02	± 0.09	0.56	± 0.05
Pb‡	64	± 14	28	± 7	Lu	0.553	± 0.031	0.083	± 0.004
					Hf	4.34	± 0.16	3.88	± 0.07
	Other elements†				Ta	0.508	± 0.011	0.500	± 0.005
Sc	20.74	± 0.16	14.30	± 0.14	Th	7.1	± 0.4	1.28	± 0.06
Ga‡	30	± 6	19.8	± 3.0	U	8.63	± 0.09	0.918	± 0.024
Sr‡	1465	± 72	48.1	± 2.4					
Au	< 0.12		0.027	± 0.007					

*Column 1 minus column 2 is the amount of an element that dissolved in the acid or was lost in the firing: abundance of residue = 44.5 percent. †Elements V, Ag, and In are at least 20 percent and all other "enhanced elements" are at least a factor of 3 more abundant in the boundary residue than in the other residues. All "depleted elements" are at least 20 percent less abundant in the boundary residue than in the other residues. "Other elements" do not show a consistent pattern of boundary residue abundances relative to the others. ‡Measured by hard XRF (*43*). §Flux monitors were used in the NAA measurements of these elements. The indicated errors are applicable for comparing the two entries for a given element, but calibration uncertainties of possibly 10 to 20 percent must be considered when the values are used for other purposes.

extraterrestrial material at the time of the extinctions, or whether it was formed by the normal, slow accumulation of meteoritic material (*19*), followed by concentration in the boundary rocks by some identifiable mechanism.

There is prima facie evidence for an abnormal influx in the observations that the excess iridium occurs exactly at the time of one of the extinctions; that the extinctions were extraordinary events, which may well indicate an extraordinary cause; that the extinctions were clearly worldwide; and that the iridium anomaly is now known from two different areas in western Europe and in New Zealand. Furthermore, we will show in a later section that impact of a 10-km earth-crossing asteroid, an event that probably occurs with about the same frequency as major extinctions, may have produced the observed physical and biological effects. Nevertheless, one can invent two other scenarios that might lead to concentration of normal background iridium at the boundary. These appear to be much less likely than the sudden-influx model, but we cannot definitely rule out either one at present.

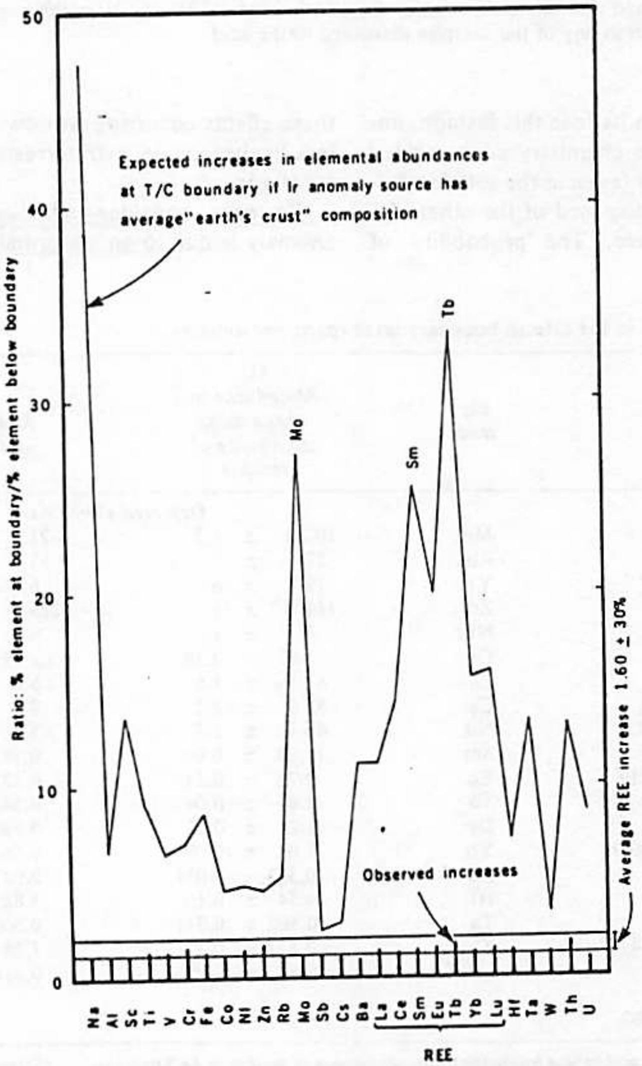
The first scenario requires a physical or chemical change in the ocean waters at the time of the extinctions, leading to extraction of iridium resident in the seawater. This would require iridium concentrations in seawater that are higher than those presently observed. In addition, it suggests that the positive iridium anomaly should be accompanied by a compensating negative anomaly immediately above, but this is not seen.

The second scenario postulates a reduction in the deposition rate of all components of the pelagic sediment except for the meteoritic dust that carries the concentrated iridium. This scenario requires removal of clay but not of iridium-bearing particles, perhaps by currents of exactly the right velocity. These currents must have affected both the Italian and Danish areas at exactly the time of the C-T extinctions, but at none of the other times represented by our samples. We feel that this scenario is too contrived, a conclusion that is justified in more detail elsewhere (*23*).

In summary, we conclude that the anomalous iridium concentration at the C-T boundary is best interpreted as indicating an abnormal influx of extraterrestrial material.

Negative Results of Tests for the Supernova Hypothesis

Considerable attention has been given to the hypothesis that the C-T extinctions were the result of a nearby supernova (*11*). A rough calculation of the distance from the assumed supernova to the solar system, using the measured surface density of iridium in the Gubbio



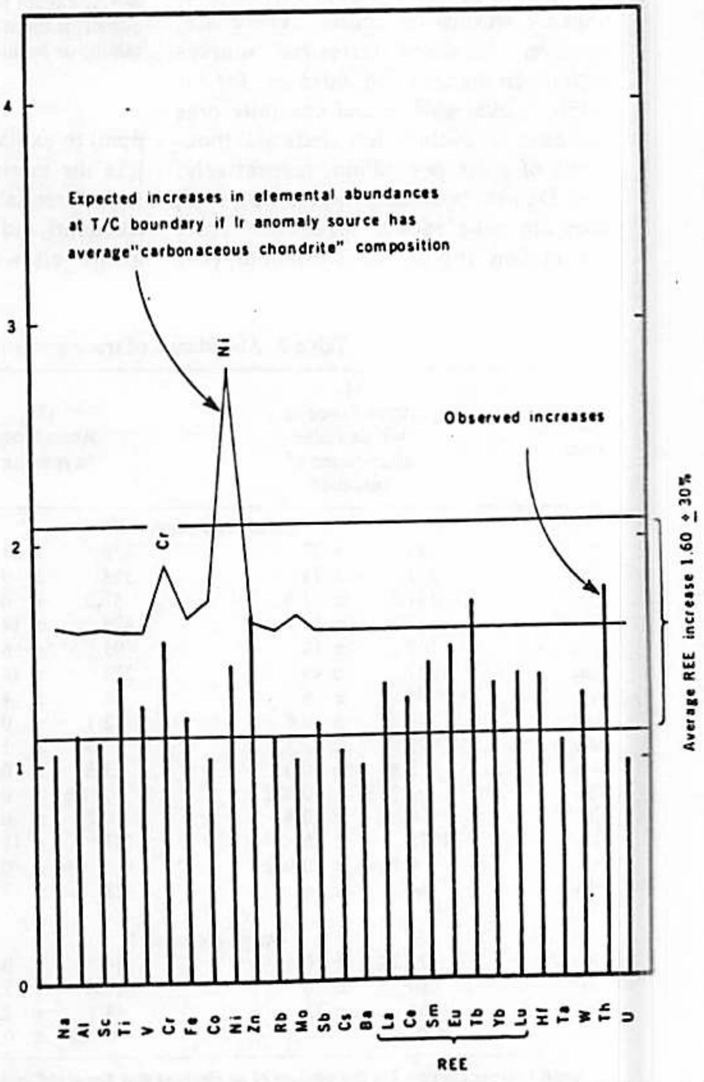


Fig. 10 (left). Comparison of observed elemental abundance patterns in the Gubbio section samples with average patterns expected for crustal material (*46*). Fig. 11 (right). Comparison of observed elemental abundance patterns in the Gubbio section samples with patterns expected for carbonaceous chondrites (*46*).

boundary layer and the amount of iridium expected to be blown off in the supernova explosion, gives about 0.1 light-year. The probability is about 10^{-9} (*49*) that, during the last 100 million years, a supernova occurred within this distance from the sun. Any mechanism with such a low a priori probability is obviously a one-time-only theory. Nevertheless, because the theory could be subjected to direct experimental tests, it was treated as a real possibility until we obtained two other independent pieces of evidence that forced us to reject it.

Elements heavier than nickel can be produced in stars only by neutron capture followed by beta decay. The most intense source of neutrons so far postulated is that produced by the gravitational collapse of the core of a star that leads immediately to a supernova explosion. In this environment the rapid capture of neutrons ("*r* process") leads to the formation of the heaviest known isotopes. The slower capture of neutrons by heavy isotopes in highly evolved stars ("*s* process") leads to a different mix of isotopes (*50*).

One heavy isotope in particular offered the possibility of testing the supernova hypothesis; this is ^{244}Pu, with a half-life of 80.5 × 10^6 years. The explosion of a supernova should send out an expanding shell of newly created heavy elements, with a ratio of Ir atoms to ^{244}Pu atoms equal to about 10^3. This value is inferred from the existence of an anomaly in the meteoritic abundance of heavy xenon isotopes that is interpreted as being due to the fission of ^{244}Pu (*51*). Any ^{244}Pu incorporated in the earth at the time of the creation of the solar system, about 4.7 billion years ago, would have decayed by 58 half-lives, or by a factor of 10^{17}, which would make it quite undetectable in the Gubbio section by the most sensitive techniques available. If the C-T extinctions were due to a supernova, and if this were the source of the anomalous Ir, each Ir atom should have been accompanied by about 10^{-3} ^{244}Pu atom, and this ^{244}Pu would have decayed by only a factor of 2.

Plutonium-244 is easily detected both by mass spectrometry and by NAA. The former is more sensitive, but the latter was immediately available. In NAA, which we utilized, ^{244}Pu is converted to ^{245}Pu, which has a half-life of 10 hours and emits many characteristic gamma rays and x-rays. Plutonium was chemically separated from 25- and 50-g batches of boundary clay and from a 50-g batch of bedding clay from below the C-T boundary, and nearly "mass-free" samples were obtained—no carriers were added. Chemical separations were also performed on the plutonium fraction after the neutron irradiation. No significant gamma radiation was observed, other than that associated with the plutonium isotopes. In order to measure our chemical yields, Gubbio acid-soluble and acid-insoluble residues were spiked with small amounts of ^{238}Pu tracer. This plutonium isotope is easily detectable through its alpha decay, as its half-life is only 87.7 years. In addition, one of the samples was spiked with ^{244}Pu. Figure 12a shows the gamma-ray spectrum of the sample spiked with about 20 picograms of ^{244}Pu; it indicates both the sensitivity of NAA for the detection of ^{244}Pu and the freedom of the purified sample from other elements that might interfere with the detection of ^{244}Pu. The plutonium isotopic ratios in this sample and in the tracer were also measured with a single-direction-focusing mass spectrometer 5 feet in radius.

No ^{244}Pu was detected in the Gubbio samples (Fig. 12b), with a detection limit of less than 10 percent of the amount that would be expected to accompany the measured iridium if a supernova were re-

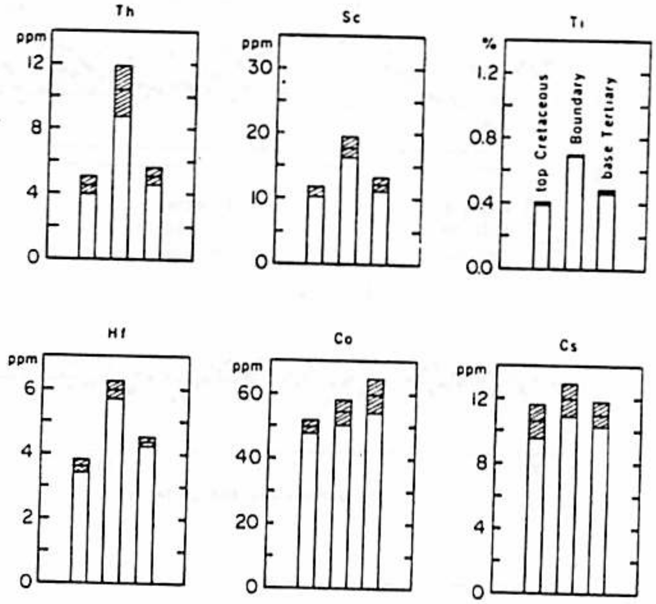


Fig. 9. Some of the element abundances measured in acid-insoluble residues of Cretaceous, boundary layer, and Tertiary rocks near Gubbio. Data include all samples from that area measured within 19 m of the boundary. There were four samples from each of the three layers; the crosshatched areas are the standard deviations. The abundance patterns for samples from ~ 27 km north of Gubbio are similar to those shown.

Table 3. Whole-rock composition of the Gubbio and Danish boundary layers (percent).

Element* or mineral	Abundance in boundary layer†		
	Gubbio (Contessa) measured	Denmark	
		Measured	Normalized
SiO_2	27.7 ± 0.6	29.0 ± 0.6	
Al_2O_3	12.19 ± 0.15	8.01 ± 0.17	
FeO‡	4.53 ± 0.05	4.35 ± 0.04	
MgO	1.10 ± 0.07	3.07 ± 0.10	
CaO	22.6 ± 0.4	23.1 ± 0.4	
Na_2O	0.1806 ± 0.0036	0.0888 ± 0.0018	
K_2O	2.46 ± 0.20	0.38 ± 0.04	
TiO_2	0.521 ± 0.022	0.324 ± 0.016	
S^{2-}	Not detected	~ 1.1	
PO_4^{3-}	Not detected	0.92 ± 0.09	
CO_2§	17.7 ± 0.3	18.4 ± 0.3	
Σ Trace elements	~ 0.2	~ 0.3	
Sum	89.2 ± 0.8	90.3 ± 1.0	
Difference‖	10.8 ± 0.8	9.7 ± 1.0	
Calcite		~ 90	41.5 (norm)
Quartz		5-7	~ 3
Pyrite		~ 5	~ 2
Illite		2-3	~ 1

*Abundance values are for element expressed as form shown. †Elements Si, Ca, Mg, S, P, and Gubbio Ti were measured by soft XRF (*42*). Some S may be lost in this sample preparation procedure. The Denmark Ti was measured by hard XRF (*43*). All other measurements were by NAA. Mineral analyses were done by M. Ghiorso and I. S. E. Carmichael by x-ray diffraction. ‡Total Fe expressed as FeO. §The CO_2 abundance was calculated from the Ca abundance by assuming all Ca was present as the carbonate. ‖The difference is mainly water and organic material.

sponsible for the latter. The ocean, however, can produce chemical and physical changes in depositing materials as well as diagenetic alterations in the deposited sediments, so the absence of measurable ^{244}Pu is not an absolutely conclusive argument.

The second method that was used to test whether a supernova was responsible for the iridium anomaly involved a measurement of the isotopic ratio of iridium in the boundary material. Iridium has two stable isotopes, 191 and 193, which would be expected to occur in about the same relative abundances, 37.3 to 62.7 percent, in all solar system material because of mixing the protosolar gas cloud. However, different supernovas should produce iridium with different isotopic ratios because of differences in the contributions of the *r* and *s* processes occasioned by variations in neutron fluxes, reaction times, and so on, from one supernova to the next. According to this generally accepted picture, solar system iridium is a mixture of that element produced by all the supernovas that ejected material into the gaseous nebula that eventually condensed to form the sun and its planets. A particular supernova would produce Ir with an isotopic ratio that might differ from that of solar system material by as much as a factor of 2 (*52*).

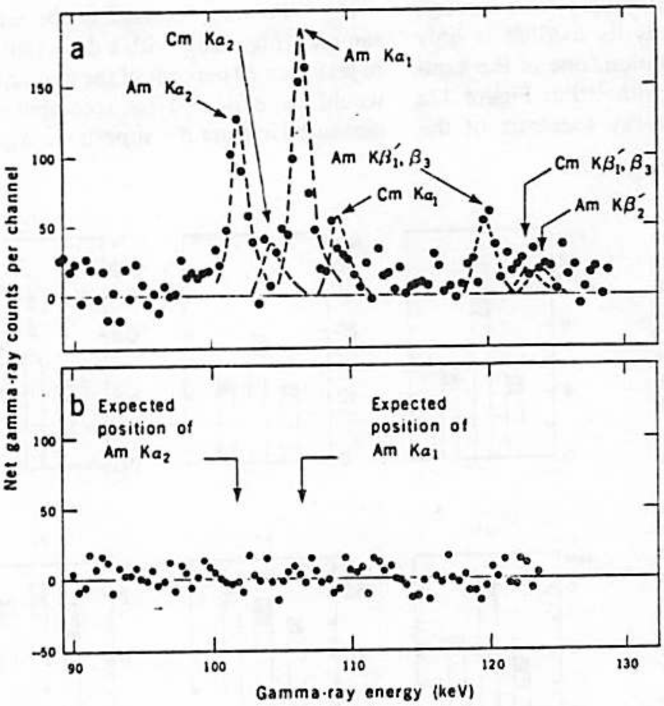


Fig. 12. Gamma-ray spectra of Pu fractions from acid-insoluble residues of irradiated boundary layer clay samples from Gubbio. (a) Sample had been spiked with ^{244}Pu and ^{238}Pu containing relatively small amounts of ^{239}Pu, ^{240}Pu, and ^{242}Pu. Dashed lines show expected energies and abundances of ^{245}Pu and equilibrated daughter radiations normalized to the 327.2-keV gamma ray of ^{245}Pu (not shown). (b) Sample had been spiked with ^{238}Pu containing relatively small amounts of ^{239}Pu, ^{240}Pu, and ^{242}Pu. No ^{244}Pu was detected.

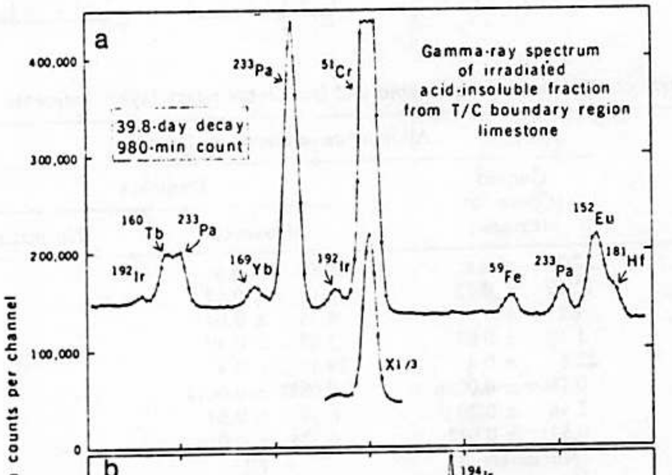


Fig. 13. (a) Gamma-ray spectrum of irradiated acid-insoluble residue from boundary layer clay at Gubbio without chemistry before or after irradiation. (b) Same as above with chemistry before and after irradiation used in isotopic ratio determinations. Counting periods, decay periods, and chemical yields are different for the two spectra.

We therefore compared the isotopic ratio of Ir from the C-T boundary clay with that of ordinary Ir, using NAA. This is a new technique (*23*), which we developed because of the extreme difficulty of determining Ir isotope ratios by mass spectrometry. In our earlier analytical work we used only the 74-day ^{192}Ir, made from ^{191}Ir by neutron capture. But in this new work we also measured the 18-hour ^{194}Ir made from the heavier Ir isotope, and extensive chemical separations before and after the neutron irradiations were necessary. Figure 13 contrasts a typical gamma-ray spectrum of the kind used in the isotopic ratio measurement with one used in an Ir abundance determination. This comparison demonstrates the need for chemical purification of the iridium fraction as well as the lack of major interfering radiations.

The final result is that the isotopic ratio of the boundary Ir differs by only 0.03 ± 0.65 percent (mean + 1 standard deviation) from that of the standard. From this, we conclude that the ^{191}Ir/^{193}Ir ratio in the boundary layer and the standard do not differ significantly by more than 1.5 percent. Therefore the anomalous Ir is very likely of solar system origin, and did not come from a supernova or other source outside the solar system (*53*)—for example, during passage of the earth through the galactic arms. [In a very recent paper, Napier and Clube suggest that catastrophic events could arise from the latter (*54*).]

The Asteroid Impact Hypothesis

After obtaining negative results in our tests of the supernova hypothesis, we were left with the question of what extraterrestrial source within the solar system could supply the observed iridium and also cause the extinctions. We considered and rejected a number of hypotheses (*23*); finally, we found that an extension of the meteorite impact hypothesis (*55, 56*) provided a scenario that explains most or all of the biological and physical evidence. In brief, our hypothesis suggests that an asteroid struck the earth, formed an impact crater, and some of the dust-sized material ejected from the crater reached the stratosphere and was spread around the globe. This dust effectively prevented sunlight from reaching the surface for a period of several years, until the dust settled to earth. Loss of sunlight suppressed photosynthesis, and as a result most food chains collapsed and the extinctions resulted. Several lines of evidence support this hypothesis, as discussed in the next few sections. The size of the impacting object can be calculated from four independent sets of observations, with good agreement among the four different diameter estimates.

Earth-Crossing Asteroids and Earth Craters

Two quite different data bases show that for the last billion years the earth has been bombarded by a nearly constant flux of asteroids that cross the earth's orbit. One data base comes from astronomical observations of such asteroids and a tabulation of their orbital parameters and their distribution of diameters (*57*). Öpik (*58*) computed that the mean time to collision with the earth for a given earth-crossing asteroid is about 200 million years. To a first approximation, the number of these objects with diameters greater than *d* drops roughly as the inverse square of *d*. E. M. Shoemaker [cited in (*59, 60*)] and Wetherill (*60*) independently estimated that there are at present about 700 earth-crossing asteroids with diameters greater than 1 km (Apollo objects), so there should be about seven with diameters greater than 10 km. This assumes that the power law with exponent −2 extends from the accessible 1-km-diameter range into the 10-km-diameter range. If one accepts the numbers given above, the mean time to collision for an earth-crossing asteroid with a diameter of 10 km or more would be 200 million years divided by 7, or ~ 30 million years. In a more sophisticated calculation, Shoemaker (*61*) estimates that a mean collision time of 100 million years is consistent with a diameter of 10 km, which is the value we will adopt. A discussion of cratering data, which leads to similar estimates, is given in Grieve and Robertson's review article (*62*) on the size and age distribution of large impact craters on the earth. Rather than present our lengthy justification (*23*) for the estimates based on the cratering data, we will simply report the evaluation of Grieve (*63*), who wrote: "I can find nothing in your data that is at odds with your premise." Grieve also estimates that the diameter of the crater formed by the impact of a 10-km asteroid would be about 200 km (*63*). This section of our article has thus been greatly condensed now that we have heard from experienced students of the two data bases involved.

Krakatoa

The largest well-studied terrestrial explosion in historical times was that of the island volcano, Krakatoa, in the Sunda Strait, between Java and Sumatra (*64*). Since this event provides the best available data on injection of dust into the stratosphere, we give here a brief summary of relevant information.

On 26 and 27 August 1883, Krakatoa underwent volcanic eruptions that shot an estimated 18 km^3 of material into the atmosphere, of which about 4 km^3 ended up in the stratosphere, where it stayed for 2 to 2.5 years. Dust from the explosion circled the globe, quickly giving rise to brilliant sunsets seen worldwide. Recent measurements of the ^{14}C injected into the atmosphere by nuclear bomb tests confirm the rapid mixing (about 1 year) between hemispheres (*65*). If we take the estimated dust mass in the stratosphere (4 km^3 times the assumed low density of 2 g/cm^3) and spread it uniformly over the globe, it amounts to 1.6×10^{-3} g/cm^2. This layer did not absorb much of the incident radiation on a "straight-through" basis. However, if it were increased by a factor of about 10^3 (a rough prediction of our theory), it is most probable that the sunlight would be attenuated to a high degree.

Since the time for the colored sunsets to disappear after Krakatoa is frequently given as 2 to 2.5 years, we have assumed that the asteroid impact material in the stratosphere settled in a few years. Thus, 65 million years ago, day could have been turned into night for a period of several years, after which time the atmosphere would return relatively quickly to its normal transparent state.

What happened during the Krakatoa explosions can be expected to happen to a much greater extent during the impact of a large asteroid. An interesting difference is that extreme atmospheric turbulence would follow the impact. The asteroid would enter the atmosphere at roughly 25 km/sec and would "punch a hole" in the atmosphere about 10 km across. The kinetic energy of the asteroid is approximately equivalent to that of 10^8 megatons of TNT.

Size of the Impacting Object

If we are correct in our hypothesis that the C-T extinctions were due to the impact of an earth-crossing asteroid, there are four independent ways to calculate the size of the object. The four ways and the results obtained are outlined below.

1) The postulated size of the incoming asteroid was first computed from the iridium measurements in the Italian sections, the tabulated Ir abundances (*66*) in type I carbonaceous chondrites (CI), which are considered to be typical solar system material, and the fraction of erupted material estimated to end up in the stratosphere. If we neglect the latter fraction for the moment, the asteroid mass is given by $M = sA/f$, where s is the surface density of Ir (measured at Gubbio to be 8×10^{-9} g/cm^2), A is the surface area of the earth, and f is the Ir mass fractional abundance in CI meteorites (0.5×10^{-6}). This preliminary value of the asteroid mass, 7.4×10^{16} g, is then divided by the estimated fraction staying in the stratosphere, 0.22, to give $M = 3.4 \times 10^{17}$ g. The "Krakatoa fraction," 0.22, is used simply because it is the only relevant number available. It could differ seriously from the correct value, however, as the two explosions are of quite different character. At a density of 2.2 g/cm^3 (*67*), the diameter of the asteroid would be 6.6 km.

2) The second estimate comes from data on earth-crossing asteroids and the craters they have made on the earth's surface. In a sense, the second estimate comes from two quite different data bases—one from geology and the other from astronomy. Calculations of the asteroid diameter can be made from both data bases, but they will not really be independent since the two data bases are known to be consistent with each other. As shown in an earlier section, the most believable calculation of the mean time between collisions of the earth and asteroids equal to or larger than 10 km in diameter is about 100 million years. The smaller the diameter the more frequent are the collisions, so our desire to fit not only the C-T extinction, but earlier ones as well, sets the mean time between ex-

tinctions at about 100 million years and the diameter at about 10 km.

3) The third method of estimating the size of the asteroid comes from the possibility that the 1-cm boundary layer at Gubbio and Copenhagen is composed of material that fell out of the stratosphere, and is not related to the clay that is mixed in with the limestone above and below it. This is quite a surprising prediction of the hypothesis, since the most obvious explanation for the origin of the clay is that it had the same source as the clay impurity in the rest of the Cretaceous and Tertiary limestone, and that it is nearly free of primary $CaCO_3$ because the extinction temporarily destroyed the calcite-producing plankton for about 5000 years. But as discussed earlier, the material in the boundary layer is of a different character from the clay above and below it, whereas the latter two clays are very similar. To estimate the diameter of the asteroid, one can use the surface density of the boundary layer (about 2.5 g/cm^2), together with an estimate of the fraction of that material which is of asteroidal origin. The asteroid diameter is then calculated to be 7.5 km. The numbers used in this calculation are the following: clay fraction in the boundary layer, 0.5; density of the asteroid, 2.2 g/cm^3; mass of crustal material thrown up per unit mass of asteroid, ~ 60 (*63*); fraction of excavated material delivered to the stratosphere, 0.22 (from the Krakatoa measurements). If one uses different numbers, the diameter changes only by the cube root of the ratio of input values.

The first and the third methods are independent, even though they both depend on measurements made on the boundary material. This can best be appreciated by noting that if the Ir abundance were about the same in the earth's crust as it is in meteorites, the iridium anomaly seen in Fig. 5 would not exist. Therefore, method 1 would not exist either. The fact that method 3 could still be used is the indicator of the relative independence of the two methods.

4) The fourth method is not yet able to set close limits on the mass of the incoming asteroid, but it leads to consistent results. This method derives from the need to make the sky much more opaque than it was in the years following the Krakatoa explosion. If it is assumed that the Krakatoa dust cloud attenuated the vertically incident sunlight by about 3 percent, then an explosion involving 33 times as much material would reduce the light intensity to $1/e$. The stratospheric mass due to an explosion of the magnitude calculated in the three earlier methods—about 1000 times that of Krakatoa—would then be expected to reduce the sunlight to $\exp(-30) = 10^{-13}$. This is, of course, much more light attenuation than is needed to stop photosynthesis. But the model used in this simplistic calculation assumes that the dust is a perfect absorber of the incident light. A reasonable albedo coupled with a slight reduction in the mass of dust can raise the light intensity under the assumed "optical depth" to 10^{-7} of normal sunlight, corresponding to 10 percent of full moonlight.

Although it is impossible to make an accurate estimate of the asteroid's size from the Krakatoa extrapolation, it would have been necessary to abandon the hypothesis had a serious discrepancy been apparent. In the absence of good measurements of the solar constant in the 1880's, it can only be said that the fourth method leads to asteroid sizes that are consistent with the other three.

Until we understand the reasons for the factor of 10 difference in Ir content of the boundary clay between Denmark and Italy, we will be faced with different values for the asteroid diameter based on the first method. The "Danish diameter" is then 6.6 km × $10^{1/3}$ = 14 km. The second and third estimates are unchanged: the second does not involve measurements made on the boundary layer, and the third uses the thickness of the clay, which is only slightly greater in Denmark than in Italy. The fourth method is based on such an uncertain attenuation value, from Krakatoa, that it is not worth recalculating. We conclude that the data are consistent with an impacting asteroid with a diameter of about 10 ± 4 km.

Biological Effects

A temporary absence of sunlight would effectively shut off photosynthesis and thus attack food chains at their origins. In a general way the effects to be expected from such an event are what one sees in the paleontological record of the extinction.

The food chain in the open ocean is based on microscopic floating plants, such as the coccolith-producing algae, which show a nearly complete extinction. The animals at successively higher levels in this food chain were also very strongly affected, with nearly total extinction of the foraminifera and complete disappearance of the belemnites, ammonites, and marine reptiles.

A second food chain is based on land plants. Among these plants, existing individuals would die, or at least stop producing new growth, during an interval of darkness, but after light returned they would regenerate from seeds, spores, and existing root systems. However, the large herbivorous and carnivorous animals that were directly or indirectly dependent on this vegetation would become extinct. Russell (*2*) states that "no terrestrial vertebrate heavier than about 25 kg is known to have survived the extinctions." Many smaller terrestrial vertebrates did survive, including the ancestral mammals, and they may have been able to do this by feeding on insects and decaying vegetation.

The situation among shallow marine bottom-dwelling invertebrates is less clear; some groups became extinct and others survived. A possible base for a temporary food chain in this environment is nutrients originating from decaying land plants and animals and brought by rivers to the shallow marine waters.

We will not go further into this matter, but we refer the reader to the proceedings of the 1976 Ottawa meeting on the C-T extinctions. This volume reproduces an extensive discussion among the participants of what would happen if the sunlight were temporarily "turned off" (*5*, pp. 144–149). Those involved in the discussion seemed to agree that many aspects of the extinction pattern could be explained by this mechanism, although a number of puzzles remained.

We must note, finally, an aspect of the biological record that does not appear to be in accord with the asteroid impact hypothesis or with any sudden, violent mechanism. Extinction of the foraminifera and nannoplankton occurs within reversed geomagnetic polarity zone Gubbio G− in the Gubbio section (*30*). Butler and co-workers (*68*, *69*) have studied the nonmarine sequence of the San Juan Basin of New Mexico and have found a polarity sequence that appears to be correlated with the reversal sequence at Gubbio. In the San Juan Basin, the highest dinosaur fossils are found in the normal polarity zone (anomaly 29) that follows what is identified as the Gubbio G− zone. It would thus appear that the dinosaur and foram-nannoplankton extinctions were not synchronous. (Extinctions occurring in the same polarity zone in distant sections would not establish either synchroneity or diachroneity.) Three comments on the San Juan Basin work have been published (*70*) calling attention to the possibility of an unconformity at the boundary, in which case the correlation of the magnetic polarity zones could be in error and the extinctions might still be synchronous. Lindsay *et al.* (*69*) argue strongly against a major hiatus, but admit that "the case is not completely closed." Russell (*71*) has noted stratigraphic evidence against a diachronous extinction in the continental and marine realm.

Resolution of the question of whether the extinctions could have been synchronous will depend on further paleomagnetic studies. In the meantime we can state that the asteroid impact hypothesis predicts that the apparently diachronous timing of the foram-nannofossil and dinosaur extinctions will eventually be shown to be incorrect.

Problems in Boundary Clay Composition

One would expect from the simplest considerations of our hypothesis that the boundary layer resulted from crustal material (enriched in certain elements by the asteroidal matter) that was distributed worldwide in the stratosphere and then fell into the ocean. This material would be subjected to chemical and physical processes in the atmosphere and then in the ocean, which would alter the composition. The enhancements of metals having water-insoluble sulfides in the Danish C-T boundary compared to the Italian might be related to an anaerobic environment during deposition of the former and an aerobic one for the latter. Hydrogen sulfide can be produced by bacteria in oxygen-deficient waters, and this would precipitate those metals if they were available. This would not, however, explain the striking depletion of some trace elements in the Danish boundary or its very high Ir abundance. If chondritic Ir with an abundance of ~ 500 ppb were diluted 60-fold with crustal material, the Ir abundance should be ~ 8 ppb rather than the 65 ppb observed. Possible solutions to these difficulties may arise when better estimates of the extent of mixing of asteroidal and terrestrial material in the atmosphere are made, and when the boundary layer chemistry is studied at additional locations and a better understanding of the marine chemistry is achieved.

Implications

Among the many implications of the asteroid impact hypothesis, if it is correct, two stand out prominently. First, if the C-T extinctions were caused by an impact event, the same could be true of the earlier major extinctions as well. There have been five such extinctions since the end of the Precambrian, 570 million years ago, which matches well the probable interval of about 100 million years between collisions with 10-km-diameter objects. Discussions of these extinction events generally list the organisms affected according to taxonomic groupings; it would be more useful to have this information given in terms of interpreted ecological or food-chain groupings. It will also be important to carry out iridium analyses in complete stratigraphic sections across these other boundaries. However, E. Shoemaker (private communication) predicts that if some of the extinctions were caused by the collision of a "fresh" comet (mostly ice), the Ir anomaly would not be seen even though the extinction mechanism was via the same dust cloud of crustal material, so the absence of a higher Ir concentration at, for example, the Permian-Triassic boundary would not invalidate our hypothesis. According to Shoemaker, cometary collisions in this size range could be twice as frequent as asteroidal collisions.

Second, we would like to find the crater produced by the impacting object. Only three craters 100 km or more in diameter are known (*62*). Two of these (Sudbury and Vredefort) are of Precambrian age. For the other, Popigay Crater in Siberia, a stratigraphic age of Late Cretaceous to Quaternary and a potassium-argon date of 28.8 million years (no further details given) have been reported (*72, 73*). Thus, Popigay Crater is probably too young, and at 100-km-diameter probably also too small, to be the C-T impact site. There is about a 2/3 probability that the object fell in the ocean. Since the probable diameter of the object, 10 km, is twice the typical oceanic depth, a crater would be produced on the ocean bottom and pulverized rock could be ejected. However, in this event we are unlikely to find the crater, since bathymetric information is not sufficiently detailed and since a substantial portion of the pre-Tertiary ocean has been subducted.

References and Notes

1. D. A. Russell, *Geol. Assoc. Can. Spec. Rep. 13* (1975), p. 119.
2. ———, in (*5*), p. 11.
3. M. B. Cita and I. Premoli Silva, *Riv. Ital. Paleontol. Stratigr. Mem. 14* (1974), p. 193.
4. D. A. Russell, *Annu. Rev. Earth Planet. Sci.* 7, 163 (1979).
5. K-TEC group (P. Béland *et al.*), *Cretaceous-Tertiary Extinctions and Possible Terrestrial and Extraterrestrial Causes* (Proceedings of Workshop, National Museum of Natural Sciences, Ottawa, 16 and 17 Nov. 1976).
6. T. Birkelund and R. G. Bromley, Eds., *Cretaceous-Tertiary Boundary Events*, vol. 1, *The Maastrichtian and Danian of Denmark* (Symposium, University of Copenhagen, Copenhagen, 1979); W. K. Christiansen and T. Birkelund, Eds., *ibid.*, vol. 2, *Proceedings*.
7. H. Tappan, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 4, 187 (1968); T. R. Worsley, *Nature (London)* 230, 318 (1971); W. T. Holser, *ibid.* 267, 403 (1977); D. M. McLean, *Science* 200, 1060 (1978); *ibid.* 201, 401 (1978); S. Gartner and J. Keany, *Geology* 6, 708 (1978).
8. E. G. Kauffman, in (*6*), vol. 2, p. 29.
9. A. G. Fischer, in (*6*), vol. 2, p. 11; ——— and M. A. Arthur, *Soc. Econ. Paleontol. Mineral. Spec. Publ. 25* (1977), p. 19.
10. J. F. Simpson, *Geol. Soc. Am. Bull.* 77, 197 (1966); J. D. Hays, *ibid.* 82, 2433 (1971); C. G. A. Harrison and J. M. Prospero, *Nature (London)* 250, 563 (1974).
11. O. H. Schindewolf, *Neues Jahrb. Geol. Palaeontol. Monatsh.* 1954, 451 (1954); *ibid.* 1958, 270 (1958); A. R. Leoblich, Jr., and H. Tappan, *Geol. Soc. Am. Bull.* 75, 367 (1964); V. I. Krasovski and I. S. Shklovsky, *Dokl. Akad. Nauk SSSR* 116, 197 (1957); K. D. Terry and W. H. Tucker, *Science* 159, 421 (1968); H. Laster, *ibid.* 160, 1138 (1968); W. H. Tucker and K. D. Terry, *ibid.*, p. 1138; D. Russell and W. H. Tucker, *Nature (London)* 229, 553 (1971); M. A. Ruderman, *Science* 184, 1079 (1974); R. C. Whitten, J. Cuzzi, W. J. Borucki, J. H. Wolfe, *Nature (London)* 263, 398 (1976).
12. S. Gartner and J. P. McGuirk, *Science* 206, 1272 (1979).
13. A. Boersma and N. Schackleton, in (*6*), vol. 2, p. 50; B. Buchardt and N. O. Jorgensen, in (*6*), vol. 2, p. 54.
14. L. Christensen, S. Fregerslev, A. Simonsen, J. Thiede, *Bull. Geol. Soc. Den.* 22, 193 (1973).
15. N. O. Jorgensen, in (*6*), vol. 1, p. 33, vol. 2, p. 62; M. Renard, in (*6*), vol. 2, p. 70.
16. H. P. Luterbacher and I. Premoli Silva, *Riv. Ital. Paleontol. Stratigr.* 70, 67 (1964).
17. H. Pettersson and H. Rotschi, *Geochim. Cosmochim. Acta* 2, 81 (1952).
18. V. M. Goldschmidt, *Geochemistry* (Oxford Univ. Press, New York, 1954).
19. J. L. Barker, Jr., and E. Anders, *Geochim. Cosmochim. Acta* 32, 627 (1968).
20. R. Ganapathy, D. E. Brownlee, P. W. Hodge, *Science* 201, 1119 (1978).
21. A. M. Sarna-Wojcicki, H. R. Bowman, D. Marchand, E. Helley, private communication.
22. J. H. Crocket and H. Y. Kuo, *Geochim. Cosmochim. Acta* 43, 831 (1979).
23. These are briefly discussed in L. W. Alvarez, W. Alvarez, F. Asaro, H. V. Michel, *Univ. Calif. Lawrence Berkeley Lab. Rep. LBL-9666* (1979).
24. A description of the NAA techniques is given in Alvarez *et al.* (*23*), appendix II; I. Perlman and F. Asaro, in *Science and Archaeology*, R. H. Brill, Ed. (MIT Press, Cambridge, Mass., 1971), p. 182.
25. These limestones belong to the Umbrian sequence, of Jurassic to Miocene age, which has been described in V. Bortolotti, P. Passerini, M. Sagri, G. Sestini, *Sediment. Geol.* 4, 341 (1970); A. Jacobacci, E. Centamore, M. Chiocchini, N. Malferrari, G. Martelli, A. Micarelli, *Note Esplicative Carta Geologica d'Italia (1:50,000), Foglio 190: "Cagli"* (Rome, 1974).
26. H. P. Luterbacher and I. Premoli Silva, *Riv. Ital. Paleontol. Stratigr.* 68, 253 (1962); I. Premoli Silva, L. Paggi, S. Monechi, *Mem. Soc. Geol. Ital.* 15, 21 (1976).
27. S. Monechi, in (*6*), vol. 2, p. 164.
28. D. V. Kent, *Geology* 5, 769 (1977); M. A. Arthur, thesis, Princeton University (1979).
29. O. Renz, *Eclogae Geol. Helv.* 29, 1 (1936); *Serv. Geol. Ital. Mem. Descr. Carta Geol. Ital.* 29, 1 (1936).
30. M. A. Arthur and A. G. Fischer, *Geol. Soc. Am. Bull.* 88, 367 (1977); I. Premoli Silva, *ibid.*, p. 371; W. Lowrie and W. Alvarez, *ibid.*, p. 374; W. M. Roggenthen and G. Napoleone, *ibid.*, p. 378; W. Alvarez, M. A. Arthur, A. G. Fischer, W. Lowrie, G. Napoleone, I. Premoli Silva, W. M. Roggenthen, *ibid.*, p. 383; W. Lowrie and W. Alvarez, *Geophys. J. R. Astron. Soc.* 51, 561 (1977); W. Alvarez and W. Lowrie, *ibid.* 55, 1 (1978).
31. Locations of the sections studied are: (i) Bottaccione Gorge at Gubbio: 43°21.9'N, 12°35.0'E (0°7.9' east of Rome); (ii) Contessa Valley, 3 km northwest of Gubbio: 43°22.6'N, 12°33.7'E (0°6.6' east of Rome); (iii) Petriccio suspension bridge, 2.3 km west-southwest of Acqualagna: 43°36.7'N, 12°38.7'E (0°11.6' east of Rome); (iv) Acqualagna, road cut 0.8 km southeast of town: 43°36.7'N, 12°40.8'E (0°13.7' east of Rome); and (v) Gorgo a Cerbara: 43°36.1'N, 12°33.6'E (0°6.5' east of Rome). We thank E. Sannipoli, W. S. Leith, and S. Marshak for help in sampling these sections.
32. J. H. Crocket, J. D. McDougall, R. C. Harriss, *Geochim. Cosmochim. Acta* 37, 2547 (1973).
33. Location: 55°16.7'N, 12°26.5'E. We thank I. Bank and S. Gregersen for taking W.A. to this locality.
34. A. Rosenkrantz and H. W. Rasmussen, *Guide to Excursions A42 and C37* (21st International

Geological Congress, Copenhagen, 1960), part I, pp. 1-17.
35. F. Surlyk, in (6), vol. 1, p. 164.
36. C. Heinberg, in (6), vol. 1, p. 58.
37. H. W. Rasmussen, in (6), vol. 1, p. 65; P. Gravesen, in (6), vol. 1, p. 72; U. Asgaard, in (6), vol. 1, p. 74.
38. E. Hakansson and E. Thomsen, in (6), vol. 1, p. 78.
39. S. Floris, in (6), vol. 1, p. 92.
40. I. Bang, in (6), vol. 1, p. 108.
41. K. Perch-Nielsen, in (6), vol. 1, p. 115.
42. Soft x-ray fluorescence measurements for major element determinations were made by S. Flexser and M. Sturz of Lawrence Berkeley Laboratory.
43. Hard x-ray fluorescence measurements for trace element determinations were made by R. D. Giauque of Lawrence Berkeley Laboratory.
44. F. G. Walton Smith, Ed. *CRC Handbook of Marine Science* (CRC Press, Cleveland, 1974), vol. 1, p. 11.
45. H. A. Wollenberg *et al.*, *Univ. Calif. Lawrence Berkeley Lab. Rep. LBL-7092*, revised 1980.
46. *Encyclopaedia Britannica* (Benton, Chicago, ed. 15, 1974), vol. 6, p. 702.
47. K. K. Turekian, *Oceans* (Prentice-Hall, Englewood Cliffs, N.J., 1976), p. 122.
48. J. H. Crocket, *Can. Mineral.* 17, 391 (1979); J. R. Ross and R. Keays, *ibid.*, p. 417.
49. I. S. Shklovsky, *Supernovae* (Wiley, New York, 1968), p. 377.
50. D. D. Clayton, *Principles of Stellar Evolution and Nucleosynthesis* (McGraw-Hill, New York, 1968), pp. 546-606.
51. D. N. Schramm, *Annu. Rev. Astron. Astrophys.* 12, 389 (1974).
52. C. F. McKee, personal communication.
53. These observations were reported at the American Geophysical Union meeting in May 1979 and at the Geological Society of America meeting in November 1979 [W. Alvarez, L. W. Alvarez, F. Asaro, H. V. Michel, *Eos* 60, 734 (1979); *Geol. Soc. Am. Abstr. Programs* 11, 350 (1979)].
54. W. M. Napier and S. V. M. Clube, *Nature (London)* 282, 455 (1979).
55. H. C. Urey, *ibid.* 242, 32 (1973).
56. E. J. Öpik, *Ir. Astron. J.* 5 (No. 1), 34 (1958).
57. E. M. Shoemaker, J. G. Williams, E. F. Helin, R. F. Wolfe, in *Asteroids*, T. Gehrels, Ed. (Univ. of Arizona Press, Tucson, 1979), pp. 253-282.
58. E. J. Öpik, *Adv. Astron. Astrophys.* 2, 220 (1963); *ibid.* 4, 302 (1964); *ibid.* 8, 108 (1971). These review articles give references to Öpik's extensive bibliography on meteorites, Apollo objects, and asteroids.
59. C. R. Chapman, J. G. Williams, W. K. Hartmann, *Annu. Rev. Astron. Astrophys.* 16, 33 (1978).
60. G. W. Wetherill, *Sci. Am.* 240 (No. 3), 54 (1979).
61. E. M. Shoemaker, personal communication.
62. R. A. F. Grieve and P. B. Robertson, *Icarus* 38, 212 (1979).
63. R. A. F. Grieve, personal communication.
64. G. J. Symons, Ed., *The Eruption of Krakatoa and Subsequent Phenomena* (Report of the Krakatoa Committee of the Royal Society, Harrison, London, 1888).
65. I. U. Olson and I. Karlen, *Am. J. Sci. Radiocarbon Suppl.* 7 (1965), p. 331; T. A. Rafter and B. J. O'Brien, *Proc. 8th Int. Conf. Radiocarbon Dating* 1, 241 (1972).
66. U. Krähenbühl, *Geochim. Cosmochim. Acta* 37, 1353 (1973).
67. B. Mason, *Space Sci. Rev.* 1, 621 (1962-1963).
68. R. F. Butler, E. H. Lindsay, L. L. Jacobs, N. M. Johnson, *Nature (London)* 267, 318 (1977); E. H. Lindsay, L. L. Jacobs, R. F. Butler, *Geology* 6, 425 (1978).
69. E. H. Lindsay, R. F. Butler, N. M. Johnson, in preparation.
70. W. Alvarez and D. W. Vann, *Geology* 7, 66 (1979); J. E. Fassett, *ibid.*, p. 69; S. G. Lucas and J. K. Rigby, Jr., *ibid.*, p. 323.
71. D. A. Russell, *Episodes 1979 No. 4* (1979), p. 21.
72. V. L. Masaytis, M. V. Mikhaylov, T. V. Selivanovskaya, *Sov. Geol. No.* 6 (1971), pp. 143-147; translated in *Geol. Rev.* 14, 327 (1972).
73. V. L. Masaytis, *Sov. Geol. No. 11* (1975), pp. 52-64; translated in *Int. Geol. Rev.* 18, 1249 (1976).
74. It will be obvious to anyone reading this article that we have benefited enormously from conversations and correspondence with many friends and colleagues throughout the scientific community. We would particularly like to acknowledge the help we have received from E. Anders, J. R. Arnold, M. A. Arthur, A. Buffington, I. S. E. Carmichael, G. Curtis, P. Eberhard, S. Gartner, R. L. Garwin, R. A. F. Grieve, E. K. Hyde, W. Lowrie, C. McKee, M. C. Michel (who was responsible for the mass spectrometric measurements), J. Neil, B. M. Oliver, C. Orth, B. Pardoe, I. Perlman, D. A. Russell, A. M. Sessler, and E. Shoemaker. One of us (W.A.) thanks the National Science Foundation for support, the other three authors thank the Department of Energy for support, and one of us (L.W.A.) thanks the National Aeronautics and Space Administration for support. The x-ray fluorescence measurements of trace elements Fe and Ti by R. D. Giauque and of major elements by S. Flexser and M. Sturz were most appreciated. We appreciate the assistance of D. Jackson and C. Nguyen in the sample preparation procedures. We are grateful to T. Lim and the staff of the Berkeley Research Reactor for many neutron irradiations used in this work. We also appreciate the efforts of G. Pefley and the staff of the Livermore Pool Type Reactor for the irradiations used for the Ir isotopic ratio measurements.

No-Tillage Agriculture

Ronald E. Phillips, Robert L. Blevins, Grant W. Thomas
Wilbur W. Frye, Shirley H. Phillips

For over 100 years, agriculture has relied upon the moldboard plow and disk harrow to prepare soil to produce food. Without the moldboard plow and disk it would not have been possible to control weeds and to obtain the yields necessary to provide favorable economic returns from agriculture. Weeds are strong competitors with food crops for water and plant nutrients, and it was not until plant growth regulators were introduced in the late 1940's that attention was turned to no-tillage agriculture. From plant growth regulators selective herbicides were developed, and these increased the feasibility of growing many different crops without tilling the soil (*1-3*).

In this article, we define conventional tillage as moldboard plowing followed by disking one or more times. By this method one obtains a loose, friable seedbed in the surface 10 centimeters of soil. We define the no-tillage system (*4*) as one in which the crop is planted either entirely without tillage or with just sufficient tillage to allow placement and coverage of the seed with soil to allow it to germinate and emerge. Usually no further cultivation is done before harvesting. Weeds and other competing vegetation are controlled by chemical herbicides. Soil amendments, such as lime and fertilizer, are applied to the soil surface.

In pasture management, chemicals are used as a substitute for tillage, herbicides being used to restrict growth and competition of undesirable plants during the establishment of the newly seeded crop or to supress growth of grasses and allow establishment of legumes. Row crop production with the no-tillage system is almost always carried out by planting the crop into soil covered by a chemically killed grass sod or with dead plant residues of a previous crop. For example, in continuous no-tillage corn (*Zea mays*) production, the soil surface at the time of planting is covered with corn stalk residues of the previous corn crop. In double-cropped soybeans, the soil at the time of planting is covered with residues of a recently harvested small-grain crop such as barley or wheat.

The land area used for row crops and forage crops grown by the no-tillage system has increased rapidly during the past 15 years. In 1974, the U.S. Department of Agriculture (*5*) estimated that the amount of cropland in the United States under no-tillage cultivation was 2.23 million hectares, and that 62 million hectares or 45 percent of the total U.S. cropland (*6*) will be under the no-tillage system by 2000. An estimated 65 percent of the seven major annual crops (corn, soybeans, sorghum, wheat, oats, barley, and rye) will be grown by the no-tillage system by the year 2000 and 78 percent by the year 2010 (*5*). In Kentucky there were 44,000, 160,400, and 220,000 ha of no-tillage corn and soybeans grown in 1969, 1972, and 1978, respectively.

R. E. Phillips and G. W. Thomas are professors and R. L. Blevins and W. W. Frye are associate professors in the Department of Agronomy, College of Agriculture, University of Kentucky, Lexington 40546. S. H. Phillips is assistant director of the Cooperative Extension Service for Agriculture, Lexington 40546.

0036-8075/80/0606-1108$01.50/0

Raup, D.M. & Sepkoski, J.J.
6. "Mass extinctions in the marine fossil record"
Science
Published by A.A.A.S
Used by permission

hukla, D. Randall, D. Straus, Y. Sud, L.
i, *Winter and Summer Simulations with the*
S Climate Model (National Aeronautics
Space Administration/Goddard Space Flight
er, Greenbelt, Md., in press).
harney, W. J. Quirk, S. H. Chow, J. Korn-
, *J. Atmos. Sci.* 34, 1366 (1977).

iber 1981

Mass Extinctions in the Marine Fossil Record

Abstract. *A new compilation of fossil data on invertebrate and vertebrate families indicates that four mass extinctions in the marine realm are statistically distinct from background extinction levels. These four occurred late in the Ordovician, Permian, Triassic, and Cretaceous periods. A fifth extinction event in the Devonian stands out from the background but is not statistically significant in these data. Background extinction rates appear to have declined since Cambrian time, which is consistent with the prediction that optimization of fitness should increase through evolutionary time.*

A number of mass extinctions have "reset" major parts of the evolutionary system during the Phanerozoic. However, the precise timing and magnitude of these events has been difficult to measure because data from the fossil record are fragmentary. Comprehensive and accurate data on extinct species have always been unobtainable, and therefore most workers have been forced to investigate extinctions at the level of genera, families, and orders, with family-level data generally preferred as the best compromise between sampling limitations and taxonomic uncertainty (*1*). Historically, the three best summaries of familial data from the fossil record have been those of Newell (*2*), Cutbill and Funnell (*3*), and Valentine (*4*). But even with these data sets, identification of specific mass extinctions has been difficult and often subjective because of taxonomic problems and especially stratigraphic imprecision. Many macroevolutionary phenomena including mass extinctions have characteristic time scales that are geologically rather short (less than several tens of millions years) and can become lost or grossly distorted when analyzed without adequate stratigraphic control.

We now present a new analysis of extinctions based on a more comprehensive and accurate data set for marine animal families. Marine vertebrates as well as invertebrates and protozoans are included, and the data benefit from compilation of taxonomic and stratigraphic investigations far beyond traditional sources (*5*). The compilation encompasses approximately 3300 fossil marine families, of which about 2400 are extinct. Times of extinction for 87 percent of the families have been resolved to the level of the stratigraphic stage (mean duration, 7.4×10^6 years), and most of the remaining data has been resolved to stratigraphic series (mean duration, 20×10^6 years).

The rates of extinction calculated from the familial data plotted against geologic time are illustrated in Fig. 1. Each point was calculated as follows: the number of families that became extinct in each of the 76 post-Tommotian (early Lower Cambrian) stages (*6*) was divided by the estimated duration of the stage (*7*); these initial rates were then modified by adding extinction rates calculated from the lower resolution series-level data to the appropriate stages. Calculations were made separately for "shelly" taxa and for rarely preserved taxa (*8*). The effect of this segregation was negligible in most cases so that the data for rarely preserved animals are not included with most points in Fig. 1. For four stages, however, addition of rarely preserved families increased calculated extinction rates by more than 0.5 family per million years. These are the stages that contain the four major Lagerstätten of the Phanerozoic marine record: Burgess Shale (Cambrian, Templetonian), Hunsrück Shale (Devonian, Siegenian), Mazon Creek concretions (Carboniferous, Moscovian), and Solnhofen Limestone (Jurassic, "Tithonian"). The combined rates for shelly and rarely preserved families for these four stages are indicated in Fig. 1 by X's with the rates for shelly families shown below. Only the Burgess Shale (Templetonian) stands out on the plot.

The distribution of the 76 points for shelly animals in Fig. 1 suggests that two rates of extinction have been operative through the Phanerozoic. (i) Normal, or background, extinction: the majority of points fall in a rather tight cluster at extinction rates less than 8.0 extinctions per million years. (ii) Mass extinction: several points stand out as being considerably higher than the background and show a maximum of 19.3 familial extinctions per million years.

The problem of determining rigorously which points in Fig. 1 should be considered mass extinctions can be approached as a simple data analysis problem of identifying trends and outliers. As an initial step, we computed a linear regression (not shown) for all 76 extinction points as a function of geologic time and then searched for significant departures from this line. Four points (or 5 percent of the data) fell above the one-sided 99 percent confidence interval. These points, which are circled in Fig. 1, are (per million years) the Ashgillian (19.3 fm), Guadalupian (14.0 fm), Dzhulfian (15.7 fm), and Maestrichtian (16.3 fm). A fifth point, the Norian (10.8 fm), fell

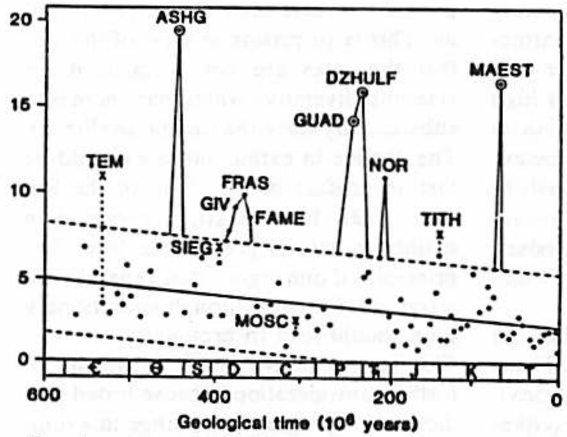


Fig. 1. Total extinction rate (extinctions per million years) through time for families of marine invertebrates and vertebrates. The plot shows statistically significant mass extinctions late in the Ordovician (ASHG), Permian (GUAD-DZHULF), Triassic (NOR), and Cretaceous (MAEST). An extinction event in the late Devonian (GIV-FRAS-FAME) is noticeable but not statistically significant. Circled points are those where the departure from the main cluster is highly significant ($P < .01$); X's indicate those cases where inclusion of rarely preserved animal groups substantially increases the calculated extinction rate (the point directly below the X is the rate calculated without the rarely preserved groups). The figure also shows a general decline in background extinction rate through time. The regression line is fit to the 67 points having extinction rates less than eight families per 10^6 years, and the dashed lines define the 95 percent confidence band for the regression. Abbreviations: TEM, Templetonian; ASHG, Ashgillian; SIEG, Siegenian; GIV, Givetian; FRAS, Frasnian; FAME, Famennian; MOSC, Moscovian; GUAD, Guadalupian; DZHULF, Dzhulfian; NOR, Norian; TITH, Tithonian; MAEST, Maestrichtian.

0036-8075/82/0319-1501$01.00/0

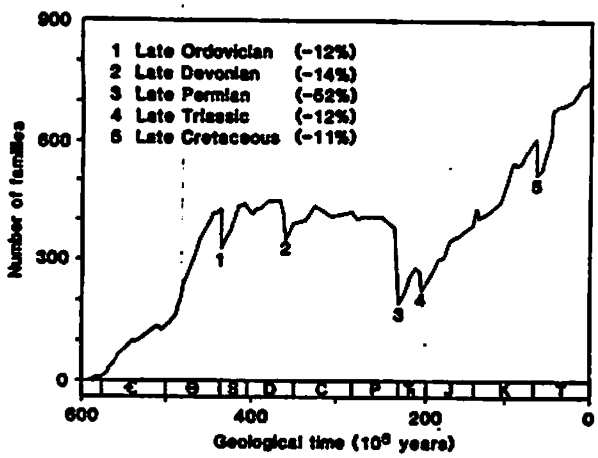


Fig. 2. Standing diversity through time for families of marine vertebrates and invertebrates. Rarely preserved groups are not included. Five mass extinctions, indicated by numerals, are recognizable by abrupt drops in the diversity curve. The relative magnitudes of these drops (measured from the stage before to the stage after the extinction event) are given in parentheses in the upper left. All mass extinctions but No. 2 (Devonian) are statistically significant in Fig. 1 and three (Nos. 1, 3, and 5) are highly significant ($P < .01$).

above the one-sided 95 percent confidence interval. Some or all of these mass extinctions have been recognized previously but without consistency and without statistical testing.

The outlying points identified above also can be recognized as major perturbations in marine diversity. The Phanerozoic diversity curve compiled from the familial data is shown in Fig. 2. Five extinction events are seen as sharp drops in standing diversity. Four of these (counting the Guadalupian and Dzhulfian as a single event) match the statistically significant outliers in Fig. 1. The fifth, labeled "2" in Fig. 2, is a late Devonian extinction that has been recognized by previous workers. This extinction does not appear as a statistically significant event in Fig. 1 because the family extinctions are distributed over two stages, the Frasnian and the preceding Givetian, which have a combined duration of about 15 million years (*9*). This smearing of extinctions may represent sampling error in that failure to identify the actual time of extinction will almost always push apparent extinctions backward in time. Alternatively, the smearing may reflect a real phenomenon—an extinction "event" that took place over millions of years. The continuation of high extinction rates into the Famennian is consistent with this hypothesis. However, it should be noted that, on the basis of other information, McLaren (*10*) suggested a meteorite impact as one possible explanation for the Frasnian extinctions.

In summary, five mass extinctions are clearly defined in the familial data. These extinctions occurred in the Late Ordovician (Ashgillian), Late Devonian (Givetian-Frasnian), Late Permian (Guadalupian-Dzhulfian), Late Triassic (Norian), and Late Cretaceous (Maestrichtian). The occurrence of these major extinctions near the ends of geologic periods simply reflects the fact that the stratigraphers who established the geologic time scale in the first half of the 19th century chose major faunal breaks as boundaries for the principal subdivisions.

With the major Phanerozoic events isolated in Fig. 1, a more accurate assessment of the nature of background extinction can be made. Although some smaller but well-known extinction events may remain hidden in Fig. 1 (*11*), the residual cluster of points suggests that background rates have been declining since the early Paleozoic. The solid line in Fig. 1 is a linear regression fitted to the 67 extinction rates for shelly animals after removal of the major extinction events; the dashed lines, which envelop nearly all these points, represent the 95 percent confidence band for the regression. The correlation coefficient for the regression is .47, which can be considered statistically significant if problems of time series and data selection are ignored. The slope of the regression line is nontrivial and indicates that the total rate of background extinction has decreased from about 4.6 to 2.0 fm per million years since the Early Cambrian. This is surprising in view of the fact that the rates are not normalized for standing diversity, which has increased substantially since the Cambrian (Fig. 2). The decline in extinction rates could be just an artifact of the "pull of the Recent" (*12*). In contrast, a decrease in extinction rate is predictable from first principles if one argues that general optimization of fitness through evolutionary time should lead to prolonged survival. This is speculative but it is worthy of further consideration because broad predictions of progressive change in evolutionary dynamics are so rarely realized when tested with data.

The decline in background extinction rate from the Early Cambrian to the Recent means that approximately 710 family extinctions did not occur that would have if the Cambrian rate had been sustained. This number is essentially identical to the amount by which familial diversity increased over that interval (680 families) (Fig. 2). This suggests that the net increase in standing diversity through the Phanerozoic may have been more an effect of decrease in extinction than increase in origination.

In conclusion, our analysis shows that major mass extinctions are far more distinct from background extinction than has been indicated by previous analyses of other data sets. Four mass extinctions are statistically significant events and are likely to represent phenomena qualitatively different from the background. The data do not tell us, of course, what stresses caused the mass extinctions. The extinctions were short-lived events in geological time, but the data do not have the resolving power to show whether the events were also short-lived in human or ecological time.

