Los Alamos Science

LOS ALAMOS NATIONAL LABORATORY

NUMBER 16 – 1988

UNSOLVED PROBLEMS IN THE SCIENCE OF LIFE

The images on the cover symbolize the wide range of problems discussed in this issue:

Monkeys have been used to study the processing of visual information in the brain (Life Nature Library, *The Primates,* photograph by Yale Joel (c) 1965 Time-Life Books, Inc.).

A comet such as Comet West, photographed on March 8, 1976, by Dr. John W. Harvey with the Case Western Reserve/Kitt Peak National Observatory Burrell-Schmidt telescope, may have been responsible for one of our planet's catastrophic extinctions of life (National Optical Astronomy Observatories).

This Apollo 17 view of the earth (NASA) symbolizes the delicate fabric of physical coincidences that has led both to our existence and our consciousness of that existence.

Will we speak to other intelligent civilizations with, for example, this translation of binary code transmitted by the Arecibo radio telescope? Depicted (from the bottom moving upward) are the radio telescope, the sun and planets with the earth offset, a human form, the double helix of the DNA molecule, and the chemical formulae for the constituent compounds of DNA (Carl Sagan and Frank Drake).

It was an unusually stimulating day and a half at Los Alamos when two Nobel Laureates in physiology, a leading paleontologist, and a leading bio-astrophysicist came together to discuss "Unsolved Problems in the Science of Life," the topic of the second in a series of special meetings sponsored by the Fellows of the Laboratory. Just like the first one on "Creativity in Science," this colloquium took us into a broader arena of ideas and viewpoints than is our usual daily fare. To contemplate the evolution and mysteries of intelligent life from the speakers' diverse points of view at one time, in one place was indeed a rare experience.

George Wald began by reciting a litany of "accidents" of nature that have made life possible and juxtaposing these against our ignorance about the nature of consciousness. He then proposed a point of view heretical for a scientist—namely that consciousness or mind is ever present throughout the universe, operating as a complement to matter and causing the little "accidents" that are really no accident at all.

David Hubel took exception to this mystical approach and emphasized instead that mind or consciousness will eventually be understood by simply digging in and finding out one step at a time how the brain works. He then took us on a dazzling tour of the early stages of the visual system, demonstrating his many discoveries about the variety and specificity of the neural circuits responsible in part for form, movement, and color perception. Dr. Hubel commented before his talk that he had stuck needles into about 10,000 neurons in the course of his research. That is what he means by digging in!

Paleontologist Jack Sepkoski changed our focus from the mysteries of intelligent life to the mass extinctions of life that seem to have occurred periodically on this planet every 26 million years. While their causes remain a mystery, extinctions have clearly been an important factor in the evolution of new life forms.

Finally Frank Drake used a physicist's logic to convince us that we are not alone in the universe—but also that the economics of energy almost certainly precludes the chance of a visit from the extraterrestrials. Instead our best chance of contact is through radio signals, and such an effort is vigorously under way.

Well—the talks left everyone teeming with thoughts about the mysteries of the brain, the nature of consciousness, the fragility of our prominence on this planet, and the readiness of our culture to meet beings from another planet. We hope this issue devoted to the proceedings of the colloquium have a similarly stimulating effect on our readers.

The written versions of the talks were based directly on the transcripts and on the visuals provided by the speakers. We thank George Wald, David Hubel, Jack Sepkoski, and Frank Drake for their help in preparing this volume. We also thank the principal organizers of the conference, Art Cox, Ed Flynn, Carl Orth, and Mudundi Raju, for reviewing the transcripts and Mark Bitensky for moderating the morning of very lively discussion that followed the formal presentations. ■

Necia Grant Cooper

Los Alamos
Science
LOS ALAMOS NATIONAL LABORATORY

NUMBER 16

UNSOLVED PROBLEMS IN THE SCIENCE OF LIFE

PROCEEDINGS OF THE FELLOWS COLLOQUIUM JULY '86

FORMAL TALKS

Cosmology of Life and Mind
by George Wald 2

Does a complementarily principle of mind and matter underlie the very structure of our universe?

Vision and the Brain
by David H. Hubel 14

How does the brain use its complex web of neurological connections to process visual information? Fundamental circuit elements that respond to either shape, movement, or color are part of the answer.

Extinctions of Life
by J. John Sepkoski, Jr. 36

Was the transition from dinosaurs to mammals the result of a rare and random catastrophic event or was it just another in a series of periodic extinctions?

The Search for Extraterrestrial Life
by *Frank Drake* 50

Our galaxy may be populated with thousands upon thousands of intelligent and highly technical civilizations. If so, where are they?

PANEL DISCUSSION

Unsolved Problems-A Morning of Questions and Answers
moderated by Mark Bitensky 70

Los Alamos Science Staff

Editor
Necia Grant Cooper

Science Writers
Roger Eckhardt
Nancy Shem

Design, Illustration, and Production
Gloria Sharp
Katherine Norskog

Circulation
Dixie McDonald

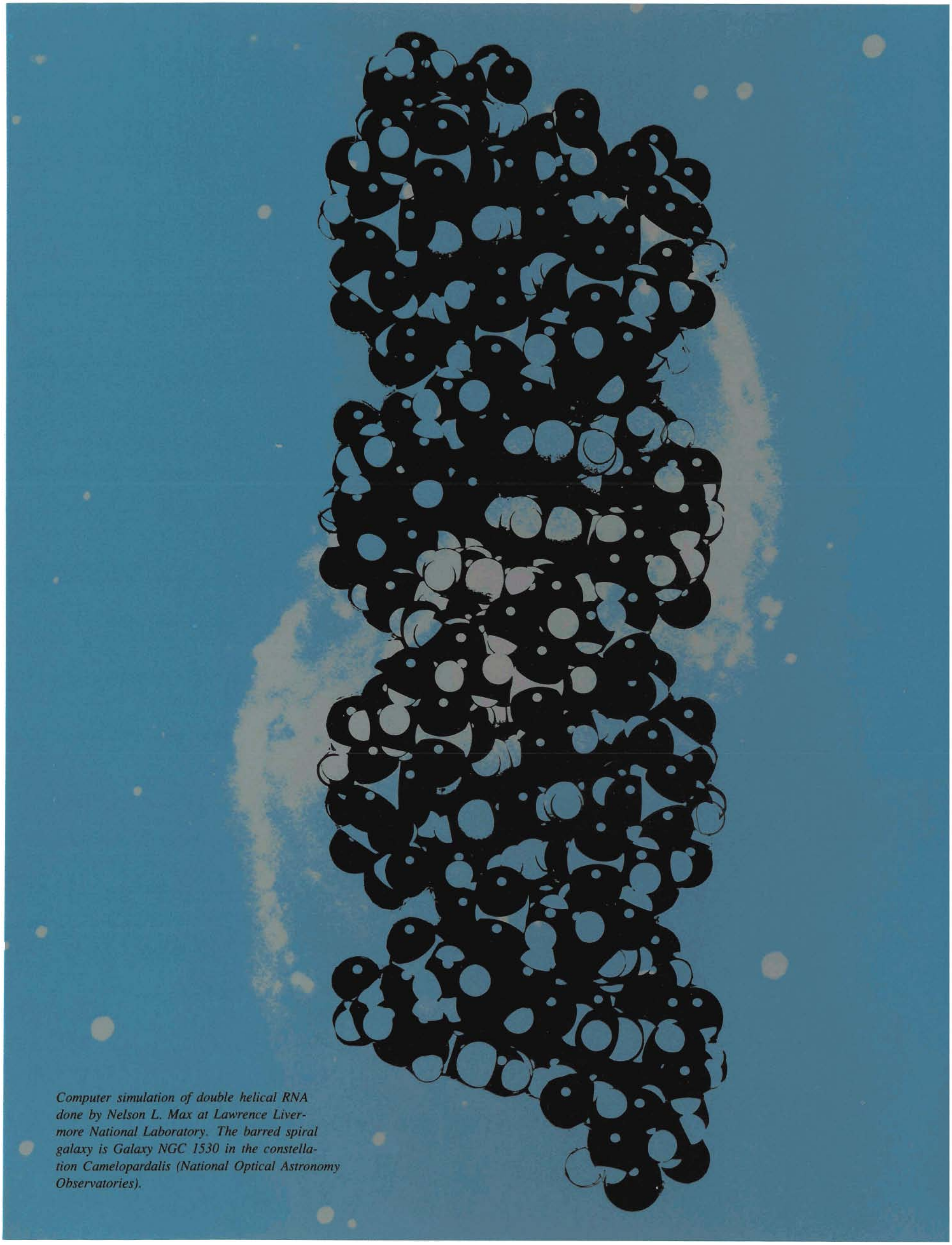
Other Contributors

IS-9 Photography Group

Printing
Jim E. Lovato
Guadalupe D. Archuleta

Address Mail to
Los Alamos Science
Mail Stop M708
Los Alamos National Laboratory
Los Alamos, New Mexico 87545

Los Alamos Science is published by Los Alamos National Laboratory, an Equal Opportunity Employer operated by the University of California for the United States Department of Energy under contract W-7405-ENG-36.



Computer simulation of double helical RNA done by Nelson L. Max at Lawrence Livermore National Laboratory. The barred spiral galaxy is Galaxy NGC 1530 in the constellation Camelopardalis (National Optical Astronomy Observatories).

Cosmology of Life and Mind

by George Wald

I am coming toward the end of my life as a scientist facing two great problems. Both are rooted in science, and I think I approach both as only a scientist would. Yet I also think that both problems are irrevocably unassimilable as science. That is not strange, because one involves cosmology and the other consciousness. I will begin with the cosmology.

We have realized for some time that we live in a *historical* universe, one in which not only living organisms but stars and galaxies are born, mature, grow old, and die. There is good reason to believe the universe to be permeated with life—a universe in which life arises, given enough time, *wherever* the conditions exist that make life possible.

How many such places are there? I like Arthur Eddington's old dictum: 10^{11}, or a hundred billion, stars make a galaxy, 10^{11} galaxies make a universe. Our own galaxy, the Milky Way, has about 10^{11} stars. With the earth nearing five billion people, a lot of us are feeling crowded—but a hundred billion stars are in the Milky Way. It is a vast thing: light, traveling at 186,000 miles per second, takes about 100,000 years to cross it from edge to edge. Yet vast though it is, the Milky Way is just a tiny spot in the universe we know. The lowest reasonable estimate of the fraction of stars in the Milky Way with a planet that could support life is 1 per cent, or a billion such places just in the Milky Way. With a billion such galaxies already in view of our telescopes, the lowest estimate of the number of places in the known universe that could support life is on the order of one billion billion, or 10^{18}.

My main theme is that if *any* of a large and increasingly recognized number of physical properties of this universe were different from what it is, life, which seems so prevalent, would be impossible, here or anywhere. I will outline a skeleton of this argument and give it structure by climbing the scale of the states of organization of matter. So I start with the elementary particles.

For the most part, our universe is made of four kinds of elementary particles: neutrons, protons, electrons, and particles of radiation called photons. (I leave out neutrinos, which interact only negligibly with matter, and also the hundreds of particles that come out of high-energy nuclear reactions.) The first three—protons, electrons, and neutrons—exist not only as particles but as antiparticles. The particles constitute matter; the antiparticles antimatter. If one looks at objects far out in the universe, one cannot be sure whether they are made of matter or antimatter, for all our information arrives via radiation, and photons do not differentiate. They are, as we say, their own antiparticles.

Why do we have a universe of matter at all? In 1952 I was giving the Vanuxem Lectures at Princeton University on the origins of life and biochemical evolution. Albert Einstein, whom I had come to know, was walking with me before the first lecture and asked, "Why do you think the natural amino acids are all left-handed?" As you know, all amino acids except the simplest, glycine, exist in two geometries that are mirror images of each other—like right and left hands. However, all the *natural* amino acids happen to be left-handed. Einstein went on to say, "I have wondered for years how the electron came out to be negative. Negative and positive are perfectly symmetrical principles in physics, so why is the electron negative?" All I could think of was: the negative electron won in the fight. I said, "That is exactly what I think of those left-handed amino acids—they won in the fight." But he was talking about a different fight—the fight between matter and antimatter. As he said, these types of matter are perfectly symmetrical. Thus, the neatest idea of what went into the big bang at the start of the known universe were equal amounts of matter and antimatter.

In the fantastic compression of the initial stages of the big bang, there must have been a wild fire storm. Whenever a particle of matter contacts its antimatter partner, mutual annihilation results and the masses of both particles are converted to radiation. Thus, at the end of the big bang there should have been a universe of radiation with *neither* matter nor antimatter. In fact, Arno Penzias and Robert Wilson of Bell Laboratories discovered a background of microwave radiation filling the universe that comes equally from all directions and is thought to be the residue of the fire storm in the big bang. The radiation is identical with the radiation that would come off a black body, say a piece of black iron, at the very cold temperature of 2.8 degrees above absolute zero, or approximately −270 degrees centigrade.

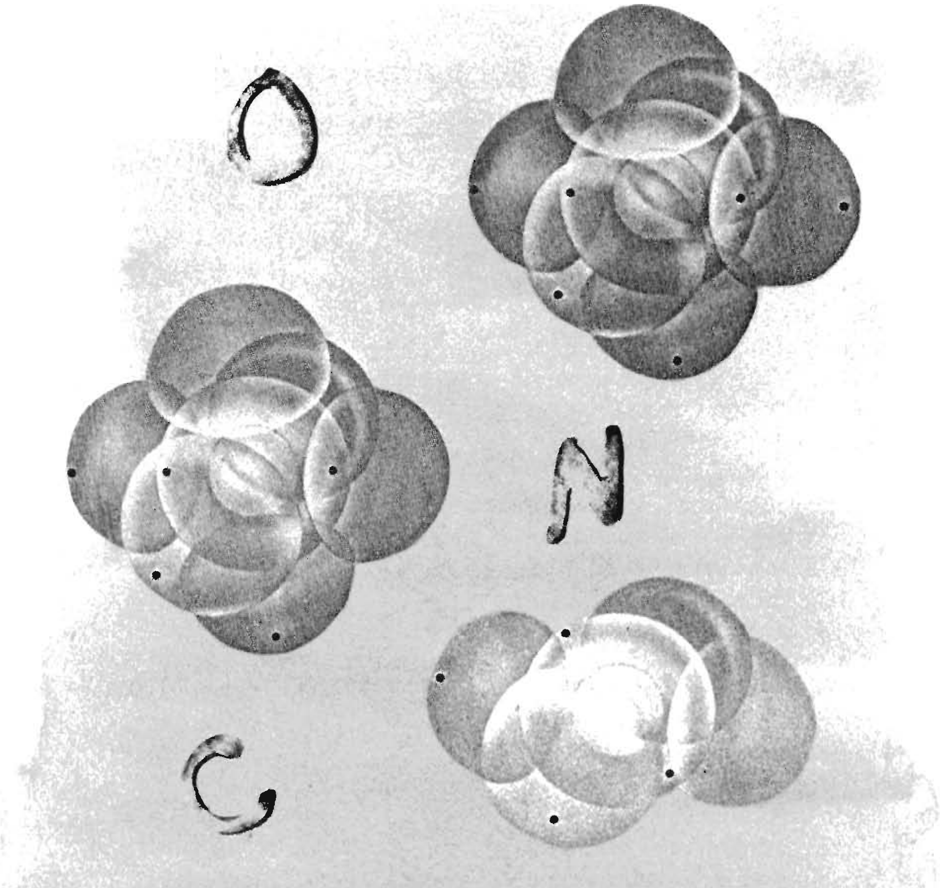
At the end of the big bang, there should have been a universe of radiation with neither matter nor antimatter.

One now realizes there are roughly a billion times as many photons of that residual radiation moving around in the universe as there are particles with mass. So we have to modify our neat idea to include a little discrepancy, a little mistake if you will: for every billion parts of antimatter involved in the big bang there were one billion and one parts of matter. Thus, when the fire storm of mutual annihilation had exhausted itself, one part in one billion of matter was left over. This residue constitutes the matter of our universe, that is, the galaxies and stars and planets and us. This little one part per billion mistake is the first element in my story.

Now how is it that we find ourselves in a universe well supplied with protons and electrons as well as neutrons? The reason is that free neutrons—neutrons outside of atomic nuclei and outside of highly dense neutron stars—disintegrate with a half life of 10.6 minutes into an electron, a proton, and radiation. If you start with a collection of free neutrons, ten minutes later half are still neutrons, but the other half is everything else you need to make a universe like ours.

Why does the reaction go in that direction? Only because a neutron is a tiny bit more massive than a proton plus an electron. Any such reaction has to go in the direction of lower mass. But the loss of mass in this case is less than one part in a thousand—in fact, eight parts in ten thousand. But what if the reaction went the other way? If it did, we would be in a universe of neutrons. The neutrons would have long since mopped up all the protons and electrons, and we would not have the chemical elements, molecules, new radiation, or, of course, life. Another small but vital discrepancy.

We need to consider two further properties of elementary particles: their masses and electric charges. The



A strange attribute critical to the properties of carbon, nitrogen, and oxygen is the fact that these are the only elements that form real double and triple chemical bonds.

nuclei of all atoms are made of protons and neutrons, which are heavy particles—each almost two thousand times the mass of an electron. The result is that almost the entire mass of an atom is concentrated in a nucleus that holds its position no matter what the electrons roaming around the periphery are doing. This fact is very important because it is the reason anything stays put in the universe. What would our universe be like if the nuclear particles and the electrons were somewhat closer together in mass? The motions of any one particle would produce reciprocal motions in the others; they would revolve around each other, and all matter would be fluid, none would be solid. Could indeed such atoms form stable bonds? You would not have molecules whose shapes you could draw with great confidence. This fact is critical because the shape of a molecule—the way one molecule fits into another—means everything in living organisms.

Here is another extraordinary circumstance. Although there is an enormous difference in mass between the proton and electron—one thousand eight hundred forty times—the magnitude of their electric charge is apparently identical. Why is it that the proton and the electron, which are so unlike in every other regard, have the same numerical charge?

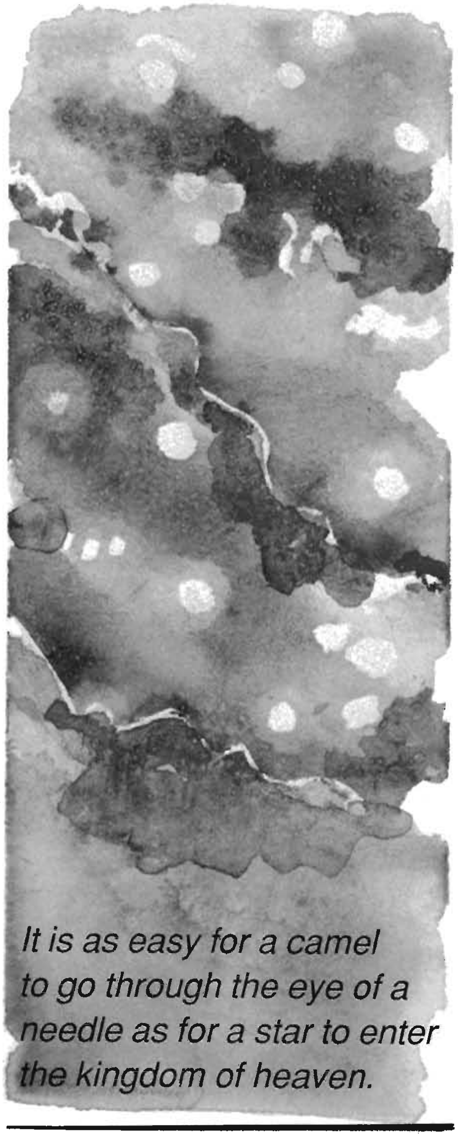
Is this a legitimate scientific question? In 1959 two of the world's most distinguished astrophysicists, R.A. Lyttleton and Herman Bondi, published a long paper in the *Proceedings of the Royal Society of London* in which they proposed that the proton and the electron differ in charge by the almost infinitesimal amount $2 \times 10^{-18}e$, where e is the tiny charge on either particle. One's first thought is who gives a damn about two billion billionths, but Lyttleton and Bondi explained that this tiny difference would result in a net charge on all particles, and thus there would be a net repulsion between all matter in the universe. Their hypothesis would account for the observed expansion of the universe. The only trouble I have with this idea is that the universe would do *nothing* but expand. Such a tiny difference in charge is enough to completely overwhelm the force of gravity that brings matter together, and so there would be no galaxies, no stars, no planets, and, worst of all, no physicists.

Before the ink was dry on that paper, John King and his group at the Massachusetts Institute of Technology were searching for a measurable difference in charge. By now they have shown that any difference has to be less than $10^{-20}e$. However, the growing consensus for the existence of quarks, which have *fractional* charge, has not made the equivalence of charge on the electron and proton any easier to understand. The electron is an indivisible unitary particle—an electron is an electron is an electron—whereas a proton consists of three quarks, two up and one down. It is a little strange that the sum of the quark charges is exactly equal to the charge of an electron.

Let us move up a step in organization to the elements. Of the 92 natural elements, 99 per cent of living matter is made of just four: hydrogen, oxygen, nitrogen, and carbon. I think it has to be this way wherever life arises in the universe because those four elements have unique properties critical to the existence of life. There are no other elements like them in the periodic table. Although I studied chemistry a long time ago, I suspect some of the same silly things are still being said. We were told that if you move vertically down a column of elements in the periodic table, those elements repeat properties. Well, any kid with a chemistry set knows better. Under oxygen is sulfur; try breathing sulfur sometime. Under nitrogen is phosphorus; there isn't any phosphorus in that kid's chemistry set because it is too dangerous: it bursts spontaneously into flames when exposed to air. Under carbon is silicon; there is about 130 times as much silicon in the crust of the earth as carbon. Then why are we made of carbon?

A strange attribute critical to the properties of these four elements is that carbon, nitrogen, and oxygen are the only elements that form real double and triple chemical bonds. What is the importance of this for life? Well, just compare two molecules that, based on the positions of their central atoms in the periodic table, should be very much alike: carbon dioxide and silicon dioxide. Carbon dioxide is a symmetrical molecule in which the carbon atom is tied to two adjacent oxygen atoms by double bonds. Those multiple bonds completely saturate the combining tendencies of all three atoms, and carbon dioxide can float off into the air as a perfectly happy and independent molecule and dissolve in the waters of the earth. Those are the places where living organisms find their carbon.

Silicon dioxide cannot form a double bond. Thus each silicon atom is tied to each oxygen with a single bond, leaving four half-formed bonds, or lone electrons, two on the silicon and one on each of the oxygens. These electrons are just dying to combine with something, but with what? Each silicon dioxide molecule combines with its neighbors until an enormous supermolecule has formed—in fact, a rock. The reason quartz is so hard to break is that you have to break a lot of chemical bonds. That is why silicon is fine for making rocks, whereas carbon is fine for making living organisms. One can make similar arguments for oxygen and nitrogen.

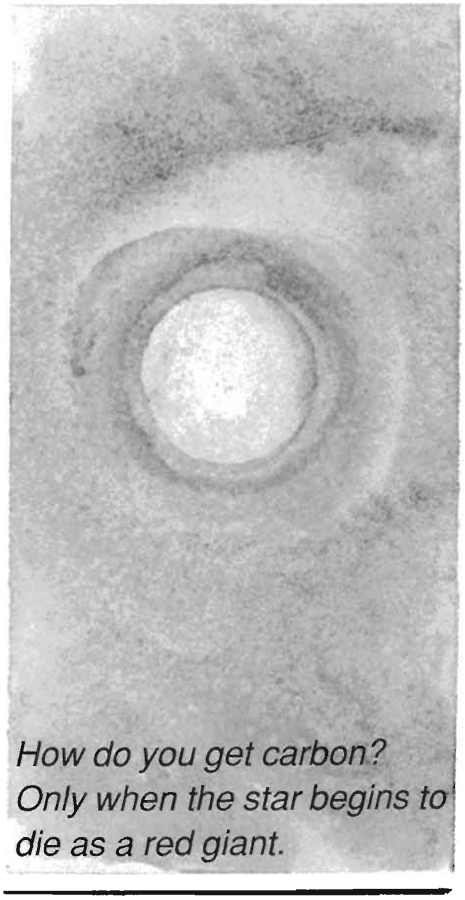


It is as easy for a camel to go through the eye of a needle as for a star to enter the kingdom of heaven.

Now we move up another step and examine molecular organization. The most important molecule, by far, in living organisms is water. But water is the strangest molecule in the whole of chemistry, and its strangest property is that ice floats. If ice did not float, I doubt there would be life. Everything contracts on cooling, including water down to 4 degrees centigrade. However, between 4 degrees and the freezing point at 0 degrees, water expands so rapidly that ice is less dense than liquid water, and it floats. If water shrank as it cooled like everything else, colder water would be heavier and would keep sinking. Freezing would begin not at the top of the lake or ocean but at the bottom, and, in the end, the body of water would freeze solid, a disaster for underwater life. Where I live the best time to go fishing is in the winter. You take your fishing equipment in one hand and a bottle of whiskey in the other and cut yourself a hole in the ice. Up to that point the fish were having a ball, getting along fine down there. Another problem that would arise if large bodies of water froze solid is that a big chunk of ice takes forever to melt. With a relatively thin skin of ice on top, the first warm weather melts it, spring arrives, and everything is happy again.

Now I take a big jump to the stars. It is as easy for a camel to go through the eye of a needle as for a star to enter the kingdom of heaven. The needle's eye in this case is the first step in the fusion of hydrogen to helium. Every main-sequence star lives by fusing hydrogen to helium. A physicist at Oak Ridge during the Manhattan project who became an administrator and then an Episcopal priest was once quoted in the *New Yorker* as having said, "God must love hydrogen bombs because He made so many of them in the form of stars." The man should have known better, both as a physicist and a priest, because you can make stars out of hydrogen but you cannot make hydrogen bombs out of hydrogen. You have to use the rare, heavy isotopes of hydrogen in bombs. A mixture, say fifty-fifty of deuterium and tritium, is needed because the conversion of ordinary hydrogen to deuterium is perhaps the slowest reaction known. It takes a hundred billion years, which is the only reason stars last so long. They are *not* hydrogen bombs, although once you get to deuterium even a star could explode. As a result, stars last a long time, and life has a chance to start evolving at those with suitable planets.

Why is the conversion of hydrogen to deuterium so slow? The nuclei of normal hydrogen are simply positively charged protons, and even at the temperatures of main-sequence stars, say around five million degrees, the collision of two protons will most likely result in their just bouncing off each other. The rare event that has to occur if such a collision is to generate deuterium is for one of the protons to disintegrate and change to a neutron *as* it collides with the other proton. That is an improbable event. But main-sequence stars have lots of time and just keep slowly turning sets of four hydrogen nuclei into pairs of deuterium nuclei and then into helium nuclei. The slight loss of mass in the reaction is turned into radiation, which is our sunlight.



How do you get carbon? Only when the star begins to die as a red giant.

How do you get carbon? The first thought is just to keep adding protons. This will not work because if one proton is added to helium, the result is a mass-five isotope, and there is no atomic nucleus with mass five. What is the path around this barrier? Well, the only alternative is to fuse helium nuclei, but that reaction requires a very much higher temperature, say a hundred million degrees, which is only achieved when the star begins to die as a red giant. When the core of a red giant gets that hot, the helium nuclei begin to fuse.

From this point on it should just be simple arithmetic, but there is another barrier. When two helium nuclei fuse, the result is a mass-eight isotope of beryllium, which is one of the most unstable atomic nuclei to exist, disintegrating in 10^{-16} second. Fortunately again, there just happens to be an excited state of the carbon-12 nucleus whose energy is equal to the energy in a beryllium-8 nucleus *plus* a helium-4 nucleus *plus* the kinetic energy at the temperature at which these nuclei can collide. This wild coincidence is a fortunate energy resonance that turns a very improbable reaction into a very efficient one. So beryllium-8 fuses with helium-4 to make carbon-12. The important point is that there are multiple barriers in the synthesis of the elements, but each of these barriers is overcome in a very ingenious way.

Once carbon is formed in a red giant, two protons can be added to the carbon-12 nucleus to give mass 14, which brings nitrogen into the universe. Add helium-4 to the carbon-12 and you have mass 16, which brings oxygen into the universe. The story goes on and on this way, but eventually such stars grow unstable and explode, sending their material off into space. Finally, suns and planets such as ours grow out of this material.

Eventually red giants grow unstable and explode, sending their material off into space. Suns and planets such as ours grow out of this material.

Now just think! Life, wherever it arises in the universe, has to invent a way to keep going, and that way must depend on the energy given off by a nearby star. As we know, life on the earth runs on sunlight through the process of photosynthesis. How do we get our sunlight? We get it from the various reactions of the elements that constitute life itself. The first way is to fuse hydrogen to helium—the proton-proton chain. The second way uses a catalytic process—the carbon-nitrogen-oxygen cycle—which starts by fusing carbon with 2 protons to yield nitrogen-14, then picks up 2 more protons to give oxygen-16, then splits the oxygen nucleus into helium and a carbon nucleus. The net result of both processes is exactly the same: four hydrogens have been turned into a helium. The four elements—carbon, nitrogen, oxygen, and hydrogen—that are the chief constituents of life on the earth are also vitally important to the source of energy that supports that life. Along with helium, these four are the most plentiful elements in the universe.

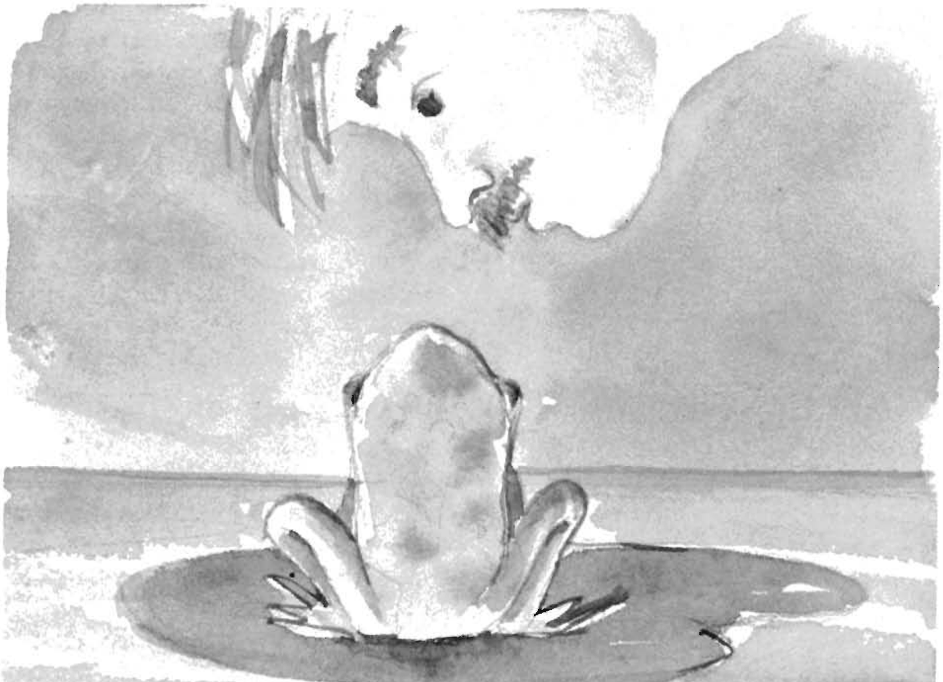
The last cosmic element in my story is equally strange, but was worked out by one of the brightest physicists alive, Stephen Hawking. There are two great forces operating in the universe: the force of dispersion and expansion powered by the big bang and the force of aggregation powered by gravity. It is all very strange because the forces are *exactly* in balance in our universe. You would think the ratio of the two could be anything, but they are *exactly* equal.

Hence we find ourselves in a very strange universe that, as a whole, is expanding but that also has islands here and there within which gravity is holding things together. For example, our own galaxy, the Milky Way, is in a rather smallish local cluster with the Andromeda galaxy and some smaller galaxies. Within our cluster there is no expansion. Our knowledge about the expanding universe comes, of course, from measurements of Doppler shifts of the light from distant sources. In general, the farther out you look the redder is the shift, indicating an overall expansion. However, the *first* spectral shift ever observed, by the American astronomer Slipher back in 1912, was not a red shift. He was looking at Andromeda in our local cluster, and he observed a *blue* shift because Andromeda is moving *toward* us at about 125 kilometers per second. To see the red shift from the earth you have to look beyond our local cluster, out to a radius of about two million light years, to where everything is expanding.

Now what if the two forces were *not* in balance, and gravity dominated instead? Our universe would still begin with a big bang, but gravity would slow the expansion until eventually the universe reached a limit. What would then follow would be the big crunch, which would either allow insufficient *time* for life to arise and evolve or would all too quickly destroy it. On the other hand, say the forces of dispersion dominated. Then matter would just fly apart without aggregation, and there would be no stars or planets. There would be no *place* for life. Fortunately, the two forces are in exact balance.

Let me summarize the first problem. We find ourselves in a universe of prolific, abundant life, but the only way this seems possible is for it to be a very peculiar universe. Any imaginative intelligence can dream up many alternative universes, any of which could be a fine, stable, but *lifeless* universe. Our living universe is a very particular universe in that the more one knows of its physics the more one sees how finely balanced and intricately meshed it is—as if it were *intended* to breed life. The fact that so many barriers and problems are solved so precisely seems pretty strange. Of course, from our self-centered point of view, these particular solutions represent the best way to make a universe. But what I want to know is how did the universe find that out? Which brings me to my next problem, that of consciousness.

For me the problem of consciousness was unavoidable because I have spent most of my scientific life working



If you do everything you can with frog retinas and human retinas, you find they are fantastically alike. Now I know that I see. Does a frog see?

on vision. I learned my business on the retinas of frogs. If you do everything you can with frog retinas and human retinas, you find they are fantastically alike. In both cases there are two types of receptors—rods and cones—there are three primary cell layers, and the chemistry of visual pigments is very much alike. Everything is very much alike.

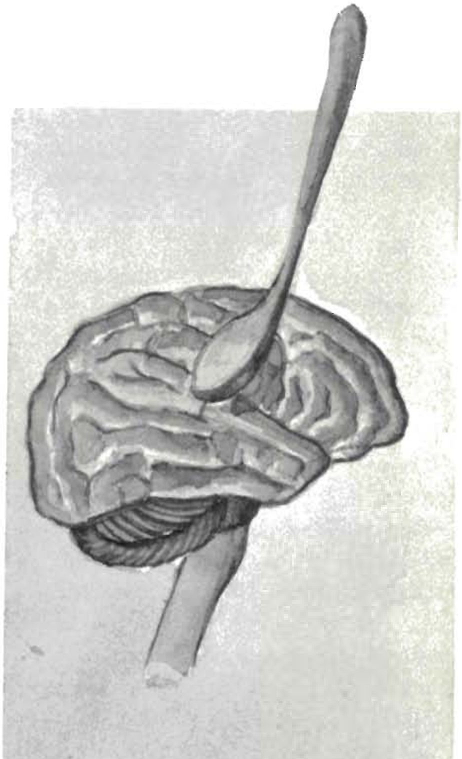
But I know that *I see*. Does a frog see? It reacts to light, but so does a photoelectrically activated garage door. Does the frog know it is reacting? Is it self-aware? There is nothing I can do as a scientist to answer that kind of a question. Absolutely nothing. So during the time I worked on all kinds of animal eyes, this problem was lurking in the background. I was occupied with easier questions then, but now the problem has moved to the foreground. Let me tell you what I can about it.

Although there is nothing I can do as a scientist to identify either the presence or absence of consciousness, I am quite convinced when I deal with another person that he *is* conscious because he is so much like me. The only primary data are what transpires in my own consciousness. Yet another person also gives all kinds of evidences of consciousness—especially in that he tells me about it.

I think that probably all mammals are conscious. I think that probably birds are conscious—why else would they sing? Frogs I worry about, fish even more. I worked on the beautiful iridescent eyes of scallops, which have about eighty eyes. They are perhaps the most complex eyes anatomically in the animal kingdom and are magnificent. But I never saw any indication that a scallop *used* its eyes. I also worked on the eyes of worms from the Bay of Naples. These are worms that live in warm seas with great bulging eyes that have everything you could ask for in an eye, but, again, there was no indication that they saw anything. We could find no behavioral responses to light at all.

Although I may think my dog has consciousness, there is nothing I can do as a scientist to bolster such thoughts to the level of evidence. How about the garage door: Does it resent opening when I flash it a signal? I think not. Does a computer that just beat a good human chess player feel elated? Again, I think not. But there is nothing I can do about obtaining evidence for those conclusions either. Consciousness gives us no signals—none. Not even a signal that it is present, let alone *what* is in it. That is the problem.

Now since consciousness gives us no signals, it is an embarrassment to scientists, particularly biologists. Biologists are made very uncomfortable by the subject, because they have always thought, as I did, that consciousness is a property of higher organisms and, as such, ought to be something they know about and can explain, at least partially, to other scientists. But they have nothing to say. It is an embarrassment. One way out is to declare that consciousness is nonexistent. P. W. Bridgman, for example, once said that consciousness was just a way of talking. He believed that for anything to be real it had to have an operational definition. There are no operations that define consciousness. In the same discussion the behavioral psychologist B. F. Skinner held that consciousness is in a private world, whereas science is in a public world, so consciousness cannot enter science, and we can forget about it. The difficulty is: no consciousness, no science, no reality. It is not some iffy phenomenon that we just project on reality; it is at the base, at the foundations.



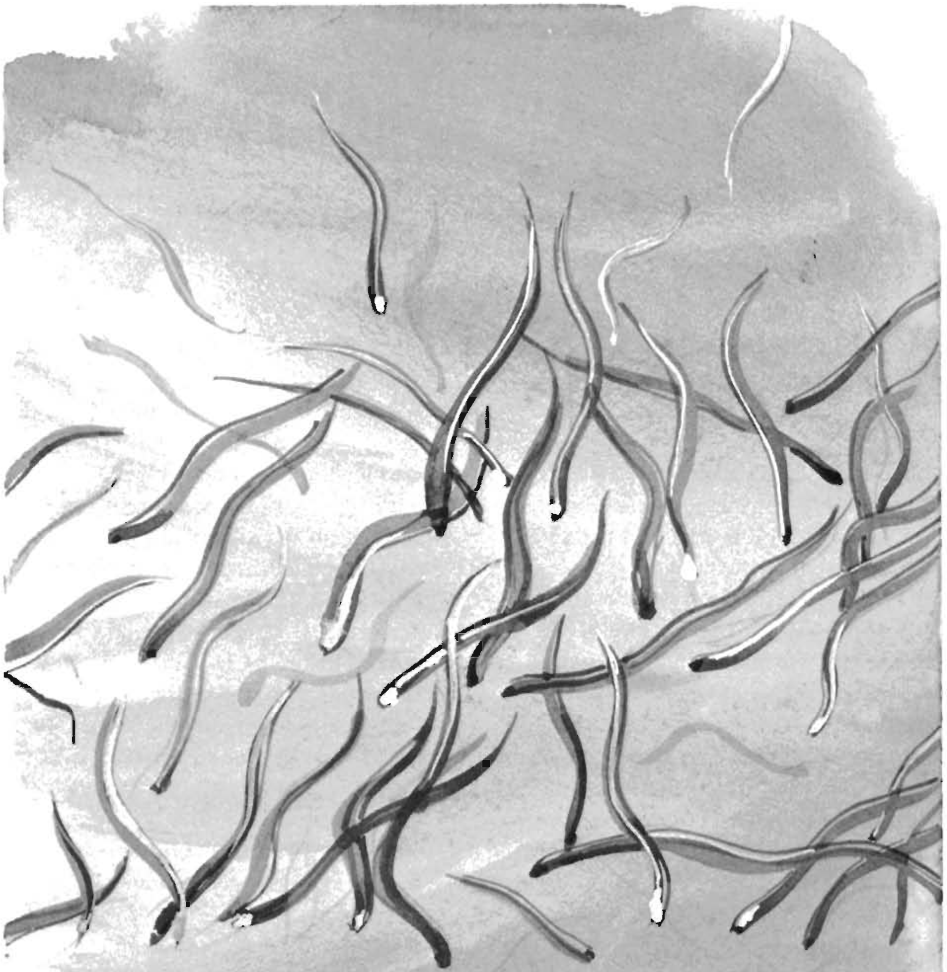
Since consciousness gives us no signals, it is an embarrassment to scientists, particularly biologists.

Now I want to raise a strange question. Since consciousness is not definable and gives no signals, where is it? The famous brain surgeon Wilder Penfield from McGill University in Montreal had absolutely unique opportunities to work with the exposed brains of unanesthetized patients. The exposed brain, by the way, feels no pain, and on one occassion Wilder said to me that once the brain is exposed he could operate on it with a spoon. Wilder was exploring the human brain for therapeutic purposes, and always for the sake of the patient, but, among other things, he searched for the center of consciousness. During one discussion with him, I asked why he thought consciousness was in the brain? He chuckled and said, "Well, I'll keep on trying." Then a couple of years later I met him again and he said, "I'll tell you one thing, it's not in the cerebral cortex."

Sometime later people became interested in the reticular formation located in the brain stem of mammals. This part of the brain is an arousal center, and, for a while, people were saying that *it* is the center of consciousness. Incidentally the reticular formation is very low down in the brain stem. The next lower part of the system is the spinal cord.

The trouble with all such arguments is that it is analogous to pulling one of the transistors out of your TV set and saying that the transistor is the source of the program because the program stopped. The reality is that the process takes a lot of machinery and it is hard to know if you are dealing just with some of the machinery or a real source.

The problem, however, is deeper than just having trouble finding the center of consciousness. How can you talk about the location of something that gives no identifiable physical signals. It is absurd. Consciousness has no location. The problem is similar to the controversy that, for a time, surrounded Heisenberg's uncertainty principle. The question was whether technology—the measurement process—was failing or whether reality was just that way. Most physicists now agree with Bohr that, yes, that is the way reality is. You can't specify the position and motion of an electron because *it doesn't have* a specific position and motion. That is the way it is with consciousness. It has no location.



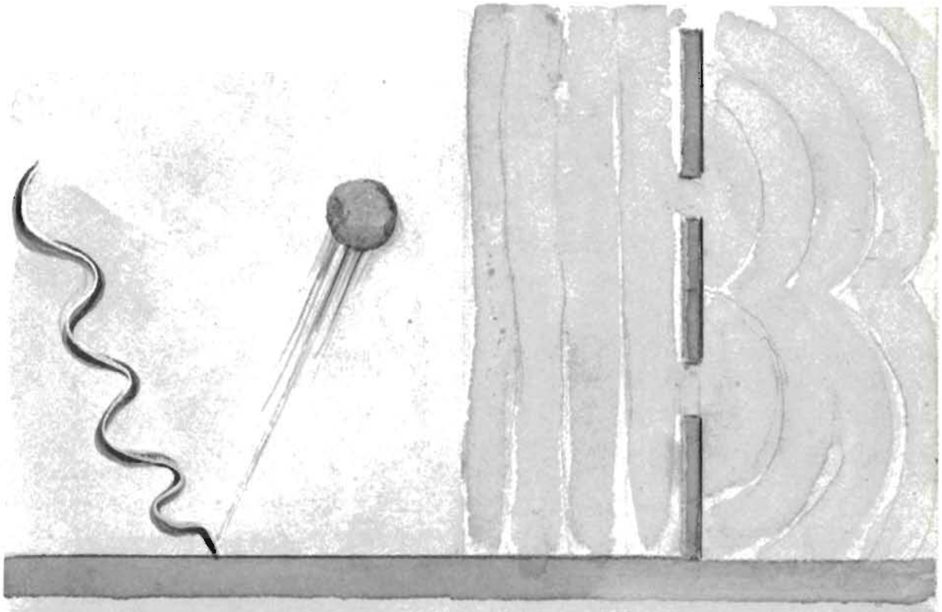
It takes the European eels three years to get back home, but there is as yet no record of a baby eel ever getting balled up and coming to the wrong continent.

A few years ago it occurred to me that these two problems—a universe that breeds life by overcoming obstacles with many special tricks and a consciousness that has no location—could be put together. At the time I was both elated and embarrassed. I was embarrassed because the thought seemed so strange to me as a scientist. But I was also elated because, as an experimentalist, I have learned that if an experiment gives you a beautiful result, enjoy it! Heaven knows whether such results will ever happen again. At any rate, within a couple of weeks, I realized that I was in the best of company.

What was the thought? Previously I had always thought of consciousness, or mind, as something that required a particularly complex central nervous system and was present only in the highest organisms. The thought now was that mind had been there all the time, and the reason this is a life-breeding universe is that the pervasive, constant presence of mind had guided the universe that way.

I was once talking to Bohr, when, to my amazement, he told a story about the love life of eels, which I think may help illustrate what I am now trying to say. Bohr's father, Christian Bohr, was a very fine physiologist, and Bohr had a great interest in biology. There are certain so-called freshwater eels that grow in fresh water for five to fifteen years but, on reaching sexual maturity, leave and migrate into the ocean. At this point they will never eat again. At best they are excellent food for us, since they are all good muscle. There are two species in the Atlantic that come, respectively, from the European and the American shores, but both migrate to overlapping areas in the South Atlantic close to Bermuda. This region is the deepest and saltiest part of the ocean, and it is where the eels spawn at great depths and die. All of them die, but the larval eels make their way back alone to their freshwater homes. It takes the American eels about fifteen months to reach our shores and come up the rivers. It takes the European eels *three years* to get back home, but there is as yet no record of a baby eel ever getting balled up and coming to the wrong continent. Bohr told all this and then said a wonderful thing: "It is just because they do not know where they are going that they always do it perfectly."