DAVID M. RAUP
Field Museum of Natural History,
Chicago, Illinois 60605

J. JOHN SEPKOSKI, JR.
Department of Geophysical Sciences,
University of Chicago,
Chicago 60637

References and Notes

1. J. W. Valentine, *J. Paleontol.* 48, 549 (1974).
2. N. D. Newell, *Sci. Am.* 208, 77 (1963); *Geol. Soc. Am. Spec. Pap.* 89, 63 (1967).
3. J. L. Cutbill and B. M. Funnel, in *The Fossil Record*, W. B. Harland *et al.*, Eds. (Geological Society of London, London, 1967), p. 791.
4. J. W. Valentine, *Palaeontology* 12, 684 (1969).
5. The data were compiled from R. C. Moore *et al.*, Eds. *Treatise on Invertebrate Paleontology* (Geological Society of America and Univ. of Kansas Press, Lawrence, 1953–1979); W. B. Harland *et al.*, Eds., *The Fossil Record* (Geological Society of London, London, 1967); A. S. Romer, *Vertebrate Paleontology* (Univ. of Chicago Press, Chicago, 1966); also, 380 additional papers and monographs. A complete listing of these data is scheduled to appear in J. J. Sepkoski, Jr., *Milwaukee Pub. Mus. Contrib. Biol. Geol.*, in press.
6. Vendian and Tommotian points were excluded from the analysis because these intervals have exceptionally low diversities and therefore exceptionally low extinction rates.
7. The geologic time scale used is a composite based on a number of recently published stage-level time scales and differs only slightly from that used by J. J. Sepkoski, Jr., *Paleobiology* 5, 222 (1979).
8. Four general kinds of animals were considered to have low fossilization potential: (i) soft-bodied animals without mineralized skeletons (for example: Nemertina, Priapulida, and Sipunculida as well as many Hydrozoa, Scyphozoa, and Polychaeta); (ii) animals with lightly sclerotized skeletons (such as many Crustacea); (iii) rarely reported fossil animals with multielement skeletons that dissociate rapidly after death (such as Octocorallia and Holothuroidea as well as some Asterozoa and Osteichthyes); and (iv) deep-sea animals with extremely poor fossil records (such as some Crinoidea, Chondrichthyes, and Osteichthyes).
9. The Frasnian might still be considered statistically distinct in Fig. 1. If each point in that figure is considered an independent event, then the probability that three of the nine highest points would be clustered about the Frasnian point is quite low ($P = .002$).
10. D. J. McLaren, *J. Paleontol.* 44, 801 (1970).
11. "Minor" mass extinctions, which do not appear as noticeable perturbations in Figs. 1 and 2,

include at least three to five events in the Cambrian [A. R. Palmer, *J. Paleontol.* 39, 149 (1965); M. E. Taylor, *ibid.* 42, 1319 (1968); J. H. Stitt, *ibid.* 45, 178 (1971)], an Early Jurassic (Toarcian) event [A. Hallam, *Paleobiology* 3, 58 (1977)], a terminal Eocene event [H. Tappan and A. R. Loeblich, *Geol. Soc. Am. Spec. Pap.* 127, 247 (1971); A. G. Fischer and M. A. Arthur, *Soc. Econ. Paleontol. Mineral* 25, 19 (1977)], and possible events in the latest Jurassic (Tithonian) and early Late Cretaceous (Cenomanian) (see A. G. Fisher and M. A. Arthur, above).

12. D. M. Raup, *Carnegie Mus. Nat. Hist. Bull.* 13, 85 (1979).
13. We thank R. K. Bambach for critical comments on this work.

28 September 1981; revised 31 November 1981

Lumber Spill in Central California Waters: Implications for Oil Spills and Sea Otters

Abstract. *A large quantity of lumber was spilled in the ocean off central California during the winter of 1978, and it spread through most of the range of the threatened California sea otter population within 4 weeks. The movement rates of lumber were similar to those of oil slicks observed elsewhere. These observations indicate that a major oil spill could expose significant numbers of California sea otters to oil contamination.*

The California population of the sea otter [*Enhydra lutris nereis* (Merriam)] was listed as "threatened" in 1977 (*1*) pursuant to the Endangered Species Act of 1973 (*2*). The listing was based on the possibility that a major oil spill could occur within the sea otter range and could kill a significant portion of the population, placing it in danger of extinction. This concern arises from the known sensitivity of the species to oil contamination (*3*). A spill of gasoline and diesel oil nearshore in the Kurile Islands, U.S.S.R., spread through 40 km of coastline and killed over 100 sea otters (*4*). Concern for the status of the California sea otter is heightened by the lack of evidence of significant population growth since 1973 (*5*).

It is difficult to project the critical day-to-day movements of floating oil near the sea otter range on the basis of existing oceanographic data. Surface current patterns off central California (San Francisco to Point Conception) have been examined with several techniques (*6–9*). The principal result is the description of mean flow patterns on a seasonal scale. However, studies of drogues and remote imagery have shown that short-term departures from mean seasonal drift may be frequent in the California current system (*6, 7, 10, 11*). Such departures involve tidal oscillations and mesoscale meanders and eddies (*6, 7, 10, 11*). The prediction of the direction of drift of floating oil is further complicated by the dominant role of wind stress at the air-sea interface (*12*). As far as we know, there are no records of major oil spills off central California on which to base predictions of oil drift. We know of no published studies of day-to-day movements of other floating materials off central California over an appropriately small time scale.

In this report we describe the movements and beaching of a large volume of lumber spilled off central California in the winter of 1978. Floating materials such as drift cards and plastic sheets have been used successfully by others in modeling the movements of oil on the sea surface (*13*). Our data provide a first approximation of the disposition of the floating component of a large oil spill occurring under similar conditions of weather and sea. Information of this kind is needed if we are to understand the potential impacts of oil spills on the California sea otter population and to develop management plans for improving the status of the population, now numbering about 1800 animals (*5*).

A cargo of 2×10^6 board feet of finished lumber (volume equivalent to 2.9×10^4 barrels of oil) spilled from a barge under tow in heavy weather 40 km west of Point Sur on 12 February 1978 (*14*). The spilled lumber was hazardous to navigation (*15*) and was therefore monitored by aircraft and merchant vessels in subsequent weeks (*16*). We compiled a record of observations of the floating lumber at sea along with sightings of lumber washed ashore after the spill.

Much of the spilled lumber remained in a single large patch that moved first toward the coast and then southeastward, parallel to the shoreline, during the first 10 days after the spill (Fig. 1). By 24 February the major patch was within 7 km of shore near Point Estero and remained relatively close to shore for the balance of the observation period (through March). Other patches of floating lumber were seen off Monterey, Point Lobos, Cape San Martin, and Point Arguello during the survey period. Beached lumber was found throughout two sections of coastline within the sea otter range, a northern section of about

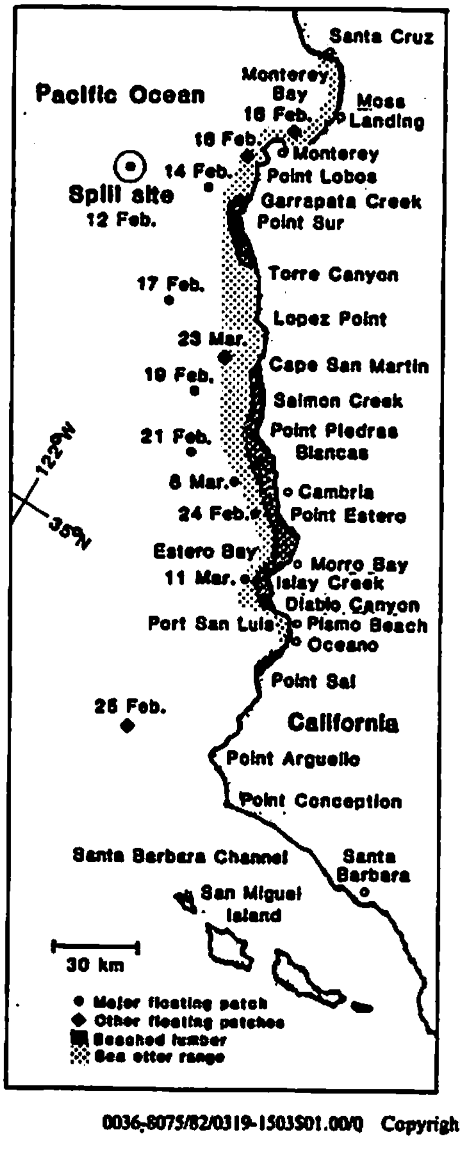


Fig. 1. Sightings of the major patch and smaller patches of floating lumber and areas of significant lumber beaching in and near the range of the sea otter population in California. Observations were made after the spillage of 2×10^6 board feet of lumber off Point Sur on 12 February 1978. All sightings of beached lumber were recorded between 12 February and 31 March 1978. The times and the positions corresponding to each sighting of the major patch of floating lumber are as follows: 12 February, 0650 GMT 36°15'N, 122°25'W (spill site); 14 February, 1930 GMT, 36°19'N, 122°07'W; 17 February, 0130 GMT, 35°55'N, 121°51'W; 19 February, 2030 GMT, 35°40'N, 121°37'W; 21 February, 0136 GMT, 35°29'N, 121°29'W; 24 February, 2100 GMT, 35°27'N, 121°05'W; 8 March, 1816 GMT, 35°30'N, 121°16'W; 11 March, 2006 GMT, 35°18'N, 120°58'W. Mean wind speed (meters per second) and the direction (relative to true north) between sightings of the major floating patch, during the period when the patch was closest to shore, were as follows: 21 to 24 February, 3.57, 317°; 24 February to 8 March, 0.90, 154°; 8 to 11 March, 4.30, 308° [wind data are from the shore station at Point Piedras Blancas (*18, 19*)].

0036-8075/82/0319-1503$01.00/0

S.J. Gould
File of Letters
Used by permission

November 23, 1982

Luis W. Alvarez
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Dear Dr. Alvarez,

Many thanks for sending me a copy of your fascinating talk of last April. It arrived just in time to update my now annual lecture on the Alvarez' hypothesis for the Cretaceous extinction.

If I may make one tiny factual correction, the last passenger pigeon died in 1913 -- so that unfortunate extirpation extended into our century. I sense your frustration at the methodological weakness of many of my colleagues and can only apologize for a general professional naivete, but also assure you that we are not all so unappreciative of proper analytical methods. Unfortunately, there is a lamentable but historically based tendency for paleontologists to have a very naive and literal view of the nature of empirical evidence (this is particularly strange since we are trained from professional infancy to regard the geological record as extremely imperfect and to avoid the necessity for all manner of catastrophy thereby). Geologists have traditionally been the great empirical field men -- suspicious of all effete laboratory calculation and mathematical formalization. This silly attitude has unfortunately spilled over into an age that should have rendered it extinct. Thus, when a paleontologist sees his last dinosaur bone he tends to interpret it as the truly final occurrence. You, of course,(and I concur entirely in your analysis), see the last empirical occurrence as a sample and ask the obvious question: given this last bit of actual evidence, what is the most probable extent above. I think that with the work of Dave Raup and others a more analytical spirit is beginning to pervade paleontology -- but it will take a generation or more to spread widely and, until then, I am afraid you will have to put up with what must be an infuriating lack of methodological sophistication.

For all the praise of the last paragraph, however, I do have one general criticism, or rather query. I think that you have made an excellent if not overwhelming case for your impact and cannot imagine any explanation of the Cretaceous extinction that will not include it as a predominant factor. But I do not understand why you state or at least imply that the whole facies of events that constitute the Cretaceous extinction must be attributed to this single cause -- especially when your asteroid can have as much impact (intellectual I mean) under the synergistic view that I have tried to suggest in the two published comments.I have made on the subject. I accept all you say about the coincidence of dinosaur, plant, and marine extinctions. But why do you not comment apon the extensive documentation of invertebrate paleontology that many groups meeting their end in the extinction were in decline for millions of years before — ammonites, and various groups of clams for example. This decline does not in the least compromise the essential role of your asteroid, but merely explains why its impact could have such a profound effect. That is, when asteroids or other impacting bodies hit at a time when the biosphere is already weakened for other reasons then a profound extinction can ensue; but, if a body hits the earth during diverse and stable biological times, then perhaps we have a smaller extinction (your Eocene-Oligocene event perhaps). This notion doesn't reduce your asteroidsto anything like a coup de grace because, without it, nothing much would have happened to the weakened Cretaceous biosphere and dinosaurs would probably still rule the earth, and I sure wouldn't be sitting in front of my dictating machine tonight. I think such a perspective would sooth paleontological sensibilities and still give your asteroid all its glory.

In any case thanks so much for making this fundamental paleontological subject a science and not a field for endless and fruitless speculation. Your theme has pleased me so much that I toot your trumpet in many places, not only before my class but also (and even en francais) et the College de France last April.

Sincerely,

Stephen Jay Gould
Professor of Geology

/ap

Lawrence Berkeley Laboratory

University of California
Berkeley, California 94720
Telephone 415/486-4000
FTS: 451-4000

December 2, 1982

Dr. Stephen Jay Gould
Museum of Comparative Zoology
The Agassiz Museum
Harvard University
Cambridge, MA 02138

Dear Steve:

It was good of you to write such a long supportive letter concerning my NAS talk. I well remember our earlier correspondence, and the fine article you wrote about our original work--you were the first person outside our little group who took our work seriously.

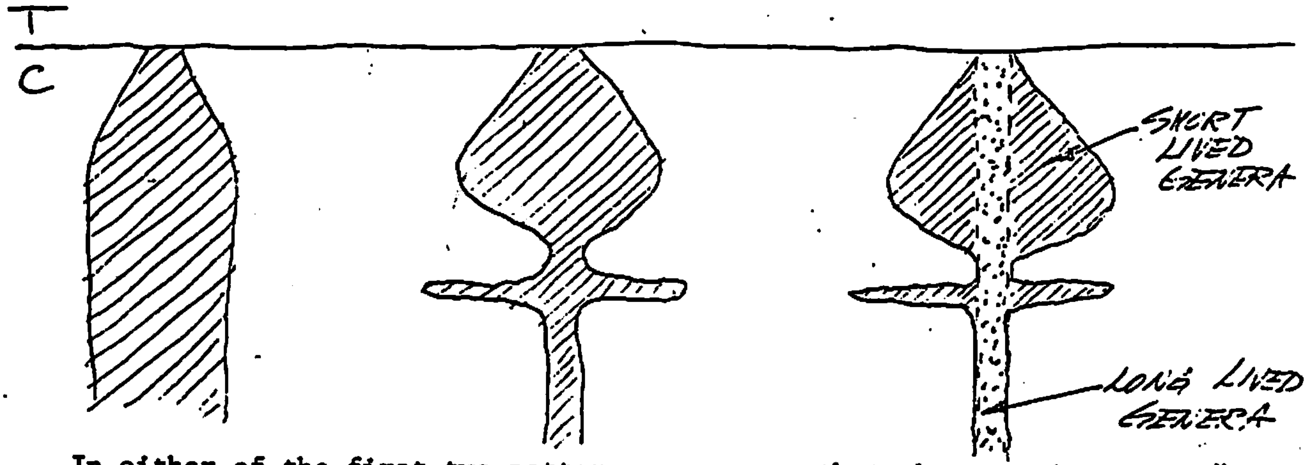
I'm glad to be set straight about the passenger pigeon. My most vivid memory about that extinction is an article I read, a good many years ago, that attributed the extinction to the invention and use of the telegraph, together with hoards of farmers with shotguns. The point that was new to me was that when a flock of birds was seen heading "that-a-way", the local telegraph operator would alert the people in that direction, and they'd be waiting with their shotguns loaded. So that memory will explain why I incorrectly said "The passenger pigeon disappeared in the last century." The picture I remember from that article showed a wood-burning locomotive, on a prairie, with lots of characters from a western movie, standing around with shotguns at the ready.

I appreciate what you say about the inability of some paleontologists to accept analytical arguments. I agree with you that Dave Raup is leading a counterrevolution, and you'll be interested to know that he wrote me a letter in much the same vein as yours. In fact, I've had several letters from paleontologists, explaining to me that although some of their colleagues had blind spots where it came to data analysis, the writers approved of what we were doing, and wished me well.

I'll now address your "one general criticism or rather query". You say, "But why do you not comment on the extensive documentation of invertebrate paleontology that many groups meeting their end in the extinction were in decline for millions of years before--ammonites, and various groups of clams for example." Well the reason I didn't mention that "extensive documentation" is that as far as I can learn, it is quite improperly drawn, just as I now feel sure that the same "extensively documented" evidence that the dinosaurs were "on the way out", before the asteroid hit is wrong. Bill Clemens makes a big point of this decline, but when one examines his best evidence, he finds it doesn't support Bill's strongly held beliefs. Dale Russell also has convincing evidence that the dinosaurs were "blind-sided", but most of our paleontological critics dismiss Dale's data because he is a confirmed catastrophist. But Bill Clemens is certainly not, and I find it most impressive that he can present not a shred of evidence to back up his belief. (And as you know, I've met with Bill in 12 weekly sessions, each lasting all morning, and frequently punctuated with requests for data to support his claims.) In the last of our sessions, he came up with data that, as I mentioned in my talk, contradicted his position, and was consistent with the view that there was no precursor to the extinction; neither the population density nor the diversity of dinosaurs showed signs of a decline before the extinction.

I did know at the time of my talk that the ammonite data had recently been re-examined by Prof. Peter Ward, a UC Davis paleontologist, who showed that ammonites could be divided into two varieties, which I'll call the specialists (short lived) and the generalists (long lived). Here is a diagram from his lecture at Berkeley--given the same day I gave my talk in Washington. (Actually this is a slide of Walt's, adapted from that of Peter Ward.)

Possible Patterns of Ammonite Abundance



In either of the first two patterns, one sees that the ammonites were "in decline" toward the C-T boundary. But the second pattern shows that the ammonites both waxed and waned in time, just as other taxonomic groups do, so no special importance should be attached to the fact that they were in a waning phase at the C-T boundary. That is a statistical argument, but the clincher was provided by Peter Ward who produced the third pattern, that he says corresponds to the geological record. Here, he separates the diagrams into two parts—the dark part corresponds to the short-lived specialists, that radiate to take advantage of changes in the environment, and then lose out when the environment turns against them. The dotted central section shows the long-lived generalists that went on through the whole Cretaceous, with very little change in population or diversity, and not paying any attention to changes in the environment. To Peter Ward the impressive fact is that these phlegmatic (if that is an acceptable word) ammonites were wiped out suddenly, together with the opportunists, and none of any sort made it into the early Tertiary. I find these observations quite impressive, and certainly counter to what you call the "extensively documented" evidence about ammonites. As Josh Billings once (approximately) said, "Our trouble isn't what we don't know, but what we know that isn't so." The reason I didn't mention this in my talk is that I knew it from a "personal communication", that Peter Ward didn't announce publically until the recent GSA meeting in New Orleans. Walt gave his talk the next day, so he was able to use the slide to which I just referred, with Peter Ward's permission.

The same thing has recently happened in a re-examination of the brachiopod record, by Finn Surlyk, from Copenhagen. Here is what he says in the abstract to a paper he has just finished:

> We have studied the brachiopods of the supposedly most "ideal" section in Denmark: Nye Kløv, to test the extinction models in a major marine invertebrate group (Surlyk & Johansen 1981). Our results show an extinction pattern compatible with that reported for coccoliths and pelagic foraminifera. About 20 species occur in the Upper Maastrichtian. Extinction is abrupt and there is no warning in the form of decreasing diversity or early extinction of specialized groups. The basal few metres of the Danian are almost devoid of brachiopods, and a Danian brachiopod fauna starts almost as abruptly as the Maastrictian disappeared. The new fauna is similar to the Maastrictian as regards density and diversity, and at maximum five species are common to both faunas.
>
> Taken at face value this extinction pattern suggests a common cause with the oceanic microplankton extinction, e.g. that the effects of a meteorite impact caused simultaneous, abrupt extinction.

In conclusion, I'll say that the evidence that I've examined most closely —in the pelagic sediments and in the dinosaur regions of North America--show a sudden extinction, without warning, at the iridium layer, within reasonable experimental errors, and quite inconsistent with theories that say the extinctions weren't related to the asteroid impact. And now I've mentioned two additional studies that I have to take "on faith" as being good solid science, because I'm not in a position to evaluate the evidence myself. All four of these extinctions seem to me to be of a catastrophic nature. There may of course be lots of evidence of gradual extinctions near the C-T boundary, but all I can say is that I haven't seen evidence for such events. If you know of such evidence, I hope you'll send it to us, because we are really coming to believe that the asteroid impact did trigger the C-T extinction, and that had it not occurred, you "wouldn't be sitting in front of (your) dictating machine tonight".

And now that I've answered your query in some detail, I'd be interested to learn why you think we "have made an excellent but not overwhelming case for (the) impact--". My personal assessment is that now, after more than three years of exposure to the best criticisms that the experts can bring to bear, the asteroid impact scenario has not only held up, but has increased immeasurable in credibility. No one has suggested another mechanism to bring in all the seven siderophiles, in the same abundance ratios one finds in chondritic meteorites (undifferentiated solar system material), but nowhere else on earth. So I'd give our hypothesis an A+, rather than the A you apparently assign it. (I've always had a reputation as a "tough grader", so perhaps the answer is simply that you are even tougher.)

It's always a pleasure to have an opportunity to exchange ideas with you. I hope that one of these days, we'll have a chance to meet, and talk more at length. And finally, congratulations on your MacArthur fellowship. As my young associate, Rich Muller, said when he was interviewed after receiving his fellowship, "Now I'll have enough money so I won't have to teach, or do research-- but I'll continue to do both, because I enjoy them so much." I'm glad to see that you're still writing your essays for Natural History--most probably for the same reason.

Very sincerely,

Luie

Luis W. Alvarez

p.s. In rereading your letter, I find I didn't comment on your remarks about the smaller Eocene-Oligocene event. We've now found iridium layers at seven places, worldwide, but all in pelagic sections. We are surprised to see that some E-O sites have more nanograms per cm^2 of Ir than do some C-T sites, so the impact events weren't nearly an order of magnitude different in size. Walt has the following as a working hypothesis, with absolutely no experimental foundation: When we find the E-O Ir boundary layer on continental sites, in Europe, for example, there is a distinct possibility that several vertebrate extinctions that are known to have occurred in a several million year time span will be shown to be precisely synchronous with the Ir layer. At the present time, there are no good time markers that can be trusted to show synchroneity to the desired accuracy, over a wide geographical area, so events that may turn out to be synchronous with the E-O Ir boundary layer could not previously have been so identified. So they now make up part of the "background noise" in Dave Raup's "family extinction plot". Walt's hope is that we may be able to show that the E-O event also led to a much more severe extinction than just that of the radiolaria.

The other event that we're still working on is the Permian-Triassic boundary layer from two places in China. At the moment, the Ir abundance in that layer is considerably less than in the sedimentary rocks above and below the layer, which are in turn compatible with the Ir content one expects from the rain of meteoric dust--ablated in the atmosphere. (It was the search for this Ir that got us all into the "Ir business", nearly five years ago.) If these observations hold up, it will show that the P-T boundary layer was emplaced at a more rapid rate than the rocks above and below it. Then we'll try to decide if it was caused by a cometary impact, or by a worldwide volcanic dust cloud. In either case, it will take all the analystical skills that Frank Asaro and Helen Michel can bring to bear on it, to get the answer.

December 8, 1982

Luis W. Alvarez
Lawrence Berkeley Laboratory
University of California
Berkeley, CA 94720

Dear Luis,

Many thanks for your long and fascinating letter which gave me the enormous, if somewhat lazy, ~~privileges~~ privilege of not having to prepare half a lecture so I could discuss "the latest" with my class. And I thought Harvard students were the only people who ever asked for an upgrade from A to A+ !

I would like to come back at you one more time on the issue of extinctions in marine invertebrates for several million years before the dramatic and sudden event that you demonstrate so well at the very close of Cretaceous times. I would also preface these comments by saying once again that I view my theme as entirely complementary to yours, and not detracting one whit from the reality or supreme importance of your extraterrestrial impact. (I speak entirely of shallow-water marine invertebrates with their excellent fossil record, not of the dinosaurs -- where I have agnostic attitudes on complete suddenness versus general decrease followed by sudden termination.) You replied to me by speaking of ammonites and making two arguments: 1. that since ammonites waxed and waned several times during their Mesozoic history (as indeed they did, and quite dramatically) their waning over several million years of the latest Cretaceous has no general significance; and 2. that Peter Ward's division of ammonites into specialists and generalists, with different extinction patterns in each, supports your view about suddenness as the only game in town. I would respond:

1. If ammonites were the only group that waned during the latest Cretaceous I would agree with you entirely -- even in stochastic models some groups are always on the wane. My claim (or rather that of traditional paleontology) is rather that too many groups were simultaneously waning during this time to attribute merely to coincidence among groups that have a history of marked variation in diversity. I am no expert on late Cretaceous faunas, but I have always understood that many groups of the classical, well-preserved record of shallow-water marine invertebrates simultaneously declined throughout the late Cretaceous -- including, just among clams, the reef-building rudistids, the trigonians, and the inoceramids.

2. I am delighted by Ward's results and regard them as a confirmation for my synergistic claims. He shows that the long-lived generalist forms do not vary much in diversity and terminate suddenly, whereas the short-lived specialists are the only ones that contribute to the late Cretaceous waning

... 2

phase. To me, this is confirmation of the classical view that generally deteriorating environments caused a marked and simultaneous decline in many groups of invertebrates. After all, when climates begin to deteriorate it is the specialist genera that differentially feel the effect. You see, in asserting the reality of these simultaneous late Cretaceous declines, I am not in any way trying to detract from the reality of a sudden termination at the end of the Cretaceous. The generalists disappear in the sudden event -- the specialists decline more gradually. But remember that this more gradual decline of the specialists is an important part of the classical data that have always been considered to be a major part of the story of what we call the great Cretaceous extinction.

So I'll push my point once again. There was a sudden termination and it was produced by an extrateerestrial impact. Without it, there would not have been an extinction of unusual magnitude -- nothing to talk about in the profoundest sense that we wouldn't have evolved without it. But the Cretaceous event was able to become one of the five great extinctions in the history of life precisely because the biosphere was already weakened for other reasons when your asteroid hit. Thus, and on the same theme, I would offer a different explanation for the problem you are now having with the Eocene-Olegocene extinction. I gather from your last paragraph that the iridium levels during this event were almost as high as those of the Cretaceous termination -- leaving the anomaly that the Eocene-Olegocene extinction event seems so much smaller in magnitude than the Cretaceous event. You are trying to explain thissby arguing that several extinctions now spread out over several million years by paleontologists might actually represent the same simultaneous event. Perhaps you are right. But I would offer the alternative view that the Eocene-Olegocene extinction was smaller simply because the impacting body struck an earth whose biosphere was at that time in a robust and diverse state.

The Permian information is fascinating. Do you have any insight about the great intra-Devenian event?

Sincerely,

Stephen Jay Gould

/ap

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO · SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF GEOLOGY AND GEOPHYSICS

BERKELEY, CALIFORNIA 94720

December 9, 1982

Professor Steven Jay Gould
Museum of Comparative Zoology
The Agassiz Museum
Harvard University
Cambridge, Massachusetts 02138

Dear Steve,

My father showed me his recent exchange of correspondence with you. I am writing to add a few details regarding the question of the abruptness with which various invertebrates died out at the Cretaceous-Tertiary boundary. I had been under the impression, as you apparently are, that they had been in decline for a long while, and that if the impact contributed to their extinction it was only by disposing of the last pitiful remnants, which were probably on the way out anyway. Recently, however, I have been looking at the records of the invertebrates in the most complete sections available, and have found a rather different picture, at least in the cases of the ammonites, brachiopods, bryozoans and bivalves.

Dad covered the ammonites in some detail, but I should clarify a couple of points. The sketch he reproduced on p. 2 should not be blamed on Peter Ward. It is my attempt to portray three possible models for ammonite extinction. The left-hand sketch is the usual conception, in which the impact played only a minor, mopping-up role. If the center sketch were correct, the impact might have been critical, because the ammonites might have recovered, as they had previously done. The right-hand sketch is a model in which the impact was critical because it wiped out the long-lived genera which otherwise survive each abundance minimum. This picture corresponds with the data on longevities of ammonite lineages presented by Ward and Signor (GSA Abs. with Progr., 14, 642, 1982). The best record of ammonites across the boundary comes from Denmark; the attached range diagram is from Birkelund (C-T Boundary Events, Copenhagen Sympos., v. 1, p. 51-57, 1979). The seven genera that make it to the last zone are all found in abundance in the hardground immediately beneath the iridium-rich fish clay. Peter Ward looked up their ranges for me and 6 of the 7 are among the most long-ranging Mesozoic ammonite genera (Peter had no range information on the other genus, Hoploscaphites). It looks to me as if this is telling us that the impact is responsible for the ammonite extinction, which probably would not have occurred otherwise. I just got back from Boulder, where I discussed this at length with Erle Kauffman. Erle seemed quite convinced by this ammonite evidence. He said that in the past, paleontologists have been impressed by the fact that only the generalists were left when the extinction occurred, and considered this to be evidence for a gradual extinction. But he suggested that in a case like the ammonites, one should now take the view that the sudden disappearance of the generalists was the most important event.

Dad also covered the brachiopod story. In his January, 1982, AAAS talk, Finn Surlyk repeatedly stressed the lack of prior ecological crisis in the sudden crash of microbrachiopods from 24 to 3 or 6 species in the most complete Danish section, Nye Kløv. I enclose the whole abstract Dad quoted. ←

sorry— it has disappeared

Recently I have been looking at the Bivalve record in the Stevns Klint section. Heinberg (C-T Boundary Events Sympos., Copenhagen, v.1, p.58-64, 1979) got good molds of aragonitic bivalves from the hardground just below the Fish Clay. He was impressed by how many typically Tertiary genera appear just <u>before</u> the Fish Clay. But note that 12 of 13 of these genera are dominantly aragonitic, and 10 are not seen again until some time later, so their first appearance in the hardground is probably just a matter of preservation of the molds there. On the other hand, 10 genera last appear in the hardground. Five are aragonitic, so they do not count. The other 5 are calcitic, so their extinctions are valid. In conclusion, of 16 calcitic genera in the hard ground, 5 go extinct exactly at the boundary. But they are all long-lived genera. And of the 11 calcitic genera in the hardground sample that survived the boundary event, every one made it to the Recent.

Finally the bryozoans. In the 1979 Copenhagen volume (p. 78-91), Hakansson and Thomsen summarized the bryozoan record across the C-T boundary at Nye Kløv, the most complete of the Danish sections (figure enclosed). The reduction in number of species of cheilostone bryozoans per 250-specimen sample across the boundary is impressive (fig. 2B), but could just be temporary environmental effect. But the truncation of species (fig. 2D) is more important: of 71 latest Cretaceous species, only 11 survived the boundary! (We need to be sure these are all permanent disappearances, and I do not yet have the species lists available.)

There are hints of similarily abrupt effects in some of the other invertebrate groups, but these are the ones for which I can make the best case from the available literature.

I am impressed with the control a ruling theory has over the way we view our data. For example, Hakansson and Thomsen's bryozoan record screams catastrophe, but they introduce it with this statement: "It is evident, however, that the turnover in the bryozoan fauna at the Maastrichtian/Danian boundary is in no way comparable to the dramatic changes noted in other groups. . ." (Of course I hope I'm not a horrible example of the control <u>my</u> ruling theory has over <u>my</u> perceptions!)

We're in the process of writing this up now, and I hope it will be coauthored by a couple of authorities on the invertebrate record. I'd be very interested in your comments and suggestions.

All the best.

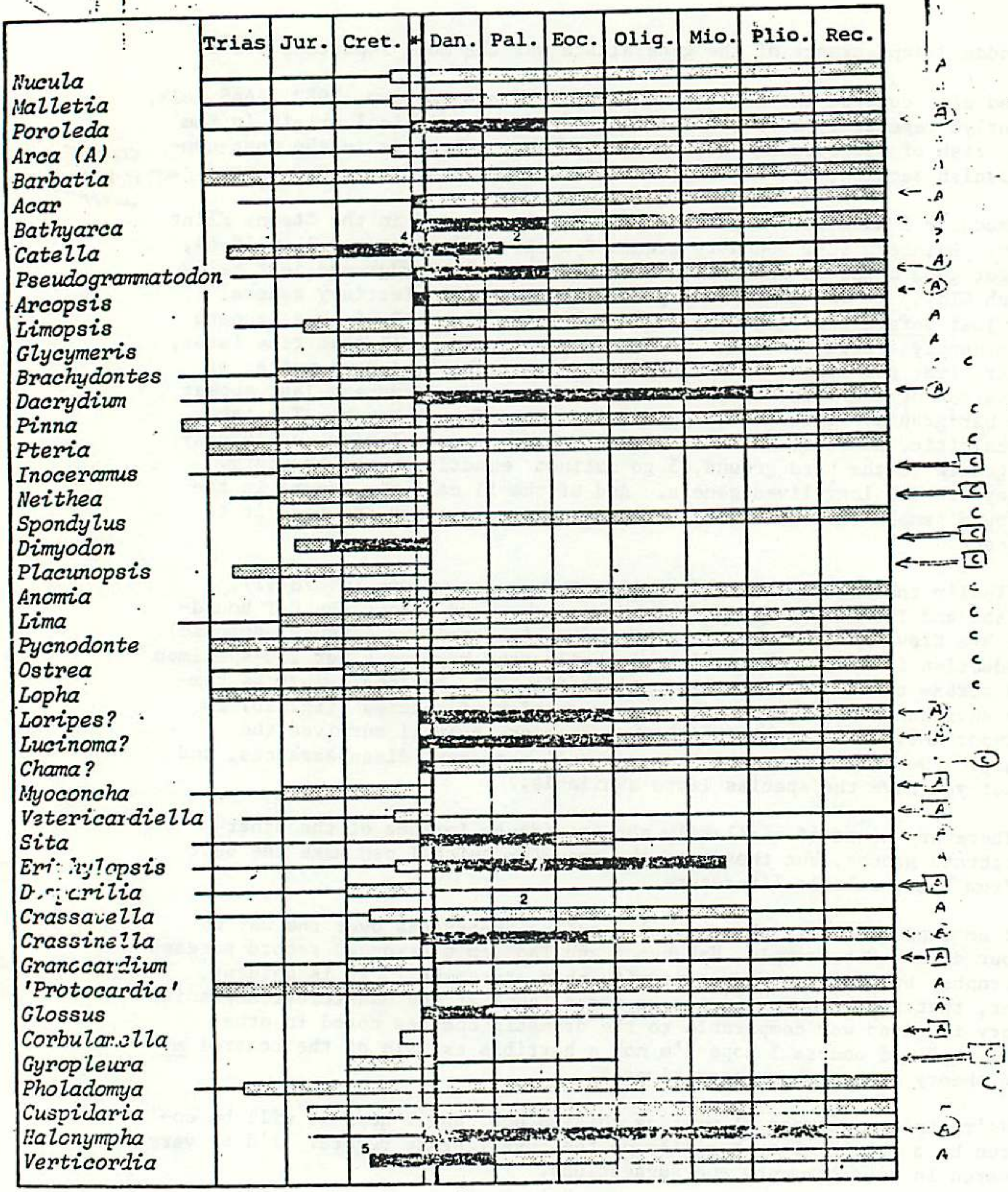
Sincerely,

Walter

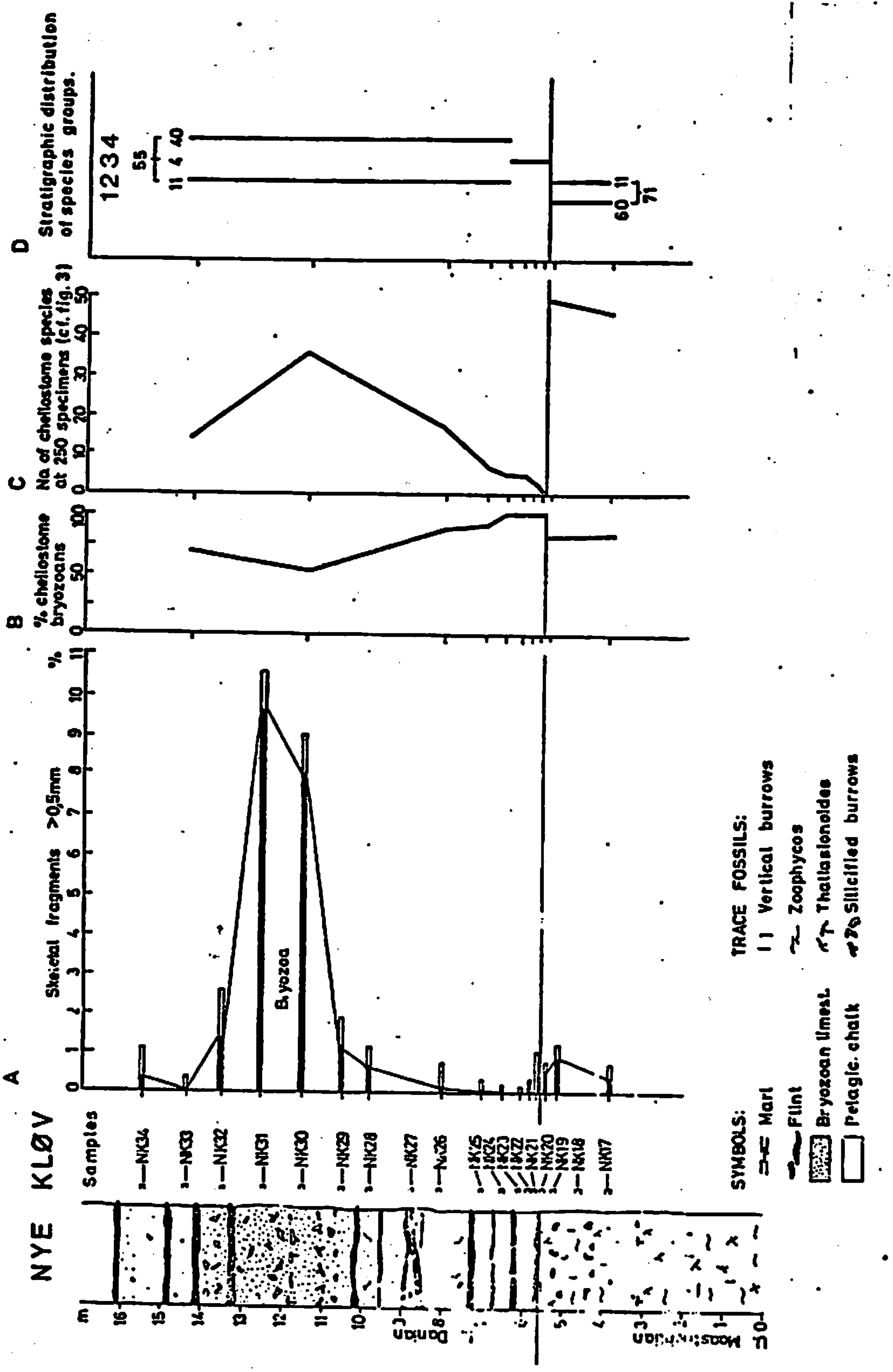
Walter Alvarez

WA/je
Enclosure

Heinberg 1979



Håkansson & Thomsen 1979



January 10, 1983

Walter Alverez
Department of Geology and Geophysics
University of California at Berkeley
Berkeley, CA 94720

Dear Walter,

I think, from the first paragraph of your most interesting letter of December 9th, that you are misinterpreting my claim somewhat. Remember that I make it from the stands of someone who has always supported your hypothesis (I was pleased that your father referred to me in his first letter as the first paleontologist outside your "little group" who took the asteroid seriously). Thus, my point in bringing up the gradual decline of invertebrate groups has not been to deny that there was a true catastrophe ~~at~~ at the very end of Cretaceous times. From my support for your hypothesis and my general negative feelings about gradualism at all levels, I have always felt comfortable with the idea of a true catastrophy. Thus, I do not claim that "if the impact contributed to their extinction it was only by disposing of the last pitiful remnants, which were probably on the way out anyway." I am arguing, rather, that some other set of forces was acting over millions of years to reduce the diversity of many invertebrate groups to a point where a generally weakened biosphere was able to show much more effect from a catastrophic impact than would be the case in more healthy ecological times.

I am delighted by all your evidence that specialist genera declined throughout the Late Cretaceous leaving generalist forms to meet their end in the catastrophe itself. The sudden disappearance of the generalists certainly does support your hypothesis, and we all rejoice for that. I only point out that the gradual decline of all those specialists is part of the classic data of the Cretaceous extinction and is part of the general phenomenon — a part that cannot be attributed to your asteroid. This is why I say that we need a synergistic theory — one that acknowledges the catastrophe as truly sudden and major in effect but that also recognizes the catastrophe as occurring at a peculiar biological time that enhanced its effect.

Sincerely,

Stephen Jay Gould

/ap

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO

SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF GEOLOGY AND GEOPHYSICS

BERKELEY, CALIFORNIA 94720

February 1, 1983

Professor Stephen Jay Gould
Museum of Comparative Zoology
The Agassiz Museum
Harvard University
Cambridge, Massachusetts 02138

Dear Steve,

Thanks for your January 10th letter. You are right. One needs to distinguish between two views of the role of the pre-extinction decline. The first view, with "pitiful remnants, which were probably on the way out anyway", I should not attribute to you; it was my understanding of what many invertebrate and dinosaur palentologists had said in response to the impact hypothesis. Your synergistic view, with the ~~scarcity~~ severity of an impact-generated extinction being related to the general health of the biota at the time, appeals to me.

The evidence relevant to a synergistic theory looks pretty complex. In the case of the ammonites, the impact seems to have been necessary for the extinction, as it truncated the long-ranging taxa. Perhaps the preceding decline was necessary also; if there had been more of the short-ranging taxa around, one of them might, by chance, have survived. At this stage, I have the impression that a prior decline is not always necessary. The forams and coccoliths seem to have been flourishing at the time of the impact, and at least in the North Sea basin, the cheilostome bryozoans and the brachiopods do not show a prior decline. Erle Kauffman has been showing me the evidence for a sharp faunal break at the Cenomanian-Turonian boundary, where no one would put a prior decline. We're going to run Erle's samples for Ir, and they should be relevant to the question.

But the thing that has really struck me is the fact that for the four groups I mentioned, the data show abrupt truncations, when I had been assured that the extinctions were entirely gradual.

Sincerely,

Walter

Walter Alvarez

WA/je

50th ANNIVERSARY 1931-1981

Lawrence Berkeley Laboratory
University of California • Berkeley, California 94720

February 10, 1983

Professor Stephen Jay Gould
Museum of Comparative Zoology
The Agassig Museum
Harvard University
Cambridge, MA 02138

Dear Steve:

I enjoyed reading your article on SETI in Discover, last night, and it reminded me that I hadn't answered your letter of December 8, 1982. (I have heard a lot of Casey Stengel stories in my time, but the one about the catcher was a new one that really cracked me up.)

It seems to me that we each accept the facts that the other produces -- you believe that "there was a sudden termination and it was produced by an ET impact." And without it, you feel the Cretaceous declines you see wouldn't have cleaned the slate to the point that we could have evolved. I agree with those points, but my scientific intuition -- battle-tested by 50 years in the front-line trenches -- tells me that as the years go by people will point increasingly to the impact as the underlying cause of the extinctions and will refer less and less frequently to the enviornmental changes that preceded the impact, and that certainly led many taxa to decline on a million year time scale. I've lost enough similar bets in my lifetime to make me not too dogmatic in that prediction, but nonetheless, I feel fairly confident that a century from now, the impact will be given the credit (or blame) for the extinctions, and the environmentally related slow declines will rate only a footnote.

I don't know enough paleontology to comment on your belief that the reason the C-T extinction was so much more severe than that at the E-O boundary was "because the biosphere was already weakened for other reasons when your asteroid hit." I am impressed by the curves from the recent Science article by Pollack, Toon et al. (<u>219</u>, 287-9 (1983)) that show the continental temperatures dropping to less than 0°F, and staying below 20°F for a few months, with severely darkened skies for several months. The final sentence in their abstract says, "Extinctions in the ocean may have been caused primarily by the temporary cessation of photosynthesis, but those on land may have been primarily

Telephone: (415) 486-4000 FTS 451-4000

induced by a combination of lowered temperatures and reduced light." I don't see how some prior weakening of the biosphere could have much influence in turning such a consequent and really terrible catastrophe into an appreciably worse one. (But as you know, I never thought about such matters until very recently.)

We haven't found an Ir enhancement at a continental Eocene-Oligocene site, and we're just getting ready to measure some irradiated limestone samples from the Gubbio Valleys. When we find it there, we'll know its paleomagnetic age; since all our previous E-O Ir and microtektite signals come from rotary-drilled deep sea cores, we don't yet even know whether the earth's field was normal or reversed. We hope to get some piston cores that will settle that question. Then when we know the number of the zone and its sign, we'll be better prepared to look for the layer in European continental sites. Then we may learn if there is any merit to Walt's suggestion that a larger E-O extinction has been lost in the background noise, through lack of precise dating information.

The only data on the recent intra-Devonian event comes from Los Alamos, where Carl Orth has systematically searched for Ir through several millions of years of sediments, and found no enhancement in the sections his paleontological friends have suggested he examine.

It's always a pleasure to exchange ideas with you.

Very sincerely,

Luie

Luis W. Alvarez

LWA/bk

March 1, 1983

Luis W. Alvarez
Lawrence Berkeley Laboratory
University of California at Berkeley
Berkeley, CA 94720

Walter Alvarez
Department of Geology and Geophysics
University of California at Berkeley
Berkeley, CA 94720

Dear Luis and Walter,

I think that with your recent letters, we have reached a point of nearly complete agreement about the facts, while we continue -- perhaps from the lethargy of traditions in our professions -- to place somewhat different emphases upon various aspects of the case. (Though even here we don't differ all that much since I suspect that Dave Raup and I are probably the invertebrate paleontologists most inclined to be friendly to notions of catastrophe -- and this for different reasons, Dave for his interest in random processes, me for my long campaign against the bias of gradualism).

Walter, I surely agree that the impact was absolutely necessary for extinctions since the long-ranging taxa would not have been exterminated otherwise. But I am pleased that you might also allow that, had other causes not removed so many short-ranging taxa, someone would have had a much better chance of sneaking through. In any case, my main point was only that the gradual declines have, for more than a century, been classically considered as part of the data set of the Cretaceous extinction -- and that the declines are coincident within too many groups to be interpreted as fortuitous.

Luis, I think I share your intuition that when the dust has settled on all this (meaning, of course, the metaphorical dust and not your literal cloud which, I take it, settled some time ago) the catastrophic removal will be seen as by far the more interesting phenomenon. Whether this reflects human psychology or biological importance I am not so sure -- though I suspect the latter as well, just as you do. But I also suspect that the gradual declines will not simply be a footnote -- at least so long as paleontologists remain so interested in coincident patterns in the history of life.

Sincerely,

Stephen Jay Gould

/ap

December 15, 1983

Walter Alvarez
Department of Geology and Geophysics
University of California at Berkeley
Berkeley, CA 94720

Dear Walter,

I am glad you liked the piece. I think you put your finger squarely on the most interesting and important issue underlying any attempt to formulate a general theory of mass extinction. Ever since I developed punctuated equilibrium, I have been fascinated with the general style of change that it represents as merely a small and specific hypothesis for one subdiscipline. (Thom's theory of catastrophes, and Kuhn's notion of scientific revolutions constitutes other examples of similar approaches to the general nature of change). Once you substitute for gradualistic assumptions the notion that stubborn structure is a property of most systems, then you almost have to think of change as a flip or rapid transition from one state to another when a stress too profound to absorp hits the system. Thus, I completely agree that the main issue is to try to determine which kinds of perturbations can be encompassed by systems, permitting them to return to their former state at least approximately, and which will engender transitions something fundamentally new. The differing strength of extinctions in response to what may be common sets of inputs may testify to this differential behavior.

While it may be true that not many paleontologists have joined Thom Schopf in actually switching our area of active research to genetics, many of us are nonetheless keenly interested in bridging this evident gap. I am very hopeful that a meeting set for early January in Miami to bring together geneticists an interested in evolution with macroevolutionists might start a rapprochement. By the way, how do you feel about your hypothesis serving as a major input to formulating the nuclear winter scenario.

Sincerely,

/ap

Stephen Jay Gould

February 24, 1984

Dr. David Raup
Department of Geophysical Sciences
University of Chicago
5734 S. Ellis Avenue
Chicago, Illinois 60637

Dear Dave,

Thanks for the Gingerich piece, which I had not seen. I confess that I don't really see the point (though perhaps this only means that my biases are showing). Darwin must have sketched that diagram in 15 minutes one night after a beer and his customary cigarette. I don't see what a careful quantitative analysis of its branching angles can accomplish. I do take the point to heart that we have documented primarily stasis as the empirical support for p.e. and that stasis is not absent from Darwin's own thought (but, dear me, it is not pervasive, as p.e. contends -- just look again at that blasted diagram -- all within-lineage change is moving somewhere slowly; there is no stasis at all) -- and that the punctuations (which are mainly represented by gaps in evidence) are largely unsupported empirically. But I don't see how he can say that we have no direct evidence at all for the punctuations themselves, especially given the brouhaha over Williamson's African lake data, which purport to illustrate the punctuations directly and empirically. Also, I guess it bothers me as a kind of win-your-own-argument-by-definition to define stasis, as Gingerich actually does explicitly in his paper, as zero-rate gradualism. Finally, I did locate the letter from him re artifacts and plotting variables against their reciprocal. On this I think we agree that he really doesn't understand our point. How can we get it across?

I've been following the latest on cratering frequencies and companion stars, if only in the newspapers and by phone with Walter Alvarez. How I wish I could be with you to discuss this all next weekend. I'd be on the next plane out if I weren't leaving for an M.C.Z. sponsored tour to India the very next day and singing in a concert the day before. What can I say? It's the damned most exciting thing that's ever happened in paleontology in my personal experience and I can only thank you and Jack for getting it started. I think that your 26 million year cycle will hold, whatever the outcome of various astronomical proposals. What I love about your paper is that it is such a statistical tour de force in finding ways to catch and affirm the cycle -- I had my grad students read and discuss it as an exemplary methodological study, quite apart from its fascinating conclusions. As you say, the total independence of cratering data from the empirical basis of your cycles is the strongest point in favor of the whole scheme -- though I confess to being just a teensy bit worried when Walter tells me that he could only find the cratering cycle by tossing out all inadequately dated impact structures and focussing only on those well specified (a good procedure of course, but if he know where the 26 million year peaks fell in advance, that could be a powerful abbconscious bias in deciding which craters to accept or reject). I'm also rooting for the companion star, a much niftier idea than Rampino's excursions (and, I gather, and of course this is most important, eminently testable by good old empirical scanning -- if only in the infrared; after

...2

all, Pluto was found by the same basic procedure of someone deciding it had to be there and then patiently scanning). But, as we agree, it's the same point I tried to make central in my Diswover dinosaur piece (and that the public just doesn't understand thanks, in part, to the failuresoof science writers to understand what science is really about): the excitement lies in the fruitfulness and testability, not in the attention-grabbing nasture of the claims. Look, as you said yourself, Raupandd Sepkoski isn't even published yet and look what's happened in the way of testable proposals!

Finally, of course, and from our own parochialism, I am most delighted by all this because, as you so well stated, it elevates our"fuddy-duddy" science right up there with the hottest of hot shots in the rigorous, experimental, mathematical etc. "hard" sciences. I love a quote from Nature written in 1969 and showing what we have to fight: "Scientists in general might be excused for assuming that most geologists are paleontologists and most paleontologists have staked out a square mile as their life's work. A revamping of the geologist's thage is hadly needed." Well, they'll never be able to say such a thing again. The Wells coral growth line basiness helped in the same way, but got stalled in the unwillingness of organisms to obey (or rather, record) astronomical periodicities clearly in their daily growth. Your data set operates at a totally different level of generality that cannot fail to pick up the major externally-driven effects.

Sincerely,

/ap

Stephen Jay Gould

April 23, 1984

Dr. David M. Raup
Department of Geophysical Sciences
University of Chicago
5734 S. Ellis Avenue
Chicago, Illinois 60637

Dear Dave,

Yes, impacts extend their impact, continuing to prove the fruitfulness of science at its best against the sterility of previous speculative traditions for catastrophic hypotheses.

For once, I am slightly one up on you concerning latest rumors of the P-T anomaly. I spoke to Norman in New York last week, and he thinks the "boundary" is really a paraconformity (but then paraconformity has been his distinctive hypothesis and special love for so many years). Even if so, an iridium anomaly is still what it is, even if the timing lacks precision for lack of continuity. The Frasnian-Fammenian report is exciting. Do you know any more about it? Perhaps we are coming to a general theory of mass extinction after all. Or are we on the verge of an embarras de richesses -- if iridium anomalies start turning up helter-skelter at non-extinction non-boundaries as well. But this suggestion has already become a virtual cliche (though it must be tested seriously by total sampling of a column), and I expect a good correction of chemical anomalies with your 26-million year cycles.

Nemesis should be renamed Sive, for the Hindu God of destruction (who also preserves and creates) and whose most beautiful manifestation is the Nataraja, or cosmic dance, which cycles in a ring of fire and creates as it destroys (Nemesis is too negative). Now that we engage in arcane debates about its name, maybe we'll even find the thing.

Sincerely,

/ap

Stephen Jay Gould

THE UNIVERSITY OF CHICAGO
DEPARTMENT OF THE GEOPHYSICAL SCIENCES
5734 S. ELLIS AVENUE
CHICAGO • ILLINOIS 60637

April 10, 1984

Professor Stephen Jay Gould
Museum of Comparative Zoology
Harvard University
Cambridge, MA 02138

Dear Steve:

Thanks for your note.

Al Fischer called me a few days before he visited Cambridge and he was in the frenzy of a 10-year old. I have worked some with his data and will do more. Although his periodic lows do not look compelling, we have to take seriously Al's intuition and judgement. [By the way, my preliminary analysis of Al's plots give 37 instead of 39 million years.]

Rumors of new iridium anomalies are flying. As you may know, the Chinese have announced a Permo-Triassic anomaly. Norman Newell and Don Boyd were at the meeting and saw the section in the field. I have talked with Don but not Norman. Don found it reasonably convincing. And it appears that the Los Alamos group has finally found (in Australia) a good Frasnian-Fammenian anomaly (in an offshore sequence). If these two anomalies turn out to be real, then the basic impact-extinction link may have been made beyond reasonable doubt. This, in turn, should make it easier for people to absorb, or at least entertain, suggestions of periodicity in extinction.

Sincerely yours,

Dave

David M. Raup
Chairman

DMR/gy

March 5, 1984

Richard A. Muller
Department of Physics
University of California at Berkeley
Berkeley, CA 94720

Dear Dr. Muller,

Just a quick note of thanks and congratulations before I depart later today on a month long trip to India (I would have gladly joined you in your session this past weekend had it not been for this eminent departure). I have been (as always) in close touch with Dave Raup during the past several months and had a long conversation by phone with Walter Alverez last week — so I have been aware of all the excitement and following it with avidity. (I enclose a recent column just published in February, but already largely out of date.) The implications, as you say, are enormous and any traditional Darwinian who does not see it in precisely the terms you set out simply has his head in the sand. Traditional notions of adaptation, survival of the fittest, and our faith in some general improvement through the history of life, albeit fitful, must all be fundamentally reassessed if all traditional clocks are so violently reset by events for which no good Darwinian organism can possibly prepare.

I am rooting hard for your companion star. For pure empirics, it would be the discovery of the century. This of course is the beauty of your proposal, whether right or wrong — it gives you something to do, something to look for. So don't bother answering this; just go and look for nemesis.

I want to write a column on all this but won't proceed until I clear everything with you since I would want to use material from your unpublished preprints. I wouldn't write it till I return in April and it would then not be published until August or September. I'll call you when I return to get some sense of timing for your papers.

Sincerely,

Stephen Jay Gould
Professor of Geology

/ap
Encl.

Lawrence Berkeley Laboratory
University of California
Berkeley, California 94720
Telephone 415/486-4000
FTS: 451-4000

Sept 27, 1982

Dear Prof. Gould:

It's been almost 3 years since I last wrote you, to tell you what we'd been finding, and thinking about. A lot has happened since then, and as this tale tells, I now feel sure that the asteroid *did* lead directly to the extinction of the dinosaurs, plus lots of other animals and plants. I'd be interested in your comments.

Sincerely,

Luis W. Alvarez

Lawrence Berkeley Laboratory
1 Cyclotron Road Berkeley, California 94720
(415) 486-4000 • FTS 451-4000

Blg. 50 Rm. 232
23 May 1984

Dr. Stephen Jay Gould
Museum of Comparitive Zoology
Harvard University
Cambridge Massachusetts 02138

Dear Steve,

Thank you for your letter and the preprint, which I assume you are submitting to Natural History magazine. I enjoyed it very much, as I have enjoyed many of your previous articles and books.

If our comet shower model is correct, then it seems to me that we can reconcile our view with that of many of your paleontological colleagues who have refused to accept the original impact model. They argue, as does Stanley in the current SCIENTIFIC AMERICAN, that the extinctions were not sudden but drawn out over 1 to 2 million years. This always seemed incompatible with the impact model, and was the source of much "controversy" (on which you always seemed to fight on our side). But in the comet shower model we expect each extinction to be, in fact, a series of one to perhaps two dozen individual events, separated by roughly 100,000 periods of respite. Many of these individul events would be unresolved, particularly if they affected the less abundant species, and they would appear to the paleontologist to be an extended single extinction. The iridium would first be found for the largest of the impacts, and only later as sensitivity improved would the individual events be found. If one looks at the C/T and the E/O boundary iridium data, there does appear to be secondary peaks perhaps 100,000 yr away, although their significance is uncertain.

Do you believe that there is clear paleontological evidence that the C/T extinction was not a single abrupt event, but that major species in fact died out over a period of a million years or more? I will be very interested in your opinion, as will Luis Alvarez. We have been lately debating the interpretations of various paleontological evidence, a debate for which we are both woefully prepared.

As for your suggestion of "Siva": it is a name I had seriously considered. I don't know if you read my preprint, or the edited version which was finally published, so I enclose a new copy of the preprint. You will note in footnote 10 I suggested four possible names, two of which are very nonwestern: Nemesis, Kali, Indra, and George. This footnote was shortened and placed into the main text by NATURE without my permission. Siva was a possibility for just the reasons you mention, but she had one devasting disadvantage for physicists: the name is already in use! At Livermore the huge laser used for controlled thermonuclear fusion, the largest laser in the world, is named "Shiva". This is so well known among physicists that it really preempts the use of the name. Of course Siva (or Shiva) was originally, like Nemesis, a goddest of death and destruction which in the later course of Hinduism evolved into a giver of life as well. But Nemesis has a similar schitzophrenia, and was also worshiped for her positive aspects. She was not a destroyer of the righteous or happy, but rather of the wealthy, powerful, and of the proud. To a suppressed people (politically or financially) she could be counted on for their eventual salvation. I understand that even today the people of Greece do not like to have their compliments on the beauty of their children. I think our culture could benefit from a similar humility. As Roger Molander said recently in an editorial in the New York Times, "our worst enemy is decidedly not other human beings."

This whole topic has become unbelieveably exciting: Paleontology, Evolution, Nuclear Physics, Astronomy, Geology, all essential; all working together! I shall be in Boston on June 18 and 19 for the SETI conference; I hope to see you there. I look forward to our future discussions on this and other subjects.

Sincerely yours,

Rich

Richard A. Muller
Professor

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO SANTA BARBARA • SANTA CRUZ

DEPARTMENT OF PHYSICS BERKELEY, CALIFORNIA 94720

27 February 1984

Professor Stephen Jay Gould
Harvard University
Cambridge, Mass.

Dear Prof. Gould:

I enclose copies of my papers on the hypothesized solar companion and on the discovery of periodic impact craters on the earth. It appears to me that although the exisence of the solar companion is still speculation, that the periodicity of comet showers and their correlation with mass extinctions is now established. This is bound to have a major effect on our understanding of evolution. Being a non-expert in this area, I feel free to speculate.

Instead of our evolution taking place in the rather benign environment which we had previously assumed, with only a random disruption from an Apollo asteroid every few hundred million year, we now know that the earth is hit in a rather regular way. After 25 to 27 million years of standard evolution, with the species fighting to find which is fittest, the earth is subjected to a million years of trauma, which none of the inhabitants had any reason to prepare for. During this time the earth is hit by perhaps two dozen comets, with intervals of a few tens of thousands of years in between. Species which could not have had a chance to compete are given a new opportunity. Species which were well-established are suddenly thrown into chaos; the advantages they had gained over the previous 25 million years wiped out in a few months. Virtually every creature on the earth dies, and for the few who survive the race is on to repopulate the world, assuming that a mate can be found.

Stagnation, under such circumstances, seems improbable if not impossible. The comet showers serve a role for species similar to that played by "death" for individuals. (If not for death, you and I never would have attained tenure; paramecia don't die; death may have been invented by nature to prevent stagnation on the short term.) It seems possible that if not for the companion star (which I prefer to call "nemesis"; see footnote 10 in the paper by Davis, Hut, and myself) that the world would still be dominated by trilobites.

I am sorry you didn't come to the MacArthur fellows meeting last year; I found it very interesting. If you come this year, perhaps we can talk more about some of these ideas, and you can correct some of my more obvious errors. I have enjoyed much of your semipopular writings, and look forward to meeting you.

Best regards.

Sincerely yours,

Rich M.

Richard A. Muller
Professor

THE UNIVERSITY OF CHICAGO
DEPARTMENT OF THE GEOPHYSICAL SCIENCES
5734 S. ELLIS AVENUE
CHICAGO · ILLINOIS 60637

February 21, 1984

Professor Stephen Jay Gould
Museum of Comparative Zoology
Harvard University
Cambridge, MA 02138

Dear Steve:

Thanks for sending the Cerion manuscript. Alas, I have not gotten to it yet. But some long stretches on airplanes coming up will help.

Enclosed is a Gingerich piece which he may not have sent you. It is in review with Systematic Zoology. I think he suggests an interesting direction for thinking and analysis but, as usual, Phil has great trouble with execution.

I liked your thing on dinosaurs in Discover.

The extinction business is gathering speed. One of these days, it might make an interesting exercise in the History of Science--no matter how it comes out. New developments fall into two classes: (a) analyses of the terrestrial cratering record and (b) astrophysical explanations of periodic extinction.