As you can see, I feel that our growing scientific knowledge—whether it be



One has as little reason to ask for the presence of matter without its complementary aspect of mind as to ask for particles that are not also simultaneously waves.

about the actions of elementary particles or the actions of eels—points unmistakably to the idea of a pervasive mind intertwined and inseparable from the material universe. This thought may sound pretty crazy, but such thinking is not only millennia old in the eastern philosophies but arose again and again among the monumental generation of physicists in the first half of this century.

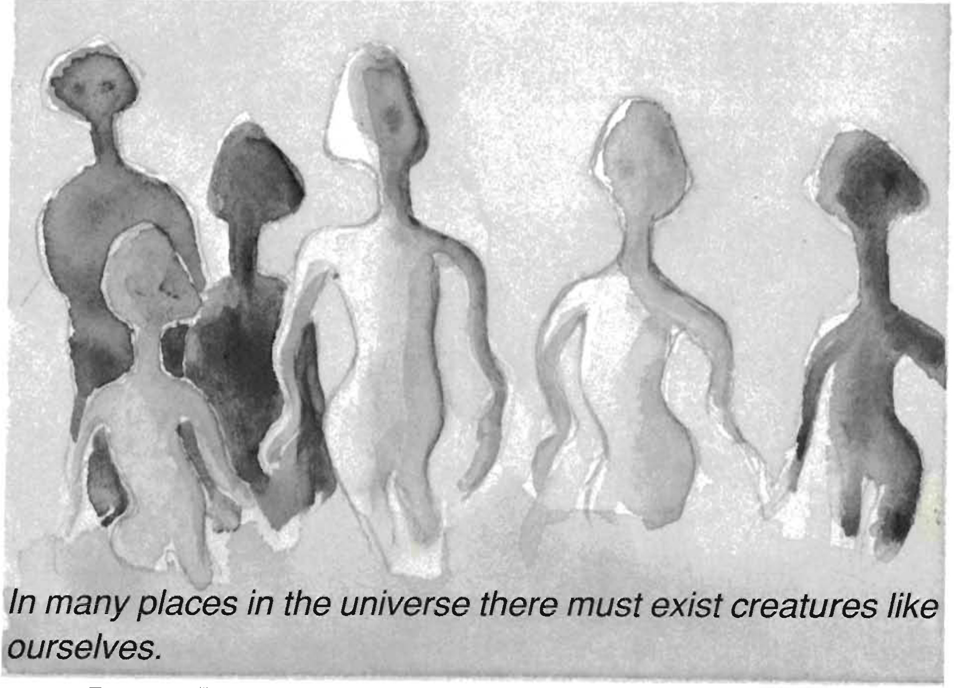
In 1928 Eddington said, "The stuff of the world is mind-stuff ... The mind-stuff is not spread in space and time. It is not something that you are going to get into science ... Recognizing that the physical world is entirely abstract and without 'actuality' apart from its linkage to consciousness, we restore consciousness to the fundamental position ..." Eddington was a pretty good physicist in his time.

Von Weizsäcker talked about his Identity Hypothesis, which he felt to be a unique but intelligible interpretation of quantum theory. "Consciousness and matter," he said, "are different aspects of the same reality."

However, the quote I like best is that of Wolfgang Pauli, who said, "To us ... the only acceptable point of view appears to be the one that recognizes *both* sides of reality—the quantitative and the qualitative, the physical and the psychical—as compatible with each other, and can embrace them simultaneously. It would be most satisfactory if *physis* and *psyche* (i.e., matter and mind) could be seen as the complementary aspects of the same reality." Just realize what Pauli is saying to us: one has as little reason to ask for the presence of matter without its complementary aspect of mind as to ask for particles that are not also simultaneously waves.

Athough this matter of mind embarrasses biologists, it is much easier to talk with physicists about it because they tend to deal with mind, day in and day out. The nineteenth-century scientists harped on the idea of an external world that can be observed without disturbing it. That world was truly objective because one could experiment without entering or affecting the part of the world being observed. However, at the very center of modern physics is the realization that you cannot keep yourself out of the experiment, and, in fact, all scientific observations are ultimately subjective.

There is a simple example of the entry of consciousness into physics experiments. Any physicist setting up an experiment on radiation, or elementary particles for that matter, must decide beforehand which set of properties—particle or wave—they intend to find. If a wave experiment is set up, they get a wave answer. If a particle experiment is set up, they get a particle answer. One cannot get both answers in one experiment.



In many places in the universe there must exist creatures like ourselves.

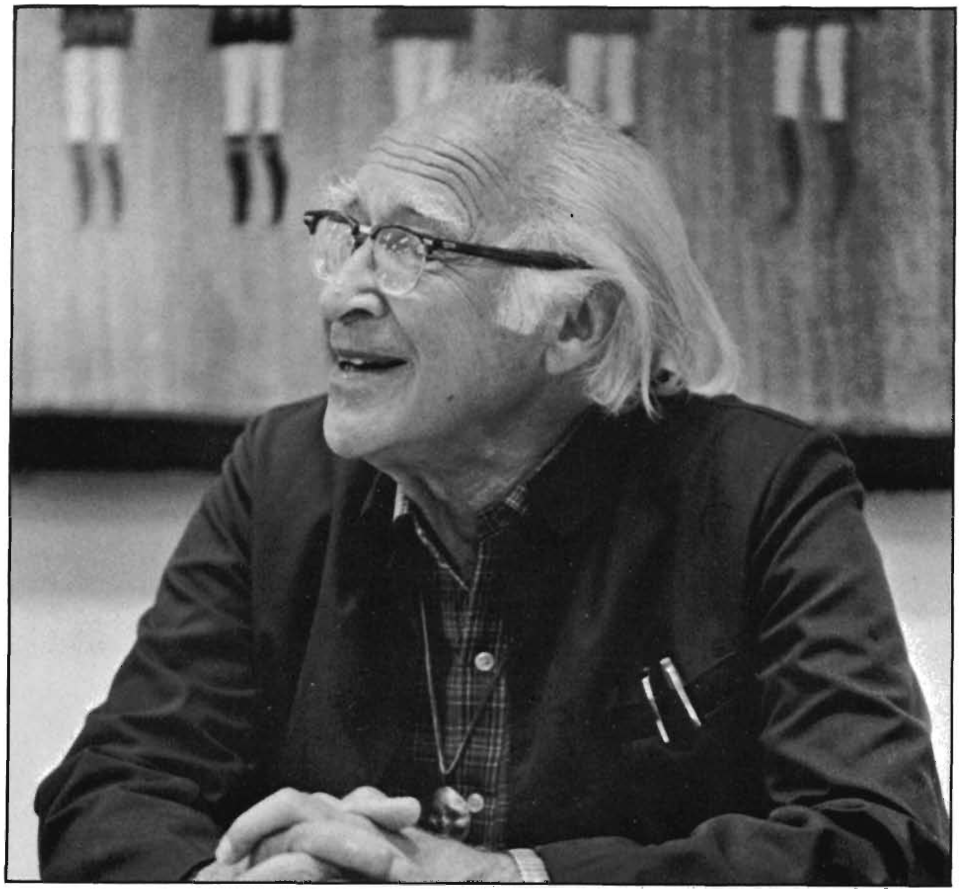
I think we live in a world of chance—without chance there are no phenomena—but not a world of accident. The universe has this weird fitting together. Arriving at this point of view I ask myself, what for? If mind was there all the time why would it take the trouble to make matter? One possible answer is, of course, at the heart of the anthropic principle, which, briefly, is that the universe has a design that makes it certain there will be physicists.

The driving force of evolution, according to Charles Darwin, is what he called natural selection. It has three components: the constant production of variations, both advantageous and disadvantageous; a mechanism for inheritance; and a competitive element. As a result, variations that work better are retained whereas variations that work less well are discarded.

In many places in the universe there must exist creatures like ourselves. By this I do not mean they are like us anatomically—former creatures on the earth were different anatomically from the current ones. But they would be like us in the creation of art, science, and technology. In some of these places they should have developed far beyond us. After all, what is ten million years in cosmic time? Such creatures form societies and invent languages and writing that form mechanisms for cultural inheritance. Those creatures make cultures, and those cultures are constantly pouring out variations, advantageous and disadvantageous. With libraries and educational systems, each generation does not have to start from scratch as regards its culture. Then there is the competition of cultures. Some rise, flourish, then disappear; yet aspects of that culture may be retained because they work better.

So one has a new kind of natural selection and a new mechanism of evolution that does not replace but rather adds to the ongoing anatomical and physiological evolution. This new phase of evolution now includes means for the independent evolution of consciousness. The prospect of this independent evolution of the pervasive, ever-present mind gives our species a transcendent worth and dignity and tells us our place in the universe: it is to know and create, and to try to understand, as we alone can do under our sun. ■

George Wald received his Ph.D. in zoology from Columbia University in 1934 and then joined Harvard University, where he has been ever since. He was the first to identify vitamin A in the retina, and most of our knowledge today regarding processes by which retinal pigments in the human eye convert light into sight comes from his work and that of his associates. These discoveries have had a profound effect on sight restoration of children, especially children in developing countries where blindness is, unfortunately, a common problem. Among Dr. Wald's many prestigious awards and honors is the 1967 Nobel Prize in Physiology or Medicine.



Further Reading

R. J. Tayler. 1979. The neutron half-life and cosmology. *Nature* 282:559–560.

L. N. Bondarenko, V. V. Kurguzov, Yu. A. Prokof'ev, E. V. Rogov, and P. E. Spivak. 1978. Measurement of the neutron half-life. *JETP Letters* (english translation of *Pis'ma v Zhurnal Eksperimental' noi i Teoreticheskoi Fiziki*) 28:303–307.

Fred Hoyle. 1975. *Astronomy and Cosmology: A Modern Course*, pp. 400–402 and p. 603. San Francisco: W. H. Freeman and Company.

R. A. Lyttleton and H. Bondi. 1959. On the physical consequences of a general excess of charge. *Proceedings of the Royal Society of London, Series A* 252:313–333.

John G. King. 1960. Search for a small charge carried by molecules. *Physical Review Letters* 5:562–565.

H. Fredrick Dylla and John G. King. 1973. Neutrality of molecules by a new method. *Physical Review A* 7:1224–1229.

George Wald. 1964. The origins of life. *Proceedings of the National Academy of Sciences* 52:595–611.

E. Margaret Burbridge, G. R. Burbridge, William A. Fowler, and F. Hoyle. 1957. Synthesis of the elements in stars. *Reviews of Modern Physics* 29:547–650.

H. A. Bethe. 1939. Energy production in stars. *Physical Review* 55:434–456.

C. B. Collins and S. W. Hawking. 1973. Why is the universe isotropic? *Astrophysical Journal* 180:317–334.

R. H. Dicke and P. J. E. Peebles. 1979. The big bang cosmology—Enigmas and nostrums. In *General Relativity: An Einstein Centenary Survey*, edited by S. W. Hawking and W. Israel, pp. 504–517. Cambridge: Cambridge University Press.

John Archibald Wheeler. 1974. The universe as home for man. *American Scientist* 62:683–691. (Also in *The Nature of Scientific Discovery: A Symposium Commemorating the 500th Anniversary of the Birth of Nicolaus Copernicus*, edited by Owen Gingerich, pp. 262–295. Washington, D.C.: Smithsonian Institute.)

Virginia Trimble. 1977. Cosmology: man's place in the universe. *American Scientist* 65:76–86.

George Wald and Stephen Rayport. 1977. Vision in annelid worms. *Science* 196:1434–1439.

D. R. Griffin. 1981. *The Question of Animal Awareness*. New York: Rockefeller University Press.

P. W. Bridgman. 1959. *The Way Things Are*, pp. 215–219. New York: Viking Press.

Eugene P. Wigner. 1967. Two kinds of reality. In *Symmetries and Reflections: Scientific Essays of Eugene P. Wigner*, pp. 185–199. Bloomington: Indiana University Press.

Erwin Schrödinger. 1958. *Mind and Matter*, p. 1. Cambridge, England: At the University Press.

George Wald. 1958. Foreword to *The Fitness of the Environment*, by L. J. Henderson, pp. xxiii–xxiv. Boston: Beacon.

A. S. Eddington. 1928. *The Nature of the Physical World*, pp. 276–277 and p. 332. New York: The Macmillan Company; Cambridge, England: At the University Press.

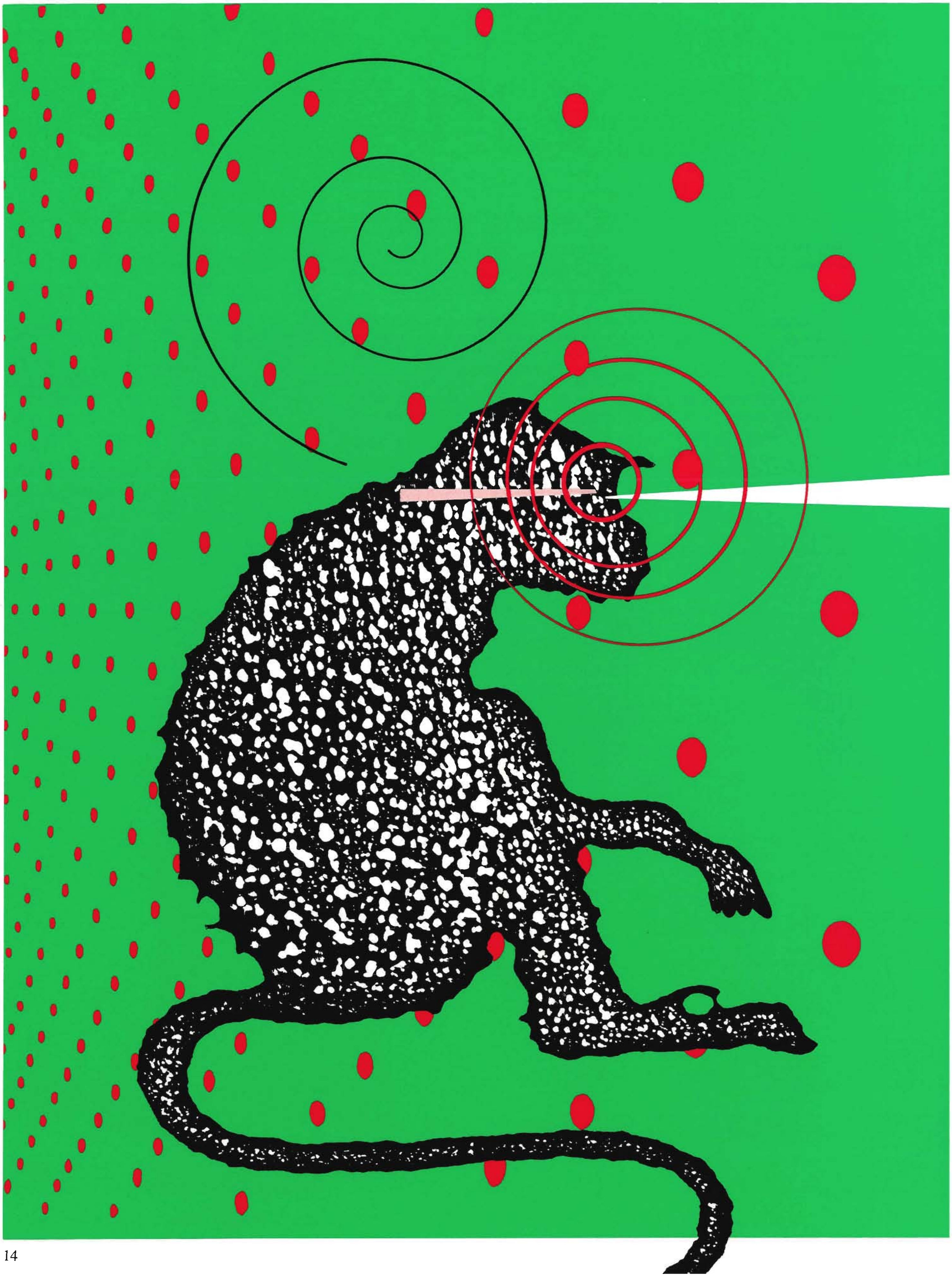
C. F. von Weizsäcker. 1980. *The Unity of Nature*, translated by F. J. Zucker, p. 252. New York: Farrar, Straus, Giroux.

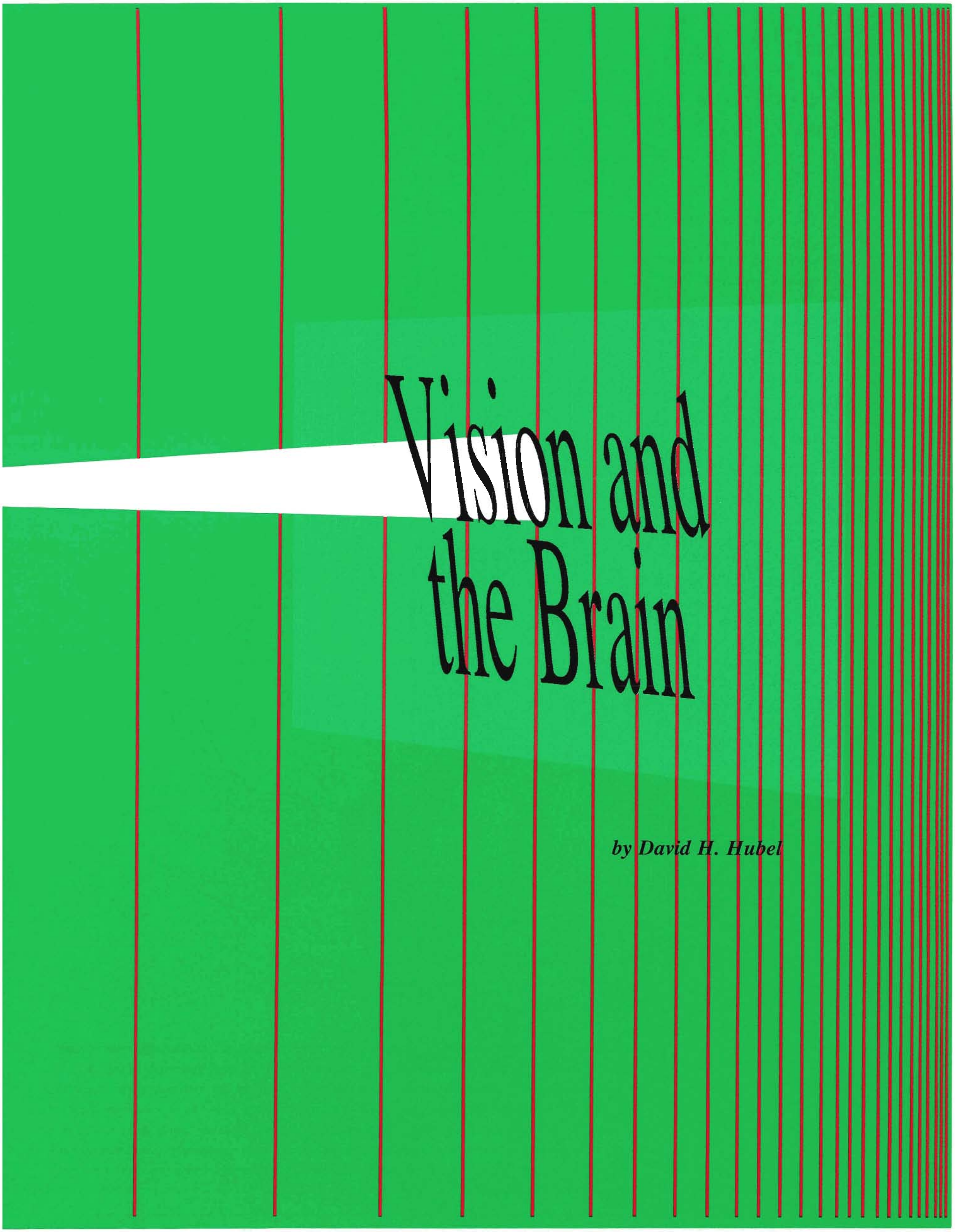
Wolfgang Pauli. 1955. In *The Interpretation of Nature and the Psyche*, edited by C. G. Jung and W. Pauli, pp. 208–210. New York: Bollingen.

R. H. Dicke. 1961. Dirac's Cosmology and Mach's principle. *Nature* 192:440–441.

B. Carter. 1974. In *Confrontation of Cosmological Theories with Observation*, edited by M. S. Longair, p. 291. Dordrecht: D. Reidel Publishing Company.

B. J. Carr and M. J. Rees. 1979. The anthropic principle and the structure of the physical world. *Nature* 278:605–612.





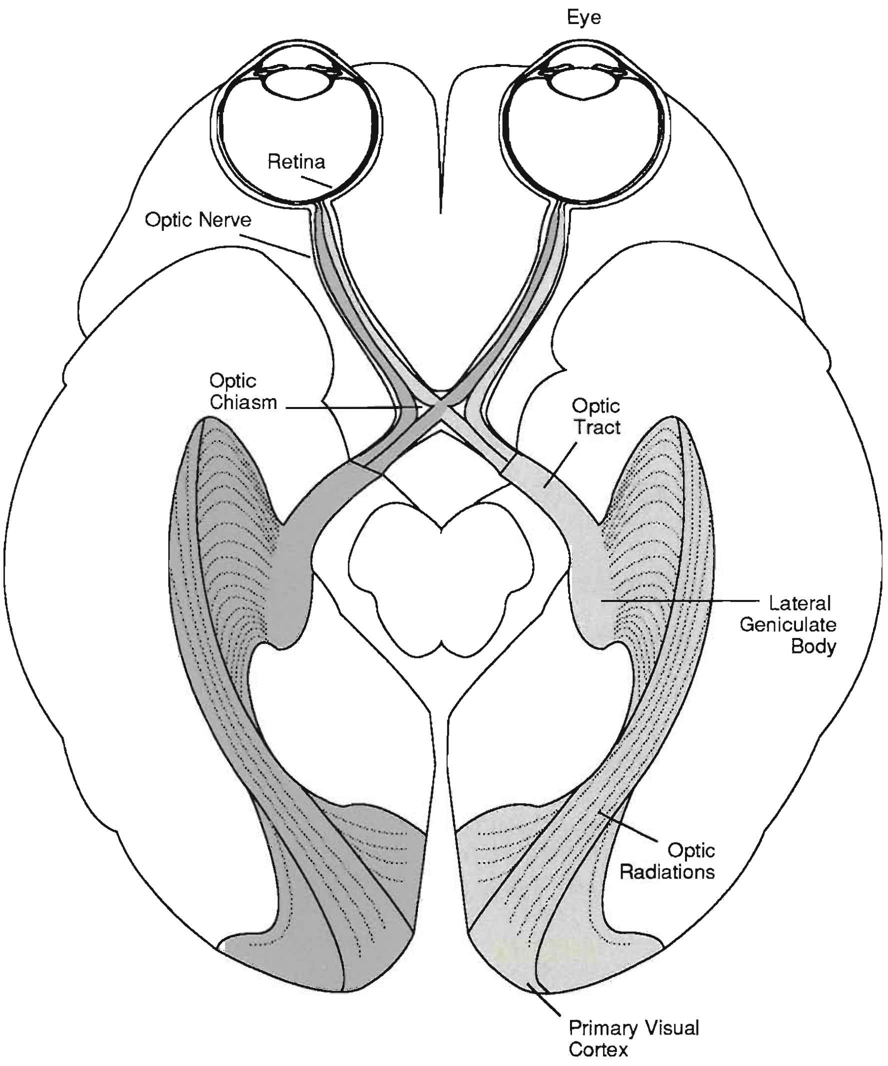
Vision and the Brain

by David H. Hubel

Let me begin by stating that we all know the brain is complex. This statement is so obvious as to be almost insulting. But why do we think it is complex?

I can think of several reasons. One is that the brain does many complex things, or so we like to think. We walk, we talk, we perceive, we play the piano or the violin, and we do things like cough, vomit, and sneeze. Another reason is that the brain contains 10^{11} cells—give or take a factor of 10—literally an astronomical number. But anyone who has seen the size of a human liver and has looked at any part of it under a microscope will probably guess that it too has 10^{11} cells. There might be five basic types of cells in the liver, with each cell of a given type doing more or less the same thing. Nevertheless even the most committed hepatic physiologist would never suggest that the liver is more complicated than the brain. The number of cells obviously doesn't tell the story. If on the other hand you are aware that each brain cell, on the average, transmits information in the form of nerve impulses to maybe a thousand other cells and at the same time can receive information from about as many others, then the number indicating complexity goes up very quickly. Now we are getting into very large astronomical numbers, maybe even cosmic. It is very hard to think about such a complex structure, and our terms for summing up what it does, like perception and consciousness, are woefully inadequate. But it won't help us to complain about the complexity: we simply have to dig in, select one part to study, and see where that gets us.

I am aware that the basic topic at this colloquium is the future of science, and I was supposed to talk about future research on the brain. That is a little hard to do if one doesn't have a basis to go on. I think that even this group of scientists is not fully aware



THE VISUAL PATHWAY

Fig. 1. The human brain and eyes seen from below. About 1 million optic nerve fibers come from each eye. At the optic chiasm half the fibers from each eye cross to the opposite hemisphere and travel back on the optic tracts to the lateral geniculate bodies. There the information is relayed to lateral geniculate cells, whose fibers pass back through the brain, in the optic radiations, to the primary visual (or striate) cortex. Note that because of this pattern of wiring, each hemisphere of the brain gets input from both eyes, and a given hemisphere, say the left, gets input from the two left half retinas, and consequently the right half of the visual world, from both eyes.

of what is going on in the rather arcane fields of brain neurophysiology and neuroanatomy. So instead of speculating about the future, I want to tell you some of what we've learned about the part of the brain concerned with vision. We know more about vision in the mammalian brain than about any other aspect of the central nervous function. The topic is a rich one, and even to give you a rough idea about it I will have to go into some technical details. But the main thing that I want to convey is a flavor for the sort of research that is going on now and for the sort of concrete facts we are learning about visual perception.

The mammalian visual system is remarkably similar among the different primates, so although the work I shall describe has been done on the macaque monkey and the squirrel monkey, the results apply almost unchanged to the visual system of the human brain. Let's begin by looking at the layout of the visual system from the eyes to the primary visual cortex at the back of the brain.

In Fig. 1 we are looking at the human brain from below. At the top of the figure are the two eyes out of the back of which come the optic nerves. Behind the lens of each eye is the retina, which contains a mosaic of 125 million light detectors called rods and cones. These light receptors make synaptic contact with other nerve cells in the retina; that is, their nerve fibers, or axons, split into a few or many branches that end on a second set of cells. These in turn have branches that end on a third set of cells called the retinal ganglion cells. The axons of these ganglion cells bundle together to form the optic nerves. About a million optic nerve fibers extend out from each eye. Some of the fibers stay on the same side of the brain, and some of them cross onto the other side. All their terminals end up in one of the two nests of cells known as the lateral geniculate bodies—geniculates for short—each containing roughly 1.5 million cells. The geniculates are deep in the head roughly between your two ears. They have a rather simple structure, in that any particular geniculate cell gets input from the optic nerve fibers and sends its output through the substance of the brain to what is called the primary visual cortex, or striate cortex, located at the back of the brain. Axons of the cells in the primary visual cortex project to a neighboring area, which then feeds into two or three other areas, and so on. Figure 2 shows a diagram of the visual pathway, which is actually made up of many millions of nerve cells.

Over the last twenty years people have come to understand fairly well how individual nerve cells work (Fig. 3). Without going into any of the rich detail, let me just say that one cell sends messages to another by events called nerve impulses. Whether a given cell fires or not depends on the sum of what it's told to do by other cells, some of whose impulses excite the cell in question, others of which inhibit it. The inhibitory influences are very important, as you will see.

Now let's consider what kinds of messages are sent through the visual pathway once light reaches the retina. It has been known since 1950 that a great deal of processing goes on between the light detectors and the optic nerves, and that the optic nerves carry rather sophisticated messages to the brain. Stephen Kuffler, the person most responsible for working this out, was my boss for some years. One of his favorite sayings, which fits the topic of all of these talks very well, is that the hardest thing to predict is the future. Now you can understand why I am keeping rather quiet about the future of brain research.

One of Kuffler's main contributions was to show that optic nerve fibers are carrying complex information of the following sort. Suppose I have an anesthetized animal facing a screen a couple of meters away and I put a microelectrode near or into one of the animal's optic nerve fibers. Then I shine lights

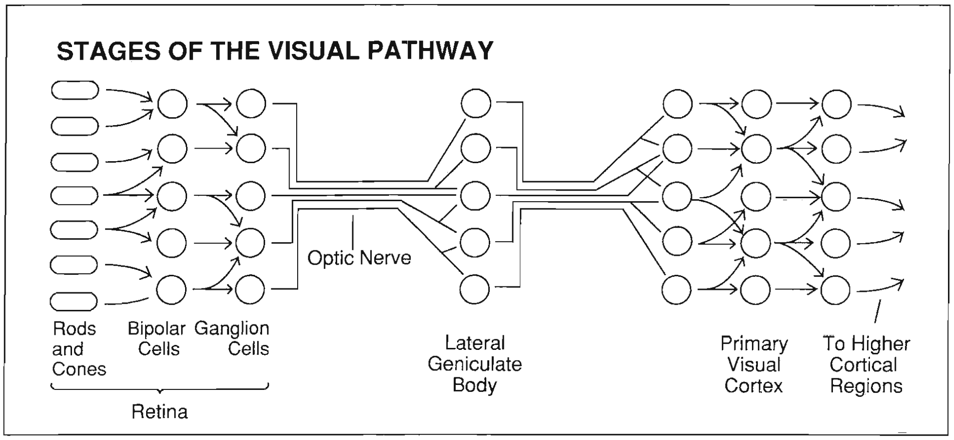


Fig. 2. Schematic illustration of the visual pathway from the retina to the higher cortical areas. At each stage one cell receives input from many cells at the preceding stage and passes information on to many cells at the next stage. Although only a small number of neurons are shown, each stage contains millions of neurons.

on the screen and ask what type of signal must reach the retina to influence this particular fiber (Fig. 4). It turns out that the area of retina influencing a typical optic nerve fiber is limited in extent, typically a circle about 1 millimeter in diameter. This region is called the receptive field of the cell. If we confine the light reaching this receptive field to a very small central region, we can drive each ganglion cell to produce up to 50 or so impulses during the onset (first tenth of a second) of stimulation. The cell then continues to fire at an average rate of up to 100 times per second. On the other hand, if we illuminate the whole receptive field, the cell responds at a much slower rate. Many of the optic nerve fibers hardly fire at all if you bathe the whole screen in light. That means that the ganglion cells are not interested in the amount of light hitting the retina but rather are interested in contrast. Each cell is making a spatial comparison between the amount of light hitting one tiny central region of its receptive field and the amount falling on the immediate surround. Illumination of the center excites the cell, and illumination of the surround inhibits it. Consequently these ganglion cells are described as having concentric "on"-center and "off"-surround receptive fields. Actually, there are two kinds of center-surround cells, those with "on" centers and "off" surrounds and those with "off" centers and "on" surrounds. The first respond best to light spots on a dark background, and the latter to dark spots on a light background (Fig. 5).

If we now do a similar experiment with the cells in the lateral geniculate body we find that they respond in roughly the same way. Each individual cell takes care of a small region of retina, and the way that region influences the geniculate cell may again be described as a concentric center-surround receptive field. (Notice that the term receptive field refers to both the region of the retina influencing a given cell as well as the nature of that influence.) Thus the kind of information that the geniculate cells send to the primary visual cortex is of a very special kind. I should point out that hundreds or thousands of retinal ganglion cells and hundreds or thousands of geniculate cells are taking care of one small region of the retina. That is, these cells have receptive fields whose centers overlap within this one small retinal region. So one small spot on the screen will activate thousands and thousands of cells.

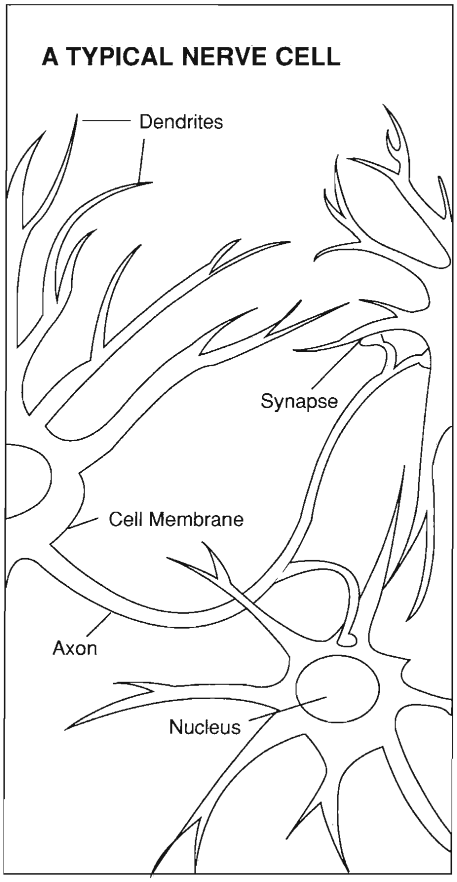


Fig. 3. A typical nerve cell may have from a few to over a thousand branches called dendrites, which receive signals from other nerve cells. Depending on the sum of the inputs (excitatory and inhibitory) the cell body fires or does not; that is, it sends out an electrical impulse that travels down the axon at a speed of about 10 meters per second. A nerve cell that fires rapidly does so at roughly the rate of a machine gun (an average of 15 times per second). When the impulse reaches the terminals of the axon, chemical messengers called neurotransmitters are released that provide input to the next set of cells. The relay of signals from one cell to the next takes place at specialized sites of contact known as synapses.

When we reach the primary visual cortex we find several stages of information processing. The cells at the earliest stage work roughly like the geniculate cells: they have center-surround receptive fields. At the second stage there are cells whose receptive fields are similar to the center-surround cells, in having an excitatory region and an inhibitory region. The geometries of these receptive fields, however, are different. They are designed to see either a light line on a dark background, a dark line on a light background, or an edge between dark and light. Moreover the line must have a certain orientation. The various receptive fields of these so-called simple cells and the response of one of them are shown in Fig. 6.

An engineer can give you a perfectly plausible diagram for how to get from a center-surround cell to an orientation-selective one. Depending on the engineer the diagram may differ, but the simplest circuit is to imagine a master cell getting input from a lot of center-surround cells, each of which differs in the positions of their receptive-field centers. If these receptive field centers are arrayed along a line, then the master cell gives a maximum response when the stimulus covers all of the center and only a small part of the surround of each of these ancestral cells, as shown in Fig. 7. A line of light made by a slit does the trick best of all. The precise circuit for this sort of thing is not known, and I am not going to discuss the possible circuits any more here.

Beyond the second stage cells work

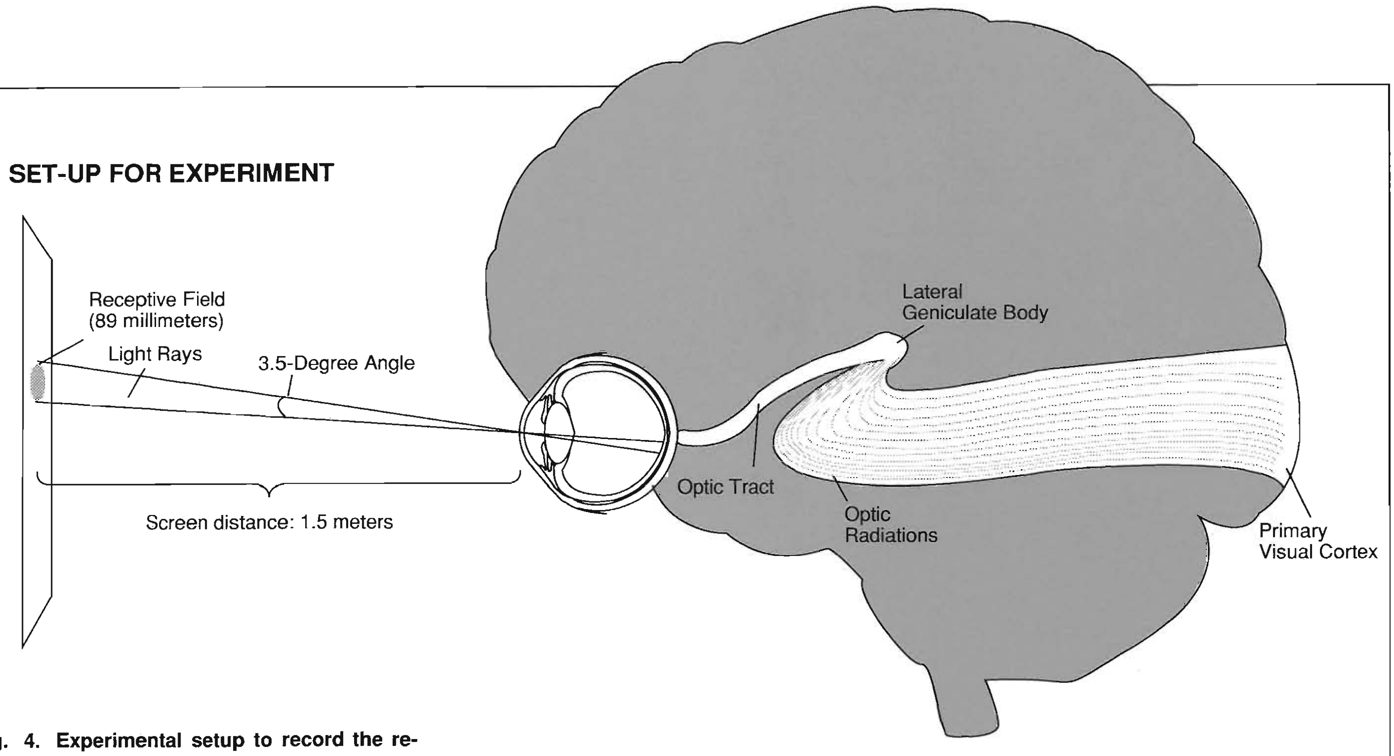


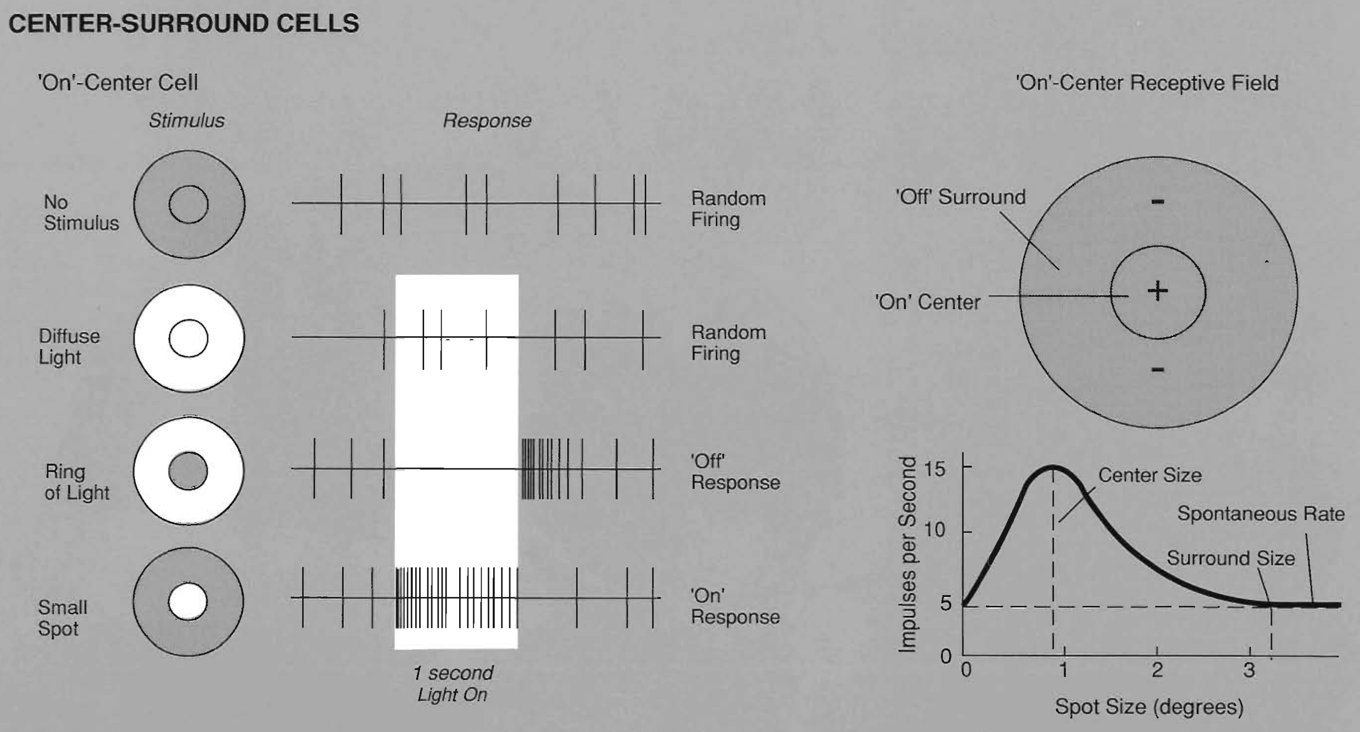
Fig. 4. Experimental setup to record the response of a single nerve cell to visual stimuli. An anesthetized animal is facing a screen about 4.5 feet away. Stimuli projected onto the screen are focused by the lens of the eye onto the retina. From there signals travel along the visual pathway (seen here from the side).

The receptive field of a single ganglion cell in the retina is a circular area about 1 millimeter in diameter, which subtends an angle of about 3.5 degrees. At a distance of 1.5 meters, a distance of 89 millimeters on the screen corresponds to 1 millimeter on the retina.

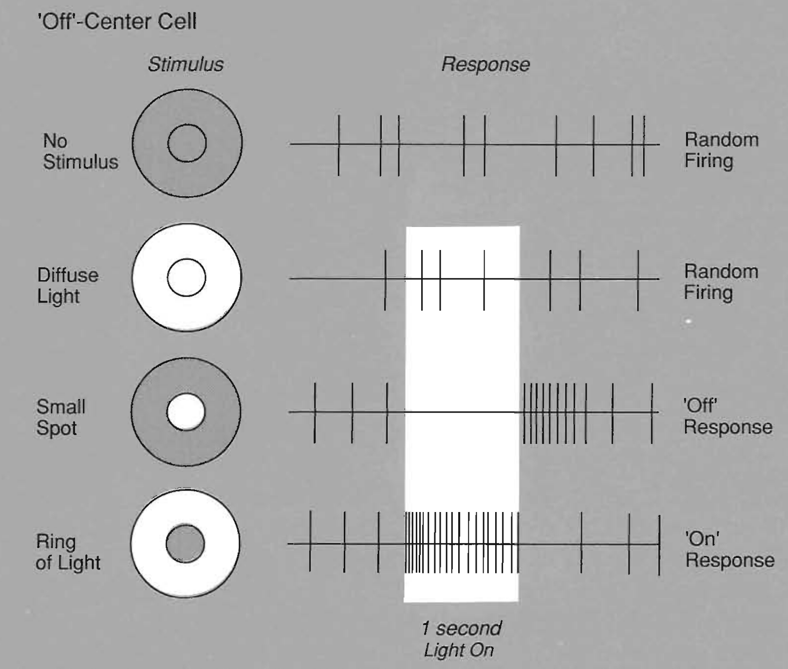
in even more complicated ways. Just as before, any given cell takes care of a small region in the retina corresponding to a small region of its visual field, for example, a portion of the screen in the experiments I described above. But these so-called complex cells don't respond to bright small spots anywhere in their receptive fields. What they like is a line moving across the region. They are also very fussy about the orientation of the line. Most of the cells are so fussy that if you change the direction of the line by more than 20 to 30 degrees they don't respond at all. Again, by "line" I mean either a light line on a dark background, a dark line on a light background, or an edge or boundary, say, between dark and light. Some cells are finicky about which of these three they respond to; others are less so. Some of them are also fussy about the direction in which the line is moving.

Now let me describe in a little more detail the experiments that demonstrate the response of these complex cells. As usual, we have an anesthetized animal facing the screen and an electrode poked into a single complex cell in the primary visual cortex. The electrode is connected to an audio monitor and each impulse from the cell is recorded as a click. The stimulus is a straight line of bright light on a screen. We project light through a slit to create this pattern of illumination, and we sweep the slit across the screen in a direction perpendicular to its orientation and repeat this for all orientations—as if we were painting light on the screen. In this way we are able to map out three things about the cell's receptive field. First we map out the receptive field of the cell, the area of the screen over which the cell can be influenced. When the line of light is moving over the cell's receptive field, the sound from the audio monitor becomes a din of clicks—the nerve cell is firing at machine-gun speed—but as the line goes past the region the sound dies down to a few isolated clicks and then to silence. As we change the orientation of the line between vertical and horizontal, the intensity of the sound corresponding to the number of impulses per second clearly varies, so

continued on page 24



The graph shows the response of a single "On"-center geniculate cell versus spot size. The spot changes from one that is smaller than the center to one that covers the entire receptive field. The maximum response occurs for a spot that covers the "On" center only (1 degree). As the spot grows larger and invades the inhibitory region the number of impulses decreases until the spot size is 3 degrees (the size of the entire receptive field). Further increase in spot size does not change the cell's response.



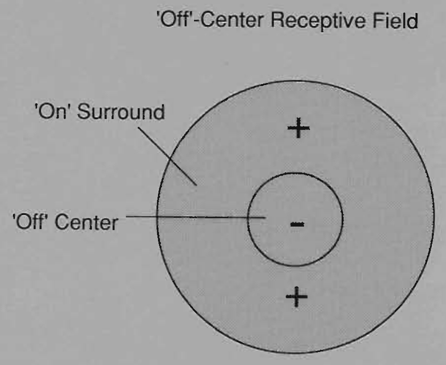
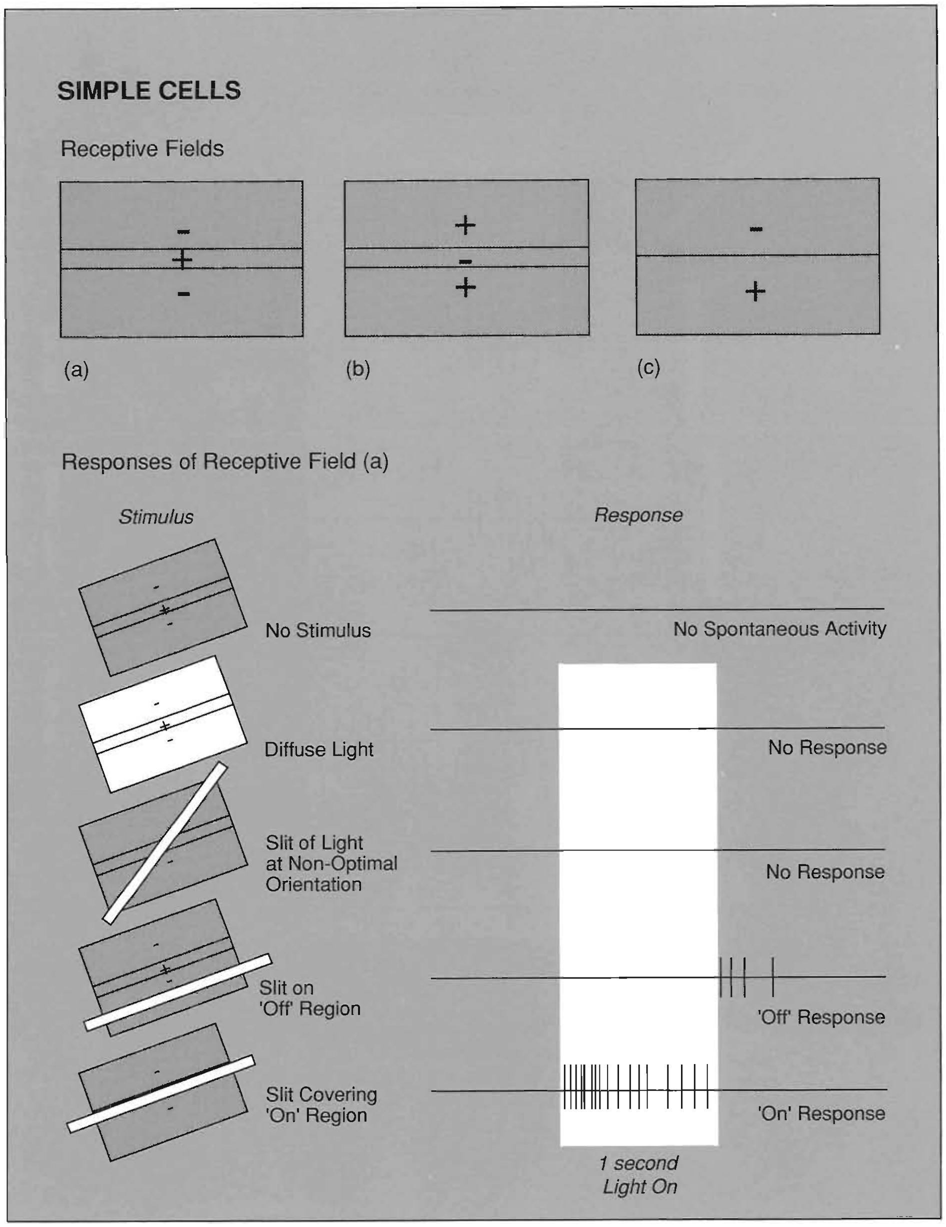
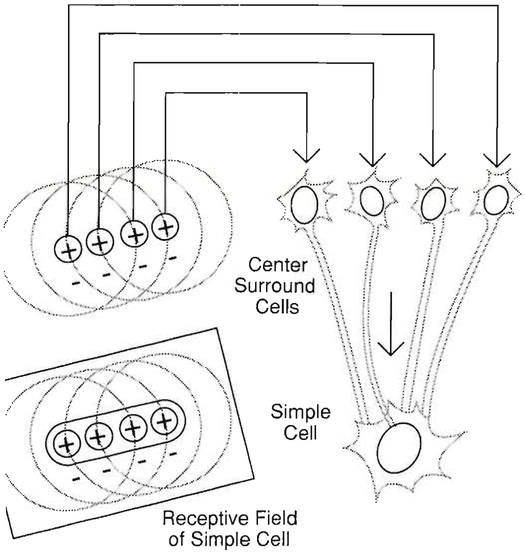


Fig. 5. Responses of cells with center-surround receptive fields to various stimuli. The recordings are from geniculate cells, which have the same pattern of response but fire much more slowly than retinal ganglion cells. The response line is 2.5 seconds long. Each vertical line represents a single impulse in the geniculate cell being recorded. With no stimulus the cell fires slowly (a few times per second) and more or less at random. With diffuse light covering the entire receptive field, the cell again fires at random and just a little faster. With a small bright spot covering just the center, the cell fires rapidly. A ring of light covering the surround stops the cell from firing. When the stimulus is turned off the cell produces a brisk burst of impulses lasting about a second. The receptive field is thus described as an excitatory ("On") center and an inhibitory ("Off") surround. The responses of the "Off" center, "On" surround cells are also shown.

Fig. 6. The receptive fields of simple cells in the cortex are designed so that the cell responds to (a) dark lines on a light background, (b) light lines on a dark background, or (c) a straight edge between dark and light regions. Like the center- surround cells, the simple cells have excitatory and inhibitory domains. A small spot anywhere in the receptive field will give a small response, either inhibitory or excitatory depending on its location, but the maximum response is obtained by stimulating the entire excitatory region and neither of the inhibitory regions. The job is done best by a slit of light whose width is about 2 minutes of arc.

Five recordings from a cell designed to see a light line on a dark background. This particular cell exhibits no spontaneous activity (some do) and also does not fire if the whole field is illuminated. When a light line fills the excitatory region the cell produces a maximum "On" response. If the light line is moved to the inhibitory region, the cell fires after the stimulus is removed. If the line covers only a small part of the excitatory region and a proportionally small part of the inhibitory region, the cell again fails to fire.





"SIMPLE" ORIENTATION-SELECTIVE CELLS

Fig. 7. One possible circuit for converting a center-surround response to an orientation-selective response. The orientation-selective cell receives input from a series of center-surround cells whose centers overlap and are arrayed along a line. When summed together, the center-surround receptive fields have an excitatory region that looks like a long narrow rectangle, flanked by inhibitory regions on either side. The receptive field looks very much like (a) in the preceding figure.

COMPLEX CELLS

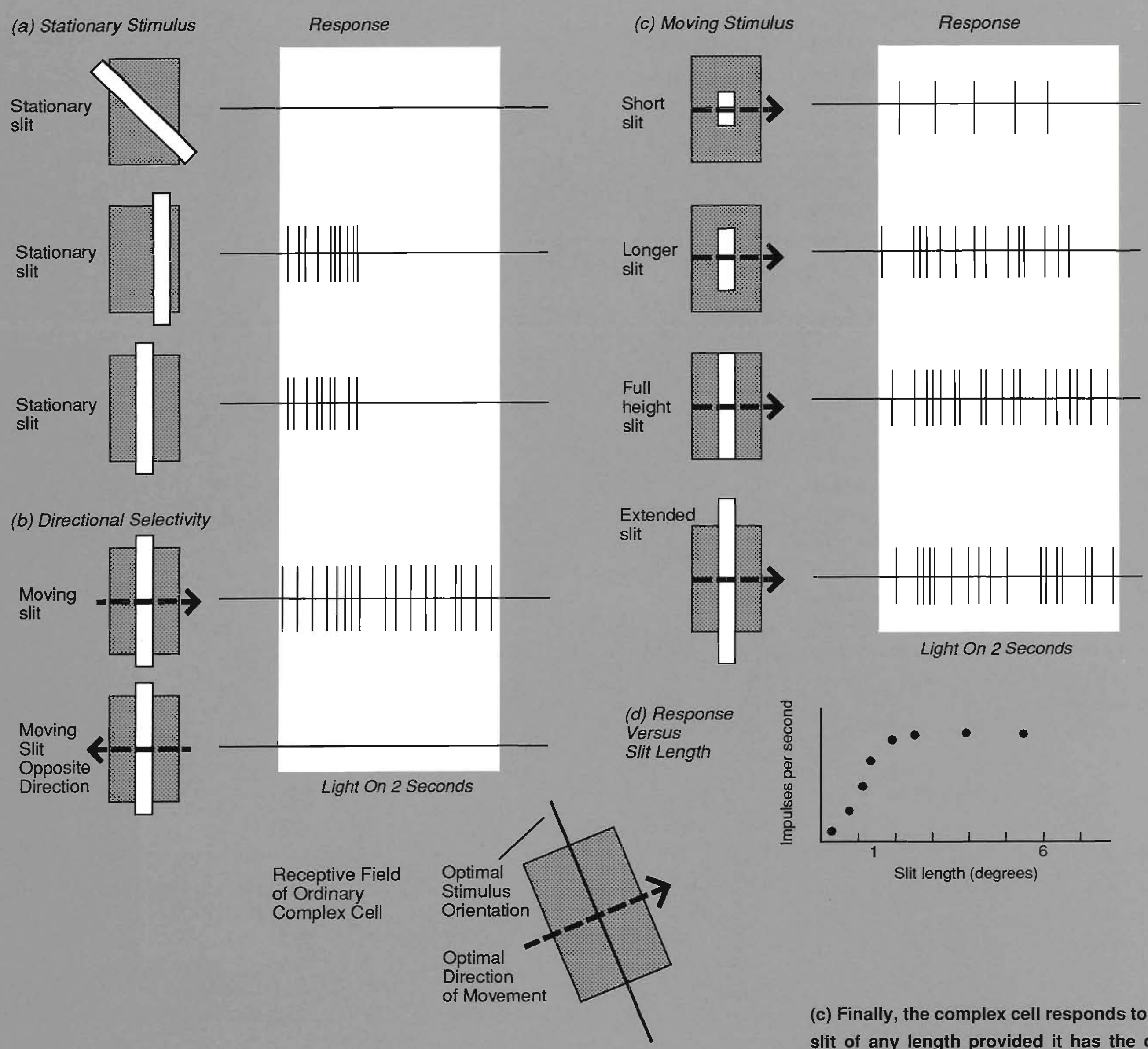


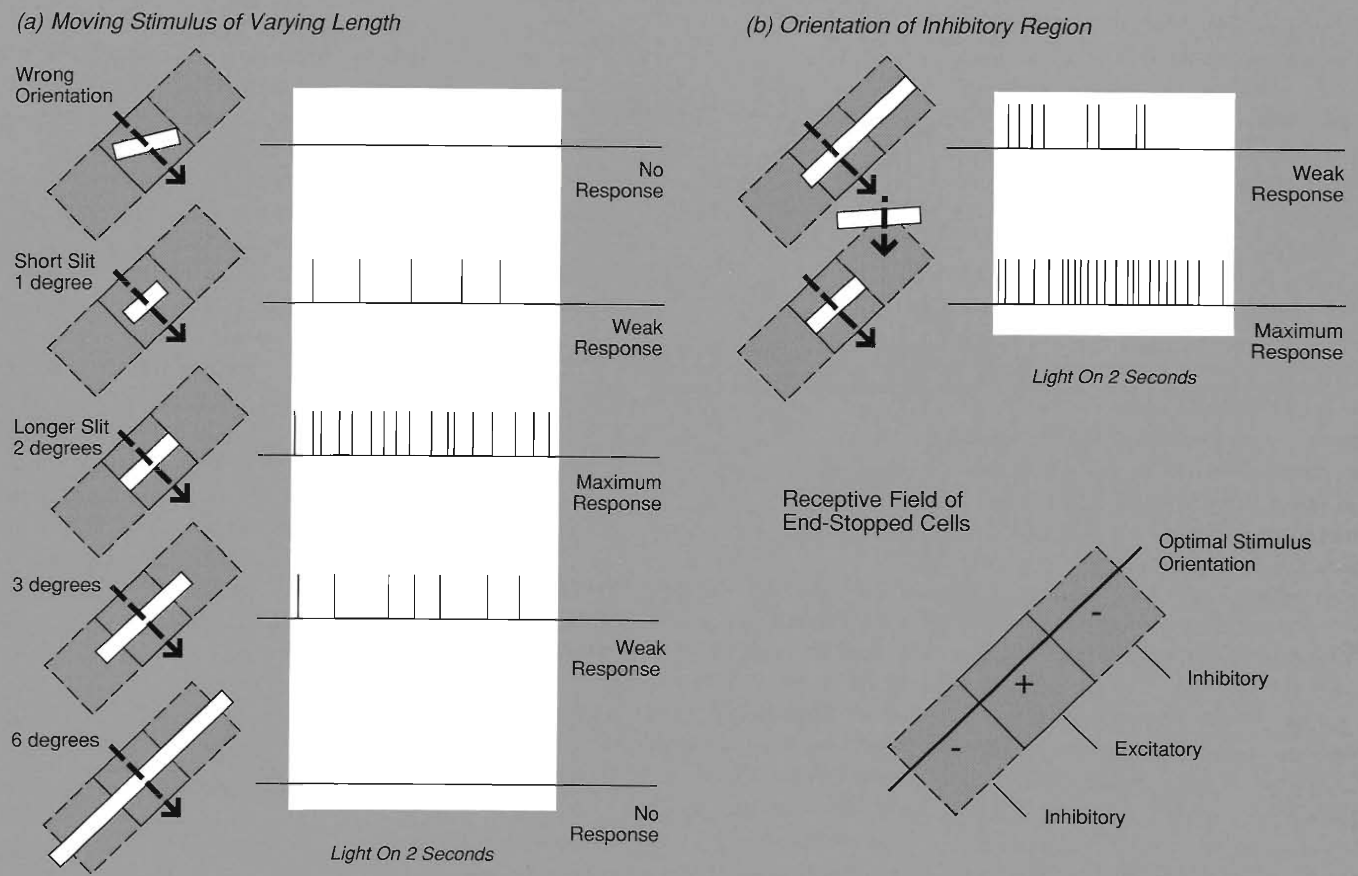
Fig. 8. A complex cell in the cortex generalizes the response of a simple cell. (a) The complex cell responds to a long narrow slit of light placed anywhere in its receptive field (rectangle) provided the orientation of the line is correct. Changing the orientation by ± 10 degrees will produce a weaker response; a larger rotation gives no response at all. The optimal width of the slit is 2 minutes of arc, the same as for the simple cell.

(b) The complex cell responds best to a slit of light moving perpendicular to its direction of orientation. For about 20 percent of these complex cells, the slit must move in one direction only. Motion in opposite direction fails to produce a response.

(c) Finally, the complex cell responds to a light slit of any length provided it has the correct orientation but the maximum response is obtained when the slit is at least long enough to cross the receptive field (2 degrees).

(d) The graph shows the response versus slit length for an ordinary complex cell. The response increases until the slit length is 2 degrees (the width of the receptive field) and then remains constant as the slit length is increased still more.

END-STOPPED CELLS



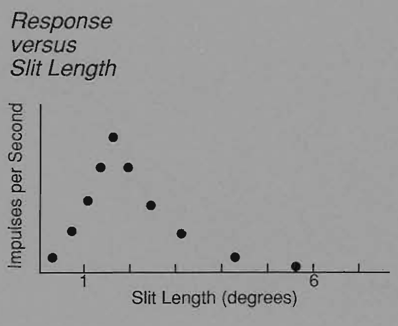


Fig. 9. (a) End-stopped cells, like complex cells, respond to lines with a specific orientation moving perpendicular to that orientation and as the line gets longer the response gets stronger. But unlike what one finds with complex cells, this cell shows a weaker response when the line is longer than about 2 degrees. At a line length of 6 degrees, the cell gives no response at all from the excitatory region. Evidently the (activating) excitatory region is flanked by inhibitory ones on either side as shown in the diagram of the receptive field. Each record has a 2 second duration. The graph summarizes the results of varying the slit length.

(b) These records show that the optimal orientation for the inhibitory region is the same as for the excitatory region. In the second record a nonoptimally oriented slit in the inhibitory region does not reduce the response to an optimally oriented slit of optimal length in the excitatory region.

continued from page 19

we can map out a preferred orientation for the line as well as a range of orientations over which the cell gives some response. Finally we swing the light back and forth, and the dramatic alternation between rapid firing and silence tells us that the cell responds to motion in one particular direction and not to motion in the opposite direction. The experiment is so clear that we made a movie of the stimulus on the screen and simultaneously recorded the audio signal generated by the cell's response. In the movie we draw in the receptive field, the preferred orientation of the line, and the preferred direction of motion of the line as these become evident from the audio response. Some results of the experiments on directionally selective complex cells are shown in Fig. 8.

We made a similar movie for another type of complex cell that is fussy about the length of the line. We call that an "end-stopped" cell. The cell responds very well to a short line but very badly to a long line. Apparently inhibition plays an important role in the functioning of end-stopped cells, just as it does in the functioning of center-surround cells. The receptive fields of these cells extend beyond the region where you get a big response. But the only way you can know that is to make the line longer and find that then you don't get any response at all from the cell. Evidently stimulating the region beyond the short line has the effect of inhibiting the cell, and if you inhibit the cell as much as you excite it, then it just sits there and does nothing. Figure 9 shows results of an experiment as well as a diagram of the end-stopped cell's receptive field.

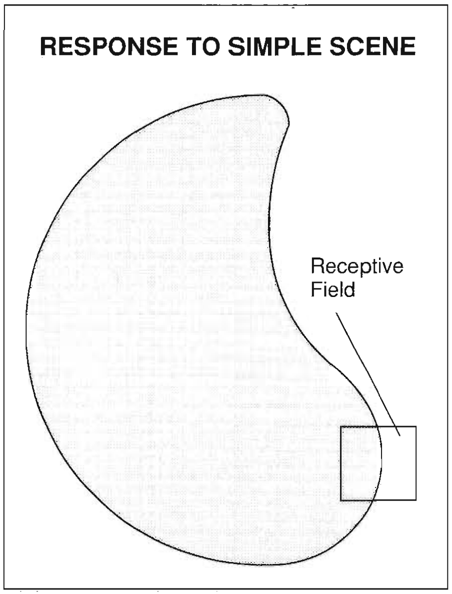


Fig. 10. Diagram to illustrate the effect on typical cortical cells of a simple scene, a kidney-shaped dark patch on a white background. If we fix our gaze on a point toward the center, we can imagine that the only cortical cells to be affected will be ones whose receptive fields are cut by the boundary, and whose optimal stimulus orientation happens to be appropriate. For example, a cell whose receptive field is vertically oriented and is stimulated by the boxed region in the figure will be activated. A cell whose field is entirely inside or outside the blob will be unaffected by the stimulus.

Now let me add one more thing to try to get at why we think the Almighty would have given us end-stopped cells. Suppose you sit back in your chairs and look at the form in Fig. 10. If you fix your gaze on a point toward the center of the form, millions of complex cells in your brain are being activated by the borders. In fact the *only* cells that are going to be tied up by this stimulus are the cells that are mediating the borders. It turns out that as you go farther and farther toward the center of the form, things are arranged in such ingenious ways that the number of cells required to give information about the interior becomes less and less. If you only consider end-stopped cells, the number of cells that are tied up by this stimulus is rather small. Only the end-stopped cells whose receptive fields happen to coincide with the regions of high curvature will respond. (Remember high curvature is essentially equivalent to small line segments in a very small region. Figure 11 shows how the receptive field is designed to respond to curves.) These phenomena are very counter-intuitive. You wouldn't think that your vivid impression of this homogeneous form would be conveyed at some stage of the brain by cells that aren't even telling you anything about the interior. It is the fact that information is coming from the borders and no additional contradictory information is coming from the inside to tell you that the contrast has changed, that lets you know that the whole form is filled. At first sight you may find this a hard pill to swallow, but it happens to be the way the brain works. If an engineer wants to build an image-processing device, he would probably invent a very similar design. He has to pay for all the transistors that take care of the innards,

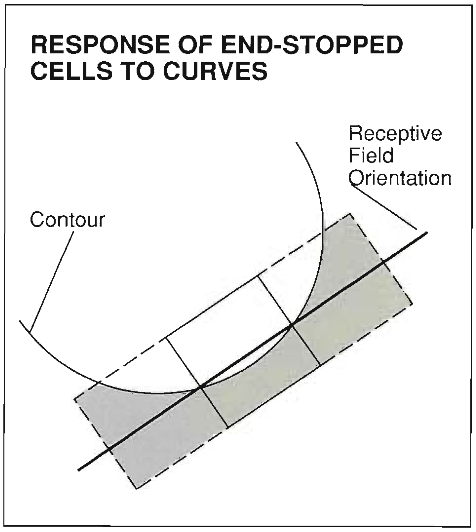


Fig. 11. This diagram shows how end-stopped cells are well designed to respond to curves. The segment of the curve passing through the excitatory region is oriented to cause a large response, whereas the segments through the inhibitory regions have the wrong orientation to inhibit the response. The sum total will thus be a large response.

and so he wants a machine that uses as few of them as possible.

So far we have discussed cells that distinguish contrast, or differences in brightness, and are therefore involved, loosely speaking, in the perception of form. The subject that I want to discuss during the remainder of this talk concerns our intuition that visual perception is not a unitary thing but must be subdivided. When we look at a scene, we are not necessarily conscious of the various subdivisions of perception. But suppose you ask an average person to break up vision into its parts. Most people would probably say you have form, color, movement, depth, maybe texture, and a few others. If you gave the right cues, even a Boston taxi driver would give you some list like this, so you do not need a scientific or neurobiological background to come up with it. This subdivision is intuitively reasonable to us. Now it turns out—and it didn't have to be so—that the brain divides up vision pretty much according to this list.

Let's take a look at the anatomy of the lateral geniculate body because that is the first place where the division of the visual pathway is rather obvious. Figure 12 is a cross section through the lateral geniculate body. From one side to the other is about 3 or 4 millimeters, and because of the stain that was used each dot is a cell body. This picture alone tells you a great deal about the geniculate's structure. It is rather like a tiny jelly roll consisting of a number of layers, one topping the other and all rolled up in a sort of curved way so that each cross section parallel to the one in Fig. 12 would look very much the same. Each layer in the geniculate receives input from either the right eye or the left. Thus, in the right geniculate body shown here, all the cells in the top layer get their input from the right eye, all cells in the next layer get their input from the left eye. The whole sequence of inputs from top to bottom goes right, left, right, left, left, right. Why the order changes near the bottom nobody knows; it may just have been to make it hard for us to remember.

The main feature that I want you to notice is the obvious difference between the two ventral (underneath) layers and the remaining four dorsal (upper) ones. You can see that the cells in the two ventral layers are bigger and more thinly scattered. If you looked at this cross section with more powerful methods you would see other differences. It has been clear for a century that the geniculate is subdivided into these two distinct regions; the ventral layers are called magnocellular because the cells are big, and the four dorsal layers are called the parvocellular layers, "parvo" for small. Moreover these two kinds of cells get their inputs from two different kinds of cells in the retina. The magno get their inputs from big retinal cells, and the parvo from small ones. At later stages in the cortex, these two branches of the visual pathway, magno and parvo, don't merge but keep their separateness and seem to have different functions. Although both magno and parvo geniculate cells have receptive fields with a center-

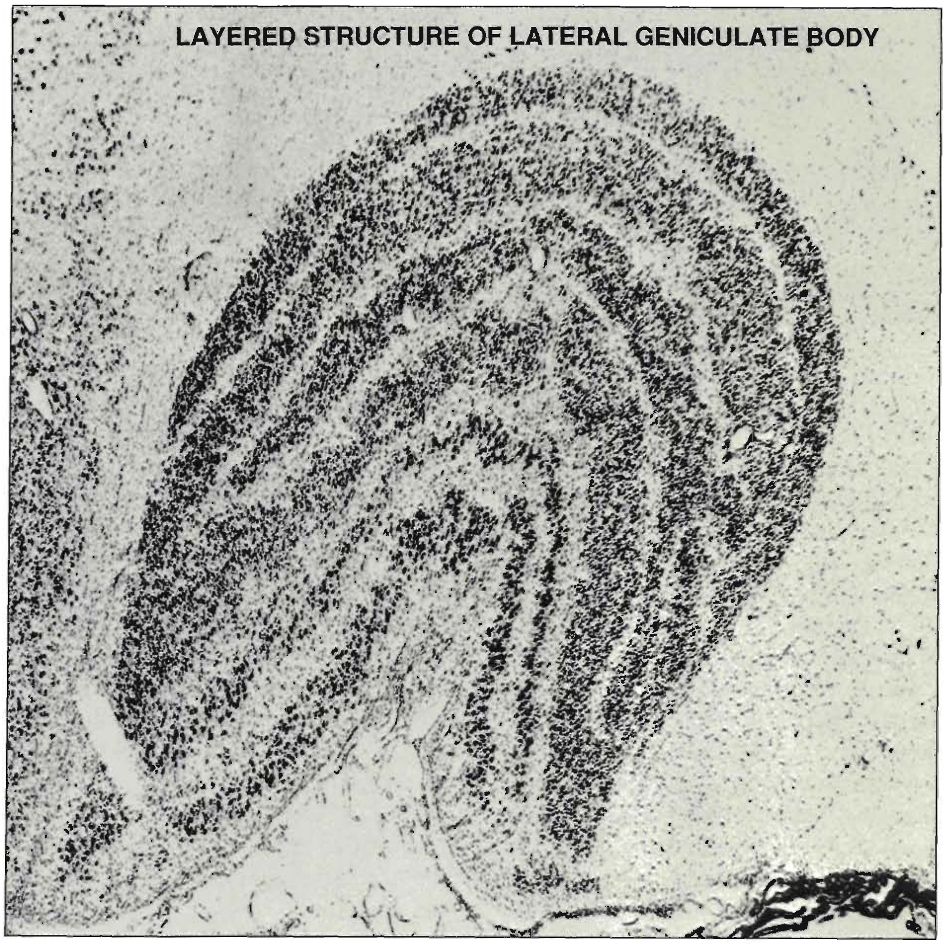


Fig. 12. Microphotograph of a cross section through the right lateral geniculate body of a macaque monkey. Each dot is a cell body, stained with cresyl violet dye. Each of the six layers gets input from one eye only. A human lateral geniculate body would look almost identical to this.

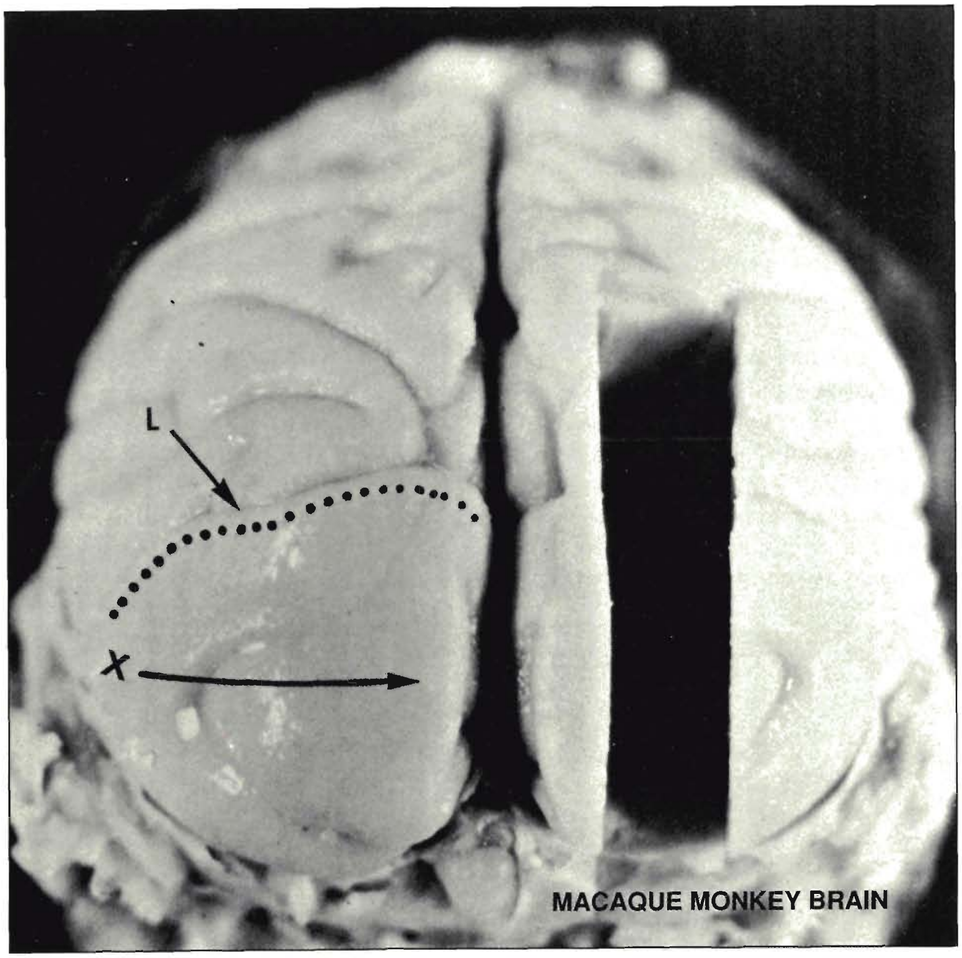


Fig. 13. A macaque monkey brain viewed from behind. The region in the foreground (below the dotted line) is the primary visual cortex, the part that is exposed on the surface. L is a deep cortical fissure (the lunate sulcus). X is the representation of the center of gaze. As we proceed from X in the direction of the arrow, the part of the visual field mapped goes to the right along the horizon.

surround target-like arrangement, the receptive field centers of the magno cells tend to be bigger than the field centers of the parvo cells. It is as though each magno cell got its input in parallel from all three kinds of color cones and are therefore color blind. The parvo cells, on the other hand, are very strongly color sensitive. They respond to color as though their centers got their input from one color cone only and the surround from one of the other two color cones only. A second difference concerns sensitivity to small changes in luminous intensity: the magno cells are much more sensitive than the parvo. When the receptive-field center is just 5 percent brighter than the surround, the magno cells respond very well, whereas the parvo cells won't respond until the intensity difference between center and surround is at least 20 percent.

Now let's see what happens to these two pathways in the visual cortex. To do so we need to look at the anatomy of the cortex more closely. Figure 13 shows what a macaque monkey brain looks like if you remove the top of the skull. The primary visual cortex (also called visual area 1) occupies most of the area below the dotted line, but part of it is tucked underneath in a slightly complicated way. Its area is about the size of a credit card, and it has a thickness of two millimeters—thicker than the average credit card, unless you have the gold kind.

If you cut out a chunk of the cortex (as in Fig. 13) on the right side and walk into the hole and turn left, you would see the cross-section of the primary visual cortex shown in Fig. 14. The richly layered structure is made visible by an appropriate stain. The axons of the geniculate cells come up vertically through the lower layers of the cortex, branch again and again, and finally terminate in layer 4C, the very dark layer a bit more than halfway down. The 4C cells send their output to the upper layers of the cortex, and the upper layers send their outputs to other regions of the brain as well as to other layers in the primary visual cortex. We will be particularly interested in the projection to visual area 2, which borders visual area 1 (that is, it is right above the dotted line in Fig. 13). The projection is done in a very precise way, so that a tiny region in visual area 1 will send its output to a tiny region in visual area 2.

I said earlier that the two branches of the visual pathway, parvocellular and magnocellular, maintain their separateness in the cortex. In particular the magnocellular layers of the geniculate transmit impulses to the top half of layer 4C (4Cα), which in turn transmits its output to layer 4B. Layer 4B then sends *its* output to visual area 2. The parvocellular layers transmit impulses to the bottom half of 4C (4Cβ), which transmits its output to the deep part of layers 2 and 3. The output from layers 2 and 3 again goes to visual area 2. That's as much as I want to say about these two pathways until later when

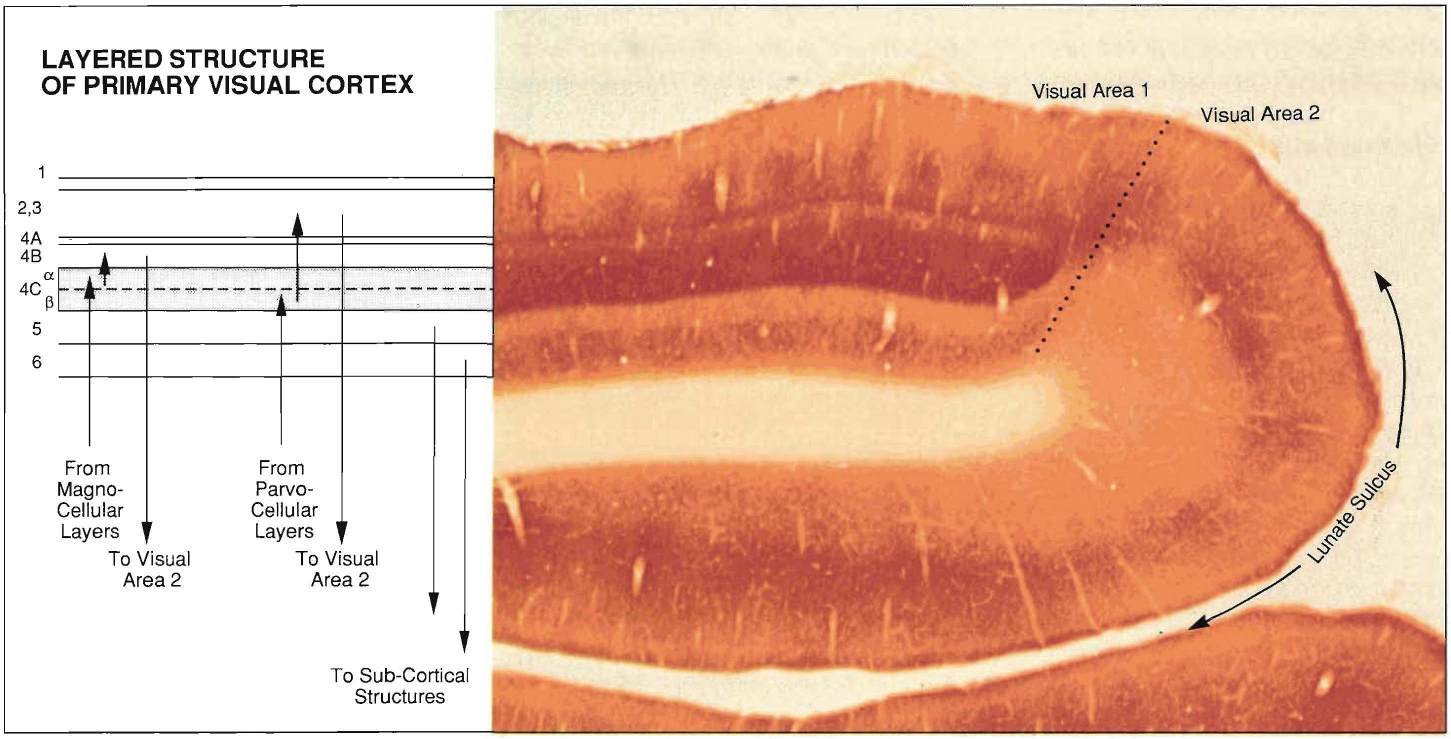


Fig. 14. Here we see a low-power cross section of the primary visual cortex, roughly what we would see if we were to walk into the cleft cut in the right hemisphere of Fig. 13 and look to our left. The cortical layers can be clearly seen. The pattern of layers changes as we go from the primary visual cortex to visual area 2. The transition between visual areas 1 and 2 occurs at the dotted line in Fig. 13. The deep fissure known as the lunate sulcus is visible in Fig. 14 just to the right of the transition between visual areas 1 and 2. About a dozen blobs can be seen above the very deeply stained layer 4.

Diagram at left shows the input to the cortical layers from the lateral geniculate body and the output to other regions of the brain.

we discuss visual area 2. About 1978 we and others began to suspect that, at least at the cortical level, there is a third branch of the visual pathway. By using a stain for cytochrome oxidase we were able to distinguish periodic regions in the upper layers (layers 2 and 3) of the cortex that were staining darker (see Fig. 14). In Fig. 15, which is a view of the cortex from above, these darkly stained periodic areas look like rather punctate oval regions about one-half millimeter apart. We call them "blobs" because of their appearance.

Around 1979 Margaret Livingstone and I were able to record the responses from individual cells in the blobs by driving a microelectrode parallel to the upper layers of the cortex. We had thought that all the cells in the upper layers are like the two kinds of complex cells I described before, either ordinary complex or end-stopped. It turned out, however, that every time we got our electrode into a blob, we

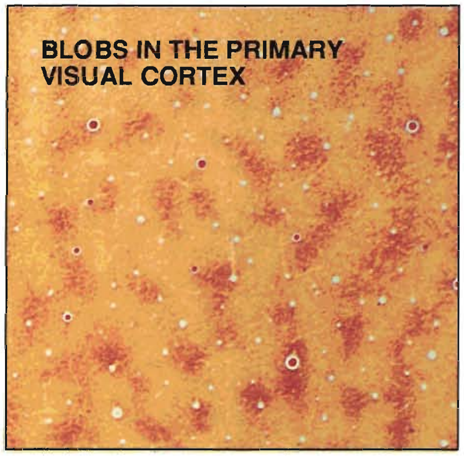


Fig. 15. Low-power picture of cytochrome oxidase blobs, in a piece of primary visual cortex cut in a plane parallel to the layers, above layer 4. We are viewing the blobs as if we were standing above the cortex in Fig. 14, looking down. Blobs are about 0.5 millimeter apart.

were likely to record five or six cells in a row that had absolutely no orientation specificity. They seemed to be of the center-surround style. About half of them ignored color, but the other half were richly involved in color and worked in a very specific way.

I want to discuss the variable color, so I will now describe how a color-sensitive blob cell works. It has a receptive field, with center and surround (Fig. 16). The center is likely to get its input from two types of cones in opposition, so that, for example, illuminating red cones has the effect of exciting the center and illuminating green cones inhibits it. If you happen to stimulate the center of the receptive field, the cell will turn on or off, depending on whether the light is red or green. If you use white light the inputs counteract each other because white light contains both long and middle wavelengths. The cell just doesn't respond at all to white light. The surround works in just the opposite direction, red inhibiting and green exciting. Now the ramifications of this are many. For example a small red spot tells the cell to fire faster but if the spot is large, it has no effect. The only thing that will drive this cell to fire faster is a small red spot, a red edge, or a long red line that stimulates all of the center and very little of the surround.

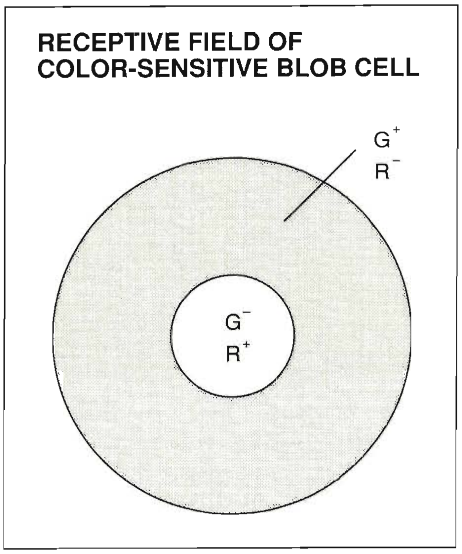


Fig. 16. Diagram of the receptive field of a red-on green-off double opponent cell. The cell's firing rate is increased by a small red spot and slowed by a small green spot. Large spots have no effect.

I am not going to spell out all the implications of this, but one consequence is that our perception of color involves not only the wavelengths coming from the object we are looking at, but the difference between the wavelengths coming from that object and the wavelengths coming from other objects in the scene. Thus *space* is involved as well as *wavelength.* For many years Edwin Land has been presenting demonstrations aimed at convincing people that space is involved in color perception just as much as wavelengths. I will now give you a kind of watered down version of one of these demonstrations. It will at least help you to understand why a conservative Canadian-born would come to this meeting in such a garish, bright red tie.

Following in the footsteps of George Wald, I will take off my coat, but for this additional reason: to show you the full glory of this tie. Everyone I think would agree, except those who are frankly color blind, that this is a red tie. What I propose to do is first of all bathe myself in long wavelength light, that is, what we are used to thinking of as red—and see what the tie looks like. That is what I will do, so perhaps we can turn off all the light in the room, and I mean *all.* Now if we can have the red projector turned on, I think you will agree that if anything the tie appears to be a kind of anemic red. It is a pale ghost of what it was before, and yet the wavelengths that are coming to your eyes from the tie are just what they were before; they have not changed at all. If we now turn the red light out and turn on the short wavelength (blue) light, the tie looks very dark, naturally, because a red tie by definition is one that does not reflect back short wavelengths. So the tie isn't shining back much of anything to you. Next we can turn off the blue light and turn the red light back on, and you again see the anemic, pinkish, washed-out red. If I were now to add the short wavelengths, I think you know that nothing new is going to come from the tie. So let's add the blue. Now you see the tie and at least from where I stand it bursts forth again in all its glory. Yet what could be more counter-intuitive than that? This is so counter-intuitive that people made fun of Land for the first twenty years of his presentations, saying that he was using magic and things like that. But these are very real phenomena, not magic, unless you want to think of biology as magic and insist that only physics is real.

Why would the Almighty wire up our brain in such a way as this? I think the answer is reasonably simple. If we are out under the blue sky and look at a colored object and then come in here and look at it under tungsten light, our estimation of the color stays remarkably constant. To realize that this constancy in color perception is not a trivial thing, try going outside and taking a picture of a white shirt and then coming inside and rephotographing it under tungsten light. You get a pink shirt on the one hand and a white shirt on the other, or else you get a perfectly good white shirt under the tungsten and a blue shirt under the blue sky. The light source makes a big difference to the camera. The camera isn't equipped to factor out the light source. Our brains are so equipped, by some mechanism like the blob cells that compare wavelengths in different regions of their receptive fields. The result is that the perceived color remains the same despite changes in the light source, and our brain does the job so well that it is hard to convince ourselves that it really is solving a problem. The

VISUAL AREAS 1 AND 2

Fig. 17. A section parallel to the layers of the visual cortex of a squirrel monkey at much lower power than Fig. 15. To the lower left in the figure is the primary visual cortex, with blobs that are 0.5 millimeter apart. To the right is visual area 2, with thick, thin, and pale stripes running at right angles to the border between visual areas 1 and 2.

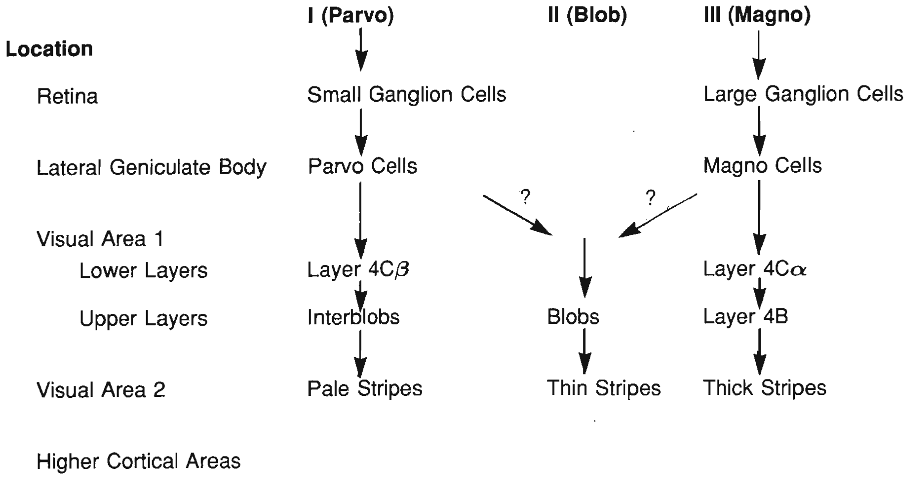
white is white. Why shouldn't white look white wherever you go?

It is indeed a complicated question, and many questions on perception have just that kind of complexity. *Exactly* how these blob cells solve the problem is not clear. One can go a certain way, but it is a complicated theoretical question and I think it hasn't been worked out satisfactorily so far.

The last topic that I want to talk about is visual area 2. If you stain visual area 2 for cytochrome oxidase and look at it, you don't see blobs but rather a pattern of dark stripes alternating with pale areas (Fig. 17). The stripes extend the full length of visual area 2, which is about 8 or 10 millimeters, and they appear every 4 or 5 millimeters. Furthermore the dark stripes are of two types, alternately thick and thin. This pattern gave us a hint that if we were to record from these regions we might find physiological differences. It also gave us a hint that the connections between visual areas 1 and 2 might have something to do with these blob and non-blob regions. We were able to show through anatomical work that the blobs project to the thin stripes and *only* to the thin stripes. I wish there were time to show you this convincingly, but it would take a while and it is a bit technical. We also found that many blobs project to a single thin stripe and inter-blob regions don't project to thin stripes at all. Furthermore the traffic is two-way. Blobs connect to thin stripes, which connect back to the blobs. The inter-blob regions in visual area 1 project to the interstripes, the pale regions of visual area 2. And finally the thick stripes in visual area 2 get input from layer 4B in visual area 1, the terminus of the magnocellular system. To sum it up, the magnocellular branch of the visual pathway is represented in the thick dark stripes, the parvo- interblob branch by the pale stripes, and the blobs are represented by the thin dark stripes. We don't know the source of the input to the blobs, but it's probably from magnocellular and parvocellular branches of the visual pathway. The three pathways have in some ways kept their separateness right up to visual area 2, and as I'll describe below, each one seems designed to perform a different function (Fig. 18).