On the first, two groups have found (independently) a strong periodic distribution in time of large impact craters. The period is essentially the same as that Jack and I found for the biological record and it is in phase. The analysis involves none of the problems imposed by the stage-level chronology and the dataset could not be more different from that for the biological record. Furthermore, there is no crater for the terminal Cretaceous so the datasets are not even tied to each other at that point. This goes a long way to confirm the periodicity of extinction and also suggests a common mechanism in large body impact. The relative lack of craters between extinction points suggests (sic) that the idea of a continual random flux of impacts may be totally wrong. All this is most exciting. The conventional wisdom has always been that impacts are randomly distributed in time. And this wisdom prevented Gene Shoemaker, Jack and me from really trying to see anything else in the data. We did test the data statistically, but with so little enthusiasm that we found nothing. The successful analyses were made by a group at Goddard headed by Rampino and one associated with Luis Alvarez at Berkeley. Jack and I have repeated the analysis using different methods and the results are striking and highly significant. Both groups submitted manuscripts to Nature at about the same time. The Rampino paper also re-analyses our extinction data and comes to essentially the same conclusions and concludes that impacts caused the major extinctions. This is one of several major reversals we have seen: Rampino has been one of the strongest opponents of the Alvarez model.

On the astrophysical side, the two papers mentioned above plus several others suggest explanations for periodic increases in impact rate. All manuscripts were submitted to Nature within a period of a few weeks in December and January and it will be interesting to see what the editor does with them. The astrophysical explanations fall roughly into two categories: (a) systematic movement of our solar system relative to the galactic plane such that our comets (Oort Cloud) are jostled on a periodic basis and (b) the proposition that our sun has a companion star on a very eccentric orbit which periodically causes the Oort Cloud to be screwed up. Both scenarios call for comet showers. On the galactic plane hypothesis, there is some fairly good evidence for periodicity which is in broad agreement with the extinction/cratering cycle. The double star idea has absolutely no confirmation but is, to me, the most intriguing. Given the 26-million year period, the astrophysicists can calculate back to the orbit, mass, etc. of the companion star. Because we cannot see it, it must be a black dwarf. And I gather that the calculated mass is at least reasonable. Infrared observations might pick up the companion, especially because we know approximately where to look. So, it is possible that this hypothesis could be confirmed by direct observation.

An amusing aspect is that all the papers referred to above are based on preprints of the Raup-Sepkoski paper or on press accounts. Our paper comes out this month.

The foregoing is about where things stand now, but things are changing pretty fast. A bunch of us are getting together at Berkeley in a couple of weeks to brainstorm. It is only a matter of time until the sedimentologists-geochemists get more deeply into the act-in the context of matching the cycle with anoxic events, isotopic data, and so on. The risk of spurious associations and correlations will increase greatly. This can be tolerated as long as the basic extinction and cratering analysis is real and valid. If the whole thing is wrong, it could take a generation of research to get rid of it. But I am much more confident of the basic truth than before.

What do you think? Jack and I have been so tied down with time series analysis that we have taken only passing glances at the evolutionary implications. And the implications could put people to work for a long time.

If (sic) what I have described and implied turns out to be really real, and especially if the double star idea works, it carries an interesting message for what a traditional, fuddy-duddy science like paleontology can do for a socalled hard science like astrophysics. We will see.

With best regards, I am

Sincerely yours,

Dave

David M. Raup
Chairman

DMR/gy

P.S. I enclose a pot-boiler about extraterrestrial intelligence which I think you were asked to write.

16 March 1984, Volume 223, Number 4641 **SCIENCE**

Alvarez, W. et. al.
"Impact theory of mass extinc-
tions and the fossil record"
Science
Published by the A.A.A.S
Used by permission

Impact Theory of Mass Extinctions and the Invertebrate Fossil Record

Walter Alvarez, Erle G. Kauffman, Finn Surlyk
Luis W. Alvarez, Frank Asaro, Helen V. Michel

Summary. There is much evidence that the Cretaceous-Tertiary boundary was marked by a massive meteorite impact. Theoretical consideration of the consequences of such an impact predicts sharp extinctions in many groups of animals precisely at the boundary. Paleontological data clearly show gradual declines in diversity over the last 1 to 10 million years in various invertebrate groups. Reexamination of data from careful studies of the best sections shows that, in addition to undergoing the decline, four groups (ammonites, cheilostomate bryozoans, brachiopods, and bivalves) were affected by sudden truncations precisely at the iridium anomaly that marks the boundary. The paleontological record thus bears witness to terminal-Cretaceous extinctions on two time scales: a slow decline unrelated to the impact and a sharp truncation synchronous with and probably caused by the impact.

In his 1969 presidential address to the Paleontological Society concerning the abrupt Late Devonian faunal turnover, McLaren (*1*) proposed that the impact of a large extraterrestrial body on the earth could trigger a mass extinction. This idea was almost completely ignored. A different extraterrestrial mechanism, explosion of a nearby supernova, was invoked by a few researchers to explain mass extinctions, particularly the one that is used to mark the boundary between the Cretaceous and the Tertiary (*2*). This idea also failed to receive much support. Throughout the 1970's, the weight of informed paleontological opinion held that the terminal-Cretaceous extinction was gradual on a time scale of (1 to 10) × 10^6 years, could best be accounted for by such environmental factors as climatic deterioration and a lowering of sea level, and clearly contradicted the predictions of a sudden, extraterrestrial triggering mechanism (*3*).

In 1978, anomalous concentrations of iridium were found in precise correspondence with the Cretaceous-Tertiary (K/T) boundary, as defined by a very abrupt change in calcareous microplankton in pelagic limestone sequences (*4*). After negative tests of the supernova hypothesis, the iridium anomaly was interpreted as indicating the impact of an asteroid or comet nucleus, roughly 10 km in diameter (*5*). Iridium anomalies have now been found worldwide in all but two of the approximately 50 essentially complete boundary sections that have been studied by researchers at seven different laboratories (*6*). The interpretation that the iridium anomaly is due to an impact is supported by probability arguments (*5, 7*), ratios among platinum-group elements (*6, 8*), the mineralogical and isotopic composition of the K/T boundary clay where it occurs (*9*), the presence of sanidine spherules (*10*) that are thought to have originated as droplets of impact melt (*11*), and calculations which show that such an impact would produce severe environmental stresses, including darkness (*5, 12*), abrupt temperature increase (*13*) or decrease, or both (*12*), and massive production of nitric acid in the atmosphere (*14*).

Subsequently a second iridium anomaly has been located, close to the Eocene-Oligocene boundary (*15*). This anomaly has now been found in seven deep-sea cores, coinciding precisely or nearly so with the North American microtektite horizon (*16*), which provides direct evidence for a major impact. [Keller *et al.* have recently reported evidence for more than one microtektite horizon in this part of the section (*17*), in which case the situation may be more complex than previously recognized.] Substantial extinctions of mammals (*18*) and various other faunal and floral groups (*19*) occurred at or near the Eocene-Oligocene boundary; these extinctions were not comparable in magnitude or stratigraphic precision to those at the K/T boundary, and it is not yet clear if there is any relation between impacts and worldwide extinctions in the middle Tertiary (*20*).

Despite the growing evidence that large impacts produce iridium anomalies and that at least the dramatic K/T calcareous plankton extinction occurred at exactly the same horizon as the iridium and the spherule concentration, the classical paleontological view has been repeatedly restated—that the terminal-Cretaceous extinctions were largely gradational and most were not related to any impact event (*21*). If this were true, it would strongly contradict the impact explanation for extinctions in its present form.

This article gives a different view of the terminal-Cretaceous fossil record of various marine invertebrate groups, which has been recognized by a few workers for some time but has not yet been widely published or accepted. It has generally been held that most invertebrate groups did not undergo serious extinction at the K/T boundary, and that groups that did become extinct at the boundary, notably the ammonites, had

---

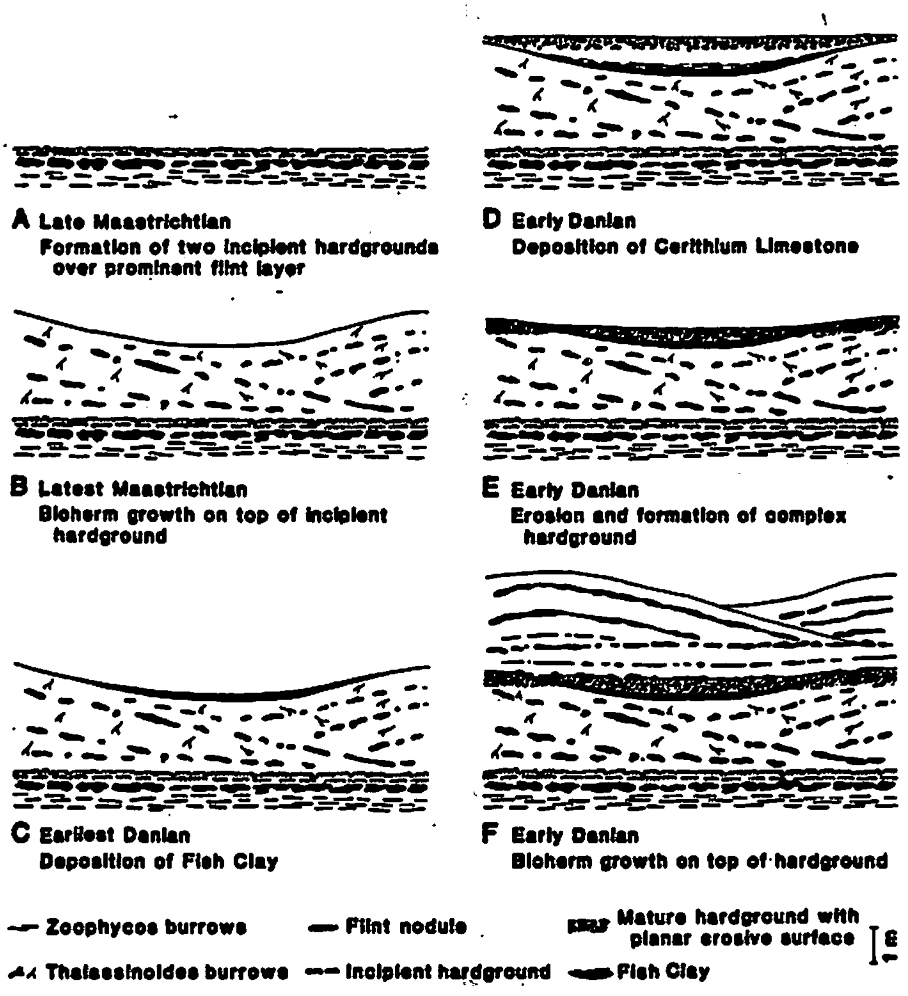
W. Alvarez is with the Department of Geology and Geophysics, University of California, Berkeley 94720. E. G. Kauffman is with the Department of Geological Sciences, University of Colorado, Boulder 80309. F. Surlyk is with the Geological Survey of Greenland, Øster Voldgade 10, DK-1350 Copenhagen K, Denmark. L. W. Alvarez, F. Asaro, and H. V. Michel are with the Lawrence Berkeley Laboratory, University of California, Berkeley 94720.

been in decline for some millions of years, a decline that was much more important in their disappearance than the effects of the suggested terminal-Cretaceous asteroidal impact. We will show that published data on the fossil record strongly contradict this view.

Recent developments in research on the terminal-Cretaceous extinctions have been widely viewed in terms of a conflict between gradualistic and catastrophic interpretations. This notion is counterproductive and should be discarded. It seems evident to us that major biotic turnovers occur on two completely different time scales. The classical paleontological view recognizes the importance of gradual turnovers on a time scale of 10^5 to 10^7 years. The evidence for changes of this kind is so strong that their existence cannot be denied by any reasonable person. The novelty resulting from the recent work on the K/T boundary is the recognition that very rapid turnovers can also occur, with characteristic times of 1 to 10^3 years. In accepting the evidence for very rapid turnovers, one need not reject the reality of gradual turnovers.

We present evidence that very rapid turnovers occurred in a number of invertebrate groups whose terminal-Cretaceous records were previously thought to show only gradual changes or none at all. We intend to establish beyond question the reality and importance of rapid faunal turnovers in these groups. By focusing on the rapid changes, we do not mean to imply that gradual changes are unimportant. However, there is clearly a need for reevaluation of the evidence relevant to gradual versus sudden extinctions in many different fossil groups.

It seems probable that the biologic effects of a large impact may depend on the conditions of diversity and robustness in which various groups find themselves at the moment of impact. Detailed paleontological and stratigraphic studies suggest that the initiation of terminal-Cretaceous extinctions for many groups may have been associated with widespread deterioration of global marine environments during large-scale eustatic fall and epicontinental regression, rapid shifts in ocean chemistry and circulation patterns, and rapid climate and temperature fluctuations during the middle and late Maastrichtian, within the last 1×10^6 years of the Cretaceous. Depleted lineages of typical groups such as ammonites, inoceramid bivalves, reef organisms (including rudists), and marine reptiles, in many cases represented mainly by generalized, long-lived stocks and endangered species of specialists in low population numbers, would have been highly susceptible to abrupt extinction in response to the same event that so dramatically eliminated the calcareous microplankton at a peak in their Cretaceous evolution.



A Late Maastrichtian
Formation of two incipient hardgrounds over prominent flint layer

B Latest Maastrichtian
Bioherm growth on top of incipient hardground

C Earliest Danian
Deposition of Fish Clay

D Early Danian
Deposition of Cerithium Limestone

E Early Danian
Erosion and formation of complex hardground

F Early Danian
Bioherm growth on top of hardground

Fig. 1. Interpreted depositional history of the informal Cretaceous-Tertiary boundary section at Stevns Klint, Denmark, prepared by F. Surlyk. The critical point is that the prominent hardground developed well after the boundary event and does not provide evidence for a break in sedimentation at the boundary. Gravesen (*49*) has reviewed the gradual recognition that the boundary itself is undulating, so that in some places the hardground affects Cretaceous sediments and in other places Tertiary sediments.

Methodology

In attempting to decide whether the asteroidal impact inferred from geochemical information had a significant effect on particular fossil groups, one encounters three notable difficulties. The first problem is that taxonomic diversity is generally presented in the literature as number of taxa per stage or, at best, per substage [that is, a resolution of (1 to 5) × 10^6 years]. Thus, even if a notable difference in the number or composition of taxa can be shown to exist between the Maastrichtian and the Danian, one cannot say, in the great majority of boundary sequences, whether the change took place over a span of 1 to 10 years, as predicted by the impact hypothesis, or over as much as (0.1 to 5) × 10^6 years, in agreement with the gradualistic extinction hypothesis.

The second problem concerns the taxonomic level chosen for the analysis. At a high taxonomic level (family, order, class), a few major Cretaceous groups, such as the inoceramids, belemnites, and ammonites, show substantial or complete extinction right at the K/T boundary. Yet the real fabric of extinction reflected by species- and genus-level patterns of diversity is hidden at this level of analysis. At these lower taxonomic levels it is difficult to find broadly representative samples of the total diversity and composition, and there is the possibility of artifacts in the reported data resulting from differing criteria for naming genera and species, or from the different intensity with which various sections have been collected and studied.

Finally, it has long been recognized that many K/T sections are incomplete across the boundary, with gaps ranging from one or a few fossil zones to several stages. For our purposes, sections with gaps of as much as, say, one substage are simply irrelevant, and any detectable gap reduces the resolution with which an extinction can be said to be instantaneous. There is room within a gap of 1×10^6 years or more for either a graded or a sudden extinction. On the other hand, there is usually no way to detect gaps of 10^3 to 10^4 years or less, which places an ultimate limit on our resolution.

To reduce the effects of these three problems as much as possible, we have based our conclusions on a small number of outstanding sections that have little or no recognizable stratigraphic discontinuity across the boundary, are richly fossiliferous in both microbiota and macrofauna, and have been carefully studied with the use of high-resolution stratigraphy and modern recovery techniques. For the most part, such sections are located in Denmark (*22*) and Spain (*23*). Although this approach excludes an enormous amount of potentially significant data, it also reduces the possibility, on the one hand, of smearing out and thus missing the record of an abrupt extinction, and, on the other hand, of seeing an abrupt event where one does not exist, as a result of stratigraphic gaps or artifacts in the diversity data.

Our conclusions are based not on new data but on an alternate view of existing, published data. To emphasize this, our illustrations are largely taken from the published literature.

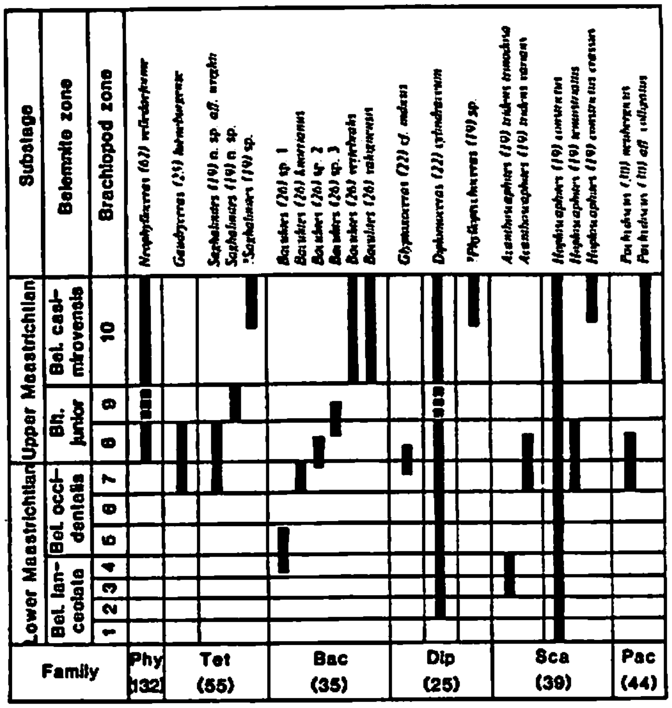


Fig. 2. Ranges of Maastrichtian ammonites in Denmark [after Birkelund (*31*); ranges (in millions of years within parentheses) are from (*30*)]. Families are as follows: *Phy*, Phylloceratidae; *Tet*, Tetragonitidae; *Bac*, Baculitidae; *Dip*, Diplomoceratidae; *Sca*, Scaphitidae; and *Pac*, Pachydiscidae. The families and genera found here are long lived; *Bel.*, *Belemnella*; *Blt.*, *Belemnitella*. The distributions for all ammonites are strongly skewed toward the long-lived end; the mode of generic longevities for all ammonites is 5×10^6 years; the mode of familial longevities is 10×10^6 years (*30*).

Continuity in the Danish Sections

The famous cliff section at Stevns Klint, south of Copenhagen, presently serves as an informal type section for the K/T boundary. It is sometimes suggested that this section is not suitable for detailed K/T boundary studies because of the presence of a hardground, indicative of a hiatus, precisely at the K/T boundary. However, on the basis of existing literature (*22*) and the common knowledge of Danish geologists (including F. Surlyk), it is clear that the hardground is actually younger than the boundary (Fig. 1).

The K/T boundary sections of Denmark are characterized by uniform carbonate lithologies. The actual boundary is marked by a thin marl bed, and locally a hardground is developed about 0.5 m above the marl. The boundary sequence at Stevns Klint is somewhat more complex than the majority of the sections. This is due to the wavy biohermal nature of the top of the Maastrichtian (Fig. 1). The crests of the bioherms are truncated and cemented by the early Danian hardground, and the cemented layer thus consists of alternating Maastrichtian and Danian chalk. The critical point is that the hardground formed well after the K/T boundary event, and its presence does not imply a stratigraphic gap at the boundary.

Probably the best of the Danish sections is at Nye Kløv in Jutland (*24*). This site has similar lithologies in the top Cretaceous and the basal Tertiary, a well-defined, 3-cm boundary clay layer, and absolutely no sign of a hardground.

Ammonites

The widespread, rapidly evolving ammonites are the best biochronological indices for the Mesozoic, and their final extinction occurs in association with the iridium peak at the end of the Cretaceous. It has been widely accepted for some time that a reduction in the diversity of the ammonites had been in progress for several million years prior to the extinction (*25*) and possibly as long as (15 to 20) × 10^6 years (*26*). The apparent decline may be due to the fact that there are very few upper Maastrichtian sections to sample, and most of those are in chalk in which the aragonitic ammonites are poorly preserved; careful statistical testing of this question is needed. Nevertheless, there is a widespread view that the ammonite extinction was a gradual process and that, if a boundary impact event did occur, it merely removed the last remnants of the ammonites, a group so far gone in decline that it was probably doomed anyway. However, in the light of detailed information on the diversity, ecology, and evolutionary longevity of ammonites at the generic level and on the occurrence of ammonites up to the K/T boundary in the essentially complete Danish sections, this gradualistic view of ammonitic extinction requires significant revision.

Cretaceous ammonite diversity through time was marked by alternating maxima and minima, a pattern recognized by numerous investigators (*25*, *27*–*30*). Since the ammonites had recovered from drastic diversity minima, for example, in the vicinity of the Triassic-Jurassic boundary and in the Albian-Cenomanian and Coniacian-Santonian boundary zones, the decline prior to the final extinction loses some of its significance. The ammonites might well have recovered from that decline as well, had it not been for the blow of the impact event, because ecologically generalized, cosmopolitan forms dominated the final ammonite assemblages.

But the case is even more striking in the light of Ward and Signor's recognition (*29*, *30*) that ammonite genera and families fall into two groups: (i) long-lived taxa which neither increase nor decrease dramatically in diversity through time, and (ii) short-lived taxa which account for most of the variation in abundance and diversity of ammonites through time, and which therefore have attracted the most attention because of their biostratigraphic usefulness. In view of this pattern, extinction of a few long-lived taxa at a time of low diversity could be fatal to the entire ammonite group, whereas a gradual decline of the short-lived taxa, although superficially more impressive, would not necessarily lead to extinction. Ammonites had survived the extinction of short-ranging taxa several times during the Mesozoic.

With this background, Birkelund's data on ammonite occurrence in the Danish Maastrichtian (*31*) are of particular importance (Fig. 2). Whereas only nine species, representing seven genera (*Neophylloceras*, *Saghalinites*, *Baculites*, *Diplomoceras*, *?Phylloptychoceras*, *Hoploscaphites*, and *Pachydiscus*) and six families, are found in the uppermost Maastrichtian of Denmark (brachiopod zone 10), Birkelund specifically noted that representatives of all seven genera occur in the uppermost bed of the Cretaceous, that is, immediately below the iridium-rich Fish Clay which coincides with the terminal Cretaceous plankton extinction. Preservation is good in this bed because it was lithified in some places by the episode of hardground formation during the early Danian. Ward and Signor (*30*) have given durations for the six nonqueried genera, ranging from (19 to 62) × 10^6 years (mean, 30 × 10^6 years), which places all six far out on the long-lived tail of the strongly skewed longevity distribution (mode = 5 × 10^6 years). Birkelund noted the abundance of ammonites in the very uppermost Maastrichtian (*31*):

> The topmost Maastrichtian chalk of Stevns Klint, where locally lithified by a [Danian] hardground, contains many mature *Hoploscaphites* and *Baculites* of normal size and rare mature specimens of other genera. Together with these there are abundant juvenile specimens of scaphites and baculites, and juvenile *Saghalinites* sp., *?Phylloptychoceras* sp. and *Pachydiscus* sp. have also been collected. . . . In conclusion it is important to stress that there is no evidence to suggest that the ammonites were affected by poor living conditions to cause dwarfing or stunting (as supposed by Wiedmann, 1969, for a late Maastrichtian fauna of Zumaya, Spain); on the contrary, they seem to have lived perfectly well to the end.

Ward and Wiedmann (*32*) have recently restudied the section at Zumaya in northern Spain, the only other locality with a K/T boundary ammonite record comparable to the one at Stevns Klint. Ammonites are common up to about 15 m below the iridium-rich boundary layer and are entirely gone by 12 m below the boundary. At Zumaya, as at Stevns Klint, dwarfing does not occur among the latest Cretaceous ammonites.

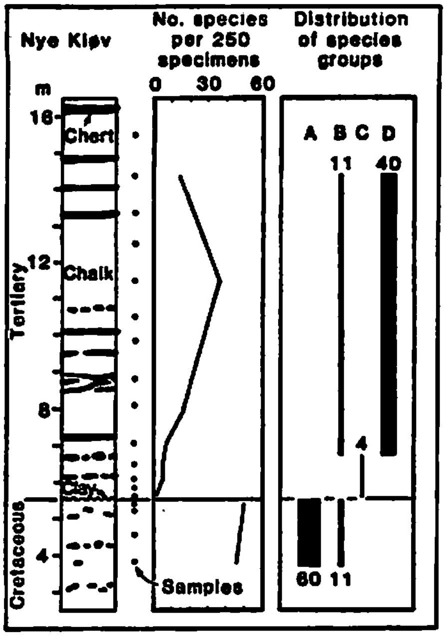


Fig. 3. Data on cheilostomate bryozoans at Nye Kløv, Denmark [simplified from Håkansson and Thomsen (*34*), figure 2]. The critical point here is that 60 out of 71 latest Cretaceous species (85 percent) disappear precisely at the Cretaceous-Tertiary boundary clay. For more detail, see Birkelund and Håkansson (*35*), figure 8.

It is not easy to understand why even the long-ranging ammonites disappeared at Zumaya, although they were thriving until the end at Stevns Klint. Nevertheless, in our view the pattern at Stevns Klint strongly supports the idea that the terminal-Cretaceous impact caused the final extinction of the ammonites. The prior decline in short-ranging taxa certainly weakened the ammonites as a group, but the common long-ranging taxa in Denmark were evidently in no difficulty whatsoever until the impact occurred. This suggests a subtle but significant change in the conclusion that should be drawn from the paleontological data. Formerly the extinction of all short-ranging ammonites was taken to mean that a gradual extinction was nearing completion, with the final dying out of remaining long-ranging taxa a predictable denouement. In the view presented here, the extinction of those long-ranging taxa was the critical and unpredictable event—without it the ammonites probably would not have disappeared.

Bryozoans

The Danish sections are very useful for studies of bryozoan evolution across the K/T boundary. Bryozoans are of such importance in the Danish uppermost Cretaceous and basal Tertiary that lithologies are distinguished by the quantity of bryozoan skeletal material they contain, ranging from chalk through bryozoan chalk to bryozoan limestone (*33*).

The cheilostomate bryozoans were evolving rapidly at this time. Håkansson and Thomsen studied (*34*) this group across the boundary, upgrading the work of Voigt (*33a*); their work is based on the section at Nye Kløv (*24*), which has little or no lithologic difference between the top Cretaceous and the basal Tertiary. Unfortunately, they do not present species identifications but only diversity data. Figure 3 shows the cheilostome bryozoan record at Nye Kløv (*34*, *35*). The number of species in a 250-specimen sample, based on rarefaction analysis, drops significantly at the boundary, from about 50 species in the uppermost Maastrichtian to one species in the lowest sample of the Danian, then recovers substantially in the first few meters of the Danian (Fig. 3). Although not necessarily indicative of a mass extinction, this diversity crash certainly indicates that the North Sea shelf area was hit by a very sudden ecological crisis, from which it gradually recovered.

Figure 3 shows that there is also a major difference in the species composition across the boundary. Of 71 species found in the top Maastrichtian, only 11 are found in the Danian; that is, 60 species (85 percent) disappear immediately below the iridium anomaly (*36*). This event can best be explained as a sudden, major extinction, directly or indirectly caused by the impact.

The abruptness of this extinction distinguishes it from an inferred prior decline in the cheilostomate bryozoans. This apparent decline was stressed by Håkansson and Thomsen (*34*), and by Håkansson (*37*) who estimated that the cheilostome fauna earlier in the late Maastrichtian approached "200 species in the Danish chalk alone (a good portion of which are undescribed)," compared to the 71 species seen in the highest Maastrichtian of Nye Kløv. We note, however, that Nye Kløv is a benthos-poor locality and that the inferred decline is probably an artifact that arises from comparing a single locality with the entire Danish chalk.

Brachiopods

For a number of years, Surlyk (*38*, *39*) has carried out detailed studies of the brachiopods of the Upper Cretaceous chalk of Denmark. He has noted (*39*, p. 48): "The brachiopods have proved to be of great biostratigraphical value in the chalk of NW Europe because many of the species have a limited vertical distribution; occur in great numbers, and are easy to determine." Recently, Surlyk and Johansen (*40*) and Johansen (*41*) have extended these studies upward into the Danian, based on samples from the section at Nye Kløv. Most of the brachiopod species encountered are very small (a few millimeters), and hundreds or thousands of individuals can be obtained from each sample (*40*), so statistical problems do not affect the determination of biostratigraphic ranges. Surlyk and Johansen summarized their findings as follows (*40*, p. 112).

> Our results show an extinction pattern compatible with that reported for coccoliths and pelagic foraminifera. About 20 species occur in the Upper Maastrichtian. Extinction is abrupt and there is no warning in the form of decreasing diversity or early extinction of specialized groups. The basal few meters of the Danian are almost devoid of brachiopods, and a Danian brachiopod fauna starts almost as abruptly as the Maastrichtian disappeared. The new fauna is similar to the Maastrichtian as regards density and diversity, and at maximum five species are common to both faunas.

Surlyk and Johansen (*40*, *42*) have made three points about this mass extinction of brachiopods. First, it was very sudden and coincided exactly with the abrupt, nearly complete extinction of coccoliths and planktonic foraminifera. Second, there was absolutely no decline in diversity of brachiopods prior to the sudden extinction. Third, the synchroneity of this event with the extinction of plankton (coccoliths and foraminifera) and bryozoans suggests a common cause, that is, that both are direct effects of the impact. They noted, however, that the brachiopod fauna was highly specialized and adapted to life on a chalky bottom composed primarily of coccoliths and foraminifera and suggested that the brachiopod extinction may have been a result of the plankton extinction, which destroyed the habitat the brachiopods required. A similar cause might be proposed for the large bryozoan extinction; the substrate requirements of bryozoans are similar to those of brachiopods. This point is very important. We would stress that our purpose in this article is to show that mass extinction did occur in a number of fossil groups, at essentially the same time as the inferred impact. The exact sequence of extinctions, and their causal relations to the impact itself and to each other, will be a difficult problem for future work.

Bivalves

Among bivalves, the cosmopolitan and rapidly evolving inoceramids and the reef-building rudists disappeared at or near the end of the Cretaceous. The maximum development of rudists occurred during the late Campanian and early to middle Maastrichtian, and this group declined rapidly during the last few million years of the Cretaceous (*27*, *43*).

These diversity declines raise the question whether the bivalve record is comparable with that of the ammonites. The best information comes from the fossiliferous Danish sections, but the record is complicated by a mineralogical problem concerning the relative stability of calcite and aragonite shells, as discussed in the legend to Fig. 4. The majority of bivalve shells are wholly or predominantly aragonite (*44*), a mineral that is unstable; shell dissolution occurs early in chalk diagenesis. Aragonitic fossils are commonly absent or poorly preserved in chalks such as those of the Danish boundary sequence. Calcite dominates the shells of a smaller number of bivalve taxa (for example, oysters and scallops) and is an important component of shells with mixed mineralogy, where it normally forms the outer layer. Calcite is stable and selectively preserved in chalks and limestones. Thus, the fossil record of most chalks is highly distorted by the artificial dominance of calcite-shelled taxa unless the carbonates were tightly cemented early in diagenesis, forming hardgrounds.

Heinberg (*45*) has studied the bivalve fauna at Stevns Klint, in the uppermost Maastrichtian sediments lithified by the Danian-age hardground and in the adjacent strata, and his range chart is reproduced in Fig. 4. The 45 genera in the hardground represent a broad ecological and taxonomic spectrum of bivalves, representing four of the five subclasses and eight of the 11 orders existing in the Cretaceous. Both of the major groups of bivalves that disappeared at the end of the Cretaceous, the inoceramids and the rudists, are represented in the highest Maastrichtian collection, by *Tenuipteria* [= "*Inoceramus*" sensu lato of Heinberg (*45*)] and *Gyropleura*, respectively. No rudists or rudist-dominated reefs are known from the latest Maastrichtian of Cretaceous tropical environments, but one species of *Gyropleura*, one of the

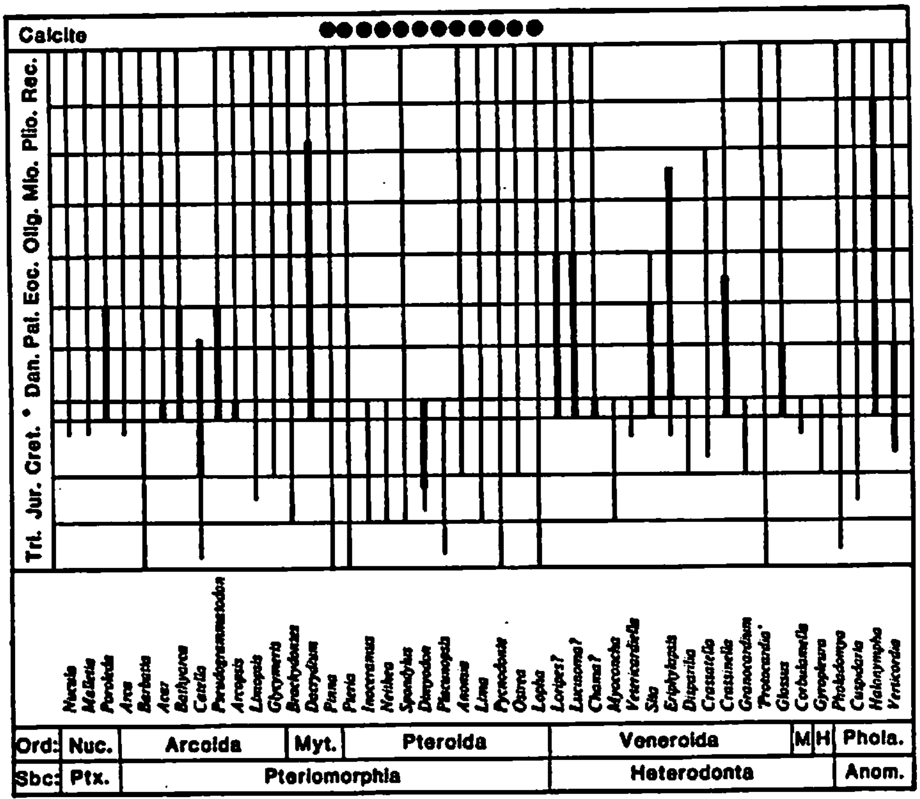


Fig. 4. Ranges of bivalve genera found in the strata near the Cretaceous-Tertiary boundary at Stevns Klint, Denmark [based on Heinberg (*45*), figure 1]. The asterisk marks the uppermost Cretaceous layer lithified by the Danian hardground and the black dots at the top mark calcite-dominated genera. Genera are grouped by order (*Nuc.*, Nuculoida; *Myt.*, Mytiloida; *M*, Myoida; *H*, Hippuritoida; *Phola.*, Pholadomyoida) and subclass (*Ptx.*, Palaeotaxodonta; *Anom.*, Anomalodesmata). Most bivalve shells are either dominantly calcitic or dominantly aragonitic. Dominantly aragonitic shells are easily dissolved and are thus absent in most chalks; they are likely to be preserved only as molds in hardgrounds, so their reported ranges in carbonate facies are not reliable. This is emphasized by the heavy lines, which show new range extensions based on the hardground fauna. Of the 12 genera that are dominantly calcitic and therefore have reliable ranges in the Danish chalk, four go extinct at the iridium-bearing boundary clay; the other eight are still living.

most generalized and widespread of small epibiont rudists, does occur in the uppermost Cretaceous chalks of Denmark, immediately beneath the iridium-bearing Fish Clay. Similarly, the Inoceramidae show a long, gradual diversity decline since the Turonian, with a final rapid decline in the Maastrichtian, but are still represented at the Danish top Maastrichtian by one species of *Tenuipteria*.

Heinberg noted that 13 typically Tertiary genera first appear in the lithified top beds of the Maastrichtian. He therefore concluded that the terminal Cretaceous faunal turnover in bivalves was under way before the boundary as marked by the Fish Clay (subsequently found to be iridium-rich), indicating a graded faunal transition. This conclusion and the known history of rudist and inoceramid decline have reinforced the common impression of a gradual faunal transition at the end of the Cretaceous (*46*).

However, an examination of Heinberg's data shows that, of the 13 genera that first appear in the lithified top Maastrichtian, 12 are dominantly aragonitic and the other is an uncertain identification ("*Chama*?") of a genus with more than 30 species that are aragonitic and only two that are calcitic (*47*). Of the 13 genera, ten had their ranges substantially extended back in time (thick lines in Fig. 4) when they were recognized by Heinberg in the hardground. Because these taxa are rarely preserved in chalk, which dominates the Upper Cretaceous of western Europe, we take this to mean that their earlier recognition was most probably limited by preservation, and that the oldest specimen found in Denmark probably does not represent the origination of the group. These aragonitic taxa almost certainly represent older Cretaceous groups that survived the extinction event, rather than the sudden advanced appearance of Tertiary forms before the mass extinction.

The question thus arises whether the last occurrences of bivalves in the lithified top Maastrichtian are also meaningless. Of the ten genera last appearing in this hardground (that is, apparent terminal-Cretaceous extinctions), five are aragonitic and their ranges should be questioned for the reasons given above. However, four definitely had partially calcitic shells, and their apparent extinctions at this level are therefore probably real, and one, *Gyropleura*, probably had calcite layers (*48*). The importance of these four extinctions is underlined by Fig. 4, which shows that these genera, and the families to which they belong, had lived for very long time spans prior to their extinction. As a further indication of the importance of this extinction, we note that, of the 12 bivalve genera in the hardground sample that are definitely calcite-bearing and should therefore have reliable longevity records, four went extinct at the level of the iridium anomaly, and the other eight have all survived to the Recent. Moreover, Heinberg's data (*45*) are at the generic level, which may well mask a more dramatic extinction at the species level, such as that seen in the brachiopods and the bryozoans.

This reinterpretation of the bivalve record leads to the conclusion that, although some groups, such as the rudists and inoceramids, underwent a prominent Maastrichtian evolutionary decline, the terminal-Cretaceous extinction itself was probably very sudden, unpredictable, and of major importance in the disappearance of these groups.

Conclusion

The four groups of marine invertebrates reviewed here show unmistakable evidence for sudden and unanticipated extinction of small to large numbers of taxa in precise correspondence with the iridium anomaly at the K/T boundary. There can be little doubt that these extinctions resulted in some way from the inferred impact, although it is not presently possible to trace the chains of direct and indirect causality. It will be important to make this kind of detailed evaluation for other invertebrate groups, such as gastropods, corals, and echinoderms, which are currently regarded as having been little affected by the terminal-Cretaceous extinction event. Detailed paleontological study of as many groups as possible, in the very best available sections, is thus an important challenge in the next few years.

References and Notes

1. D. J. McLaren, *J. Paleontol.* 44, 801 (1970).
2. O. H. Schindewolf, *Neues Jahrb. Geol. Palaeontol. Monatsh.* 1954, 457 (1954); *ibid.* 1958, 270 (1958); A. R. Loeblich, Jr., and H. Tappan, *Geol. Soc. Am. Bull.* 75, 367 (1964); V. I. Krasovski and I. S. Shklovshy, *Dokl. Akad. Nauk SSSR* 116, 197 (1957); K. D. Terry and W. H. Tucker, *Science* 159, 421 (1968); H. Laster, *ibid.* 160, 1138 (1968); W. H. Tucker and K. D. Terry, *ibid.*, p. 1138; D. Russell and W. H. Tucker, *Nature (London)* 229, 553 (1971); M. A. Ruderman, *Science* 184, 1079 (1974); R. C. Whitten, J. Cuzzi, W. J. Borucki, J. H. Wolfe, *Nature (London)* 263, 398 (1976).
3. H. Tappan, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 4, 187 (1968); T. R. Worsley, *Nature (London)* 230, 318 (1971); W. T. Holser, *ibid.* 267, 403 (1977); D. M. McLean, *Science* 200, 1060 (1978).
4. W. Alvarez, L. W. Alvarez, F. Asaro, H. V. Michel, *Eos* 60, 734 (1979); in *Cretaceous-Tertiary Boundary Events Symposium*, W. K. Christensen and T. Birkelund, Eds. (University of Copenhagen, Copenhagen, 1979, vol. 2, p. 69; *Geol. Soc. Am. Abstr. Programs* 11, 378 (1979).
5. L. W. Alvarez, W. Alvarez, F. Asaro, H. V. Michel, *Univ. Calif. Lawrence Berkeley Lab. Rep. LBL-9666* (1979); *Science* 208, 1095 (1980).
6. W. Alvarez, L. W. Alvarez, F. Asaro, H. V. Michel, *Geol. Soc. Am. Spec. Pap. 190* (1982), p. 305.
7. G. W. Wetherill and E. M. Shoemaker, *ibid.*, p. 1; P. R. Weissman, *ibid.*, p. 15; R. A. F. Grieve, *ibid.*, p. 25.
8. R. Ganapathy, *Science* 209, 921 (1980); F. Asaro, *Syllogeus* 39, 7 (1982).
9. M. Kastner, quoted by F. Asaro, *Syllogeus* 39, 6 (1982); M. Kastner, F. Asaro, H. V. Michel, W. Alvarez, abstract for *Meeting on Glass in Planetary and Geological Phenomena* (Alfred University, August 1983) (Alfred University, Alfred, N.Y., in press); D. J. DePaolo, F. T. Kyte, B. D. Marshall, J. R. O'Neil, J. Smit, *Earth Planet. Sci. Lett.* 64, 356 (1983); J. M. Luck and K. K. Turekian, *Science* 222, 613 (1983); B. F. Bohor, *Clay Minerals Soc. Annu. Meet. Program* 20, 48 (1983).
10. J. Smit and J. Klaver, *Nature (London)* 292, 47 (1981); A. Montanari, W. Alvarez, F. Asaro, H. V. Michel, L. W. Alvarez, *Geol. Soc. Am. Abstr. Programs* 14, 569 (1982); J. C. Varenkamp and E. Thomas, *Geol. Soc. Am. Spec. Pap. 190* (1982), p. 461.
11. A. Montanari, R. L. Hay, W. Alvarez, F. Asaro, H. V. Michel, L. W. Alverez, *Geology* 11, 668 (1983).
12. O. B. Toon *et al.*, *Geol. Soc. Am. Spec. Pap. 190* (1982), p. 187; J. B. Pollack, O. B. Toon, T. P. Ackerman, C. P. McKay, R. P. Turco, *Science* 219, 287 (1983).
13. C. Emiliani, E. B. Kraus, E. M. Shoemaker, *Earth Planet. Sci. Lett.* 55, 317 (1981).
14. J. S. Lewis, G. H. Watkins, H. Hartman, R. G. Prinn, *Geol. Soc. Am. Spec. Pap. 190* (1982), p. 215.
15. F. Asaro, L. W. Alvarez, W. Alvarez, H. V. Michel, *Conference on Large Body Impacts and Terrestrial Evolution: Geological, Climatological, and Biological Implications* (Lunar and Planetary Institute, Houston, 1981), p. 2 (abstract); R. Ganapathy, *Science* 216, 885 (1982); W. Alvarez, F. Asaro, H. V. Michel, L. W. Alvarez, *ibid.*, p. 886.
16. B. P. Glass and M. J. Zwart, in *Stratigraphic Micropaleontology of Atlantic Basin and Borderlands*, F. M. Swain, Ed. (Elsevier, Amsterdam, 1977), p. 553; B. P. Glass, D. L. DuBois, R. Ganapathy, *13th Lunar Planet. Sci. Conf.* (1982), p. A425.
17. G. Keller, S. D'Hondt, T. L. Vallier, *Science* 221, 150 (1983).
18. H. G. Stehlin, *Soc. Géol. Fr. Bull. Ser. 4* 18, 488 (1909); D. E. Savage and D. E. Russell, *Mammalian Paleofaunas of the World* (Addison-Wesley, Reading, Pa., 1983); M. Brunet, *Geobios. Mem. Spec.* 1, 11 (1977).
19. C. Cavalier *et al.*, *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 36, 223 (1981).
20. B. P. Glass and J. R. Crosbie, *Am. Assoc. Pet. Geol. Bull.* 66, 471 (1982); D. R. Prothero, C. R. Denham, H. G. Farmer, *Geology* 10, 650 (1982).
21. W. Clemens and J. D. Archibald, *Mem. Soc. Geol. Fr. New Ser.* 139, 67 (1980); J. D. Archibald and W. A. Clemens, *Am. Sci.* 70, 377 (1982); W. A. Clemens, J. D. Archibald, L. J. Hickey, *Paleobiology* 7, 293 (1981); L. J. Hickey, *Nature (London)* 292, 529 (1981); T. J. M. Schopf, *Science* 211, 571 (1981); *Geol. Soc. Am. Spec. Pap. 190* (1982), p. 415; W. A. Clemens, *ibid.*, p. 407; C. B. Officer and C. L. Drake, *Science* 219, 1383 (1983); E. G. Kauffman, in *Cretaceous-Tertiary Boundary Events Symposium*, W. K. Christensen and T. Birkelund, Eds. (University of Copenhagen, Copenhagen, 1979), vol. 2, p. 29; E. Voigt, *ibid.*, p. 38.
22. T. Birkelund and R. G. Bromley, Eds., *Cretaceous-Tertiary Boundary Events Symposium*, (University of Copenhagen, Copenhagen, 1979), vol. 1.
23. S. F. Percival and A. G. Fischer, *Evol. Theory* 2, 1 (1977).
24. E. Håkansson and J. M. Hansen, in (*22*), p. 171.
25. J. Wiedmann, *Biol. Rev.* 44, 563 (1969).
26. P. W. Signor, personal communication.
27. E. G. Kauffman, in *Catastrophes in Earth History: The New Uniformitarianism* (Princeton Univ. Press, Princeton, N.J., in press).
28. W. J. Kennedy, in *Patterns of Evolution*, A. Hallam, Ed. (Elsevier, Amsterdam, 1977), p. 251.
29. P. D. Ward and P. W. Signor, *Geol. Soc. Am. Abstr. Programs* 14, 642 (1982).
30. ———, *Paleobiology*, in press.
31. T. Birkelund, in (*22*), p. 51.
32. P. D. Ward and J. Wiedmann, in *Symposium on Cretaceous Stage Boundaries* (Copenhagen, October 1983) (University of Copenhagen, Copenhagen, in press).

33. R. G. Bromley, in (*22*), p. 16.
33a. E. Voigt, *Leopoldina* 6, 379 (1930); *Geologie* 8, 1 (1959); *Ber. Dutsch. Ges. Geol. Wiss. Reihe A*, 12, 479 (1967); *Bur. Rech. Géol. Min. Mém.* 77, 45 (1972).
34. E. Håkansson and E. Thomsen, in (*22*), p. 78.
35. T. Birkelund and E. Håkansson, in (*22*), p. 373.
36. M. A. Nazarov *et al.*, *13th Lunar Planet. Sci. Conf.* (1982), p. 580.
37. E. Håkansson, personal communication.
38. F. Surlyk, thesis, University of Copenhagen (1969); *Bull. Geol. Soc. Den.* 20, 152 (1970); *Newsl. Stratigr.* 1, 7 (1970); *Biol. Skr. Dan. Vidensk. Selsk.* 19, 1 (1972); *Bull. Geol. Soc. Den.* 22, 219 (1973); *Palaeogeogr. Palaeoclimatol. Palaeoecol.* 15, 185 (1974).
39. ———, in (*22*), p. 48
40. ——— and M. B. Johansen, *Abstr. 3rd Eur. Meet. Int. Assoc. Sedimentol.* (1982), p. 112.
41. M. B. Johansen, thesis, University of Copenhagen (1982).
42. F. Surlyk and M. B. Johansen, *Abstr. 148th Meet. AAAS* (1982), p. 49; *Science* 223, 1174 (1984).
43. E. G. Kauffman and N. F. Sohl, in *Encyclopedia of Paleontology*, R. W. Fairbridge and D. Jablonski, Eds. (Dowden, Hutchinson & Ross, Stroudsburg, Pa., 1979), p. 723.
44. J. D. Taylor, W. J. Kennedy, A. Hall, *Bull. Br. Mus. (Natl. Hist.) Zool.* (Suppl. 3) (1969); W. J. Kennedy, J. D. Taylor, A. Hall, *Biol. Rev.* 44, 499 (1969).
45. C. Heinberg, in (*22*), p. 58.
46. T. Birkelund and R. Bromley, in (*22*), p. 5.
47. W. J. Kennedy, N. J. Morris, J. D. Taylor, *Palaeontology* 13, 379 (1970).
48. W. J. Kennedy and J. D. Taylor, *Proc. Geol. Soc. London, No. 1645* (1968), p. 325.
49. P. Gravesen, in (*22*), p. 8.
50. We thank P. D. Ward, P. W. Signor, and T. A. Hansen for helpful criticism. This work was supported by NSF grant EAR-81-15858.

Hydrogen-Evolving Solar Cells

Adam Heller

Summary. Sunlight is directly converted to chemical energy in hydrogen-evolving photoelectrochemical cells with semiconductor electrodes. Their Gibbs free energy efficiency of solar-to-hydrogen conversion, 13.3 percent, exceeds the solar-to-fuel conversion efficiency of green plants and approaches the solar-to-electrical conversion efficiency of the best *p-n* junction cells. In hydrogen-evolving photoelectrodes, electron-hole pairs photogenerated in the semiconductor are separated at electrical microcontacts between the semiconductor and group VIII metal catalyst islands. Conversion is efficient when the island diameters are small relative to the wavelengths of sunlight exciting the semiconductor; when the island spacings are smaller than the diffusion length of electrons at the semiconductor surface; when the height of the potential energy barriers that separate the photogenerated electrons from holes at the semiconductor surface is raised by hydrogen alloying of the islands; when radiationless recombination of electron-hole pairs at the semiconductor-solution interface between the islands is suppressed by controlling the semiconductor surface chemistry; and when the semiconductor has an appropriate band gap (1.0 to 1.8 electron volts) for efficient solar conversion.

Optimal conversion of sunlight by a single-threshold converter, whether semiconductor-based or molecular, requires an energy gap near 1.4 electron volts (*1–3*). Tandem systems, based on two semiconductors or on two light-harvesting molecules, require materials with gaps near 1.8 and 1.0 eV (*4*). At normal solar irradiance and at 27°C, the thermodynamic limit to the solar conversion efficiency is 27 percent for a single converter and 36 percent for tandem cells (*4*). For nonconcentrated sunlight the actual efficiency that has been attained is 21.9 percent (*5*). Although the thermodynamic efficiency limits for semiconductor and molecular systems are the same, all efficient systems today are semiconductor-based.

The foundations of the science of semiconductor-based photoelectrodes were laid between 1955 and 1970 (*6–16*). In 1972 it was shown that when an oxygen-evolving metal anode is replaced by the stable semiconducting anode *n*-TiO_2 (*17*), a substantial part of the electrical energy required for the electrolysis of water is conserved (*18*). Unfortunately, the band gap (E_{BG}) of *n*-TiO_2 is 3.0 eV, so the excitation spectrum of this material is confined to the ultraviolet. Consequently, sunlight could not be efficiently converted in early cells based on *n*-TiO_2 or *n*-$SrTiO_3$ (E_{BG} = 3.2 eV) photoanodes (*19–24*).

Photoanodes with band gaps appropriate for efficient solar conversion were first reported in 1960 (*9*), before the introduction of the concept of power-producing ("regenerative") cells (*25*). However, these were quite unstable. Ideas for their stabilization emerged in 1966 (*26*), but the introduction of relatively stable photoanodes took nearly a decade (*27–50*).

Simultaneously, the second key problem of photoelectrodes, the reduction of quantum yield from radiationless recombination of electron-hole pairs at the semiconductor-solution interface, was addressed. It was shown that recombination can be reduced by properly controlling the interfacial chemistry (*37, 51–66*). These advances led to the first power-producing photoelectrochemical cells of greater than 10 percent solar conversion efficiency (*51, 52, 63, 64, 67*). Stabilization and reduction of surface recombination also opened the way to efficient and direct photoelectrochemical cells for producing hydrogen. These cells are the subject of this article.

Photoelectrodes

Hydrogen can be evolved in either photocathode- or photoanode-based cells. Photocathodes and photoanodes are based on electrical contacts between semiconductors and electrolytes, between semiconductors and metallic electrocatalysts, or between semiconductors and both electrolytes and catalysts. Associated with these contacts is a barrier, ψ_B, which separates the photogenerated electrons (e^-) and holes (h^+) (Fig. 1). Hydrogen evolution at photocathodes requires the presence of a catalyst on or near the surface. When the catalyst is on the surface of the semiconductor, photogenerated electrons diffuse or drift to the interface of the semiconductor and the catalyst (Fig. 1a), where they reduce adsorbed protons to form hydrogen, according to $2e^- + 2H^+ \rightarrow H_2$. Electrical neutrality is maintained by transport of holes through both the bulk of the semiconductor and its back electrical contact to the anode, where they oxidize either dissolved anions such as chloride, through the reaction $2h^+ + 2Cl^- \rightarrow Cl_2$, or water itself, through the reaction

---

Adam Heller is a member of the technical staff and head of the Electronic Materials Research Department at Bell Laboratories, Murray Hill, New Jersey 07974.

SPECIATION

Schopf, T.J.M. (ed.)
Models in Paleobiology--
Eldridge & Gould "Punctuated
equilibria: An alternative to
phylletic gradualism" (c) 1972
W.H. Freeman-Used by permissior

5

PUNCTUATED EQUILIBRIA: AN ALTERNATIVE TO PHYLETIC GRADUALISM

Niles Eldredge • *Stephen Jay Gould*

Editorial introduction. Moving from populations to species, we recall that the process of speciation as seen through the hyperopic eyes of the paleontologist is an old and venerable theme. But the significance of "gaps" in the fossil record has been a recurrent "difficulty," used on the one hand to show that spontaneous generation is a "fact," and on the other hand to illustrate the "incompleteness" of the fossil record. Some have expressed a third interpretation, which views such gaps as the logical and expected result of the allopatric model of speciation.

Bernard's Eléments de Paléontologie (1895) discusses the existence of gaps in the fossil record as follows, p. 25, English edition:

> Still it remains an indisputable fact that in the most thoroughly explored regions, those where the fauna is best known, as, for instance, the Tertiary of the Paris basin, the species of one bed often differ widely from those of the preceding, even where no stratigraphic gap appears between them. This is easily explained. The production of new forms usually takes place within narrowly limited regions. It may happen in reality that one form evolves in the same manner in localities widely separated from each other, and farther on we shall see examples of this; but this is not generally the case, the area of the appearance of species is

> usually very circumscribed. This fact has been established in the case of certain butterflies and plants. The diversity having once occurred, the new types spread often to great distances, and may be found near the present form without crossing with it or presenting any trace of transition.
>
> The same phenomenon must have taken place in former epochs. It is then only by the merest chance that geologists are able to locate the origin of the species they have under consideration; if, furthermore, the phenomena of erosion or metamorphism have destroyed or changed the locality in question, direct observation will not furnish any means of supplying the missing links of the chain.

Although this has been pointed out nicely by Bernard—and moreover, any number of paleontologists will tell you that this is what they teach—comprehension and application are two different things. And indeed, the fossil record has been interpreted by many to show just the opposite. J. B. S. Haldane's classical *The Cause of Evolution*, published in 1932, contains the following passage (p. 213):

> But [Sewall] Wright's theory [that evolution is most rapid in populations large enough to be reasonably variable, but small enough to permit large changes in gene frequencies due to random drift] certainly supports the view taken in this book that the evolution in large random-mating populations, which is recorded by paleontology, is not representative of evolution in general, and perhaps gives a false impression of the events occurring in less numerous species.

Thus an extremely eminent student of the evolutionary process considered that the known fossil record supported the view of evolution proceeding as a stately unfolding of changes in large populations.

The interpretation supported by Eldredge and Gould is that allopatric speciation in small, peripheral populations *automatically* results in "gaps" in the fossil record. Throughout their essay, however, runs a larger and more important lesson: *a priori* theorems often determine the results of "empirical" studies, before the first shred of evidence is collected. This idea, that theory dictates what one sees, cannot be stated too strongly.

Statement

In this paper we shall argue:

(1) The expectations of theory color perception to such a degree that new notions seldom arise from facts collected under the influence of old pictures of the world. New pictures must cast their influence before facts can be seen in different perspective.

(2) Paleontology's view of speciation has been dominated by the picture of "phyletic gradualism." It holds that new species arise from the slow and steady transformation of entire populations. Under its influence, we seek unbroken fossil series linking two forms by insensible gradation as the only complete mirror of Darwinian processes; we ascribe all breaks to imperfections in the record.

(3) The theory of allopatric (or geographic) speciation suggests a different interpretation of paleontological data. If new species arise very rapidly in small, peripherally isolated local populations, then the great expectation of insensibly graded fossil sequences is a chimera. A new species does not evolve in the area of its ancestors; it does not arise from the slow transformation of all its forbears. Many breaks in the fossil record are real.

(4) The history of life is more adequately represented by a picture of "punctuated equilibria" than by the notion of phyletic gradualism. The history of evolution is not one of stately unfolding, but a story of homeostatic equilibria, disturbed only "rarely" (i.e., rather often in the fullness of time) by rapid and episodic events of speciation.

The Cloven Hoofprint of Theory

Innocent, unbiased observation is a myth.
P. B. Medawar (1969, p. 28)

Isaac Newton possessed no special flair for the turning of phrases. Yet two of his epigrams have been widely cited as guides for the humble and proper scientist—his remark in a letter of 1675 written to Hooke: "If I have seen farther, it is by standing on the shoulders of giants," and his confusing comment of the *Principia* (1726 edition, p. 530): "hypotheses non fingo"—[I frame no hypotheses]. The first is not his own; it has a pedigree extending back at least to Bernard of Chartres in 1126 (Merton, 1965). The second is his indeed, but modern philosophers have offered as many interpretations for it as the higher critics heaped upon Genesis 1 in their heyday (see Mandelbaum, 1964, p. 72 for a bibliography).

Although most scholars would now hold, with Hanson (1969, 1970, see also Koyré, 1968), that Newton meant only to eschew idle speculation and untestable opinion, his phrase has traditionally been interpreted in another light—as the credo of an inductivist philosophy that views "objective" fact as the primary input to science and theory as the generalization of this unsullied information. For example, Ernst Mach, the great physicist-philosopher, wrote (1893, p. 193): "Newton's reiterated and emphatic protestations that he is not concerned with hypotheses as to the causes of phenomena, but has simply to do with the investigation and transformed statement of *actual facts* . . . stamps him as a philosopher of the *highest* rank."

Today, most philosophers and psychologists would brand the inductivist credo as naive and untenable on two counts:

(1) We do not encounter facts as *data* (literally "given") discovered objectively. All observation is colored by theory and expectation. (See Vernon, 1966, on the relation between expectation and perception. For a radical view, read Feyerabend's (1970) claim that theories act as "party lines" to force observation in preset channels, unrecognized by adherents who think they perceive an objective truth.)

(2) Theory does not develop as a simple and logical extension of observation; it does not arise merely from the patient accumulation of facts. Rather, we observe in order to test hypotheses and examine their consequences. Thus, Hanson (1970, pp. 22–23) writes: "Much recent philosophy of science has been dedicated to disclosing that a 'given' or a 'pure' observation language is a myth-eaten fabric of philosophical fiction. . . . In any observation statement the cloven hoofprint of theory can readily be detected."

Yet, inductivist notions continue to control the methodology and ethic of practicing scientists raised in the tradition of British empiricism. In unguarded moments, great naturalists have correctly attributed their success to skill in hypothesizing and power in imagination; yet, in the delusion of conscious reflection, they have usually ascribed their accomplishments to patient induction. Thus, Darwin, in a statement that should be a motto for all of us (letter to Fawcett, September 18, 1861, quoted in Medawar, 1969), wrote:

> About thirty years ago there was much talk that geologists ought only to observe and not theorize; and I well remember someone saying that at this rate a man might as well go into a gravel-pit and count the pebbles and describe the colours. How odd it is that anyone should not see that all observation must be for or against some view if it is to be of any service.

Yet, in traditional obeisance to inductivist tenets, he wrote in his autobiography that he had "worked on true Baconian principles, and without any theory collected facts on a wholesale scale" (see discussion of this point in Ghiselin, 1969a; Medawar, 1969; and de Beer, 1970).

Almost all of us adhere, consciously or unconsciously, to the inductivist methodology. We do not recognize that all our perceptions and descriptions are made in the light of theory. Leopold (1969, p. 12), for example, claimed that he could describe and analyze the aesthetics of rivers "without introduction of any personal preference or bias." He began by generating "uniqueness" values, but abandoned that approach when the sluggish, polluted, murky Little Salmon River scored highest among his samples. He then selected a very small subset of his measures for a simplified type of multivariate scaling. As he must have known before he started, Hells Canyon of the Snake River now ranked best. It cannot be accidental that the article was written by an opponent to applications then before the Federal Power Commission for the damming of Hells Canyon. (It is no less fortuitous that so many philosophers, Hegel and Spencer in particular, generated ideal states by pure reason that mirrored their own so well.)

In paleontology, even the most "objective" undertaking, the "pure" description of fossils, is all the more affected by theory because that theory is unacknowledged. We describe part by part and are led, subtly but surely, to the view that complexity is irreducible. Such description stands against a developing science of form (Gould, 1970a, 1971a) because it both gathers different facts (static states rather than dynamic correlations) and presents contrary comparisons (compendia of differences rather than reductions of complexity to fewer generating factors). D'Arcy Thompson, with his usual insight, wrote of the "pure" taxonomist (1942, p. 1036), "when comparing one organism with another, he describes the differences between them point by point and 'character' by 'character.' If he is from time to time constrained to admit the existence of 'correlation' between characters . . . yet all the while he recognizes this fact of correlation somewhat vaguely, as a phenomenon due to causes which, except in rare instances, he can hardly hope to trace; and he falls readily into the habit of thinking and talking of evolution as though it had proceeded on the lines of his own description, point by point and character by character."

The inductivist view forces us into a vicious circle. A theory often compels us to see the world in its light and support. Yet, we think we see objectively and therefore interpret each new datum as an independent confirmation of our theory. Although our theory may be wrong, we cannot confute it. To extract ourselves from this dilemma, we must bring in a more adequate theory; it will not arise from facts collected in the old way. Paleontology supported creationism in continuing comfort, yet the imposition of Darwinism forced a new, and surely more adequate, interpretation upon old facts. Science progresses more by the introduction of new world-views or "pictures"* than by the steady accumulation of information.

This issue is central to the study of speciation in paleontology. We believe that an inadequate picture has been guiding our thoughts on speciation for 100 years. We hold that its influence has been all the more tenacious because paleontologists, in claiming that they see objectively, have not recognized its guiding sway. We contend that a notion developed elsewhere, the theory of allopatric speciation, supplies a more satisfactory picture for the ordering of paleontological data.

---

* We have no desire to enter the tedious debate over what is, or is not, a "model," "theory," or "paradigm" (Kuhnian, not Rudwickian). In using the neutral word "picture," we trust that readers will understand our concern with alternate ways of seeing the world that render the same facts in *different* ways.

Phyletic Gradualism: Our Old and Present Picture

Je mehr sich das palaeontologische Material vergrössert, desto zahlreicher und vollständiger werden die Formenreihen.