When we record from visual area 2 we find that the cells in the thick stripes are very orientation-selective and very movement-sensitive, more so than the other regions, so we have reason to think that these cells are involved in perception of movement. They are also involved in stereoscopic depth perception, something that we don't find in visual area 1. With both eyes open these cells are very fussy about the distance to the screen: a given cell responds only if the screen is at the appropriate distance. Evidently the input from both eyes must have a specific alignment for the cell to fire (the principles of stereopsis are explained in Fig. 19). If the screen is not at that distance the cell simply ignores the stimulus and doesn't work at all. There are three categories of cells involved in stereopsis: "near" cells, which respond to stimuli at distances closer than *d*; the distance at which your gaze is fixed; "far" cells, which respond to stimuli at distances greater than *d*; and cells with no disparity, which respond

BRANCHES OF THE VISUAL PATHWAY



Properties	I (Parvo)	II (Blob)	III (Magno)
Spatial Resolution	**high**	low	low
Orientation Selectivity	**yes**	no	**yes**
Movement Sensitivity	**yes**	no	**yes**
Directionality	no	no	**yes**
Stereopsis	no	no	**yes**
Contrast Sensitivity	low	**high**	**high**
Color	no	**yes**	no

CONCLUSIONS

Pathway I, the parvocellular, is characterized by high spatial resolution, orientation selectivity, and end-stopping. We guess that this pathway is concerned with high-resolutions form perception.

Pathway II, the blob system, is concerned with color, but not with movement, stereoscopic depth perception, or form.

Pathway III, the magnocellular, exhibits systematic selectivity for movement and disparity between inputs from the right and left eye. This pathway thus seems concerned with movement and depth perception.

Fig. 18. Three separate branches in the visual pathway. Results of human psychophysical tests support the conclusions above.

to stimuli at the distance *d*. The cells in the thick stripes are thus concerned with stereoscopic depth perception as well as movement.

As you may imagine, the thin stripes, which we know are the terminals of the blobs, are color-coded, and about half of them are involved in the same sort of color problem (color constancy) we described for the blobs. They are thus a continuation of the blob system. The pale regions, finally, are full of end-stopped cells. In visual area 1, you may find that 20 percent of the cells are end-stopped, but in visual area 2 more like 80 percent are end-stopped. So we think that the pale regions are concerned predominantly with form perception. Form perception is a complicated concept, and I am using the term loosely.

All this has many consequences for perception, and I will have time to discuss only two of them before we close. One relates to the fact that the thick stripes are interested in stereoscopic depth perception and get their input from the magnocellular layers. Since, as far as we know, the magnocellular layers are not concerned with color at all, we would predict that stereoscopic depth perception and the perception of movement do not involve color. These predictions can be tested on humans, and I want to show you a few examples of the kinds of tests we have been doing.

First I will show you a simple demonstration to get across the idea that movement doesn't have much to do with color, nor with form. These three parts of vision seem to function independently. You perhaps know that if you turn on a spot on an oscilloscope screen, then turn it off and immediately turn on another spot a small distance away from the first, turn that spot off, and then continue to alternate back and forth, you get a vivid impression of movement from one spot to the other. Psychologists call this apparent movement. This illusion is used in neon signs and in airplane landing strips and of course in cinematography. For some years psychologists have been playing with the array of spots shown in Fig. 20. First you flash two spots on diagonal corners of a square, then turn them off and flash two spots on the other diagonal corners, and so on. Most people looking at this array have the impression that the spots are going up and down, but you can just as well have the impression that the dots

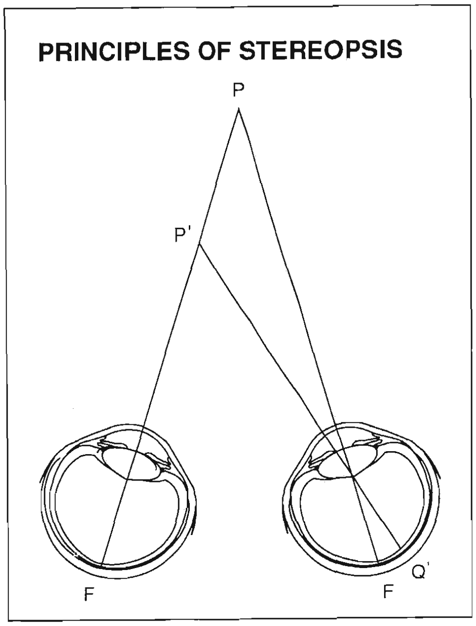


Fig. 19. When an observer fixes his gaze on a point P, the images of P in the two eyes fall on the two foveas. A second point P' closer to the observer than P has its two images, in this case F and Q', displaced outwards relative to the distance between the two foveas. Similarly a point more distant than P will have its images displaced inwards. Relative horizontal displacement of the two retinal images, for near or far objects, is interpreted by the brain as relative depth. This was discovered in 1838 by Sir Charles Wheatstone, who also invented the Wheatstone bridge—and the concertina.

are going from side to side. If you look at the oscilloscope screen long enough, your perception may flip and you will see the spots going up and down rather than horizontally or vice versa. If you have trouble making the flip, say from vertical to horizontal motion, you can block the two bottom dots with your hand, then take your hand away and the dots will seem to be going from side to side rather than up and down. If all of this is so, then you would think that if we made the two top spots X's and the two bottom ones O's you would see them going horizontally; X would go to X and O would go to O. But that's not at all what happens. You are just as happy to see an X going up and down and turning into an O and then back to an X. Similarly if you make two of them green and the other two red, it seems to have no influence on your perception of movement. You are perfectly happy to see a green spot jump over and become red. It is no problem. This seems to suggest that movement perception is quite different than color or form perception.

Now I want to turn to depth perception. I can't show you anything about stereoscopic depth perception without fitting everyone with polarized glasses, but I can show you examples of how other kinds of depth perception, perhaps all types, are colorblind. There are many other cues to depth: occlusion, parallax, movement, and so on. They all seem to fail if the figure you are looking at contains color borders but no change in intensity across these borders. I hope to convince you of that.

To do these demonstrations we use equiluminosity. We take a picture that has red and green areas and try to bal-

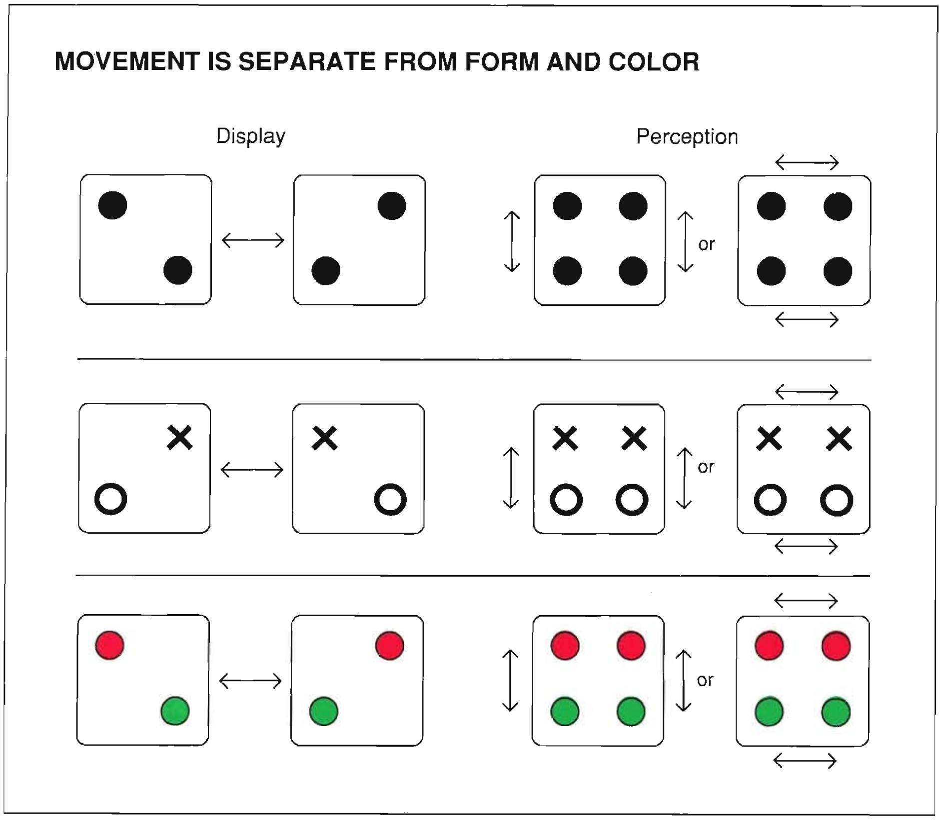


Fig. 20. The film display shows two dots that alternate between diagonally opposite corners of a square. The perception one gets is of dots moving vertically (up and down) or horizontally (from side to side). If the dots are changed to O's and X's, one sees an O going up and down and turning into an X and back to an O or X's and O's going from side to side. It is possible to flip from one perception to the other. Similarly if the dots are green and red one sees a green dot jump and become a red one and jump back and become green again. Thus movement perception appears to be distinct from color and form.

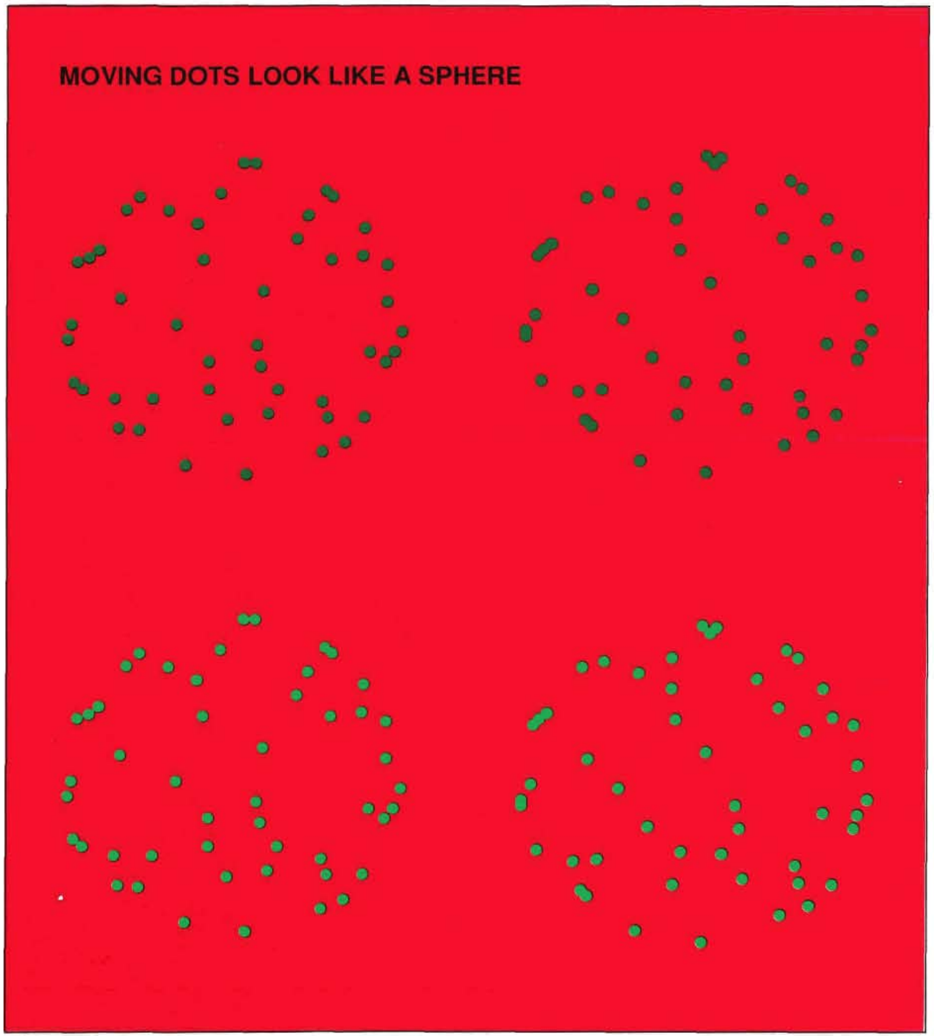


Fig. 21. Dots on a transparent rotating sphere are projected onto a red plane. Two successive images are shown in the figures. When the dots are dark green, the moving dots give the impression of a rotating sphere. When the green dots become equiluminous with the red background the impression of the sphere (that is, of depth) deteriorates, and the dots appear to be moving at random.

ance the two colors so that they are equal in luminance, that is, equally bright, roughly speaking. If you get them equally bright and, say, the system for detecting depth or movement is colorblind, you should stop seeing that phenomenon whereas you would see it perfectly well with a black and white image.

I have a demonstration in which a bunch of dots are pasted on a rotating transparent sphere, and the moving image is projected on to a red plane (Fig. 21), so the dots on the plane shift in their position in such a way as to imitate the movement of dots pasted on the rotating transparent sphere. Everyone who looks at this film has the impression of a moving sphere. The dots start out dark green, and then gradually become a brighter and brighter green. When you get the brightness of the red background and green dots balanced, your capacity to deduce shape from movement disappears, and all the dots seem to be moving all over the place in a random fashion. You lose the impression of the sphere. Then as I increase the brightness of the green still more, the impression of the rotating sphere returns. This demonstration illustrates why we think that our ability to deduce shape from movement requires luminance differences. When the luminance is equal, color borders alone are *not* enough to give the shape because the visual pathway for color perception is separate from the pathway for perceiving differences in luminance.

In 1984 Cavanagh and Favreau discovered that if you had red and green stripes moving slowly downward on an oscilloscope screen and you change the intensities of the red and green, you can come to some balance for which the impression of movement deteriorates. That is, the red and green stripes seem to be moving more slowly or not moving at all. I want to demonstrate this. What I have here is two sine wave gratings 180 degrees out of phase, one red and one green. We have blocked off half the TV screen with red cellophane, so that on the left half of the screen you see red and black stripes and on the right half you see red and green (Fig. 22). The stripes are moving slowly down the screen. We start with the green stripes lower in intensity than the red and gradually build up the intensity of the green. We have found out that different people have different balance points; that is, the red and green are perceived as equally bright at different relative intensities. This makes the effect hard to demonstrate, and is the reason that I'm slowly building up the green: everyone will get a balance at one or another level of green.

Now if you look at the junction be-

tween the two halves of the screen, of course you will see that both sets of stripes are moving at the same rate. But if you look away a little bit, say a few feet above the whole demonstration along the middle, and you ask yourself which is moving faster, I think that you will agree that the red and black stripes on the left seem to be moving faster than the red and green stripes on the right. As the relative brightness changes, there may even be some point when the red and green stripes appear to be stationary. But as the film proceeds, we can satisfy ourselves by looking at the border that both sets of stripes are moving at exactly the same rate. This demonstration clearly suggests that the perception of movement relies heavily on differences in luminance, or brightness. Color cues alone are not reliable.

Fig. 22. Two sets of moving stripes, one red and black, the other red and green, show how movement perception deteriorates when the red and green stripes are equiluminous. Although the stripes move down the screen at equal velocity, when the red and green stripes are equiluminous, they seem to move more slowly than the red and black stripes.

Maybe I have convinced you that perception is more complicated than the word might imply. We get to know better what the words imply by getting a deeper understanding of what is behind them. ■

Questions and Answers

Question: You have focused on a part of the brain that is transducing external reality inward and allowing us to see it. Have you tried to move into parts of the brain that are doing more abstract things such as problem solving?
Hubel: What I have discussed today represents the kind of investigation that has been done so far, and it is obviously very far from explaining how you recognize a face, a boat, a hat, or any familiar image. We are very far from understanding what we call shape recognition, to say nothing about more abstract things, such as language, speech, and maybe the most difficult of all, problem solving. At the moment, the problem of getting to Mars is easy compared to that of understanding how the brain solves problems. That is more like the problem of getting to a planet in the Andromeda galaxy, which is difficult indeed given our life span and the speed of light. On the other hand, understanding perception is not impossible in principle—but we are still a very long way off.

Question: When you look at other parts of the brain, are there any initial clues about how they are organized?
Hubel: I think there are some, but you might not accept them as bona fide clues. If you look at the organization of the entire cortex, not just visual area 1, you might ask whether the areas responsible for problem solving—the frontal lobes or parietal lobes or something like that—are organized differently. To a first approximation they are amazingly similar. Maybe the interesting differences about the organization are still concealed from us because we haven't looked at these areas in the right way, but we simply don't know. The part of the brain that I've talked about comes, roughly speaking, hard-wired: everything I have said today is true in a newborn monkey. Orientation specificity of

individual cells, for example, doesn't take any learning at all. In contrast the areas that are important for languages don't come hard-wired, at least in the sense that we are not born knowing German or any other languages.

Question: Is the auditory part of the brain hard-wired in a manner similar to the visual?
Hubel: We know enough about the auditory system to deduce that inhibition is again going to be important in titrating out excitation to produce stimulus specificity. Although we know a great deal about the response of primary auditory nerve fibers to auditory stimuli, we know very little about the central auditory system, except in a few animals like the bat that use their audition so differently that it may not even be pertinent to our understanding of language. For some reason audition in the central nervous system has been much more difficult to explore, and research has gotten off to a much slower start, but it should prove to be every bit as interesting as vision. From a superficial examination the auditory apparatus in the cortex looks not too dissimilar from the visual cortex. It has an input and an output; it has layers—but this similarity may be just like the similarity of the boxes housing the television set and the personal computer. They look superficially similar, but they are very different, except that both are crummy technologically (I mean the TV and computer!).

Question: Are there physiological differences among people or animals that lead to different responses to the same stimuli?
Hubel: The similarities are more striking than the differences. You have to get into the realm of color and other rather specific things before you find differences between a squirrel monkey and a macaque and a cat. Even though I didn't let on at the time, some of the demonstrations of complex cells that I showed you were actually done in a cat. Only an expert would know that they weren't done in a monkey. The apparatus seems very similar at these early stages in the brain.

Question: Are the cells you have described involved in dreams?
Hubel: My guess is that they are not involved. Dreams are more likely to involve cells that are several or many stages farther into the nervous system, probably in the temporal lobe. The Penfield work, in which the temporal lobe was stimulated and dream-like sequences were produced in epileptic subjects, suggests that this structure is very much involved in vision and dreaming.

Question: What about anesthetics?
Hubel: For the experiments I have described, we use a general anesthetic so that the animal is unconscious. But the same things can be tested in waking animals if you implant the electrode and then train the animal to keep its gaze riveted on the screen. In waking, purring cats, for example, or in animals that are walking around and not unhappy, you see no great differences in the response of these cells. General anesthetics probably work primarily on the reticular system, a system deep in the brain, because when you knock out that system either by lesions or concussion or general anesthetic, you lose consciousness. But the anesthetic doesn't have nearly as strong an effect on the visual regions we have tested. The cells do tend to fire more slowly and sluggishly, and we keep checking in chronically prepared animals to be sure that we are not looking at some artifact of the anesthetic. At the moment we have no doubts that the stimulus specificity we have demonstrated is independent of the anesthetic. What is not clear is how far we will be able to penetrate into the nervous system without encountering problems related to the anesthetic. For those experiments we must use chronically prepared animals, which is far more time-consuming and in the end requires more animals.

Question: Are the cells in the visual cortex used for other brain functions?
Hubel: We have not the slightest hint that these cells are involved in auditory stimuli or other sensory inputs except maybe in a very secondary way. If a chronically prepared animal is drowsy and you arouse it, say, with a ringing bell, the cells do respond better but that doesn't mean the response is specifically to an auditory signal. These cells are really very specialized. We have no guarantee that in all cases we have found the optimal stimulus, but after several years of work we are more and more convinced. We have tried numerous things and many of our enemies have too. It is a good thing about enemies; they check up on you.

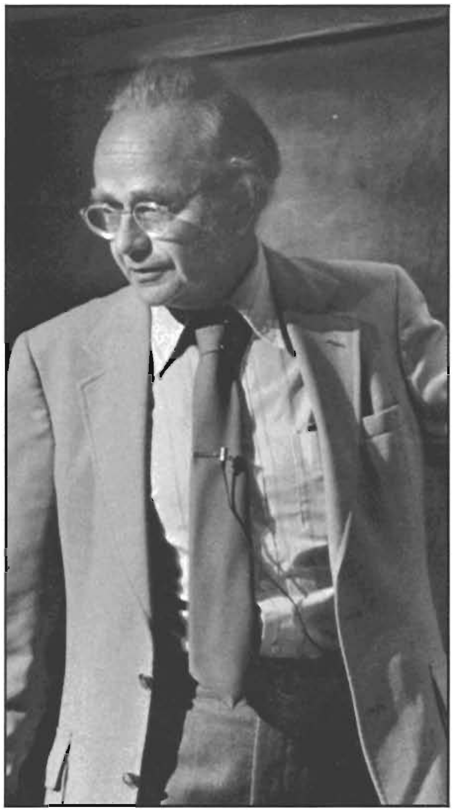
Question: Do you expect technical advances that might change the rate of progress of your understanding?
Hubel: Technical advances are certainly increasing the rate of progress in anatomy. The methods for revealing the complex, elegant systems of connected structures, such as blobs and stripes have been revolutionized in the last ten or fifteen years. Advances in physiology have come more slowly but I'd really be a pessimist if I thought that big improvements wouldn't come sooner or later.

Question: Do drugs change the sort of picture you've shown us?
Hubel: We haven't looked very much at the effects of drugs, but other people have shown that certain drugs interfere with the neurotransmitter that is largely responsible for inhibition. In particular, when the drug called bicuculine is

dumped on the cortex, the cells lose their orientation specificity and respond to all orientations. Many groups of people doing pharmacological studies are trying to unravel the visual circuitry by identifying the transmitters responsible for specific responses. This work is still in the very early stages and has yet to solve significant problems, but many people are optimistic about it.

Question: Have you tried to use some complex visual form as a stimulus and compare the responses of someone familiar with this form to someone who is not?

Hubel: We could do something like that by having a person look at something and asking whether the cells fire better when he is paying attention and so on. To do such an experiment with any effectiveness, we would have to be able to record from single human cells without going through the skull, but no technique for that is even in sight. Perhaps twenty years from now we will be able to do such things without requiring an operation. Experiments with waking monkeys have been done, but the animals require a great deal of training. The best work has been done on comparing attentiveness. If the animal is attentive to one part of his visual field and not to another, do the stimuli in the two respective places work differently? They do work differently, but showing that has been a struggle.



David H. Hubel was born in Canada of American parents, grew up in Montreal, and did honors in mathematics and physics at McGill College. He graduated from McGill Medical School and received training in neurology at Montreal Neurological Institute and Johns Hopkins Hospital. He began his studies of vision in the Neuropsychiatry Division of Walter Reed Army Institute of Research and then returned to Johns Hopkins Hospital to join the laboratory of Stephen Kuffler. There he began a collaboration with Torsten Wiesel that lasted over twenty years and led in 1981 to their receiving the Nobel Prize in Medicine and Physiology. He has been at Harvard Medical School since 1959 and is now the John Enders Professor of Neurobiology. Dr. Hubel descibed himself in *Les Prix Nobel* (Stockholm: Almqvist & Wiksell International, 1982) as follows: "Since the age of five I have spent a disproportionate amount of time on music, for many years the piano, then recorders, and now the flute. I do woodworking and photography, own a small telescope for astronomy, and I ski and play tennis and squash. I enjoy learning languages, and have spent untold hours looking up words in French, Japanese and German dictionaries. In the laboratory I enjoy almost everything, including machining, photography, computers, surgery—even neurophysiology."

Further Reading

David H. Hubel. 1988. *Eye, Brain, and Vision.* Scientific American Library series number 22. New York: Scientifc American Library.

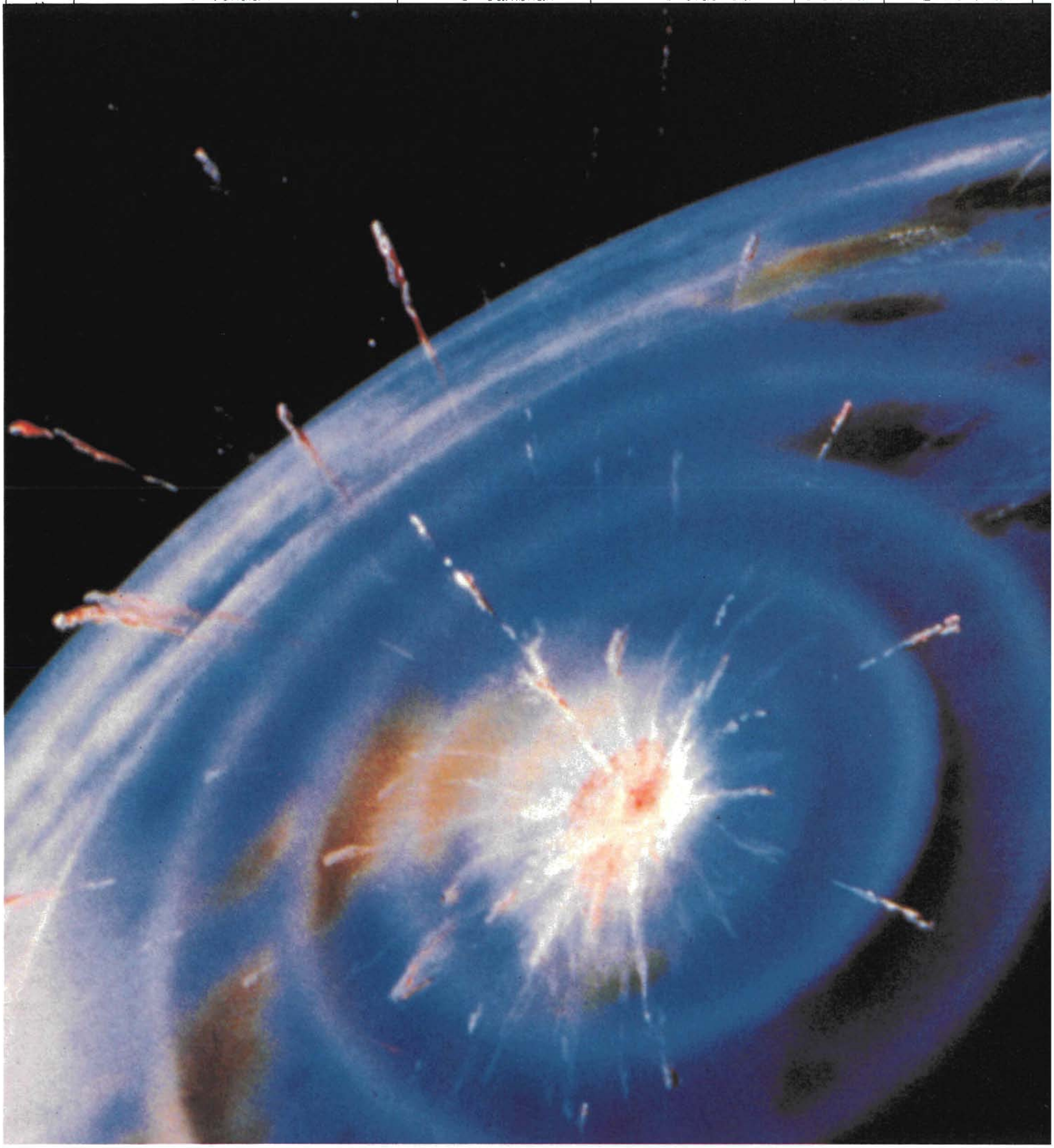
M. S. Livingstone and D. H. Hubel. 1987. Connections between Layer 4B of area 17 and thick cytochrome oxidase stripes of area 18 in the squirrel monkey. *The Journal of Neuroscience* 7: 3371–3377.

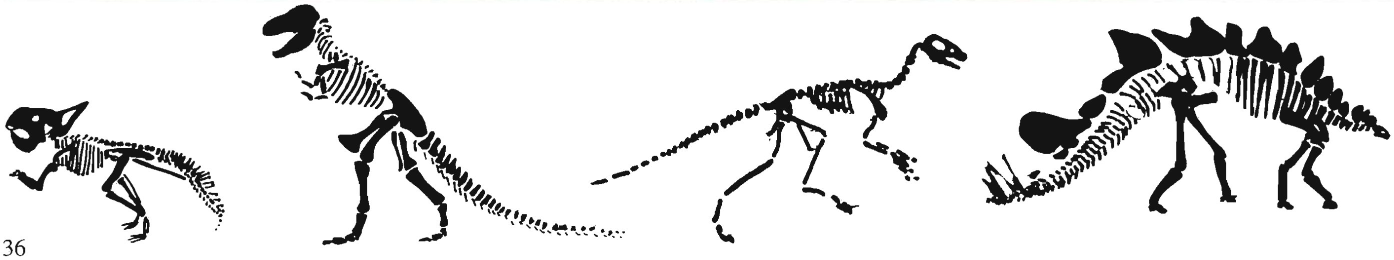
M. S. Livingstone and D. H. Hubel. 1987. Segregation of form, color, and stereopsis in primate area 18. *The Journal of Neuroscience* 7: 3378–3415.

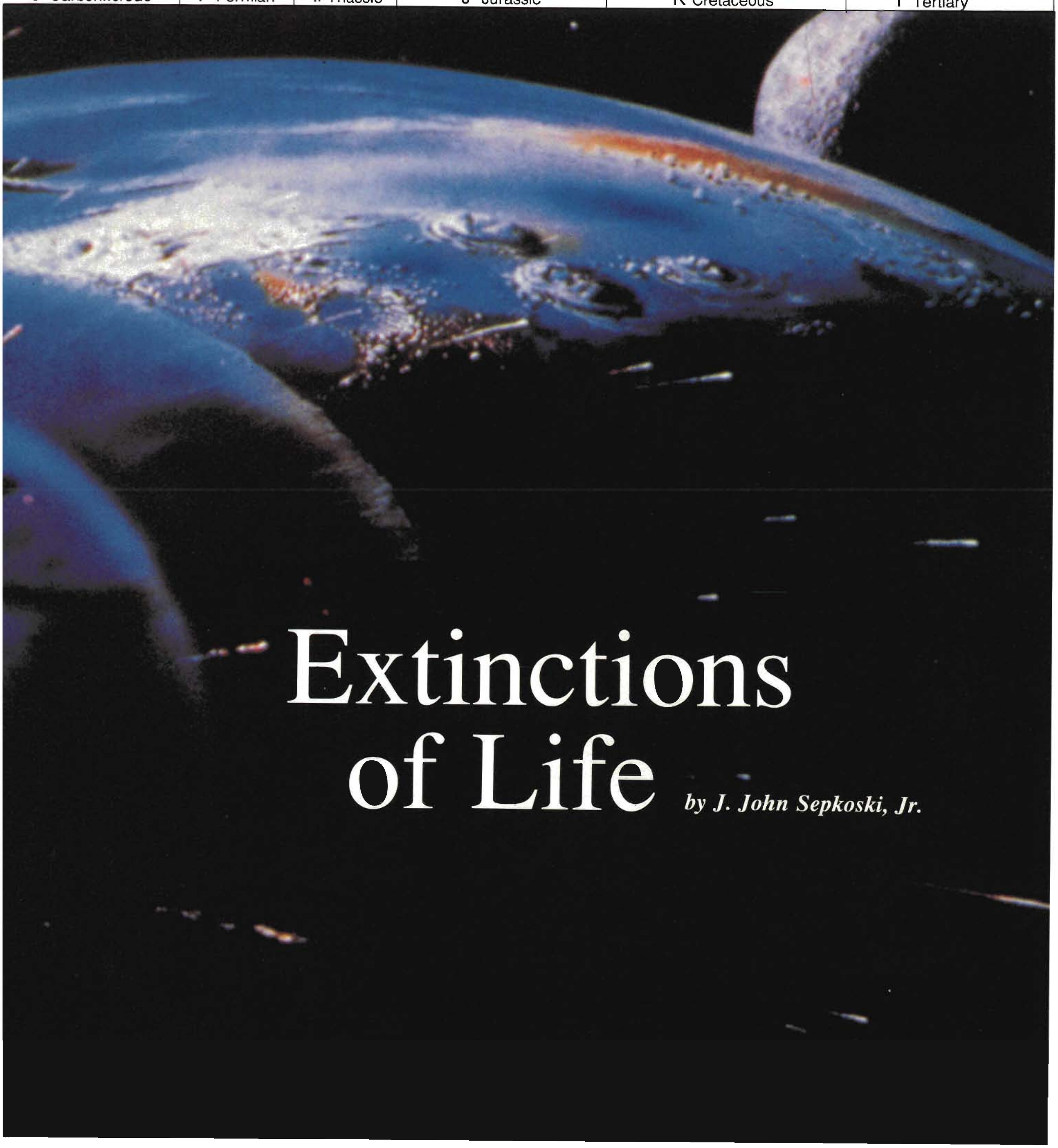
M. S. Livingstone and D. H. Hubel. 1987. Psychophysical evidence for separate channels for the perception of form, color, movement, and depth. *The Journal of Neuroscience* 7: 3416–3468.

We wish to thank Scientific American Library for permitting us to use material from *Eye, Brain, and Vision* in Figs. 1–11, 14, and 19. Figures 20 and 21 were adapted from the article "Psychophysical evidence for separate channels for the perception of form, color, movement, and depth."

Precambrian		Paleozoic				
	V Vendian	Є Cambrian	O Ordovician	S Silurian	D Devonian	

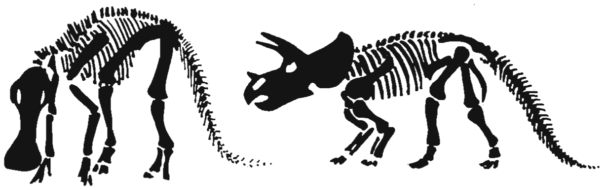


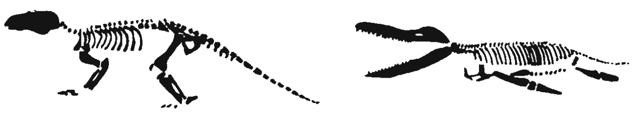




Extinctions of Life

by J. John Sepkoski, Jr.

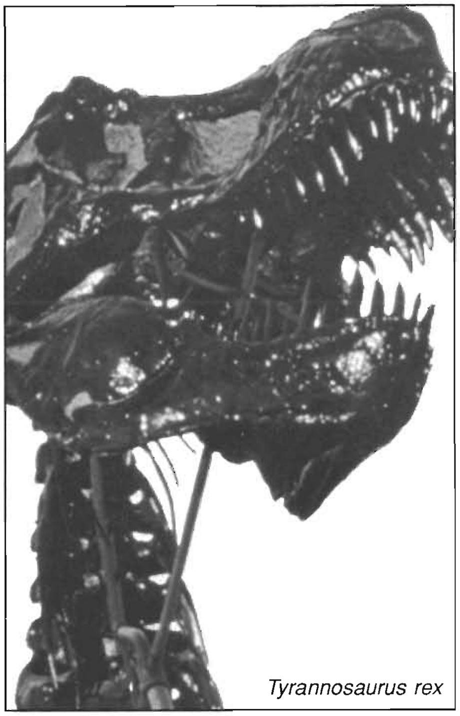




It is a delight to be here and to talk about extinctions of life, although some of you might find that title incongruous. We usually use the word *life* to refer to the collective properties of living organisms. So *extinction of life* suggests perhaps annihilation of *all* life. However, the study of extinctions is in its infancy, and in new fields, where there is much more ignorance than understanding, we often use order-of-magnitude estimates, ballpark guesses, and first approximations. Given that, the title is okay, since, to a first approximation, life *is* extinct. Probably more than 99 percent of all species that have ever lived on this planet have disappeared. The richness of the biota around us reflects only a slight excess of speciation over extinction.

Despite its magnitude and its apparent importance in the evolution of life, we know very, very little about what extinction is, as either a phenomenon or a process. How does a particular species become extinct? What array of processes are operative during an extinction? How frequently are extinctions catastrophic? How can we predict what species or what kinds of species will become extinct in a given situation? And, how can we manage the biota to control extinction in the present and future world? These are some questions that we are not sure how to answer. But they are certainly of vital contemporary importance. As more and more of the earth's surface is altered and reengineered, we are facing unprecedented levels of extinction, unprecedented at least in historical time. And as we face the possibility of nuclear winter, we need to know what that might do to the biota. Finally, from the standpoint of pure science, we want to understand how extinction has influenced the history of life on this planet and perhaps be able to make statements concerning the evolution of living systems elsewhere in the universe. So we need to know something about extinction—how it operates and what results it produces.

Courtesy Department Library Services, American Museum of Natural History. Neg./Trans. No. V/C2827



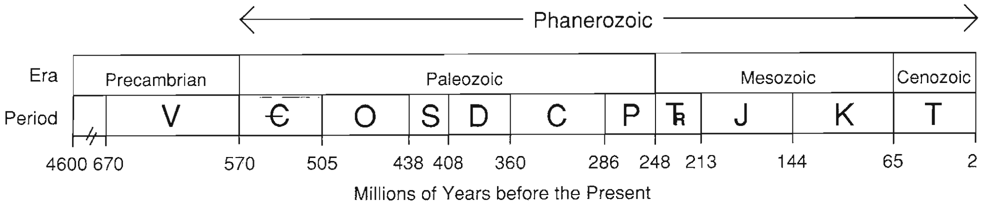
Tyrannosaurus rex

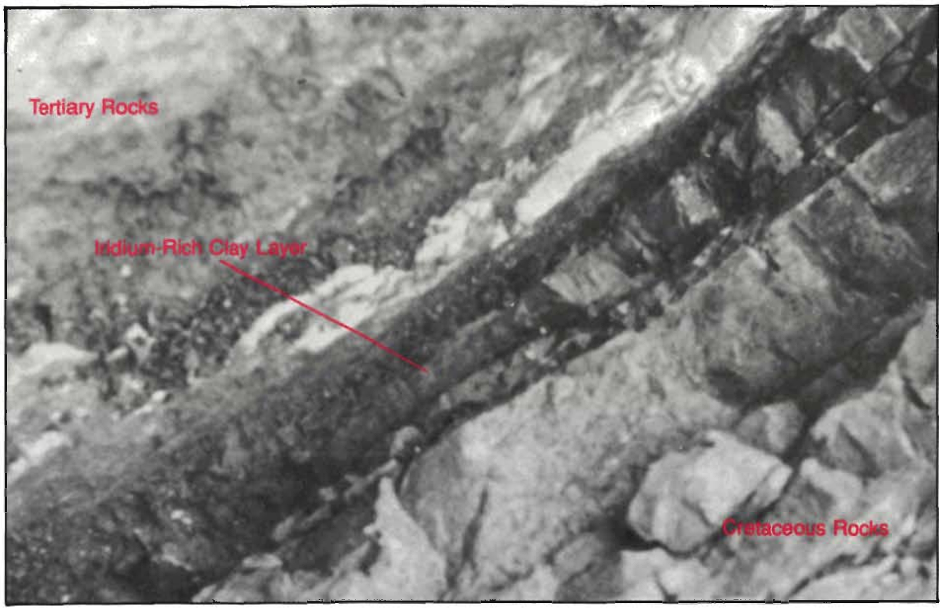
When we think about extinction, the image that immediately comes to mind is the dinosaur. Dinosaurs have been known for well over a century now, the first fossils having been recognized in the 1820s. The early conceptions about dinosaurs were that they were a strange group of animals. They were very large animals, thought perhaps to be too big for terrestrial ecosystems. They were thought to be cold-blooded, like most modern reptiles, and therefore too slow. They were thought to have too small brains and therefore to be too dumb. In a nutshell, dinosaurs were thought to have all of the characteristics that an extinct group of animals ought to have, and their disappearance seemed perfectly understandable. That of course led to the use of the epithet *dinosaur* for anything that is beyond its time and ought to be gone. I hope my students never refer to me as a dinosaur.

Many of the old ideas about dinosaurs have changed radically through research of the last few decades. We now know that not all dinosaurs were large, although the average size was fairly great. Some dinosaurs were the size of birds, and, in fact, some dinosaurs were the ancestors of birds. (Some people make the statement that dinosaurs are not extinct; they have simply taken to the trees.) We know from their morphology that some dinosaurs were very active and were probably not cold-blooded. They may have been as homeothermic as you and I are. From studies of trackways of dinosaurs as well as some of their morphological features, people have argued that dinosaurs weren't incredibly dumb animals. Some of them traveled in organized herds and probably showed some fairly complex behaviors.

Finally, we know that dinosaurs were the dominant large animals on land for about 150 million years, twice the span during which mammals have held that position. Dinosaurs arose in the late Triassic, at about the same time that mammals appeared. They then dominated the large-animal adaptive zone until they became extinct rather rapidly at the end of the Cretaceous.

The research of the last few decades turned the disappearance of this very symbol of extinction into very much of an enigma. Many speculations were





IRIDIUM-RICH DEPOSIT AT CRETACEOUS-TERTIARY BOUNDARY

Fig. 1. Close-up of the iridium-rich clay layer at the boundary between Cretaceous and Tertiary rocks in a stratigraphic section near Gubbio, Italy. The high iridium content of the clay (see Fig. 2) is attributed to the impact with the earth of an extraterrestrial body. Since discovery of the Gubbio anomaly in 1978, deposits similarly rich in iridium have been found at Cretaceous-Tertiary boundaries worldwide. (Photo courtesy of Alessandro Montanari, Department of Geology and Geophysics, University of California, Berkeley.) ◀

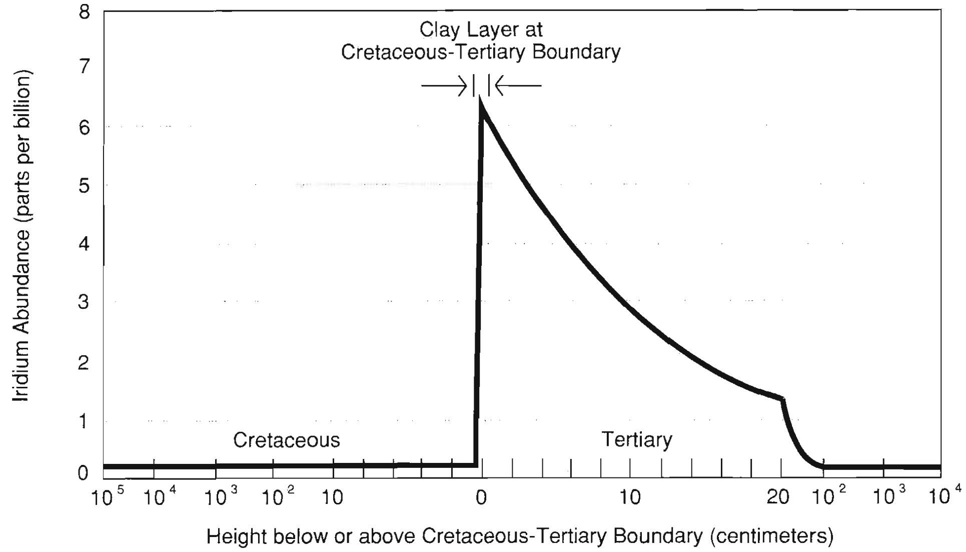
published on what circumstances might have caused dinosaurs to become extinct, but none seemed very satisfying, at least not until a discovery by Luis and Walter Alvarez in 1979.

Most of you are probably familiar with that discovery. Walter Alvarez was looking at some stratigraphic sections, near Gubbio in central Italy, that span the Cretaceous-Tertiary boundary. He saw a peculiar clay layer, 1 to 2 centimeters thick, sandwiched between older Cretaceous rocks and younger Tertiary rocks (Fig. 1). Walter was curious about the clay and sent it back to his father for analysis. Luis, Frank Asaro, and Helen Michel performed a number of geochemical analyses of the clay and found that it contained an excess of iridium (Fig. 2). The excess was far too large to explain on the basis of terrestrial surface sources, which are highly depleted in iridium. They hypothesized that the excess iridium was due to the impact of a large—perhaps 10 kilometers in diameter—extraterrestrial object on the last day of the Cretaceous.

Now an impact by a 10-kilometer-diameter object would wreak havoc on the earth. Various scenarios, which differ quantitatively but agree qualitatively, suggest that huge amounts of dust were thrown into the atmosphere, blocking out sunlight for perhaps three months. The impact may have first heated the atmosphere and then cooled it. It may have produced large amounts of nitrogen oxides, which would rain down as nitric acid. The list of damages can go on and on.

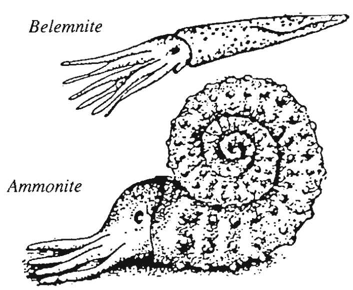
THE ALVAREZ IRIDIUM ANOMALY

Fig. 2. A plot, versus height above or below the Cretaceous-Tertiary boundary, of iridium abundance in various stratigraphic sections from the vicinity of Gubbio, Italy. The abundance rises abruptly at the end of the Cretaceous to a value some twenty-five times greater than the background level and then falls back to that level within approximately 15,000 years. (Figure adapted from "Current status of the impact theory for the terminal Cretaceous extinction" by Walter Alvarez, Luis W. Alvarez, Frank Asaro, and Helen V. Michel. In Silver and Schultz 1982, 305–315.) ▼



Various climatic and chemical models suggest that the earth wouldn't have been a very pleasant place on the last night of the Cretaceous, that is, the three-month night that followed the impact. Photosynthesis by green plants would have been shut down, and large herbivores that fed upon them would have starved, as would the carnivores that stalked the plant eaters. The expected result would be extinctions. We don't have an equation relating the species that would become extinct to the size of the impacting body. The one thing we do know is that if things got as bad as the models predict, more kinds of animals than just dinosaurs should have become extinct. And indeed that is what the fossil record shows. The flying reptiles, which had a long history in the Mesozoic, vanished at the end of the Cretaceous. In the oceans the large marine reptiles, such as the plesiosaurs, disappeared. So did a large number of marine invertebrates, including the ammonites (well-known marine fossils of the Mesozoic), almost all of the belemnites, and a large variety of clams, snails, crabs, bryozoans, and brachiopods.

Thus a whole suite of organisms became extinct at the same time that the dinosaurs did. From the fossil record we can estimate that about 45 percent of marine animal genera became extinct at the end of the Cretaceous. Extrapolating down to the species level leads to estimates that 60 to 75 percent of marine species became extinct in the last 2 million years or less of the Cretaceous period. So whatever happened was indeed quite devastating to the marine biota.



How do we know what became extinct? How do we make quantitative estimates of the magnitudes of mass extinctions? Paleontologists use two basic methods to study mass extinctions and other events in geologic history. The traditional method is to collect information about the types and numbers of fossils in the various strata of outcrops or core samples and then to determine the times when the various fossil taxa first appeared, flourished, and then disappeared. Such data are then used to assess patterns of origination and extinction and perhaps to test hypotheses concerning those phenomena.

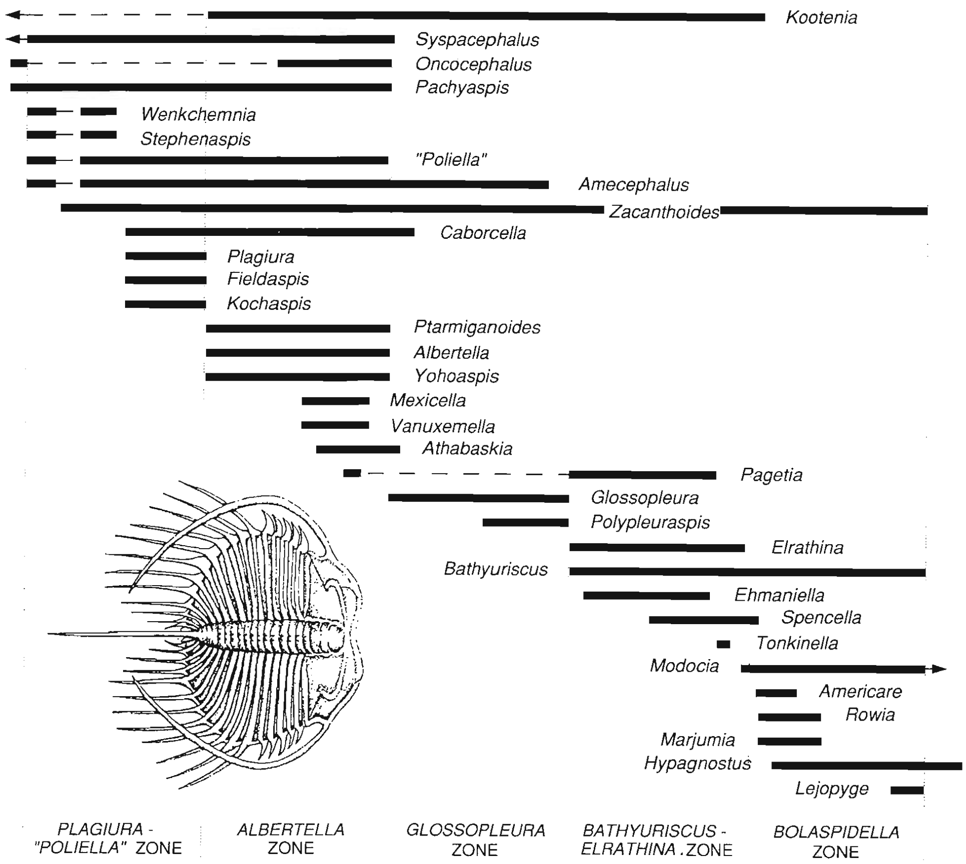
This "normal" methodology gives many details about extinction, such as the abundance of an organism before its disappearance and the time scale of its disappearance. But usually such data are available only for a single group—a single order or class or even phylum—in a rather local region of the earth. And amassing the data is very labor-intensive. Despite a century and a half of work by paleontologists worldwide, we still have detailed data on patterns of extinction for only a small number of localities, a small number of time intervals, and a small number of taxa.

To sidestep the gaps in the detailed paleontological data—and to supplement them—a second way of studying mass extinctions has been developed. This second way has been the subject of my work. Rather than studying detailed information on local patterns of extinction over relatively short time intervals, I am trying to discern global patterns over longer time intervals. My approach is analogous to deducing the population demographics of ancient peoples from the spotty records available. What records have been unearthed are assembled and correlated, as well as possible considering the many records that are missing. The focus is not on individuals but on some higher group—families, perhaps, or tribes.

Like historical census data, the fossil record is incomplete, covering only a small sample of the earth's biota. Still, it contains a huge number of species from all parts of the world—too much data to assess well. We therefore usually work at higher taxonomic levels, such as the genus or the family. We lose resolution doing that but sometimes get a better overall picture, because a genus, say, is included in our data set even if all but one of its species are missing from the fossil record.

I have attempted to obtain data on all animals but have concentrated most of my attention on marine organisms. The reason for doing so is that, although terrestrial organisms, such as dinosaurs, flying reptiles, and giant mammals, are certainly more spectacular, our fossil record for them is far poorer than that for marine organisms. After all, land is an area of net erosion, as you can certainly see in the environs of Los Alamos. The oceans are areas of net sedimentation. They end up with a larger and more complete fossil record that, for various historic and economic reasons, has been far better explored and far better studied.

The detailed data collected by paleontologists are usually presented as "biostratigraphic range charts." Figure 3 is an example showing data for the occurrence of trilobite genera in Middle Cambrian strata in western North America. Note that even this study dealt not with species but with genera. Note also that the geologic zones are not plotted according to scale. That is, the time interval spanned by each zone is not the same, although each is allotted an equal space on the chart. We don't have good estimates of the duration of those geologic time intervals since our methods for determining time in the Cambrian are not accurate enough. Furthermore,



BIOSTRATIGRAPHIC RANGE CHART

Fig. 3. This chart presents paleontologic data for the time ranges of trilobite genera through the stratigraphic zones of the Middle Cambrian period in western North America. Dashed lines indicate lack of field data. (Figure adapted from *The Cambrian System in the Southern Canadian Rocky Mountains, Alberta and British Columbia (Second International Symposium on the Cambrian System, Guidebook for Field Trip 2)*, compiled by James D. Aiken, edited by Michael E. Taylor, 31. Denver, Colorado: U.S. Geological Survey, International Union of Geological Sciences, Geological Survey of Canada, 1981.) ▶

geologic time intervals usually can be accurately characterized only over local areas.

Putting together data for all fossil marine taxa from all over the world, we come up with something like a small telephone book. Figure 4 is a page from such a compilation giving first and last appearances in the fossil record for Cambrian and Ordovician trilobite families. The data set I have assembled covers about 3500 marine families and about 30,000 marine genera.

To develop some picture of extinction patterns from such a data set, the simplest thing to do is to count the number of families or genera that are present in each time interval. In the case of families, 77 standard geologic time intervals compose the last 600 million

EXTINCTION DATA FOR TRILOBITE FAMILIES

Fig. 4. A page from a summary of data on the appearance and disappearance worldwide of marine families. The data shown are those for trilobites. The abbreviations in parentheses denote subdivisions of the Cambrian and Ordovician geologic periods. (Figure adapted from *A Compendium of Fossil Marine Families* by J. John Sepkoski, Jr. Milwaukee Public Museum Contributions in Biology and Geology Number 51. Milwaukee, Wisconsin: Milwaukee Public Museum Press, 1982.) ◀

Class Trilobita	
Order Agnostida (= Miomera)	
Agnostidae	Ꞓ (Boto) — O (Ashg)
Clavagnostidae	Ꞓ (uMid) — Ꞓ (Dres)
Condylopygidae	Ꞓ (Boto) — Ꞓ (uMid)
Diplagnostidae	Ꞓ (mMid) — O (Trem)
Discagnostidae	Ꞓ (Dres)
Eodiscidae	Ꞓ (Atda) — Ꞓ (uMid)
Pagetiidae	Ꞓ (Atda) — Ꞓ (mMid)
Phalacromidae	Ꞓ (uMid) — Ꞓ (Dres)
Sphaeragnostidae	O (Ashg)
Trinodidae	Ꞓ (Dres) — O (Ashg)
Order Redlichiida	
Abadiellidae	Ꞓ (Atda) — Ꞓ (lMid)
Bathynotidae	Ꞓ (Boto) — Ꞓ (lMid)
Chengkouiidae	Ꞓ (Boto)
Daguinaspididae	Ꞓ (Atda)
Despujolsiidae	Ꞓ (Atda)
Dolerolenidae	Ꞓ (Atda) — Ꞓ (Boto)
?Ellipsocephalidae	Ꞓ (Atda) — Ꞓ (mMid)
Emuellidae	Ꞓ (lMid)
Gigantopygidae	Ꞓ (Boto)
Hicksiidae	Ꞓ (Boto)
Kueichowiidae	Ꞓ (Boto)
Longduiidae	Ꞓ (Boto)
Mayiellidae	Ꞓ (Boto)
Neoredlichiidae	Ꞓ (Atda) — Ꞓ (Boto)
Olenellidae	Ꞓ (Atda) — Ꞓ (mMid)
Paradoxididae	Ꞓ (Atda) — Ꞓ (uMid)
Protolenidae	Ꞓ (lTom) — Ꞓ (mMid)
Redlichiidae	Ꞓ (Atda) — Ꞓ (mMid)
Saukiandidae	Ꞓ (Boto)
Yinitidae	Ꞓ (Atda) — Ꞓ (Boto)
Yunnanocephalidae	Ꞓ (Atda)

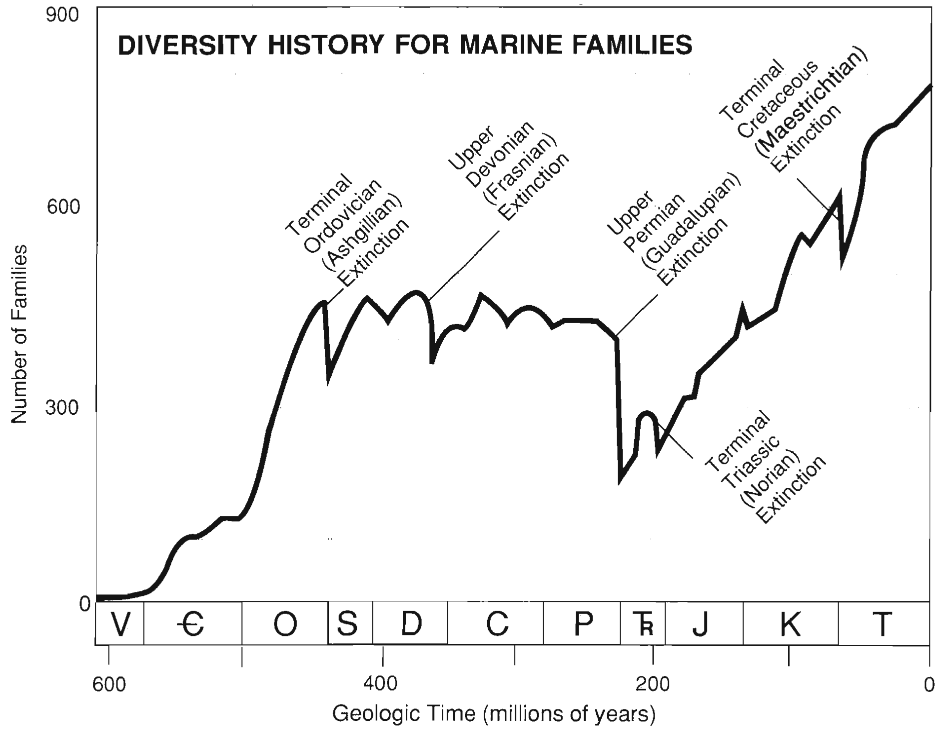
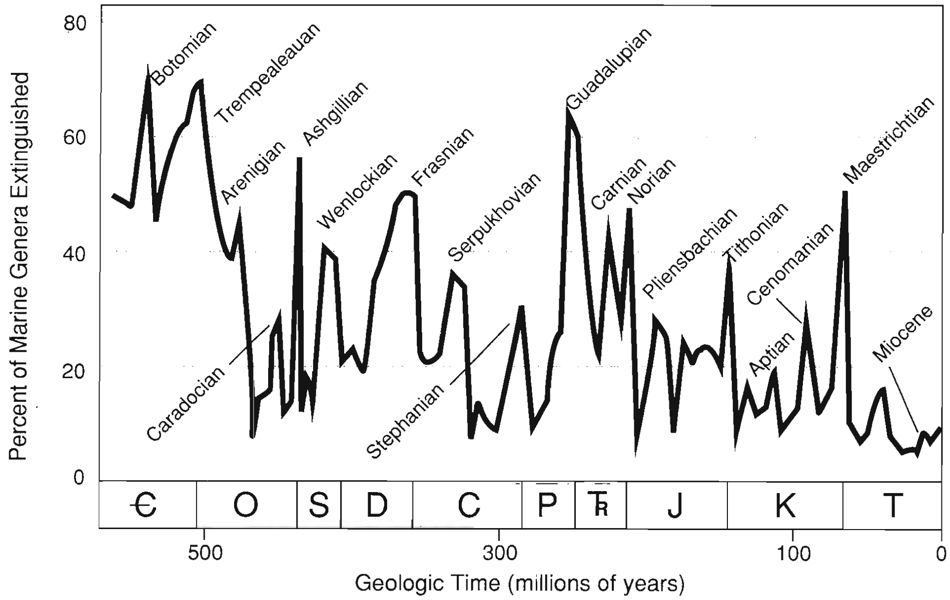


Fig. 5. This history of marine animal diversity reveals five principal mass extinctions, of which the upper Permian, or Guadalupian, was by far the most devastating. Lesser extinction events are also visible. (Figure adapted from "Mass extinctions in the Phanerozoic oceans: A review" by J. John Sepkoski, Jr. In Silver and Schultz 1982, 283–289.) ◀

EXTINCTION RATE HISTORY FOR MARINE GENERA

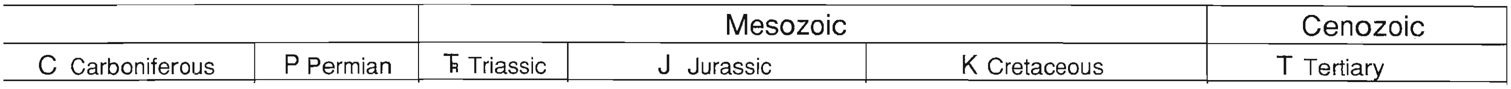
Fig. 6. This history of extinction rates shows more clearly than the diversity curve (Fig. 5) the many extinction events experienced by marine fauna. (Figure adapted from "Phanerozoic overview of mass extinction" by J. J. Sepkoski, Jr. In *Patterns and Processes in the History of Life (Report of the Dahlem Workshop on Patterns and Processes in the History of Life, Berlin 1985, June16–21)*, edited by D. M. Raup and D. Jablonski, 277–295. Berlin: Springer-Verlag, 1986.) ▼



years, which is often referred to as the Phanerozoic, the eras of geologic time for which evidence of animal life on the earth is abundant. For genera the data base I have is a little better, composed of about 100 time intervals (attained by carefully subdividing some of the longer standard intervals).

Figure 5 is a plot of the number of marine animal families versus time interval. The big mass extinctions show up as large and rapid drops in the number of families. As you can see, the terminal Cretaceous, or Maestrichtian, extinction, the one that led to the demise of the dinosaurs on land, was fairly rapid but not excessively large. About 17 percent of marine animal families disappeared in that time interval. Because the disappearance of a family requires the disappearance of every genera and species within the family, a family kill of about 17 percent corresponds to a genus kill of about 45 percent and a species kill of around 60 to 75 percent.

The terminal Cretaceous event certainly isn't the only large mass extinction we see in Fig. 5. And it certainly isn't the largest. The largest was the Guadalupian at the end of the Permian, when about 55 percent of marine families became extinct. Virtually every order and class of marine organisms lost an extensive number of families. Going through the same sort of extrapolation,



we find that about 80 percent of marine genera and perhaps more than 95 percent of marine species disappeared at the end of the Permian period. Other major events visible in Fig. 5 include one at the end of the Ordovician, which is probably the second largest extinction of marine animal fauna. But it is not that much larger than the one at the end of the Cretaceous. Another extinction occurred in the late Devonian, and another in the late Triassic, right on the tail of the Guadalupian extinction.

In addition to the large mass extinctions, many smaller extinction events have occurred—in fact, around two dozen. Simple diversity data don't reveal the smaller extinctions, but other metrics of extinction intensity do.

Figure 6 shows one such metric, a plot of the extinction rate for marine genera in each of the hundred or so sampling intervals spanning the Phanerozoic. Most of the spikes, or local maxima, correspond to extinction events. The larger spikes—the Maestrichtian, the Norian, the Guadalupian, the Frasnian, and the Ashgillian—are the same major mass extinctions that we see in the familial diversity data (Fig. 5). Many of the other spikes have been recognized by paleontologists in detailed field data on localized regions and restricted groups of organisms.

TIMING OF MARINE GENERA EXTINCTIONS

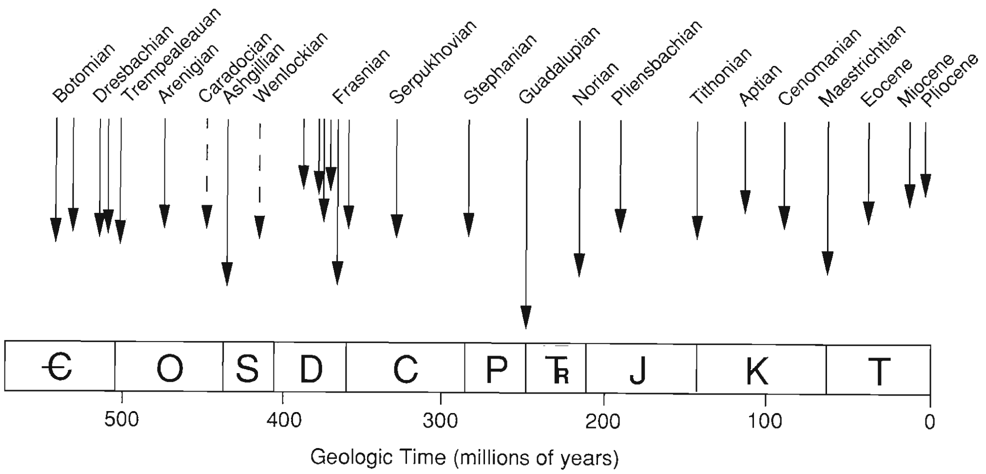


Fig. 7. At least during the most recent 300 million years of geologic time, extinctions have occurred with considerable regularity, as this display of the data of Fig. 6 reveals. The lengths of the arrows indicate the magnitudes of the extinction rates. (Figure adapted from "Phanerozoic overview of mass extinction" by J. J. Sepkoski, Jr. In *Patterns and Processes in the History of Life (Report of the Dahlem Workshop on Patterns and Processes in the History of Life, Berlin 1985, June 16–21)*, edited by D. M. Raup and D. Jablonski, 277–295. Berlin: Springer-Verlag, 1986.)

The data of Fig. 6, particularly when displayed as in Fig. 7, reveal a very interesting feature of extinctions—a remarkable regularity in their timing during the past 300 million years. That observation was first made by Al Fisher in the late seventies and was then rediscovered by my colleague David Raup and me about five years ago when we were looking at the family data.

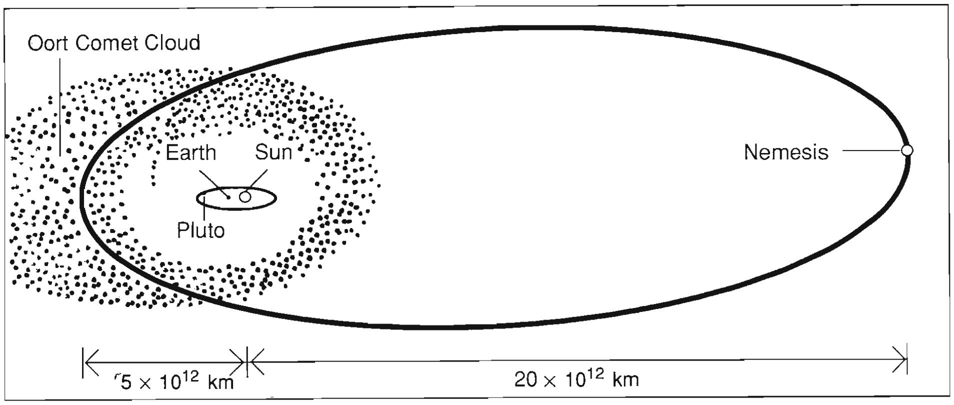
Figure 8 is another attempt to portray the regularity. Here I have simply assigned a cycle number to extinction events during the last 250 million years and plotted the cycle numbers against the estimated times of the events. Note the good fit of the data points to a straight line, which indicates a constant, or stationary, periodicity. Dave Raup and I have performed a variety of analyses and have found that the probability of such a periodic extinction pattern occurring at random is extremely low. A stationary periodicity describes the extinction events far better than any sort of random or semi-random model we can conceive of. I am quite convinced that, at least over the last 250 million years of the earth's history, extinctions have occurred with a stationary periodicity of 26 million years.

That observation, however, does not agree with traditional views of mass extinctions, which implicitly assume that each extinction event was produced independently by some random environmental perturbation or perhaps by a random coincidence of several environmental variables. And, since each extinction event was independent of the others, it therefore could be studied independently. But if the extinction events recur regularly, they cannot be independent of one another, at least not in terms of their timing. Perhaps we are dealing with a series of events caused by a single, ultimate forcing agent that has clock-like behavior.

When Dave Raup and I published that speculation, we didn't know what the agent was. However, one event, the terminal Cretaceous mass extinction, was known to be associated with the impact of a large extraterrestrial object with the earth. If an impact caused one mass extinction in the periodic sequence, perhaps impacts caused all the others as well.

The idea that most mass extinctions, at least over the last 250 million years, are the result of impacts of one or more extraterrestrial bodies leads of course to the next question: What could be the cause of regularly periodic impacts? Several hypotheses have been offered;

the best known is the Nemesis, or so-called death-star, hypothesis (Fig. 9), which was put forward independently by several groups. The idea is that the sun is not alone, that it is accompanied by a small companion star in a highly elliptic orbit with an orbital periodicity of 26 million years or so. That companion, Nemesis, is usually far from the sun, but during the small portion of its period when it is passing through the Oort Cloud, it scatters up to a billion comets into the inner solar system. Jack Hills has calculated that, out of that billion or so comets, perhaps an average of about two dozen of various masses hit the earth, wreaking havoc and causing extinction of many species on land and in the ocean.



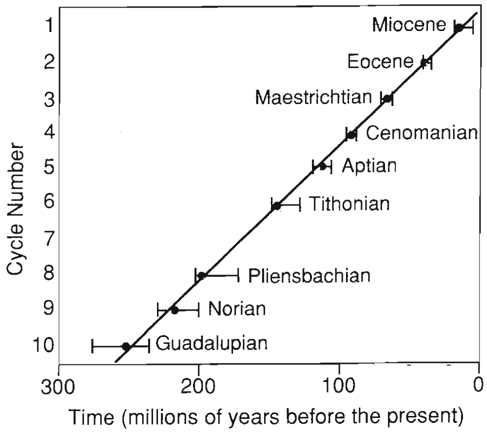
THE NEMESIS HYPOTHESIS

Fig. 9. The Nemesis hypothesis has been proposed as an explanation for the apparent regular periodicity of extinctions. According to that hypothesis, Nemesis, a companion star of the sun, scatters comets into the inner solar system when it passes through the Oort Cloud every 26 million years. The impacts of a small number of the scattered comets with the earth cause the observed extinctions.

Several years ago we were all very excited about such ideas, but time has tempered our excitement somewhat. Some of the predictions of the models are now looking a little cloudy, if you will permit me. Carl Orth, Frank Kyte, and others have failed to find iridium or other geochemical anomalies appearing consistently with the various periodic extinctions. Although an iridium anomaly and microtektites are associated with the Eocene extinction event, the one that occurred about 26 million years after the end of the Cretaceous, there is no good evidence of impact signatures at many of the others. Also, Nemesis has not yet been found, and there are some unresolved theoretical problems with the death-star hypothesis, expecially about the stability of the companion star's orbit.

REGULAR PERIODICITY OF MESOZOIC EXTINCTIONS

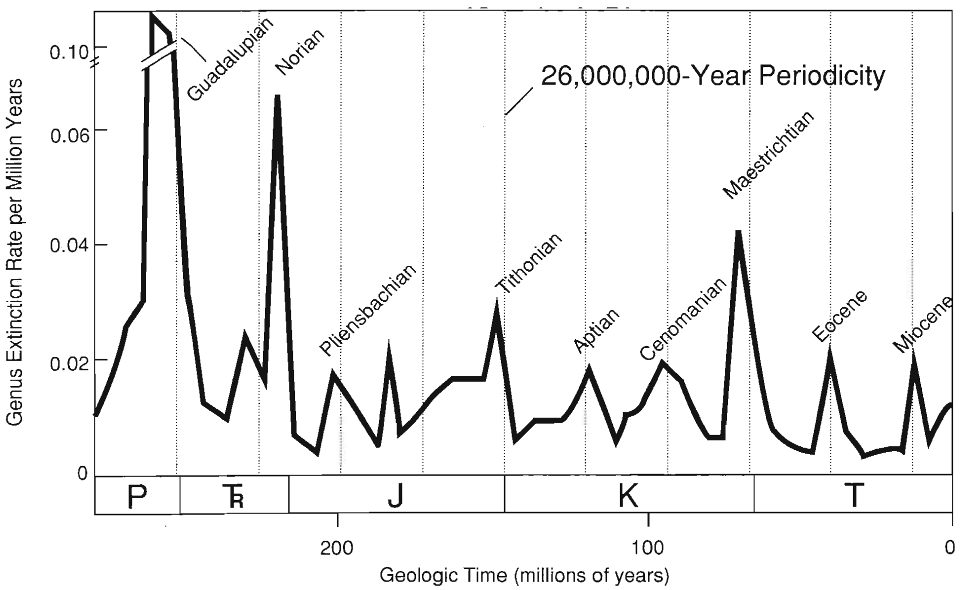
Fig. 8. The data points in this graph consist of "cycle numbers" assigned to the Mesozoic and Cenozoic extinction events and the times of their occurrence. The good fit of the points to a straight line indicates that the extinctions are regularly periodic. (Figure adapted from "Periodicity in marine extinction events" by J. John Sepkoski, Jr., and David M. Raup. In *Dynamics of Extinction*, edited by David K. Elliott, 3–36. New York: John Wiley & Sons, 1986.)



The only thing that I think has survived is the regular periodicity of the extinctions. To me that still looks good, especially now that some of the gaps in the periodic series have been filled in. But some better data have also led to new observations and new questions.

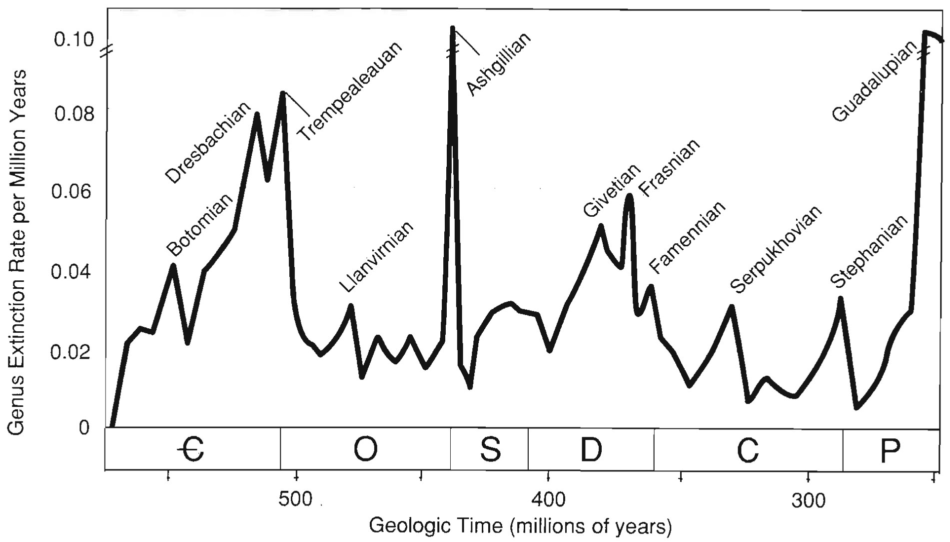
One of the more remarkable observations that come from the latest generation of data is shown in Fig. 10, a plot of the per genus extinction rate per million years. That metric is essentially the probability of extinction per time interval. Figure 10 seems to show a remarkable uniformity not only in the timing but also in the magnitude of the smaller extinction events. Within the resolution of the data, the smaller events are identical in amplitude. In addition to the smaller, constant-amplitude events, we have a few outliers, particularly the Maestrichtian, Norian, and Guadalupian events. Perhaps—and this is pure speculation now—the impact or whatever it was that happened at the end of the Cretaceous, say, was simply coincidental with a peak of extinction produced by an independent periodic forcing agent, and the combination of the two caused absolute havoc. But if the impact had occurred in a trough between periodic events, it would have caused a much smaller, aperiodic extinction event.

Figure 11 is a similar plot for the Paleozoic era. The extinction peaks in the Permian and Carboniferous periods still give an impression of some regularity in their timing. There is a little more vari-



COMPARISON OF 26,000,000-YEAR PERIODICITY AND MESOZOIC EXTINCTION PEAKS

Fig. 10. This superposition of a 26,000,000-year periodicity on data for genus extinction rates during the Mesozoic shows how closely such a regular periodicity fits the extinction peaks. Note also the similarity in magnitude among most of the extinction rate peaks. (Figure adapted from Sepkoski 1986.)

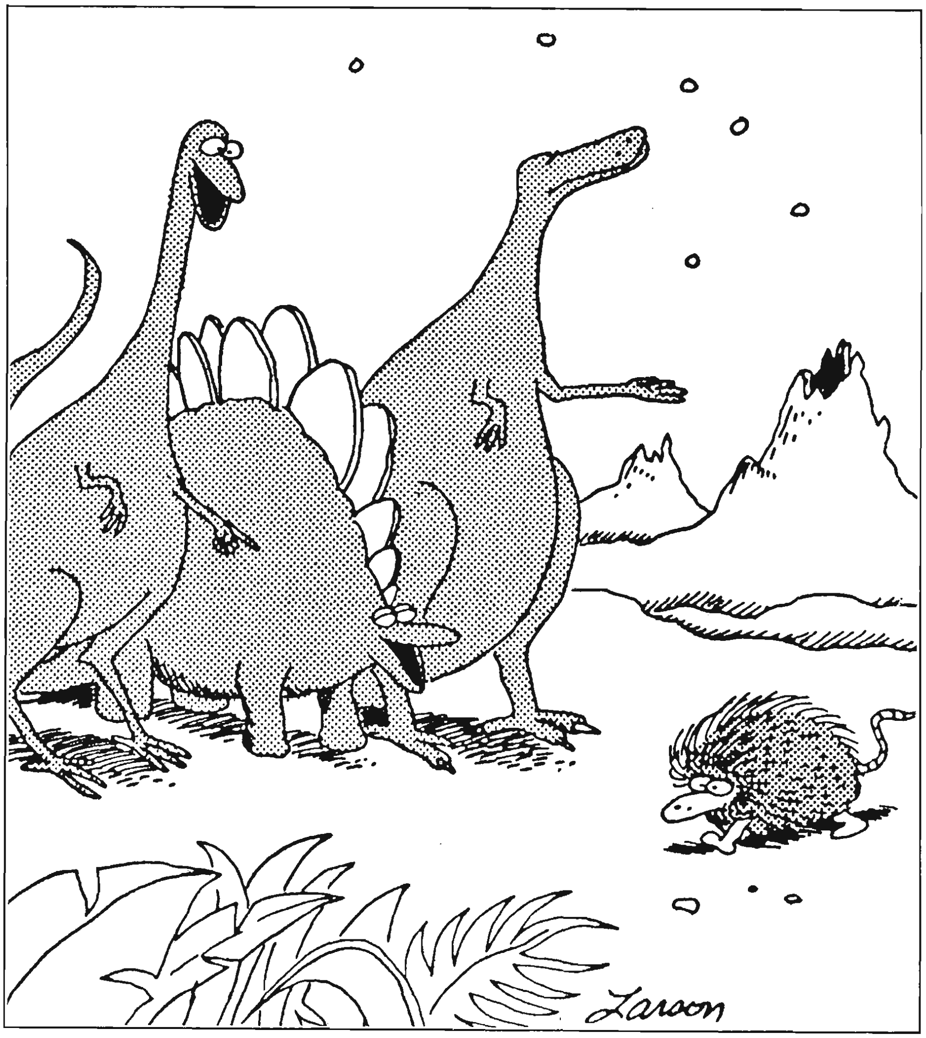


PALEOZOIC EXTINCTION PEAKS

Fig. 11. During the Permian and Carboniferous periods of the Paleozoic era the peaks of the genus extinction rate history exhibit a fairly regular periodicity but one closer to 30 to 35 million rather than 26 million years. In contrast, the earlier extinction peaks (during the Devonian, Silurian, Ordovician, and Cambrian periods) seem to lack any periodicity. (Figure adapted from Sepkoski 1986.)

ation in the timing, but then our ability to estimate geologic time during that era isn't so good. However, our best estimates suggest that the spacing between the Permian and Carboniferous events is on the order of 30 to 35 million years, somewhat longer than the 26-million-year spacing between the Mesozoic events. Perhaps that indicates a variable periodicity. Back beyond the Carboniferous the pattern seems to break down into chaos. It is not clear whether the lack of pattern, or at least of periodic pattern, represents problems with the fossil data or with our ability to tell geologic time accurately. It is also possible that the apparently chaotic pattern reflects a combination of periodic and aperiodic events. And it is eminently possible that there is no periodicity at all in the Paleozoic.

Despite the many unanswered questions about extinction, one thing is clear: Many extinction events have occurred, some of them rather large. And that fact raises a question that's not easy to answer: What are the effects of those frequent extinction events on the course of evolution, on the history of the earth's biota? Our feeling is that the effects were more profound than the simple elimination of various



This cartoon, from Beyond the Far Side, is reproduced with the permission of Chronicle Features.

taxa, such as the "outmoded" dinosaurs. Indeed, the extinction events may have had some very constructive effects.

Looking back at Fig. 5, we see that the number of marine families rises rapidly to a sort of equilibrium during the Paleozoic era. That equilibrium is punctuated by extinction events of various amplitudes but the system seems to rebound and to fill up again rather quickly. Then the great Permian mass extinction seems to destabilize the system, and the subsequent number of families rises above the former equilibrium value. But, in fact, arguments can be made that diversity was already increasing before that event, and what appears to be a great increase in the number of families during the Mesozoic and early Cenozoic eras is a combination of rebound from the Permian event and a natural rise that would eventually have moved asymptotically toward a greater equilibrium value.

The reason the system fills up is that the whole-ocean ecosystem is finite in terms of habitat space and other resources. Therefore it can hold only a limited number of kinds of animals. And the reason the system rebounds very quickly after an extinction event is that ecospace has been opened up, which leads to a very rapid radiation into specialized taxa. Even during the long-term rise in diversity during the Mesozoic and Cenozoic, we see rapid rebounds after the Norian and Maestrichtian mass extinctions. So those large mass extinctions are opening up ecospace and promoting very rapid evolution in their wake.

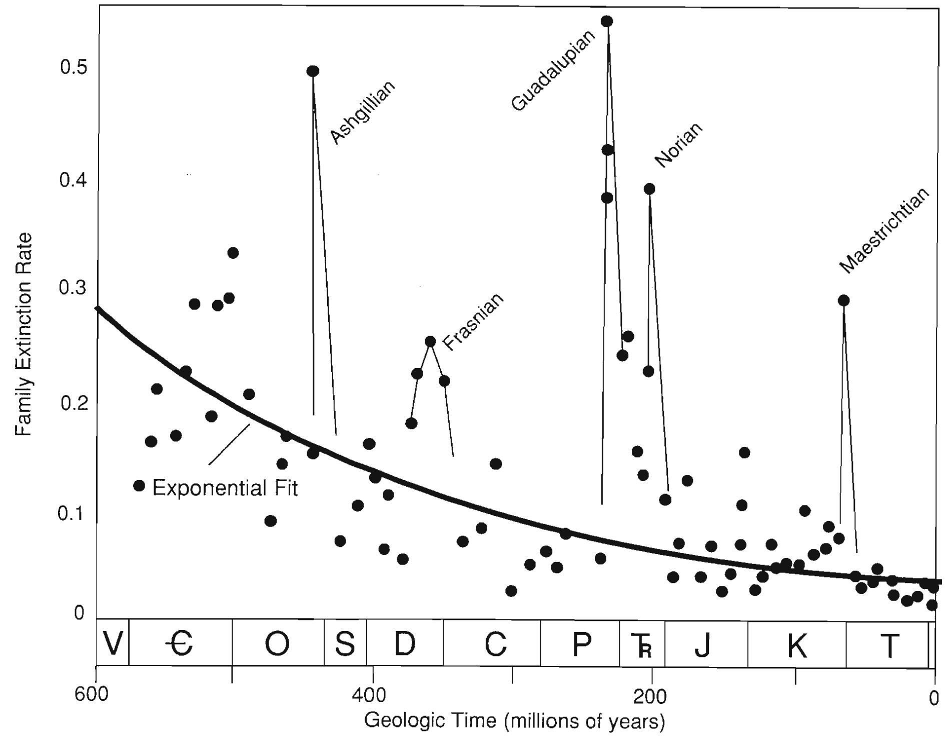
Let's look at evolutionary innovation, that is, at the appearance of new kinds of animals, in the marine system. We find that after the Ordovician period nearly two-thirds of the new taxonomic orders that appeared in the oceans originated during the rebounds that followed extinction events. Those rebounds, though, constitute only one-third of the time span. So mass extinctions increase evolutionary innovation by a factor of about 2 and in that sense seem to be filling a creative, constructive role. However, the best example of this by far is seen not in the oceans but on land.

Did you know that your ancestors were vermin? During most of its history, the mammal class consisted of tiny quivering vermin living in the interstices of a dinosaurian world. Mammals have been pre-eminent only for the last 65 million years, that is, only following the rapid extinction of dinosaurs. Within approximately a dozen million years of the early Tertiary, virtually every modern order of mammal—from mice to whales, from bats to elephants—appeared in terrestrial ecosystems. It was as if an inhibiting force on innovative mammalian evolution had been lifted with elimination of the dinosaurs.

This constructive role of mass extinction might be absolutely necessary in the earth's evolutionary system and perhaps in evolutionary systems elsewhere in the universe, as George Wald certainly argued and I'm sure Frank Drake will argue.

Another feature of extinction in general that increases the evolutionary importance of the large mass extinctions is the following. In some of the graphs shown previously, you may have noticed a secular decline in the "background" extinction rate through the Phanerozoic. (Background extinction is total extinction minus that occurring during the big mass extinctions.) The rates tend to be very high early in the Cambrian and decline through the later Phanerozoic. Figure 12 shows how a simple exponential fits that decline for marine families. The decline suggests that marine taxa are becoming more and more resistant to whatever processes cause extinction, at least at the family

		Mesozoic			Cenozoic
C Carboniferous	P Permian	₮ Triassic	J Jurassic	K Cretaceous	T Tertiary



DECLINE OF BACKGROUND EXTINCTION

Fig. 12. Extinction is an ever-present feature of geologic history. The background extinction (that is, total extinction minus the large peaks of extinction) shows a decline throughout the Phanerozoic that is fitted quite well by a simple exponential. Such a decline has implications for evolutionary innovation. (Figure adapted from "Some implications of mass extinction for the evolution of complex life" by J. John Sepkoski, Jr. In *The Search for Extraterrestrial Life: Recent Developments (Proceedings of the 112th Symposium of the International Astronomical Union held at Boston University, Boston, Mass., U.S.A., June 18–21, 1984)*, edited by Michael D. Papagiannis, 223–232. Dordrecht, Holland: D. Reidel Publishing Company, 1985.)

level. We might speculate that background extinction will asymptotically grind to a halt. If that should happen and if no more mass extinctions occur, there would be very little potential for evolutionary innovation or for further evolutionary development of the ecosystem. The evolutionary machine might not halt completely, but it would certainly slow down without major mass extinctions to reset it. Thus extinctions may be a necessary force in the development of complex life and, from what we see of patterns at the end of the Cretaceous, perhaps even for the appearance of consciousness in an evolutionary system.

I hope I have shown that our understanding of extinction is still very limited and that this aspect of the science of life presents numerous unsolved problems. ■

Questions and Answers

Question: Has anybody tried to correlate the dates of large craters with those of extinction events? Also, a lot of meteors are carbonaceous chondrites, which probably wouldn't be expected to contain much iridium. So wouldn't it be a mistake to say that if you don't find iridium there was no impact?
Sepkoski: In 1982 Greeve published a compendium of the best estimates of crater ages at that time. An analysis by Walter Alvarez and Rich Muller suggested a periodicity in those crater ages that wasn't too different from the periodicity we see in extinction events. Since then a lot of the crater dates were cleaned up, and on reanalysis the periodicity didn't look as good. But several manuscripts now in press or review [and subsequently published] indicate that a periodicity in crater ages has been essentially refound. If it is assumed that maybe 50 to 65 percent of the craters are due to random cratering events, perhaps impacts of Apollo asteroids or something of that nature, the timing of the rest of the craters looks quite periodic statistically. However, the periodicity is about 30 million years, which isn't the same as 26 million years. Also, over at least the most recent part of the extinction time series, the crater dates are out of phase by about 9 million years.

Your second question would best be answered by an expert on meteorites, which I am not. But it is my understanding that virtually all meteorites, except eucrites, are enriched in iridium relative to earth crustal rocks, often by several orders of magnitude.

Question: Is the type of extinction due to humans the same as that of the older extinctions?
Sepkoski: I think that the advent of humans has probably caused two mass extinctions. There was certainly a major extinction on land—but not in the

oceans—about twelve thousand years ago. The Holarctic continents, South America, and Australia lost their large mammal fauna then. According to some pretty good arguments, now coupled with some pretty good evidence, that extinction event was related to the appearance of fairly efficient hunting bands at the end of the last ice age. Of course, the extinction was an aperiodic event, and so I would expect some extraordinary agent, such as human predation, to have been responsible. Like many earlier mass extinctions, the event twelve thousand years ago affected large terrestrial animals but not marine fauna. There are also good arguments that in historical times we have entered a second mass extinction that is much more extensive in terms of the kinds of organisms that are being affected. It is difficult as yet to get good information on what kinds of organisms are being affected at present, so comparisons with older events are tenuous.

The closing statements I made about the beneficial effects of extinction may need a litttle clarification. As a paleontologist, an evolutionary paleobiologist, I am looking at how the whole evolutionary system behaves over vast spans of time—tens of millions of years. That is very different from processes that happen over human time scales of days, weeks, and years. I fear that some of the animals and plants disappearing right now may be very useful for a variety of purposes. We shouldn't be too relaxed to see them disappear before we can characterize them better and know what the short-term ramifications of their extinction are. The rebounds from mass extinctions, which take place over 10 million years or so, may be good from the standpoint of a large-scale evolutionary system such as the entire biosphere of the earth. But, from the human standpoint, the first few decades or centuries after the initiation of an extinction event may in fact be quite catastrophic. Thus, my comments about the constructive aspects of extinction are meant to give solace.

Courtesy Department Library Services, American Museum of Natural History. Neg./Trans. No. V/C2627

Cotylorhynchus

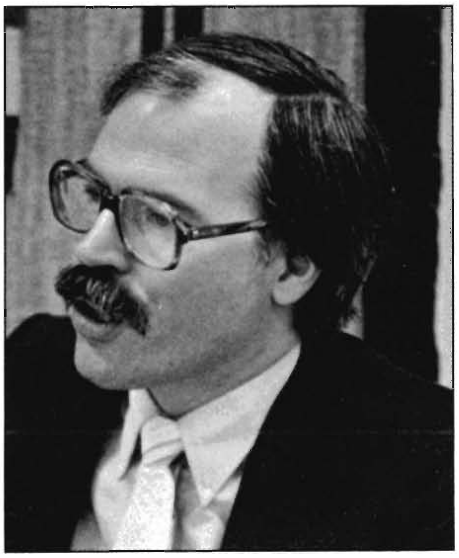
Question: Are there correlations between the changes in the earth's magnetic field and the extinction events?
Sepkoski: Work by David Raup and several others suggests that the reversals of the earth's magnetic field over the last 200 or 250 million years show a periodicity. But there is some question as to whether that periodicity is stationary or nonstationary. Also the purported periodicity isn't the same as that of the extinction events. It is about 30 million years, more in tune in both frequency and phase with the cratering periodicity than with the extinction periodicity.