Zittel, 1895, p. 11

Charles Darwin viewed the fossil record more as an embarrassment than as an aid to his theory. Why, he asked (1859, p. 310), do we not find the "infinitely numerous transitional links" that would illustrate the slow and steady operation of natural selection? "Why then is not every geological formation and every stratum full of such intermediate links? Geology assuredly does not reveal any such finely graduated organic chain; and this, perhaps, is the gravest objection which can be urged against my theory" (1859, p. 280). Darwin resolved this dilemma by invoking the great inadequacy of surviving evidence (1859, p. 342): "The geological record is extremely imperfect and this fact will to a large extent explain why we do not find interminable varieties, connecting together all the extinct and existing forms of life by the finest graduated steps. He who rejects these views on the nature of the geological record, will rightly reject my whole theory."

Thus, Darwin set a task for the new science of evolutionary paleontology: to demonstrate evolution, search the fossil record and extract the rare exemplars of Darwinian processes—insensibly graded fossil series, spared somehow from the ravages of decomposition, non-deposition, metamorphism, and tectonism. Neither the simple testimony of change nor the more hopeful discovery of "progress" would do, for anti-evolutionists of the catastrophist schools had claimed these phenomena as consequences of their own theories. The rebuttal of these doctrines and the test for (Darwinian) evolution could only be an *insensibly graded fossil sequence*—this discovery of all transitional forms linking an ancestor with its presumed descendant (*figure 5-1*). The task that Darwin set has guided our studies of evolution to this day.*

In titling his book *On the Origin of Species by Means of Natural Selection*, Darwin both identified this event as the keystone of evolution and stated his belief in its manner of occurrence. New species can arise in only two ways: by the transformation of an entire population from one state to another (phyletic evolution) or by the splitting of a lineage (speciation). The second process must occur: otherwise there could be no increase in numbers of taxa and life would cease as lineages became extinct. Yet, as Mayr (1959) noted, Darwin muddled this distinction and cast most of his discussion in terms of phyletic

---

* Beliefs in "saltative" evolution, buttressed by de Vries' "mutation theory," collapsed when population geneticists of the 1930's welded modern genetics and Darwinism into our "synthetic theory" of evolution. The synthetic theory is completely Darwinian in its identification of natural selection as the efficient cause of evolution.

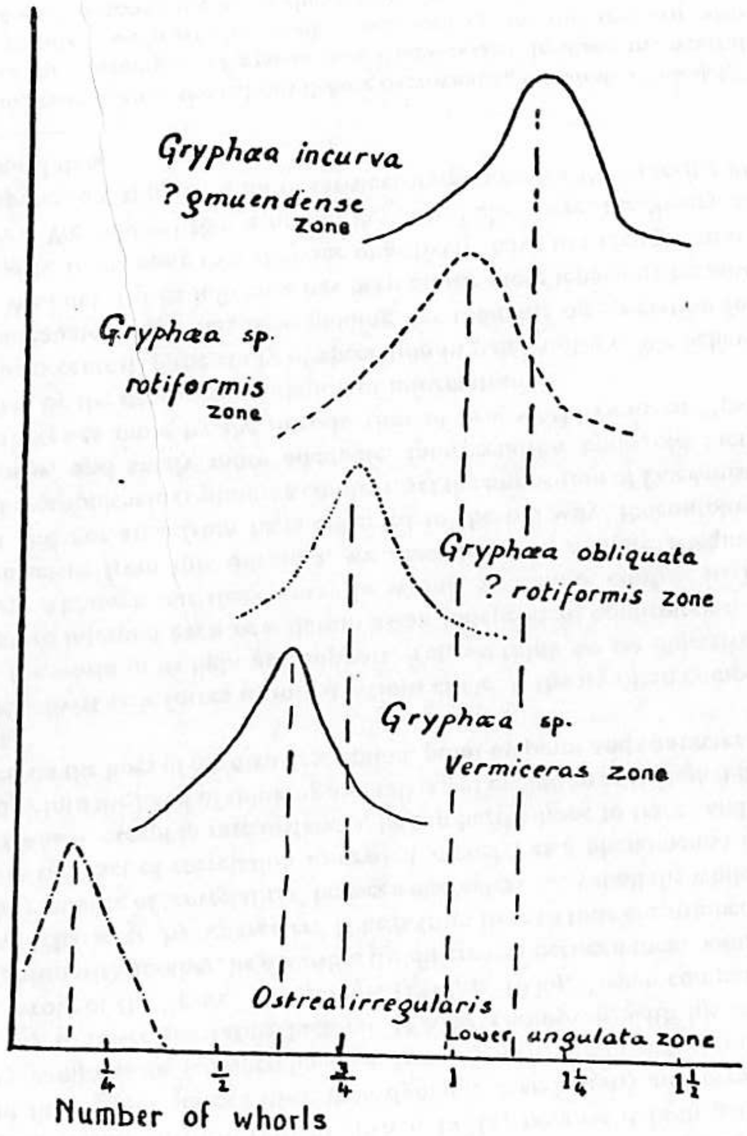


Figure 5-1: The classic case of postulated phyletic gradualism in paleontology. Slow, progressive, and gradual increase in whorl number in the basal Liassic oyster *Gryphaea.* From Trueman, 1922; figure 5.

evolution. His insistence on insensibly graded sequences among fossils reflects this emphasis, for if species arise by the gradual transformation of entire populations, an even sequence of intermediates should indeed be found. When Darwin did discuss speciation (the splitting of lineages), he continued to look through the glasses of transformation: he saw splitting largely as a sympatric process, proceeding slowly and gradually, and producing progressive divergence between forms. To Darwin, therefore, speciation entailed the same expectation as phyletic evolution: a long and insensibly graded chain of intermediate forms. Our present texts have not abandoned this view (*figure 5-2*), although modern biology has.

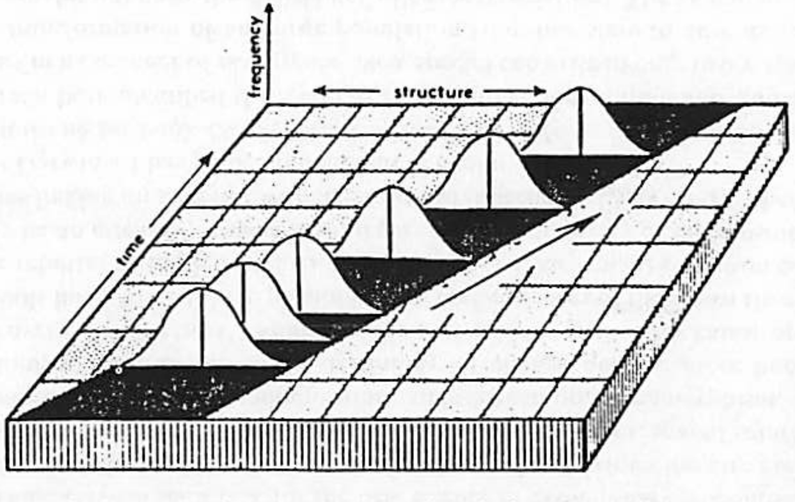


Figure 5-2: A standard textbook view of evolution *via* phyletic gradualism. From Moore, Lalicker, and Fischer, 1952; figure 1-14.

In this Darwinian perspective, paleontology formulated its picture for the origin of new taxa. This picture, though rarely articulated, is familiar to all of us. We refer to it here as "phyletic gradualism" and identify the following as its tenets:

(1) New species arise by the transformation of an ancestral population into its modified descendants.

(2) The transformation is even and slow.

(3) The transformation involves large numbers, usually the entire ancestral population.

(4) The transformation occurs over all or a large part of the ancestral species' geographic range.

These statements imply several consequences, two of which seem especially important to paleontologists:

(1) Ideally, the fossil record for the origin of a new species should consist of a long sequence of continuous, insensibly graded intermediate forms linking ancestor and descendant.

(2) Morphological breaks in a postulated phyletic sequence are due to imperfections in the geological record.

Under the influence of phyletic gradualism, the rarity of transitional seri remains as our persistent bugbear. From the reputable claims of a Cuvier or an Agassiz to the jibes of modern cranks and fundamentalists, it has stood as the bulwark of anti-evolutionist arguments: "For evolution to be true, there had to be thousands, millions of transitional forms making an unbroken chain" (Anon., 1967—from a Jehovah's Witnesses pamphlet).

We have all heard the traditional response so often that it has become imprinted as a catechism that brooks no analysis: the fossil record is extreme imperfect. To cite but one example: "The connection of arbitrarily selected 'species' in a time sequence, in fact their complete continuity with one another, is to be expected in all evolutionary lineages. But, *fortunately*, becaus of the imperfect preservation of fossil faunas and floras, we shall meet relatively few examples of this, no matter how long paleontology continues" (Eaton, 1970, p. 23—our italics; we are amused by the absurdity of a claim that we should rejoice in a lack of data because of the taxonomic convenienc thus provided).

This traditional approach to morphological breaks merely underscore what Feyerabend meant (see above) in comparing theories to party lines, fo it renders the picture of phyletic gradualism virtually unfalsifiable. Th picture prescribes an interpretation and the interpretation, viewed improperly as an "objective" rendering of data, buttresses the picture. We hav encountered no dearth of examples, and cite the following nearly at random Neef (1970) encountered "apparent saltation in the *Pelicaria* lineage" (p. 464; a group of Plio-Pleistocene snails. Although he cites no lithologic or geographic data favoring either interpretation, the picture of phyletic gradualism prescribes a preference: "It is likely that the discontinuity . . . is due to a period of non-deposition. . . . The possibility that the apparent saltations in the *Pelicaria* lineage are due to the migration of advanced forms from smal nearby semi-isolated populations and that deposition of the Marima Sandstone was continuous cannot be entirely excluded" (1970, p. 454).

Moreover, the picture's influence has many subtle extensions. For instance

(1) It colors our language. We are compelled to talk of "morphological breaks" in order to be understood. But the term is not a neutral descriptor. it presupposes the truth of phyletic gradualism, for a "break" is an interruption of something continuous. (Under a deVriesian picture, for example. "breaks" are "saltations"; they are real and expressive of evolutionary processes.)

(2) It prescribes the cases that are worthy of study. If breaks are artificial. the sequences in which they abound become, *ipso facto*, poor objects for evolutionary investigation. But surely there is something insidious here: if breaks are real and stand against the picture of phyletic gradualism, then the picture itself excludes an investigation of the very cases that could place it in jeopardy.

If we doubt phyletic gradualism, we should not seek to "disprove" it "in the rocks." We should bring a new picture from elsewhere and see if it provides a more adequate interpretation of fossil evidence. In the next section, we express our doubts, display a different picture, and attempt this interpretation

But before leaving the picture of phyletic gradualism, we wish to illustrat its pervasive influence in yet another way. Kuhn (1962) has stressed the impact of textbooks in molding the thought of new professionals. The "normal science" that they inculcate is "a strenuous and devoted attempt to force nature into the conceptual boxes supplied by professional education" (1962 p. 5).

Before the "modern synthesis" of the 1930's and 40's, English-speakin invertebrate paleontologists were raised upon two texts—Eastman's translation of Zittel (1900) and that venerable *Gray's Anatomy* of British works Woods' *Palaeontology* (editions from 1893 to 1946, last edition reprinted fiv times before 1958 and still very much in use). Both present an orthodo version of phyletic gradualism. In a classic statement, ending with the sentenc that serves as masthead to this section, Zittel wrote (Eastman translation 1900, p. 10):

> Weighty evidence for the progressive evolution of organisms i afforded by fossil transitional series, of which a considerable number ar known to us, notwithstanding the imperfection of the palaeontologic record. By transitional series are meant a greater or lesser number o similar forms occurring through several successive horizons, and constituting a practically unbroken morphic chain . . . With increasin abundance of palaeontological material, the more numerous and mor complete are the series of intermediate forms which are brought to ligh

The last edition of Woods (1946) devotes three pages to evolution; all but tw paragraphs (one on ontogeny, the other on orthogenesis) to an exposition o phyletic gradualism (one page on the imperfection of the record, another o some rare examples of graded sequences).

Our current textbooks have changed the argument not at all. Moore Lalicker and Fischer (1952, p. 30), in listing the fossil record among "evidences of evolution," have only this to say about it: "Although lack o knowledge is immeasurably greater than knowledge, many lineages amon fossils of various groups have been firmly established. These demonstrate th transformation of one species or genus into another and thus constitu documentary evidence of gradual evolution." And Easton (1960, p. 34), citin the apotheosis of our achievements, writes: "An evolutionary series represents the peak of scientific accomplishment in organizing fossil invertebrate It purports to show an orderly progression in morphologic changes amon related creatures during successive intervals of time."

That these older texts hold so strongly to phyletic gradualism should surprise no one; harder to understand is the fact that virtually all modern texts repeat the same arguments even though their warrant had disappeared, as we shall now show, with the advent of the allopatric theory of speciation.

The Biospecies and Punctuated Equilibria: A Different Picture of Speciation

> Habits of thought in the tradition of a science are not readily changed, it is not easy to deviate from the customary channels of accumulated experience in conventionalized subjects.
>
> G. L. Jepsen, 1949, p. v

An irony. The formulation of the biological species concept was a major triumph of the synthetic theory (Mayr, 1963, abridged and revised 1970, remains the indispensable source on its meaning and implications). Since paleontology has always taken its conceptual lead from biology (with practical guidance from geology), it was inevitable that paleontologists should try to discover the meaning of the biospecies for their own science.

Here we meet an ironic situation: the taxonomic perspective—one of our persistent albatrosses—dictated an approach to the biospecies. Instead of extracting its insights about evolutionary processes, we sought only its prescriptions for classification. We learned that species are populations, that they are recognized in fossils by ranges of variability not by correspondence to idealized types. The "new systematics" ushered in the revolution in species-level classification that Darwin's theory had implied but not effected. In paleontology, its main accomplishment has been a vast condensation and elimination of spurious taxa established on typological criteria.

But the new systematics also rekindled a theoretical debate unsurpassed in the annals of paleontology for its ponderous emptiness: What is the nature of a paleontological species? In this reincarnation: can taxa designated as biospecies be recognized from fossils? Biologists insisted that the biospecies is a "real" unit of nature, a population of interacting individuals, reproductively isolated from all other groups. Yet its reality seemed to hinge upon what Mayr calls its "non-dimensional" aspect: species are distinct at any moment in time, but the boundaries between forms must blur in temporal extension—a continuous lineage cannot be broken into objective segments. Attempts to reconcile or divorce the non-dimensional biospecies and the temporal "paleospecies" creep on apace (Imbrie, 1957; Weller, 1961; McAlester, 1962; Shaw, 1969; and an entire symposium edited by Sylvester-Bradley, 1956); if obfuscation is any sign of futility, we offer the following as a plea for the termination of this discussion: "Such a plexiform lineage . . . constitutes a chronospecies (or paleospecies), and it is composed of many successional polytypic morphospecies ('holomorphospecies'), each of which is in theory the paleontological equivalent of a neontological biospecies" (Thomas, 1956, p. 24).

The discussion is futile for a very simple reason: the issue is insoluble: it is not a question of fact (phylogeny proceeds as it does no matter how we name its steps), but a debate about ways of ordering information. When Whitehead said that all philosophy was a footnote to Plato, he meant not only that Plato had identified all the major problems, but also that the problems were still debated because they could not be solved. The point is this: the hierarchical system of Linnaeus was established for his world: a world of discrete entities. It works for the living biota because most species are discrete at any moment in time. It has no objective application to evolving continua, only an arbitrary one based on subjective criteria for division. Linnaeus would not have set up the same system for our world. As Vladimir Nabokov writes in *Ada* (1969, p. 406): "Man . . . will never die, because there may never be a taxonomical point in his evolutionary progress that could be determined as the last stage of man in the cline turning him into Neohomo, or some horrible throbbing slime."

Then does the biospecies offer us nothing but semantic trouble? On one level, the answer is no because it can be applied with great effectiveness to past time-planes. But on another level, and this involves our irony, we must avoid the narrow approach that embraces a biological concept only when it can be transplanted bodily into our temporal taxonomy. The biospecies abounds with implications for the operation of evolutionary processes. Instead of attempting vainly to name successional taxa objectively in its light (McAlester, 1962), we should be applying its concepts. In the following section, we argue that one of these concepts—the theory of allopatric speciation—might reorient our picture for the origin of taxa.

Implications of allopatric speciation for the fossil record. We wish to consider an alternate picture to phyletic gradualism; it is based on a theory of speciation that arises from the behavior, ecology, and distribution of modern biospecies. First, we must emphasize that mechanisms of speciation can be studied directly only with experimental and field techniques applied to living organisms. No theory of evolutionary mechanisms can be generated directly from paleontological data. Instead, theories developed by students of the modern biota generate predictions about the course of evolution in time. With these predictions, the paleontologist can approach the fossil record and ask the following question: Are observed patterns of geographic and stratigraphic distribution, and apparent rates and directions of morphological change, consistent with the consequences of a particular theory of speciation? We can apply and test, but we cannot generate new mechanisms. If discrepancies are found between paleontological data and the expected patterns, we may be able to identify those aspects of a general theory that need improvement. But we cannot formulate these improvements ourselves.*

During the past thirty years, the allopatric theory has grown in popularity to become, for the vast majority of biologists, *the* theory of speciation. Its only serious challenger is the sympatric theory. Here we discuss only the implications of the allopatric theory for interpreting the fossil record of sexually-reproducing metazoans. We do this simply because it is the allopatric, rather than the sympatric, theory that is preferred by biologists. We shall therefore contrast the allopatric theory with the picture of phyletic gradualism developed in the last section.

Most paleontologists, of course, are aware of this theory, but the influence of phyletic gradualism remains so strong that discussions of geographic speciation are almost always cast in its light: geographic speciation is seen as the slow and steady transformation of two separated lineages—i.e., as *two* cases of phyletic gradualism (*figure 5–3*). Raup and Stanley (1971, p. 98), for example, write:

> Let us consider populations of a species living at a given time but not in geographic contact with each other. . . . Two or more segments of the species thus evolve and undergo *phyletic* speciation independently. . . . The distinction between phyletic and geographic speciation is to some extent artificial in that both processes depend on natural selection. The critical difference is that phyletic speciation is accomplished in the absence of geographic isolation and geographic speciation requires geographic isolation (italics ours).

The central concept of allopatric speciation is that new species can arise only when a small local population becomes isolated at the margin of the geographic range of its parent species. Such local populations are termed *peripheral isolates*. A peripheral isolate develops into a new species if *isolating mechanisms* evolve that will prevent the re-initiation of gene flow if the new form re-encounters its ancestors at some future time. As a consequence of the allopatric theory, new fossil species do not originate in the place where their ancestors lived. It is extremely improbable that we shall be able to trace the gradual splitting of a lineage merely by following a certain species up through a local rock column.

Another consequence of the theory of allopatric processes follows: since selection always maintains an equilibrium between populations and their local environment, the morphological features that distinguish the descendant

---

* The rate and direction of morphological change over long periods of time is the most obvious kind of evolutionary pattern that we can test against predictions based on processes observed over short periods of time by neontologists. We try to do this in the next section.

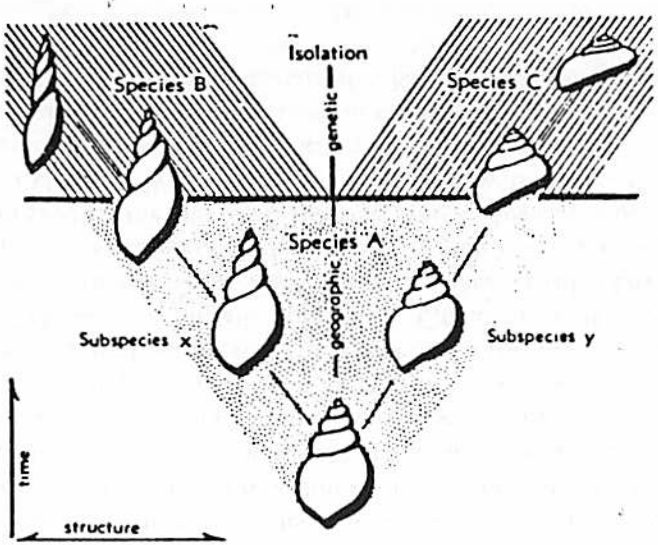


Figure 5–3: A hypothetical case of geographic speciation viewed from th perspective of phyletic gradualism—slow and gradual transformation i two lineages. From Moore, Lalicker, and Fischer, 1952; figure 1–15.

species from its ancestor are present close after, if not actually prior to, th onset of genetic isolation. These differences are often accentuated if the tw species become sympatric at a later date (character displacement—Brown an Wilson, 1956). In any event, most morphological divergence of a descendan species occurs very early in its differentiation, when the population is sma and still adjusting more precisely to local conditions. After it is fully estab lished, a descendant species is as unlikely to show gradual, progressiv change as is the parental species. Thus, in the fossil record, we should n expect to find gradual divergence between two species in an ancestra descendant relationship. Most evolutionary changes in morphology occur i a short period of time relative to the total duration of species. After th descendant is established as a full species, there will be little evolutionar change except when the two species become sympatric for the first time.

These simple consequences of the allopatric theory can be combined into a expected pattern for the fossil record. Using stratigraphic, radiometric, c biostratigraphic criteria (for organisms other than those under study), w establish a regional framework of correlation. Starting with these correlation patterns of geographic (not stratigraphic) variation among samples of fossi should appear. Tracing a fossil species through any local rock column, s long as no drastic changes occur in the physical environment, should produc *no* pattern of constant change, but one of oscillation in mean values. Closel related (perhaps descendant) species that enter the rock column should appear suddenly and show no intergradation with the "ancestral" species in morphological features that act as inter-specific differentia. There should be no gradual divergence between the two species when both persist for some time to higher stratigraphic levels. Quite the contrary—it is likely that the two species will display their greatest difference when the descendant first appears. Finally, in exceptional circumstances, we may be able to identify the general area of the ancestor's geographic range in which the new species arose.

Another conclusion is that time and geography, as factors in evolution, are not so comparable as some authors have maintained (Sylvester-Bradley, 1951). The allopatric theory predicts that most variation will be found among samples drawn from different geographic areas rather than from different stratigraphic levels in the local rock column. The key factor is adjustment to a heterogeneous series of micro-environments vs. a general pattern of stasis through time.

In summary, we contrast the tenets and predictions of allopatric speciation with the corresponding statements of phyletic gradualism previously given:

(1) New species arise by the splitting of lineages.

(2) New species develop rapidly.

(3) A small sub-population of the ancestral form gives rise to the new species.

(4) The new species originates in a very small part of the ancestral species' geographic extent—in an isolated area at the periphery of the range.

These four statements again entail two important consequences:

(1) In any *local* section containing the ancestral species, the fossil record for the descendant's origin should consist of a sharp morphological break between the two forms. This break marks the migration of the descendant, from the peripherally isolated area in which it developed, into its ancestral range. Morphological change in the ancestor, even if directional in time, should bear no relationship to the descendant's morphology (which arose in response to local conditions in its isolated area). Since speciation occurs rapidly in small populations occupying small areas far from the center of ancestral abundance, we will rarely discover the actual event in the fossil record.

(2) Many breaks in the fossil record are real: they express the way in which evolution occurs, not the fragments of an imperfect record. The sharp break in a local column accurately records what happened in that area through time. Acceptance of this point would release us from a self-imposed status of inferiority among the evolutionary sciences. The paleontologist's gut-reaction is to view almost any anomaly as an artifact imposed by our institutional millstone—an imperfect fossil record. But just as we now tend to view the rarity of Precambrian metazoans as a true reflection of life's history rather than a testimony to the ravages of metamorphism or the lacunae of Lipalian intervals, so also might we reassess the smaller breaks that permeate our Phanerozoic record. We suspect that this record is much better (or at least much richer in optimal cases) than tradition dictates.

Problems of phyletic gradualism. In our alternate picture of phyletic gradualism, we are not confronted with a self-contained theory from modern biology. The postulated mechanism for gradual uni-directional change is "orthoselection," usually viewed as a constant adjustment to a uni-directional change in one or more features of the physical environment. The concept of orthoselection arose as an attempt to remove the explanation of gradual morphological change from the realm of metaphysics ("orthogenesis"). It does *not* emanate from *Drosophila* laboratories, but represents a hypothetical extrapolation of selective mechanisms observed by geneticists.

Extrapolation of gradual change under selection to a complete model for the origin of species fails to recognize that speciation is primarily an ecological and geographic process. Natural selection, in the allopatric theory, involves adaptation to local conditions and the elaboration of isolating mechanisms. Phyletic gradualism is, in itself, an insufficient picture to explain the origin of diversity in the present, or any past, biota.

Although phyletic gradualism prevails as a picture for the origin of new species in paleontology, very few "classic" examples purport to document it. A few authors (MacGillavry, 1968, Eldredge, 1971) have offered a simple and literal interpretation of this situation: *in situ*, gradual, progressive evolutionary change is a rare phenomenon. But we usually explain the paucity of cases by a nearly-ritualized invocation of the inadequacy of the fossil record. It *is* valid to point out the rarity of thick, undisturbed, highly fossiliferous rock sections in which one or more species occur continuously throughout the sequence. Nevertheless, if most species evolved according to the tenets of phyletic gradualism, then, no matter how discontinuous a species' occurrence in thick sections, there should be a shift in one or more variables from sample to sample up the section. This is, in fact, the situation in most cases of postulated gradualism: the "gradualism" is represented by dashed lines connecting known samples. This procedure provides an excellent example of the role of preconceived pictures in "objectively documented" cases. One of the early "classics" of phyletic gradualism, Carruthers' (1910) study of the Carboniferous rugose coral *Zaphrentites delanouei* (Milne-Edwards and Haime) and its reinterpretation by Sylvester-Bradley (1951), is of this kind. We do not say that the analysis is incorrect; the *Z. delanouei* stock may have evolved as claimed. We merely wish to show how the *a priori* picture of phyletic gradualism has imposed itself upon limited data.

How pervasive, then, is gradualism in these quasi-continuous sequences? A number of authors (including, *inter alia*, Kurtén, 1965, MacGillavry, 1968, and Eldredge, 1971) have claimed that most species show little or no change throughout their stratigraphic range. But though it is tempting to conclude that gradual, progressive morphological change is an illusion, we recognize that there is little hard evidence to support either view.

As a final, and admittedly extreme, example of *a priori* beliefs in phyletic gradualism, we cite the work of Brace (1967) on human evolution. This is all the more instructive since most paleoanthropologists, in reversing an older view that Brace still maintains, now claim that hominid evolution involves speciation by splitting as well as phyletic evolution by transformation (seen especially in the presumed coexistence of two australopithecine species in the African lower Pleistocene—Howell, 1967; Tobias, 1965; Pilbeam, 1968; Pilbeam and Simons, 1965). Brace (1967) has claimed that the fossil record of man includes four successive "stages" in direct ancestral-descendant relation. These are the Australopithecine (with two successive "phases"—the australopithecus and paranthropus), the Pithecanthropus, the Neanderthaloid, and, finally, the Modern Stage. In discussing the history of paleoanthropology, Brace shows that most denials of ancestral-descendant relationships among hominid fossils stem from a desire to avoid the conclusion that *Homo sapiens* evolved from some "lower," more "brutish" form. But Brace has lumped all such analyses under the catch phrase "hominid catastrophism." Hominid catastrophism, according to Brace, is the denial of ancestral-descendant relationships among fossils, with the invocation of extinction and subsequent migrations of new populations that arose by successive creation. Such views are, of course, absurd, but Brace would include *all* cladistic interpretations of the hominid record within "hominid catastrophism." To view hominid phylogeny as a gradual, progressive, unilineal process involving a series of stages, Brace claims, is the interpretation most consonant with evolutionary theory. His interpretation of phylogeny may be correct (though most experts deny it), but he is seriously wrong to claim that phyletic gradualism is the picture most consistent with modern biological thought. Quite apart from the issue of probable overlap in the ranges of his stages, it would be of great interest to determine the degree of stasis attained by them during any reasonably long period of time.

Application of allopatric concepts to paleontological examples. At this point, there is some justification for concluding that the picture of phyletic gradualism is poorly documented indeed, and that most analyses purporting to illustrate it directly from the fossil record are interpretations based on a preconceived idea. On the other hand, the alternative picture of stasis punctuated by episodic events of allopatric speciation rests on a few general statements in the literature and a wealth of informal data. The idea of *punctuated equilibria* is just as much a preconceived picture as that of phyletic gradualism. We readily admit our bias towards it and urge readers, in the ensuing discussion, to remember that our interpretations are as colored by our preconceptions as are the claims of the champions of phyletic gradualism by theirs. We merely reiterate: (1) that one must have some picture of speciation in mind, (2) that the data of paleontology cannot decide which picture is mor
adequate, and (3) that the picture of punctuated equilibria is more in accor
with the process of speciation as understood by modern evolutionists.

We could cite any number of reported sequences that fare better und
notions of allopatric processes than under the interpretation of phylet
gradualism that was originally applied. This is surely true for all or part of th
three warhorses of the English literature: horses themselves, the Cretaceo
echinoid *Micraster*, and the Jurassic oyster *Gryphaea*. Simpson (1951) h
shown that the phylogeny of horses is a luxuriant, branching bush, not t
ladder to one toe and big teeth that earlier authors envisioned (Matthew an
Chubb, 1921). Nichols (1959) believes that *Micraster senonensis* was a migra
from elsewhere and that it did not arise and diverge gradually from
cortestudinarium as Rowe (1899) had maintained. Hallam (1959, 1962) h
argued that the transition from *Liostrea* to *Gryphaea* was abrupt and th
neither genus shows *any* progressive change through the basal Liassic zon
contrary to Trueman's claim (1922, p. 258) that: "It is doubtful whether a
better example of lineage of fossil forms could be found." Gould (1971b a
in press) has confirmed Hallam's conclusions. Hallam interprets the sudd
appearance of *Gryphaea* as the first entry into a local rock column of
species that had evolved rapidly elsewhere. He writes (1962, p. 574): "T
interpretation is more in accord with the experience of most invertebrate pa
ontologists who, despite continued collecting all over the world and an ev
increasing amount of research, find 'cryptogenic' genera and species far mo
commonly than they detect gradual trends or lineages. The sort of evolutio
tentatively propose for *Gryphaea* could in fact be quite normal among
invertebrates." We agree.

We choose, rather, to present two examples from our own work which
believe are interpreted best from the viewpoint of allopatric speciation.
prefer to emphasize our own work simply because we are most familiar w
it and are naturally more inclined to defend our interpretations.

Gould (1969) has analyzed the evolution of *Poecilozonites bermuden*
zonatus Verrill, a pulmonate snail, during the last 300,000 years of
Bermudian Pleistocene. The specimens were collected from an alternat
sequence of wind-blown sands and red soils. Formational names, domin
lithologies, and glacial-interglacial correlations are given in *table 5-1*.

The small area and striking differentiation of stratigraphic units in
Bermudian Pleistocene permit a high degree of geographic and tempo
control. *P. bermudensis* (Pfeiffer) is plentiful in all post-Belmont formatio
in addition, one subspecies, *P.b. bermudensis*, is extant and available for stu
in the laboratory.

Distinct patterns of color banding differentiate an eastern from a west
population of *P. bermudensis zonatus*. The boundary between these
groups is sharp, and there are no unambiguous cases of introgression.

bermudensis zonatus was divided into two stocks, evolving in parallel with little gene flow between them, throughout the entire interval of Shore Hills to Southampton time. Both eastern and western *P.b. zonatus* became extinct sometime after the deposition of Southampton dunes; they were replaced by *P.b. bermudensis*, a derivative of eastern *P.b. zonatus* which had been evolving separately in the area of St. George's Island since St. George's time. Gould (1969, 1970b) has discussed the parallel oscillation of several morphological features in both stocks of *P.b. zonatus*; these are adaptive shifts in response to glacially-controlled variations in climate. Both stocks exhibit stability in other features that serve to distinguish them from their nearest relatives. There is no evidence for any gradual divergence between eastern and western *P.b. zonatus*.

Several samples of *P. bermudensis* share many features that distinguish them from *P. bermudensis zonatus*. These characters can be arranged in four categories: color, general form of the spire, thickness of the shell, and shape of the apertural lip. The ontogeny of *P.b. zonatus* illustrates the interrelation of these categories. Immature shells of *P.b. zonatus* are weakly colored, relatively wide, lack a callus, and have the lowest portion of the outer apertural lip at the umbilical border. This combination of character states is exactly repeated in the large mature shells of non-*zonatus* samples of *P. bermudensis*. Since every ontogenetic feature developed at or after the fifth whorl in non-*zonatus* samples is attained by whorls 3-4 in *P.b. zonatus*, Gould (1969) concludes that the non-*zonatus* samples of *P. bermudensis* are derived by paedomorphosis from *P.b. zonatus*.

These paedomorphic samples range through the entire interval of Shore Hills to Recent. The most obvious hypothesis would hold that they constitute a continuous lineage evolving separately from *P.b. zonatus*. Gould rejects this and concludes that paedomorphic offshoots arose from the *P.b. zonatus* stock at four different times; the arguments are based on details of stratigraphic and geographic distrubution, as well as on morphology.

Figure 5-4 summarizes the history of splitting in the *P.b. zonatus* lineage. The earliest paedomorph, *P.b. fasolti* Gould, occurs in the Shore Hills Formation within the geographic range of eastern *P.b. zonatus*. *P.b. fasolti* and the contemporary population of eastern *P.b. zonatus* share a unique set of morphological features including, *inter alia*, small size at any given whorl, low spire, relatively wide shell, and a wide umbilicus. These features unite the Shore Hills paedomorph and non-paedomorph, and set them apart from all post-Shore Hills *P. bermudensis*.

In the succeeding Harrington Formation, paedomorphic samples of *P. bermudensis* lived in both the eastern and western geographic regions of *P.b. zonatus*. The eastern paedomorph, *P.b. sieglindae* Gould, may have evolved from the Shore Hills paedomorph, *P.b. fasolti*. However, both *P.b. sieglindae*

Table 5-1. Stratigraphic column of Bermuda.

Formation	*Description*	*Interpretation*
Recent	Poorly developed brownish soil or crust	Interglacial
Southampton	Complex of eolianites and discontinuous unindurated zones	„
St. George's	Red paleosol of island wide extent	Glacial
Spencer's Point	Intertidal marine, beach and dune facies	Interglacial
Pembroke	Extensive eolianites and discontinuous unindurated zones	„
Harrington	Fairly continuous unindurated layer with shallow water marine and beach facies	„
Devonshire	Intertidal marine and poorly developed dune facies	„
Shore Hills	Well-developed red paleosol of island-wide extent	Glacial
Belmont	Complex shallow water marine, beach and dune facies	Interglacial
Soil (?)	A reddened surface rarely seen in the Walsingham district	Glacial?
Walsingham	Highly altered elolianites	Interglacial

and the contemporaneous population of eastern *P.b. zonatus* lack the distinc tive features of all Shore Hills *P. bermudensis* and a more likely hypothesi holds that the features uniting all post-Shore Hills *P. bermudensis* wer evolved only once. If this is the case, *P.b. sieglindae* is a second paedomorphi derivative of eastern *P.b. zonatus*.

P.b. sieglindae differs from its contemporary paedomorph *P.b. siegmun* Gould in that each displays the color pattern of the local non-paedomorph Very simply, *P.b. sieglindae* is found in eastern Bermuda and shares th banding pattern of eastern *P.b. zonatus*, while *P.b. siegmundi* is found i western Bermuda and has the same color pattern as western *P.b. zonatus*. I addition, both *P.b. sieglindae* and *P.b. siegmundi* evolved at the periphery c the known range of their putative ancestors. The independent derivation c the two Harrington paedomorphs from the two stocks of *P.b. zonatus* seem clear.

Finally, the living paedomorph, *P.b. bermudensis*, first appears in the St George's Formation on St. George's Island. While St. George's Island i within the geographic range of eastern *P.b. zonatus*, it is far removed from th

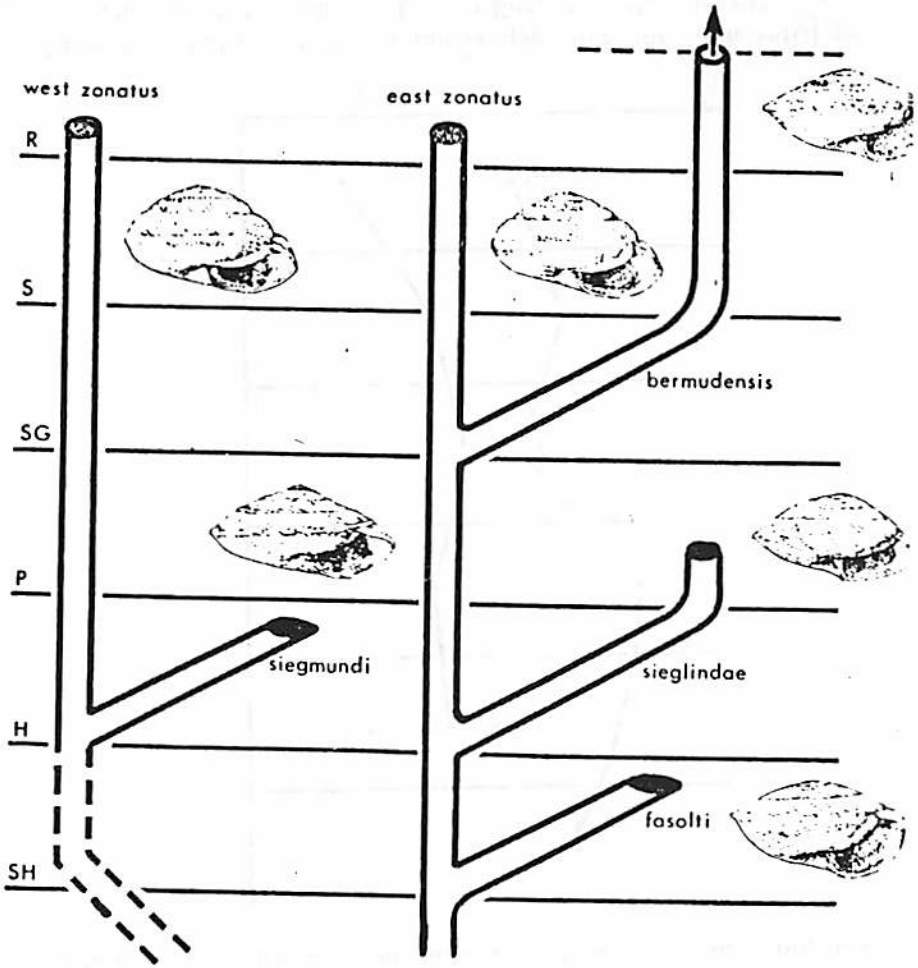


Figure 5-4: Reconstruction of the phylogenetic history of *P. bermudensis* showing iterative development of paedomorphic subspecies. SH—Shore Hills; H—Harrington; P—Pembroke; SG—St. George's; S—Southampton; R—Recent. From Gould, 1969; figure 20.

area in which *P.b. sieglindae* arose and lived. Gould concludes that *P.b. sieglindae* was a short-lived population that never enjoyed a wide geographic distribution; he estimates that the Pembroke population's range did not exceed 200 meters. Although there is little morphological evidence to support it, Gould recognizes a fourth paedomorphic subspecies, *P.b. bermudensis*, derived directly from (eastern) *P.b. zonatus*. The conclusion is based upon geographic and stratigraphic data.

Gould (1969) has advanced an adaptive explanation for the four separate origins of paedomorphic populations from *P.b. zonatus*. This explanation, based on the value of thin shells in lime-poor soils, need not be elaborated here. What is important, for our purposes, is to emphasize that the reconstruction of phylogenetic histories for the paedomorphs involves (1) attention to geographic data (the allopatric model), (2) discontinuous stratigraphic occurrence (a more literal interpretation of the fossil record), and (3) formal arguments based on morphology. It is entirely possible, from morphological data alone, to interpret the three paedomorphs of the eastern *zonatus* area as a gradational biostratigraphic series. *Figure 5-5* shows a tempting interpretation of phyletic gradualism for "lower eccentricity," an apertural

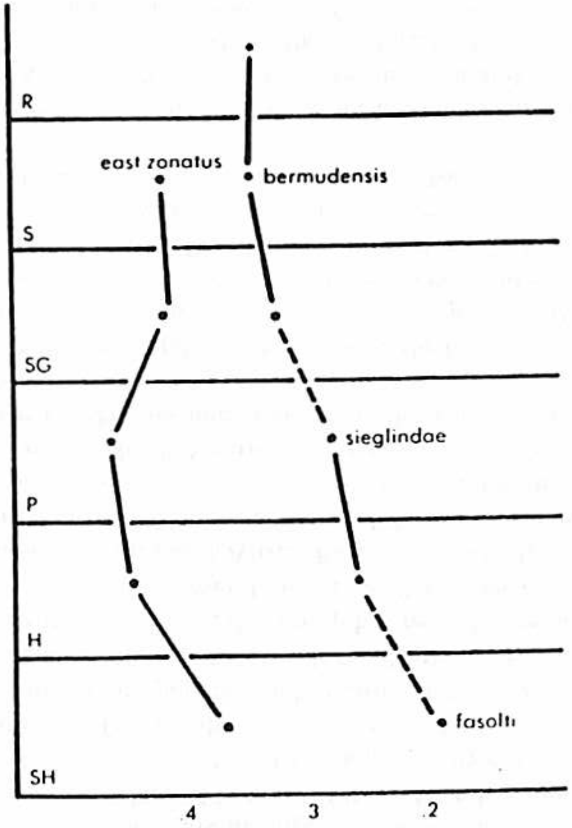


Figure 5-5: Plot of means of mean sample values of "lower eccentricity" in *P. bermudensis*. Dashed lines show the phylogeny of the three paedomorphs of eastern *zonatus* as a direct ancestral-descendant sequence, and offer a tempting instance of phyletic gradualism. Abbreviations as in *figure 5-4*.

variable. Values gradually increase through time. *Figure 5–6*, however, co founds this interpretation by showing that stratigraphic variability in "diffe ential growth ratio" within both *P.b. sieglindae* and *P.b. bermudensis* varies a direction *opposite* to the net stratigraphic "trend": *P.b. fasolti—P. sieglindae—P.b. bermudensis*: this could be read to indicate that each sul species is unique. In fact, neither graph affords sufficient evidence to warrai either conclusion. Morphology, stratigraphy, and geography must all t evaluated.

The phylogenetic history of the trilobite *Phacops rana* (Green) froı the Middle Devonian of North America (Eldredge, 1971; 1972) provid another example of the postulated operation of allopatric processes. As i *Poecilozonites bermudensis*, full genetic isolation was probably not estal lished between "parent" and "daughter" taxa; this conclusion, based on

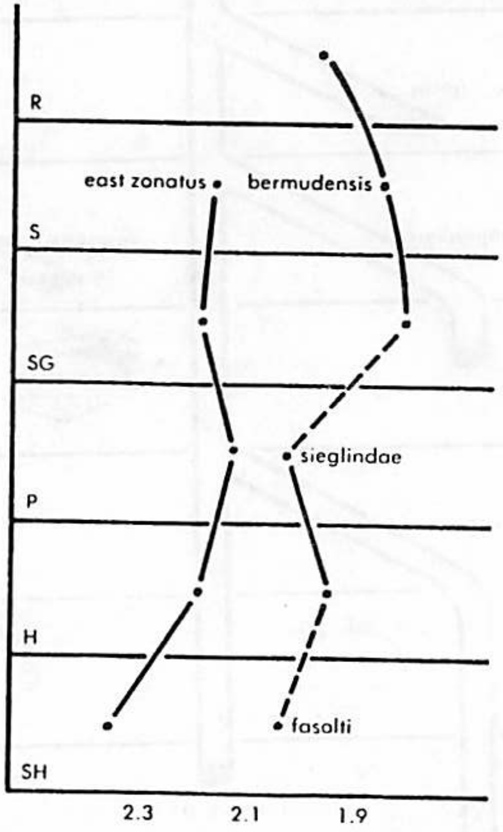


Figure 5–6: Plot of means of mean sample values for "differential growth ratio" in *P. bermudensis*. Dashed lines show the interpretation of the phylogeny of the three paedomorphs as a direct ancestral-descendant sequence. Abbreviations as in *figure 5–4*.

inferences from morphological variability, may be unwarranted. For our purposes, it does not matter whether we are dealing with four subspecies of *P. rana*, or four separate species of *Phacops*, including *P. rana* and its three closest relatives. The basic mode of evolution underlying the group's phylogenetic history as a whole is the same in either case.

Features of eye morphology exhibit the greatest amount of variation among samples of *P. rana*. Lenses are arranged on the visual surface of the eye in vertical dorso-ventral files (Clarkson, 1966). A stable number of dorso-ventral files, characteristic of the entire sample in any population, is reached early in ontogeny. The number of dorso-ventral (d.-v.) files is the most important feature of interpopulational variation in *P. rana*.

The closest known relative of *P. rana* is *P. schlotheimi* (Bronn) *s. l.*, from the Eifelian of Europe and Africa; this group has recently been revised by C. J. Burton (1969). In addition, several samples of *P. rana* have been found in the Spanish Sahara in northwestern Africa (Burton and Eldredge, in preparation). *P. schlotheimi* and the African specimens of *P. rana* are most similar to *P. rana milleri* Stewart and *P. rana crassituberculata* Stumm, the two oldest subspecies of *P. rana* in North America. All these taxa possess 18 dorso-ventral files. Eldredge (1972) concludes that 18 is the primitive number of d.-v. files for all North American *Phacops rana*.

Figure 5–7 summarizes relationships among the four subspecies of *P. rana* without regard to stratigraphic occurrence. The oldest North American *P. rana* occurs in the Lower Cazenovian Stage of Ohio and central New York State. All have 18 d.-v. files. Populations with 18 d.-v. files (*P. rana milleri* and *P. rana crassituberculata*) persist into the Upper Cazenovian Stage in the epicontinental seas west of the marginal basin in New York and the Appalachians.

Of the two samples the one that displays intra-populational variation in d.-v. file number occurs in the Lower Cazenovian of central New York. Some specimens have 18 d.-v. files, while others reduce the first d.-v. file to various degrees; a few lack it altogether. *All P. rana* from subsequent, younger horizons in New York and adjacent Appalachian states have 17 dorso-ventral files. Apparently, 17 d.-v. file *P. rana rana* arose from an 18 d.-v. file population on the northeastern periphery of the Cazenovian geographic range of *P. rana*. Seventeen d.-v. file *P. rana* persist, unchanged in most respects, through the Upper Cazenovian, Tioughniogan, and Taghanic Stages in the eastern marginal basin. Seventeen d.-v. file *P. rana rana* first appears in the shallow interior seas at the beginning of the Tioughniogan Stage, replacing the 18 d.-v. file populations that apparently became extinct during a general withdrawal of seas from the continental interior. All Tioughniogan *P. rana* possess 17 dorso-ventral files.

A second, similar event involving reduction in dorso-ventral files occurred during the Taghanic. Here again, a variable population inhabited the eastern

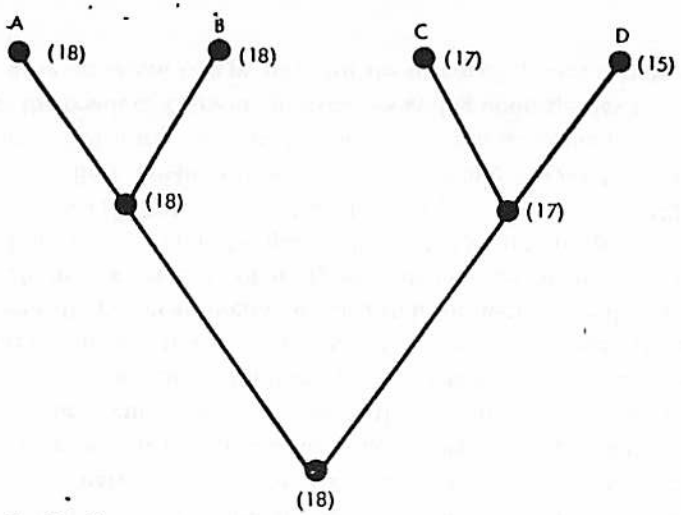


Figure 5–7: Outline of relationships of four subspecies of *Phacops rana*. A—*Phacops rana crassituberculata* Stumm; B—*Phacops rana milleri* Stewart; C—*Phacops rana rana* (Green); D—*Phacops rana norwoodensis* Stumm. Numbers in parentheses refer to number of dorso-ventral files typical of subspecies or hypothesized to characterize condition of common ancestor.

marginal basin in New York. This suggests that, once more, reduction in d.-v. files occurred allopatrically on the periphery of the known range of *P. rana rana*. The subsequent spread of stabilized, 15 d.-v. file *P. rana norwoodensis* through the Taghanic seas of the continental interior was instantaneous in terms of our biostratigraphic resolution. *Figure 5–8* summarizes this interpretation of the history of *P. rana*.

Under the tenets of phyletic gradualism, this story has a different (and incorrect) interpretation: the three successional taxa of the epeiric seas form an *in situ* sequence of gradual evolutionary modification. The sudden transitions from one form to the next are the artifact of a woefully incomplete fossil record. Most evolutionary change occurred during these missing intervals: fill in the lost pieces with an even dotted line.

If the interpreter pays attention to geographic detail, however, quite a different tale emerges, one that allows a more literal reading of the fossil record. Now the story is one of stasis: no variation in the most important feature of discrimination (number of d.-v. files)—actually a complex of highly interrelated variables) through long spans of time. Two samples displaying intra-populational variation in numbers of d.-v. files identify relatively "sudden" events of reduction in files on the periphery of the species' geographic range. These two samples, moreover, have a very short stratigraphic, and very restricted geographic, distribution.

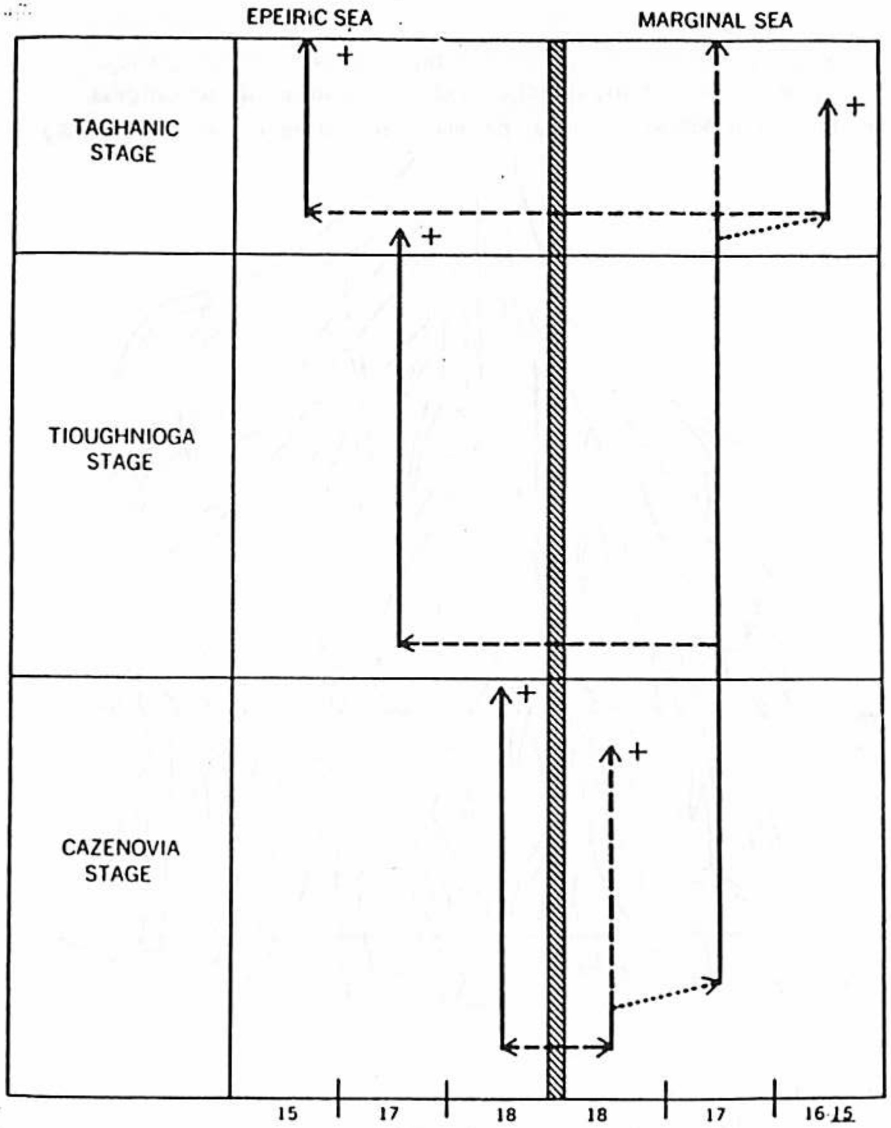


Figure 5–8: Hypothesized phylogeny of the *Phacops rana* stock in the Middle Devonian of North America. Numbers at the base of the diagram refer to the population number of dorso-ventral files. Dotted lines: origin of new (reduced) number of d.-v. files in a peripheral isolate; horizontal dashed lines: migration; vertical solid lines: presence of taxon in indicated area; dashed vertical lines: persistence of ancestral stock in a portion of the marginal sea other than that in which the derived taxon occurs. Crosses denote final disappearance; for fuller explanation, see text.

Our two examples, so widely separated in scale, age, and subject, have muc in common as exemplars of allopatric processes. Both required an attentio to details of *geographic* distribution for their elucidation. Both involved *more literal* reading of the fossil record than is allowed under the unconscio guidance of phyletic gradualism. Both are characterized by *rapid* evolutionar events punctuating a history of stasis. These are among the expected co sequences if most fossil species arose by allopatric speciation in small, per pherally isolated populations. This alternative picture merely represents th application to the fossil record of the dominant theory of speciation in moder evolutionary thought. We believe that the consequences of this theory ar more nearly demonstrated than those of phyletic gradualism by the foss record of the vast majority of Metazoa.

Some Extrapolations to Macroevolution

Before 1930, paleontology sought a separate theory for the causes of macr evolution. The processes of microevolution (including the origin of specie were deemed insufficient to generate the complexity and diversity of life, eve under the generous constraint of geological time; a variety of special caus were proposed—vitalism, orthogenesis, racial "life" cycles, and univers acceleration in development to name just a few.

However, the advent of the "modern synthesis" inspired a reassessmen that must stand as the major conceptual advance in 20th-century paleontol ogy. Special explanations for macroevolution were abandoned for a simpl fying theory of extrapolation from species-level processes. All evolutionar events, including those that seemed most strongly "directed" and greatl extended in time, were explained as consequences of mutation, recombinatio selection, etc.—i.e., as consequences only of the phenomena that produc evolution in nature's real taxon, the species. (The modern synthesis receive its name because it gathered under one theory—with population genetics a its core—the events in many subfields that had previously been explained b special theories unique to that discipline. Such an occurrence marks scientifi "progress" in its truest sense—the replacement of special explanatio carrying little power in prediction or extension with general theories, rich i implications and capable of unifying a diverse set of phenomena that ha seemed unrelated. Thus Simpson (1944, 1953) did for paleontology wh Dobzhansky (1937) had done for classical genetics, Mayr (1942) for syst matics, de Beer (1940) for development, White (1954) for cytology, an Stebbins (1950) for systematic botany—he exemplified the phenomena of h field as the result of Darwinian processes acting upon species.)

We have discussed two pictures for the origin of species in paleontology. I the perspective of a species-extrapolation theory of macroevolution, w should now extend these pictures to see how macroevolution proceeds under their guidance. If actual events, as recorded by fossils, fit more comfortably with the predictions of either picture, this will be a further argument for that picture's greater adequacy.

Under phyletic gradualism, the history of life should be one of *stately unfolding*. Most changes occur slowly and evenly by phyletic transformation; splitting, when it occurs, produces a slow and very gradual divergence of forms (Weller's (1969) tree of life—reproduced as *figure 5–9*—records the extrapolation of this partisan view, not a neutral hatrack for the fossils themselves). We have already named our alternate picture for its predicted extrapolation—*punctuated equilibria*. The theory of allopatric speciation implies

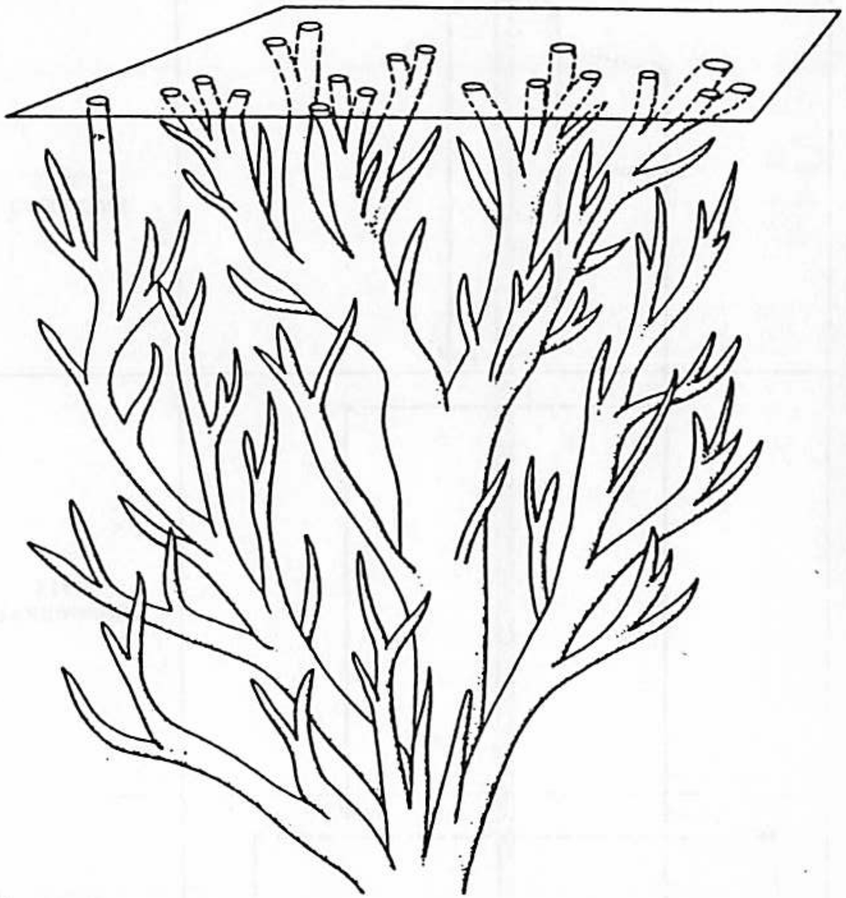


Figure 5–9: The "Tree of Life" viewed from the perspective of phyletic gradualism. Branches diverge gradually one from the other. A slow and relatively equal rate of evolution pervades the system. From Weller, 1969; figure 637.

that a lineage's history includes long periods of morphologic stability, punc tuated here and there by rapid events of speciation in isolated subpopulations

We now consider two phenomena of macroevolution as case studies of ou extrapolated pictures. The first is widely recognized as anomalous under th unconscious guidance of stately unfolding; it emerges as an expectation unde the notion of punctuated equilibria. The second phenomenon seems, super ficially, to have an easier explanation under stately unfolding, but we sha argue that it has a more interesting interpretation when viewed with th picture of punctuated equilibria.

(1) "*Classes*" *of great number and low diversity*

To many paleontologists, nothing is more distressing than the curren situation in echinoderm systematics. Ubaghs (1967), in his contribution to th *Treatise on Invertebrate Paleontology*, recognizes 20 classes and at least on has been added since then—Robison and Sprinkle's (1969) ctenocystoids Yet, although all appeared by the Ordovician, only five survived the Devonian Moreover, although each class has a distinct Bauplan, many display a diversity often considered embarrassingly small for so exalted a taxonomic rank—the *Treatise* describes eight classes with five or fewer genera; five o these include but a single genus (as does the new ctenocystoids).

There are two aspects to this tale that fit poorly with the traditional view of stately unfolding:

(1) The presence of 21 classes by the Ordovician, coupled with their presumed monophyletic descent, requires extrapolation to a common ancestor uncomfortably far back in the Precambrian if Ordovician diversity is the apex of a gradual unfolding. Yet current views of Precambrian evolution will not happily accommodate a complex metazoan so early (Cloud, 1968).

(2) We expect that successively higher ranks of the taxonomic hierarchy will contain more and more taxa: a class with one genus is anomalous and we are led either to desperate hopes for synonymy or, once again, to our old assumption—that we possess a fragmentary record of a truly diverse group. Yet this expectation is no consequence of the logic of taxonomy (which demands only that each taxon be *as* inclusive as the lower ones it incorporates); it arises, rather, from a picture of stately unfolding. In *figure 5–9*, a new higher taxon attains its rank *by virtue of* its diversity—an evenly progressing, evenly diverging set of branches cannot produce such a taxon with limited diversity, for a lineage "graduates" from family to order to class only as it persists to a tolerable age and branches an acceptable number of times.

With the picture of punctuated equilibria, however, classes of small membership are welcome and echinoderm evolution becomes more intriguing than bothersome. Since speciation is rapid and episodic, repeated splitting during short intervals is likely when opportunities for full speciation following isolation are good (limited dangers of predation or competition in peripheral environments, for example—a likely Lower Cambrian situation). When these repeated splits affect a small, isolated lineage; when adaptation to peripheral environments involves new modes of feeding, protection, and locomotion; and when extinction of parental species commonly follows the migration of descendants to the ancestral area, then very distinct phenons with few species will develop. Since higher taxa are all "arbitrary" (they reflect no interacting group in nature, but rather a convenient arrangement of species that violates no rule of monophyly, hierarchical ordering, etc.), we believe that they should be defined by morphology. Criteria of diversity are too closely tied to partisan pictures; morphology, though not as "objective" as some numerical taxonomists claim, is at least more functional for information retrieval.

(2) *Trends*

Trends, or biostratigraphic character gradients, are frequently mentioned as basic features of the fossil record. Sequences of fossils, said to display trends, range from the infraspecific through the very highest levels of the taxonomic hierarchy. Trends at and below the species level were discussed in the previous section, but the relation between phyletic gradualism and trends among related clusters of species—families or orders—remains to be examined.

Many, if not most, trends involving higher taxa may simply reflect a selective rendering of elements in the fossil record, chosen because they seem to form a morphologically-graded series coincident with a progressive biostratigraphic distribution. In this sense, trends may represent simple extrapolations of phyletic gradualism.

But a claim that all documented trends are just unwarranted extrapolations based on a preconception would be altogether too facile an explanation for the large numbers of trends cited in the literature. For this discussion, we accept trends as a real and important phenomenon in evolution, and adopt the simple definition given by MacGillavry (1968, p. 72): "A trend is a direction which involves the *majority* of related lineages of a group" (our italics).

If trends are real and common, how can they be reconciled with our picture, in which speciation occurs in peripheral isolates by adaptation to local conditions and the perfection of isolating mechanisms? The problem may be stated in another way: Sewall Wright (1967, p. 120) has suggested that, just as mutations are stochastic with respect to selection within a population, so might speciation be stochastic with respect to the origin of higher taxa. As a slight extension of that statement, we might claim that adaptations to local conditions by peripheral isolates are stochastic with respect to long-term, net directional change (trends) within a higher taxon as a whole. We are left with a bit of a paradox: to picture speciation as an allopatric phenomenon, involving rapid differentiation within a general, long-term picture of stasis, is to deny the picture of directed gradualism in speciation. Yet, superficially at least, this directed gradualism is easier to reconcile with valid cases of long-term trends involving many species.