Question: The effects of an impact that creates a crater a couple hundred kilometers in diameter are obviously horrendous—far worse than those of a nuclear war. So how can it be that an extinction is not associated with every large crater?
Sepkoski: An extinction of some magnitude could well be associated with every large crater, but that doesn't mean we would see such a correlation in the data. I could, for instance, sweep that whole question under the rug by simply saying that our ability to date craters is still rudimentary, not nearly approaching even our ability to date fossils. The problem may also lie in the loss of resolution we incur by dealing with higher taxonomic levels. Remember that even the impact at the end of the Cretaceous, which spread a 1-centimeter dust layer over the entire face of the earth, eliminated only about 17 percent of the animal families in the oceans, and on land it eliminated only about 10 percent of the vertebrate families. So at the family level the biosphere seems rather insensitive to perturbation. The combination of a small response and imperfections in the data for higher taxonomic levels could obliterate any observable response. Alternatively, absence of a marked response in association with an impact or the like could mean that the impact was completely out of phase with the periodic extinction force. We are trying to use statistical models to sort out these problems and to learn how to start attacking when we see associations, but we are really just beginning.

Question: Are there any explanations for the rebound phenomenon, and does the nature of the animals that survive an extinction provide information about the nature of the extinction force?
Sepkoski: That is a very good question. One thing that we know from looking at radiations, including rebounds, is that evolution can go on extraordinarily rapidly, at least on geologic time scales. If the rate of evolution across the Precambrian-Cambrian boundary had continued to the present, the oceans would now contain on the order of 10^{27} families, in contrast to the about 3×10^3 that in fact they do contain. (We would essentially have bouillabaisse from New York to London.) What we

see as normal rates of evolution through most of the fossil record seem to be very, very damped, which I suspect is just a crowding effect. The clearing of ecospace by the extinction of a lot of species may take the brakes off evolution, so that the initial, unconstrained evolutionary rates are again in effect, rapidly refilling the open ecospace. The evolutionary rates during the rebounds can be of about the same magnitude as that across the Precambrian-Cambrian boundary, when animals were first appearing in large numbers in the marine system.

At this time only a few systematic studies of victims and survivors of mass extinctions exist, and so little can be deduced from them about the nature of extinctive forces. At the end of the Cretaceous, small animals and animals in detritus-based food chains preferentially survived, which seems consistent with impact scenarios. On the other hand, warm-blooded, high-energy birds also survived, which seems problematic. Whatever the forces, David Jablonski recently completed a study for the Cretaceous that suggests the rules of the game change during mass extinctions: Victims of those events do not have the same sort of properties as species that are vulnerable to extinction during normal "background" times. Thus, mass extinctions represent more than simply intensification of extinction; they represent real changes in the nature of extinctive forces.



J. John Sepkoski, Jr., received a B.S. in geology from the University of Notre Dame in 1970 and a Ph.D. in geological sciences from Harvard University in 1977. After serving from 1974 to 1978 as an Instructor and then Assistant Professor in the Department of Geological Sciences at the University of Rochester, he moved to the Department of the Geophysical Sciences at the University of Chicago, where he is now a Professor of Paleontology. He is also a Research Associate at the Field Museum of Natural History. In 1983 he received the Charles Schuchert Award from the Paleontological Society. He has served as co-editor of *Paleobiology* and is a consulting editor for McGraw-Hill's *Encyclopedia of Science and Technology*. He is a member of the American Association for the Advancement of Science, the Paleontological Society, the Society of Sigma Xi, and the Society for the Study of Evolution.

Further Reading

Luis W. Alvarez, Walter Alvarez, Frank Asaro, and Helen V. Michel. 1980. Extraterrestrial cause for the Cretaceous-Tertiary extinction. *Science* 208: 1095–1108.

Walter Alvarez. 1986. Toward a theory of impact crises. *EOS* 67: 649, 653–655, 658.

Walter Alvarez, Erle G. Kauffman, Finn Surlyk, Luis W. Alvarez, Frank Asaro, and Helen V. Michel. 1984. Impact theory of mass extinctions and the invertebrate fossil record. *Science* 223: 1135–1141.

Marc Davis, Piet Hut, and Richard A. Muller. 1984. Extinction of species by periodic comet showers. *Nature* 308: 715–717.

Donald Goldsmith. 1985. *Nemesis: The Death-Star and Other Theories of Mass Extinction.* New York: Walker and Company.

Stephen Jay Gould. 1984. The cosmic dance of Siva. *Natural History* August 1984, 14–19.

David M. Raup. 1986. *The Nemesis Affair: A Story of the Death of Dinosaurs and the Ways of Science.* New York: W. W. Norton & Company.

David M. Raup. 1986. Biological extinction in Earth history. *Science* 231: 1528–1533.

David M. Raup. 1987. Mass extinction: A commentary. *Palaeontology* 30: 1–13.

David M. Raup and J. John Sepkoski, Jr. 1982. Mass extinctions in the marine fossil record. *Science* 215: 1501–1503.

David M. Raup and J. John Sepkoski, Jr. 1984. Periodicity of extinctions in the geologic past. *Proceedings of the National Academy of Sciences of the United States of America* 81: 801–805.

David M. Raup and J. John Sepkoski, Jr. 1986. Periodic extinction of families and genera. *Science* 231: 833–836.

J. John Sepkoski, Jr. 1982. Mass extinctions in the Phanerozoic oceans: A review. In *Geological Implications of Impacts of Large Asteroids and Comets on the Earth*, edited by Leon T. Silver and Peter H. Schultz, 283–289. Geological Society of America Special Paper 190. Boulder, Colorado: The Geological Society of America, Inc.

J. John Sepkoski, Jr. 1985. Some implications of mass extinction for the evolution of complex life. In *The Search for Extraterrestrial Life: Recent Developments (Proceedings of the 112th Symposium of the International Astronomical Union held at Boston University, Boston, Mass., U.S.A., June 18–21, 1984)*, edited by Michael D. Papagiannis, 223–232. Dordrecht, Holland: D. Reidel Publishing Company.

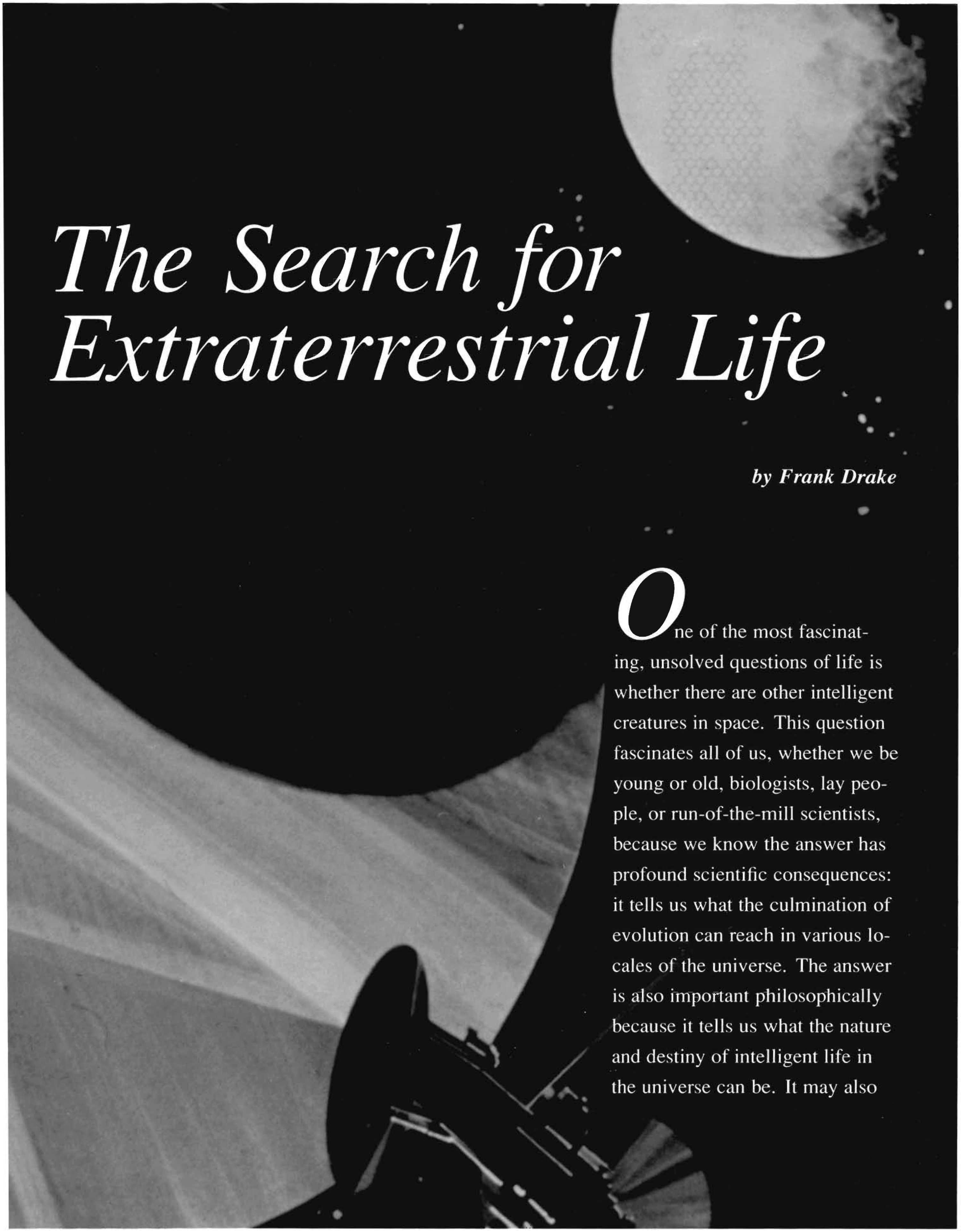
J. John Sepkoski, Jr. 1986. Global bioevents and the question of periodicity. In *Global Bio-Events: A Critical Approach (Proceedings of the First International Meeting of the IGCP Project 216: "Global Biological Events in Earth History")*, edited by Otto H. Walliser, 47–61. Berlin: Springer-Verlag.

J. John Sepkoski, Jr., and David M. Raup. 1986. Periodicity in marine extinction events. In *Dynamics of Extinction*, edited by David K. Elliott, 3–36. New York: John Wiley & Sons.

Leon T. Silver and Peter H. Schultz, editors. 1982. *Geological Implications of Impacts of Large Asteroids and Comets on the Earth.* Geological Society of America Special Paper 190. Boulder, Colorado: The Geological Society of America.

Steven M. Stanley. 1987. *Extinction.* New York: Scientific American Books, Inc., Scientific American Library.

Artist's conception of space-based radiotelescope shielded from radio transmissions on the earth (NASA Ames Research Center).



The Search for Extraterrestrial Life

by Frank Drake

One of the most fascinating, unsolved questions of life is whether there are other intelligent creatures in space. This question fascinates all of us, whether we be young or old, biologists, lay people, or run-of-the-mill scientists, because we know the answer has profound scientific consequences: it tells us what the culmination of evolution can reach in various locales of the universe. The answer is also important philosophically because it tells us what the nature and destiny of intelligent life in the universe can be. It may also answer personal questions about the significance of our role in the universe and what we can accomplish. Presumably, an answer might also show us how to obtain important scientific and technological information from other civilizations—knowledge we would not otherwise gain except by hundreds of years of hard, expensive effort. Thus the question of whether extraterrestrial intelligence exists is important for reasons that range from the most profound philosophical level to the very practical and technical levels.

Nowadays we generally break the question into two parts. First we ask about our expectations for intelligent life in space based on what we know of biology, the arrangement of the universe, and the laws of physics. This part guides us about the difficulty of the search, the possible closeness of the nearest civilizations, the expected number of civilizations in the galaxy, and the range of life that might be found. These answers in turn serve as guidance for the second, more practical question: What is the most promising way to search for that life?

The Number N of Detectable Civilizations

$$N = R f_s f_p n_e f_l f_i f_c L$$

where

R = average rate of star formation in galaxy (stars/year)
f_s = fraction of stars that are "good suns"
f_p = fraction of good stars with planetary systems
n_e = number of planets per star within ecoshell
f_l = fraction of ecoshell planets on which life develops
f_i = fraction of living species that develop intelligence
f_c = fraction of intelligent species achieving electromagnetic communication
L = lifetime in electromagnetic communicative phase (years)

How Many Other Worlds?

The number of detectable civilizations N that might exist in our galaxy can be represented by the nice, neat equation shown above. In this equation the "good suns" associated with f_s are stars whose production of light has the length and constancy needed for the evolution of intelligent life. Also, the fraction of intelligent species f_c that achieve electromagnetic communication are those that develop the technology needed to make them detectable in the galaxy.

The equation plays two roles. First, it tells us what we need to know: Looking at it, we see what the other unanswered questions about life are. Second, because the equation is a simple product—there are no exponential or logarithmic parameters, no gamma functions, no sines or cosines, or such—it tells us that all the things we need to know are equally important. The chain of factors is only as strong as its weakest link.

The product of all the factors except the last one gives the rate of production of potentially detectable civilizations in the galaxy. Using the best numbers we have, which frequently are only crude guesses, we arrive at a rate of about one per year. If we now multiply by the mean longevity of highly technical civilizations L, we find the total number of detectable civilizations in the galaxy: $N \approx L$.

Now all these factors are, in fact, questions. The rate of star formation is an astronomical question, and we happen to know the answer quite well—about twenty per year since the birth of the galaxy. Realistically, this number is the only one we know well. We know the fraction of good suns fairly well, but the fraction of stars that have planets is currently seen only through a glass darkly. Intimations of planet formation in the disks of dust around nearby stars, in the detection of a brown dwarf around a star, and so forth suggest that more than 10 per cent of the stars have planetary systems. As yet, though, we haven't confirmed that estimate to any degree of certainty. Models of planetary formation suggest there will be two planets in each system suitable for life, but there is no physical evidence other than our own system to support such an answer to this astrophysical question.

Now we come to the questions about life. What fraction of potential life-bearing planets give rise to life? The experiments that have been done now in a host of laboratories show a multitude of pathways giving rise to the chemicals of life. Moreover, if that is not good enough, these chemicals seem to be carried in from outer space by comets and asteroids. It is pretty clear that we know quite a bit about this question: the fraction must be very close to one.

The next three factors—the fraction of living systems that give rise to intelligence, the fraction of those that give rise to technology, and finally the longevity—are fascinating in themselves, and I will dwell on these factors a little more than on the previous ones. In fact, we usually quickly wave our hands over the last factor and move on, but it is one of the most important unanswered questions about life.

Is Intelligence Inevitable?

Until recently the fraction of systems that give rise to intelligence was quite controversial. Intelligence was believed to be the artifact of a limited series of freak events in the course of evolution and thus occurred only rarely in systems of living things. In other words, it was thought to be an anomaly in the universe—a view that some people still hold. However, over the years evidence has increased to support the idea that intelligence will arrive by one or another of many evolutionary paths. If one examines the fossil record, the only feature that one always sees increasing in power or size is the brain. We have had larger creatures, faster flying creatures, faster running creatures, more vicious creatures, but we have never had creatures on the surface of the earth that were smarter than the ones we have today. In fact, many studies of the cranial capacity of creatures have shown the increase in intelligence to be very monotonic over the course of time.

Figure 1 shows the distribution of brain mass versus body weight both for reptiles and mammals living now and for archaic reptiles and mammals that lived about seventy million years ago in the period before the Cretaceous-Tertiary boundary. Surprisingly, we find for each group not a scatter diagram but a well-defined relationship in which brain mass goes essentially as body weight to the two-thirds power, a rule that has applied through all evolutionary eras for which we have good data.

BRAIN-BODY WEIGHT RATIO

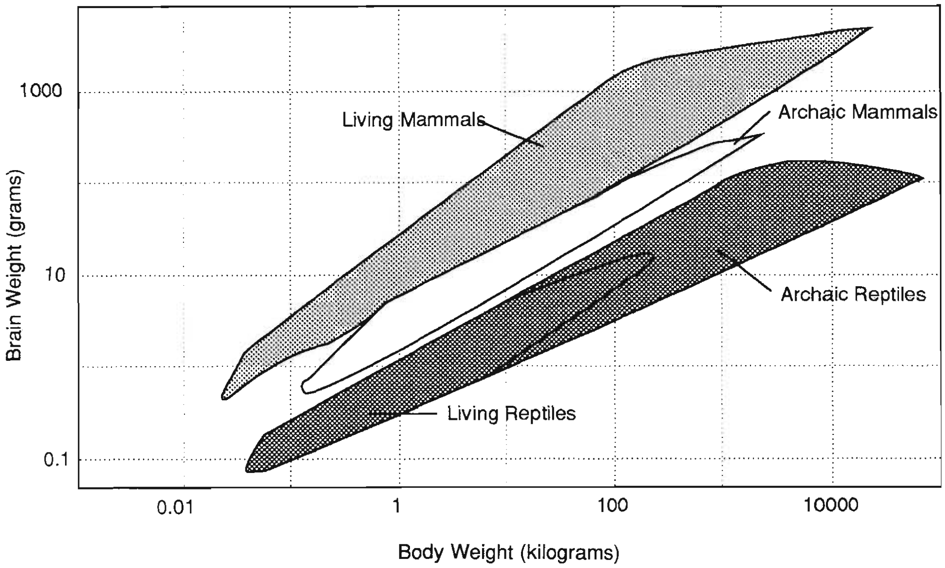


Fig. 1. A logarithmic plot of brain weight versus body weight for living mammals, living reptiles, archaic mammals, and archaic reptiles, where archaic means 70 million years ago during the age of the dinosaurs. The plot shows that each group follows the typical two-thirds power relationship. The average statistical increase in brain mass for a given body weight over those 70 million years has been on the order of a factor of 10, but the 20-fold increase for the mammals is significantly larger than the 4-fold increase for the reptiles.

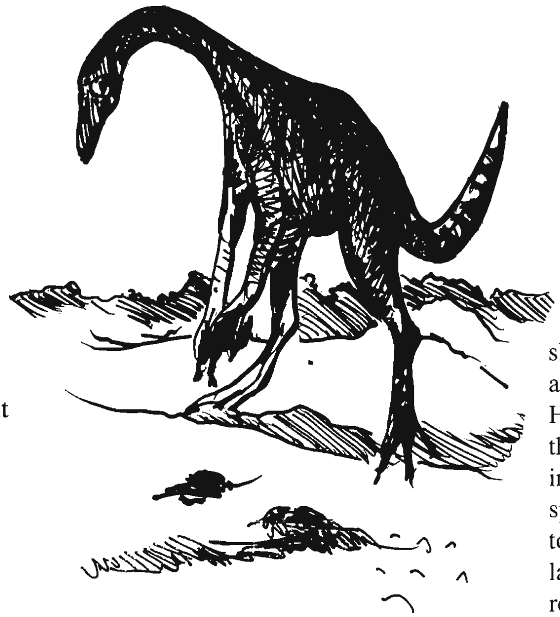
Each of the four groups shown in Fig. 1 occupies a clearly defined region. For example, the archaic reptiles (the dinosaurs) were, of course, much larger than current reptiles, and they occupy a region higher both in body weight and brain weight than the region occupied by the living reptiles. But when we compare mammals to reptiles *at a given body weight*, we find the archaic mammals have approximately four times the brain mass of the archaic reptiles, whereas the living mammals have twenty times the brain mass of the living reptiles. Although the change is more dramatic for the mammals than for the less evolutionarily advanced reptiles, there has been a definite increase for both. One can make plots using different or more finely divided time intervals, and one always finds a monotonic increase in brain size with time.

The host of statistical evidence showing the increase supports the idea that, in one way or another, an intelligent creature will appear on a life-supporting planet. In fact, if competition did not take its toll, there would most likely eventually be a *large* number of intelligent creatures on each planet.

An interesting bit of evidence for this last conjecture is the recent discovery of dinosaurs that had relatively large brain masses. Their brains were not as large as those of humans by any stretch of the imagination, but the mass was larger than that of the typical dinosaur brain. These creatures include one known as an ostrich dinosaur that lived in Mongolia and the northern United States, one with the delightful name of *Stenonychosaurus inegualis* that stood about six feet tall and weighed a little over a hundred pounds, and my favorite, a cute critter called *Saurornithoides*, which means reptile with feet like a bird.

The saurornithoides (Fig. 2) had about the same height and weight as we do, stood on its feet using its forearms as we use our arms, and had the equivalent of an opposable thumb. It used its intelligence and its "hands" to catch its favorite food, rat-like animals—the things that Jack Sepkoski calls vermin—that were the original mammals, our ancestors. Most interesting, the saurornithoides had a rather large brain case: Its brain mass was not the few grams typical of dinosaurs but was of the order of a hundred grams. Although smaller than that of a human infant, that mass is getting there, folks. If the saurornithoides had had another ten or twenty million years, it could well have been the first intelligent creature on earth. Your mommy and daddy, your spouse, your girlfriend or boyfriend would have looked like the creature in Fig. 2. We would have had to redesign the furniture, and we would all be looking forward to a dinner of vermin. But that was not to be. Before saurornithoides became intelligent enough to preserve itself from catastrophe, a catastrophe happened, and the earth was left to the vermin. Sixy-five million years later the vermin are us.

The saurornithoides are an example of creatures that could have led to a very weird species of intelligent life on the earth, although an intelligent species can probably be much weirder depending upon who wins the evolutionary race. Regardless, we now expect intelligence to be capable of arising over and over again on a planet. This expectation, of course, leads one to ask: If we wipe ourselves out what will be the next intelligent creature? I have a favorite candidate, although no one ever believes me when I say it's the squirrel. But anyone with a bird feeder knows they are very intelligent. So there it is, standing in the wings—on its two hind feet using its forepaws—ready to take over if things go wrong.



A SAURORNITHOIDES

Fig. 2. This relatively intelligent dinosaur used its "hands" and intelligence to catch vermin, our mammalian ancestors.

Space Colonies?

What about the development of technology? Such development happened independently three times on earth: in the Middle East, in China, and in Central America. In each case, population pressure triggered the development of organized agriculture, which, in turn, created the need for tools and then artisans to construct the tools. Of course, the artisans quickly learned they could make weapons, and then it was only an instant in cosmic time from stone axes to video tape recorders and motorcycles. It seems that technology should be very common in the universe.

What becomes of technical civilizations? Many people feel our civilization is very primitive in the sense that, as we have seen in the history of many other civilizations, there is a great panorama of technical development ahead of us. One scenario is the colonization of space, which could create living room for literally billions and billions of people. Today we have the technology to build great space colonies, although they would be very expensive. A typical colony might be twenty miles long and several miles in diameter, and it would slowly rotate in space with its long axis pointed toward the sun (Fig. 3). Huge mirrors, inclined to reflect sunlight through windows, would create daylight inside. Initially, one might think that a space colony would be a horrible place to live, but it could be delightful with landscapes, lakes, rivers, and a force resembling the earth's gravity due to the slow rotation (a few revolutions a minute) of the colony. Moreover, you could have a bug-free environment and made-to-order weather: open the mirrors a lot to have Hawaii week or close the mirrors most of the way to have an Aspen skiing week.

In this scenario civilizations eventually leave their dreadful home planets and colonize space, where literally tens or thousands of millions live in wondrous splendor. What do they do in these colonies? They make *other* colonies. This very heady concept creates a picture of almost everyone living in colonies. The only people left on the earth are park rangers because the planet has been made a national park. People take their kids down on the Gray Line Space Shuttle to show them how terrible it was for their grandmommy and granddaddy who had to endure tornadoes and storms and mosquitoes and other horrible things.

Whether or not the colonization of space is the course of the future for civilizations is important for our equation because it affects our estimates of both the longevity of civilizations and how brilliantly they might shine to the universe. Notice that my discussion has now become unscientific because, for the first time, I am talking about something we don't know has actually happened in the universe. Everything up to this point we know happened at least once.

Carrying this idea further, one can't avoid thinking that colonization must occur elsewhere and intelligent civilizations are destined to colonize the stars

SPACE COLONIES

Fig. 3. The space colonies depicted here are about twenty miles in length and simulate gravity by slowly rotating around their long axes. Huge petal-like mirrors reflect sunlight onto idyllic landscapes inside, and food is grown in the pods in a ring at the end of each colony.

of our galaxy. Such a thought may have occurred first to none other than Enrico Fermi, who, one day at Los Alamos about forty years ago, started going around asking: "Where are they?" He had made an analysis similar to mine, estimating how many civilizations might be out there and how fast they were created. Given the continuous march of technology he felt that, even at slow interstellar flight speeds, a civilization could populate the entire galaxy in only a hundred million years—a small fraction of the age of the galaxy. This estimate suggests that the first intelligent civilization to embark on a such an enterprise should have already taken over the whole galaxy and thus arrived at the earth. Where are they? Some people, including Eric Jones at Los Alamos, have taken this idea very seriously and worked very hard exploring the possibilities and obstacles to interstellar colonization.

Now is it possible that something is wrong and we are, in fact, alone in the galaxy? Everything we know seems to indicate that colonists should be out there, but, if so, why have they not yet come to the earth? Well, there are a variety of solutions to this Fermi paradox. Perhaps there are cosmic hazards to space travel of which we are unaware. Or perhaps the colonization follows a diffusion equation. If instead of moving outward radially, colonization is best described by a random walk, then it takes about the age of the galaxy to colonize the entire galaxy. In other words, the colonists should not be here yet—but tomorrow they will come out of the sky!

Galactic Colonization?

The argument I favor is simply that it makes better sense to colonize in your own system than to endure the costs and hazards of going to other stars. There may indeed be enormous numbers of civilizations of great technical prowess that don't bother to come to earth in person. In other words, I assume that an intelligent civilization will colonize space *only* if it gets a good bang for its buck, that is, only if the quality of life for its expenditure is equivalent to the lifestyle it would get for the same amount of resources, energy, or money in its own system. But, you say, we do not know what the cost of interstellar colonization is or what the propulsion systems are and what they cost, and so forth. However, the way to get a minimal cost is to use the minimum kinetic energy required for interstellar colonization. Such an approach yields a cost figure in the absence of any knowledge about the actual physics of the propulsion systems. In other words, we assume the energy used per colonist is no greater than the energy required to give a good life to that colonist back in the home system and *all* of the energy is used in the kinetic energy of the spacecraft. Such assumptions surely yield a very conservative lower limit.

If E is the energy ratio per colonist, we equate that to kinetic energy to get the velocity of the spacecraft ($v = \sqrt{2E/m}$). What is a reasonable energy ratio for a human being? The most energy-rich country in the world is the United States, which, during a recent year, consumed a total of about 10^{20}

joules of energy. The population is on the order of 250 million, which yields about 4×10^{11} joules per person per year. If a lifetime is about a hundred years, lifetime energy consumption per person is around 4×10^{13} joules. This energy ratio takes you to Europe, buys you ice cream cones and corn on the cob, makes your car go, and all that.

What spacecraft velocity does this limit yield? If we allow a spacecraft mass of ten tons per colonist—about the per-passenger mass of a typical airliner—the velocity of the spacecraft is ninety kilometers per second. That velocity is pretty fast—about ten times the velocity of our Voyager or Viking spacecraft. The time to go ten light years, which is probably the minimum distance to a suitable star, is forty thousand years. Now that picture is a little discouraging—you are sitting in a DC-9 for forty thousand years eating airline food and watching the same movies over and over again.

If we abandon this approach and, instead, assume a velocity high enough to get to the star in a hundred years, that is, a couple of generations, the energy required *per colonist* is $2 \times 10^5 E$. In other words, a trip that takes place in a reasonable amount of time uses the same energy as does a good life in the home system for 200 thousand people. What makes this approach even more unrealistic is that, besides not allowing for inefficiencies in the production of fuel and in the propulsion system, we have arrived at the distant star at a very high speed and have no energy remaining with which to stop. We just go whistling through the system and out the other side! If we take proper rocket-mass ratios and so forth, the actual energy *per colony* for a hundred colonists is about 200 million E, that is, the same energy needed to support the entire population of the United States for their lifetime. To launch a mission we would have to shut down America for a hundred years.

Much less energy is required to build space colonies in one's own system. The energy required to accelerate ten tons to a velocity sufficient to go ten light years in a hundred years—the example I just discussed—is about 10^{18} joules. The energy required to put ten tons in lunar orbit, that is an orbit in the solar system with an orbital velocity that will keep it from falling, is about 6×10^{11} joules. The ratio of energy required for interstellar as compared to solar system colonization is thus about 10^7. Although one might argue that very advanced civilizations will have access to much greater energy resources, a factor of 10^7 is very hard to overcome.

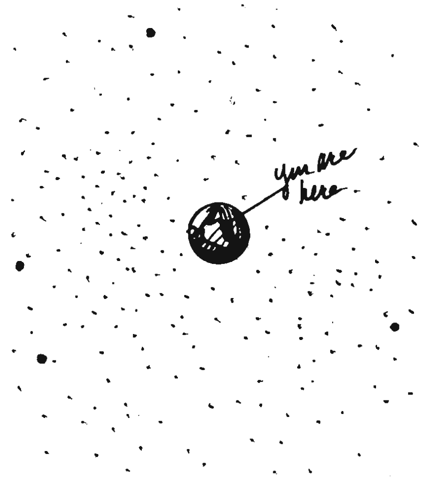
Lord Kelvin said that you don't know anything in science until you put numbers to it. I feel the idea of colonization is one for which we have to pay attention to the numbers. A dumb civilization might go to the stars, but a smart one will, I think, stay right at home. There is enough energy from the sun to support, believe it or not, somewhere between 10^{20} and 10^{22} human beings, depending upon the quality of life. That number is about 10^{12} times as many people as there are now—which seems like enough. A civilization can do all that at home for about one ten-millionth the cost of doing it at the stars. Where are they? They are probably living in great splendor in their own home systems, and they are there in great numbers for us to find.

(By the way, I would recommend the analysis of the energy required for interstellar colonizations as a project for the Economics and International Studies Section here at Los Alamos. I ought to also mention that that section includes the esteemed economist Robert Drake, who happens to be my brother.)

Notice that my argument raises one of the important questions of life for which we do not know the answer: What is the limit of energy that intelligent civilizations manipulate? Do civilizations go to the stars, colonize their own systems, or stay on their own planets? To answer that we need to know how much energy they generate, manipulate, and store. Of course, there may not be a single answer to the question; it may depend on motivations, anatomies, or who knows. However, the answer is crucial to understanding what may become of intelligent civilizations, what we might find in space, and what we should be looking for. The question thus affects *how* we search for civilizations.

Where Do We Search?

Suppose we start our search outwards in space. What is the best strategy? Where should we look to find civilizations fastest? One's intuition says we should look at the nearest stars similar to the sun because the signals will be strongest from them. Surprisingly, that is wrong. If you look at normal stars in



FAINT AND BRIGHT STARS

Fig. 4. Because intrinsically bright stars are relatively rare, their distance from the earth is typically much larger than that of the dimmer stars.

the night sky, the brightest and easiest to detect are, with few exceptions, not the closest. Bright stars are in fact very distant; moreover, most of the twenty nearest stars are very faint and invisible to the naked eye (Fig. 4 and Table 1). This strange situation can be understood by looking at the distribution of intrinsic brightness: faint stars are plentiful and bright stars are rare (see "Are They Near or Far?"). As a result, the typical distance of a bright star from the earth is much greater than the typical distance of a faint star. But if the intrinsically bright stars are bright enough, as they indeed are, they will still outshine the brightest of the nearby stars.

The same thing is true of cosmic radio sources. The brightest, or apparently brightest, cosmic radio sources are not the nearest ones in our own galaxy but the most distant, the quasars. So again we have a situation where the easiest things to detect are not the closest but the farthest.

Thus it may not make sense to look at the nearest stars after all. Perhaps, instead, we should look at the *most* stars we can to locate the intrinsically brightest civilizations. Whether this is right depends again on the maximum amount of energy a civilization can manipulate. Is there an enormous range of energies? If so, the right strategy is to look for the rare but intrinsically bright civilization. On the other hand, if civilizations all operate at about the same energy level, the right strategy is to look at the nearest stars. Unfortunately, we can't know which answer is right until we have found other civilizations.

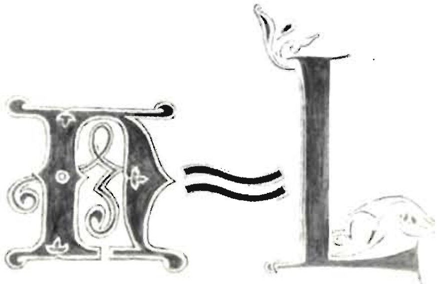
Table 1

Of these stars, the three that are *both* bright and near are shown in bold.

The Twenty Brightest Stars	The Twenty Nearest Stars
Achernar	Groombridge 34
Aldebaran	L726-8
Capella	τ Ceti
Rigel	ϵ Eridani
Betelgeuse	**Sirius**
Canopus	Luyten +5° 1668
Sirius	**Procyon**
Procyon	Wolf 359
Pollux	Lalande 21185
Regulus	Ross 128
Acrux	**α Centauri**
Spica	Barnard's star
Agena	Cincinnati 2456
Arcturus	Ross 154
α Centauri	61 Cygni
Antares	Lac 8760
Vega	ϵ Indi
Altair	L789-6
Deneb	Lacaille 9352
Formalhaut	Ross 248

What Will We Find?

Now as I said earlier, this type of reasoning coupled with our best estimates for the various factors leads to the prediction that the number of detectable civilizations in space is roughly of the order of L, the average longevity in years of civilizations in a communicative phase:



$$N \approx L$$

I've used an elaborate font for the equation here because, in science, when you don't know something very well, it's more impressive and compelling to write it with fancy letters.

It is, of course, arrogant to try to say anything about the value of L. We have been a detectable civilization for only about forty years now—since the advent of television! Studies have shown that the strongest and most detectable signs of the earth, by far, are our television broadcasts, and we have become brighter every year as the power of these broadcasts has increased. About three hundred stars are now receiving Kukla, Fran, and Ollie and Uncle Miltie, and each year about ten more stars join in. Despite our terrible ignorance about the value of L, there is a relevant point, as we will see, that comes out of any discussion of this factor.

We make an estimate for L by imagining possible ways that civilizations might terminate (Table 2). This process is a guess, and everyone is invited to make their own estimates or invent other ways that civilizations might reach the end of the line.

The first category is total destruction, that is, MAD is actually invoked on the planet. The detectable lifetime for a civilization that ends with such an event might typically be fifty years. Let's guess that the probability of the event happening might be 10 per cent—that is, one in ten systems destroy themselves through war.

Another event is the cosmic accident such as a large asteroid crashing to the surface of the earth. The time between such collisions is very long—millions of years—so the probability for this event is very small.

Another possibility is the degeneration of culture; that is, the quality of life simply goes down as the world drifts into a subsistence culture. The television stations, of course, get turned off. Such civilizations may be detectable for 10^4 years, and let's guess that there's a 10 per cent probability of this occurring.

Although that last category requires very wild guesses, the next is more reasonable: becoming invisible due to superior technology. If we detect civilizations by their television programs and they all go to fiber optics or cable tele-

Table 2

An estimate of L, the lifetime of a technologically advanced civilization. In this case the total weighted value for L of 12,000 years is almost entirely the result of the "no limitation" civilization, a rare society ($P_{\text{exist}} = 10^{-3}$) that achieves a long lifetime (10^7 years), say by dodging the perils of nuclear war and building large numbers of space colonies in their own planetary system.

Limitation On L	L_c (yr)	P_{exist}	$L_c \cdot P_{\text{exist}}$ (yr)
Total Destruction	50	0.1	5
Cosmic Accident	10^6	10^{-6}	1
Degeneration of Culture	10^4	0.1	1000
Invisibility Due to Superior Technology	10^3	0.7	700
Abandonment of Technology	10^3	0.1	100
No Limitation	10^7	10^{-3}	10,000
		1.0	L = 12,000 yr

vision, they may vanish from the scene. In fact, for people like me, cable television is very bad news. The second most impressive sign of our existence, by the way, are the military radars of the Soviet Union and the United States, so cable television and peace on earth are both bad news. I'll be optimistic about this category, giving it a seventy per cent chance of happening and a lifetime of a thousand years before the civilization actually becomes invisible.

Abandonment of technology is, in a way, related to degeneration of culture, and I've given this possibility similar numbers: a thousand-year lifetime and a 10 per cent probability of happening.

Finally we have the "no limitation" category: civilizations that build space colonies and transmit to those colonies, to their interstellar spacecraft, and what not. They may exist for 10 million years, but we will guess that only rarely—one in a thousand times—does a civilization accomplish this.

Now, to get the total L we multiply the lifetime in years times the probability for each category, then sum—in other words, we calculate the weighted mean value of L. Using my guesses the answer is 12,000 years. The main point I want to make, however, is that the final value comes almost entirely from the contribution of the no-limitation category—a category to which we assign a probability of 10^{-3}! In other words, the result is strongly influenced by something we know very little about, something for which we may know the exponent by only a factor of two or three. The long-lived civilizations control L, and all the other categories amount to only a drop in the bucket.

What happens if you carry this to an extreme and assume that in some cases civilizations achieve immortality? By the way, immortality for a living species is not out of the question. Jack Sepkoski says that extinction is good because it makes it possible for evolution to speed up. Likewise, death within a species is good because it allows more members of the species to pass through that ecological niche, raising the chances for favorable mutations. At some time on the earth there could well have been an immortal species, but if there were also species that were *not* immortal, the ones that died evolved and got better and better until they ate all the immortal ones. That's why we don't see any immortal species today.

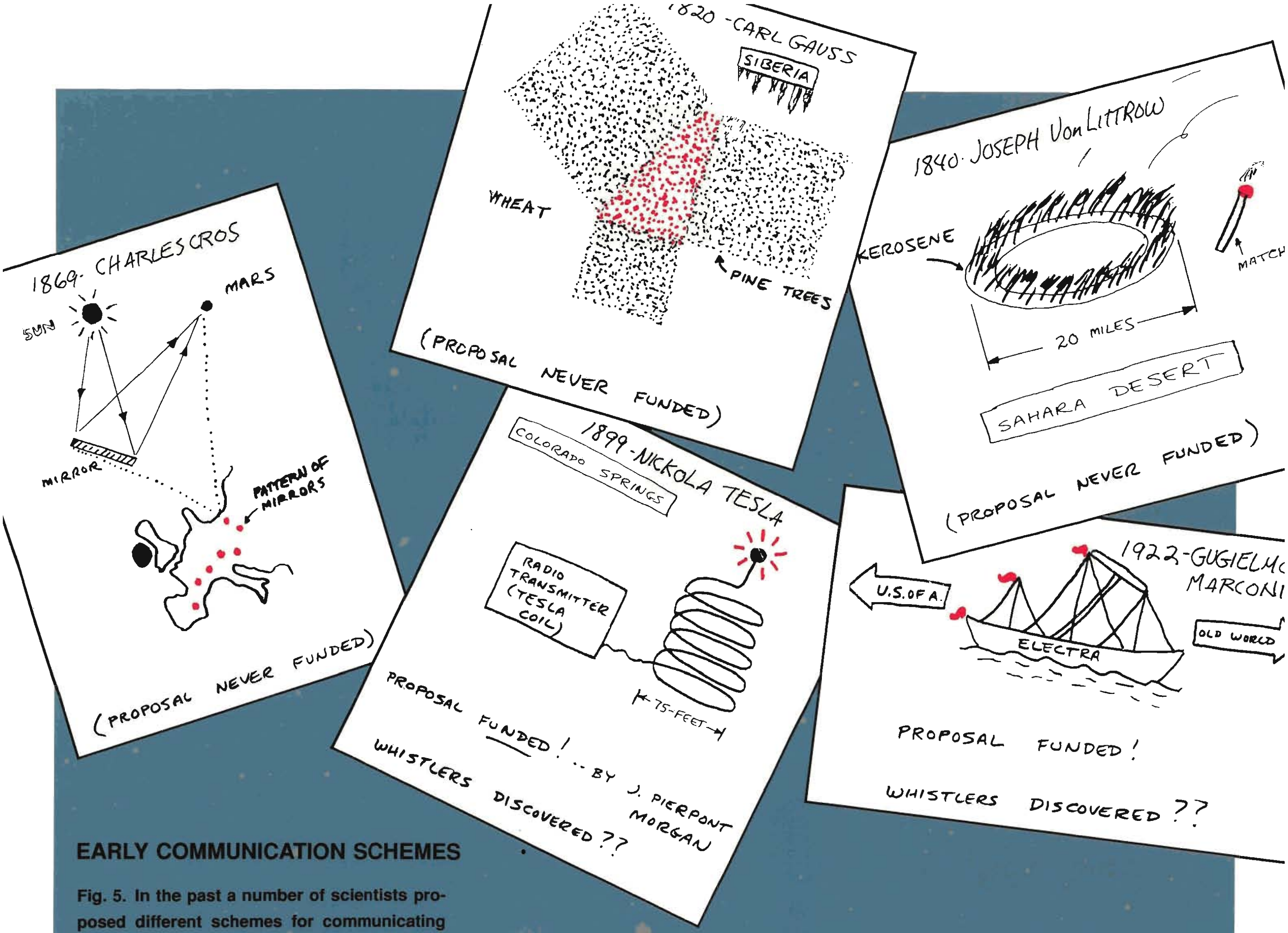
Although evolution selects for death—and death is a good thing to have until the species is intelligent enough to look after such things—there is nothing in physics or biology that requires death; it's an artifact of evolution. It could well be that there are civilizations clever enough to undo that part of evolution and become immortal.

Say we make the same analysis as before except represent the immortal civilization as one with a lifetime on the order of the age of the galaxy or the age of the oldest of the suitable stars (Table 3). Given a probability of 1 per cent—one in a hundred civilizations that actually achieve immortality—we get a value for L of 10 million—essentially all from the immortal civilizations. If we divide the weighted lifetime for each category by the total weighted lifetime L, we get the probability of discovering a civilization in any particular category. For example, the probability that you are going to find a civilization that is about to self-destruct is 5×10^{-7}. The

Table 3

A second estimate of L in which the "no limitation" society has been replaced with the "immortal" society, that is, one with a lifetime equal to that of the galaxy ($L_c = 10^9$ years). We quess that 1 per cent of the technological civilizations achieve this state of biological immortality. The last column, which gives the probability of detecting each particular type of society P_{detect}, shows that immortal societies, if they exist, are the ones that will almost certainly be found.

Limitation On L	L_c (yr)	P_{exist}	$L_c \cdot P_{\text{exist}}$ (yr)	P_{detect}
Total Destruction	50	0.1	5	$5 \cdot 10^{-7}$
Cosmic Accident	10^6	10^{-6}	1	10^{-7}
Degeneration	10^4	0.1	1000	10^{-4}
Invisibility	10^3	0.7	700	$7 \cdot 10^{-5}$
Abandonment of Technology	10^3	0.1	100	10^{-5}
Immortality	10^9	0.01	10^7	**0.9998**
			L = 10,002,000 yr	



EARLY COMMUNICATION SCHEMES

Fig. 5. In the past a number of scientists proposed different schemes for communicating with extraterrestrial civilizations.

probability of finding a civilization that will go invisible is 7×10^{-5}. Once again, my main point is that even if only a small percentage of civilizations are very long-lived, they're the ones we're going to find. We'll find the very old ones, the very technically competent ones.

To some people, such as George Wald, this last idea is worrisome. He thinks such a discovery may be a great blow to us because the superiority of the other civilization will be destructive to our self-image. Nevertheless, the idea gives guidance to our search. We should expect to find the civilizations that are very different from us and that are practicing technology very different from what we are used to.

Now where does this analysis lead us? In general, we do not adopt values of 10 million for L; we adopt the more conservative figure of about ten thousand years. If that figure is accurate, there are on the order of ten thousand civilizations in the galaxy, about one in ten million stars has a civilization we can detect, and the nearest civilizations are about a thousand light years distant.

Rockets or Radio?

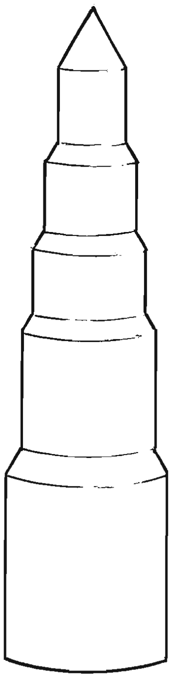
So now we ask the next great unsolved question of life: *What* is the most promising way to search a thousand light years and at least ten million stars? First, let's review some old ideas about communicating with extraterrestrial civilizations. The great mathematician Carl Friedrich Gauss proposed cutting a pattern into the forest of Siberia: a central region planted in wheat in the form of a right triangle and a square of pine trees adjacent to each side (Fig. 5a). He thought the pattern would be visible to powerful telescopes at least across the solar system and maybe far out in the universe. If so, this would prove not only that there was intelligent life on earth but that we understood the Pythagorean theorem! The proposal was never funded.

Another great scientist, in this case the physicist Joseph von Littrow, had a similar idea (Fig. 5b). He proposed digging big trenches in the Sahara Desert, perhaps in the form of circles and triangles twenty miles across. He would then apply a very sophisticated technology by filling the trenches with kerosene and lighting them with a match, thus making flaming geometric figures visible across the solar system. The proposal was never funded.

The French physicist Charles Cros suggested using mirrors to reflect sunlight to Mars (Fig. 5c). The mirrors would be placed across some sophisticated part of the world, such as Europe, in a pattern that the extraterrestrials would recognize as, say, the big dipper, again revealing intelligent life on earth. Again the proposal was never funded.

Not too long ago and not too far from

SPEC SHEET FOR THE ULTIMATE ROCKET



Fuel: Matter-Antimatter
Cruising Speed: 0.7*c*
Flight Plan: Round Trip between Earth and Another Star

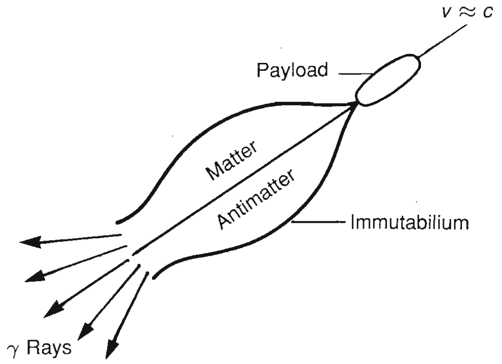
	Tons	Energy (Years of U.S. Electrical Consumption)
Payload	1000	**
Fourth Stage	1400	21,000
Third Stage	3400	51,000
Second Stage	8200	123,000
First Stage	20,000	305,000
Total	34,000	500,000

Fig. 7. Numbers such as those presented here imply that it may be considerably more beneficial to stay within one's own planetary system than to colonize the galaxy.

Los Alamos, Nickola Tesla finally got onto the right track by sending radio messages (Fig. 5d). He built one of the largest Tesla coils in the history of the world in Colorado Springs. It was 75 feet in diameter and about 150 feet high. Funded by J. Pierpont Morgan, it succeeded in standing people's hair on end for miles around when it was turned on. He actually received signals—strange, regular chirps that sounded very intelligent—and he believed he had detected another civilization. Knowing the frequencies at which the device received, we now think he discovered a phenomenon called whistlers: radio waves that propagate very slowly in the magnetosphere of the earth.

Figure 5e is another project that was funded! Guglielmo Marconi, the inventor of radio, also listened for signals from outer space and heard the same chirping sounds that Tesla had heard. He too thought he had discovered signals from other worlds, but again he was probably reporting whistlers.

What about rockets? Many of us who have gone to the movies think that rockets are the way to communicate with other worlds. But, as we've already seen, the speeds developed by chemical rockets will mean literally millions of years to go a thousand light years and return. So we have to use the sort of thing invented at Los Alamos.



THE ULTIMATE ROCKET

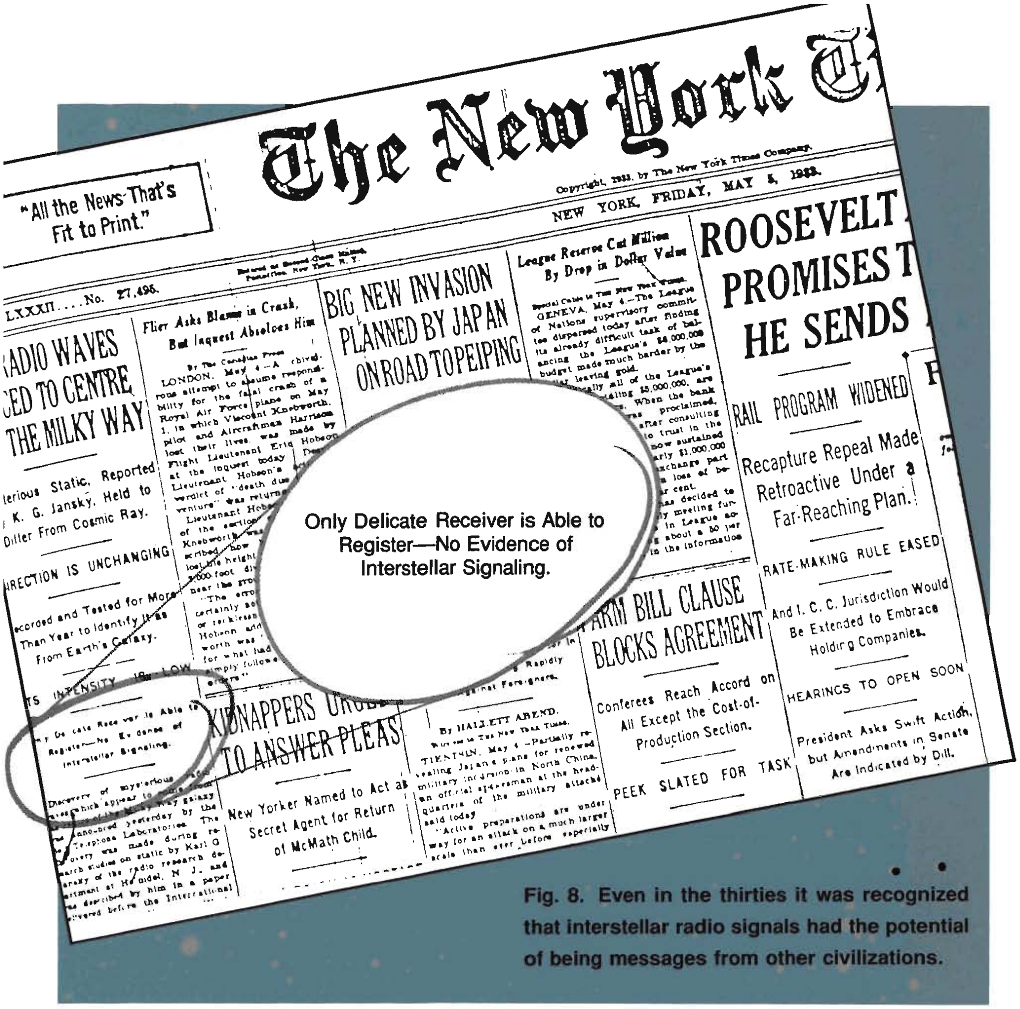
Fig. 6. Although the matter-antimatter rocket is simple in concept, a tricky technical problem remains to be solved: a material—here called immutabilium—needs to be invented that will hold both fuels until they are needed!

The ultimate rocket (Fig. 6) is very simple: it has two tanks that contain matter and antimatter. Of course, there's a small technical problem—what do you build the rocket with so that the whole device doesn't just go bang! The material is called *immutabilium.* We know its name, but its invention is left as a technical problem to you. As you know, when matter meets antimatter there is complete annihilation and a great big blast of gamma rays.

Now such rockets can go at nearly the speed of light, thus simulating what Captain Kirk and Mister Spock do every night on television when they zip from one planet to another within the hour. However, even if we solved the technical problems of building a matter-antimatter rocket, it wouldn't be very practical.

Figure 7 illustrates a rocket designed for a payload of a thousand tons and a cruising speed of seven-tenths the speed of light. It takes off, cruises close to the speed of light for as far as you want to go, then lands. Somebody gets out, looks around, takes some pictures, gets back on, returns to earth, again at seven-tenths the speed of light, and lands. Why seven-tenths? According to special relativity that's the speed at which the crew ages at the same rate it's traveling—the crew will be ten years older if they travel ten light years.

The rocket weighs thirty-four thousand tons. Of that, thirty-three thousand tons is fuel, half of which is matter. We get the matter by connecting our garden hose to the tank and filling it up. But the other sixteen thousand five hundred tons is antimatter, which has to be made. We don't know how to make that amount of antimatter or how to store it or how to make the immutabilium. But even if we did accomplish all that, it would take at least as much energy as there is in sixteen thousand five hundred tons of matter. That mass times the velocity of light squared equals five hundred thousand years of the total electric power production in the United States—



"All the News That's Fit to Print."

The New York T

LXXXII....No. 27,496.

Copyright, 1933, by The New York Times Company.

NEW YORK, FRIDAY, MAY 5, 1933.

ADIO WAVES
CED TO CENTRE
THE MILKY WAY

lerious Static, Reported
K. G. Jansky, Held to
Differ From Cosmic Ray.

IRECTION IS UNCHANGING

ecorded and Tested for More
Than Year to Identify It as
From Earth's Galaxy.

TS INTENSITY IS LOW

Only Delicate Receiver is Able to Register—No Evidence of Interstellar Signaling.

Flier Asks Blame in Crash, But Inquest Absolves Him

BIG NEW INVASION
PLANNED BY JAPAN
ON ROAD TO PEIPING

League Reserve Cut Million
By Drop in Dollar Value

ROOSEVELT
PROMISES T
HE SENDS

RAIL PROGRAM WIDENED

Recapture Repeal Made
Retroactive Under a
Far-Reaching Plan.

RATE-MAKING RULE EASED

And I. C. C. Jurisdiction Would
Be Extended to Embrace
Holding Companies.

HEARINGS TO OPEN SOON

President Asks Swift Action,
but Amendments in Senate
Are Indicated by Dill.

ARM BILL CLAUSE
BLOCKS AGREEMENT

Conferees Reach Accord on
All Except the Cost-of-
Production Section.

PEEK SLATED FOR TASK

KIDNAPPERS URGE
TO ANSWER PLEAS

New Yorker Named to Act as
Secret Agent for Return
of McMath Child.

By HALLETT ABEND.

Fig. 8. Even in the thirties it was recognized that interstellar radio signals had the potential of being messages from other civilizations.

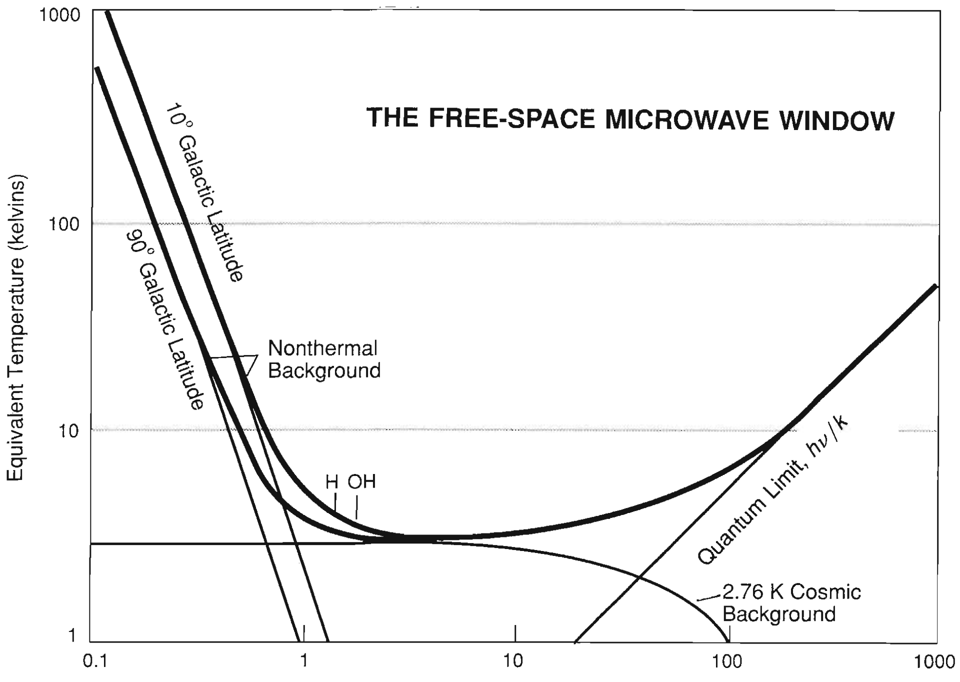


Fig. 9. The optimum electromagnetic frequencies for interstellar search and communication lie between the galactic nonthermal background at low frequencies and the quantum limit $h\nu/k$ that rises at higher frequencies. The bottom part of the minimum is the 2.76-kelvin cosmic background, a residue of the big bang. FM, television, and radar frequencies lie in this window also. Since a great deal of the nonthermal background radiation is a result of relativistic electrons orbiting in magnetic fields throughout our galaxy, the strength of this background varies with the galactic latitude of the observation.

for one mission, and we need ten million missions to explore the galaxy! If that isn't enough, the rocket has a bad side effect when it takes off: it incinerates one hemisphere of the earth. At least there'd no longer be any problem finding a site for nuclear waste disposal, but the Sierra Club would object.

Anyway, the analysis shows that Captain Kirk and Mister Spock have lied to us. You can't call up Scotty and order warp seven to go anywhere in the galaxy in two minutes. Whether or not one transfers things through space depends, of course, on how much energy a civilization can manipulate and whether one goes slowly rather than fast. However, I think the answer here is that you don't transport things through space, you transport information.

Is there a cheap way to transport information at the speed of light? The answer is yes. One uses electromagnetic radiation as guessed by Tesla and Marconi as long ago as 1933. When the *New York Times* announced the discovery of cosmic radio emission by Jansky at Bell Telephone Laboratories (Fig. 8), the lowest headline said: "No evidence of interstellar signaling." Even then they wondered if radio was the means by which one world would find another.

The Cosmic Haystack

Since that time a great deal of thought has been given to the subject, and we have repeatedly arrived at the conclusion that radio waves are best. This idea is correct not because of the status of our technology or the particular prowess that we have at certain wavelengths compared to others but rather because of fundamental limitations due to the arrangement of the universe and the laws of physics. For instance, the second law of thermodynamics sets limits on noise levels that can't be overcome by any technology. Certain minima in the noise levels, however, lead to optimum

RADIO TELESCOPES

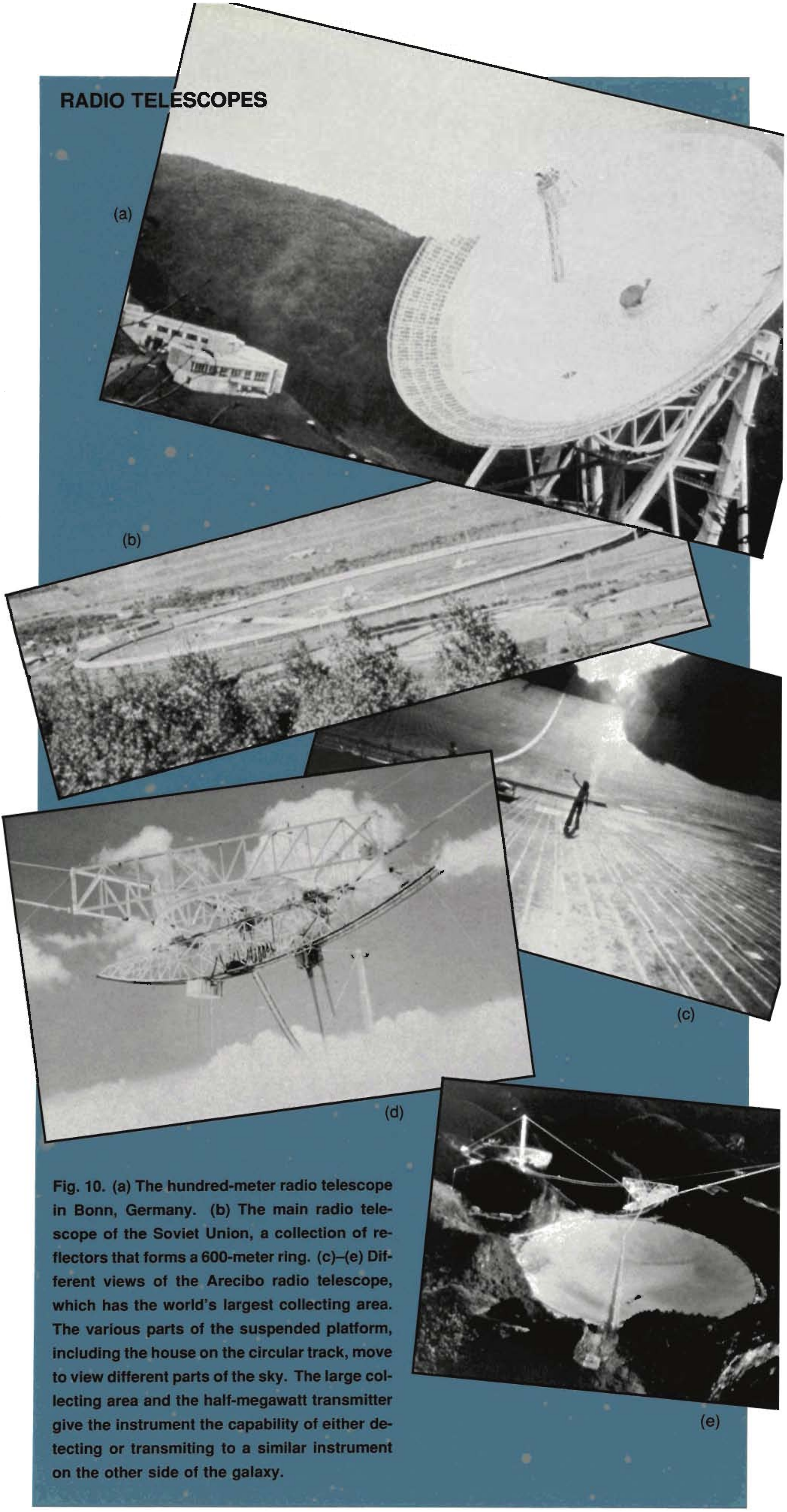


Fig. 10. (a) The hundred-meter radio telescope in Bonn, Germany. (b) The main radio telescope of the Soviet Union, a collection of reflectors that forms a 600-meter ring. (c)–(e) Different views of the Arecibo radio telescope, which has the world's largest collecting area. The various parts of the suspended platform, including the house on the circular track, move to view different parts of the sky. The large collecting area and the half-megawatt transmitter give the instrument the capability of either detecting or transmiting to a similar instrument on the other side of the galaxy.

frequencies for interstellar search and communication.

One of the limits is the quantum nature of light. Light comes in packets, and $h\nu/k$ gives the equivalent temperature of the noise associated with this quantum aspect of light (Fig. 9). A second source of noise is radio emission of relativistic electrons orbiting in the magnetic fields of our galaxy. These electrons produce radiation with a steeply rising spectrum that essentially jams radio telescopes. The third source of radiation (it's interesting that it plays a role) is that left over from the big bang. It's normally called the cosmic background and has an equivalent temperature of about 3 kelvins.

Every advanced civilization can sum these curves precisely as we have and arrive at the very pronounced minimum in the microwave region where wavelengths are on the order of a few centimeters. This minimum is true for us, true for every civilization. In fact, I suspect the curve in Fig. 9 has been shown many times in our galaxy looking exactly like it does there, except the letters are written in funny ways that we wouldn't understand. The dish antennas that people have in their back yards operate at that same point precisely because it has the minimum noise and the greatest sensitivity, making the receivers cheaper. The upshot is that any intelligent civilization will be using this region copiously for its own communications and perhaps to communicate with other civilizations.

Likewise, *we* can search with the most sensitivity in this region, and our searches have been concentrating there. We happen, of course, to have good instruments for this purpose. The Bonn telescope in Germany (Fig. 10a) is a hundred meters across and can detect signals from distances of hundreds of light years. The main radio telescope of the Soviet Union in the northern Caucasus (Fig. 10b) is a 600-meter-diameter

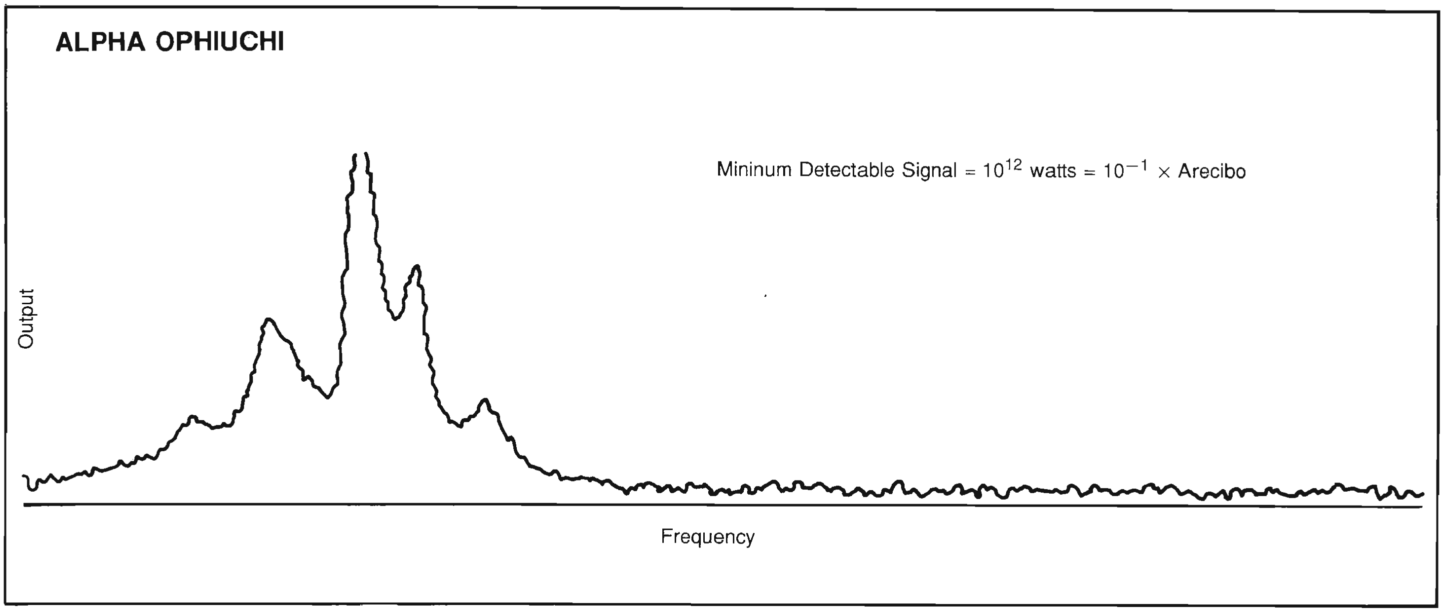


Fig. 11. For this search the Arecibo telescope in the receiving mode was pointed at α Ophiuchi, a star 54 light years from the earth. A 3000-channel spectrometer was used with output integrated over a time of twenty minutes (the jogs in the spectrum are the boundaries between adjacent channels). The large signals are emissions of neutral hydrogen in the star's atmosphere at the 21-centimeter wavelength. No other signals are present even though a radio telescope emitting at a tenth of the power of an Arecibo would have been detectable.

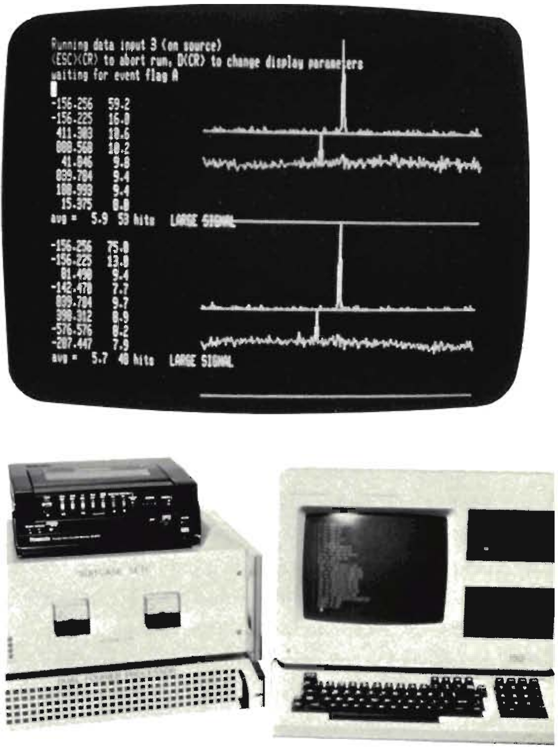
ring of reflectors, each 10 meters high. All the reflectors are moved and tilted under computer control to focus radiation on a central point.

The telescope with the largest collecting area in the world is the Arecibo radio telescope (Fig 10c–e), which is 1000 feet across with 20 acres of collecting area—more combined collecting area than all the other telescopes in the world put together. The radiation is focused on a suspended platform fifty stories in the air. The three towers holding the platform are each 10 feet taller than the Washington Monument. And if that doesn't give you a feel for its size, the bowl would hold 357 million boxes of corn flakes. The 38,778 surface panels, each about 2 square meters in size, are all put in place to an accuracy of about 1 millimeter.

The telescope has a 0.5-megawatt radar transmitter, and when that signal is focused by the big dish, the power density in the beam is equivalent to what could be radiated, without the dish, by a twenty-million-million-watt transmitter. Such power is twenty times the total electric power production of the earth. Thus, when Arecibo is tranmitting, it produces the strongest directed signal leaving the earth. In fact, the signal is about a million times brighter than the radio emission of the sun. Civilizations can outshine their stars. The signal is detectable by a similar instrument, not just from a distance of a thousand light years, but from anywhere in the galaxy. So we can reach out the required thousand light years and touch someone.

Figure 11 shows data from a search in which the Arecibo telescope, in the receiving mode, was pointed for twenty minutes at α Ophiuchi, which is fifty-four light years distant. The 3000-channel spectrometer being used detected large signals from neutral hydrogen at the 21-centimeter wavelength, but no signals that could be interpreted as other life. Had there been an Arecibo at α Ophiuchi, it might have made a signal as large as the neutral hydrogen signal, so it is easy to detect manifestations of life that are even weaker than we ourselves manifest. Sensitivity is not the problem.

The problem is dealing with the large number of frequency channels in the radio window of optimum sensitivity. Only in recent years have we begun to cope by making use of modern computer technology. The first step in this direction (Fig. 12) was a system so small it was called suitcase SETI (Search for ExtraTerrestrial Intelligence). The system costs only $20,000, uses a personal computer, a video tape recorder for data acquisition, and a custom-made Fourier transformer that allows one to



```
Running data input 3 (on source)
(ESC)(CR) to abort run, D(CR) to change display parameters
waiting for event flag A

-156.256   59.2
-156.225   16.0
 411.303   10.6
 088.568   10.2
  41.046    9.8
 039.704    9.4
 100.993    9.4
  15.375    8.8
avg =  5.9  53 hits   LARGE SIGNAL

-156.256   75.0
-156.225   13.0
  81.490    9.4
-142.470    7.7
 039.704    9.7
 398.312    8.9
-576.576    8.2
-207.447    7.9
avg =  5.7  40 hits   LARGE SIGNAL
```

THE SUITCASE SETI

Fig. 12. Pictured is a Search for ExtraTerrestrial Intelligence system that costs only $20,000 but is capable of measuring 128,000 frequency channels simultaneously. An improved version capable of handling 8 million channels is now in continuous operation at Harvard University.

measure 128,000 channels simultaneously. When this system is searching, it picks out the band in the spectrum with the most intense signal and then increases signal resolution by displaying channels just in that region. The system, developed by Paul Horowitz, has now been expanded to 8 million channels and is being used continuously at Harvard University to search for extraterrestrial signals.

When we examine the search problem carefully, we see that we have a large n-dimensional search to explore, where n is of the order of seven. We must deal realistically with the ten-million-star problem, the thousand-light-year problem, and the fact that there are literally tens or hundreds of millions of possible frequency channels, even in the relatively narrow band of optimum wavelengths. Furthermore, what signal format is appropriate, pulses or continuous wave? Is the signal on all the time or only occasionally? Is it polarized?

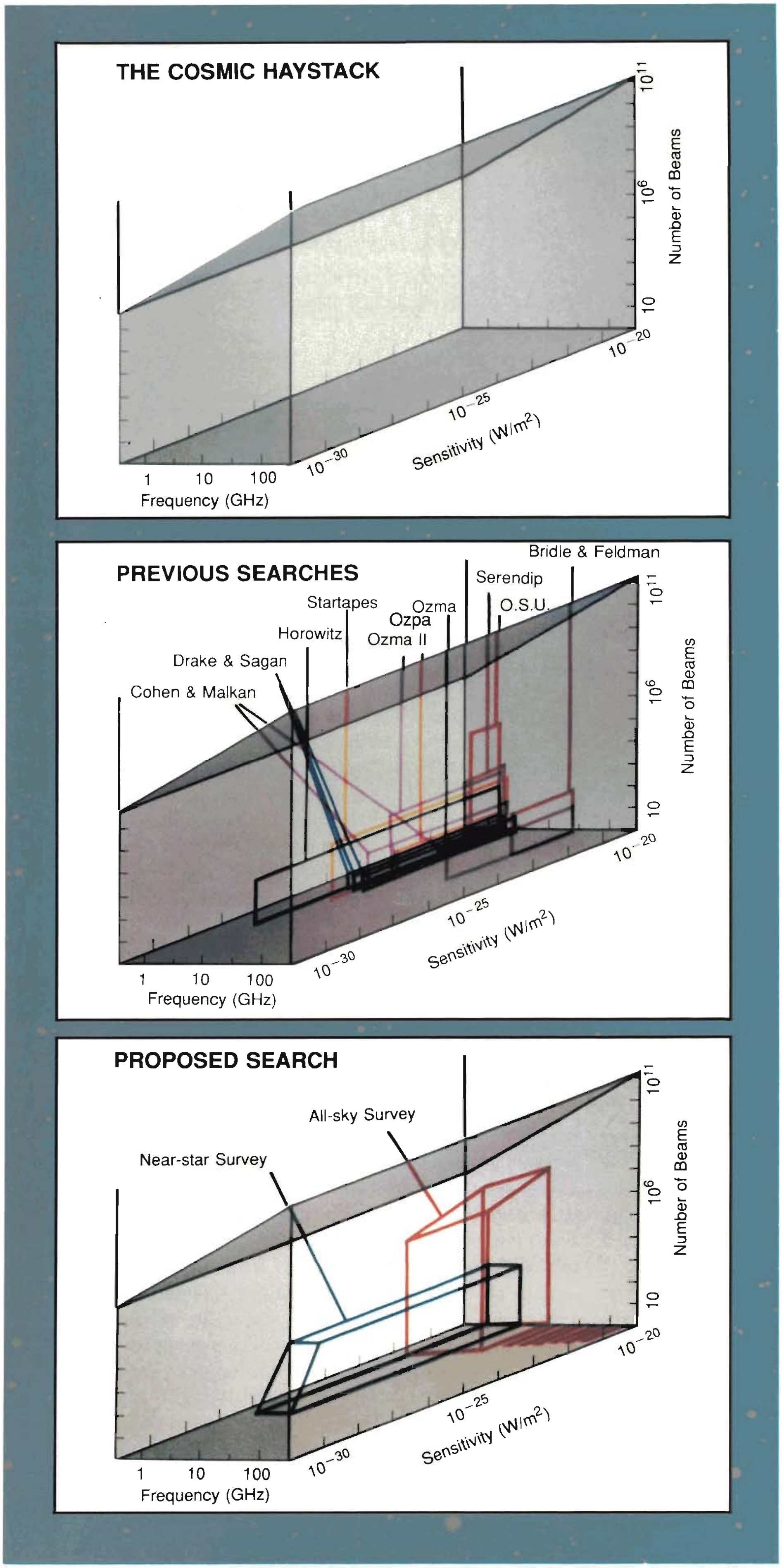
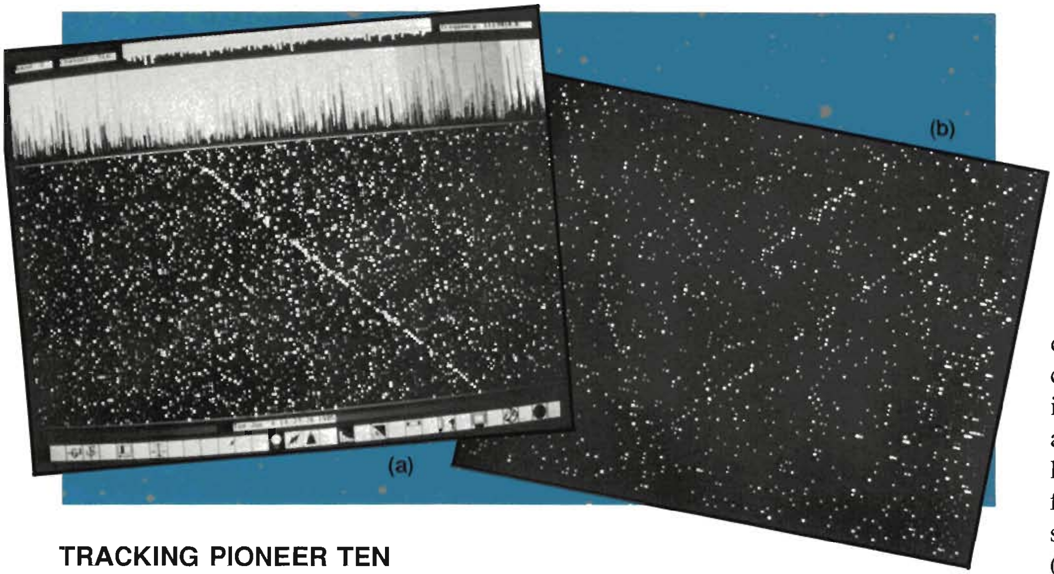


Fig. 13. (a) The *three* most difficult variables to cover in the search for extraterrestrial intelligence are signal strength at the earth (represented by sensitivity in watts per square meter), frequency coverage (in gigahertz), and receiving direction (represented by the number of beams examined). (b) The volume of the cosmic haystack covered in previous searches is actually very small: the width of each search volume shown is gossamer thin in frequency and, moreover, is plotted along a logarithmic scale. (c) The volume of cosmic haystack to be covered in a proposed ten-year search includes an all-sky survey that will examine most of the brighter stars in the sky over a large range of frequencies and a near-star survey that will examine fewer stars (nearby solar types) over a larger range of brightnesses. Although the frequency range examined for the latter survey will be smaller than for the all-sky survey, it will still be much greater than in previous searches.



TRACKING PIONEER TEN

Fig. 14. (a) The 1-watt signal emanating from Pioneer Ten (currently beyond Pluto 3.3 billion miles from the earth) is monitored by displaying a 128,000-channel spectrum horizontally every second, allowing the eye to pick out the signal as a diagonal line. The multichannel spectrum analyzer that generated these data is essentially what will be used at the NASA Ames Research Center to search for extraterrestrial intelligence. (b) Another example, a simulation, shows how this computerized system can detect signals that are even weaker (by a factor of ten or a hundred) than those from Pioneer Ten. (Photos courtesy NASA.)

(This last question is easy since one need measure only two polarizations.) The hardest problems to deal with are frequency coverage, signal strength at the earth, and direction, or the stars we point our receiver at. These variables—sensitivity, frequency, and number of places searched—make up what we call *the cosmic haystack* (Fig. 13a). Somewhere in the faint signals of that haystack are the diamonds: signals from other civilizations. Figure 13b shows the searches that have actually been done, starting in 1960 with project Ozma, in which only two stars were looked at. Although the volume of the cosmic haystack that has been searched may look impressive, in fact, it isn't. The width of each of the search volumes is gossamer thin in frequency, and, besides, the scale is logarithmic. So we have hardly touched the cosmic haystack, and it is not surprising that we have not yet detected the signal of an extraterrestrial civilization.

We are currently putting together a search funded by NASA and operated from the NASA Ames Research Center that will cover a much greater volume of the cosmic haystack (Fig. 13c). In fact, it will do ten million times more searching than all previous searches put together. It will contain two components: an all-sky survey to cover the possibility that the easiest civilizations to find are the farthest by looking at every star in the sky, and a high-sensitivity, near-star search that will be successful if the nearest civilizations are the easiest to detect. For such an effort we must have an enormous frequency coverage, which will be done using a multichannel spectrum analyzer that is a broad-band, 8-million-channel system with an overall bandwidth of 250 megahertz connected to a dedicated VAX computer and disk system.

The project goal is to search the volume of the cosmic haystack shown in Fig. 13c for ten years. The system is currently being debugged and improved at the Goldstone tracking station at NASA using a 100-foot dish antenna. In fact, the system has already been used to detect the most distant intelligent signal ever received at the earth: the one coming from the Pioneer Ten spacecraft, which is currently beyond Pluto at a distance of 3.3 billion miles, radiating a total power of 1 watt. Detecting 1 watt 3.3 billion miles from the earth using a 100-foot dish with a pretty good maser is outstanding!

The detection process is impressive. In this case, a Sun computer searches 128,000-channels of the output of the multichannel spectrum analyzer for strong signals. On finding a signal, the data are restricted to about a thousand channels in that region, and a spectrum is taken once per second. These spectra are represented as adjacent horizontal lines; the eye can pick out the signal from Pioneer Ten as a line of points slanting very clearly through the data (Fig. 14a). If you had only a single spectrum, you couldn't be sure of the signal because of other equally strong points. Only with the ensemble and the diagonal line is the signal's presence clear, which is why you need a computer. It must search not for signals in individual channels but for patterns. In fact, the signal could be ten or a hundred times fainter than the one from Pioneer Ten and still be automatically detected by the system (as in the simulations of Fig. 14b).

Heretofore a human being could do this job but now only a computer can. For example, Fig. 15a is a raster of information like that shown in Fig. 14. Is there evidence of an intelligent signal there? I'll even tell you that it consists not of a steady signal but of five equally spaced pulses along a diagonal line. See if you can find the points, then turn the page for the answer (Fig. 15b). Embarrassed? The NASA system is

A SIGNAL?

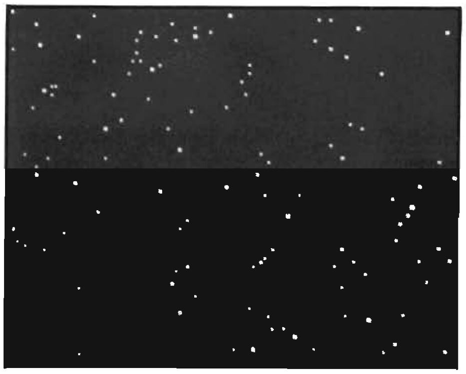


Fig. 15a. Can you spot the signal in this data? Although not continuous, the signal consists of five equally spaced pulses along a diagonal line. When you've completed your analysis, turn the page to see the computer's selection.

THE SIGNAL REVEALED

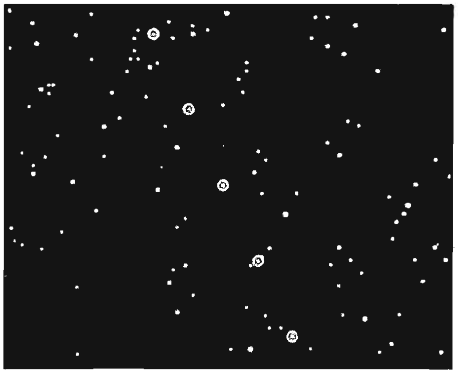


Fig. 15b. The NASA system will be programed to locate patterns such as this one among scattered, random signals.

THE VOYAGER MESSAGE

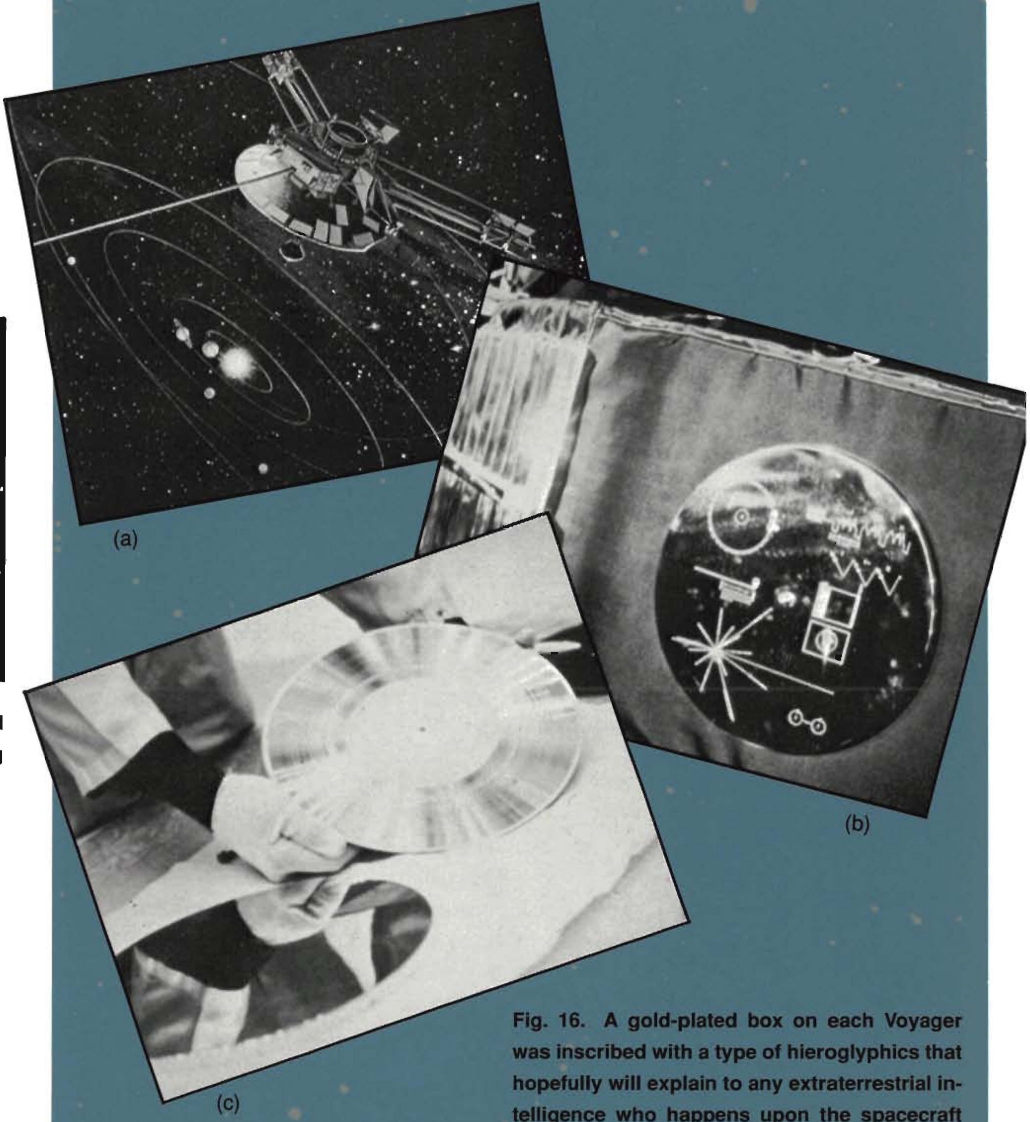


Fig. 16. A gold-plated box on each Voyager was inscribed with a type of hieroglyphics that hopefully will explain to any extraterrestrial intelligence who happens upon the spacecraft its origin. The message also explains how to use the record inside, which contains recorded greetings, an hour and a half of various kinds of music, sounds of the earth, and ten minutes of pictures (see Fig. 17). (Courtesy NASA.)

programmed to find this type of information in real time, and the system will be used in a very powerful search for, hopefully, as many years as needed.

That is where we stand with our searching. As you recognize, we could do a lot more if we spent a lot more money. There wasn't a lot of hardware in the figures, but that is what we've been able to buy with the funds coming from NASA over the last few years.

Messages

I now want to address one last grand unanswered question of life: Can we communicate with them? A number of messages have already been sent into space: two on Pioneer X and XI, two on Voyager I and II, and a radio message sent from Arecibo. How do we communicate with another civilization in a way that we think will work despite not having a common language (fluent Galactic) and no prior contact? My example is a message that is in outer space on both Voyager I and II. (One of these spacecraft recently went by Uranus on its way to Neptune; the other is flying out of the solar system.) Each craft has a gold-plated box inscribed with rather strange hieroglyphics (Fig. 16). Those hieroglyphics hopefully tell another intelligence that there's a record in the box. The record contains one and a half hours of music ranging from Brahms and Beethoven to ethnic music and Blind Willie Nelson—Earth's greatest hits. There are greetings in fifty languages and a long section on the sounds of the earth that range from volcanoes and rainstorms to a baby crying. Finally there are ten minutes of recorded television pictures.