MacGillavry's definition of a trend removes part of the problem by using the expression "majority of related lineages." This frees us from the constraint of reconciling *all* events of adaptation to local conditions in peripheral isolates, with long-term, net directional change.

A reconciliation of allopatric speciation with long-term trends can be formulated along the following lines: we envision multiple "explorations" or "experimentations" (see Schaeffer, 1965)—i.e., invasions, on a stochastic basis, of new environments by peripheral isolates. There is nothing inherently directional about these invasions. However, a subset of these new environments might, in the context of inherited genetic constitution in the ancestral components of a lineage, lead to new and improved efficiency. Improvement would be consistently greater within this hypothetical subset of local conditions that a population might invade. The overall effect would then be one of net, apparently directional change: but, as in the case of selection upon mutations, the initial variations would be stochastic with respect to this change (*figure 5-10*). We postulate no "new" type of selection. We simply state a view of long-term, superficially "directed" phenomena that is in accord with the theory of allopatric speciation, and also avoids the largely untestable concept of orthoselection.

Conclusion: Evolution, Stately or Episodic?

Heretofore, we have spoken of the morphological stability of species in time without examining the reasons for it. The standard definition of a biospecies—as a group of actually or potentially reproducing organisms sharing a common gene pool—specifies the major reason usually cited: gene flow. Since the subpopulations of a species adapt to a range of differing local environments, we might expect these groups to differentiate, acquire isolating mechanisms and, eventually, to form new species. But gene flow exerts a homogenizing influence "to counteract local ecotypic adaptation by breaking up well-integrated gene complexes" (Mayr, 1963, p. 178). The role of gene flow is recognized in the central tenet of allopatric speciation: speciation occurs in *peripheral* isolates because only geographic separation from the parental species can reduce gene flow sufficiently to allow local differentiation to proceed to full speciation.

Recently, however, a serious challenge to the importance of gene flow in species' cohesion has come from several sources (Ehrlich and Raven, 1969, for example). Critics claim that, in most cases, gene flow is simply too restricted to exert a homogenizing influence and prevent differentiation. This

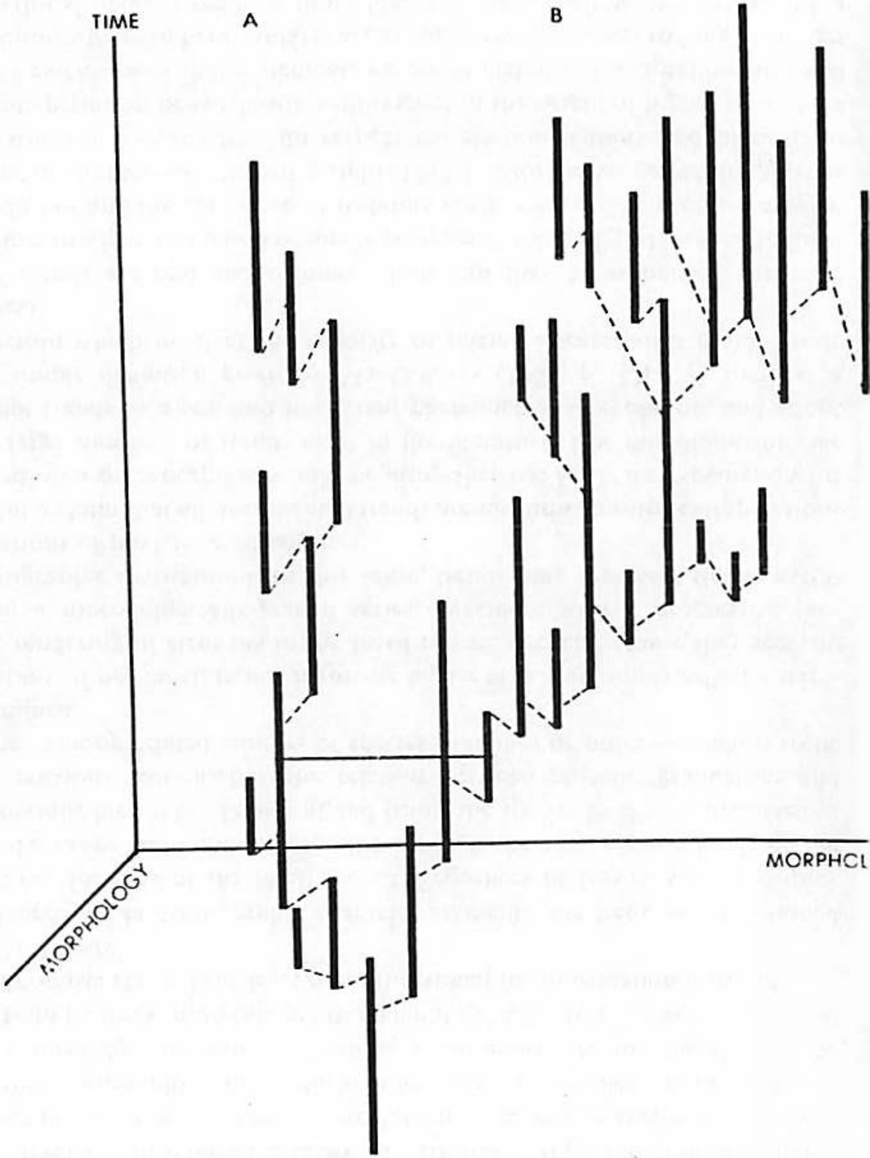


Figure 5-10: Three-dimensional sketch contrasting a pattern of relati stability (A) with a trend (B), where speciation (dashed lines) is occurrin in both major lineages. Morphological change is depicted here along t horizontal axes, while the vertical axis is time. Though a retrospecti pattern of directional selection might be fitted as a straight line in (B), t actual pattern is stasis within species, and differential success of speci exhibiting morphological change in a particular direction. For furth explanation, see text.

produces a paradox: why, then, are species coherent (or even recognizable)? Why do groups of (relatively independent) local populations continue to display a fairly consistent phenotype that permits their recognition as a species? Why does reproductive isolation not arise in every local population? Why is the local population itself not considered the "real" unit in evolution (as some would prefer—Sokal and Crovello, 1970, p. 151, for example)?

The answer probably lies in a view of species and individuals as homeostatic systems—as amazingly well-buffered to resist change and maintain stability in the face of disturbing influences. This concept has been urged particularly by Lerner (1954) and Mayr (1963), though the latter still gives more weight to gene flow than many will allow. Lerner (1954, p. 6) recognizes two types of homeostasis, mediated in both cases, he believes, by the generally higher fitness of heterozygous vs. homozygous genotypes: (1) ontogenetic self-regulation (developmental homeostasis) "based on the greater ability of the heterozygote to stay within the norms of canalized development" and (2) self-regulation of populations (genetic homeostasis) "based on natural selection favoring intermediate rather than extreme phenotypes." In this view, the importance of peripheral isolates lies in their small size and the alien environment beyond the species border that they inhabit—for only here are selective pressures strong enough and the inertia of large numbers sufficiently reduced to produce the "genetic revolution" (Mayr, 1963, p. 533) that overcomes homeostasis. The coherence of a species, therefore, is not maintained by interaction among its members (gene flow). It emerges, rather, as an historical consequence of the species' origin as a peripherally isolated population that acquired its own powerful homeostatic system. (We regard this idea as a serious challenge to the conventional view of species' reality that depends upon the organization of species as ecological units of *interacting* individuals in nature. If groups of nearly-independent local populations are recognized as species only because they share a set of homeostatic mechanisms developed long ago in a peripheral isolate that was "real" in our conventional sense of interaction, then some persistent anomalies are resolved. The arrangement of many asexual groups into good phenetic "species," quite inexplicable if interaction is the basis for coherence, receives a comfortable explanation under notions of homeostasis.)

Thus, the challenge to gene flow that seemed to question the stability of species in time ends by reinforcing that stability even more strongly. If we view a species as a set of subpopulations, all ready and able to differentiate but held in check only by the rein of gene flow, then the stability of species is a tenuous thing indeed. But if that stability is an inherent property both of individual development and the genetic structure of populations, then its power is immeasurably enhanced, for the basic property of homeostatic systems, or steady states, is that they resist change by self-regulation. That local populations do not differentiate into species, even though no external bar prevents it, stands as strong testimony to the inherent stability of species in time.

Paleontologists should recognize that much of their thought is conditioned by a peculiar perspective that they must bring to the study of life: they must look down from its present complexity and diversity into the past: their view must be retrospective. From this vantage point, it is very difficult to view evolution as anything but an easy and inevitable result of mere existence, as something that unfolds in a natural and orderly fashion. Yet we urge a different view. The norm for a species or, by extension, a community is stability. Speciation is a rare and difficult event that punctuates a system in homeostatic equilibrium. That so uncommon an event should have produced such a wondrous array of living and fossil forms can only give strength to an old idea: paleontology deals with a phenomenon that belongs to it alone among the evolutionary sciences and that enlightens all its conclusions—time.

Science - 28 August, 1981 - v. 213, no. 4511

Stebbins, G. and Ayala, F.
"Is a new evolutionary synthesis necessary?"
Science
Published by the A.A.S
Used by permission

Is a New Evolutionary Synthesis Necessary?

G. Ledyard Stebbins and Francisco J. Ayala

Summary. The current (synthetic) theory of evolution has been criticized on the grounds that it implies that macroevolutionary processes (speciation and morphological diversification) are gradual. The extent to which macroevolution is gradual or punctuational remains to be ascertained. Macroevolutionary processes are underlain by microevolutionary phenomena and are compatible with the synthetic theory of evolution. But microevolutionary principles are compatible with both gradualism and punctualism; therefore, logically they entail neither. Thus, macroevolution and microevolution are decoupled in the important sense that macroevolutionary patterns cannot be deduced from microevolutionary principles.

The current theory of evolution, known as the "modern synthesis" (*1*), has been challenged by some scientists. Gould, for example, has written that "The modern synthesis, as an exclusive proposition, has broken down on *both of its fundamental claims*: extrapolationism (gradual allelic substitution as a model for all evolutionary change) and nearly exclusive reliance on selection leading to adaptation" (*2*; emphasis added). Gould goes on to voice the need for a "new and general evolutionary theory [that] will embody [the] notion of hierarchy and stress a variety of themes either ignored or explicitly rejected by the modern synthesis." Similar statements have been made by a few others (*3*, *4*).

Many evolutionists would be surprised to see identified as the two "fundamental claims" of the modern synthesis those listed by Gould and most would not agree that the modern synthesis has "broken down." The impression that a "straw man" has been erected is confirmed when one discovers that the proposed new "themes" (*2–4*) are part and parcel of the modern synthesis (*1*, *5–12*). However, the critics' appeal to the pluralistic structure of evolutionary theory, to the hierarchical nature of evolutionary processes, and to the distinctive contributions made by the study of macroevolutionary phenomena deserve attention.

Mutation and Selection

Genetic changes underlie the evolution of organisms: mutations are the ultimate source of the genetic variation that makes possible the evolutionary process. "Genetic mutations are changes in the hereditary materials. . . . They can be classified in one of two major categories: *gene* (or point) *mutations*, which affect only one or a few nucleotides within a gene; and *chromosomal mutations* (or aberrations), which affect the number of chromosomes, or the number or the arrangement of genes in a chromosome" (*1*, p. 57). Gene mutations are the source of allelic variation; chromosomal mutations make possible the evolution of the amount and organization of the hereditary material (DNA). It is misleading to criticize the modern synthesis on the alleged grounds of its exclusive reliance on "point mutations (micromutations)" (*2*). Chromosomal mutations (traditionally known as chromosomal abnormalities or aberrations) played an essential role in the development of the modern synthesis and remain one of its pivots (*6*, *9–10*).

The frequencies of genes and gene arrangements change through the generations (evolve) owing to four processes: mutation, migration (gene flow), random drift, and natural selection. These four elementary processes of genetic change were already characterized by the early theorists who established the mathematical foundations of the modern synthesis (*5*, *12*, *13*). The highly organized character of organisms and their obvious adaptations are largely the result of natural selection operating under a variety of constraints, but responding to the demands of the environment. The constraints include the existing structure of organisms (and, hence, past history), the genetic variation available, as well as the particular circumstances of the physical and biotic components of the environment.

Without natural selection, populations of organisms would disintegrate over the generations because mutation and drift are random with respect to adaptation. But it is not correct that, for the synthetic theory, "All genetic change is adaptive"; or that "genetic drift certainly occurs—but only in populations so small and so near the brink that their rapid extinction will almost certainly ensue" (*2*, p. 120; *3*, pp. 20, 23–25). Controversy concerning the relative importance of random drift vis-à-vis natural selection has existed from the very beginning among the proponents of the modern synthesis. Fisher (*12*), for example, minimized the importance of random drift, but this is an important and decisive process in Wright's "shifting balance" version of the modern synthesis (*5*). Some evolutionists have relegated the importance of random drift to restricted—but by no means trivial—circumstances, such as "founder effects," which occur when a population is derived from only a few colonizers (*7*).

During the last decade no other issue has been more actively debated among evolutionists than the role of random drift. Molecular studies have shown that protein polymorphisms are pervasive in natural populations and that protein changes accompany the evolution of species (*14*). The neutrality theory of protein evolution proposes that evolution at the molecular level is largely due to random drift rather than being impelled by natural selection (*15*). But many evolutionists maintain that natural selection plays an essential role even at the molecular level (*14*). The "selectionist" and "neutralist" views of molecular evolution are competing hypotheses within the framework of the synthetic theory of evolution (*16*).

The Origin of Species

Living beings do not represent a continuum of all possible gene combinations generated at random, but are rather naturally grouped into species: arrays of populations between which intermediates are rare or absent. The distinctness of species is preserved by reproductive isolating mechanisms, that is, by biologically determined impediments to gene exchange, such as ethological or ecological

---

G. L. Stebbins is professor emeritus and F. J. Ayala is professor in the Department of Genetics, University of California, Davis 95616.

0036-8075/81/0828-0967$01.00/0

differentiation and hybrid sterility. In sexually reproducing organisms, species may be defined as "groups of interbreeding natural populations that are reproductively isolated from other such groups" (*6, 7*).

Interest in the process of speciation has recently burgeoned among paleontologists who sponsor the punctuated equilibria model (*3, 4, 17*), which is contrasted with the gradualistic model of macroevolution. Proponents of the punctualist model argue that, according to paleontological evidence, "species have tended to last for such long intervals of geological time that, once formed, they must have evolved very slowly. . . . This condition, when compared to the rapid pace of large-scale evolution, implies that most sizable evolutionary steps in the history of life must have occurred cryptically from a paleontological vantage point, during the rapid origination of certain species from small, localized populations of pre-existing species" (*3*, p. 3).

Whether macroevolution occurs according to the punctualist or the gradual model is something to be decided empirically. Certainly both modes have occurred in evolution, and the question then is their relative importance and the identification of factors that determine one or the other mode. Our primary concern here is, however, not this issue but rather whether any inconsistencies exist between the punctualist mode of evolutionary change and the synthetic theory's understanding of the speciation process.

We note, first, that the alleged relevance of punctuational evolution to speciation is based, at least in part, on two misunderstandings. The first one is a definitional artifact: paleontologists recognize species by their different morphologies as preserved in the fossil record (*18*). Thus, speciation events yielding little or no morphologically different products go totally unrecognized. Sibling (that is, morphologically indistinguishable) species are common in many groups of insects, in rodents, and in other well-studied organisms (*6, 7*). Speciation as seen by the paleontologist always involves substantial morphological change *because only when such change has occurred is the paleontologist able to recognize the presence of a new species.* The second misunderstanding concerns the time scale. When punctualists argue that paleontological evidence indicates that speciation is a rapid process (*3, 4, 17*), they are using a geological time scale. Instantaneous events in the paleontological scale, as in the transition between different geological strata, may involve thousands, at times many thousands, of years. In the microevolutionary scale of the population biologist, a thousand years is a long time, not an instant.

A more fundamental point is that rapid speciation, even in the microevolutionary scale, is not inconsistent with, and has been postulated by, the synthetic theory. Lewis's model of "saltational" speciation (*19*) and White's "stasipatric" speciation (*20*) are speciation models proposing that new species can arise in a few generations, as a result of the reproductive isolation produced by translocations and other chromosomal mutations. Polyploidy is the limiting case of rapid speciation—requiring only one or two generations—through chromosomal change (*6, 10*). Rapid speciation without chromosomal reorganization has not only been postulated by the proponents of the modern synthesis (*6, 7, 11*) but has been studied experimentally. A notable example—an incipient neospecies that arose in a *Drosophila paulistorum* culture, sometime between 1958 and 1963—was the subject of investigation by Dobzhansky for many years (*21*). In brief, the synthetic theory recognizes that there are a number of modes of speciation and that some of them, variously grouped under such terms as "saltational," "rapid," or "quantum" modes, require only a few generations and are effectively instantaneous in the geological time scale (*22*).

According to Gould, "The most exciting entry among punctuational models for speciation in ecological time is the emphasis, now coming from several quarters, on chromosomal alterations as isolating mechanisms" (*2*, p. 123). The role of chromosomal mutations in speciation is, like other important empirical questions, a subject of continued investigation and increased understanding. But the works to which Gould refers represent only the most recent accomplishments of a continuum that extends back to the 1930's (*6, 9, 10, 23*). Summarizing in 1950 the earlier work, Stebbins (*10*) concluded that (i) the most effective chromosome barriers of reproductive isolation come from the accumulation of small chromosomal changes; (ii) these changes may accumulate in a short time, such as 50 to 100 generations, to the point of resulting in reproductive isolation; (iii) these small changes occur largely independently of changes in the genes affecting external morphology; and, therefore, (iv) morphologically undifferentiated species may exhibit substantial chromosomal differences.

From Microevolution to Macroevolution

We come now to what has been called "the central question" posed by the proponents of punctualism, namely, "whether the mechanisms underlying microevolution can be extrapolated to explain macroevolution" (*24*). The argument has been succinctly expressed as follows: "if species originate in geological instants and then do not alter in major ways, then evolutionary trends cannot represent a simple extrapolation of allelic substitution within a population" (*2*, p. 125).

The question raised is the general issue of reduction as it applies to the different levels of the evolutionary process. Evolutionary trends are high-level phenomena predicated from events that encompass different species, as well as genera and higher taxa, and that extend over long periods of time. Microevolutionary studies are, on the contrary, concerned with evolutionary changes in populations that occur within "instants" of geological time. But, as so often happens with questions of reductionism, the issue of "whether the mechanisms underlying microevolution can be extrapolated" to macroevolution involves separate issues that must be distinguished in order to arrive at a satisfactory resolution.

Three separate questions, at least, are involved: (i) whether microevolutionary processes *operate* (and have operated in the past) throughout the different taxa in which macroevolutionary phenomena are observed; (ii) whether the microevolutionary processes identified by population geneticists (mutation, chromosomal change, random drift, natural selection) can account for the morphological changes and other macroevolutionary phenomena observed in higher taxa or, rather, whether additional kinds of genetic processes need to be postulated; and (iii) whether evolutionary trends and other macroevolutionary patterns can be predicted from knowledge of microevolutionary processes.

These distinctions may perhaps become clearer if we state them as they might be formulated by a biologist concerned with the question whether the laws of physics and chemistry can be extrapolated to biology. The first question would be whether the laws of physics and chemistry apply to the atoms and molecules present in living organisms. The second question would be whether biological phenomena can be accounted for as the result of interactions between atoms and molecules according to the laws known to physics and chemistry or whether the workings of organisms require additional kinds of interactions between atoms and molecules. The third question would be whether living phenomena can be predicted from the laws of physics and chemistry.

As to the first question, it is unlikely that any paleontologist would claim that mutation, chromosome change, drift, natural selection, and other microevolutionary processes do not apply to each of the populations of the higher taxa that are considered in macroevolution. There is, of course, an added dimension—macroevolutionists are largely concerned with phenomena of the past. Direct observation of microevolutionary processes in populations of long-extinct organisms is not possible. But there is no reason to believe that the processes of mutation, random drift, and natural selection, or the nature of the interactions between organisms and the environment would have been different in nature for, say, Paleozoic brachiopods and ostracoderms than for modern molluscs and fishes. Extinct and living populations—like different living populations—may have experienced quantitative differences in the relative importance of one or another process, but the processes could hardly have been different in kind.

The Origin of Differences Between Higher Taxa

The second question raised above has more substantive implications than the first. Can the microevolutionary processes studied by population geneticists account for macroevolutionary phenomena or do we need to postulate new kinds of genetic processes? The large morphological (phenotypic) changes observed in evolutionary history, and the rapidity with which they appear in the geological record, is one major matter of concern. Another issue is stasis—the apparent persistence of species, with little or no morphological change, for hundreds of thousands or millions of years. The apparent dilemma is that microevolutionary processes apparently yield small but continuous changes, while macroevolution as seen by punctualists occurs by large and rapid bursts of change followed by long periods without change.

Forty years ago Goldschmidt argued that the incompatibility is real: "The decisive step in evolution, the first step towards macroevolution, the step from one species to another, requires another evolutionary method than that of sheer accumulation of micromutations" (*25*). The specific solution postulated by Goldschmidt, that is, the occurrence of systemic mutations, yielding hopeful monsters, can be excluded in view of current genetic knowledge, but the issue raised by him deserves attention.

Single-gene or chromosome mutations may have large effects on the genotype because they act early in the embryo and their effects become magnified through development. Single-gene "macromutations" have been carefully analyzed, for example, in *Drosophila melanogaster*—mutations such as "bithorax" and the homeotic mutants that transform one body structure, for example, antennae, into another, such as legs. These large-effect mutations are not incompatible with the synthetic theory. Whether the kinds of morphological differences that characterize different taxa are due to such "macromutations" or to the accumulation of several mutations with small effect has been examined particularly in plants where fertile interspecific, and even intergeneric, hybrids can be obtained. The results of numerous studies do not support the hypothesis that the establishment of macromutations is necessary for divergence at the macroevolutionary level (*10, 23*). In animals, even a familial character, the presence of three ocelli in drosophilids, can be changed by artificial selection, demonstrating that a family-distinctive trait can be produced by the accumulation of small mutations present in natural populations (*26*). Moreover, Lande has convincingly shown that major morphological changes, such as in the number of digits or limbs, can occur in a geologically rapid fashion through the accumulation of mutations each with a small effect (*27*). In general, the evidence from plants as well as from animals supports Fisher's (*12*) theoretical argument that the probability of incorporation of a mutation in a population is inversely proportional to the magnitude of the mutation's effect on the phenotype.

Nevertheless, rapid phenotypic evolution may be caused by relatively slight genetic changes that affect critical stages of development. Alberch (*28*) has described differences in the number and position of tarsal bones in salamanders of the genus *Plethodon*. It is not known at present whether only one mutation, or several with additive effects, is involved. But the important point is that only a few of the possible genetic changes can, in interaction with the rest of the genome, yield a functional phenotype; and, therefore, the organ can change in only one or very few directions. Phenotypic evolution is thus directed along certain channels that may be followed by separate lineages. To what extent canalization of development restricts the possible directions of morphological evolution is a question as yet unsolved.

How often mutations with large phenotypic effects are involved in the origin of new taxa is also an unsolved question. The punctualists' thesis that such mutations may have been largely responsible for macroevolutionary change is based on the rapidity with which morphological discontinuities appear in the fossil record (*2, 3*). But the alleged evidence they present does not necessarily support the proposition. Microevolutionists and macroevolutionists use different time scales. The "geological instants" during which speciation and morphological shifts occur may involve intervals of the order of 50,000 years. There is little doubt that the gradual accumulation of small-effect mutations may yield sizable morphological changes during periods of that length. Anderson's study of body size in *Drosophila pseudoobscura* may serve as an example (*29*). Large populations, derived from a single set of parents, were set up at different temperatures and allowed to evolve on their own. A gradual, genetically determined, change in body size ensued, with flies kept at lower temperature becoming larger than those kept at higher temperatures. After 12 years, the mean size of the flies from the population kept at 16°C had become, when tested under standard conditions, approximately 10 percent greater than the size of the flies from the populations at 27°C; the change of mean value being greater than the standard deviation in size at the time when the tests were made. Assuming ten generations per year, the populations diverged at an average rate of 8×10^{-4} of the mean value per generation.

Paleontologists have emphasized the "extraordinary high net rate of evolution that is the hallmark of human phylogeny" (*3*). Interpreted in terms of the punctualist hypothesis, human phylogeny would have occurred as a succession of jumps, or geologically instantaneous saltations, interspersed by long periods without morphological change. Could these bursts of phenotypic evolution be due to the gradual accumulation of small changes? Consider cranial capacity, the character undergoing the greatest relative amount of change. The fastest rate of net change occurred between 500,000 years ago, when our ancestors were represented by *Homo erectus*, and 75,000 years ago, when Neanderthal man had acquired a cranial capacity similar to that of modern humans. In the intervening 425,000 years, cranial capacity evolved from about 900 cubic centimeters in Peking man to about 1400 cubic centimeters in Neanderthal people. Let us assume that the increase in brain size occurred in a single burst at the rate observed in *Drosophila pseudoobscura* of 8×10^{-4} of the mean value per generation. The change from 900 to 1400 cubic centimeters could have taken place in 540 generations or, if we assume 25 years per generation, in 13,500 years. Thirteen thousand years are, of course, a geological instant. Yet, this evolutionary "burst" could have taken place by gradual accumulation of small-effect mutations at rates compatible with those observed in microevolutionary studies (*30*).

We now raise the question of "stasis," the long-term persistence of species without morphological change. According to the model of punctuated equilibria, most phenotypic macroevolutionary change occurs in rapid bursts followed by long periods of stasis, during which little if any morphological change takes place. Phenotypic stability is compatible with microevolutionary processes; it ensues from stabilizing selection (*1, 6, 8*). Stebbins (*10*) in 1950 pointed out the morphological similarity, in forest trees and some herbs, between populations that have been separated from each other for millions of years. According to Dobzhansky (*31*), a successful morphology may persist unchanged for extremely long periods of time, even through speciation events. Some sibling species in *Drosophila* diverged from each other millions of years ago, yet their morphologies have remained identical to each other and to their ancestral species (*32*). Dobzhansky postulated that evolution in such cases continues, however, at the physiological or biochemical level; a prediction confirmed by recent molecular studies (*33*, p. 587).

Whether the phenomenon of paleontological stasis is as common as claimed by the punctualists needs to be carefully examined (*34*). As indicated by Levinton and Simon (*18*), paleontological taxonomy at the species level "requires the identification of species-specific characters which are invariant with time." Stasis may often be only apparent, as another artifact of the definition of species used.

Reduction, Hierarchy, and Macroevolution

We have just argued that the macroevolutionary patterns proposed by the model of punctuated equilibria—short periods of rapid phenotypic change followed by long spans of morphological stasis—are compatible with the theory of population genetics concerning microevolutionary processes. But does the theory predict that macroevolution will be punctuational? This is the third question formulated above, and the answer can only be no. The theory of population genetics is compatible with both punctualism and gradualism. Logically, therefore, it does not entail either. Whether macroevolution occurs predominantly according to the model of punctuated equilibria or to the model of phyletic gradualism is a question to be decided by studying macroevolutionary patterns, not by inference from our knowledge of microevolutionary processes (*35*).

Levinton and Simon (*18*) have written that "the implications of [the species-selection model proposed by the punctualists] should be of immediate concern to population biologists studying microevolutionary phenomena because it claims to negate the importance of population level phenomena in long term evolution," and they have gone on to "question the [punctualists'] belief that microevolution is decoupled from macroevolution." Statements of this kind need to be clarified. We have established above that at least three different issues are at stake, and have stated our solution to the first two issues. We may reiterate our points paraphrasing the terminology used by Levinton and Simon. Population level phenomena *are* important to long-term evolution because the populations in which macroevolutionary patterns are observed are the same populations that evolve at the microevolutionary level. Moreover, the study of microevolutionary phenomena *is* important to macroevolution because any theory of macroevolution that is correct must be compatible with well-established microevolutionary principles and theories; and indeed we have argued that the model of punctuated equilibria is compatible with the theory of population genetics. In these two senses—identity at the level of events and compatibility of theories—macroevolution cannot be decoupled from microevolution. But there is one sense (which epistemologically is most important) in which macroevolution and microevolution are decoupled, namely, in the sense that macroevolution is an autonomous field of study that must develop and test its own theories. In other words, macroevolutionary theories are not reducible (at least at the present state of knowledge and probably in principle) to microevolutionary theories.

Gould (*2*, p. 121) has pointed out that the study of evolution embodies a concept of hierarchy—a world constructed not as a smooth and seamless continuum, permitting simple extrapolation from the lowest level to the highest, but as a series of ascending levels, each bound to the one below it in some ways and independent in others . . . 'emergent' features not implicit in the operation of processes at lower levels, may control events at higher levels."

The world of life is hierarchically structured. There is a hierarchy of levels: from atoms, through molecules, organelles, cells, tissues, organs, multicellular individuals and populations, to communities. Time adds another dimension of the hierarchy, with the interesting consequence that transitions from one level to another occur such that as time proceeds the descendants of a single species may include several species, genera, families, and so forth. Hierarchical organization often is such that the phenomena at a given level cannot be inferred from knowledge of the phenomena at a lower level of the hierarchy. Statements about "emergent" features imply this inability to predict from one level of organization to another. Consider, for example, the question whether water has emergent properties relative to its components, hydrogen and oxygen. One could argue that among the properties of hydrogen and oxygen one must include their ability to combine according to the formula H_2O and to exhibit the properties attributed to water. Proceeding accordingly, one could claim that the properties of oxygen and hydrogen include those of hemoglobin and other proteins as well as human speech and abstract thought, because oxygen and hydrogen have these properties when combined with other given atoms in certain ways. But this is a definitional maneuver that contributes little to the understanding of the relationships between complex systems and their constituent parts.

The consideration that is important is whether the properties of a complex object or system can be inferred from the study of component parts in isolation. It is for this reason that we do not usually include among the properties of hydrogen those of water, ethyl alcohol, proteins, or human beings (*36*).

The question of whether macroevolution is an autonomous field of knowledge is more appropriately posed in terms of the relationships between macroevolutionary and microevolutionary theories, rather than in terms of hierarchy of levels or emergent properties (*37*). The study of macroevolution is autonomous with respect to microevolutionary studies if the theories, hypotheses, and models of macroevolution cannot be "reduced" to the microevolutionary theories, hypotheses, or models. Two conditions are jointly necessary and sufficient for the reduction of one branch of science to another: derivability and connectability (*38*). The condition of derivability requires that the laws and theories of the branch of science to be reduced be derived as logical consequences from the laws and theories of some other branch of science. The condition of connectability requires that the distinctive terms of the secondary branch of science be redefined in the language of the branch of science to which it is reduced—this redefinition of terms is, of course, necessary in order to analyze the logical connections between the theories of the two branches of science.

Microevolutionary processes, as now known, are compatible with the two models of macroevolution—punctualism and gradualism. From microevolutionary knowledge, we cannot infer which one of those two macroevolutionary patterns prevails. Hence, the condition of derivability is not satisfied. Needless to say, the conflict between punctualism and gradualism is not the only macroevolutionary issue that cannot be decided by logical inference from microevolutionary principles. Consider, for example, the question of rates of morphological evolution. Three groups of crossopterygian fishes flourished during the Devonian. The lungfishes (Dipnoi) changed little for hundreds of millions of years and remain as relics. The coelacanths became highly successful in the open ocean until the Cretaceous, then declined and stagnated, leaving only the relictual *Latimeria*. The rhipidistians, in contrast, evolved into the amphibians, reptiles, and, finally, birds and mammals (*39*). Models to explain divergent rates of morphological evolution must incorporate factors other than microevolutionary principles, including rates of speciation and the environmental and biotic conditions that may account for successions of morphological change in some but not other lineages.

Distinctive macroevolutionary theories and models have been advanced concerning such issues as rates of morphological evolution, patterns of species extinctions, and historical factors regulating taxonomic diversity. As long as these theories and models are compatible with the theories and laws of population biology, the decision as to which one among alternative hypotheses is correct cannot be reached by recourse to microevolutionary principles. Such a decision must rather be based on appropriate tests with the use of macroevolutionary evidence (*3, 4, 18*). Thus, macroevolution is an autonomous field of evolutionary study and, in this epistemologically very important sense, macroevolution is decoupled from microevolution (*40*).

References and Notes

1. A recent summary of the theory of evolution can be found in Th. Dobzhansky *et al.*, *Evolution* (Freeman, San Francisco, 1977).
2. S. J. Gould, *Paleobiology* 6, 119 (1980).
3. S. M. Stanley, *Macroevolution, Pattern and Process* (Freeman, San Francisco, 1979).
4. E. S. Vrba, *S. Afr. J. Sci.* 76, 61 (1980).
5. S. Wright, *Genetics* 16, 97 (1931); *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1968–1978), vols. 1 to 4.
6. Th. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937); *Genetics of the Evolutionary Process* (Columbia Univ. Press, New York, 1970).
7. E. Mayr, *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942); *Animal Species and Evolution* (Harvard Univ. Press, Cambridge, Mass., 1963).
8. G. G. Simpson, *Tempo and Mode in Evolution* (Columbia Univ. Press, New York, 1944).
9. M. J. D. White, *Animal Cytology and Evolution* (Cambridge Univ. Press, London, 1945); *Modes of Speciation* (Freeman, San Francisco, 1978).
10. G. L. Stebbins, *Variation and Evolution in Plants* (Columbia Univ. Press, New York, 1950).
11. H. L. Carson, *Am. Nat.* 109, 83 (1975); G. L. Bush, *Annu. Rev. Ecol. Syst.* 6, 339 (1975).
12. R. A. Fisher, *The Genetic Theory of Natural Selection* (Clarendon, Oxford, 1930).
13. J. B. S. Haldane, *The Causes of Evolution* (Harper, New York, 1932).
14. F. J. Ayala, Ed., *Molecular Evolution* (Sinauer, Sunderland, Mass., 1976).
15. M. Kimura, *Nature (London)* 217, 624 (1968); J. L. King and T. H. Jukes, *Science* 164, 788 (1969); M. Kimura and T. Ohta, *Nature (London)* 229, 467 (1971).
16. F. J. Ayala, in *Molecular Evolution and Polymorphism*, M. Kimura, Ed., (National Institute of Genetics, Mishima, Japan, 1977), p. 73.
17. N. Eldredge, *Evolution* 25, 156 (1971); S. J. Gould and N. Eldredge, *Paleobiology* 3, 115 (1977).
18. J. S. Levinton and C. M. Simon, *Syst. Zool.* 29, 130 (1980).
19. H. Lewis, *Science* 152, 167 (1966).
20. M. J. D. White, *ibid.* 159, 1065 (1968).
21. Th. Dobzhansky and O. Pavlovsky, *Genetics* 55, 141 (1967); Th. Dobzhansky, *Science* 177, 664 (1972).
22. We have characterized quantum speciation as "rapid, requiring only a few generations. The ancestor of new species . . . may consist of only one or a few individuals. . . . Quantum speciation usually and perhaps always includes one or more stochastic or chance events" (*1*, pp. 198–199).
23. J. Clausen, *Stages in the Evolution of Plant Species* (Cornell Univ. Press, Syracuse, N.Y., 1951); V. Grant, *Plant Speciation* (Columbia Univ. Press, New York, 1971).
24. R. Lewin, *Science* 210, 883 (1980).
25. R. Goldschmidt, *The Material Basis of Evolution* (Yale Univ. Press, New Haven, Conn., 1940), p. 183.
26. J. M. Smith and K. C. Sondhi, *Genetics* 45, 1039 (1960).
27. R. Lande, *Evolution* 30, 314 (1976); *ibid.* 32, 73 (1978).
28. P. Alberch, *Am. Zool.* 20, 653 (1980).
29. W. W. Anderson, *Evolution* 27, 278 (1973).
30. The mean cranial capacity of modern man is 1342 cm^3, with the standard deviation, based on several thousand individuals, being between 169 and 200 cm^3 [P. V. Tobias, *The Brain in Hominid Evolution* (Columbia Univ. Press, New York, 1971) p. 43]. Thus the standard deviation for cranial capacity in modern humans is greater than 10 percent of the mean—giving a coefficient of variation greater than that observed for body size in *Drosophila pseudoobscura* [see (*29*)].
31. Th. Dobzhansky, *Am. Nat.* 90, 337 (1956).
32. The common occurrence of sibling species provides little support to the punctualists' postulated association between speciation events and morphological evolution.
33. F. J. Ayala, M. L. Tracey, D. Hedgecock, R. C. Richmond, *Evolution* 28, 576 (1974).
34. T. J. M. Schopf, *Paleobiology* 6, 380 (1980).
35. T. Hallam, *ibid.* 4, 16 (1978); R. Lande, *ibid.* 6, 233 (1980). We do not imply that all models of macroevolution are compatible with the theory of population genetics, but simply that the two models under consideration are, in the present state of knowledge, compatible with the theory. Some models of macroevolution, such as orthogenesis, are not compatible with our knowledge of microevolution. The conceptual field of macroevolutionary models compatible with population genetics is less than the universe of possible models, but is not reduced to only one such model.
36. Notice that, under the proposal herein made, whether or not a property is considered emergent depends on the state of knowledge; what appeared at one point as an emergent property might later be predictable from knowledge of the component parts. Notice also that the question of emergent properties is not limited to the world of life.
37. The reason is that the hierarchical differentiation of the subject matter is neither necessary nor sufficient for the autonomy of scientific disciplines. It is not necessary, because entities at a single hierarchical level can be the subject of diversified disciplines: cells are appropriate subject of study for cytology, genetics, immunology, and the like. In fact, identical events can be the subject of concern of different disciplines. The writing of this note can be studied by a physiologist interested in the muscular and nerve processes involved, by a philosopher interested in the epistemological question raised, by a psychologist concerned with thought processes, and so on. Moreover, hierarchical differentiation of subject matter is not a sufficient condition for the autonomy of the concerned disciplines; relativity theory applies all the way from subatomic particles to planetary motions, and genetic laws apply to multicellular organisms as well as to cellular and even subcellular entities.
38. E. Nagel, *The Structure of Science* (Harcourt, Brace, & World, New York, 1961); F. J. Ayala, *Am. Sci.* 56, 207 (1968).
39. E. H. Colbert, *Evolution of the Vertebrates* (Wiley-Interscience, New York, ed. 3, 1980).
40. This does not imply, however, that macroevolutionary studies cannot be incorporated into the synthetic theory of evolution. Quite to the contrary, the modern theory of evolution is called "synthetic" because it incorporates knowledge from diverse autonomous disciplines, such as genetics, ecology, systematics, and paleontology.

41. Supported by NIH grant 1-P01-GM-2221.

Gould, S.J.
"Darwinism and the expansion
of evolutionary theory"
Science
Published by the A.A.A.S

Used by permission

Darwinism and the Expansion of Evolutionary Theory

Stephen Jay Gould

Ben Sira, author of the apocryphal book of Ecclesiasticus, paid homage to the heroes of Israel in a noted passage beginning, "let us now praise famous men." He glorified great teachers above all others, for their fame shall eclipse the immediate triumphs of kings and conquerors. And he argued that the corporeal death of teachers counts for nothing—indeed, it should be celebrated—since great ideas must live forever: "His name will be more glorious than a thousand others, and if he dies, that will satisfy him just as well." These sentiments express the compulsion we feel to commemorate the deaths of great thinkers; for their ideas still direct us today. Charles Darwin died 100 years ago, on 19 April 1882, but his name still causes fundamentalists to shudder and scientists to draw battle lines amidst their accolades.

What Is Darwinism?

Darwin often stated that his biological work had embodied two different goals (*1*): to establish the fact of evolution, and to propose natural selection as its primary mechanism. "I had," he wrote, "two distinct objects in view; firstly to show that species had not been separately created, and secondly, that natural selection had been the chief agent of change" (*2*).

Although "Darwinism" has often been equated with evolution itself in popular literature, the term should be restricted to the body of thought allied with Darwin's own theory of mechanism, his second goal. This decision does not provide an unambiguous definition, if only because Darwin himself was a pluralist who granted pride of place to natural selection, but also advocated an important role for Lamarckian and other nonselectionist factors. Thus, as the 19th century drew to a close, G. J. Romanes and A. Weismann squared off in a terminological battle for rights to the name "Darwinian"—Romanes claiming it for his eclectic pluralism, Weismann for his strict selectionism (*3*).

If we agree, as our century generally has, that "Darwinism" should be restricted to the world view encompassed by the theory of natural selection itself, the problem of definition is still not easily resolved. Darwinism must be more than the bare bones of the mechanics: the principles of superfecundity and inherited variation, and the deduction of natural selection therefrom. It must, fundamentally, make a claim for wide scope and dominant frequency; natural selection must represent the primary directing force of evolutionary change.

I believe that Darwinism, under these guidelines, can best be defined as embodying two central claims and a variety of peripheral and supporting statements more or less strongly tied to the central postulates; Darwinism is not a mathematical formula or a set of statements, deductively arranged.

1) The creativity of natural selection. Darwinians cannot simply claim that natural selection operates since everyone,

---

The author is a professor of geology at the Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138.

0036-8075/82/0423-0380$01.00/0

including Paley and the natural theologians, advocated selection as a device for removing unfit individuals at both extremes and preserving, intact and forever, the created type (*4*). The essence of Darwinism lies in a claim that natural selection is the primary directing force of evolution, in that it creates fitter phenotypes by differentially preserving, generation by generation, the best adapted organisms from a pool of random variants (*5*) that supply raw material only, not direction itself. Natural selection is a creator; it builds adaptation step by step.

Darwin's contemporaries understood that natural selection hinged on the argument for creativity. Natural selection can only eliminate the unfit, his opponents proclaimed; something else must create the fit. Thus, the American Neo-Lamarckian E. D. Cope wrote a book with the sardonic title *The Origin of the Fittest* (*6*), and Charles Lyell complained to Darwin that he could understand how selection might operate like two members of the "Hindoo triad"—Vishnu the preserver and Siva the destroyer—but not like Brahma the creator (*7*).

The claim for creativity has important consequences and prerequisites that also become part of the Darwinian corpus. Most prominently, three constraints are imposed on the nature of genetic variation (or at least the evolutionarily significant portion of it). (i) It must be copious since selection makes nothing directly and requires a large pool of raw material. (ii) It must be small in scope. If new species characteristically arise all at once, then the fit are formed by the process of variation itself, and natural selection only plays the negative role of executioner for the unfit. True saltationist theories have always been considered anti-Darwinian on this basis. (iii) It must be undirected. If new environments can elicit heritable, adaptive variation, then creativity lies in the process of variation, and selection only eliminates the unfit. Lamarckism is an anti-Darwinian theory because it advocates directed variation; organisms perceive felt needs, adapt their bodies accordingly, and pass these modifications directly to offspring.

Two additional postulates, generally considered part and parcel of the Darwinian world view, are intimately related to the claim for creativity, but are not absolute prerequisites or necessary deductive consequences: (i) *Gradualism*. If creativity resides in a step-by-step process of selection from a pool of random variants, then evolutionary change must be dominantly continuous and descendants must be linked to ancestors by a long chain of smoothly intermediate phenotypes. Darwin's own gradualism precedes his belief in natural selection and has deeper roots (*8*); it dominated his world view and provided a central focus for most other theories that he proposed, including the origin of coral atolls by subsidence of central islands, and formation of vegetable mold by earthworms (*9, 10*). (ii) *The adaptationist program*. If selection becomes creative by superintending, generation by generation, the continuous incorporation of favorable variation into altered forms, then evolutionary change must be fundamentally adaptive. If evolution were saltational, or driven by internally generated biases in the direction of variation, adaptation would not be a necessary attribute of evolutionary change.

The argument for creativity rests on relative frequency, not exclusivity. Other factors must regulate some cases of evolutionary change—randomness as a direct source of modification, not only of raw material, for example. The Darwinian strategy does not deny other factors, but attempts to circumscribe their domain to few and unimportant cases.

2) Selection operates through the differential reproductive success of individual organisms (the "struggle for existence" in Darwin's terminology). Selection is an interaction among individuals; there are no higher-order laws in nature, no statements about the "good" of species or ecosystems. If species survive longer, or if ecosystems appear to display harmony and balance, these features arise as a by-product of selection among individuals for reproductive success.

Although evolutionists, including many who call themselves Darwinians, have often muddled this point (*11*), it is a central feature of Darwin's logic (*12*). It underlies all his colorful visual imagery including the metaphor of the wedge (*13*, p. 67), or the true struggle that underlies an appearance of harmony: "we behold the face of nature bright with gladness," but . . . (*13*, p. 62). Darwin developed his theory of natural selection by transferring the basic argument of Adam Smith's economics into nature (*14*): an ordered economy can best be achieved by letting individuals struggle for personal profits, thereby permitting a natural sifting of the most competitive (laissez-faire); an ordered ecology is a transient balance established by successful competitors pursuing their own Darwinian edge.

Summary. The essence of Darwinism lies in the claim that natural selection is a creative force, and in the reductionist assertion that selection upon individual organisms is the locus of evolutionary change. Critiques of adaptationism and gradualism call into doubt the traditional consequences of the argument for creativity, while a concept of hierarchy, with selection acting upon such higher-level "individuals" as demes and species, challenges the reductionist claim. An expanded hierarchical theory would not be Darwinism, as strictly defined, but it would capture, in abstract form, the fundamental feature of Darwin's vision—direction of evolution by selection at each level.

As a primary consequence, this focus upon individual organisms leads to reductionism, not to ultimate atoms and molecules of course, but of higher-order, or macroevolutionary, processes to the accumulated struggles of individuals. Extrapolationism is the other side of the same coin—the claim that natural selection within local populations is the source of all important evolutionary change.

Darwinism and the Modern Synthesis

Although Darwin succeeded in his first goal, and lies in Westminster Abbey for his success in establishing the fact of evolution, his theory of natural selection did not triumph as an orthodoxy until long after his death. The Mendelian component to the modern, or Neo-Darwinian, theory only developed in our century. Moreover, and ironically, the first Mendelians emphasized macromutations and were non-Darwinians on the issue of creativity as discussed above.

The Darwinian resurgence began in earnest in the 1930's, but did not crystallize until the 1950's. At the last Darwinian centennial, in 1959 (both the 100th anniversary of the *Origin of Species* and the 150th of Darwin's birth), celebrations throughout the world lauded the "modern synthesis" as Darwinism finally triumphant (*15*).

Julian Huxley, who coined the term (*16*), defined the "modern synthesis" as an integration of the disparate parts of biology about a Darwinian core (*17*).

Synthesis occurred at two levels: (i) The Mendelian research program merged with Darwinian traditions of natural history, as Mendelians recognized the importance of micromutations and their correspondence with Darwinian variation, and as population genetics supplied a quantitative mechanics for evolutionary change. (ii) The traditional disciplines of natural history, systematics, paleontology, morphology, and classical botany, for example (*18*), were integrated within the Darwinian core, or at least rendered consistent with it.

The initial works of the synthesis, particularly Dobzhansky's first (1937) edition of *Genetics and the Origin of Species*, were not firmly Darwinian (as defined above), and did not assert a dominant frequency for natural selection. They were more concerned with demonstrating that large-scale phenomena of evolution are consistent with the principles of genetics, whether Darwinian or not; and they therefore, for example, granted greater prominence to genetic drift than later editions of the same works would allow.

Throughout the late 1940's and 1950's, however, the synthesis hardened about its Darwinian core. Analysis of textbooks and, particularly, the comparison of first with later editions of the founding documents, demonstrates the emergence of natural selection and adaptation as preeminent factors of evolution. Thus, for example, G. G. Simpson redefined "quantum evolution" in 1953 as a limiting rate for adaptive phyletic transformation, not, as he had in 1944, as a higher-order analog of genetic drift, with a truly inadaptive phase between stabilized end points (*19*). Dobzhansky removed chapters and reduced emphasis upon rapid modification and random components to evolutionary change (*20*). David Lack reassessed his work on Darwin's finches and decided that minor differences among species are adaptive after all (*21*). His preface to the 1960 reissue of his monograph features the following statement (*22*):

> This text was completed in 1944 and . . . views on species-formation have advanced. In particular, it was generally believed when I wrote the book that, in animals, nearly all of the differences between subspecies of the same species, and between closely related species in the same genus, were without adaptive significance. . . . Sixteen years later, it is generally believed that all, or almost all, subspecific and specific differences are adaptive. . . . Hence it now seems probable that at least most of the seemingly nonadaptive differences in Darwin's finches would, if more were known, prove to be adaptive.

Mayr's definition of the synthesis, offered without rebuttal at a conference of historians and architects of the theory, reflects this crystallized version:

> The term "evolutionary synthesis" was introduced by Julian Huxley . . . to designate the general acceptance of two conclusions: gradual evolution can be explained in terms of small genetic changes ("mutations") and recombination, and the ordering of this genetic variation by natural selection; and the observed evolutionary phenomena, particularly macroevolutionary processes and speciation, can be explained in a manner that is consistent with the known genetic mechanisms (*23*).

This definition restates the two central claims of Darwinism discussed in the last section: Mayr's first conclusion, with its emphasis on gradualism, small genetic change, and natural selection, represents the argument for creativity; while the second embodies the claim for reduction. I have been challenged for erecting a straw man in citing this definition of the synthesis (*24*), but it was framed by a man who is both an architect and the leading historian of the theory, and it is surely an accurate statement of what I was taught as a graduate student in the mid-1960's. Moreover, these very words have been identified as the "broad version" of the synthesis (as opposed to a more partisan and restrictive stance) by White (*25*), a leading evolutionist and scholar who lived through it all.

The modern synthesis has sometimes been so broadly construed, usually by defenders who wish to see it as fully adequate to meet and encompass current critiques, that it loses all meaning by including everything. If, as Stebbins and Ayala claim, " 'selectionist' and 'neutralist' views of molecular evolution are competing hypotheses within the framework of the synthetic theory of evolution" (*26*), then what serious views are excluded? King and Jukes, authors of the neutralist theory, named it "non-Darwinian evolution" in the title of their famous paper (*27*). Stebbins and Ayala have tried to win an argument by redefinition. The essence of the modern synthesis must be its Darwinian core. If most evolutionary change is neutral, the synthesis is severely compromised.

What Is Happening to Darwinism

Current critics of Darwinism and the modern synthesis are proposing a good deal more than a comfortable extension of the theory, but much less than a revolution. In my partisan view, neither of Darwinism's two central themes will survive in their strict formulation; in that sense, "the modern synthesis, as an exclusive proposition, has broken down on both of its fundamental claims" (*28*). However, I believe that a restructured evolutionary theory will embody the essence of the Darwinian argument in a more abstract, and hierarchically extended form. The modern synthesis is incomplete, not incorrect.

Critique of Creativity: Gradualism

At issue is not the general idea that natural selection can act as a creative force; the basic argument, in principle, is a sound one. Primary doubts center on the subsidiary claims—gradualism and the adaptationist program. If most evolutionary changes, particularly large-scale trends, include major nonadaptive components as primary directing or channeling features, and if they proceed more in an episodic than a smoothly continuous fashion, then we inhabit a different world from the one Darwin envisaged.

Critiques of gradualist thought proceed on different levels and have different import, but none are fundamentally opposed to natural selection. They are therefore not directed against the heart of Darwinian theory, but against a fundamental subsidiary aspect of Darwin's own world view—one that he consistently conflated with natural selection, as in the following famous passage: "If it could be demonstrated that any complex organ existed, which could not possibly have been formed by numerous, successive, slight modifications, my theory would absolutely break down" (*29*).

At the levels of microevolution and speciation, the extreme saltationist claim that new species arise all at once, fully formed, by a fortunate macromutation would be anti-Darwinian, but no serious thinker now advances such a view, and neither did Richard Goldschmidt (*30*), the last major scholar to whom such an opinion is often attributed. Legitimate claims range from the saltational origin of key features by developmental shifts of dissociable segments of ontogeny (*31*) to the origin of reproductive isolation (speciation) by major and rapidly incorporated genetic changes that precede the acquisition of adaptive, phenotypic differences (*32*).

Are such styles of evolution anti-Darwinian? What can one say except "yes and no." They do not deny a creative role to natural selection, but neither do they embody the constant superintending of each event, or the step-by-step construction of each major feature, that traditional views about natural selection have advocated. If new *Baupläne* often arise in an adaptive cascade following the saltational origin of a key feature, then part of the process is sequential and adaptive, and therefore Darwinian; but the initial step is not, since selection does not play a creative role in building the key feature. If reproductive isolation often precedes adaptation, then a major aspect of speciation is Darwinian (for the new species will not prosper unless it builds distinctive adaptations in the sequential mode), but its initiation, including the defining feature of reproductive isolation, is not.

At the macroevolutionary level of trends, the theory of punctuated equilibrium (*33*) proposes that established species generally do not change substantially in phenotype over a lifetime that may encompass many million years (stasis), and that most evolutionary change is concentrated in geologically instantaneous events of branching speciation. These geological instants, resolvable (*34*) in favorable stratigraphic circumstances (so that the theory can be tested for its proposed punctuations as well as for its evident periods of stasis), represent amounts of microevolutionary time fully consistent with orthodox views about speciation. Indeed, Eldredge and I originally proposed punctuated equilibrium as the expected geological consequence of Mayr's theory of peripatric speciation. The non-Darwinian implications of punctuated equilibrium lie in its suggestions for the explanation of evolutionary trends (see below), not in the tempo of individual speciation events. Although punctuated equilibrium is a theory for a higher level of evolutionary change, and must therefore be agnostic with respect to the role of natural selection in speciation, the world that it proposes is quite different from that traditionally viewed by paleontologists (and by Darwin himself) as the proper geological extension of Darwinism.

The "gradualist-punctuationalist debate," the general label often applied to this disparate series of claims, may not be directed at the heart of natural selection, but it remains an important critique of the Darwinian tradition. The world is not inhabited exclusively by fools, and when a subject arouses intense interest and debate, as this one has, something other than semantics is usually at stake. In the largest sense, this debate is but one small aspect of a broader discussion about the nature of change: Is our world (to construct a ridiculously oversimplified dichotomy) primarily one of constant change (with structure as a mere incarnation of the moment), or is structure primary and constraining, with change as a "difficult" phenomenon, usually accomplished rapidly when a stable structure is stressed beyond its buffering capacity to resist and absorb. It would be hard to deny that the Darwinian tradition, including the modern synthesis, favored the first view while "punctuationalist" thought in general, including such aspects of classical morphology as D'Arcy Thompson's theory of form (*35*), prefers the second.

Critique of Creativity: Adaptation

The primary critiques of adaptation have arisen from molecular data, particularly from the approximately even ticking of the molecular clock, and the argument that natural populations generally maintain too much genetic variation to explain by natural selection, even when selection acts to preserve variation as in, for example, heterozygote advantage and frequency-dependent selection. To these phenomena, Darwinians have a response that is, in one sense, fully justified: Neutral genetic changes without phenotypic consequences are invisible to Darwinian processes of selection upon organisms and therefore represent a legitimate process separate from the subjects that Darwinism can treat. Still, since issues in natural history are generally resolved by appeals to relative frequency, the domain of Darwinism is restricted by these arguments.

But another general critique of the adaptationist program has been reasserted within the Darwinian domain of phenotypes (*36*). The theme is an old one, and not unfamiliar to Darwinians. Darwin himself took it seriously, as did the early, pluralistic accounts of the modern synthesis. The later, "hard" version of the synthesis relegated it to unimportance or lip service. The theme is two-pronged, both arguments asserting that the current utility of a structure permits no assumption that selection shaped it. First, the constraints of inherited form and developmental pathways may so channel any change that even though selection induces motion down permitted paths, the channel itself represents the primary determinant of evolutionary direction. Second, current utility permits no necessary conclusion about historical origin. Structures now indispensable for survival may have arisen for other reasons and been "coopted" by functional shift for their new role.

Both arguments have their Darwinian versions. First, if the channels are set by past adaptations, then selection remains preeminent, for all major structures are either expressions of immediate selection, or channeled by a phylogenetic heritage of previous selection. Darwin struggled mightily with this problem. Ultimately, in a neglected passage that I regard as one of the most crucial paragraphs in the *Origin of Species* (*37*), he resolved his doubts, and used this argument to uphold the great British tradition of adaptationism. Second, if coopted structures initially arose as adaptations for another function, then they too are products of selection, albeit in a regime not recorded by their current usage. We call this phenomenon preadaptation; as the primary solution to Mivart's taunt (*38*) about "the incipient stages of useful structures," it is a central theme of orthodox Darwinism.

But both arguments also have non-Darwinian versions, not widely appreciated but potentially fundamental. First, many features of organic architecture and developmental pathways have never been adaptations to anything, but arose as by-products or incidental consequences of changes with a basis in selection. Seilacher has suggested, for example, that the divaricate pattern of molluscan ornamentation may be nonadaptive in its essential design. In any case, it is certainly a channel for some fascinating subsidiary adaptations (*39*). Second, many structures available for cooptation did not arise as adaptations for something else (as the principle of preadaptation assumes) but were nonadaptive in their original construction. Evolutionary morphology now lacks a term for these coopted structures, and unnamed phenomena are not easily conceptualized. Vrba and I suggest that they be called exaptations (*40*), and present a range of potential examples from the genitalia of hyenas to redundant DNA.

Evolutionists admit, of course, that all selection yields by-products and incidental consequences, but we tend to think of these nonadaptations as a sort of evolutionary frill, a set of small and incidental modifications with no major consequences. I dispute this assessment and claim that the pool of nonadaptations must be far greater in extent than the direct adaptations that engender them. This pool must act as a higher-level analog of genetic variation, as a phenotypic source of raw material for further evolution. Nonadaptations are not just incidental allometric and pleiotropic effects on other parts of the body, but multifarious expressions potentially within any adapted structure. No one doubts, for example, that the human brain became large for a set of complex reasons related to selection. But, having reached its unprecedented bulk, it could, as a computer of some sophistication, perform in an unimagined range of ways bearing no relation to the selective reasons for initial enlargement. Most of human society may rest on these nonadaptive consequences. How many human institutions, for example, owe their shape to that most terrible datum that intelligence permitted us to grasp—the fact of our personal mortality.

I do not claim that a new force of evolutionary change has been discovered. Selection may supply all immediate direction, but if highly constraining channels are built of nonadaptations, and if evolutionary versatility resides primarily in the nature and extent of nonadaptive pools, then "internal" factors of organic design are an equal partner with selection. We say that mutation is the ultimate source of variation, yet we grant a fundamental role to recombination and the evolution of sexuality—often as a prerequisite to multicellularity, the Cambrian explosion and, ultimately, us. Likewise, selection may be the ultimate source of evolutionary change, but most actual events may owe more of their shape to its nonadaptive sequelae.

Is Evolution a Product of Selection Among Individuals?

Although arguments for a multiplicity of units of selection have been advanced and widely discussed (*41*), evolutionists have generally held fast to the overwhelming predominance, if not exclusivity, of organisms as the objects sorted by selection—Dawkins' (*42*) attempt at further reduction to the gene itself notwithstanding. How else can we explain the vehement reaction of many evolutionists to Wynne-Edwards' theory of group selection for the maintenance of altruistic traits (*43*), or the delight felt by so many when the same phenomena were explained, under the theory of kin selection, as a result of individuals pursuing their traditional Darwinian edge. I am not a supporter of Wynne-Edwards' particular hypothesis, nor do I doubt the validity and importance of kin selection; I merely point out that the vehemence and delight convey deeper messages about general attitudes.

Nonetheless, I believe that the traditional Darwinian focus on individual bodies, and the attendant reductionist account of macroevolution, will be supplanted by a hierarchical approach recognizing legitimate Darwinian individuals at several levels of a structural hierarchy, including genes, bodies, demes, species, and clades.

The argument may begin with a claim that first appears to be merely semantic, yet contains great utility and richness in implication, namely the conclusion advanced by Ghiselin and later supported by Hull that species should be treated as individuals, not as classes (*44*). Most species function as entities in nature, with coherence and stability. And they display the primary characteristics of a Darwinian actor; they vary within their population (clade in this case), and they exhibit differential rates of birth (speciation) and death (extinction).

Our language and culture include a prejudice for applying the concept of individual only to bodies, but any coherent entity that has a unique origin, sufficient temporal stability, and a capacity for reproduction with change can serve as an evolutionary agent. The actual hierarchy of our world is a contingent fact of history, not a heuristic device or a logical necessity. One can easily imagine a world devoid of such hierarchy, and conferring the status of evolutionary individual upon bodies alone. If genes could not duplicate themselves and disperse among chromosomes, we might lack the legitimately independent level that the "selfish DNA" hypothesis establishes for some genes (*45*). If new species usually arose by the smooth transformation of an entire ancestral species, and then changed continuously toward a descendant form, they would lack the stability and coherence required for defining evolutionary individuals. The theory of punctuated equilibrium allows us to individuate species in both time and space; this property (rather than the debate about evolutionary tempo) may emerge as its primary contribution to evolutionary theory.

In itself, individuation does not guarantee the strong claim for evolutionary agency: that the higher-level individual acts as a unit of selection in its own right. Species might be individuals, but their differential evolutionary success might still arise entirely from natural selection acting upon their parts, that is, upon phenotypes of organisms. A trend toward increasing brain size, for example, might result from the greater longevity of big-brained species. But big-brained species might prosper only because the organisms within them tend to prevail in traditional competition.

But individuation of higher-level units is enough to invalidate the reductionism of traditional Darwinism—for pattern and style of evolution depend critically on the disposition of higher-level individuals, even when all selection occurs at the traditional level of organisms. Sewall Wright, for example, has often spoken of "interdemic selection" in his shifting balance theory (*46*), but he apparently uses this phrase in a descriptive sense and believes that the mechanism of change usually resides in selection among individual organisms, as when, for example, migrants from one deme swamp another. Still, the fact of deme structure itself—that is, the individuation of higher-level units within a species—is crucial to the operation of shifting balance. Without division into demes, and under panmixia, genetic drift could not operate as the major source of variation required by the theory.

We need not, however, confine ourselves to the simple fact of individuation as an argument against Darwinian reductionism. For the strong claim that higher-level individuals act as units of selection in their own right can often be made. Many evolutionary trends, for example, are driven by differential frequency of speciation (the analog of birth) rather than by differential extinction (the more usual style of selection by death). Features that enhance the frequency of speciation are often properties of populations, not of individual organisms, for example, dependence of dispersal (and resultant possibilities for isolation and speciation) on size and density of populations.

Unfortunately, the terminology of this area is plagued with a central confusion (some, I regret to say, abetted by my own previous writings). Terms like "interdemic selection" or "species selection" (*47*) have been used in the purely descriptive sense, when the sorting out among higher-level individuals may arise solely from natural selection operating upon organisms. Such cases are explained by Darwinian selection, although they are irreducible to organisms alone. The same terms have been restricted to cases of higher-level individuals acting as units of selection. Such situations are non-Darwinian, and irreducible on this strong criterion. Since issues involving the locus of selection are so crucial in evolutionary theory, I suggest that these terms only be used in the strong and restricted sense. Species selection, for example, should connote an irreducibility to individual organisms (because populations are acting as units of selection); it should not merely offer a convenient alternative description for the effects of traditional selection upon organisms.

The logic of species selection is sound, and few evolutionists would now doubt that it can occur in principle. The issue, again and as always in natural history, is one of relative frequency; how often does species selection occur, and how important is it in the panoply of evolutionary events. Fisher himself dismissed species selection because, relative to organisms, species are so few in number (within a clade) and so long in duration (*48*):

> The relative unimportance of this as an evolutionary factor would seem to follow decisively from the small number of closely related species which in fact do come into competition, as compared to the number of individuals in the same species; and from the vastly greater duration of the species compared to the individual.

But Fisher's argument rests on two hidden and questionable assumptions. (i) Mass selection can almost always be effective in transforming entire populations substantially in phenotype. The sheer number of organisms participating in this efficient process would then swamp any effect of selection among species. But if stasis be prevalent within established species, as the theory of punctuated equilibrium asserts and as paleontological experience affirms (overwhelmingly for marine invertebrates, at least), then the mere existence of billions of individuals and millions of generations guarantees no substantial role for directional selection upon organisms. (ii) Species selection depends on direct competition among species. Fisher argues for differential death (extinction) as the mechanism of species selection. I suspect, however, that differential frequency of speciation (selection by birth) is a far more common and effective mode of species selection. It may occur without direct competition between species, and can rapidly shift the average phenotype within a clade in regimes of random extinction.

J. Maynard Smith (*49*) has raised another objection against species selection: simply, that most features of organisms represent "things individual creatures do." How, he asks, could one attribute the secondary palate of mammals to species selection? But the origin of a feature is one thing (and I would not dispute traditional selection among organisms as the probable mechanism for evolving a secondary palate), and the spread of features through larger clades is another. Macroevolution is fundamentally about the combination of features and their differential spread. These phenomena lie comfortably within the domain of effective species selection. Many features must come to prominence primarily through their fortuitous phyletic link with high speciation rates. Mammals represent a lineage of therapsids that may have survived (while all others died) as a result of small body sizes and nocturnal habits. Was the secondary palate a key to their success, or did it piggyback on the high speciation rates often noted (for other reasons) in small-bodied forms. Did mammals survive the Cretaceous extinction, thereby inheriting the world from dinosaurs, as a result of their secondary palate, or did their small size again preserve them during an event that differentially wiped out large creatures.

Evolutionary Pattern by Interaction Between Levels

The hierarchical model, with its assertion that selection works simultaneously and differently upon individuals at a variety of levels, suggests a revised interpretation for many phenomena that have puzzled people where they implicitly assumed causation by selection upon organisms. In particular, it suggests that negative interaction between levels might be an important principle in maintaining stability or holding rates of change within reasonable bounds.

The "selfish DNA" hypothesis, for example, proposes that much middle-repetitive DNA exists within genomes not because it provides Darwinian benefits to phenotypes, but because genes can (in certain circumstances) act as units of selection. Genes that can duplicate themselves and move among chromosomes will therefore accumulate copies of themselves for their own Darwinian reasons. But why does the process ever stop? The authors of the hypothesis (*45*) suggest that phenotypes will eventually "notice" the redundant copies when the energetic cost of producing them becomes high enough to entail negative selection at the level of organisms. Stability may represent a balance between positive selection at the gene level and the negative selection it eventually elicits at the organism level.

All evolutionary textbooks grant a paragraph or two to a phenomenon called "overspecialization," usually dismissing it as a peculiar and peripheral phenomenon. It records the irony that many creatures, by evolving highly complex and ecologically constraining features for their immediate Darwinian advantage, virtually guarantee the short duration of their species by restricting its capacity for subsequent adaptation. Will a peacock or an Irish elk survive when the environment alters radically? Yet fancy tails and big antlers do lead to more copulations in the short run of a lifetime. Overspecialization is, I believe, a central evolutionary phenomenon that has failed to gain the attention it deserves because we have lacked a vocabulary to express what is really happening: the negative interaction of species-level disadvantage and individual-level advantage. How else can morphological specialization be kept within bounds, leaving a place for drab and persistent creatures of the world. The general phenomenon must also regulate much of human society, with many higher-level institutions compromised or destroyed by the legitimate demands of individuals (high salaries of baseball stars, perhaps).

Some features may be enhanced by positive interaction between levels. Stenotopy in marine invertebrates, for example, seems to offer advantages at both the individual level (when environments are stable) and at the species level (boosting rates of speciation by brooding larvae and enhancing possibilities for isolation relative to eurytopic species with planktonic larvae). Why then do eurytopic species still inhabit our oceans? Suppression probably occurs at the still higher level of clades, by the differential removal of stenotopic branches in major environmental upheavals that accompany frequent mass extinctions in the geological record.

If no negative effect from a higher level suppressed an advantageous lower-level phenomenon, then it might sweep through life. Sex in eukaryotic organisms may owe its prominence to unsuppressed positive interaction between levels. The advantages of sex have inspired a major debate among evolutionists during the past decade. Most authors seek traditional explanation in terms of benefit to organisms (*50*), for example, better chance for survival of some offspring if all are not Xeroxed copies of an asexual parent, but the genetically variable products of two individuals. Some, however, propose a spread by species selection, for example, by vastly higher speciation rates in sexual creatures (*51*).

The debate has often proceeded by mutual dismissal, each side proclaiming its own answers correct. Perhaps both are right, and sex predominates because two levels interact positively and are not suppressed at any higher level. No statement is usually more dull and unenlightening than the mediator's claim, "you're both right." In this case, however, we must adopt a different view of biological organization itself to grasp the mediator's wisdom—and the old solution, for once, becomes interesting in its larger implication. We live in a world with reductionist traditions, and do not react comfortably to notions of hierarchy. Hierarchical theories permit us to retain the value of traditional ideas, while adding substantially to them. They traffic in accretion, not substitution. If we abandoned the "either-or" mentality that has characterized arguments about units of selection, we would not only reduce fruitless and often acrimonious debate, but we would also gain a deeper understanding of nature's complexity through the concept of hierarchy.

A Higher Darwinism?

What would a fully elaborated, hierarchically based evolutionary theory be called? It would neither be Darwinism, as usually understood, nor a smoothly continuous extension of Darwinism, for it violates directly the fundamental reductionist tradition embodied in Darwin's focus on organisms as units of selection.

Still, the hierarchical model does propose that selection operates on appropriate individuals at each level. Should the term "natural selection" be extended to all levels above and below organisms; there is certainly nothing unnatural about species selection. Some authors have extended the term (*48*), while others, Slatkin for example (*52*), restrict natural selection to its usual focus upon individual organisms: "Species selection is analogous to natural selection acting on an asexual population" (*52*).

Terminological issues aside, the hierarchically based theory would not be Darwinism as traditionally conceived; it would be both a richer and a different theory. But it would embody, in abstract form, the essence of Darwin's argument expanded to work at each level. Each level generates variation among its individuals; evolution occurs at each level by a sorting out among individuals, with differential success of some and their progeny. The hierarchical theory would therefore represent a kind of "higher Darwinism," with the substance of a claim for reduction to organisms lost, but the domain of the abstract "selectionist" style of argument extended.

Moreover, selection will work differently on the objects of diverse levels. The phenomena of one level have analogs on others, but not identical operation. For example, we usually deny the effectiveness of mutation pressure at the level of organisms. Populations contain so many individuals that small biases in mutation rate can rarely establish a feature if it is under selection at all. But the analog of mutation pressure at the species level, directed speciation (directional bias toward certain phenotypes in derived species), may be a powerful agent of evolutionary trends (as a macroevolutionary alternative to species selection). Directed speciation can be effective (where mutation pressure is not) for two reasons: first, because its effects are not so easily swamped (given the restricted number of species within a clade) by differential extinction; second, because such phenomena as ontogenetic channeling in phyletic size increase suggest that biases in the production of species may be more prevalent than biases in the genesis of mutations.

Each level must be approached on its own, and appreciated for the special emphasis it places upon common phenomena, but the selectionist style of argument regulates all levels and the Darwinian vision is extended and generalized, not defeated, even though Darwinism, strictly constructed, may be superseded. This expansion may impose a literal wisdom upon that famous last line of *Origin of Species*, "There is grandeur in this view of life."

Darwin, at the centenary of his death, is more alive than ever. Let us continue to praise famous men.

References and Notes

1. I have argued [*Nat. Hist.* 91, 16 (April 1982)] that a third and larger theme captures the profound importance and intellectual power of Darwin's work in a more comprehensive way: his successful attempt to establish principles of reasoning for historical science. Each of his so-called "minor" works (treatises on orchids, worms, climbing plants, coral reefs, barnacles, for example) exhibits both an explicit and a covert theme—and the covert theme is a principle of reasoning for the reconstruction of history. The principles can be arranged in order of decreasing availability of information, but each addresses the fundamental issue: how can history be scientific if we cannot directly observe a past process: (i) If we can observe present processes at work, then we should accumulate and extrapolate their results to render the past. Darwin's last book, on the formation of vegetable mold by earthworms (1881), is also a treatise on this aspect of uniformitarianism. (ii) If rates are too slow or scales too broad for direct observation, then try to render the range of present results as stages of a single historical process. Darwin's first book on a specific subject, the subsidence theory of coral atolls (1842), is (in its covert theme) a disquisition on this principle. (iii) When single objects must be analyzed, search for imperfections that record constraints of inheritance. Darwin's orchid book (1862), explicitly about fertilization by insects, argues that orchids are jury-rigged, rather than well built from scratch, because structures that attract insects and stick pollen to them had to be built from ordinary parts of ancestral flowers. Darwin used all three principles to establish evolution as well: (i) observed rates of change in artificial selection, (ii) stages in the process of speciation displayed by modern populations, and (iii) analysis of vestigial structures in various organisms. Thus, we should not claim that all Darwin's books are about evolution. Rather, they are all about the methodology of historical science. The establishment of evolution represents the greatest triumph of the method.
2. C. Darwin, *The Descent of Man* (Murray, London, ed. 2, 1889), p. 61.
3. G. J. Romanes, *Darwin, and After Darwin* (Longmans, Green, London, 1900), pp. 1–36.
4. Failure to recognize that all creationists accepted selection in this negative role led Eiseley to conclude falsely that Darwin had "borrowed" the principle of natural selection from his predecessor E. Blyth [L. Eiseley, *Darwin and the Mysterious Mr. X* (Dutton, New York, 1979)]. The Reverend William Paley's classic work *Natural Theology*, published in 1803, also contains many references to selective elimination.
5. By "random" in this context, evolutionists mean only that variation is not inherently directed towards adaptation, not that all mutational changes are equally likely. The word is unfortunate, but the historical tradition too deep to avoid.
6. E. D. Cope, *The Origin of the Fittest* (Appleton, New York, 1887).
7. L. G. Wilson, Ed., *Sir Charles Lyell's Scientific Journals on the Species Question* (Yale Univ. Press, New Haven, Conn., 1970), p. 369.
8. Darwin was convinced, for example, in part by reading a theological work arguing that extreme rapidity (as in the initial spread of Christianity) indicated a divine hand, that gradual and continuous change was the mark of a natural process [H. Gruber, *Darwin on Man* (Dutton, New York, 1974)].
9. C. Darwin, *The Structure and Distribution of Coral Reefs* (Smith, Elder, London, 1842).
10. ———, *The Formation of Vegetable Mould, Through the Action of Worms* (Murray, London, 1881).
11. The following works have done great service in identifying and correcting this confusion: G. C. Williams, *Adaptation and Natural Selection* (Princeton Univ. Press, Princeton, N.J., 1966); J. Maynard Smith, *The Evolution of Sex* (Cambridge Univ. Press, New York, 1978).
12. A persuasive case for Darwin's active interest in this subject and for his commitment to individual selection has been recently made by M. Ruse, *Ann. Sci.* 37, 615 (1980).
13. C. Darwin, *On the Origin of Species* (Murray, London, 1859).
14. S. S. Schweber, *J. Hist. Biol.* 10, 229 (1977).
15. S. Tax, Ed., *Evolution After Darwin* (Univ. of Chicago Press, Chicago, 1960), vols. 1–3.
16. J. Huxley, *Evolution, the Modern Synthesis* (Allen & Unwin, London, 1942).
17. For example: "The opposing factions became reconciled as the younger branches of biology achieved a synthesis with each other and with the classical disciplines: and the reconciliation converged upon a Darwinian center" (*16*, p. 25).
18. E. Mayr, *Systematics and the Origin of Species* (Columbia Univ. Press, New York, 1942); G. G. Simpson, *Tempo and Mode in Evolution* (Columbia Univ. Press, New York, 1944); B. Rensch, *Neuere Probleme der Abstammungslehre* (Enke, Stuttgart, 1947); G. L. Stebbins, *Variation and Evolution in Plants* (Columbia Univ. Press, New York, 1950).
19. S. J. Gould, in *The Evolutionary Synthesis*, E. Mayr and W. B. Provine, Eds. (Harvard Univ. Press, Cambridge, Mass., 1980), p. 153.
20. S. J. Gould, *Dobzhansky and the Modern Synthesis*, introduction to reprint of first (1937) edition of Th. Dobzhansky, *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1982).
21. D. Lack, *Darwin's Finches* (Harper Torchbook Edition, New York, 1960).
22. This statement appears as the first paragraph in the preface to (*21*).
23. E. Mayr, in *The Evolutionary Synthesis*, E. Mayr and W. B. Provine, Eds. (Harvard Univ. Press, Cambridge, Mass., 1980), p. 1.
24. S. Orzack, *Paleobiology* 7, 128 (1981).
25. M. J. D. White, *ibid.*, p. 287.
26. G. L. Stebbins and F. J. Ayala, *Science* 213, 967 (1981).
27. J. L. King and T. H. Jukes, *ibid.* 164, 788 (1969).
28. S. J. Gould, *Paleobiology* 6, 119 (1980).
29. C. Darwin (*13*, p. 189). On the day before publication of the *Origin of Species*, T. H. Huxley wrote to Darwin (letter of 23 November 1859): "You load yourself with an unnecessary difficulty in adopting *Natura non facit saltum* so unreservedly."
30. S. J. Gould, *The Uses of Heresy*, introduction to the republication of the 1940 edition of R. Goldschmidt, *The Material Basis of Evolution* (Yale Univ. Press, New Haven, Conn., 1982).

31. P. Alberch, *Am. Zool.* 20, 653 (1980).
32. M. J. D. White, *Modes of Speciation* (Freeman, San Francisco, 1978); G. L. Bush, S. M. Case, A. C. Wilson, J. L. Patton, *Proc. Natl. Acad. Sci. U.S.A.* 74, 3942 (1977).
33. N. Eldredge and S. J. Gould, in *Models in Paleobiology*, T. J. M. Schopf, Ed. (Freeman, Cooper, San Francisco, 1972), p. 82; S. J. Gould and N. Eldredge, *Paleobiology* 3, 115 (1977).
34. P. Williamson, *Nature (London)* 293, 437 (1981).
35. D'Arcy W. Thompson, *On Growth and Form* (Cambridge Univ. Press, New York, 1942).
36. S. J. Gould and R. C. Lewontin, *Proc. R. Soc. London Ser. B* 205, 581 (1979); G. V. Lauder, *Paleobiology* 7, 430 (1981).
37. It is the concluding comment of chapter 6, and reads, in part: "It is generally acknowledged that all organic beings have been formed on two great laws—Unity of Type, and the Conditions of Existence. . . . Natural selection acts by either now adapting the varying parts of each being to its organic and inorganic conditions of life; or by having adapted them during long-past periods of time. . . . Hence, in fact, the law of the Conditions of Existence is the higher law; as it includes, through the inheritance of former adaptations, that of Unity of Type."
38. St. G. Mivart, *On the Genesis of Species* (Macmillan, London, 1871).
39. A. Seilacher, *Lethaia* 5, 325 (1972).
40. S. J. Gould and E. S. Vrba, *Paleobiology*, in press.
41. R. C. Lewontin, *Annu. Rev. Ecol. Syst.* 1, 1 (1970).
42. R. Dawkins, *The Selfish Gene* (Oxford Univ. Press, New York, 1976).
43. V. C. Wynne-Edwards, *Animal Dispersion in Relation to Social Behavior* (Oliver & Boyd, Edinburgh, 1962).
44. M. Ghiselin, *Syst. Zool.* 23, 536 (1974); D. L. Hull, *Annu. Rev. Ecol. Syst.* 11, 311 (1980).
45. W. F. Doolittle and C. Sapienza, *Nature (London)* 284, 601 (1980); L. E. Orgel and F. H. C. Crick, *ibid.*, p. 604.
46. S. Wright, *Evolution and the Genetics of Populations* (Univ. of Chicago Press, Chicago, 1968–1978), vols. 1–4.
47. S. M. Stanley, *Macroevolution* (Freeman, San Francisco, 1979); *Proc. Natl. Acad. Sci. U.S.A.* 72, 646 (1975); also references in (*33*).
48. R. A. Fisher, *The Genetical Theory of Natural Selection* (Dover, ed. 2, New York, 1958), p. 50.
49. J. Maynard Smith, personal communication.
50. G. C. Williams, *Sex and Evolution* (Monographs in Population Biology, No. 8, Princeton Univ. Press, Princeton, N.J., 1975).
51. S. M. Stanley, *Science* 190, 382 (1975).
52. M. Slatkin, *Paleobiology* 7, 421 (1981).
53. I thank Ernst Mayr, Philip Kitcher, Montgomery Slatkin, and Steven Stanley for their most helpful comments. Malcolm Kottler kindly pointed out to me the passage from David Lack quoted in (*21, 22*).

This View of Life

Gould, S.J.
"What happens to bodies if genes act for themselves?"
<u>Natural History</u>
American Museum of Nat. Hist.

Used by permission

The Ultimate Parasite

What happens to bodies if genes act for themselves?

by Stephen Jay Gould

The uncommon good prose of scientists is more often spare than flowery. In my favorite example, James D. Watson and Francis Crick used less than a page to announce their structure of DNA in 1953. They began with the sparsest announcement: "We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest." And they ended with a reminder that they had not overlooked a major point just because they had chosen to defer its discussion: "It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material" (that is, the two strands of the double helix would pull apart and each then act as a template for the reconstitution of its partner).

Francis Crick, now a professor at the Salk Institute in southern California, has continued to generate controversial, challenging hypotheses (and he has often been right). In late 1981, he will publish a book, *Life Itself*, advocating a theory of "directed panspermia"—the idea that Earth's original life arrived as microorganisms dispatched by intelligent beings who chose not to make the long journey themselves. (Ten will get you fifty that he's wrong this time—but only fifty; he's been right too often.)

Crick has also not lost his gift for a well-turned phrase. In the presentation of his latest controversial hypothesis, published a year ago in *Nature* (April 17, 1980) with Salk colleague Leslie Orgel as first author, he outdid the last line of his 1953 paper with Watson. Orgel and Crick conclude: "The main facts are, at first sight, so odd that only a somewhat unconventional idea is likely to explain them." Indeed, the facts are so interesting, and the wondering about them so intense, that the same issue of *Nature* carried an accompanying article by Dalhousie University biologists W. Ford Doolittle and Carmen Sapienza, who had, quite independently, devised the same explanation and argued the case, in many ways, more forcefully.

What, then, are these disturbing facts? When a younger Crick determined the structure of DNA in 1953, and others cracked the genetic code a few years later, everything seemed momentarily to fall into order. The old idea of genes as beads on a string (the chromosome) seemed to gain its vindication from the Watson-Crick model. Each three nucleotides in DNA codes for an amino acid (via an RNA intermediary); a string of amino acids makes a protein. Perhaps we could simply read down a chromosome to find genes lined up, one after the other, each ready to begin the assembly of its essential part.

It was not to be so. Is it ever? We now know that the genetic material of higher organisms is vastly more complex. Many genes come in pieces, separated in DNA by sequences of nucleotides that are not transcribed into RNA. Many proteins are coded by partial sequences on two or more chromosomes. What controls regulate their assembly? (Human globin, the protein component of hemoglobin, contains alpha and beta chains—and the genes for each chain are on separate chromosomes.)

Even more disturbing (and exhilarating) is the discovery, made more than a decade ago but gathering intensity ever since, that only a small percentage of DNA codes for proteins in higher organisms—and that these are the only bits of DNA whose function we may truly understand at the moment. In humans, somewhat more than 1 percent, but not as much as 2 percent, of DNA codes for proteins. Much of the rest contains sequences that are repeated over and over again—hundreds or thousands of identical (or nearly identical) beads, sometimes following one after the other, sometimes dispersed widely over several chromosomes. Why so many copies? What do they do? The "selfish DNA" hypothesis of Doolittle, Sapienza, Orgel, and Crick provides an unusual answer to the puzzling question of why so much DNA exists in repeated copies (but I will keep you in suspense for a bit and discuss the conventional answers first).

Higher organisms contain different classes of repeated DNA. One type, called highly repeated or satellite DNA, contains short and simple sequences repeated hundreds of thousands or millions of times; 5 percent or so of human DNA falls into this class. We hardly have a clue about the origin and function of satellite DNA; neither the selfish DNA hypothesis nor the conventional hypotheses can explain it. Satellite DNA is, as they say, a "whole 'nother" story waiting to be told.

The current debate over the conventional and selfish DNA hypotheses centers upon the so-called intermediate or middle-repetitive DNA, some 15 to 30 percent of both the human and the fruit fly genome. Middle-repetitive DNA exists in tens to a few hundred copies per sequence; the copies are often widely dispersed on several chromosomes.

I have said nothing, so far, about the DNA of simpler organisms—the prokaryotic bacteria and blue-green algae, which have no nucleus and carry their DNA in a single chromosome. The DNA of prokaryote (prenucleate) organisms is "better behaved" with reference to the original hopes of the Watson-Crick model. Most bacterial DNA is single copy and protein coding, almost those beads on a string after all. But even prokaryotes are not immune to repetition. A hot topic of late concerns the presence in prokaryotes of so-called transposons, transposable elements, or more colorfully, jumping genes. These sequences of DNA, as their various names proclaim, can repeat themselves and then autonomously move about to other positions on the bacterial chromosome. They often exist in about as many copies as middle-repetitive DNA in eukaryotes (higher organisms with a nucleus and paired chromosomes). This has led many biologists to propose that at least some of the middle-repetitive DNA in higher organisms amplifies itself by the same mechanism of transposition. (The selfish DNA hypothesis assumes a correspondence between prokaryote transposons and the source of middle-repetitive DNA in eukaryotes. Some middle-repetitive DNA probably arises in other ways, and selfish DNA will therefore not explain all of it.)

Conventional arguments for the existence of middle-repetitive DNA follow the usual Darwinian perspective. Evolution is about the struggle of individuals (bodies) to leave more surviving offspring in future generations. This struggle operates by natural selection and selection is a potent editor. Major features of organisms—and some 25 percent of the genetic material cannot be minor—must exist because they provide some advantages to individuals in the struggle for life. We must, in other words, find a function for middle-repetitive DNA in terms of advantages to the bodies that carry it.

Rumblings of claims for nonadaptive and nonfunctional status have been heard from time to time (selfish DNA is the first, and more subtle, explosion for this perspective). Still, as Doolittle and Sapienza detail in their article, the overwhelming majority of proposals have hewed to Darwinian orthodoxy: they assume that middle-repetitive DNA cannot exist in such amounts unless it confers direct adaptive benefits upon organisms. (I will save myself some words from now on by simply writing "repetitive DNA" when I mean only "middle-repetitive DNA.")

The conventional adaptationist hypotheses have fallen into two classes: one, I believe, obviously wrong on (unrecognized) principle; the other undoubtedly correct in part (I do not believe that all repetitive DNA is selfish DNA). The unreasonable arguments postulate what I like to call a "retrospective significance" for repetitive DNA—that is, they justify its existence by discussing the benefits it may confer upon distant evolutionary futures.

Suppose all working genes could only exist in one copy that coded for an essential protein. How then could substantial evolutionary change ever occur? What will supply the essential protein while evolution monkeys about with the only coding sequence that produces it? But if a gene can repeat itself, then one copy might continue to code for the essential protein, leaving the other free to change. Thus, potential flexibility for evolutionary change has often been cited as the primary significance of repetitive DNA.

I have no quarrel with the idea that redundancy may supply the flexibility that evolution requires for initiating major changes. Susumu Ohno, who first popularized this idea in 1970 in a brilliant book (*Evolution by Gene Duplication*), argued that, without redundancy "from a bacterium only numerous forms of bacteria would have emerged." Duplication supplies the raw material of major evolutionary change: "The creation of a new gene from a redundant copy of an old gene is the most important role that gene duplication played in evolution."

But think about it for a moment. The argument is sound and may represent, in fact, the major *effect* of gene duplication for evolution. Yet unless our usual ideas about causality are running in the wrong direction, this flexibility simply cannot be the adaptive explanation for why repetitive DNA exists. Selection works on individuals at the moment. It cannot sense what may be of use ten million years hence in a distant descendant. The duplicated gene may make future evolutionary change possible, but selection cannot preserve it unless it confers an "immediate significance." Future utility is an important consideration in evolution, but it cannot be the explanation for current preservation. Future utilities can only be the *fortuitous effects* of other direct reasons for immediate favor.

(The confusion of *current utility* with *reasons for past historical origin* is a logical trap that has plagued evolutionary thinking from the start. Feathers work beautifully in flight, but the ancestors of birds must have evolved them for another reason—probably for thermoregulation—since a few feathers on the arm of a small running reptile will not induce takeoff. Our brains enlarged for a set of complex reasons, but surely not so that some of us could write essays about it. Interested readers may wish to consult a technical article, to appear in *Paleobiology* during 1982, that Elisabeth Vrba and I have written about this subject. We wish to restrict the term *adaptation* only to those structures that evolved for their current utility; those useful structures that arose for other reasons or for no conventional reason at all, and were then fortuitously available for other usages, we call exaptations. New and important genes that evolved from a repeated copy of an ancestral gene can only be exaptations, and their new usage cannot be the reason for the original duplication.)

The second set of adaptive arguments is legitimate in proposing an immediate selective benefit for repeated DNA. If genes move about and insert themselves on different chromosomes, for example, they may occasionally link up with other segments of DNA to form advantageous new combinations. More importantly, much DNA, while not coding for protein itself, may play a role in regulating the DNA that does. This regulatory DNA may turn other genes on and off and may determine the sequence and location of expression for the genes that do code for proteins. If repetitive DNA performs these regulatory functions, then its dispersal throughout the genome can have profound immediate effects. Inserted into a new chromosome, it may turn adjacent genes on and off in new ways and sequences. It may, for example, bring together the products of two genes that had never been in proximity. This new combination may benefit an organism (see the classic article of Roy Britten and Eric Davidson, *Quarterly Review of Biology*, 1971, pp. 111–31).

Yet, for all these efforts, the nagging suspicion remains that these adaptive explanations cannot account for all repetitive DNA. There is simply too much of it, too randomly dispersed, too apparently nonsensical in its construction, to argue that each item perseveres because natural selection had favored it in a regulatory role. The selfish DNA hypothesis proposes a fundamentally different explanation for much of this repetition. It is radical in that literal sense of getting to the roots, for it demands that we reassess some basic and usually unquestioned assumptions of evolutionary argument—what Orgel and Crick meant when they spoke of facts "so odd that only a somewhat unconventional idea is likely to explain them."

The argument is simplicity itself once you establish the frame of mind to permit it: if repetitive DNA is transposable, then why do we need an adaptive explanation for it (in conventional terms of benefits to bodies) at all? It may simply spread of its own accord from chromosome to chromosome, making more copies of itself while other "sedentary" genes cannot. These extra copies may persist, not because they confer advantages upon bodies, but for precisely the opposite reason—because bodies do not notice them. If they have no effect upon bodies, if they are (in this sense) "junk," then what is to stop their spread? They are merely playing Darwin's game, but at the wrong level. We usually think of natural selection as a struggle among bodies to leave more surviving offspring. Here certain genes have found a way, through transposability, or "jumping," to leave more copies of themselves *within* a body. Is any other explanation required? Orgel and Crick's title reflects this reversed perspective: "Selfish DNA: The Ultimate Parasite."

I can now almost hear the disappointment and anger of some readers: "That bastard Gould. He led us along for pages, and now he gives an explanation that is no explanation at all. It just plain happens, and that's all there is to it. Is this a joke or a counsel of despair?" I beg to differ from this not entirely hypothetical adversary (a composite constructed from several real responses I have received to verbal descriptions of the selfish DNA hypothesis). The explanation seems hokey only in the context of adherence to traditional views that all important features must be adaptations and that bodies are *the* agent of Darwinian processes. The radical content of selfish DNA is not the explanation itself, but the reformulated perspective that must be assimilated before the explanation confers any satisfaction.

If bodies are the only "individuals" that count in evolution, then selfish DNA is unsatisfying because it does nothing for bodies and can only be seen as random with respect to bodies. But why should bodies occupy such a central and privileged position in evolutionary theory? To be sure, selection can only work on discrete individuals with inherited continuity from ancestor to descendant. But are bodies the only kind of legitimate individuals in biology? Might there not be a hierarchy of individuals, with legitimate categories both above and below bodies: genes below, species above. (I confess to what evolutionists call a "preadaptation" for favorable response to the selfish DNA hypothesis. I have long argued that species must be viewed as true evolutionary units and that macroevolutionary trends are often powered by a "species selection" that is analogous to, but not identical with, natural selection acting upon bodies.) Selfish DNA may do nothing for bodies, but bodies are the wrong level of analysis. From a gene's point of view, transposable elements have developed a great Darwinian innovation: they have found a way to make more surviving copies of themselves (by repetition and transposition), and this, in itself, is the evolutionary *summum bonum*. If bodies don't notice this repetition, and therefore cannot suppress it by dying or failing to reproduce, then so much the better for repeating genes.

In this sense, selfish DNA is about the worst possible name for the phenomenon, for it records the very prejudice that the new structure of explanation should be combating: the exclusive focus on bodies as evolutionary agents. When we call repetitive DNA "selfish," we imply that it is acting for itself when it should be doing something else, namely, helping bodies in their evolutionary struggle. Likewise, we should

not refer to repetitive DNA as "nonadaptive," for although it may not be helping bodies, it is acting as its own Darwinian agent. I can't think of a much better term in a language replete with anthropocentric terms, but how about "self-centered DNA"—without the opprobrious overtones that "selfish" inevitably contains.

Another argument against the use of selfish DNA lies in its historical source: Richard Dawkins's book *The Selfish Gene* (1976). Dawkins argued that bodies are the wrong level of evolutionary analysis and that all evolution is nothing but a struggle among genes. Bodies are merely temporary containers for their selfish genes. Superficially, this looks like selfish DNA writ larger, hence Orgel and Crick's decision to borrow the term. In fact, the theories of selfish genes and selfish DNA could not be more different in the structures of explanation that nurture them.

Dawkins writes as a superstrict modern Darwinian, committed to the idea that all features must be interpreted as adaptations and that all of evolution is a struggle for existence among individuals at the lowest level. He merely decided that Darwinians weren't radical enough in reducing such higher-level reveries as "the good of the species" or "the harmony of nature" to the unrestrained struggle of individuals. The struggling items are one level lower—genes rather than bodies—and the Darwinian program of reduction can go even further than modern supporters had dared to hope.

Selfish DNA, on the other hand, gains its rationale from the antireductionistic belief that evolution works on a hierarchy of legitimate levels that cannot be collapsed to the first rung of the scale. Dawkins's selfish genes increase in frequency because they have effects upon bodies, aiding them in their struggle for existence. Selfish DNA increases in frequency for precisely the opposite reason—because it initially has *no effect* on bodies and therefore is not suppressed at this legitimate higher level. Dawkins's theory is an unconventional proposal to explain ordinary adaptation of bodies. Selfish DNA survives only because it makes no difference to bodies.

The idea of selfish genes is, in my opinion, fundamentally incorrect (however stimulating) because the features of bodies that make a difference in natural selection cannot be decomposed into products of individual genes. Each gene contributes to several features and each feature is built by several genes and their complex interactions. When a body survives because natural selection has favored a set of features, selection has touched so many genes in so many complex and unbreakable connections that the idea of individual genes battling for personal survival makes no sense. (See my column of December 1977, but see especially the critique of Dawkins by evolution's grandest old man, Sewall Wright—92 and still going strong—in *Evolution*, vol. 34, 1980, pp. 825–43.) But selfish DNA can spread because it makes no features in bodies. It may be adaptive at its own level, but it is invisible and nonadaptive at the level of natural selection among bodies. Hence it may accumulate at first because natural selection cannot "see" and suppress it.

But if middle-repetitive DNA is self-centered, why does it only exist in hundreds of copies within genomes? If it can spread by transposition while other genes cannot, why does it not generate millions and billions of copies, eventually crowding everything else out? What stops it? Why is it behaving as an "intelligent" parasite (enough copies to be comfortable and powerful, but not enough to destroy the host and itself), rather than as a voracious cancer?

The potential answer to this question, proposed by both sets of authors, illustrates another interesting point about the hierarchical mode of thinking that underlies the theory of self-centered DNA. In hierarchical models, levels are not independent, walled off by impenetrable boundaries from those above and below. Levels leak and interact. Arthur Koestler, whom I do not usually praise but whose commitment to hierarchy I find admirable, chose as his metaphor for hierarchy the double-faced god Janus, standing at one level but looking for connections in both directions.

Now consider different forms of selection working at levels of gene, body, and species. A transposon enters a genetic system and begins to amplify itself by replication and movement. In the process of selection among genes, it is increasing by an analog of what we would call "differential birth" in natural selection among bodies. Its increase initially produces no interaction with the level of natural selection upon bodies, and nothing suppresses its intrinsic drive to manufacture more copies.

But eventually, if its increase continues unabated, bodies must begin to notice. There is an energetic cost attached to the replication, generation after generation, of hundreds or thousands of copies of DNA sequences that do nothing for the bodies investing that energy. Bodies may not notice a few copies, but vast numbers must eventually produce a disadvantage at the good old Darwinian level of natural selection among bodies. At this point, further increases in copies of self-centered DNA will be suppressed because bodies carrying too many copies will suffer in natural selection, taking all their copies with them when they die or fail to reproduce. The usual level of tens to hundreds of copies may well represent a balance between inexorable increase at the level of selection among genes and eventual suppression at the next level of selection among bodies. Levels are connected by complex ties of feedback. My plea for a recognition of levels other than natural selection acting upon bodies is not a negation of Darwinian theory but an attempt to enrich it.

The arguments will continue for a long time. One group of scientists notes the similarity in arrangement within chromosomes of repetitive sequences in two creatures as evolutionarily distant as the toad *Xenopus laevis* and the sea urchin *Strongylocentrotus purpuratus*. This similarity refutes self-centered DNA and points to common function, since wandering transposons, beholden only to their own level, should disperse more randomly among chromosomes. Others point out that an important transposable element in yeast and another in the fruit fly *Drosophila melanogaster* are represented in different strains of the same species by about the same number of copies, but in very different positions among chromosomes. Do the different positions represent self-centered amplification and the similarity in numbers reflect suppression at the higher level of selection upon bodies?

As with all interesting questions in natural history, the solution requires an inquiry about relative frequency, not an absolute yes or no. The logic of self-centered DNA seems sound. The question remains: how important is it? How much repetitive DNA is self-centered DNA? If the answer is "way less than one percent" because conventional selection on bodies almost always overwhelms selection among genes, then self-centered DNA is one more good and plausible idea scorned by nature. If the answer is "lots of it," then we will need a fully articulated hierarchical theory of evolution. My own inclinations are, obviously, for hierarchy. Cartesian reductionism has been the source of science's triumph for 300 years; but I suspect that we have reached its limits in several areas.

We have legitimate, idiosyncratic reasons for continuing our linguistic habit of identifying "individuals" with bodies, and for granting a primacy to bodies among the objects of nature. I can't, for example, imagine any acceptable politics that does not focus upon the primacy of individual bodies—and we weep for the inhumanity of those that did not, but flourished for a time nonetheless. Nature, however, acknowledges many kinds of individuals, both great and small.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

Proc. R. Soc. Lond. B 205, 581–598 (1979)
Printed in Great Britain

The spandrels of San Marco and the Panglossian paradigm: a critique of the adaptationist programme

By S. J. Gould and R. C. Lewontin
Museum of Comparative Zoology, Harvard University,
Cambridge, Massachusetts 02138, *U.S.A.*

An adaptationist programme has dominated evolutionary thought in England and the United States during the past 40 years. It is based on faith in the power of natural selection as an optimizing agent. It proceeds by breaking an organism into unitary 'traits' and proposing an adaptive story for each considered separately. Trade-offs among competing selective demands exert the only brake upon perfection; non-optimality is thereby rendered as a result of adaptation as well. We criticize this approach and attempt to reassert a competing notion (long popular in continental Europe) that organisms must be analysed as integrated wholes, with *Baupläne* so constrained by phyletic heritage, pathways of development and general architecture that the constraints themselves become more interesting and more important in delimiting pathways of change than the selective force that may mediate change when it occurs. We fault the adaptationist programme for its failure to distinguish current utility from reasons for origin (male tyrannosaurs may have used their diminutive front legs to titillate female partners, but this will not explain *why* they got so small); for its unwillingness to consider alternatives to adaptive stories; for its reliance upon plausibility alone as a criterion for accepting speculative tales; and for its failure to consider adequately such competing themes as random fixation of alleles, production of non-adaptive structures by developmental correlation with selected features (allometry, pleiotropy, material compensation, mechanically forced correlation), the separability of adaptation and selection, multiple adaptive peaks, and current utility as an epiphenomenon of non-adaptive structures. We support Darwin's own pluralistic approach to identifying the agents of evolutionary change.

Gould, S.J. & Lewontin, R.C.
"The spandrels of San Marco:
a critique of the adaptationist
program"
Proceedings of the Royal Society
of London
Published by the Royal Society
Used by permission

1. Introduction

The great central dome of St Mark's Cathedral in Venice presents in its mosaic design a detailed iconography expressing the mainstays of Christian faith. Three circles of figures radiate out from a central image of Christ: angels, disciples, and virtues. Each circle is divided into quadrants, even though the dome itself is radially symmetrical in structure. Each quadrant meets one of the four spandrels in the arches below the dome. Spandrels – the tapering triangular spaces formed by the intersection of two rounded arches at right angles (figure 1) – are necessary architectural by-products of mounting a dome on rounded arches. Each spandrel contains a design admirably fitted into its tapering space. An evangelist sits in the upper part flanked by the heavenly cities. Below, a man representing one of the four Biblical rivers (Tigris, Euphrates, Indus and Nile) pours water from a pitcher into the narrowing space below his feet.

The design is so elaborate, harmonious and purposeful that we are tempted to view it as the starting point of any analysis, as the cause in some sense of the surrounding architecture. But this would invert the proper path of analysis. The system begins with an architectural constraint: the necessary four spandrels and their tapering triangular form. They provide a space in which the mosaicists worked; they set the quadripartite symmetry of the dome above.

Figure 1. One of the four spandrels of St Mark's; seated evangelist above, personification of river below.

Such architectural constraints abound and we find them easy to understand because we do not impose our biological biases upon them. Every fan vaulted ceiling must have a series of open spaces along the mid-line of the vault, where the sides of the fans intersect between the pillars (figure 2). Since the spaces must exist, they are often used for ingenious ornamental effect. In King's College Chapel in Cambridge, for example, the spaces contain bosses alternately embellished with

the Tudor rose and portcullis. In a sense, this design represents an 'adaptation', but the architectural constraint is clearly primary. The spaces arise as a necessary by-product of fan vaulting; their appropriate use is a secondary effect. Anyone who tried to argue that the structure exists because the alternation of rose and portcullis makes so much sense in a Tudor chapel would be inviting the same ridicule that Voltaire heaped on Dr Pangloss: 'Things cannot be other than they are...Everything is made for the best purpose. Our noses were made to carry spectacles, so we have spectacles. Legs were clearly intended for breeches, and we wear them.' Yet evolutionary biologists, in their tendency to focus exclusively on immediate adaptation to local conditions, do tend to ignore architectural constraints and perform just such an inversion of explanation.

FIGURE 2. The ceiling of King's College Chapel.

As a closer example, recently featured in some important biological literature on adaptation, anthropologist Michael Harner has proposed (1977) that Aztec human sacrifice arose as a solution to chronic shortage of meat (limbs of victims were often consumed, but only by people of high status). E. O. Wilson (1978) has used this explanation as a primary illustration of an adaptive, genetic predisposition for carnivory in humans. Harner and Wilson ask us to view an elaborate

social system and a complex set of explicit justifications involving myth, symbol, and tradition as mere epiphenomena generated by the Aztecs as an unconscious rationalization masking the 'real' reason for it all: need for protein. But Sahlins (1978) has argued that human sacrifice represented just one part of an elaborate cultural fabric that, in its entirety, not only represented the material expression of Aztec cosmology, but also performed such utilitarian functions as the maintenance of social ranks and systems of tribute among cities.

We strongly suspect that Aztec cannibalism was an 'adaptation' much like evangelists and rivers in spandrels, or ornamented bosses in ceiling spaces: a secondary epiphenomenon representing a fruitful use of available parts, not a cause of the entire system. To put it crudely: a system developed for other reasons generated an increasing number of fresh bodies; use might as well be made of them. Why invert the whole system in such a curious fashion and view an entire culture as the epiphenomenon of an unusual way to beef up the meat supply. Spandrels do not exist to house the evangelists. (Moreover, as Sahlins argues, it is not even clear that human sacrifice was an adaptation at all. Human cultural practices can be orthogenetic and drive towards extinction in ways that Darwinian processes, based on genetic selection, cannot. Since each new monarch had to outdo his predecessor in even more elaborate and copious sacrifice, the practice was beginning to stretch resources to the breaking point. It would not have been the first time that a human culture did itself in. And, finally, many experts doubt Harner's premise in the first place (Ortiz de Montellano 1978). They argue that other sources of protein were not in short supply, and that a practice awarding meat only to privileged people who had enough anyway, and who used bodies so inefficiently (only the limbs were consumed, and partially at that) represents a mighty poor way to run a butchery.)

We deliberately chose non-biological examples in a sequence running from remote to more familiar: architecture to anthropology. We did this because the primacy of architectural constraint and the epiphenomenal nature of adaptation are not obscured by our biological prejudices in these examples. But we trust that the message for biologists will not go unheeded: if these had been biological systems, would we not, by force of habit, have regarded the epiphenomenal adaptation as primary and tried to build the whole structural system from it?

2. THE ADAPTATIONIST PROGRAMME

We wish to question a deeply engrained habit of thinking among students of evolution. We call it the adaptationist programme, or the Panglossian paradigm. It is rooted in a notion popularized by A. R. Wallace and A. Weismann (but not, as we shall see, by Darwin) towards the end of the nineteenth century: the near omnipotence of natural selection in forging organic design and fashioning the best among possible worlds. This programme regards natural selection as so powerful and the constraints upon it so few that direct production of adaptation through its operation becomes the primary cause of nearly all organic form, function, and behaviour. Constraints upon the pervasive power of natural selection are recognized of course (phyletic inertia primarily among them, although immediate architectural constraints, as discussed in the last section, are rarely acknowledged). But they are usually dimissed as unimportant or else, and more frustratingly, simply acknowledged and then not taken to heart and invoked.

Studies under the adaptationist programme generally proceed in two steps:

(1) An organism is atomized into 'traits' and these traits are explained as structures optimally designed by natural selection for their functions. For lack of space, we must omit an extended discussion of the vital issue: 'what is a trait?' Some evolutionists may regard this as a trivial, or merely a semantic problem. It is not. Organisms are integrated entities, not collections of discrete objects. Evolutionists have often been led astray by inappropriate atomization, as D'Arcy Thompson (1942) loved to point out. Our favourite example involves the human chin (Gould 1977, pp. 381–382; Lewontin 1978). If we regard the chin as a 'thing', rather than as a product of interaction between two growth fields (alveolar and mandibular), then we are led to an interpretation of its origin (recapitulatory) exactly opposite to the one now generally favoured (neotenic).

(2) After the failure of part-by-part optimization, interaction is acknowledged via the dictum that an organism cannot optimize each part without imposing expenses on others. The notion of 'trade-off' is introduced, and organisms are interpreted as best compromises among competing demands. Thus, interaction among parts is retained completely within the adaptationist programme. Any suboptimality of a part is explained as its contribution to the best possible design for the whole. The notion that suboptimality might represent anything other than the immediate work of natural selection is usually not entertained. As Dr Pangloss said in explaining to Candide why he suffered from venereal disease: 'It is indispensable in this best of worlds. For if Columbus, when visiting the West Indies, had not caught this disease, which poisons the source of generation, which frequently even hinders generation, and is clearly opposed to the great end of Nature, we should have neither chocolate nor cochineal.' The adaptationist programme is truly Panglossian. Our world may not be good in an abstract sense, but it is the very best we could have. Each trait plays its part and must be as it is.

At this point, some evolutionists will protest that we are caricaturing their view of adaptation. After all, do they not admit genetic drift, allometry, and a variety of reasons for non-adaptive evolution? They do, to be sure, but we make a different point. In natural history, all possible things happen sometimes; you generally do not support your favoured phenomenon by declaring rivals impossible in theory. Rather, you acknowledge the rival, but circumscribe its domain of action so narrowly that it cannot have any importance in the affairs of nature. Then, you often congratulate yourself for being such an undogmatic and ecumenical chap. We maintain that alternatives to selection for best overall design have generally been relegated to unimportance by this mode of argument. Have we not all heard the catechism about genetic drift: it can only be important in populations so small that they are likely to become extinct before playing any sustained evolutionary role (but see Lande 1976).

The admission of alternatives in principle does not imply their serious consideration in daily practice. We all say that not everything is adaptive; yet, faced with an organism, we tend to break it into parts and tell adaptive stories as if trade-offs among competing, well designed parts were the only constraint upon perfection for each trait. It is an old habit. As Romanes complained about A. R. Wallace in 1900: 'Mr. Wallace does not expressly maintain the abstract impossibility of laws and causes other than those of utility and natural selection... Nevertheless, as he nowhere recognizes any other law or cause..., he practically concludes that, on inductive or empirical grounds, there *is* no such other law or cause to be entertained.'

The adaptationist programme can be traced through common styles of argument. We illustrate just a few; we trust they will be recognized by all:

(1) If one adaptive argument fails, try another. Zig-zag commissures of clams and brachiopods, once widely regarded as devices for strengthening the shell, become sieves for restricting particles above a given size (Rudwick 1964). A suite of external structures (horns, antlers, tusks) once viewed as weapons against predators, become symbols of intraspecific competition among males (Davitashvili 1961). The eskimo face, once depicted as 'cold engineered' (Coon *et al.* 1950), becomes an adaptation to generate and withstand large masticatory forces (Shea 1977). We do not attack these newer interpretations; they may all be right. We do wonder, though, whether the failure of one adaptive explanation should always simply inspire a search for another of the same general form, rather than a consideration of alternatives to the proposition that each part is 'for' some specific purpose.

(2) If one adaptive argument fails, assume that another must exist; a weaker version of the first argument. Costa & Bisol (1978), for example, hoped to find a correlation between genetic polymorphism and stability of environment in the deep sea, but they failed. They conclude (1978, pp. 132, 133): 'The degree of genetic polymorphism found would seem to indicate absence of correlation with the particular environmental factors which characterize the sampled area. The results suggest that the adaptive strategies of organisms belonging to different phyla are different.'

(3) In the absence of a good adaptive argument in the first place, attribute failure to imperfect understanding of where an organism lives and what it does. This is again an old argument. Consider Wallace on why all details of colour and form in land snails must be adaptive, even if different animals seem to inhabit the same environment (1899, p. 148): 'The exact proportions of the various species of plants, the numbers of each kind of insect or of bird, the peculiarities of more or less exposure to sunshine or to wind at certain critical epochs, and other slight differences which to us are absolutely immaterial and unrecognizable, may be of

the highest significance to these humble creatures, and be quite sufficient to require some slight adjustments of size, form, or colour, which natural selection will bring about.'

(4) Emphasize immediate utility and exclude other attributes of form. Fully half the explanatory information accompanying the full-scale Fibreglass *Tyrannosaurus* at Boston's Museum of Science reads: 'Front legs a puzzle: how *Tyrannosaurus* used its tiny front legs is a scientific puzzle; they were too short even to reach the mouth. They may have been used to help the animal rise from a lying position.' (We purposely choose an example based on public impact of science to show how widely habits of the adaptationist programme extend. We are not using glass beasts as straw men; similar arguments and relative emphases, framed in different words, appear regularly in the professional literature.) We don't doubt that *Tyrannosaurus* used its diminutive front legs for something. If they had arisen *de novo*, we would encourage the search for some immediate adaptive reason. But they are, after all, the reduced product of conventionally functional homologues in ancestors (longer limbs of allosaurs, for example). As such, we do not need an explicitly adaptive explanation for the reduction itself. It is likely to be a developmental correlate of allometric fields for relative increase in head and hindlimb size. This non-adaptive hypothesis can be tested by conventional allometric methods (Gould (1974) in general; Lande (1978) on limb reduction) and seems to us both more interesting and fruitful than untestable speculations based on secondary utility in the best of possible worlds. One must not confuse the fact that a structure is used in some way (consider again the spandrels, ceiling spaces and Aztec bodies) with the primary evolutionary reason for its existence and conformation.

3. TELLING STORIES

> 'All this is a manifestation of the rightness of things, since if there is a volcano at Lisbon it could not be anywhere else. For it is impossible for things not to be where they are, because everything is for the best' (Dr Pangloss on the great Lisbon earthquake of 1755 in which up to 50000 people lost their lives).

We would not object so strenuously to the adaptationist programme if its invocation, in any particular case, could lead in principle to its rejection for want of evidence. We might still view it as restrictive and object to its status as an argument of first choice. But if it could be dismissed after failing some explicit test, then alternatives would get their chance. Unfortunately, a common procedure among evolutionists does not allow such definable rejection for two reasons. First, the rejection of one adaptive story usually leads to its replacement by another, rather than to a suspicion that a different kind of explanation might be required. Since the range of adaptive stories is as wide as our minds are fertile, new stories can always be postulated. And if a story is not immediately available, one can always plead temporary ignorance and trust that it will be forthcoming, as did Costa & Bisol (1978), cited above. Secondly, the criteria for acceptance of a story are so loose that many pass without proper confirmation. Often, evolutionists use *consistency* with natural selection as the sole criterion and consider their work done when they concoct a plausible story. But plausible stories can always be told. The key to historical research lies in devising criteria to identify proper explanations among the substantial set of plausible pathways to any modern result.

We have, for example (Gould 1978) criticized Barash's (1976) work on aggression in mountain bluebirds for this reason. Barash mounted a stuffed male near the nests of two pairs of bluebirds while the male was out foraging. He did this at the same nests on three occasions at 10 day intervals: the first before eggs were laid, the last two afterwards. He then counted aggressive approaches of the returning male towards both the model and the female. At time one, aggression was high towards the model and lower towards females but substantial in both nests. Aggression towards the model declined steadily for times two and three and plummeted to near zero towards females. Barash reasoned that this made evolutionary sense since males would be more sensitive to intruders before eggs were laid than afterwards (when they can have some confidence that their genes are inside). Having devised this plausible story, he considered his work as completed (1976, pp. 1099, 1100):

> 'The results are consistent with the expectations of evolutionary theory. Thus aggression toward an intruding male (the model) would clearly be especially advantageous early in the breeding season, when territories and nests are normally defended...The initial aggressive response to the mated female is also adaptive in that, given a situation suggesting a high probability of adultery (i.e. the presence of the model near the female) and assuming that replacement females are available, obtaining a new mate would enhance the fitness of males... The decline in male–female aggressiveness during incubation and fledgling stages could be attributed to the impossibility of being cuckolded after the eggs have been laid...The results are consistent with an evolutionary interpretation.'

They are indeed consistent, but what about an obvious alternative, dismissed without test by Barash? Male returns at times two and three, approaches the model, tests it a bit, recognizes it as the same phoney he saw before, and doesn't bother his female. Why not at least perform the obvious test for this alternative to a conventional adaptive story: expose a male to the model for the *first* time after the eggs are laid.

Since we criticized Barash's work, Morton *et al.* (1978) repeated it, with some variations (including the introduction of a female model), in the closely related eastern bluebird *Sialia sialis*. 'We hoped to confirm', they wrote, that Barash's conclusions represent 'a widespread evolutionary reality, at least within the genus *Sialia*. Unfortunately, we were unable to do so.' They found no 'anticuckoldry' behaviour at all: males never approached their females aggressively after testing the model at any nesting stage. Instead, females often approached the male model and, in any case, attacked female models more than males attacked male models.

'This violent response resulted in the near destruction of the female model after presentations and its complete demise on the third, as a female flew off with the model's head early in the experiment to lose it for us in the brush' (1978, p. 969). Yet, instead of calling Barash's selected story into question, they merely devise one of their own to render both results in the adaptationist mode. Perhaps, they conjecture, replacement females are scarce in their species and abundant in Barash's. Since Barash's males can replace a potentially 'unfaithful' female, they can afford to be choosy and possessive. Eastern bluebird males are stuck with uncommon mates and had best be respectful. They conclude: 'If we did not support Barash's suggestion that male bluebirds show anticuckoldry adaptations, we suggest that both studies still had "results that are consistent with the expectations of evolutionary theory" (Barash 1976, p. 1099), as we presume any careful study would.' But what good is a theory that cannot fail in careful study (since by 'evolutionary theory', they clearly mean the action of natural selection applied to particular cases, rather than the fact of transmutation itself).

4. THE MASTER'S VOICE RE-EXAMINED

Since Darwin has attained sainthood (if not divinity) among evolutionary biologists, and since all sides invoke God's allegiance, Darwin has often been depicted as a radical selectionist at heart who invoked other mechanisms only in retreat, and only as a result of his age's own lamented ignorance about the mechanisms of heredity. This view is false. Although Darwin regarded selection as the most important of evolutionary mechanisms (as do we), no argument from opponents angered him more than the common attempt to caricature and trivialize his theory by stating that it relied exclusively upon natural selection. In the last edition of the *Origin*, he wrote (1872, p. 395):

> 'As my conclusions have lately been much misrepresented, and it has been stated that I attribute the modification of species exclusively to natural selection, I may be permitted to remark that in the first edition of this work, and subsequently, I placed in a most conspicuous position – namely at the close of the Introduction – the following words: "I am convinced that natural selection has been the main, but not the exclusive means of modification." This has been of no avail. Great is the power of steady misinterpretation.'

Romanes, whose once famous essay (1900) on Darwin's pluralism versus the panselectionism of Wallace and Weismann deserves a resurrection, noted of this passage (1900, p. 5): 'In the whole range of Darwin's writings there cannot be found a passage so strongly worded as this: it presents the only note of bitterness in all the thousands of pages which he has published.' Apparently, Romanes did not know the letter Darwin wrote to *Nature* in 1880, in which he castigated Sir Wyville Thomson for caricaturing his theory as panselectionist (1880, p. 32):

> 'I am sorry to find that Sir Wyville Thomson does not understand the principle of natural selection...If he had done so, he could not have written the following sentence in the Introduction to the Voyage of the Challenger: "The character of the abyssal fauna refuses to give the least support to the theory which refers the evolution of species to extreme variation guided only by natural selection." This is a standard of criticism not uncommonly reached by theologians and metaphysicians when they write on scientific subjects, but is something new as coming from a naturalist...Can Sir Wyville Thomson name any one who has said that the evolution of species depends only on natural selection? As far as concerns myself, I believe that no one has brought forward so many observations on the effects of the use and disuse of parts, as I have done in my "Variation of Animals and Plants under Domestication"; and these observations were made for this special object. I have likewise there adduced a considerable body of facts, showing the direct action of external conditions on organisms.'

We do not now regard all of Darwin's subsidiary mechanisms as significant or even valid, though many, including direct modification and correlation of growth, are very important. But we should cherish his consistent attitude of pluralism in attempting to explain Nature's complexity.

5. A PARTIAL TYPOLOGY OF ALTERNATIVES TO THE ADAPTATIONIST PROGRAMME

In Darwin's pluralistic spirit, we present an incomplete hierarchy of alternatives to immediate adaptation for the explanation of form, function, and behaviour.

(1) No adaptation and no selection at all. At present, population geneticists are sharply divided on the question of how much genetic polymorphism within populations and how much of the genetic differences between species is, in fact, the result of natural selection as opposed to purely random factors. Populations are finite in size and the isolated populations that form the first step in the speciation process are often founded by a very small number of individuals. As a result of this restriction in population size, frequencies of alleles change by *genetic drift*, a kind of random genetic sampling error. The stochastic process of change in gene frequency by random genetic drift, including the very strong sampling process that goes on when a new isolated population is formed from a few immigrants, has several important consequences. First, populations and species will become genetically differentiated, and even fixed for different alleles at a locus in the complete absence of any selective force at all.

Secondly, alleles can become fixed in a population *in spite of natural selection*. Even if an allele is favoured by natural selection, some proportion of population, depending upon the product of population size N and selection intensity s, will become homozygous for the less fit allele because of genetic drift. If Ns is large this random fixation for unfavourable alleles is a rare phenomenon, but if selection coefficients are on the order of the reciprocal of population size ($Ns = 1$) or smaller, fixation for deleterious alleles is common. If many genes are involved in influencing a metric character like shape, metabolism or behaviour, then the intensity of selection on each locus will be small and Ns per locus may be small. As a result, many of the loci may be fixed for non-optimal alleles.

Thirdly, new mutations have a small chance of being incorporated into a population, even when selectively favoured. Genetic drift causes the immediate loss of most new mutations after their introduction. With a selection intensity s, a new favourable mutation has a probability of only $2s$ of ever being incorporated. Thus, one cannot claim that, eventually, a new mutation of just the right sort for some adaptive argument will occur and spread. 'Eventually' becomes a very long time if only one in 1000 or one in 10000 of the 'right' mutations that do occur ever get incorporated in a population.

(2) No adaptation and no selection on the part at issue; form of the part is a correlated consequence of selection directed elsewhere. Under this important category, Darwin ranked his 'mysterious' laws of the 'correlation of growth'. Today, we speak of pleiotropy, allometry, 'material compensation' (Rensch 1959, pp. 179–187) and mechanically forced correlations in D'Arcy Thompson's sense (1942; Gould 1971). Here we come face to face with organisms as integrated wholes, fundamentally not decomposable into independent and separately optimized parts.

Although allometric patterns are as subject to selection as static morphology itself (Gould 1966), some regularities in relative growth are probably not under immediate adaptive control. For example, we do not doubt that the famous 0.66 interspecific allometry of brain size in all major vertebrate groups represents a selected 'design criterion,' though its significance remains elusive (Jerison 1973). It is too repeatable across too wide a taxonomic range to represent much else than a series of creatures similarly well designed for their different sizes. But another common allometry, the 0.2 to 0.4 intraspecific scaling among homeothermic adults differing in body size, or among races within a species, probably does not require a selectionist story though many, including one of us, have tried to provide one (Gould 1974). R. Lande (personal communication) has used the experiments of Falconer (1973) to show that selection upon *body size alone* yields a brain–body slope across generations of 0.35 in mice.

More compelling examples abound in the literature on selection for altering the timing of maturation (Gould 1977). At least three times in the evolution of arthropods (mites, flies and beetles), the same complex adaptation has evolved, apparently for rapid turnover of generations in strongly r-selected feeders on superabundant but ephemeral fungal resources: females reproduce as larvae and grow the next generation within their bodies. Offspring eat their mother from inside and emerge from her hollow shell, only to be devoured a few days later by their own progeny. It would be foolish to seek adaptive significance in paedomorphic morphology *per se*; it is primarily a by-product of selection for rapid cycling of generations. In more interesting cases, selection for small size (as in animals of the interstitial fauna) or rapid maturation (dwarf males of many crustaceans) has occurred by progenesis (Gould 1977, pp. 324–336), and descendant adults contain a mixture of ancestral juvenile and adult features. Many biologists have been tempted to find primary adaptive meaning for the mixture, but it probably arises as a by-product of truncated maturation, leaving some features 'behind' in the larval state, while allowing others, more strongly correlated with sexual maturation, to retain the adult configuration of ancestors.

(3) The decoupling of selection and adaptation.

(i) Selection without adaptation. Lewontin (1979) has presented the following hypothetical example: 'A mutation which doubles the fecundity of individuals will sweep through a population rapidly. If there has been no change in efficiency of resource utilization, the individuals will leave no more offspring than before, but simply lay twice as many eggs, the excess dying because of resource limitation. In what sense are the individuals or the population as a whole better adapted than before? Indeed, if a predator on immature stages is led to switch to the species now that immatures are more plentiful, the population size may actually decrease as a consequence, yet natural selection at all times will favour individuals with higher fecundity.'

(ii) Adaptation without selection. Many sedentary marine organisms, sponges and corals in particular, are well adapted to the flow régimes in which they live. A wide spectrum of 'good design' may be purely phenotypic in origin, largely induced by the current itself. (We may be sure of this in numerous cases, when genetically identical individuals of a colony assume different shapes in different microhabitats.) Larger patterns of geographic variation are often adaptive and purely phenotypic as well. Sweeney & Vannote (1978), for example, showed that many hemimetabolous aquatic insects reach smaller adult size with reduced fecundity when they grow at temperatures above and below their optima. Coherent, climatically correlated patterns in geographic distribution for these insects – so often taken as *a priori* signs of genetic adaptation – may simply reflect this phenotypic plasticity.

'Adaptation' – the good fit of organisms to their environment – can occur at three hierarchical levels with different causes. It is unfortunate that our language has focused on the common result and called all three phenomena 'adaptation': the differences in process have been obscured and evolutionists have often been misled to extend the Darwinian mode to the other two levels as well. First, we have what physiologists call 'adaptation': the phenotypic plasticity that permits organisms to mould their form to prevailing circumstances during ontogeny. Human 'adaptations' to high altitude fall into this category (while others, like resistance of sickling heterozygotes to malaria, are genetic and Darwinian). Physiological adaptations are not heritable, though the capacity to develop them presumably is. Secondly, we have a 'heritable' form of non-Darwinian adaptation in humans (and, in rudimentary ways, in a few other advanced social species):

cultural adaptation (with heritability imposed by learning). Much confused thinking in human sociobiology arises from a failure to distinguish this mode from Darwinian adaptation based on genetic variation. Finally, we have adaptation arising from the conventional Darwinian mechanism of selection upon genetic variation. The mere existence of a good fit between organism and environment is insufficient evidence for inferring the action of natural selection.

(4) Adaptation and selection but no selective basis for differences among adaptations. Species of related organisms, or subpopulations within a species, often develop different adaptations as solutions to the same problem. When 'multiple adaptive peaks' are occupied, we usually have no basis for asserting that one solution is better than another. The solution followed in any spot is a result of history; the first steps went in one direction, though others would have led to adequate prosperity as well. Every naturalist has his favourite illustration. In the West Indian land snail *Cerion*, for example, populations living on rocky and windy coasts almost always develop white, thick and relatively squat shells for conventional adaptive reasons. We can identify at least two different developmental pathways to whiteness from the mottling of early whorls in all *Cerion*, two paths to thickened shells and three styles of allometry leading to squat shells. All 12 combinations can be identified in Bahamian populations, but would it be fruitful to ask why – in the sense of optimal design rather than historical contingency – *Cerion* from eastern Long Island evolved one solution, and *Cerion* from Acklins Island another?

(5) Adaptation and selection, but the adaptation is a secondary utilization of parts present for reasons of architecture, development or history. We have already discussed this neglected subject in the first section on spandrels, spaces and cannibalism. If blushing turns out to be an adaptation affected by sexual selection in humans, it will not help us to understand why blood is red. The immediate utility of an organic structure often says nothing at all about the reason for its being.

6. ANOTHER, AND UNFAIRLY MALIGNED, APPROACH TO EVOLUTION

In continental Europe, evolutionists have never been much attracted to the Anglo-American penchant for atomizing organisms into parts and trying to explain each as a direct adaptation. Their general alternative exists in both a strong and a weak form. In the strong form, as advocated by such major theorists as Schindewolf (1950), Remane (1971), and Grassé (1977), natural selection under the adaptationist programme can explain superficial modifications of the *Bauplan* that fit structure to environment: why moles are blind, giraffes have long necks, and ducks webbed feet, for example. But the important steps of evolution, the construction of the *Bauplan* itself and the transition between *Baupläne*, must involve some other unknown, and perhaps 'internal', mechanism. We believe that English biologists have been right in rejecting this strong form as close to an appeal to mysticism.

But the argument has a weaker – and paradoxically powerful – form that has not been appreciated, but deserves to be. It also acknowledges conventional selection for superficial modifications of the *Bauplan*. It also denies that the adaptationist programme (atomization plus optimizing selection on parts) can do much to explain *Baupläne* and the transitions between them. But it does not therefore resort to a fundamentally unknown process. It holds instead that the basic body plans of organisms are so integrated and so replete with constraints upon adaptation (categories 2 and 5 of our typology) that conventional styles of selective arguments can explain little of interest about them. It does not deny that change, when it occurs, may be mediated by natural selection, but it holds that constraints restrict possible paths and modes of change so strongly that the constraints themselves become much the most interesting aspect of evolution.

Rupert Riedl, the Austrian zoologist who has tried to develop this thesis for English audiences (1977 and 1975, now being translated into English by R. Jefferies), writes:

> 'The living world happens to be crowded by universal patterns of organization which, most obviously, find no direct explanation through environmental conditions or adaptive radiation, but exist primarily through universal requirements which can only be expected under the systems conditions of complex organization itself... This is not self-evident, for the whole of the huge and profound thought collected in the field of morphology, from Goethe to Remane, has virtually been cut off from modern biology. It is not taught in most American universities. Even the teachers who could teach it have disappeared.'

Constraints upon evolutionary change may be ordered into at least two categories. All evolutionists are familiar with *phyletic* constraints, as embodied in Gregory's classic distinction (1936) between habitus and heritage. We acknowledge a kind of phyletic inertia in recognizing, for example, that humans are not optimally designed for upright posture because so much of our *Bauplan* evolved for quadrupedal life. We also invoke phyletic constraint in explaining why no molluscs fly in air and no insects are as large as elephants.

Developmental constraints, a subcategory of phyletic restrictions, may hold the most powerful rein of all over possible evolutionary pathways. In complex organisms, early stages of ontogeny are remarkably refractory to evolutionary change, presumably because the differentiation of organ systems and their integration into a functioning body is such a delicate process, so easily derailed by early errors with accumulating effects. Von Baer's fundamental embryological laws (1828) represent little more than a recognition that early stages are both highly conservative and strongly restrictive of later development. Haeckel's biogenetic law, the primary subject of late nineteenth century evolutionary biology, rested upon a misreading of the same data (Gould 1977). If development occurs in integrated packages, and cannot be pulled apart piece by piece in evolution, then the adaptationist programme cannot explain the alteration of developmental programmes underlying nearly all changes of *Bauplan*.

The German palaeontologist A. Seilacher, whose work deserves far more attention than it has received, has emphasized what he calls '*bautechnischer*', or *architectural*, constraints (Seilacher 1970). These arise not from former adaptations retained in a new ecological setting (phyletic constraints as usually understood), but as architectural restrictions that never were adaptations, but rather the necessary consequences of materials and designs selected to build basic *Baupläne*. We devoted the first section of this paper to non-biological examples in this category. Spandrels must exist once a blueprint specifies that a dome shall rest on rounded arches. Architectural constraints can exert a far-ranging influence upon organisms as well. The subject is full of potential insight because it has rarely been acknowledged at all.

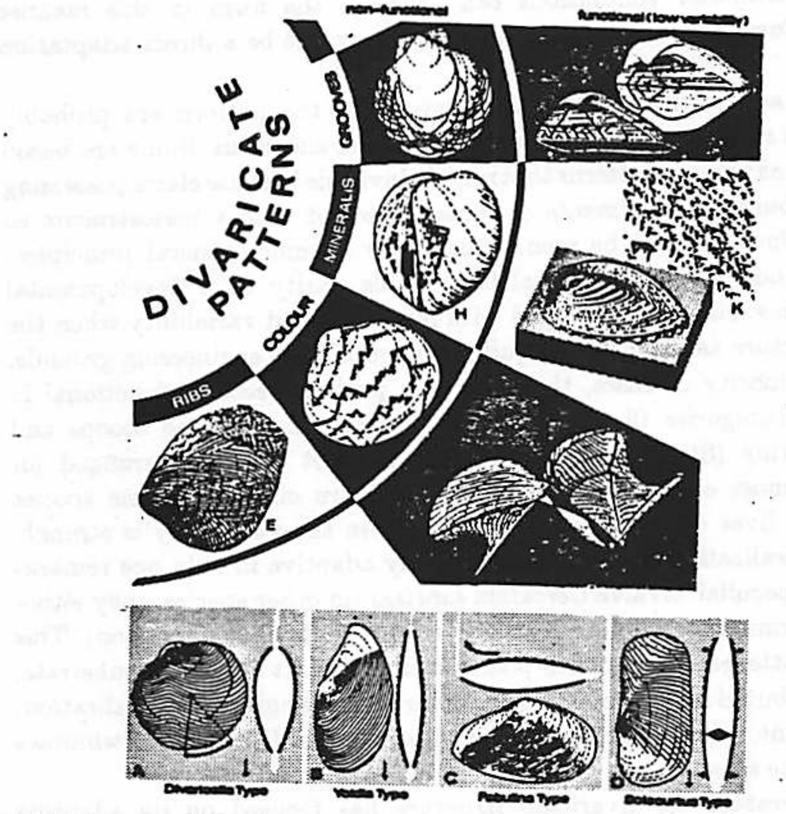


FIGURE 3. The range of divaricate patterns in molluscs. E, F, H, and L are non-functional in Seilacher's judgement. A–D are functional ribs (but these are far less common than non-functional ribs of the form E). G is the mimetic *Arca zebra*. K is *Corculum*. See text for details.

In a fascinating example, Seilacher (1972) has shown that the divaricate form of architecture (figure 3) occurs again and again in all groups of molluscs, and in brachiopods as well. This basic form expresses itself in a wide variety of structures: raised ornamental lines (not growth lines because they do not conform to the mantle margin at any time), patterns of coloration, internal structures in the mineralization of calcite, and incised grooves. He does not know what generates this pattern and feels that traditional and nearly exclusive focus on the adaptive value of each manifestation has diverted attention from questions of its genesis in growth and also prevented its recognition as a general phenomenon. It must arise from some characteristic pattern of inhomogeneity in the growing mantle, probably from the generation of interference patterns around regularly spaced centres; simple computer simulations can generate the form in this manner (Waddington & Cowe 1969). The general pattern may not be a direct adaptation at all.

Seilacher then argues that most manifestations of the pattern are probably non-adaptive. His reasons vary, but seem generally sound to us. Some are based on field observations: colour patterns that remain invisible because clams possessing them either live buried in sediments or remain covered with a periostracum so thick that the colours cannot be seen. Others rely on more general principles: presence only in odd and pathological individuals, rarity as a developmental anomaly, excessive variability compared with much reduced variability when the same general structure assumes a form judged functional on engineering grounds.

In a distinct minority of cases, the divaricate pattern becomes functional in each of the four categories (figure 3). Divaricate ribs may act as scoops and anchors in burrowing (Stanley 1970), but they are not properly arranged for such function in most clams. The colour chevrons are mimetic in one species (*Pteria zebra*) that lives on hydrozoan branches; here the variability is strongly reduced. The mineralization chevrons are probably adaptive in only one remarkable creature, the peculiar bivalve *Corculum cardissa* (in other species, they either appear in odd specimens or only as post-mortem products of shell erosion). This clam is uniquely flattened in an anterio-posterior direction. It lies on the substrate, posterior up. Distributed over its rear end are divaricate triangles of mineralization. They are translucent, while the rest of the shell is opaque. Under these windows dwell endosymbiotic algae!

All previous literature on divaricate structure has focused on its adaptive significance (and failed to find any in most cases). But Seilacher is probably right in representing this case as the spandrels, ceiling holes and sacrificed bodies of our first section. The divaricate pattern is a fundamental architectural constraint. Occasionally, since it is there, it is used to beneficial effect. But we cannot understand the pattern or its evolutionary meaning by viewing these infrequent and secondary adaptations as a reason for the pattern itself.

Galton (1909, p. 257) contrasted the adaptationist programme with a focus on constraints and modes of development by citing a telling anecdote about Herbert Spencer's fingerprints:

> 'Much has been written, but the last word has not been said, on the rationale of these curious papillary ridges; why in one man and in one finger they form whorls and in another loops. I may mention a characteristic anecdote of Herbert

> Spencer in connection with this. He asked me to show him my Laboratory and to take his prints, which I did. Then I spoke of the failure to discover the origin of these patterns, and how the fingers of unborn children had been dissected to ascertain their earliest stages, and so forth. Spencer remarked that this was beginning in the wrong way; that I ought to consider the purpose the ridges had to fulfil, and to work backwards. Here, he said, it was obvious that the delicate mouths of the sudorific glands required the protection given to them by the ridges on either side of them, and therefrom he elaborated a consistent and ingenious hypothesis at great length. I replied that his arguments were beautiful and deserved to be true, but it happened that the mouths of the ducts did not run in the valleys between the crests, but along the crests of the ridges themselves.

We feel that the potential rewards of abandoning exclusive focus on the adaptationist programme are very great indeed. We do not offer a council of despair, as adaptationists have charged; for non-adaptive does not mean non-intelligible. We welcome the richness that a pluralistic approach, so akin to Darwin's spirit, can provide. Under the adaptationist programme, the great historic themes of developmental morphology and *Bauplan* were largely abandoned; for if selection can break any correlation and optimize parts separately, then an organism's integration counts for little. Too often, the adaptationist programme gave us an evolutionary biology of parts and genes, but not of organisms. It assumed that all transitions could occur step by step and underrated the importance of integrated developmental blocks and pervasive constraints of history and architecture. A pluralistic view could put organisms, with all their recalcitrant, yet intelligible, complexity, back into evolutionary theory.

References (Gould & Lewontin)

Baer, K. E. von 1828 *Entwicklungsgeschichte der Tiere*. Königsberg: Bornträger.

Barash, D. P. 1976 Male response to apparent female adultery in the mountain bluebird: an evolutionary interpretation. *Am. Nat.* 110, 1097–1101.

Coon, C. S., Garn, S. M. & Birdsell, J. B. 1950 *Races*. Springfield, Ohio: C. Thomas.

Costa, R. & Bisol, P. M. 1978 Genetic variability in deep-sea organisms. *Biol. Bull.* 155, 125–133.

Darwin, C. 1872 *The origin of species*. London: John Murray.

Darwin, C. 1880 Sir Wyville Thomson and natural selection. *Nature, Lond.* 23, 32.

Davitashvili, L. S. 1961 *Teoriya polovogo otbora* [Theory of sexual selection]. Moscow: Akademii Nauk.

Falconer, D. S. 1973 Replicated selection for body weight in mice. *Genet. Res.* 22, 291–321.

Galton, F. 1909 *Memories of my life*. London: Methuen.

Gould, S. J. 1966 Allometry and size in ontogeny and phylogeny. *Biol. Rev.* 41, 587–640.

Gould, S. J. 1971 D'Arcy Thompson and the science of form. *New Literary Hist.* 2(2), 229–258.

Gould, S. J. 1974 Allometry in primates, with emphasis on scaling and the evolution of the brain. In *Approaches to primate paleobiology. Contrib. Primatol.* 5, 244–292.

Gould, S. J. 1977 *Ontogeny and phylogeny*. Cambridge, Mass.: Belknap Press.

Gould, S. J. 1978 Sociobiology: the art of storytelling. *New Scient.* 80, 530–533.

Grassé, P.-P. 1977 *Evolution of living organisms.* New York: Academic Press.

Gregory, W. K. 1936 Habitus factors in the skeleton of fossil and recent mammals. *Proc. Am. phil. Soc.* 76, 429–444.

Harner, M. 1977 The ecological basis for Aztec sacrifice. *Am. Ethnologist* 4, 117–135.

Jerison, H. J. 1973 *Evolution of the brain and intelligence.* New York: Academic Press.

Lande, R. 1976 Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30, 314–334.

Lande, R. 1978 Evolutionary mechanisms of limb loss in tetrapods. *Evolution* 32, 73–92.

Lewontin, R. C. 1978 Adaptation. *Scient. Am.* 239 (3), 156–169.

Lewontin, R. C. 1979 Sociobiology as an adaptationist program. *Behav. Sci.* (In the press.)

Morton, E. S., Geitgey, M. S. & McGrath, S. 1978 On bluebird 'responses to apparent female adultery'. *Am. Nat.* 112, 968–971.

Ortiz de Montellano, B. R. 1978 Aztec cannibalism: an ecological necessity? *Science N.Y.* 200, 611–617.

Remane, A. 1971 *Die Grundlagen des natürlichen Systems der vergleichenden Anatomie und der Phylogenetik.* Königstein-Taunus: Koeltz.

Rensch, B. 1959 *Evolution above the species level.* New York: Columbia University Press.

Riedl, R. 1975 *Die Ordnung des Lebendigen.* Hamburg: Paul Parey.

Riedl, R. 1977 A systems-analytical approach to macro-evolutionary phenomena. *Q. Rev. Biol.* 52, 351–370.

Romanes, G. J. 1900 The Darwinism of Darwin and of the post-Darwinian schools. In *Darwin, and after Darwin*, vol. 2, new edn. London: Longmans, Green & Co.

Rudwick, M. J. S. 1964 The function of zig-zag deflections in the commissures of fossil brachiopods. *Palaeontology* 7, 135–171.

Sahlins, M. 1978 Culture as protein and profit. *New York review of books*, 23 Nov., pp. 45–53.

Schindewolf, O. H. 1950 *Grundfragen der Paläontologie.* Stuttgart: Schweizerbart.

Seilacher, A. 1970 Arbeitskonzept zur Konstruktionsmorphologie. *Lethaia* 3, 393–396.

Seilacher, A. 1972 Divaricate patterns in pelecypod shells. *Lethaia* 5, 325–343.

Shea, B. T. 1977 Eskimo craniofacial morphology, cold stress and the maxillary sinus. *Am. J. phys. Anthrop.* 47, 289–300.

Stanley, S. M. 1970 Relation of shell form to life habits in the Bivalvia (Mollusca). *Mem. geol. Soc. Am.* no. 125, 296 pp.

Sweeney, B. W. & Vannote, R. L. 1978 Size variation and the distribution of hemimetabolous aquatic insects: two thermal equilibrium hypotheses. *Science, N.Y.* 200, 444–446.

Thompson, D. W. 1942 *Growth and form.* New York: Macmillan.

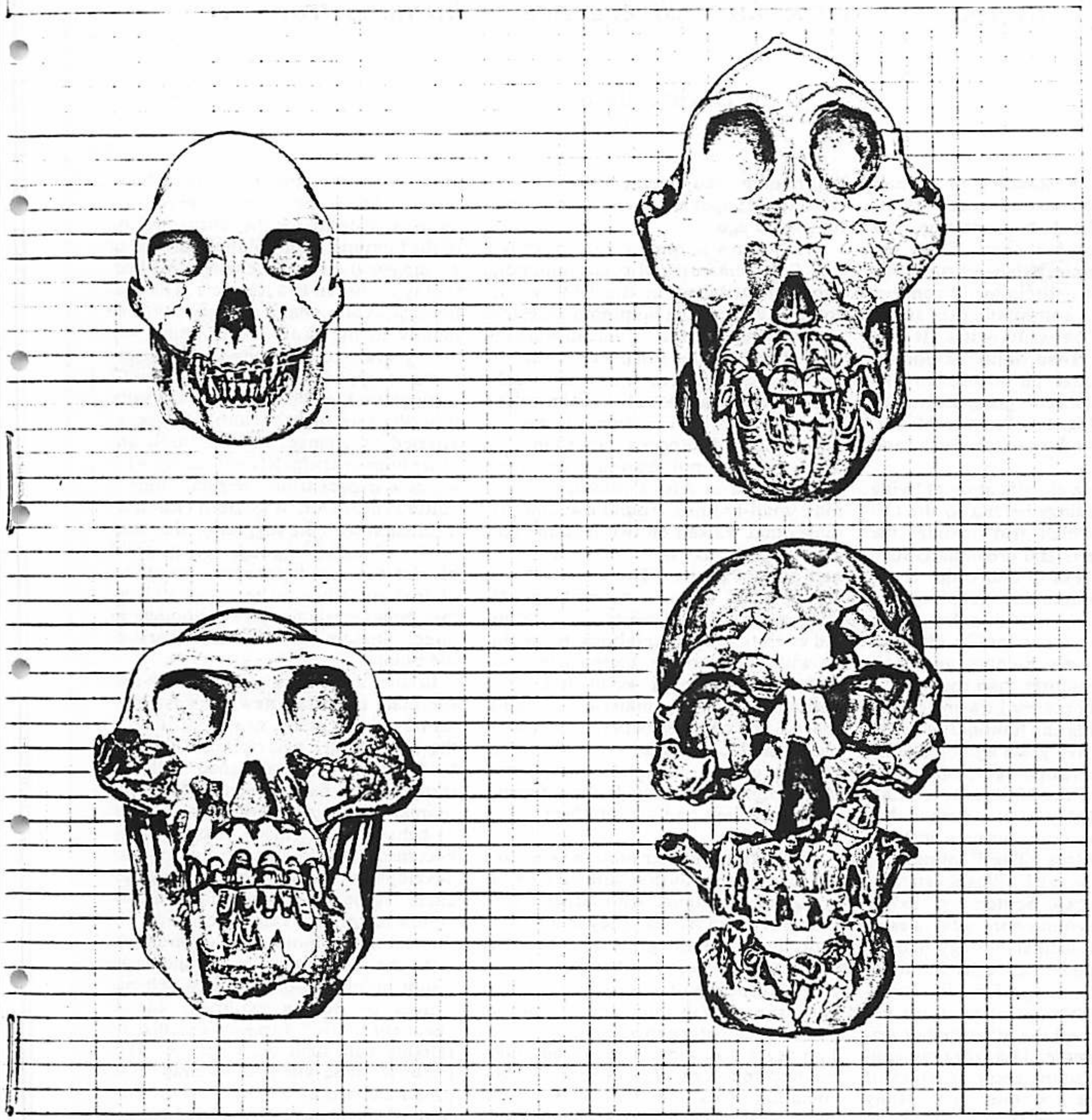
Waddington, C. H. & Cowe, J. R. 1969 Computer simulation of a molluscan pigmentation pattern. *J. theor. Biol.* 25, 219–225.

Wallace, A. R. 1899 *Darwinism.* London: Macmillan.

Wilson, E. O. 1978 *On human nature.* Cambridge, Mass.: Harvard University Press.

SCIENTIFIC AMERICAN

Pilbeam, D. "The Descent of Hominoids and Hominids"
Scientific American
W.H. Freeman and Company
Used by permission



DESCENT OF HOMINOIDS AND HOMINIDS

$2.50

March 1984

The Descent of Hominoids and Hominids

Findings over the past five years have made it possible to trace with greater assurance the divergence of the apes from the Old World monkeys and the later divergence of humans from apes

by David Pilbeam

Human understanding of human evolution has advanced greatly in the past generation and even in the past five years. Since the 1960's, for example, paleoanthropology has ceased being a discipline of concern only to physical anthropologists and a few interested paleontologists. It has evolved into a broad multidisciplinary pursuit that enlists the services of historical geologists, ecologists, molecular biologists, zoologists, students of animal behavior and even chemists and physicists.

In the Linnaean classification of living things human beings belong to the Order Primates, which also includes the apes, the monkeys and prosimians such as the lemurs. Within that order is the superfamily Hominoidea, including only humans and the apes. Within that superfamily, in turn, is the family Hominidae, including only humans and their extinct relatives closer than the apes.

Let me briefly review how the evolution of hominoids and hominids was understood some five years ago by a leading physical anthropologist, Sherwood L. Washburn of the University of California at Berkeley, and contrast that summary of the consensus then with the consensus now [see "The Evolution of Man," by Sherwood L. Washburn; SCIENTIFIC AMERICAN, September, 1978]. At that time revolutionary advances in geology and molecular biology had already challenged earlier ideas about the geography and timing of hominid evolution and the biological relations between humans and the other higher primates. For example, data from molecular biology made the reconstruction of primate descent a simpler task. Many physical anthropologists were skeptical, however, of the molecular biologists' suggestion that African apes (chimpanzees and gorillas) were closely related to humans but Asian apes (orangutans and gibbons) were not. The biggest surprise was the molecular biologists' proposal that humans had diverged from the African line of apes much less than 10 million years ago.

Washburn's summary also suggested that *Ramapithecus*, a Miocene hominoid widely distributed in the Old World, could have been the hominoid ancestral to all living great apes and to humankind. Further, the discovery in Ethiopia of *Australopithecus afarensis* (widely known from the partial skeleton called "Lucy") and in Tanzania of the famous Laetoli footprints showed that by more than 3.5 million years ago there had evolved in Africa primitive hominids: small-brained, ground-dwelling primates that walked on two legs and had canine teeth that were relatively small and did not project. The discoveries in Ethiopia and Tanzania showed bipedalism was the first and most basic of hominid evolutionary adaptations. By analogy with chimpanzees, these early hominids probably used wood and other naturally occurring materials as tools. Indeed, the contrast between the hominids' small canines and the projecting canines of the chimpanzee suggested that among the hominids the offensive role of canines had already been taken over by tools.

By the time another million years had passed, in Washburn's summary of the consensus, hominids with larger brains had evolved in Africa. They made stone tools and lived alongside a robust (that is, relatively heavy-boned and large-toothed) species of *Australopithecus* that became extinct in about another million years. Even before then, say by 1.5 million years ago, a more advanced hominid had appeared: *Homo erectus*. The association of its stone tools with animal bones was accepted as clear evidence that hunting behavior had become a hominid attribute long before the hominid brain expanded to its modern size.

Brain size and tool complexity increased only slowly between then and about 100,000 years ago, when a still more advanced hominid arose: *Homo sapiens*. These increases, however, are indirect evidence for the reorganization of the hominid brain. With the arrival of *H. sapiens* the rate of change in stone-tool types began to accelerate. The anatomical evolutionary changes are seen mainly in the skull. Then, about 40,000 years ago, anatomically modern humans appeared and spread rapidly throughout the Old World. At the same time the rate of behavioral change, as attested by changes in stone tools and other human artifacts, approached what we now expect from "normal" human cultural behavior. Washburn closed his discussion with the suggestion that fully modern human language, which possibly did not exist before the appearance of anatomically modern man, catalyzed this behavioral advance, although no direct anatomical evidence bears on the point.

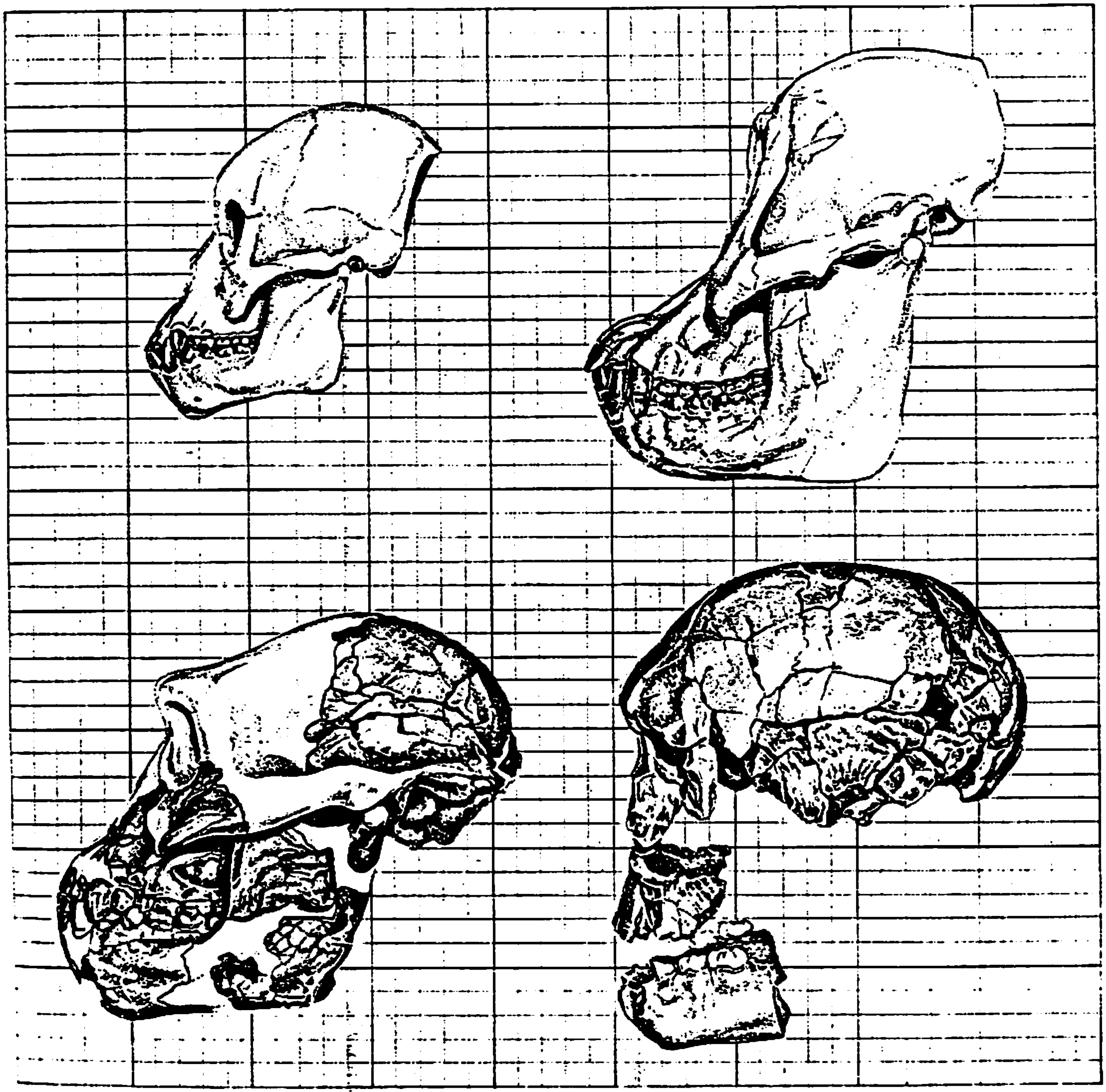
In the five years since Washburn's summary important new fossil discoveries have been made, not all of them in the field. Fuller descriptions of earlier fossils have been published. Paleoanthropologists have become better able to interpret their data through the work of behavioral scientists, and they have become increasingly aware of relevant developments in the other natural sciences. These trends have shifted the discipline toward a greater integration with the rest of the biological community.

At the same time there is increasing doubt about the degree to which narratives of human evolution can be "brought to life." Just how detailed and reliable can such accounts be? Hard problems arise when the hominids of the past are interpreted in terms of the hominids of the present. The early hominids were, after all, markedly different from any living today. In many instances, however, these differences have been ignored and early hominids have been made to seem too much like modern human beings. Let me review some of these problems, starting at the beginning with the evolution of hominoids, the group from which the hominids sprang.

Hominoid Evolution

The earlier debate between physical anthropologists and molecular biologists over the pattern and timing of hominoid evolution is now basically settled. Most paleontologists (and comparative anatomists) agree the molecular patterns showing the African apes are genetically very little different from humans and the Asian apes are about twice as different reflect the fact that the common ancestor hominids share with the chimpanzee and the gorilla was in existence only about half as long ago as the last common ancestor of all the larger hominoids.

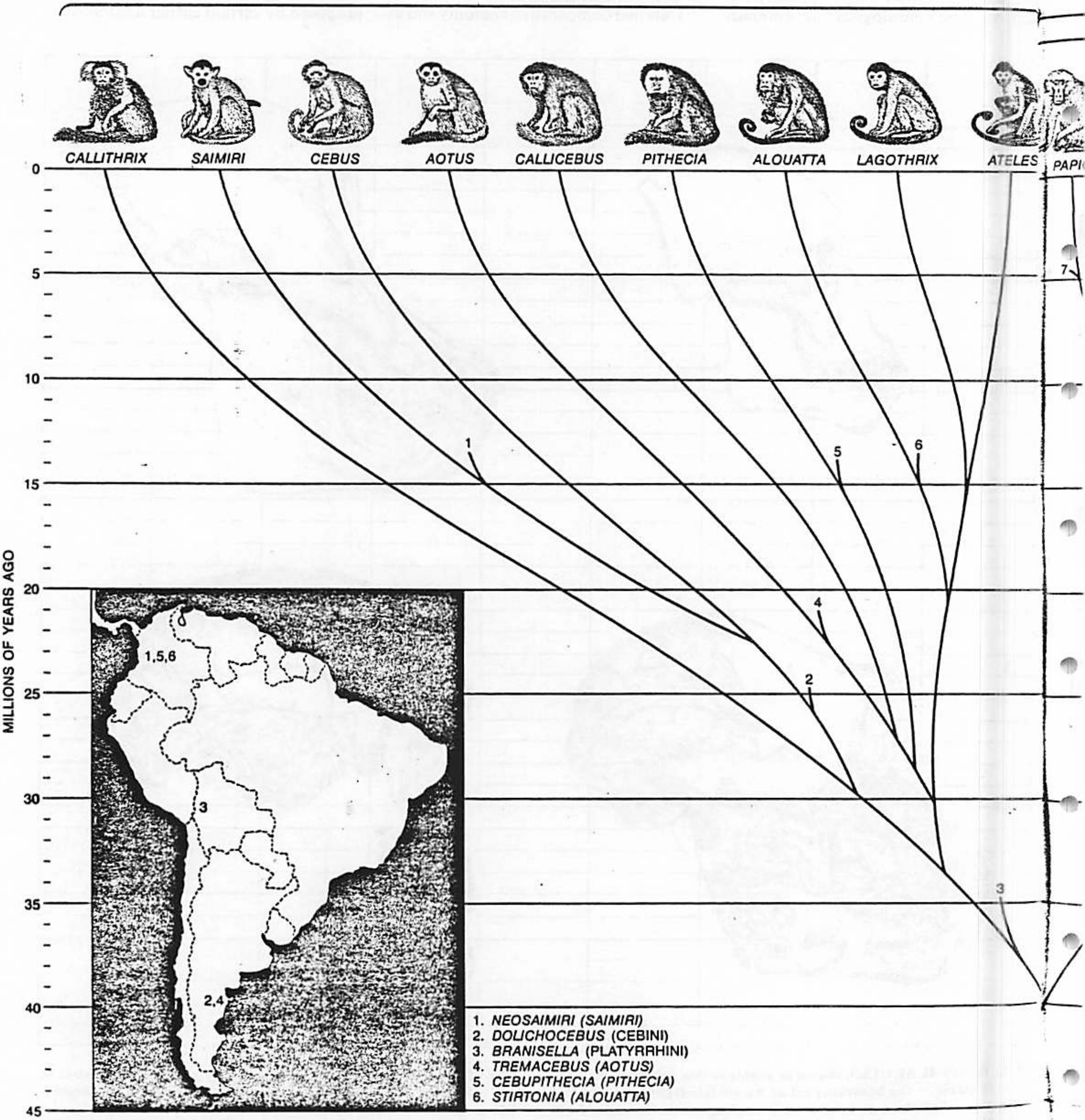
Detailed comparative anatomy shows that the patterns of morphological resemblance among living hominoids confirm the molecular findings. The great apes and the hominids do not form what is called a monophyletic group, that is, they did not share a last common ancestor after the lineage leading to humans diverged. Instead African apes and humans are monophyletic with respect to the Asian apes. (That had in fact been proposed by certain earlier anatomists.)



FOUR FOSSIL SKULLS, shown in profile in this painting, exemplify key stages in the branching off of hominoids from ancestral Old World monkey stock and the later branching off of hominids from ancestral hominoid stock. At the top left is the best-known of the early Miocene apes, *Proconsul africanus*, a baboon-size primate that is perhaps similar to the common ancestor of all apes and humans. At the top right is the substantially larger late Miocene ape *Sivapithecus*, first known from fossil teeth and jaw fragments found in foothills of the Himalayas in the 1900's. It may have been ancestral to the only living large hominoid of Asia, the orangutan. At the bottom left is the earliest-known African Pliocene hominid, *Australopithecus afarensis*, a species that walked erect in the period between four and two million years ago. At the bottom right is the much larger-brained hominid *Homo habilis*, which was first found in Kenya in 1972. Members of the species ranged southern and eastern Africa for a few hundred thousand years beginning about two million years ago.

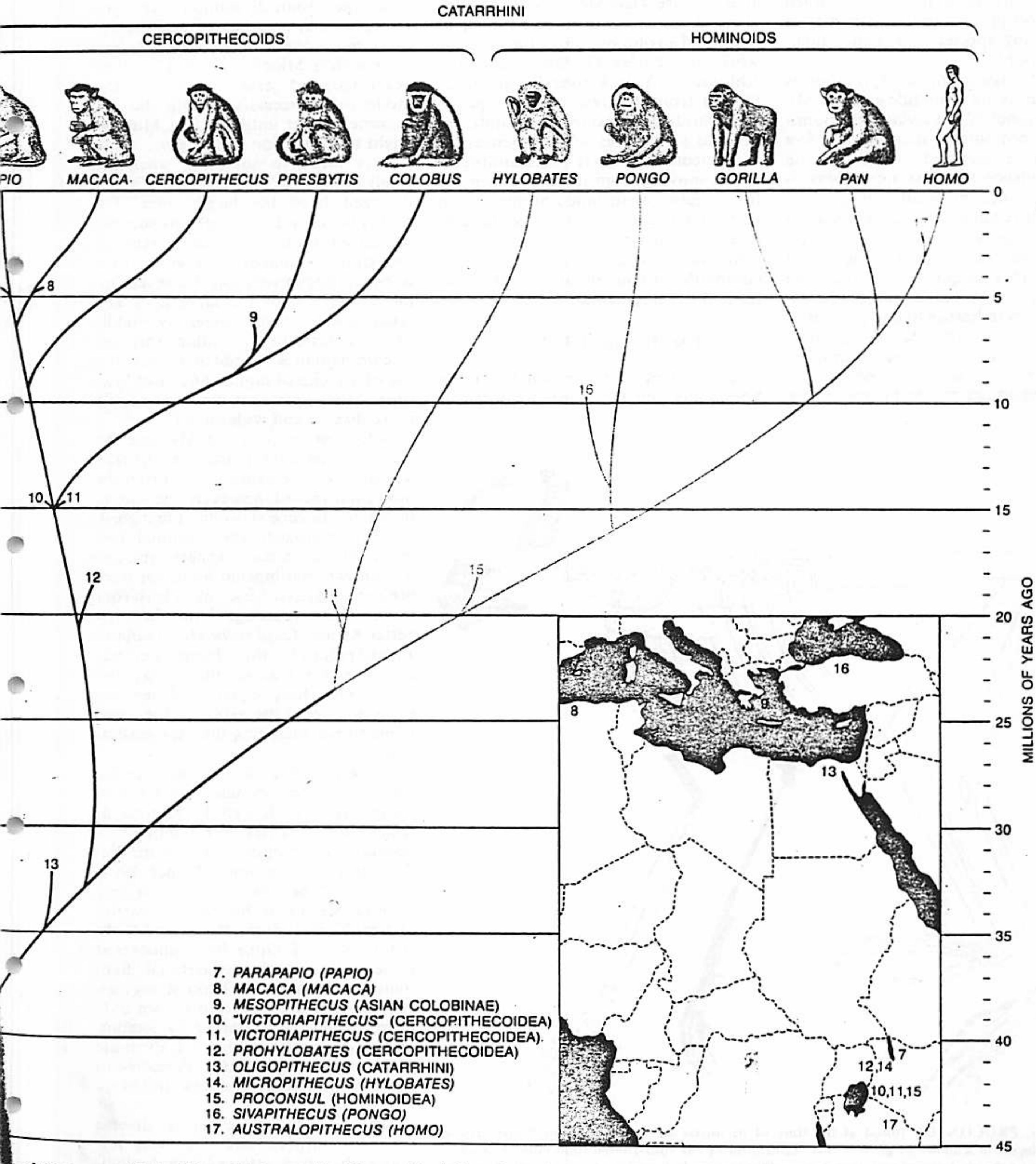
This means features common to all apes, such as the absence of a tail, must either have been retained from the last ancestor of all large hominoids or have evolved separately in the different apes. As a consequence of the molecular findings morphologists are reexamining hominoid anatomical characteristics and similarities and looking more closely at the processes of embryonic development and their genetic controls.

These "new" molecular branching patterns are not incompatible with the hominoid fossil record. The time is past when many paleontologists placed the origin of the hominids in the middle Miocene, 16 or 15 million years ago, or when some molecular biologists held that the entire hominoid radiation took place within the past 10 million years, with the hominids diverging only a brief



BRANCHING SEQUENCE of the higher primates from 45 million years ago to the present is shown under two major headings, the Platyrrhini, or the New World monkeys, at the left and the Catarrhini, or the Old World monkeys, at the right. The hominoid and hominid branches are in color at the far right. The sequences are based on such biochemical data as small differences in the amino acid sequence four million years ago. A tentative consensus would now place the radiation of all the great apes from an ancestral stock in the middle Miocene, with the hominids splitting from an African ape lineage in the late Miocene, perhaps eight or seven million years ago.

Why was the hominoid fossil record misinterpreted, at least by dimmer paleontologists such as me? There are a number of reasons. First, far too much attention was being paid to the fossil record as a source of information about evolutionary branching sequences. It is now clear that the molecular record can tell more about hominoid branching patterns than the fossil record does. (The fossil record remains, however, the only source of direct knowledge of past evolutionary states.) Misinterpretation also resulted from the fact that the hom-



of the same protein in different living species. The specific dating of the branches is based on the age of fossils and on other kinds of geochronological data. The maps at left and right indicate where the earliest fossil representative of the genus was found. The name to the right of each fossil genus indicates its relation to living forms. Thus *Sivapithecus* pairs with *Pongo* but *Proconsul* pairs with all hominoids.

inoid fossil record going back before the last four million years of hominid evolution has until recently been both fragmentary and poorly dated. It consisted almost entirely of teeth and small pieces of jaw, which are parts of the body that are seldom very informative on questions of taxonomy. For example, the equivalent parts of the body in living mammals are generally not of much help in sorting individuals into species or clustering species into higher taxonomic groups.

A third cause of misinterpretation is the very diversity of middle and late Miocene hominoids. The living large hominoids are not notably diverse: only five genera have survived. When all the fossil evidence of past hominoids is taken into account, what emerges is a group that is quite heterogeneous; each hominoid species is in many respects unique. These factors have made it hard to reconstruct by means of fossils the branching sequences of living hominoids. It is even harder to link particular fossil species to particular living ones.

Significant new Miocene hominoid fossils, however, have changed the earlier state of affairs. Skeletal parts—faces, brain cases and limbs—that were misidentified or have only now been uncovered are yielding a clearer and often surprising picture of both evolutionary relations and the nature of the early hominoids. One can even catch glimpses of how they may have lived and what their ecological relations were.

At the same time newer molecular analyses are clarifying and expanding ideas about the pattern and timing of hominoid evolution; an example is the work of Charles G. Sibley and Jon Ahlquist of Yale University, in which DNA's from different primate species are hybridized to establish the differences and similarities of their sequences of nucleotides. Slowly paleoanthropology is moving from the time when the fossil record contributed nothing to an understanding of phylogenetic patterns to a time when it is one component of a synthesis that renders data of all kinds compatible: fossil, anatomical, molecular and developmental.

The Hominoid Radiation

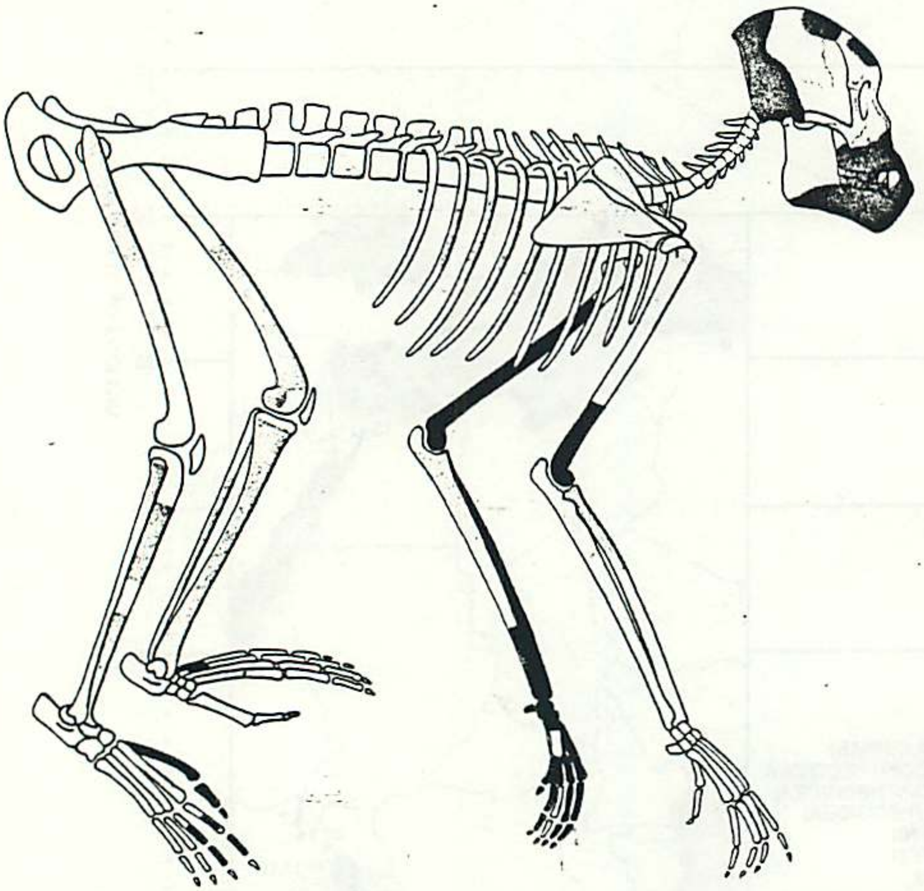
An account of the radiation of the hominoids that represents a consensus of current opinion can be summarized as follows. First, it is not clear when the great apes diverged from their relatives, the Old World monkeys. (The New World monkeys played no part in hominoid evolution.) It seems almost certain, however, that the divergence came after middle Oligocene times (some 30 million years ago) and before early Miocene times (some 20 million years ago). However that may be, the radiation of the hominoids was under way in the early or middle Miocene (20 to 15 million years ago) and certain lineages of this moderately successful group did not become extinct until the late Miocene (eight to six million years ago).

It is similarly not clear when the small hominoids, the gibbons of Asia, diverged from the larger ones. The split is possibly documented by some of the earlier Miocene primate fossils, although it is a controversial point. There is better documentation, however, for the divergence of the large apes into Asian and African species by middle Miocene times, say 16 million years ago. The orangutan is the odd and sole survivor of a group of middle and later Miocene Asian apes that were previously more diverse and widespread.

Sadly, for most of the Miocene the African hominoid branch is virtually without fossil representation. From the molecular record, however, one can infer a late Miocene splitting of first gorillas, then chimpanzees and hominids (between 10 and six million years ago), and it is known that bipedal hominids were present in eastern Africa by at least four to 3.5 million years ago. Until there is a better African fossil record (particularly a fossil record for the African apes) one can only guess about the timing and precise branching sequence of these lineages and about the nature of their early members, including the very earliest hominids.

If one goes back to Oligocene times, the African fossil primate record is comparatively rich. Elwyn L. Simons of Duke University has been hunting successfully for primate fossils in the Fayum of Egypt on and off since 1961. Since 1977 he and his collaborators John G. Fleagle of the State University of New York at Stony Brook and Richard F. Kay of Duke have uncovered some truly remarkable material. Substantial parts of the skeleton of *Aegyptopithecus*, a species hitherto known only from a skull found earlier by Simons, and *Apidium* are now in hand. Even the very sparsely represented *Parapithecus* and *Propliopithecus* are now relatively well characterized.

These fossils represent a diverse group of unspecialized tree-living primates that had attained the biological status of monkeys at least as early as middle Oligocene times, 30 million



PARTS OF PROCONSUL found at the time of its initial discovery by Tom Whitworth in 1951 are shown in a shade of gray in this restoration of the four-footed hominoid. The additional parts of the skeleton shown in color are those recently found in museum collections by Alan Walker and Martin Pickford, who prepared this restoration. The parts shown in outline only, such as the pelvis, most of the spine and parts of the skull and jaw, have not been found.

years ago. They were not, however, primitive primates like the lemurs but bore a general resemblance to the living New World monkeys. They are nonetheless too primitive and too different from the living higher primates of the Old World for one to know whether they are representative of a time before the divergence of the hominoids from the Old World monkeys or a time after the divergence.

The early Miocene of eastern Africa is a happy exception to the fossil-poor record for Africa as a whole. Collecting over the past 50 years in Kenya and Uganda has yielded some 1,000 fossils of higher primates that are between 17 and 22 million years old and that almost certainly represent hominoid lineages of a period after their divergence from the Old World monkeys. Mainly through the work of Peter Andrews of the British Museum (Natural History) the paleoecology of the period is now well known. The landscape was a mixture of forest, open woodland and grassland, with the forest predominating. The forest and woodland were inhabited by at least six hominoid species. One of them, *Proconsul africanus*, has for some time been the best-known Miocene ape, and it has now become even better known.

A small animal the size of a baboon, *Proconsul* has played a central role in thinking about hominoid evolution. The species had been represented until recently only by an almost complete forelimb, a jaw, skull fragments and a few foot bones, all of one individual, an almost complete skull of another individual and other limb fragments, teeth and jaws. Then Alan C. Walker of Johns Hopkins University and Martin Pickford of the National Museums of Kenya discovered—in museum and laboratory collections—many more pieces of the "forelimb" individual, so that it is now represented by much of its skeleton.

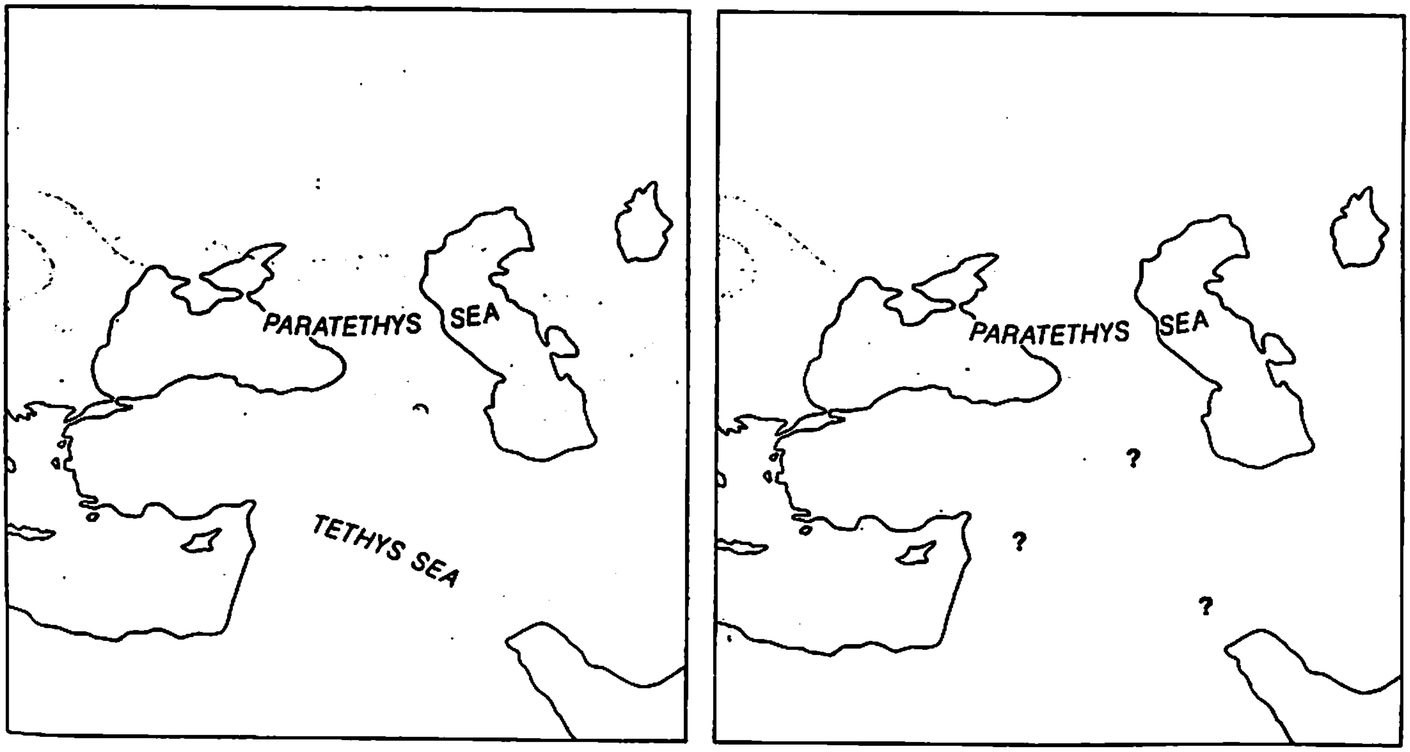
Until now *Proconsul* had been interpreted in several contradictory ways. At one extreme was the view held by some that its form was too specialized for it to be the ancestor of any living hominid and that it was possibly not even a hominoid. Others saw it as an unspecialized early hominoid and still others accepted it as being ancestral to the chimpanzee. The addition of the new material clarifies the issue to some extent. *Proconsul*, the consensus now holds, is a hominoid, albeit a very unspecialized and primitive one. It was a tree-dwelling, fruit-eating "formative ape" in which the sexes were distinctly different in size. In its total pattern it was unlike any living higher primate. Its elbow and shoulder joints and feet are like a chimpanzee's, its wrist is like a monkey's and its lumbar vertebrae are like a gibbon's. Many of its other features are unique, as is its total configuration. *Proconsul* provides a salutary lesson for students of evolution: the relations inferred for the animal have depended on what part of the body was being studied. When a fossil animal is found in fragments and over a period of time, the very order of discovery of its various parts will affect the phylogenetic interpretations, particularly in the case of a "mosaic" species such as *Proconsul*.

Dryomorphs and Ramamorphs

The end of the early Miocene, some 17 million years ago, was marked by a highly significant geologic event: continental drift linked Africa and Arabia with mainland Eurasia, thereby allowing the migration of African hominoids (and other mammals) into the rest of the Old World. The linkup initiated major mountain building: the Alps rose in Europe, the Taurus Mountains in Turkey and the Zagros Mountains in Iran. The Tethys Sea disappeared and oceanic and atmospheric circulation patterns shifted, bringing about large-scale changes in climate and habitats.

The Eurasian climate was seasonal in the middle and late Miocene, from 17 to five million years ago. The winters were milder, however, than they are today. Much of what is now grassland and even desert was then woodland, and the forests were more widespread. The mammalian fauna reflected these conditions: grassland grazers were scarce but woodland and forest browsers abundant.

The hominoid fossil record for this period in Eurasia consists mainly of jaws and teeth, so that the phylogenetic relations are still not clear. As an approximate summary, however, the hom-



JOINING OF AFRICA AND EURASIA followed the time 20 million years ago when the Tethys Sea (*color*) separated the two landmasses, as is shown at the left. In early Miocene times, some 17 million years ago, the sea shrank, as is shown at the right, leaving a land bridge that allowed animals to migrate between Africa and Eurasia. Among migrants to Eurasia were some of the African hominoids.

inoids can be divided into two broad groups. One group can informally be called the dryomorphs, after the various European species of the genus *Dryopithecus*, a primate with many primitive hominoid characteristics. The other can be called the ramamorphs, after *Ramapithecus* and similar genera. The ramamorphs were mainly Asian in distribution; they differ from other Miocene hominoids in certain characteristics that resemble those of later hominoids.

Edward Lewis, a Yale graduate student who was later with the U.S. Geological Survey, gave the name *Ramapithecus* to a fossilized fragment of a primate upper jaw from India he had found in 1932. The fossil is now known to be a little more than seven million years old. By the 1960's more primate jaws and teeth from India, Pakistan and other parts of the Old World had been found and assigned to the same genus. The animal was seen as being different from its supposedly more apelike contemporaries, *Dryopithecus* and another Indian fossil find, *Sivapithecus*. Its various remains showed that its jaws were robust and its cheek teeth large, with thick enamel caps. These features and a few others resemble those of the later African hominids of the genus *Australopithecus*. The resemblances led many, including me, to argue that *Ramapithecus* was in fact an early hominid, that the hominids had diverged from the hominoids of Africa at least 15 million years ago and therefore that the divergence date of four million years ago, based on the molecular data, was wildly incorrect.

In 1973 I began a collaborative research project with the Geological Survey of Pakistan aimed at finding additional *Ramapithecus* fossils. The fossil-rich formations where we are still at work are the product of the erosion of the rising Himalayas; they are called the Siwalik series. They vary in age from one million to 17 million years and form a thick wedge that runs along the southern flank of the great mountain range from Afghanistan to Burma. In the section of the Siwalik formation in Pakistan we have been fortunate in finding many new hominoid fossils, including specimens of both *Ramapithecus* and its relative *Sivapithecus*, and thousands of other mammalian fossils.

Our large interdisciplinary team has now built up a dated sequence of faunal changes over the 16 million years of deposition in the region and has made considerable progress in reconstructing its animal communities of the past. It is clear from the new primate material that *Ramapithecus* and *Sivapithecus* are much more alike than I had thought: they may even belong to the same genus.

In any event the new *Sivapithecus* fossils, which include a partial skull and isolated limb bones, show a number of features, particularly of the face and palate, that resemble those of the sole surviving Asian great ape, the orangutan (*Pongo pygmaeus*). They do not, as expected, resemble features of *Australopithecus*. The jaw and tooth resemblances to *Australopithecus* remain but their phylogenetic significance can be dismissed; they are probably either instances of parallel evolution or features retained from the last common ancestor of all living hominoids. At least this is my opinion and that of my colleague Steven Ward of Kent State University. Some of the hominoid fossils from Pakistan that are at least 12 million years old show similar features. If our interpretation is correct, the split between African and Asian hominoids is certainly that old and probably even older. Perhaps the split coincided approximately with the continental-drift linkup of Africa

MILLIONS OF YEARS AGO	EPOCH	1978 CONSENSUS	1984 CONSENSUS
	LATEST PLEISTOCENE	Anatomically modern *H. sapiens* appears.	Anatomically modern *H. sapiens* appears (markedly different from archaic *H. sapiens* both cranially and subcranially). Abundant stone and bone tools and other artifacts.
.045			
	LATE PLEISTOCENE	Little change from preceding phases	
.125		*H. sapiens* appears.	New tool types and techniques; significant elaboration of behavior.
	MIDDLE PLEISTOCENE	Gradual hominid evolutionary progression.	Stone tools somewhat more complex. Archaic *H. sapiens* appears (little different from *H. erectus*).
.7			Use of fire?
	EARLY PLEISTOCENE	*H. erectus* appears: robust *Australopithecus* extinct.	Stone tools somewhat more complex.
1.6			*H. erectus* present. First *H. habilis* and then *A. boisei* and *A. robustus* extinct.
	PLIOCENE	Gracile, bipedal hominids in Tanzania and Ethiopia (including *A. afarensis*).	Earliest stone tools; *A. africanus*, *A. robustus* present in South Africa. Hominids diversify; *A. boisei*, *H. habilis* coexist. Laetoli footprints. Bipedal *Australopithecus* present in East Africa.
5			
	LATE MIOCENE	African apes and African protohominids diverge.	7.5-4.5 million years ago: major climatic/faunal changes; African hominoids diversify (inferred from molecular data).
11			
	MIDDLE MIOCENE	*Ramapithecus* established as protohominoid.	African and Asian hominoid split. *Sivapithecus (Ramapithecus)* related to *Pongo?*
17			
	EARLY MIOCENE		Hominoid radiation begins.
24	OLIGOCENE	Hominoids diverge from Old World monkeys.	Catarrhine radiation begins.

CHANGING VIEWS on hominoid and hominid evolution appear on this chart, which extends from Oligocene times at the bottom to the most recent Pleistocene at the top. Numerals between the geologic subdivisions give the terminal dates of each subdivision in millions of years or fractions thereof before the present. The 1978 consensus of scholarly opinion on the subject is given at the left in the chart, together with items of fossil or archaeological evidence in support of such opinion. Today's consensus of opinion is given at the right.

and Asia and the consequent mixing of faunas 17 or 16 million years ago. Tentatively I now place the split at about 16 million years ago.

Early Hominids

The first undoubted hominids appear from perhaps four million to 3.75 million years ago in Tanzania and Ethiopia. Hominid evolution before then is obscure. Changes in fauna and habitat between 7.5 and 4.5 million years ago reflect a number of climatic and geographic events, the most spectacular one being the drying up of the Mediterranean on several occasions about 5.5 million years ago. The connection between these events and the evolution of the hominoids of Africa is not clear. Nevertheless, the footprint evidence of hominids that walked on two legs, uncovered by Mary Leakey and her colleagues at Laetoli in Tanzania, is firmly dated at 3.75 million years ago. Moreover, the almost half-complete skeleton of "Lucy" and other specimens of *Australopithecus afarensis* found by Donald Johanson and his colleagues at Hadar in Ethiopia are dated, somewhat less certainly, at between four and three million years ago. There is some taxonomic dispute about Lucy and the associated hominid remains from Hadar. Do they represent a new single species, *Australopithecus afarensis?* A northern subspecies of *A. africanus?* Two or more species? Tiptoeing past these questions, I shall give only one version, acknowledging that the situation is not clear-cut.

A. afarensis is primitive in most of its features. The adult males were probably from 50 to 100 percent larger than the females. Individuals weighed from 25 to 50 kilograms or more. Their brain was the size of an African great ape's, but with such a wide range of adult weights it is not clear whether the brain size with respect to body weight was larger or smaller. In some features of the face and palate these early hominids resembled chimpanzees. Overall, however, the skull as it has been reconstructed looks more like a female gorilla's: *A. afarensis* has larger teeth and is more robust than a chimpanzee.

The dentition includes a few plausibly primitive features, although the canine teeth of both sexes have a low crown and have lost their apelike forward projection. The cheek teeth are relatively large and are capped with a thick enamel, probably an adaptation to chewing large quantities of fruits, seeds, pods, roots and tubers, some of which may have been quite tough.

It is impossible to prove that the footprints in Tanzania were made by hominids like those represented by the Hadar fossils, but the morphology of the hip, knee and ankle joints of Lucy and her companions shows that the Hadar population was clearly bipedal. The foot bones show the same adaptation to bipedalism and an intriguing additional feature: the toe bones and metatarsals are long and curved. Compared with members of the genus *Homo* the arms of *A. afarensis* were long and the legs were short. Its hands were capable of powerful grasping. Judging by the proportions of the hand bones and the morphology of their joints, its manipulations were probably more precise than those of living chimpanzees.

The Hadar hominids lived in an area of woodland and savanna, away from the Pliocene forests, as part of a community of mammals that was structurally quite like the later communities of the Pleistocene. Such communities began to develop in eastern Africa in the later Miocene, and it is possible that they included hominids still unknown to us. By analogy with living mammals it is unlikely that the Hadar hominids were monogamous.

No stones that have been altered as tools have been found associated with the Hadar hominids. This may not, however, be entirely relevant. Suppose *A. afarensis* used the same kind of tools living chimpanzees do: stems, leaves, wood and stones casually adapted to food gathering, food processing and display behavior. Suppose, moreover, they did so more often than chimpanzees do today. Such tools would be impossible to recognize four million years later even if some near-miracle had preserved them.

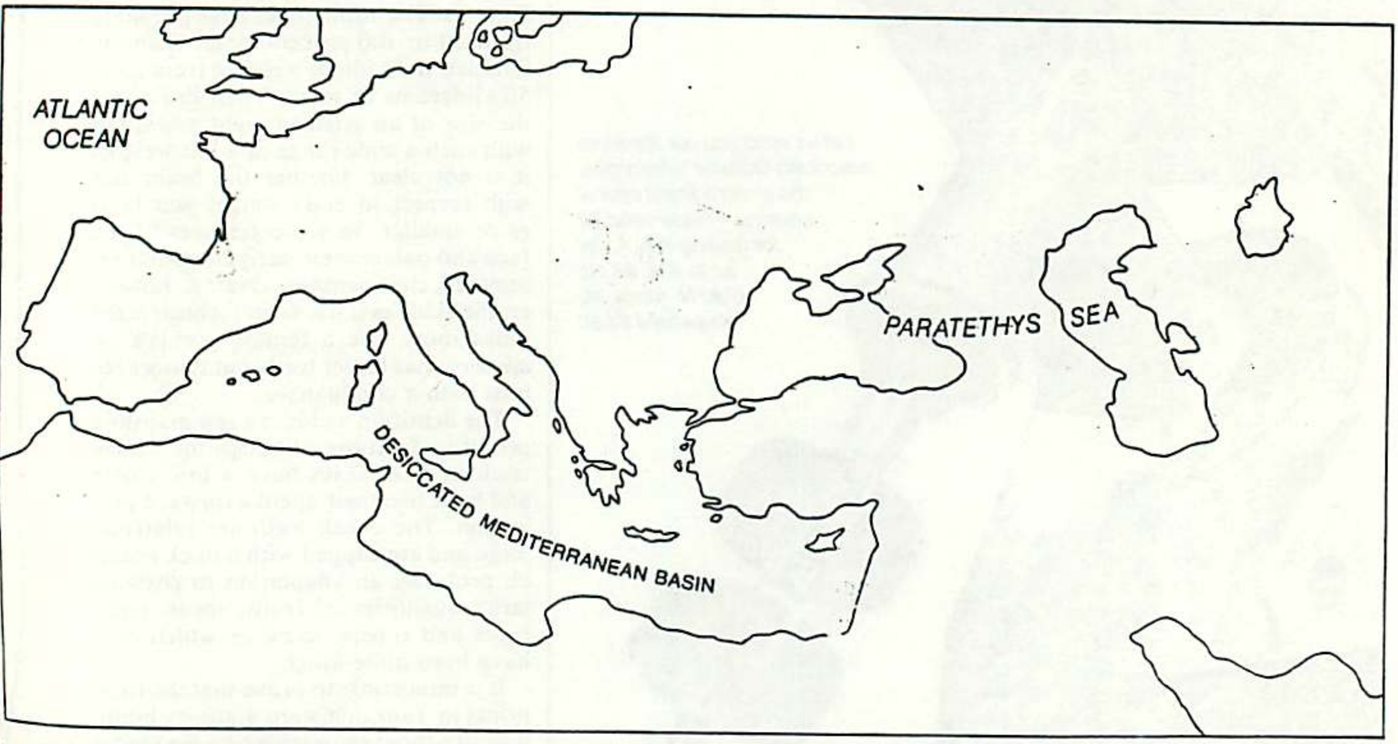
Tool use of this kind, if it may be postulated, might have been an important component of those behaviors that stimulated walking on two legs and the reduction of the canine teeth. Concepts such as these are much debated. Were the Hadar hominids protohunters and killers, opportunistic scavengers or gentle vegetarians? Did monogamy evolve along with bipedalism? Did pair-bonded males bring food to their female partner at a home base? Was gathered food, animal or plant, shared or did the Hadar and Laetoli hominids forage on an individual basis? Each of these viewpoints, to say nothing of others, has its adherents. For the present, however, any picture of what these early hominids were like as living, breathing animals must be deferred until there is more agreement on the nature of their basic adaptations. Above all, we must resist the temptation to see them as modern humans merely some distance removed from us in time.

Closing in on *Homo*

At some point between 2.5 and two million years ago African hominids underwent a modest adaptive radiation. By the later date at least two, probably three and perhaps even more species of hominids were present, although in any one part of Africa there is no evidence for more than three. In eastern Africa a species of *Australopithecus*, *A. boisei*, robust and with very large teeth, lived at the same time as a larger-brai nid whose type-fossil crani logued in Kenya as E.R.-14 cies is now most commonly *Homo habilis*, a name that was co 1964 by L. S. B. Leakey, Philip V. To as and John R. Napier, and was applied to E.R.-1470 by its discoverers, Richard Leakey and his colleagues. There was probably a third hominid species, small-toothed and small-brained, living in eastern Africa at the same time. Moreover, in southern Africa there was another species of *Australopithecus*, *A. robustus*, and possibly *H. habilis* as well. By about 1.75 million years ago, however, *H. habilis* disappeared from Africa and was replaced by an even larger-brained hominid: *Homo erectus*. A clearer picture of what kinds of animals these were is slowly emerging. The picture cannot ever, of course, be genuinely sharp, and one must avoid peering at it too closely lest, like a newspaper halftone, it dissolve into meaningless dots.

Consider *A. boisei*. A larger species than either *A. afarensis* or the *Australopithecus* species of southern Africa, it was small-brained and the male was markedly larger than the female. The species appeared more than two million years ago and survived, apparently with little morphological change, for a million years. *A. boisei* may have dug with sticks and hammered with stones, but there is no evidence that it used tools intensively.

Although *A. boisei* was larger than



LATE MIOCENE EPISODE, the drying up of the Mediterranean basin after it was sealed off from the Atlantic by mountain building, must have had catastrophic effects on the climate and ecology of Africa. The episode took place about 5.5 million years ago. Its effects on the mammalian fauna, hominoids included, are not clear, but it was perhaps a million and a half years later that hominids appeared.

other species of *Australopithecus*, it may have behaved much like them. Studies of the chewing surfaces of the teeth of various *Australopithecus* species by Alan Walker and by Frederick Grine of the State University of New York at Stony Brook suggest that all of them were basically rather apelike vegetarians, eating broadly similar diets that called for much repetitive chewing. Possibly their diet included more roots and tubers and less fruit than the diet of forest-dwelling apes.

Why did *A. boisei* evolve, and from what did it evolve? Was its evolution the result of a change in climate and habitat, perhaps a change in the availability of certain kinds of plant foods? Why did *A. boisei* become extinct after a million years of apparent stability? None of these questions can yet be answered. The fossil hunt in eastern Africa is intense, however, and one may hope *A. boisei*, its ancestor and the transition between them will eventually be better understood.

Understanding the transition to *Homo habilis* from its ancestor (which might also be the ancestor of *A. boisei*) is another matter. The various species of *Australopithecus* are difficult animals to "imagine"—to reconstruct as though they still lived—because nothing like them lives today. Still, they can be perceived dimly as an odd kind of ape. *H. habilis* is even more difficult to imagine. It is rather like *Australopithecus* in a number of ways, for example in its face and teeth, but it has a significantly larger brain, averaging about 700 cubic centimeters. There are no clear associations between the skulls of *H. habilis* and the other bones of the species, but the limb bones that are assumed to represent it, unlike those of *Australopithecus*, resemble those of later species of the genus *Homo* (with the exception of modern *H. sapiens*). It has been suggested that the resemblance reflects changes in the mode of locomotion and in the dimensions of the female pelvis demanded by the size of the newborn infant.

At about the same time that *A. boisei* and *H. habilis* appear in Africa so do the first archaeological sites: concentrations of used or altered stone, often brought from some distance away, together with animal remains. The consensus is that the sites document a shift in diet to include more animal food; *H. habilis* rather than *A. boisei* is generally implicated. The dietary change and the increase in brain size are also linked causally. Thereafter the consensus breaks down. Was *H. habilis* a hunter-gatherer who transported food to home bases to be shared and who practiced a division of labor that left hunting to the males and gathering to the females? This is to ask: Was *H. habilis* behaviorally "modern" or was it still basically vegetarian, perhaps adding some meat to its diet mainly by scavenging?

The archaeological record is itself ambiguous. The same scattering of tools and bones could be produced by opportunistic scavenging for individual consumption, with little or no food transport and sharing, or could represent home bases as sites of intense social interaction. In short, the picture of *H. habilis* remains blurred. More fossil material and more archaeological sites are needed, and even then the transition from *Australopithecus* to *Homo* is likely to be obscure. Both animals are too different from any known today. What happened to *H. habilis?* It survived for a few hundred thousand years only to be replaced by a much more durable species, *H. erectus*. Was *H. habilis* the ancestor of *H. erectus?* Although it is widely assumed to be the case, it may not be so.

Getting Around

Homo erectus is the first widely distributed hominid species. It appeared earliest in Africa, as far as is known, and it may indeed have originated there some 1.6 million years ago. Whether or not that is so, by one million years ago the species was present in southeastern and eastern Asia and survived in that area at least until 300,000 years ago. In that span of time, well over a million years, the physical record of *H. erectus* is one of prolonged morphological stability.

H. erectus resembled later species of *Homo* (modern *H. sapiens* excepted) in both body size and robustness. Larger-brained than *H. habilis* (more than 800 cubic centimeters), it had front teeth as big as those of earlier hominids, but its cheek teeth and its face were smaller. The archaeological record suggests that some *H. erectus* populations were makers of larger, symmetrically flaked stone tools: bifaces, or "hand axes." Moreover, some *H. erectus* populations may have used fire. Some later *H. erectus* populations may have commanded more sophisticated techiques of producing and modifying stone tools, but the overall impression is one of prolonged stability, in behavior as well as in morphology. Arthur Jelinek of the University of Arizona has proposed that "paleocultural" is the appropriate term for such behavior, so different is it from the rapid changes we associate with "cultural" behavior in modern humans.

Paleoanthropologists shift gears, often without realizing it, when they talk about human evolution over the past 1.5 million years. The problems in hominoid and hominid evolution I have presented up to this point concern macroevolutionary questions: issues of broad adaptations, evolutionary trends and speciation. The hominids labeled *H. erectus*, "archaic" *H. sapiens* (including the Neanderthals) and "modern" *H. sapiens* probably represent a continuum, and the patterns of change within these lineages represent in contrast microevolution. Nevertheless, some of the most interesting paleoanthropological work concerns the last great step in human evolution: the shift, about 45,000 or 40,000 years ago, from archaic *H. sapiens* to modern *H. sapiens*.

Erik Trinkaus of the University of New Mexico has shown that the Neanderthals (and probably their contemporaries in other parts of the Old World) were probably as unlike us behaviorally as they were physically. The Neanderthals' skeletons were much more robust and the muscle attachments on the bones indicate that they were much stronger than we are. Their teeth were larger and are heavily worn, probably (by analogy with living Eskimos) from being used for a variety of nonfeeding activities (such as the chewing of animal hides to soften them). The transition to modern *H. sapiens* was marked by a loss of the Neanderthal robustness in skeleton, face and dentition. There were also changes in the morphology of the female pelvis. These perhaps suggest that formerly easy births had become harder or even that the time of gestation was being reduced to our present nine months from perhaps 11 months (a period in line with predictions based on general mammalian relations among maternal body size, fetal size and length of gestation).

Important behavioral changes are also evident in the archaeological record. They include a proliferation of superior stone and bone tools, shifts in hunting patterns, in the use and control of fire, in the use of clothing, in settlement patterns, in population size, in ecological range, in art and other evidences of ritual activity. All of this points to the emergence of a species possessing modern behavioral capabilities (and potential) from an ancestral species lacking, at least by modern standards, in some significantly human characteristics.

No brief summary such as this one can do justice to the richness and complexity of the information available regarding this most recent, even if microevolutionary, transition. Yet it remains the best-documented and potentially the most understandable of all the transitions of the past 30 million years. As an outsider, working in a time span that may contain no relevant fossils at all in any particular 100,000-year period, I am ruefully envious of those who work with this most recent period. Viewed at long range it is, in a very real sense, the icing on the cake. Yet one must remember that it is only the icing. The cake itself—the many earlier stages of human evolution—remains for now much harder to digest.

Gould, S.J.
"The Hottentot Venus"
Natural History
American Museum of Nat. Hist.
Used by permission

The Hottentot Venus

A sensation from Piccadilly to Paris, she drew crowds from all classes, including scientists

by Stephen Jay Gould

I had a little friend in nursery school. I don't even remember her name. But I do recall that I offered her some secret advice one day at the playground. I told her that the enormous surrounding creatures known as adults always looked up when they walked, and that we little folk would therefore find all manner of valuable things on the ground if we only kept our gazes down. Were my paleontological predispositions already in evidence?

Carl Sagan and I both grew up in New York, both interested in biology and astronomy. Since Carl is tall and chose astronomy, while I'm short and chose paleontology, I always figured that he'd be looking up (as he did with some regularity in "Cosmos"), while I'd be sticking to my old but good advice and staring at the ground. But I one-upped him (literally) last month in Paris.

A few years back, Yves Coppens, professor at the Musée de l'Homme, took Carl on a tour of the museum's innards. There, on a shelf in storage, he found the brain of Paul Broca floating in Formalin in a bell jar. He wrote a fine essay about this visit, and then used it as the title piece of his book *Broca's Brain*. A few months ago, Yves took me on a similar tour. I held the skull of Descartes and of our mutual ancestor, the old man of Cro-Magnon. I also found Broca's brain, resting on a shelf and surrounded by other bell jars holding the brains of his illustrious scientific contemporaries—all white and all male. Yet I found the most interesting items on the shelf just above. Perhaps Carl never looked up.

This area of the museum's "back wards" holds Broca's collection of anatomical parts, to which he later made his own generous and posthumous contribution. Broca, a great medical anatomist and anthropologist, embodied the great nineteenth-century faith in quantification as a key to objective science. If he could collect enough human parts from enough human races, the resultant measurements would surely define the great scale of human progress, from chimp to Caucasian. Broca was not more virulently racist than his contemporary scientists (nearly all successful white males, of course); he was simply more assiduous in accumulating irrelevant data, selectively presented to support an a priori viewpoint.

These shelves contain a ghoulish potpourri: severed heads from New Caledonia; an illustration of foot binding as practiced upon Chinese women—yes, a bound foot and lower leg, severed between knee and ankle. And, on the shelf just above the brains, I saw a little exhibit that provided an immediate and chilling insight into nineteenth-century *mentalité* and the history of racism: in three smaller jars, I saw the dissected genitalia of three Third-World women. I found no brains of women, and neither Broca's penis nor any male genitalia grace the collection.

The three jars are labeled *une négresse*, *une péruvienne*, and *la Vénus Hottentotte*, or the Hottentot Venus. Georges Cuvier himself, France's greatest anatomist (see my column of June 1981), had dissected the Hottentot Venus upon her death in Paris late in 1815. And he went right to the genitalia for a particular and interesting reason, to which I will return after recounting the tale of this unfortunate woman.

In an age before television and movies made virtually nothing on earth exotic, and when anthropological theory assessed as subhuman both malformed Caucasians and the normal representatives of other races, the exhibition of unusual humans became a profitable business both in upper-class salons and in street-side stalls (see Richard D. Altick's *The Shows of London*, Harvard University Press, 1978, or the book, stage, and screen treatments of the "Elephant Man"). Supposed savages from faraway lands were a mainstay of these exhibitions, and none was so famous as the Hottentot Venus. (The Hottentots and Bushmen are closely related, small-statured people of southern Africa. Traditional Bushmen, when first encountered by Europeans, were hunter-gatherers, while Hottentots were pastoralists who raised cattle. Anthropologists now tend to forgo these European and somewhat derogatory terms and to designate both groups collectively as the Khoi-San peoples, a composite word constructed from each group's own name for itself.) The Hottentot Venus was a servant of Dutch farmers near Capetown, and we do not know her actual group membership. I will refer to her by her given name of Sartje.

Hendrick Cezar, brother of Sartje's "employer," suggested a trip to England for exhibition and promised to make Sartje a wealthy woman thereby. Lord Caledon, governor of the Cape, granted permission for the trip but later regretted his decision when he understood its purposes more fully. (Sartje's exhibition aroused much debate and she always had supporters, disgusted with the display of humans as animals; the show went on, but not to universal approbation.) She arrived in London in 1810 and immediately went on exhibition in Piccadilly, where she was a sensation, for reasons soon to be discussed. A member of the African Association, a benevolent society that petitioned for her "release," described the show. He first encountered Sartje in a cage on a platform raised a few feet above the floor:

> On being ordered by her keeper, she came out. . . . The Hottentot was produced like a wild beast, and ordered to move backwards and forwards and come out and go into her cage, more like a bear in a chain than a human being.

Yet Sartje, interrogated in Dutch before a court, insisted that she was not under restraint and understood perfectly well that she had been guaranteed half the profits. The show went on.

After a long tour of the English provinces, Sartje went to Paris where an animal trainer exhibited her for fifteen months, causing as great a sensation as in England. Cuvier and all the great naturalists of France visited her and she posed in the nude for scientific paintings at the Jardin du Roi. But she died of some inflammatory ailment on December 29, 1815, and ended up on Cuvier's dissecting table, rather than wealthy in Capetown.

Why, in an age deluged with human exhibitions, was Sartje such a sensation? We may offer two answers, each troubling and each associated with one of her official titles—Hottentot and Venus.

On the racist ladder of human progress, Bushmen and Hottentots vied with Australian aborigines for the lowest rung, just above chimps and orangs. (Some scholars have argued that the earliest designation applied by seventeenth-century Dutch settlers—*Bosmanneken*, or "Bushman"—was a literal translation of a Malay word well known to them—*Orang Outan*, or "man of the forest.") In this system, Sartje exerted a grim fascination, not as a missing link in a later evolutionary sense, but as a creature who straddled that dreaded boundary between human and animal and thereby taught us something about a self still present, although submerged, in "higher" creatures.

Contemporary commentators emphasized both the simian appearance and the brutal habits of Bushmen and Hottentots. In 1839, the leading American anthropologist S.G. Morton labeled Hottentots as "the nearest approximation to the lower animals. . . . Their complexion is a yellowish brown, compared by travellers to the peculiar hue of Europeans in the last stage of jaundice. . . . The women are represented as even more repulsive in appearance than the men." Mathias Guenther, in an excellent article, "The Changing Western Image of the Bushmen" (*Paideuma*, vol. 26, 1980, pp. 123–240), cites an 1847 newspaper account of a Bushman family displayed at the Egyptian Hall in London:

> In appearance they are little above the monkey tribe. They are continually crouching, warming themselves by the fire, chatting or growling. . . . They are sullen, silent and savage—mere animals in propensity, and worse than animals in appearance.

And the jaundiced account of a failed missionary in 1804:

> The Bushmen will kill their children without remorse, on various occasions; as when they are ill shaped, or when they are in want of food, or when obliged to flee from the farmers or others; in which case they will strangle them, smother them, cast them away in the desert or bury them alive. There are instances of parents throwing their tender offspring to the hungry lion, who stands roaring before their cavern, refusing to depart before some peace offering be made to him.

Guenther reports that this equation of Bushman and animal became so ingrained that one party of Dutch settlers, out on a hunting expedition, shot and ate a Bushman, assuming that he was the African equivalent of the Malay orang.

Cuvier's monograph of Sartje's dissection, published in the *Mémoires du Muséum d'Histoire Naturelle* for 1817, followed this traditional view. After discussing and dismissing various ill-founded legends, Cuvier promised to present only "positive facts"—including this description of a Bushman's life:

> Since they are unable to engage in agriculture, or even in a pastoral life, they subsist entirely on hunting and pilfering. They live in caves and cover themselves only with the skins of animals they have killed. Their only industry involves the poisoning of their arrows and the manufacture of nets for fishing.

His description of Sartje herself emphasizes any point of superficial similarity with any ape or monkey. (I hardly need mention that since people vary so much, each group must be closer than others to some feature of some other primate, without implying anything about genealogy or aptitude.) Cuvier, for example, discusses the flatness of Sartje's nasal bones: "In this respect, I have never seen a human head more similar to that of monkeys." He emphasizes various proportions of the femur (upper leg bone) as embodying "characters of animality." He speaks of Sartje's small skull (no surprise for a woman four and a half feet tall), and relegates her to stupidity according to "that cruel law, which seems to have condemned to an eternal inferiority those races with small and compressed skulls." He even abstracted a set of supposedly simian responses from her behavior: "Her movements had something brusque and capricious about them, which recall those of monkeys. She had, above all, a way of pouting her lips, in the same manner as we have observed in orang utans."

Yet a careful reading of the entire monograph belies these interpretations, since Cuvier states again and again (although he explicitly draws neither moral nor message) that Sartje was an intelligent woman with general proportions that would not lead connoisseurs to frown. He mentions, in an offhand sort of way, that Sartje possessed an excellent memory, spoke Dutch rather well, had some command of English, and was learning a bit of French when she died. (Three foreign languages ain't bad for a caged brute; I only wish that more Americans could do one-third so well.) He admitted that her shoulders, back, and chest "had grace"; and with the gentilesse of his own race, spoke of *sa main charmante* ("her charming hand").

Yet Sartje's hold over well-bred Europe did not arise from her racial status alone. She was not simply the Hottentot or the Hottentot woman, but the Hottentot *Venus*. Under all the official words lay the great and largely unsaid reason for her popularity. Khoi-San women do exaggerate two features of their sexual anatomy (or at least of body parts that excite sexual feelings in most men). The Hottentot Venus won her fame as a sexual object, and the combination of supposed bestiality and lascivious fascination focused the attention of men who could thus obtain both vicarious pleasure and a smug reassurance of superiority. (Women visitors probably experienced the second benefit in sufficiently heightened form to more than offset any absence of the first.)

Primarily—for, as they say, you can't miss it—Sartje was, in Altick's words, "steatopygous to a fault." Khoi-San women accumulate large amounts of fat in their buttocks, a condition called steatopygia. The buttocks protrude far back, often coming to a point at their upper extremity and sloping down toward the genitalia. Sartje was especially well endowed, the probable cause of Cezar's decision to convert her from servant to siren. Sartje covered her genitalia during exhibitions, but her rear end *was* the show, and she submitted to endless gaze and poke for five long years. Since European women did not wear bustles at the time, but indicated by their clothing only what nature had provided, Sartje seemed all the more incredible.

Cuvier well understood the mixed bestial and sexual nature of Sartje's fascination when he wrote that "everyone was able to see her during her eighteen-month stay in our capital, and to verify the enormous protrusion of her buttocks and the brutal appearance of her face." In his dissection, Cuvier focused on an unsolved mystery surrounding each of her unusual features. Europeans had long wondered whether the large buttocks were fatty, muscular, or perhaps even supported by a previously unknown bone. The problem had already been solved—in favor of fat—by external observation, the primary reason for her disrobing before scientists at the Jardin du Roi. Still, Cuvier dissected her buttocks and reported:

> We could verify that the protuberance of her buttocks had nothing muscular about it, but arose from a [fatty] mass of a trembling and elastic consistency, situated immediately under her skin. It vibrated with all movements that the woman made.

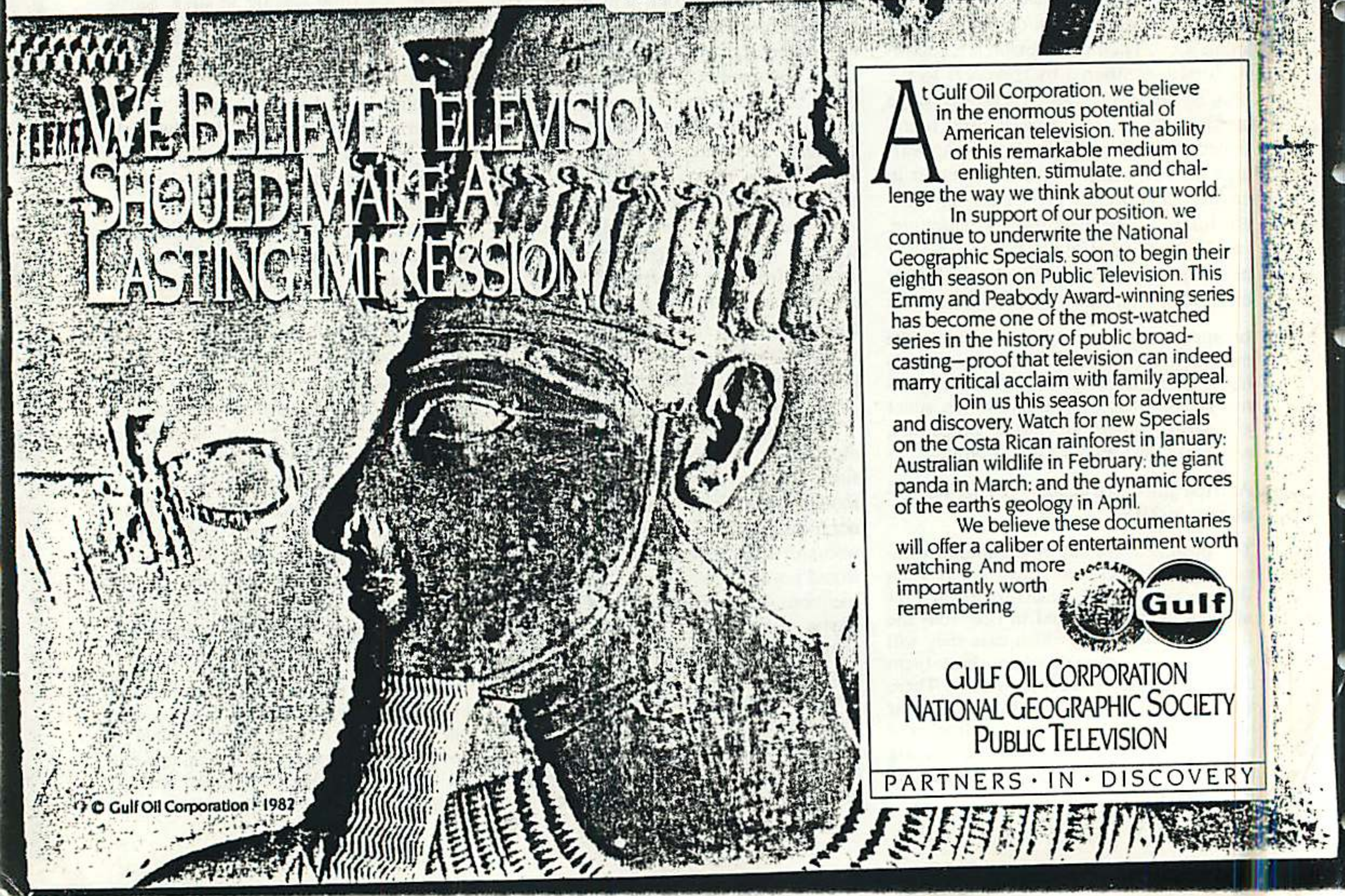
But Sartje's second peculiarity was far more enticing to scientists, all the more so because she kept it scrupulously hidden, even refusing a display at the Jardin. Only after her death would the curiosity of science be slaked.

Reports had circulated for two centuries of a wondrous structure attached directly to the female genitalia of Khoi-San women and covering these private parts with a veil of skin, the so-called *sinus pudoris*, or "curtain of shame." (If I may be permitted a short excursion into the realm of scholarly minutiae—the footnotes of more conventional academic publication—I would like to correct a standard mistranslation of Linnaeus, one that I have made myself. In his original description of *Homo sapiens*, Linnaeus provided a most unflattering account of African blacks, including the line: *feminae sinus pudoris*. This has usually been translated, "women are without shame"—a slur quite consistent with the rest of his words. Now "without shame," in Latin, should be *sine pudore*, not *sinus pudoris*. But eighteenth-century scientific Latin was written so indifferently that misspellings and wrong cases are no bar to actual intent, and the reading "without shame" has held. But Linnaeus was only saying that African women have a genital flap, or *sinus pudoris*. He was also wrong, because only the Khoi-San and a few related peoples develop this feature.)

The nature of the *sinus pudoris* had generated a lively debate, with partisans of both sides claiming eyewitness support. One party held that the *sinus* was simply an enlarged part of the ordinary genitalia; others called it a novel structure found in no other race. Some even described it as a large fold of skin hanging down from the lower abdomen itself. Cuvier resolved to find out, and made this quest the primary object of his dissection. He began his monograph by noting: "There is nothing more famous in natural history than the *tablier* (the French rendering of *sinus pudoris*) of Hottentots, and, at the same time, no feature has been the object of so many arguments." Cuvier resolved the debate with his usual elegance: the *labia minora*, or "inner lips," of the ordinary female genitalia are greatly enlarged in Khoi-San women, and may hang down three or four inches below the vagina when women stand, thus giving the impression of a separate and enveloping curtain of skin. Cuvier preserved his skillful dissection of Sartje's genitalia and wrote with a flourish: "I have the honor to present to the Academy the genital organs of this woman prepared in a manner that leaves no doubt about the nature of her *tablier*." And Cuvier's gift still stands forgotten on a shelf at the Musée de l'Homme—right above Broca's brain.

Yet while Cuvier correctly identified the nature of Sartje's *tablier*, he fell into an interesting error, born of the same false association that had been the source of Sartje's fascination—sexuality with animality. Since Cuvier regarded Hottentots as the most bestial of people, and since they had a large *tablier*, he assumed that the *tablier* would become progressively smaller as the darkness of southern Africa ceded to the light of Egypt. (In the last part of his monograph, Cuvier argues that the ancient Egyptians must have been fully Caucasian; who else could have built the pyramids?)

Cuvier knew that female circumcision was widely practiced in Ethiopia. He assumed that the *tablier* was at least half-sized among these people of intermediate hue and geography; and he further conjectured that its removal was necessary for sexual access, not that circumcision represented a custom born of power and imposed upon girls with genitalia not noticeably different from those of Euro-



pean women. "The negresses of Aby sinia," he wrote, "are inconvenienced to the point of being obliged to destroy these parts by knife and cauterization" (*par le fer et par le feu*, as he wrote in more euphonious French).

Cuvier also told an interesting tale, which needs no comment in repetition:

> The Portuguese Jesuits, who converted the King of Abyssinia and part of his people during the 16th century, felt that they were obliged to proscribe this practice [of female circumcision] since they thought that it was a holdover from the ancient Judaism of that nation. But it happened that Catholic girls could no longer find husbands, because the men could not reconcile themselves to such a disgusting deformity. The College of Propaganda sent a surgeon to verify the fact and, on his report, the reestablishment of the ancient custom was authorized by the Pope.

I needn't burden you with any detailed refutation of the general arguments that made the Hottentot Venus such a sensation. Their obvious ties to a priori prejudice, rather than to Sartje's reality, can only make us wonder what truths we now hold objective and dear will someday be exposed as equally rooted in hope, culture, and personal preference. I do, however, find it amusing that she and her people are, by modern convictions, so singularly and especially unsuited for the role she was forced to play.

If the early nineteenth century cast the Khoi-San peoples as approximations to the lower primates, they now rank among the heroes of modern social movements. Their languages, with complex clicks, were once dismissed as a guttural farrago of beastly sounds. They are now widely admired for their complexity and subtle expression. Cuvier had stigmatized the hunter-gatherer life styles of the traditional San (Bushmen) as the ultimate degradation of a people too stupid and indolent to farm or raise cattle. The same groups have become prototypes of the ecology movement for their understanding, nonexploitative, and balanced approach to natural resources. Of course, as Guenther points out in his article on the Bushman's changing image, our modern accolades may also be unrealistic. Still, if one must be exploited rather than understood, attributions of kindness and heroism sure beat accusations of animality.

Furthermore, while Cuvier's contemporaries sought physical signs of bestiality in Khoi-San anatomy, anthropologists now identify these people as perhaps the most paedomorphic of human groups. Humans have evolved by a general retardation (or slowing up) of developmental rates, leaving our adult bodies rather similar in many respects to the juvenile, but not to the adult, form of our primate ancestors—an evolutionary result called paedomorphosis, or "child shaping." On this criterion, the greater the extent of paedomorphosis, the further away from a simian past (although minor differences among human races do not translate into variations in mental or any other kind of worth). Although Cuvier searched hard to find signs of animality in Sartje's lip movements or in the form of her leg bone, her people are, in general, perhaps the least simian of all humans.

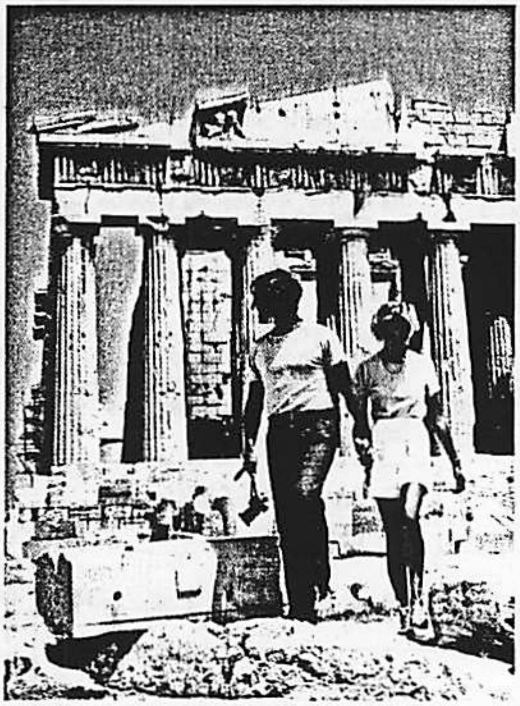
Finally, the major rationale for Sartje's popularity rests on a false premise. She fascinated Europeans because she had enlarged buttocks and genitalia and because she supposedly belonged to the most backward of human groups. It all fit for Cuvier's contemporaries. Advanced humans (read modern Europeans) are refined, modest, and sexually restrained (not to mention hypocritical for advancing such a claim). Animals are overtly and actively sexual, and so betray their primitive character. Thus, Sartje's exaggerated sexual organs are proof of her animality. But the argument is backward. Humans are the most sexually active primates, and humans have the largest sexual organs of our order. In this dubious line of argument a person with larger than average endowment is, if anything, more human.

On all accounts—mode of life, physical appearance, and sexual anatomy—London and Paris should have stood in a giant cage while Sartje watched. Still, Sartje had her posthumous triumphs. Broca inherited not only Cuvier's preparation of her *tablier*, but her skeleton as well. In 1862, he thought he had found a criterion for arranging human races by physical merit. He measured the ratio of radius (lower arm bone) to humerus (upper arm bone), reasoning that higher ratios indicate longer forearms—a traditional feature of apes. He became very hopeful that objective measurement had confirmed his foregone conclusion when blacks averaged .794 and whites .739. But Sartje's skeleton yielded .703 and Broca promptly abandoned his criterion. Had not Cuvier praised the arm of the Hottentot Venus?

Sartje continues her mastery of Mr. Broca today. His brain decomposes in a leaky jar. Her *tablier* stands above, while her well-prepared skeleton gazes up from below. Death, as the good book says, is swallowed up in victory.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.

"We never thought we could afford Greece."



In case you hadn't heard, Greece is a steal. Because the U.S. dollar is now worth 35% more than just two years ago.

A good hotel for two costs $24 a night. Lunch for two, about $6. A wool sweater, around $20. A night at the ancient Greek theatre sets you back $3.50.

Visit beautiful Athens, then cruise the seductive Greek Isles. Or charter a yacht and be master of your own destiny.

No matter where you go, tour packages are a heftier bargain than any time in recent history.

Call your travel agent.

As luck would have it, Europe's most fascinating country is also one of the least expensive.

Greece. The Classic European Vacation

¡Speak Spanish like a diplomat!

What sort of people need to learn a foreign language as quickly and effectively as possible? *Foreign service personnel,* that's who. Members of America's diplomatic corps are assigned to U.S. embassies abroad, where they must be able to converse fluently in every situation.

Now you can learn to speak Spanish just as these diplomatic personnel do— with the Foreign Service Institute's Programmatic Spanish Course. You'll learn Latin American Spanish recorded by native speakers.

The U.S. Department of State has spent thousands of dollars developing this course. It's by far the most effective way to learn Spanish at your own convenience and at your own pace.

The course consists of a series of cassettes and accompanying textbook. Simply follow the spoken and written instructions, listening and repeating. By the end of the course, you'll be learning and speaking entirely in Spanish!

This course turns your cassette player into a "teaching machine." With its unique "programmatic" learning method, you set your own pace — testing yourself, correcting errors, reinforcing accurate responses.

AUDIO-FORUM®

The FSI's Programmatic Spanish Course comes in two volumes, each shipped in a handsome library binder. You may order one or both:

☐ **Volume I: Basic.** 12 cassettes (17 hr.), manual, and 464-p. text, $115.
☐ **Volume II: Intermediate.** 8 cassettes (11½ hr.), manual, and 614-p. text, $98.

(Conn. and N.Y. residents add sales tax)

TO ORDER BY PHONE, PLEASE CALL TOLL-FREE NUMBER: 1-800-243-1234.

To order by mail, clip this ad and send with your name and address, and a check or money order — or charge to your credit card (AmEx, VISA, MasterCard, Diners) by enclosing card number, expiration date, and your signature.

The Foreign Service Institute's Spanish course is unconditionally guaranteed. Try it for three weeks. If you're not convinced it's the fastest, easiest, most painless way to learn Spanish, return it and we'll refund every penny you paid. Order today!

81 courses in 26 other languages also available. Write us for free catalog. Our 10th year.

Audio-Forum
Suite 914
On-the-Green,
Guilford, CT 06437
(203) 453-9794

Or visit our New York sales office: 145 E. 49th St., New York, N.Y. 10017 (212)753-1783

Gould, S.J.
"Carrie Buck's Daughter"
Natural History
American Museum of Nat. Hist.
Used by permission

Carrie Buck's Daughter

A popular, quasi-scientific idea can be a powerful tool for injustice

by Stephen Jay Gould

The Lord really put it on the line in his preface to that prototype of all prescription, the Ten Commandments:

> ... for I, the Lord thy God, am a jealous God, visiting the iniquity of the fathers upon the children unto the third and fourth generation of them that hate me (Exod. 20:5).

The terror of this statement lies in its patent unfairness—its promise to punish guiltless offspring for the misdeeds of their distant forebears.

A different form of guilt by genealogical association attempts to remove this stigma of injustice by denying a cherished premise of Western thought—human free will. If offspring are tainted not simply by the deeds of their parents but by a material form of evil transferred directly by biological inheritance, then "the iniquity of the fathers" becomes a signal or warning for probable misbehavior of their sons. Thus Plato, while denying that children should suffer directly for the crimes of their parents, nonetheless defended the banishment of a man whose father, grandfather, and great-grandfather had all been condemned to death.

It is, perhaps, merely coincidental that both Jehovah and Plato chose three generations as their criterion for establishing different forms of guilt by association. Yet we have a strong folk, or vernacular, tradition for viewing triple occurrences as minimal evidence of regularity. We are told that bad things come in threes. Two may be an accidental association; three is a pattern. Perhaps, then, we should not wonder that our own century's most famous pronouncement of blood guilt employed the same criterion—Oliver Wendell Holmes's defense of compulsory sterilization in Virginia (Supreme Court decision of 1927 in *Buck* v. *Bell*): "three generations of imbeciles are enough."

Restrictions upon immigration, with national quotas set to discriminate against those deemed mentally unfit by early versions of IQ testing, marked the greatest triumph of the American eugenics movement—the flawed hereditarian doctrine, so popular earlier in our century and by no means extinct today (see my column on Singapore's "great marriage debate," May 1984), that attempted to "improve" our human stock by preventing the propagation of those deemed biologically unfit and encouraging procreation among the supposedly worthy. But the movement to enact and enforce laws for compulsory "eugenic" sterilization had an impact and success scarcely less pronounced. If we could debar the shiftless and the stupid from our shores, we might also prevent the propagation of those similarly afflicted but already here.

The movement for compulsory sterilization began in earnest during the 1890s, abetted by two major factors—the rise of eugenics as an influential political movement and the perfection of safe and simple operations (vasectomy for men and salpingectomy, the cutting and tying of Fallopian tubes, for women) to replace castration and other obvious mutilation. Indiana passed the first sterilization act based on eugenic principles in 1907 (a few states had previously mandated castration as a punitive measure for certain sexual crimes, although such laws were rarely enforced and usually overturned by judicial review). Like so many others to follow, it provided for sterilization of afflicted people residing in the state's "care," either as inmates of mental hospitals and homes for the feebleminded or as inhabitants of prisons. Sterilization could be imposed upon those judged insane, idiotic, imbecilic, or moronic, and upon convicted rapists or criminals when recommended by a board of experts.

By the 1930s, more than thirty states had passed similar laws, often with an expanded list of so-called hereditary defects, including alcoholism and drug addiction in some states, and even blindness and deafness in others. It must be said that these laws were continually challenged and rarely enforced in most states; only California and Virginia applied them zealously. By January 1935, some 20,000 forced "eugenic" sterilizations had been performed in the United States, nearly half in California.

No organization crusaded more vociferously and successfully for these laws than the Eugenics Record Office, the semiofficial arm and repository of data for the eugenics movement in America. Harry Laughlin, superintendent of the Eugenics Record Office, dedicated most of his career to a tireless campaign of writing and lobbying for eugenic sterilization. He hoped, thereby, to eliminate in two generations the genes of what he called the "submerged tenth"—"the most worthless one-tenth of our present population." He proposed a "model sterilization law" in 1922, designed

> to prevent the procreation of persons socially inadequate from defective inheritance, by authorizing and providing for eugenical sterilization of certain potential parents carrying degenerate hereditary qualities.

This model bill became the prototype for most laws passed in America, although few states cast their net as widely as Laughlin advised. (Laughlin's categories encompassed "blind, including those with seriously impaired vision; deaf, including those with seriously impaired hearing; and dependent, including orphans, ne'er-do-wells, the homeless, tramps, and paupers.") Laughlin's suggestions were better heeded in Nazi Germany, where his model act served as a basis for the infamous and stringently enforced *Erbgesundheitsrecht*, leading by the eve of World War II to the sterilization of some 375,000 people, most for "congenital feeblemindedness," but including nearly 4,000 for blindness and deafness.

The campaign for forced eugenic sterilization in America reached its climax and height of respectability in 1927, when the Supreme Court, by an 8–1 vote, upheld the Virginia sterilization bill in the case of *Buck* v. *Bell*. Oliver Wendell Holmes, then in his mid-eighties and the most celebrated jurist in America, wrote the majority opinion with his customary verve and power of style. It included the notorious paragraph, with its chilling tag line, cited ever since as the quintessential statement of eugenic principles. Remembering with pride his own distant experiences as an infantryman in the Civil War, Holmes wrote:

> We have seen more than once that the public welfare may call upon the best citizens for their lives. It would be strange if it could not call upon those who already sap the strength of the state for these lesser sacrifices. . . . It is better for all the world, if instead of waiting to execute degenerate offspring for crime, or to let them starve for their imbecility, society can prevent those who are manifestly unfit from continuing their kind. The principle that sustains compulsory vaccination is broad enough to cover cutting the Fallopian tubes. Three generations of imbeciles are enough.

Who, then, were the famous "three generations of imbeciles," and why should they still compel our interest?

When the state of Virginia passed its compulsory sterilization law in 1924, Carrie Buck, an eighteen-year-old white woman, was an involuntary resident at the State Colony for Epileptics and Feeble-Minded. As the first person selected for sterilization under the new act, Carrie Buck became the focus for a constitutional challenge launched, in part, by conservative Virginia Christians who held, according to eugenical "modernists," antiquated views about individual preferences and "benevolent" state power. (Simplistic political labels do not apply in this case, and rarely do in general. We usually regard eugenics as a conservative movement and its most vocal critics as members of the left. This alignment has generally held in our own decade. But eugenics, touted in its day as the latest in scientific modernism, attracted many liberals and numbered among its most vociferous critics groups often labeled as reactionary and antiscientific. If any political lesson emerges from these shifting allegiances, we might consider the true inalienability of certain human rights.)

But why was Carrie Buck in the State Colony and why was she selected? Oliver Wendell Holmes upheld her choice as judicious in the opening lines of his 1927 opinion:

> Carrie Buck is a feeble-minded white woman who was committed to the State Colony. . . . She is the daughter of a feeble-minded mother in the same institution, and the mother of an illegitimate feeble-minded child.

In short, inheritance stood as the crucial issue (indeed as the driving force behind all eugenics). For if measured mental deficiency arose from malnourishment, either of body or mind, and not from tainted genes, then how could sterilization be justified? If decent food, upbringing, medical care, and education might make a worthy citizen of Carrie Buck's daughter, how could the State of Virginia justify the severing of Carrie's Fallopian tubes against her will? (Some forms of mental deficiency are passed by inheritance in family lines, but most are not—a scarcely surprising conclusion when we consider the thousand shocks that beset fragile humans during their lives, from difficulties in embryonic growth to traumas of birth, malnourishment, rejection, and poverty. In any case, no fair-minded person today would credit Laughlin's social criteria for the identification of hereditary deficiency—ne'er-do-wells, the homeless, tramps, and paupers—although we shall soon see that Carrie Buck was committed on these grounds.)

When Carrie Buck's case emerged as the crucial test of Virginia's law, the chief honchos of eugenics knew that the time had come to put up or shut up on the crucial issue of inheritance. Thus, the Eugenics Record Office sent Arthur H. Estabrook, their crack fieldworker, to Virginia for a "scientific" study of the case. Harry Laughlin himself provided a deposition, and his brief for inheritance was presented at the local trial that affirmed Virginia's law and later worked its way to the Supreme Court as *Buck* v. *Bell*.

Laughlin made two major points to the court. First, that Carrie Buck and her

mother, Emma Buck, were feebleminded by the Stanford-Binet test of IQ, then in its own infancy. Carrie scored a mental age of nine years, Emma of seven years and eleven months. (These figures ranked them technically as "imbeciles" by definitions of the day, hence Holmes's later choice of words. Imbeciles displayed a mental age of six to nine years; idiots performed worse, morons better, to round out the old nomenclature of mental deficiency.) Second, that most feeblemindedness is inherited, and Carrie Buck surely belonged with this majority. Laughlin reported:

> Generally feeble-mindedness is caused by the inheritance of degenerate qualities; but sometimes it might be caused by environmental factors which are not hereditary. In the case given, the evidence points strongly toward the feeble-mindedness and moral delinquency of Carrie Buck being due, primarily, to inheritance and not to environment.

Carrie Buck's daughter was then, and has always been, the pivotal figure of this painful case. As I stated before, we tend (often at our peril) to regard two as potential accident and three as an established pattern. The supposed imbecility of Emma and Carrie might have been coincidental, but the diagnosis of similar deficiency for Vivian Buck (made by a social worker, as we shall see, when Vivian was but six months old) tipped the balance in Laughlin's favor and led Holmes to declare the Buck lineage inherently corrupt by deficient heredity. Vivian sealed the pattern—*three* generations of imbeciles are enough. Besides, had Carrie not given illegitimate birth to Vivian, the issue (in both senses) would never have emerged.

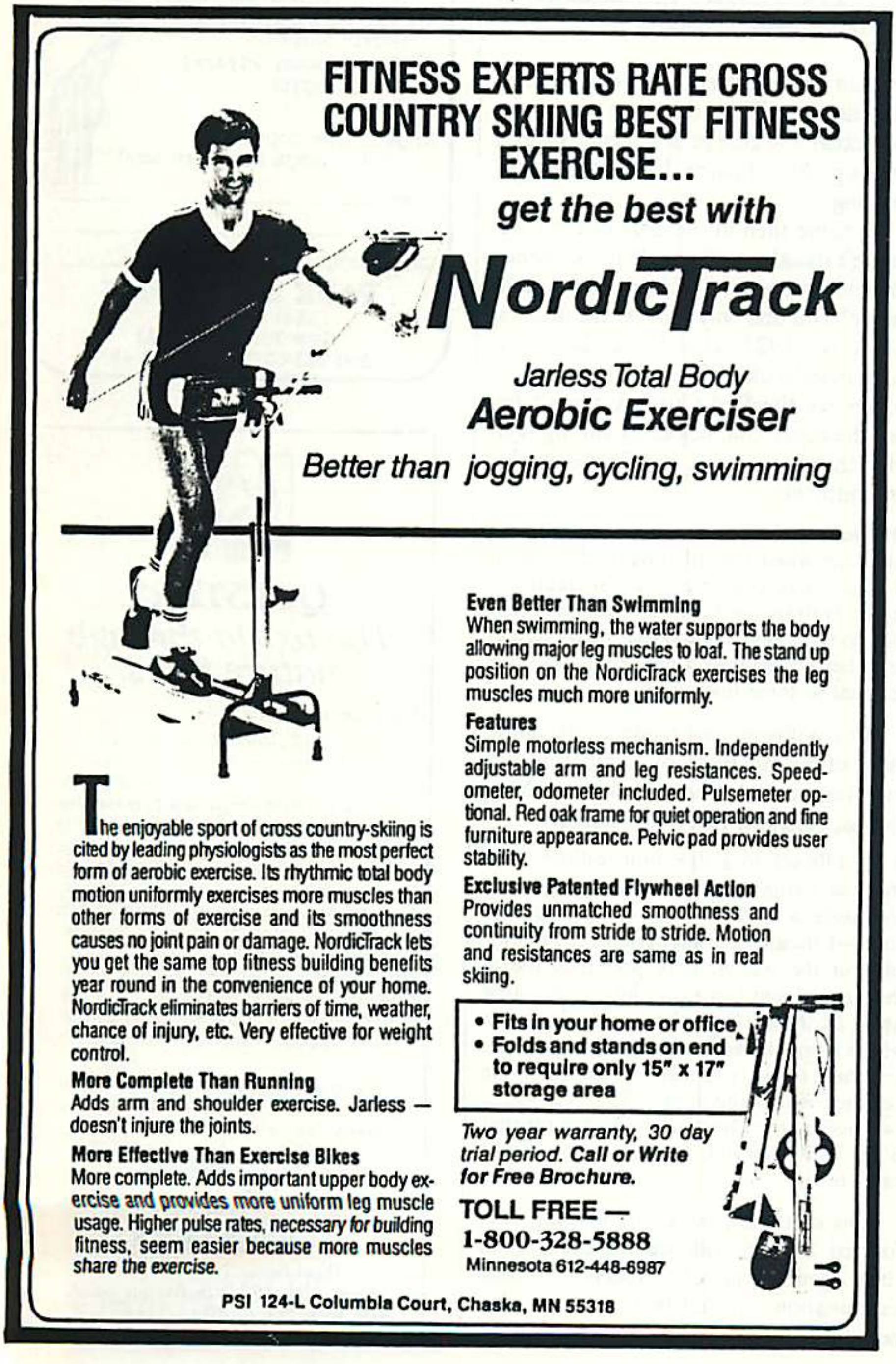
Oliver Wendell Holmes viewed his work with pride. The man so renowned for his principle of judicial restraint, who had proclaimed that freedom must not be curtailed without "clear and present danger"—without the equivalent of falsely yelling "fire" in a crowded theater—wrote of his judgment in *Buck* v. *Bell*: "I felt that I was getting near the first principle of real reform."

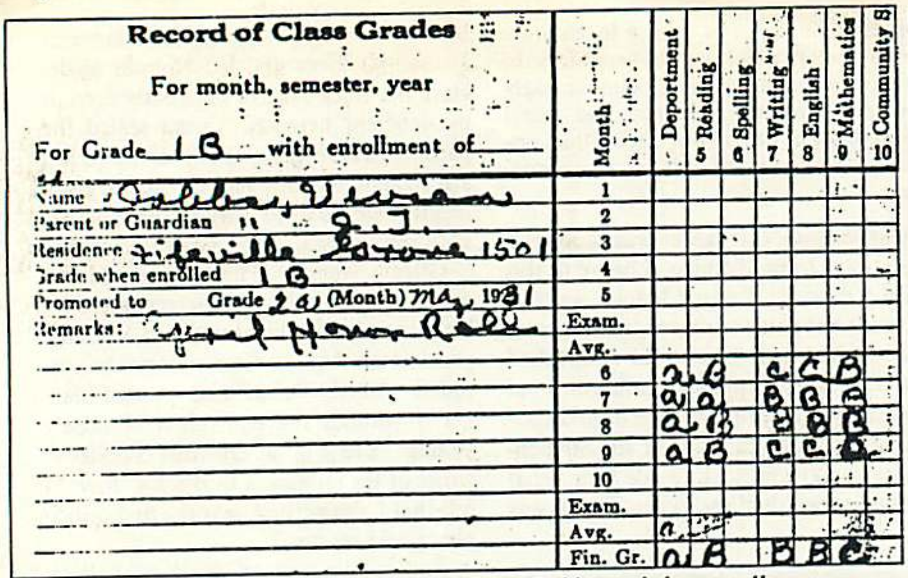
And so the case of *Buck* v. *Bell* remained for fifty years, a footnote to a moment of American history perhaps best forgotten. And then, in 1980, it reemerged to prick our collective conscience, when Dr. K. Ray Nelson, then director of the Lynchburg Hospital where Carrie Buck was sterilized, researched the records of his institution and discovered that more than 4,000 sterilizations had been performed, the last as late as 1972. He also found Carrie Buck, alive and well near Charlottesville, and her sister Doris, covertly sterilized under the same law (she was told that her operation was for appendicitis), and now, with fierce dignity, dejected and bitter because she had wanted a child more than anything else in her life and had finally, in her old age, learned why she had never conceived.

As scholars and reporters visited Carrie Buck and her sister, what a few experts had known all along became abundantly clear to everyone. Carrie Buck was a woman of obviously normal intelligence. For example, Paul A. Lombardo of the School of Law at the University of Virginia, and a leading scholar of the *Buck* v. *Bell* case, wrote in a letter to me:

> As for Carrie, when I met her she was reading newspapers daily and joining a more literate friend to assist at regular bouts with the crossword puzzles. She was not a sophisticated woman, and lacked social graces, but mental health professionals who examined her in later life confirmed my impressions that she was neither mentally ill nor retarded.

On what evidence, then, was Carrie Buck consigned to the State Colony for Epileptics and Feeble-Minded on January 23, 1924? I have seen the text of her commitment hearing; it is, to say the least, cursory and contradictory. Beyond the simple and undocumented say-so of her foster parents, and her own brief appearance before a commission of two doctors and a





Record of Class Grades

For month, semester, year

For Grade 1 B with enrollment of

Name Dobbs, Vivian
Parent or Guardian " J. T.
Residence Fifeville [illegible] 1501
Grade when enrolled 1 B
Promoted to Grade 2 A (Month) May, 1931
Remarks: [illegible] Honor Roll

Month	Deportment	5 Reading	6 Spelling	7 Writing	8 English	9 Mathematics	10 Community S
1							
2							
3							
4							
5							
Exam.							
Avg.							
6	a	B		C	C	B	
7	a	a		B	B	B	
8	a	B		B	B	B	
9	a	B		C	C	B	
10							
Exam.							
Avg.	a						
Fin. Gr.	a	B		B	B	C	

Charlottesville school record of 1931 notes Vivian Dobbs made honor roll

justice of the peace, no evidence was presented. Even the crude and early Stanford-Binet test, so fatally flawed as a measure of innate worth (see my book *The Mismeasure of Man*, although the evidence of Carrie's own case suffices) but at least clothed with the aura of quantitative respectability, had not yet been applied.

When we understand why Carrie Buck was committed in January 1924, we can finally comprehend the hidden meaning of her case and its message for us today. The silent key, again and as always, is her daughter Vivian, born on March 28, 1924, and then but an evident bump on her belly. Carrie Buck was one of several illegitimate children borne by her mother, Emma. She grew up with foster parents, J. T. and Alice Dobbs, and continued to live with them, helping out with chores around the house. She was apparently raped by a relative of her foster parents, then blamed for her resultant pregnancy. Almost surely, she was (as they used to say) committed to hide her shame (and her rapist's identity), not because enlightened science had just discovered her true mental status. In short, she was sent away to have her baby. Her case never was about mental deficiency; it was always a matter of sexual morality and social deviance. The annals of her trial and hearing reek with the contempt of the well-off and well-bred for poor people of "loose morals." Who really cared whether Vivian was a baby of normal intelligence; she was the illegitimate child of an illegitimate woman. Two generations of bastards are enough. Harry Laughlin began his "family history" of the Bucks by writing: "These people belong to the shiftless, ignorant and worthless class of anti-social whites of the South."

We know little of Emma Buck and her life, but we have no more reason to suspect her than her daughter Carrie of true mental deficiency. Their deviance was social and sexual; the charge of imbecility was a cover-up, Mr. Justice Holmes notwithstanding.

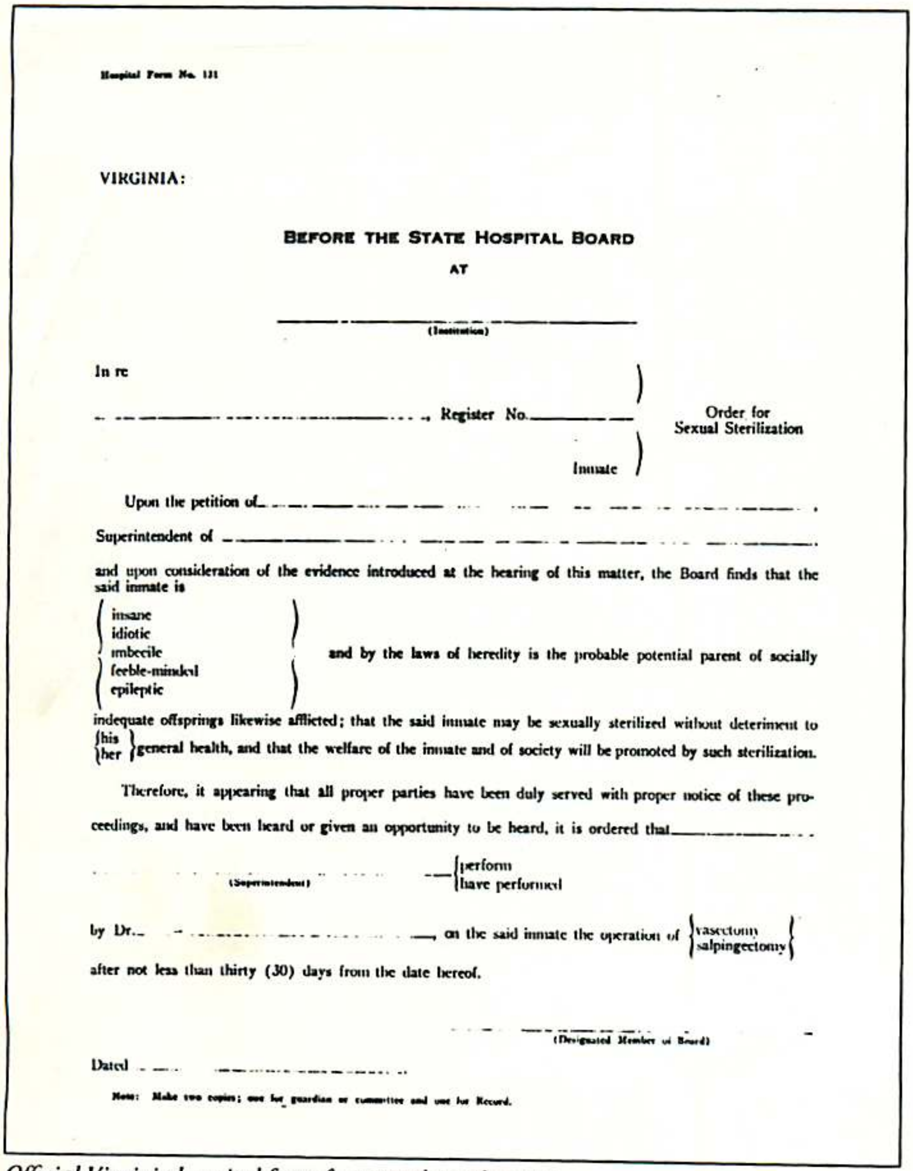
We come then to the crux of the case, Carrie's daughter, Vivian. What evidence was ever adduced for her mental deficiency? This and only this: At the original trial in late 1924, when Vivian Buck was seven months old, a Miss Wilhelm, social worker for the Red Cross, appeared before the court. She began by stating honestly the true reason for Carrie Buck's commitment:

> Mr. Dobbs, who had charge of the girl, had taken her when a small child, had reported to Miss Duke [the temporary secretary of Public Welfare for Albemarle County] that the girl was pregnant and that he wanted to have her committed somewhere—to have her sent to some institution.

Miss Wilhelm then rendered her judgment of Vivian Buck by comparing her with the normal granddaughter of Mrs. Dobbs, born just three days earlier:

> It is difficult to judge probabilities of a child as young as that, but it seems to me not quite a normal baby. In its appearance—I should say that perhaps my knowledge of the mother may prejudice me in that regard, but I saw the child at the same time as Mrs. Dobbs' daughter's baby, which is only three days older than this one, and there is a very decided difference in the development of the babies. That was about two weeks ago. There is a look about it that is not quite normal, but just what it is, I can't tell.

This short testimony, and nothing else, formed all the evidence for the crucial third generation of imbeciles. Cross-examination revealed that neither Vivian nor the Dobbs grandchild could walk or



Hospital Form No. 131

VIRGINIA:

BEFORE THE STATE HOSPITAL BOARD

AT

____
(Institution)

In re
____________________, Register No. ________ } Order for Sexual Sterilization
Inmate

Upon the petition of ____________________,

Superintendent of ____________________.

and upon consideration of the evidence introduced at the hearing of this matter, the Board finds that the said inmate is

{insane
idiotic
imbecile
feeble-minded
epileptic} and by the laws of heredity is the probable potential parent of socially indequate offsprings likewise afflicted; that the said inmate may be sexually sterilized without deteriment to {his / her} general health, and that the welfare of the inmate and of society will be promoted by such sterilization.

Therefore, it appearing that all proper parties have been duly served with proper notice of these proceedings, and have been heard or given an opportunity to be heard, it is ordered that ____________________

____________________ (Superintendent) {perform / have performed}

by Dr. ____________________, on the said inmate the operation of {vasectomy / salpingectomy}

after not less than thirty (30) days from the date hereof.

____
(Designated Member of Board)

Dated ____________________

Note: Make two copies; one for guardian or committee and one for Record.

Official Virginia hospital form for sexual sterilization

talk, and that "Mrs. Dobbs' daughter's baby is a very responsive baby. When you play with it or try to attract its attention—it is a baby that you can play with. The other baby is not. It seems very apathetic and not responsive." Miss Wilhelm then urged Carrie Buck's sterilization: "I think," she said, "it would at least prevent the propagation of her kind." Several years later, Miss Wilhelm denied that she had ever examined Vivian or deemed the child feebleminded.

Unfortunately, Vivian died at age eight of "enteric colitis" (as recorded on her death certificate), an ambiguous diagnosis that could mean many things but may well indicate that she fell victim to one of the preventable childhood diseases of poverty (a grim reminder of the real subject in *Buck* v. *Bell*). She is therefore mute as a witness in our reassessment of her famous case.

When *Buck* v. *Bell* resurfaced in 1980, it immediately struck me that Vivian's case was crucial and that evidence for the mental status of a child who died at age eight might best be found in report cards. I have therefore been trying to track down Vivian Buck's school records for the past four years and have finally succeeded. (They were supplied to me by Dr. Paul A. Lombardo, who also sent other documents, including Miss Wilhelm's testimony, and spent several hours answering my questions by mail and Lord knows how much time playing successful detective in re Vivian's school records. I have never met Dr. Lombardo; he did all this work for kindness, collegiality, and love of the game of knowledge, not for expected reward or even requested acknowledgment. In a profession—academics—so often marked by pettiness and silly squabbling over meaningless priorities, this generosity must be recorded and celebrated as a sign of how things can and should be.)

Vivian Buck was adopted by the Dobbs family, who had raised (but later sent away) her mother, Carrie. As Vivian Elaine Dobbs, she attended the Vena Public Elementary School of Charlottesville for four terms, from September 1930 until May 1932, a month before her death. She was a perfectly normal, quite average student, neither particularly outstanding nor much troubled. In those days before grade inflation, when C meant "good, 81–87" (as defined on her report card) rather than barely scraping by, Vivian Dobbs received A's and B's for deportment and C's for all academic subjects but mathematics (which was always difficult for her, and where she scored D) during her first term in Grade 1A, from September 1930 to January 1931. She improved during her second term in 1B, meriting an A in deportment, C in mathematics, and B in all other academic subjects; she was on the honor roll in April 1931. Promoted to 2A, she had trouble during the fall term of 1931, failing mathematics and spelling but receiving A in deportment, B in reading, and C in writing and English. She was "retained in 2A" for the next term—or "left back" as we used to say, and scarcely a sign of imbecility as I remember all my buddies who suffered a similar fate. In any case, she again did well in her final term, with B in deportment, reading, and spelling, and C in writing, English, and mathematics during her last month in school. This offspring of "lewd and immoral" women excelled in deportment and performed adequately, although not brilliantly, in her academic subjects.

In short, we can only agree with the conclusion that Dr. Lombardo has reached in his research on *Buck* v. *Bell*—there were no imbeciles, not a one, among the three generations of Bucks. I don't know that such correction of cruel but forgotten errors of history counts for much, but it is at least satisfying to learn that forced eugenic sterilization, a procedure of such dubious morality, earned its official justification (and won its most quoted line of rhetoric) on a patent falsehood.

Carrie Buck died last year. By a quirk of fate, and not by memory or design, she was buried just a few steps from her only daughter's grave. In the umpteenth and ultimate verse of a favorite old ballad, a rose and a brier—the sweet and the bitter—emerge from the tombs of Barbara Allen and her lover, twining about each other in the union of death. May Carrie and Vivian, victims in different ways and in the flower of youth, rest together in peace.

Stephen Jay Gould teaches biology, geology, and the history of science at Harvard University.