The instructions on the box tell how to convert the waveforms on the record to pictures. Figure 17 represents a small sample that should indicate how we think pictures can be used to communicate simply and without language. (The fact that a cat can recognize a bird through a window but not on the screen of a television gives pause because the cat could be the extraterrestrial who doesn't see things when they're presented as a flat picture. Let's hope other civilizations understand how two-dimensional pictures work.) Some of the pictures show aspects of the earth; all show things that are special about us. We don't tell them about mathematics

A GLIMPSE OF EARTH

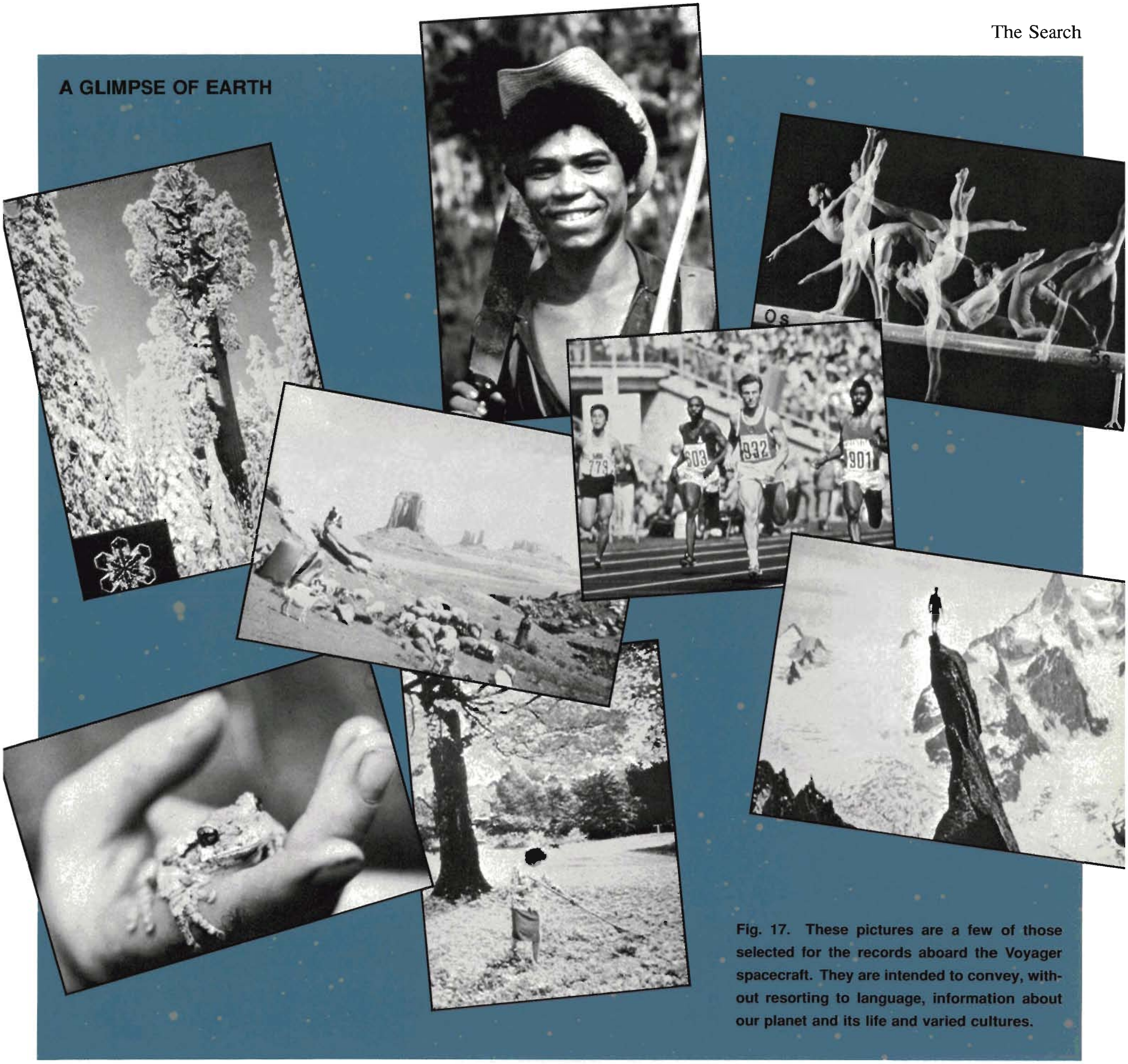


Fig. 17. These pictures are a few of those selected for the records aboard the Voyager spacecraft. They are intended to convey, without resorting to language, information about our planet and its life and varied cultures.

or the laws of physics—they know such things already.

Look at Fig. 17 and try to imagine yourself as an extraterrestrial knowing nothing about humanity. What would you make of these pictures? There are, of course, potential problems of ambiguity. In the Monument Valley picture, which of the animals are being herded, which are doing the herding? We show pictures of different ethnic groups implying that our planet is not a single homogenous society, but does the smile on the Guatemalan field hand mean that he is friendly or that he is getting ready to bite? The fact that he's carrying a machete in his right hand may reinforce the last interpretation. A nice stroboscopic picture of the famous gymnast Cathy Rigby, which has never been published except in outer space, shows the articulation of the human body and what it can do in five seconds. The mountain climbing picture was ostensibly included to prove that we are adventurous but may also show that some of us are crazy. We have tall trees and water in crystalline form on the earth. Some of our creatures have to rake things at a certain time of year when certain things fall off the trees. We have competitive sports, but again if you look carefully you'll see that all four of the creatures have one leg shorter than the other. Is this some subspecies or a second intelligent species? Looking even closer you'll see that all four creatures are four inches off the ground. Have they discovered anti-gravity on the planet earth? Some of the aliens may see the frog and say, "Thank goodness! There's the intelligent creature, and it looks just like us."

All these pictures and more can be sent over a radio link with little effort in less than one second. There is also the potential of receiving similarly rich information in a few seconds without asking a question and having to wait thousands of years for the answer. Such richness is in our future, although we probably will have to build huge radio systems to achieve that capability. But we know how to do such things now; there is no technology that we don't already have, there will just be a lot to build—billions of dollars worth. Although we could build the system on the earth, it might be better in space—large dishes shielded from the earth by huge screens that keep manmade transmissions out of the system (Fig. 18). With a diameter of 5 kilometers, the system could be one of our most idealistic and grandest projects, perhaps, in the long run, one of the best things we could do with our space transportation system. Whether or not we do this depends on how much wisdom and idealism there is on this planet, and that, of course, is one of the other great questions of life. How good are we? ■

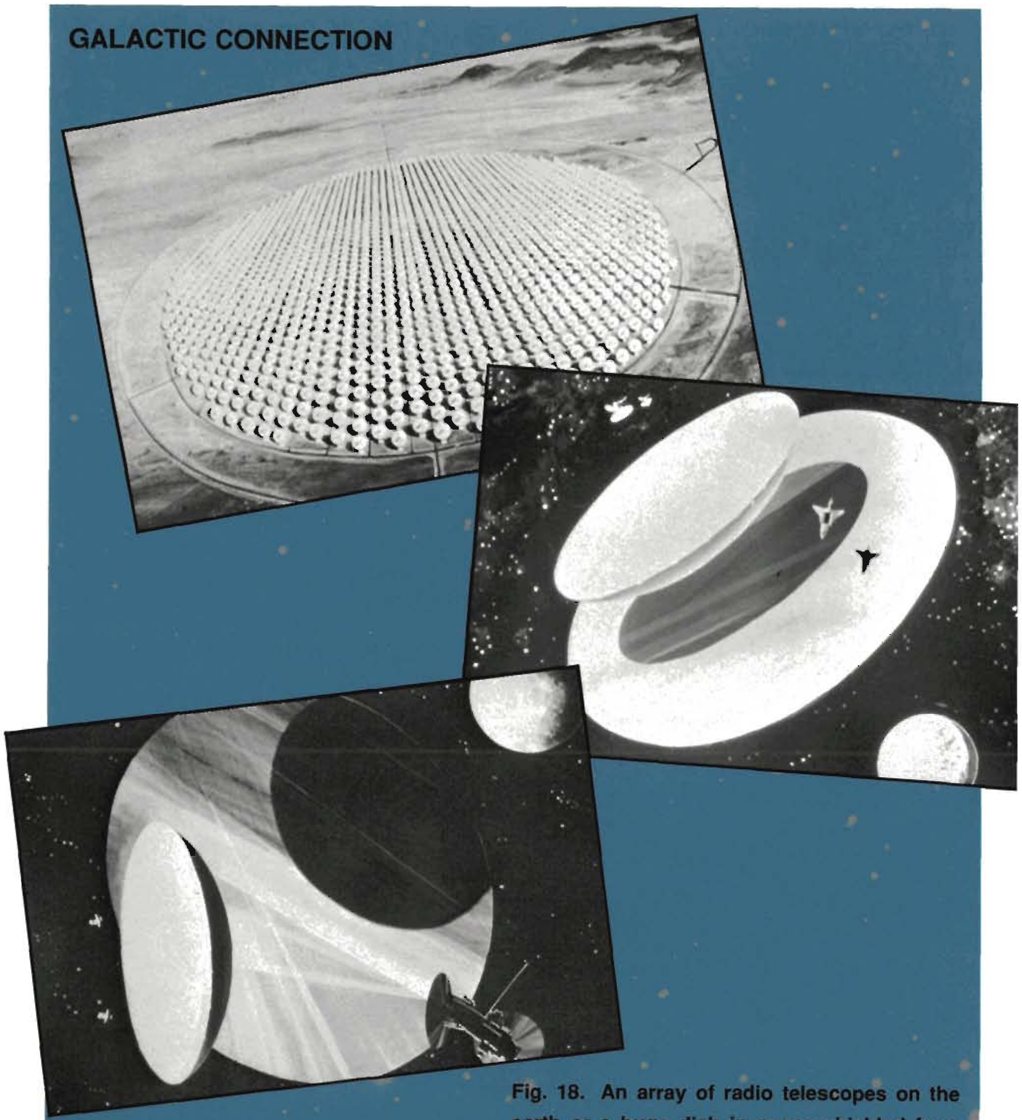


Fig. 18. An array of radio telescopes on the earth or a huge dish in space shielded from manmade transmissions on our planet may one day serve to communicate with other civilizations in the galaxy. (Courtesy NASA Ames Research Center.)

Questions and Answers

Question: Does the unit of time that is peculiar to the earth—our year—affect the results of the equation for the number of intelligent civilizations?
Drake: It would if we did things in terms of years but the number N is unitless. The equation is a rate of production in things per year times L in years. Thus years cancels out and the unit of time we use doesn't matter.

Question: Why have you chosen inefficient rockets for your examples?
Drake: You are getting into the sophistication of rockets. It's true that what is really important to the rocket is momentum ejected rather than energy, and so there are optimized versions of the antimatter-matter rocket. For example, rather than using the energy to expel gamma rays it can be used to expel hydrogen atoms that serve essentially as propelling pellets. In that way, you can increase the efficiency but only by factors of two, three, or four. Qualitatively there is no difference. With regard to the interstellar ram jet, that, of course, is a nice way to go if you can. But scoops that are hundreds of kilometers across are required to collect the hydrogen atoms, which, in turn, must be funneled to a central point and used efficiently in a fusion reactor. Whether all that technology is possible we do not know. If it were possible you could achieve pretty high speeds.

Question: Is the intent of our listening effort to receive messages or just to expand technology in this area?
Drake: We are listening in the radio spectrum for a variety of signals but signals that would all be *intentionally* transmitted. We are looking for continuous wave signals, we are looking for pulse trains, we are looking for drifting pulse trains, we are looking for polarization-modulated waves—all the various things that Maxwell's equations allow in electromagnetic radiation. It's this aspect that's special about the NASA search over previous searches. Previous ones have searched only for continuously transmitted signals at a fixed frequency. The NASA project looks for all varieties of signals, and that's what costs a lot and requires a big computer capacity.

Are They Near or Far?

How does one determine which civilizations might be more detectable, those near or those far from us? Suppose that technological civilizations radiating energy at a power level P in the range dP occupy space with a certain density $\rho(P)$. If the minimum power we can detect at the earth is $P_{\min}$, the number of civilizations that are detectable from a distance R is then given by

$$n(P) = \rho(P)\frac{4}{3}\pi R^3\left(\frac{P}{P_{\min}}\right)^{\frac{3}{2}}. \qquad (1)$$

If we now assume that the density of radiating civilizations obeys a power-law distribution, $\rho(P) = KP^{-\alpha}$, where K and α are undetermined constants, then $n(P) \propto P^{\frac{3}{2}-\alpha}$. The ratio of the detectable civilizations above and below a certain power level P_1 is thus

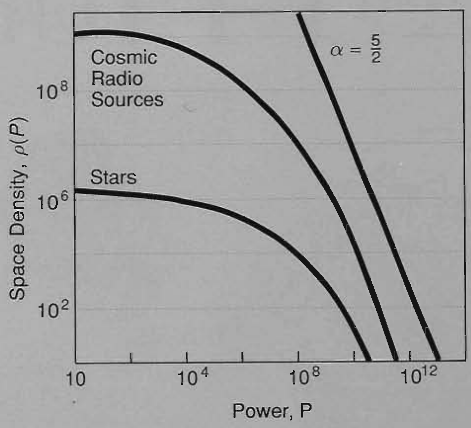
$$\begin{aligned}\frac{N_{P>P_1}}{N_{P<P_1}} &= \frac{\int_{P_1}^{\gamma P_1} n(P)\,dP}{\int_0^{P_1} n(P)\,dP}\\ &= \frac{\int_{P_1}^{\gamma P_1} P^{\frac{3}{2}-\alpha}\,dP}{\int_0^{P_1} P^{\frac{3}{2}-\alpha}\,dP} \qquad (2)\\ &= \frac{\gamma^{\frac{5}{2}-\alpha}-1}{1},\end{aligned}$$

where N is the integrated number of detectable civilizations in the specified range, P_1 is a breakpoint power level, and γP_1 is the maximum power a civilization might radiate. In general, the maximum power will be orders of magnitude larger than P_1, and $\gamma \gg 1$.

Now what does this equation tell us about the bright, detectable civilizations? Are they near or far from us? If the ratio represented by Eq. 2 is greater than 1, the number of bright civilizations detectable despite large distances from the earth will be larger than the number of dim civilizations detectable only when they are close. In other words, the brightest civilizations as seen from the earth are more likely to be far away if this ratio is greater than 1. This will be the case if

$$\gamma^{\frac{5}{2}-\alpha} > 2,$$
$$\text{or} \quad \alpha + \frac{\ln 2}{\ln\gamma} < \frac{5}{2}. \qquad (3)$$

THE LUMINOSITY FUNCTION

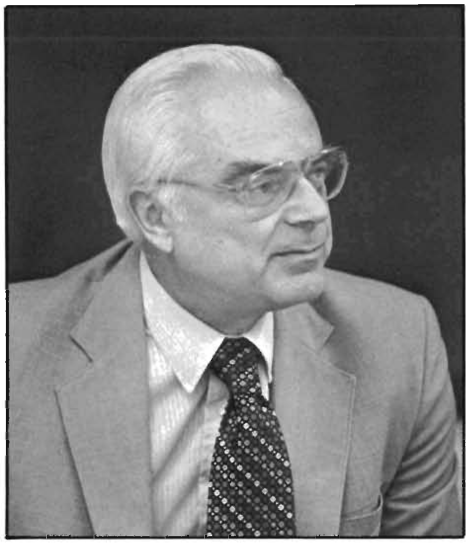


In general, the brightest power levels are orders of magnitude larger than the threshold power P_1, so the $\ln 2/\ln\gamma$ term in Eq. 3 will be negligible. We can thus simply say that if $\alpha < 5/2$, the brightest civilizations as detected at the earth are far from us. As the exponent in the power law approaches 5/2, we move to the other extreme: The ratio in Eq. 2 goes to 0. In other words, the dim stars dominate, and we will most likely find our civilizations among the close stars.

The figure above represents plots of the space density of objects emitting at power P versus that power in arbitrary units. Thus the $\alpha = 5/2$ line represents the situation of dim, near objects completely dominating as the type of object detectable at the earth. However, we see that both the plot for cosmic radio sources and, even more so, the plot for stars deviate considerably from the $\alpha = 5/2$ line, implying that it is the distant, bright civilizations that are more likely to be detected at the earth. ■

Star fields courtesy of Galen Gisler, Los Alamos National Laboratory.

Credits for photos on page 67: Monument Valley: Ray Manley, Shostal Associates; Fallen leaves: Jodi Cobb, (c) National Geographic Society; Sequoia and snowflake: photo by Josef Muench (Robert F. Sisson, (c) National Geographic Society); Tree toad: David Wickstrom; Man from Guatemala: United Nations; Mountain climber: Gaston Rébuffat; Gymnast Cathy Rigby: (c) 1971 Phillip Leonian, photographed for Sports Illustrated; Olympic sprinters: Picturepoint London.



Frank Drake earned his Bachelor of Engineering, Physics, with honors at Cornell University and his M.S. and Ph.D. in astronomy at Harvard University. While a professor at Cornell he was director of the Arecibo Observatory in Puerto Rico. From 1971 to 1981 he was the director of the National Astronomy and Ionospheric Center, and about three years ago he moved from the east to the west coast to become Dean of Natural Sciences at the University of California, Santa Cruz. Although he has done a variety of work in astrophysics, including research on pulsars and the radio noise from Jupiter, he is most widely known for his belief that intelligent life exists elsewhere in the universe. Beginning in 1960 with pioneering efforts on Project Ozma, he became a leading authority on methods to detect signals emitted by extraterrestrial life. He and Carl Sagan helped design the messages that have left our solar system inscribed on plaques and records aboard the Pioneer and Voyager spacecraft.

Unsolved Problems

Colloquium photographs by Fred Rick; posterizations by Chris Lindberg.

A morning of discussion moderated by Mark W. Bitensky

This morning we have the very pleasant opportunity to continue learning from the four dedicated students of biology who lectured yesterday on unsolved problems in the science of life. George Wald recounted the litany of anomalies that characterize the progeny of the big bang and introduced a deus ex machina—mind itself—as a driving force in evolution. David Hubel described what is known about how the detailed visual features of movement, form, and color are analyzed by the occipital cortex. John Sepkoski convinced us that extinction, like speciation, must be regarded as an integral part of evolution, playing the critical role of "making place" for newly evolving species. And Frank Drake projected a cosmos full of life and intelligence and with marvelous humor described efforts to communicate with that intelligent life.

I have consulted with our guests, and they have to a man agreed to a full and free-flowing discussion. I request only that questions and comments be clear and brief. Let us begin.

Audience: I have a question for Frank Drake. What countries are searching for extraterrestrial beings?

Drake: Two countries are making major efforts—the United States and the Soviet Union. The Soviets have been searching now for twenty years. In fact, for a long time they were the only people searching. One of their projects, which is based at the Lebedev Physical Institute in Moscow, uses an array of about five radio-frequency receivers placed across the Soviet Union. A similar network is operated from the Gorky Research Radiophysical Institute. Both institutions have, until recently, been looking for short but powerful radio-frequency pulses, a type of signal very different from what we Americans are looking for. They recognize, as we do, that one of the really difficult aspects of a search is selecting the search frequency. Their way of finessing that problem is to look for short pulses, which appear on all frequencies. Their hope is that the extraterrestrials are thinking the same way and are transmitting short pulses.

Now the problem with short pulses is that human activities—operating cars and motorcycles, for instance—produce lots of them. So the Soviets look for short pulses that are coincident in an array of widely separated telescopes. If a pulse is cosmic, it will appear at all stations, but if it is interference, it appears only at one.

So far the Soviets have detected two interesting sources of coincident short pulses. One is the sun, and nobody had known before that the sun emits short radio-frequency pulses. The other was an American reconnaissance satellite that transmits information in the form of big, short radio-frequency bursts over a broad and variable band of frequencies to hinder reception by unfriendly receivers. But the Soviets did pick the signal up, and it got them very excited until they were told what the source was.

One of the problems with the Soviet program is that their small antennas can detect only very strong signals. In fact, to be detected by their system, a source at a reasonable distance of 1000 light years must have a luminosity equal to that of the sun. So the Soviet search will detect only those civilizations with capabilities well beyond those of earthlings, and for that reason the Americans don't think it is very effective.

The Soviets are also building a 70-meter steerable, parabolic radio telescope on a mountain near Samarkand, which is to be used not only for conventional radio astronomy but also in a program similar to that of the Americans.

I should note that Canada, France, the Netherlands, and Australia have also carried out searches, but theirs have been less extensive than the Soviet and American efforts.

Audience: I have a question for Professor Hubel. What chemicals are involved in visual perception, and are the transport mechanisms electronic or ionic?

Hubel: Your question has major subheadings. One concerns how nerve impulses are transported along nerve fibers, or axons. There is a certain electric potential—about a tenth of a volt—across the membrane of the axon of a nerve at rest. But when some stimulus reaches the beginning of the axon, ion channels in the membrane there open briefly, positive ions flow into the axon, and the membrane potential changes. The potential change at the next region along the axon is somewhat less, but if it is still great enough to cause ion channels there to open, it is augmented by another influx of positive ions. Because of that positive feedback, the change in potential travels unattenuated along the length of the axon. The impulse travels along the axon like the snap of a rope at one end travels to the other end. Information, rather than anything physical, is conducted. But the transport is ionic in the sense that it involves the flow of ions rather than electrons.

When the impulse gets to the specialized structures, the terminals, at the end of the axon, the change in potential there causes release of a substance called a neurotransmitter. The transmitter diffuses to the next nerve and, by changing its permeability to ions, makes that nerve either more or less likely to fire. Between twenty-five and fifty neurotransmitters are known, although as short a time ago as about twelve years only four were known. New ones are being discovered every year. All the known neurotransmitters are very small molecules. Many, like gamma-aminobutyric acid, are amino acids. The enzyme acetylcholine and the hormones epinephrine, or adrenaline, and norepinephrine are among the most common. Why so many neurotransmitters exist is not known.

Audience: But if the transport of a nerve impulse is ionic, how can the impulse travel so fast?

Hubel: The speed of transmission, which ranges from about 1 meter per second to about 100 meters per second depending on the type of axon, is entirely predictable from such factors as the capacitance across the axon membrane and the permeability of the membrane to ions. You apply an equation not much more sophisticated than Ohm's law and out comes the transmission speed. One of the reasons nerve impulses travel so fast is the fact that axons are encased, everywhere except at particular points called nodes of Ranvier, in an insulating sheath of myelin. The flow of ions through the membrane occurs almost exclusively at those nodes, which are about a millimeter apart.

Audience: Is the same mechanism involved in the transport of audio signals?

Hubel: There are no basic differences between the transport of auditory and visual signals. Each nerve system has some very specialized cells, but essentially the same transport mechanism is involved.

Bitensky: Are the neurotransmitters small so they can diffuse rapidly, and does their variety support subtle dialogues among nerves?

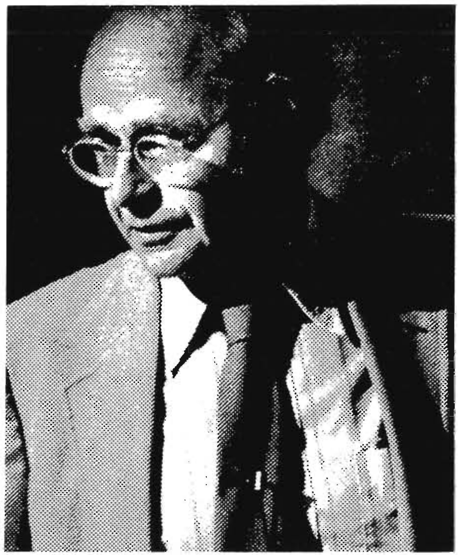
Hubel: Well, yes to the first question. The smallness of the molecules probably reflects an evolutionary drive for faster diffusion and easier release and uptake. Concerning the second question, the terminals of certain axons contain many different transmitters, so the opportunity for a much more complex dialogue exists. But I don't know of any cases in which more than two are released. Usually one is a so-called modulator, and the other is really doing the job. The modulator may change certain things, but in fact usually it is not known why more than one is released. It can be shown that one is enough to do the job.

Bitensky: Do neurons react to a variety of transmitters?

Hubel: Usually to at least two—one excitatory and one inhibitory.

Audience: My question is addressed to anyone who wishes to respond. In view of the complexities of the human nervous system, do you think computer-based artificial intelligence makes any sense?

Hubel: That is something I think about a lot because I have quite a bit of dialogue with a number of friends who work on artificial intelligence. I think that the majority of people in artificial intelligence are not trying to produce a thinking brain, or anything like one, but to build intelligent machines for image translation, robotics, and so on. Those are very worthwhile goals, so one can't object to them any more than one can object to the goals of, say, electronic engineers. On the other hand, a certain number of people in artificial intelligence are trying to learn how the brain works by developing computer programs to solve problems the brain is known to have to solve. They then ask whether the brain solves the problem the same way. Their efforts are very useful because the more people who are thinking about how the brain might work, the more guidance we have as to the type of experiments that we might do. I'm not sure whether that is the answer you want.


David H. Hubel

Bitensky: The differences between brain and computer are very striking. The brain is terribly slow compared with the computer, but the richness of its interconnections— about 10^{15} synapses—is far, far greater. Many scientists in artificial intelligence say vehemently that it is just as absurd to try to emulate the brain as it is to try to fly like a bird. Fixed-wing airplanes are quite different from birds. Certainly, many fascinating things may emerge from understanding how the brain solves various problems.

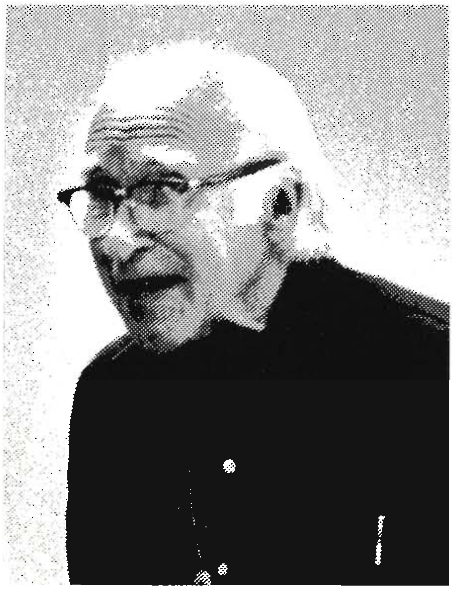
Audience: Would any of the panel care to comment on whether extrasensory perception—ESP—is an unsolved problem in the science of life?

Drake: I'll be glad to answer that one. About once a week I get a letter from someone who tells me I am wasting my time because he or she is already in contact with the extraterrestrials through ESP. My response is always to ask the person to tell me something the extraterrestrials know that we don't know already. So far I've gotten no response. Adding to my skepticism is the large number of experiments conducted daily that very conclusively refute ESP. Those experiments take place primarily in two places—Reno and Las Vegas. The odds of winning some of the games of chance played there, say blackjack or roulette, are about 1 percent lower than the odds of losing. So if even a very few people had enough ESP to foresee or influence what is going to happen even 1 percent of the time, they could become regular winners and run the casinos out of business. The entire gambling industry would collapse. As far as I'm concerned, the fact that the casinos continue raking in the money day by day proves conslusively that ESP does not exist.

Audience: My question is addressed to George Wald. Although Wilder Penfield may have been unable to locate mind as a thing in the cerebral cortex, he very definitely showed that mind as a process is located in specific hard-wired structures in the brain. So can't we say the the mind is located totally in the cerebral cortex and in the reticular formation?

Wald: I can only comment. I spend a great deal of time trying to sort out the obviously sloppy ways in which the words mind and consciousness are used. Yes, indeed, we can determine to a degree the pieces of machinery that are involved in the workings of the mind or consciousness. But where does that get us? Some great physicists have essentially said that all matter has an accompaniment of mind. What do they mean by that? They don't mean that stones are intelligent as we understand intelligence, still less that stones are self-aware as we experience self-awareness. Let me try to explain what they mean. A former professor of physiology at Harvard Medical School, Walter Cannon—whom I remember as a very wise person—wrote a book called *The Wisdom of the Body*. What did he talk about in that book? Well, he talked about the very fine regulation of the concentration of glucose in the blood, of body temperature, of the pH of body fluids, and so on. As the great physiologist Claude Bernard said, the constancy of the internal environment is the condition of a free life. We can go to the Arctic or the tropics, and we are free because of all that internal regulation. But please note that the regulation is unconscious. It has nothing to do with will or intelligence. In fact, one can only interfere with the regulation by intruding with one's intelligence. The English scientist Galton tried for one day not to draw a breath without willing it. At some point he decided he'd had enough of the willing and then was deeply embarrassed to find that his breathing stopped. If he hadn't somehow gotten through that phase, he would have probably passed out, and the unconscious regulation would have taken over again.

Now in exactly the sense that one can speak of the wisdom of the body, one can speak of the wisdom of the planet, of the solar system, of the universe. But it is *wisdom*, not intelligence, and wisdom in the sense of fitting together. I may have seemed yesterday to be disparaging silicon when I said, "And that's why silicon is good for making rocks, but to make living organisms, we need carbon." But if silicon weren't good for making rocks, we wouldn't be here. Rocks are the skeletons of the planets—so thank heavens for rocks. Things fit to a remarkable degree.


George Wald

Mind or consciousness are involved in a tremendous range of human activities. At one end is the child learning to avoid the fire. An awful lot of learning is just personal housekeeping. At the other end is mathematics. Tell me where mathematics comes from and wherein lies the rightness of mathematical thinking. One might think first of its self-consistency, but Gödel poked holes in that. I lived next door to a mathematician for a while—and I mean a creative mathematician, one who makes mathematics, not just uses it—and I never saw that man working. He spent his mornings in the bathtub and his afternoons quietly walking up and down the street with his little children. But he was a fine mathematician. Eugene Wigner wrote a nice essay asking how it is that mathematics fits physics so well. He concludes that it is simply a miracle, one for which we should be grateful.

I want to mention what Schrödinger, no mean physicist, says in the last chapter of his book *What Is Life?*. He says that he has been interested in Eastern philosophy for many years, and then he asks whether we are perhaps mistaken in thinking that there are as many minds as there are bodies. Clearly there are many bodies, but perhaps there are many fewer minds, perhaps only one. I do not believe in ESP, but I do think that the experiments done to determine whether ESP exists are laughable. They are like trying to produce a physical explanation for the existence of God. But the idea of one mind has something in it.

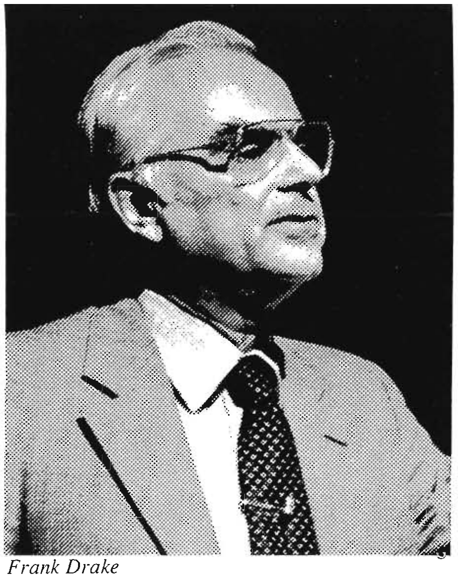
Let me say one more thing. The Judeo-Christian god *made* the universe; the Hindu supreme god Brahman *thinks* the universe. Is it possible to think reality? Theoretical physicists seem to do it. It started when Paul Dirac saw that his wave equation for the electron was satisfied by another particle of opposite electric charge, and then that particle—the positron—was discovered the next year. Now it is pretty much taken for granted that if a theory describing some known aspect of reality has alternative solutions, those solutions also have physical reality.

Hubel: I would like to respond to the original question. We humans tend to make up words that have perfectly good uses—the word *sky* is a good example—and then try to reify those words, to identify them with physical things. The mind can no more be regarded as a thing located some particular place than the sky can. But astronomers don't study "the sky" or "the heavens" or worry about where "the sky" is. They study all that we know constitutes "the sky". Some day we may regard the mind and consciousness the same way.

My hang-up with what George talked about yesterday has to do with what makes biology profoundly different from physics and the other natural sciences, namely evolution. In terms of evolution, the mechanisms responsible for Cannon's wisdom of the body are very well understood right down to the molecular level. We have no indication that any such guiding force exists in, say, physics. I'm thankful that ice floats and that carbon atoms form chains, and no one can say what a universe in which those facts didn't hold would be like. But I don't go along with invoking an all-permeating mind or consciousness to explain them. The very idea of a permeating force is a religious concept. It falls outside the realm of science. We have come a long way, thanks to scientists like Darwin, toward transcending the conflicts between science and religion. It is true that some scientists—Sherrington, Penfield, Eccles, and Schrödinger, for example—commit one part of their consciousness to science and hold in reserve some marginal part that is the source of soft statements. I find those statements disturbing because they tend to become identified as scientific statements although they are not. I'm not suggesting we should ignore everything outside science. I don't think *that* highly of science. But it's a good game to be in. It's very interesting, so interesting that I find talk about ESP rather silly. There are enough things to say gee whiz about in real science that we don't need silly things like astrology to keep ourselves happy.

Audience: In the mid thirties von Neumann suggested that consciousness might play a very significant role in the interpretation of quantum theory, in the understanding of what measurement means. That idea was followed up by London and Bauer and is being pursued to this day by Eugene Wigner. It may be at the roots of one of the great physical theories of our time. I would like Professor Wald to comment on what he feels the role of consciousness might be in future theories of matter.

Wald: First I want to respond to what David said. I am a scientist, and very glad to be one. In fact, I have spent my life pretty much as a strict constructionist in science. I certainly think that evolution is a great thing and that the wisdom of the body is understood. Perhaps natural selection was involved in the evolution of a universe in which ice floats and carbon atoms form chains. Also I believe that every thoughtful scientist realizes that science deals only with a marginal part of reality. Reality is the very big picture, and science can deal cleanly and quantitatively with only a portion of that reality. Science cannot deal with what are in many ways the most important aspects of our lives. A Harvard great, the mathematician George Birkhoff—do you know what a Harvard great is? A Harvard professor who is still there so he can tell you he is great— wrote a book called *Aesthetic Measure*. In it he presented a formula by which one could quantitatively assess the aesthetic value of a work of art, such as Beethoven's Ninth Symphony or Rembrandt's *Self-Portrait*. Then he decided to write a sonnet that would rate 100, and he did. It was a lousy sonnet. That is what the computer might do—write lousy sonnets.

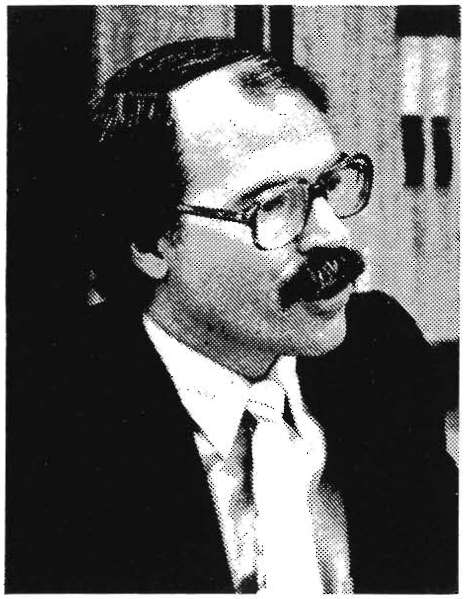


Frank Drake

Hubel: George, I think you are misconstruing what I said. I would be the first to agree that science plays a marginal role in our lives and has little if any immediate relationship to the most important things we do and say. I was expressing a negative opinion about scientists who include, as a last chapter in a supposedly scientific work, their wooly, nonscientific, difficult to understand thoughts about, say, mind and consciousness.

Wald: I don't share your negative feelings. Science is a path, one among many. It is the path to the boundaries of what we know. Of the many paths I prefer that of science to all others, perhaps because it does have boundaries. You seem to be saying that scientists shouldn't look beyond the boundaries, and if they do they should keep their wooly thoughts to themselves. Many scientists have looked beyond the boundaries—Newton, Maxwell, and so on down the line. How does one have the temerity to speak with superiority of such people? I'll admit, though, that we played that game as graduate students, saying too bad about that last chapter of Jeans', too bad about that last chapter of Eddington's.

Bitensky: It's time to move the discussion forward. I believe that David was not telling scientists to stay within the boundaries. He was saying that what lies beyond is simply not science. Now I would like Jack Sepkoski to comment



J. John Sepkoski, Jr.

on how human consciousness might affect the speciation and extinction that characterize evolution on the earth.

Sepkoski: First I want to emphasize that my comments yesterday about the constructive aspects of extinction—constructive, that is, on a time scale of several tens of millions of years—were not meant to lend support to a so-what attitude toward the effects of human activities on the biosphere. After all, we have no way of knowing whether those effects will, in the long term, be constructive or destructive, and what may be constructive to the entire system in the long term may be very destructive to individual species, even ourselves, in the short term. But it is fairly clear that massive re-engineering of the earth is causing a departure from Darwinian evolution, and genetic engineering can only bring about an even greater departure.

I should like to comment that I am less amazed by the existence of life than Professor Wald is, perhaps because of the rashness of my relative youth. Also I don't view intelligence as the pinnacle of creation, as being pre-eminent in and of itself. Intelligence is only one solution to survival, one that has been tried by a variety of organisms. Some organisms, the social insects, for example, rely on collective rather than individual intelligence. But survival of a species can be promoted by any number of tricks—by being camouflaged or showy, by being able to run fast, to reproduce quickly, to climb trees. That is why we enjoy such a rich variety of fauna and flora. But human intelligence coupled with culture is a factor very different in kind from those at work in evolution until the last few centuries, and the biosphere faces a whole new ball game. Before evolution had no purpose; it produced what could survive, not what should survive. Now the biosphere is increasingly subject to human purposes, to the uses we make of the earth. Thus, the history of the earth, which extends back some four and a half billion years, has moved into a very different era.


Mark W. Bitensky

Audience: May I ask Dr. Drake what will be the next step after signals from an extraterrestrial civilization are detected?

Drake: That will depend on what we detect. What is most likely to be detected is a signal at a signal-to-noise level so low that no information can be extracted from the signal. So we will know only that another civilization exists. But of course that will be big news in itself. Then we must do whatever is required—build a much larger radiotelescope system, for example—to obtain information about that civilization. That information may have a great influence upon our own civilization. Or it may turn out that the extraterrestrials are so different from us that learning about them will be motivated only by scientific curiosity, like learning about the ecology of elephant seals.

Audience: Dr. Sepkoski, you implied yesterday that some maximum number of species exists at any given time. What might be the mechanisms for enforcing that maximum?

Sepkoski: The maximum is a relative, not an absolute, maximum. Probably some absolute maximum exists, since the earth can support only so much biomass and the efficiency of energy transfer can be only so great. But the number of species existing at any time has never been anywhere near the limit imposed by those factors. On the earth today we see local ecological systems, particularly islands, approaching an approximate equilibrium as new species appear and existing species vanish. The equilibrium number of species can increase or decrease, however, if a pool of species is introduced that uses the habitat and its resources in an entirely different way. That has been observed, for example, on oceanic islands and in a number of habitat islands on the continents. The fossil record over hundreds of millions of years for, say, the whole oceanic ecosystem presents a very similar picture. We see an approximately constant number of families and genera. We also see the equilibrium perturbed by several large mass extinctions and then quick rebounds to the former level. That level seems to be maintained by background extinction, or slow attrition, of existing species and slow replacement by other species. The slow attrition of species was probably caused by competition among organisms for limited resources as well as by perurbations or small "catastrophes" in local ecological communities; replacement resulted from normal processes of speciation. We also see jumps in the steady-state number of taxa when a different style of fauna appears, for example, when Ordovician fauna replaced Cambrian fauna. And the animals that became dominant after the great Permian extinction did things in yet a different way. So it makes sense that the number of species in the oceans today is greater than it was 250 million years ago. We can imagine that if marine organisms found yet another way of organizing ecosystems, their number might jump even higher. We see terrestrial paral-

lels, especially among plants—plateaus of diversity maintained by balanced speciation and extinction and jumps in diversity due to new ways of doing things, particularly at the advent of angiosperms. We have some hints of the same thing going on with vertebrates and perhaps with insects, but the insect fossil record is pretty messy.

Audience: I'd like Dr. Sepkoski to comment on Fred Hoyle's theory that life at some level pervades the universe and that that cosmic life is the origin of life on the earth.

Sepkoski: I haven't thought very critically about Hoyle's version of panspermia because I find it too easy to dismiss out of hand. Many ideas of that sort are based on the notions that 4.5 billion years is not enough time to produce the diversity of life we see on the earth today or that 1 billion years—the time between the birth of the earth and the age of the earliest fossil evidence of life—is not sufficient time to produce life itself. Unfortunately, we don't have theoretical principles of evolution with which we can quantitatively predict absolute rates of evolution. All we have right now is an ability to measure relative rates of evolution in some situations. My impression from looking at the fossil record, though, is that evolution can work extraordinarily fast when it is unconstrained. In the absence of competition and crowding, mutations and other accidents produce a huge array of variations from which natural selection can produce a wonderful array of outcomes. I don't believe we need panspermia or any other means of inoculating the earth with life. But that's only my gut reaction to what I see in the fossil record, and we do need quantitative theories before we can say definitively that such hypotheses are unnecessary.

Bitensky: When we speak about evolution, we are really talking about the evolution of DNA, and there is now a lot of evidence that the shuffling of whole exons is one of the changes that occur in DNA. That shuffling allows the mixing of very big pieces of DNA and so could be responsible for very rapid evolution.

Sepkoski: Exon shuffling certainly leads to rapid rates of change in DNA, but in fact speciation doesn't require any changes in DNA. The extraordinary genetic variations among individuals of a species is more than sufficient. Mutation could cease today, and after tens of hundreds of millions of years a far different biota would inhabit the earth.

Bitensky: But the extraordinary variation is, in retrospect, a reflection of the plasticity and heterogeneity of DNA, which is made possible by shuffling.

Audience: My question has several parts and is addressed to Professor Sepkoski and Professor Drake. First, what mechanism is behind the rapid increase in number of taxa after a mass extinction? Second, is there some mechanism that prevents the simultaneous existence of more than one intelligent species? And finally, wouldn't it be very depressing for us humans to come into contact with a civilization much more advanced than ours?

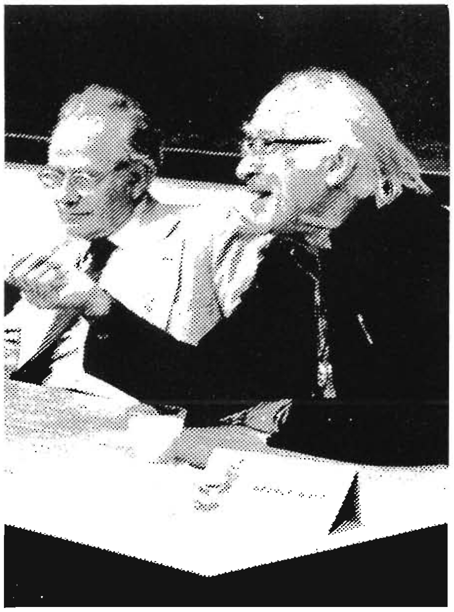
Sepkoski: In answer to your first question, we see rapid evolution following mass extinctions because of a change not in the process but in the boundary conditions. Variations occur all the time, but most of the variations don't survive. Most new species probably arise from small, local, slightly variant populations of existing species. But ecologists have learned that such local populations disappear at phenomenal rates, probably because of competitive pressures that keep them small in size and hence susceptible to extinction. But if somehow the lid of competition is lifted so that populations can expand, then the probability of their extinction goes way down. And then we see rapid increases in number of species.

Turning next to the question about a possible limit on the number of intelligent species, first we need to define intelligence. I prefer an operational definition, as a measure of the ability to control, to re-engineer, the local environment in a nonstereotyped way. I mentioned before that a variety of animals can re-engineer habitat, and they are not all even mammals. I think that competition is inevitable if more than one intelligent species exists, and in that competition only one will win, will become pre-eminent. Now I yield to Frank.

Drake: Why is there only one intelligent species on the earth? Because of the greed and selfishness of *Homo sapiens*. The fossil record indicates that at some times more than one intelligent species inhabited the earth simultaneously—Neanderthal man and Cro-Magnon man, for example. The fossil skulls of those other species often show signs of having been hit with a blunt instrument, and one suspects that it was *Homo sapiens* who was wielding the blunt instrument, getting rid of all competitors.

The graph of number of species versus brain weight—corrected for body weight—is very interesting. For aquatic creatures, particularly aquatic mammals, the curve is continuous. There are species with brains almost the size of those of the dolphins and killer whales, which have the largest brains. But the curve for terrestrial mammals is continuous only up to a certain brain size, then a gap occurs, and beyond that gap there is only one species—*Homo sapiens*. What created that gap? We did. We eliminated the competition to have the earth to ourselves. That is our nature, and not something to be very proud of. The dolphins and the killer whales have not done the same thing. Terrestrial mammals seem prone to population explosions, and the resulting population pressure leads to fierce competition. But marine mammals do not engage in population explosions. The populations of dolphins and killer whales could expand enormously, since they have very few predators, but for unknown reasons that doesn't happen. So those most intelligent marine mammals have no drive to eliminate near rivals.

In any case one intelligent species dominates the terrestrial ecosystem on the earth. What can be said about the universe? If the other intelligent creatures out there are like us, then they will want to eliminate near competitors. So when they see a new intelligent species emerging, they will stamp it out just as we stamped out the australopithecines. However, the extreme expense of interstellar travel may be our salvation, since no possible benefit could justify the cost. On the earth inferior cultures have been exploited by superior cultures, as, for example, the Europeans exploited North America and Polynesia. But getting to North America and Polynesia was easy. Going to a distant solar system for self-protection or economic reasons would cost far more than any possible benefit. So the great distances between stars and the laws of physics create a very effective and beneficial quarantine. Intelligent civilizations that far apart can neither exploit nor attack each other. We hope. But they can help each other by communicating.



Audience: Today's feats of technology would have been regarded as impossible only a few hundred years ago. So isn't it rash to say that physical contact with other intelligent life is unlikely? Maybe traveling faster than the speed of light *is* somehow possible, for example.

Drake: Yes, we certainly should not neglect the possibility that all the physics relevant to this problem is not known. History raises big red warning flags about thinking that we know everything.

Regarding the question about our egos being demolished by contacting a civilization more advanced than ours—and that is the most likely possibility—I don't consider that a problem. We all have been exposed to minds and accomplishments greater than ours. In fact, for most of us that is a continual experience. But the result is more often inspiration rather than depression. I don't believe the human brain is limited in any fundamental way and think it can emulate the power of any intelligence we may find in the universe.

Wald: We humans have stockpiled all the hardware necessary for destruction of our civilization, although at the moment it has not been used. How likely is it that other civilizations have committed suicide with similar hardware and that no one is out there for us to communicate with?

Drake: It has been said that the civilizations we detect will be those that have passed successfully through the nuclear crisis, which will occur in every civilization almost simultaneously with the development of the technology necessary for communication with other civilizations.

Bitensky: Perhaps intelligent extraterrestrial beings are waiting to communicate with us until we prove our worthiness by transcending the nuclear crisis.

Wald: The supposition on the part of many people is that the civilizations we might contact would be benign. Is our civilization benign? We grow viruses in our closest mammalian relatives, we slaughter bottle-nosed dolphins by the millions, and we are far from benign even to our fellow humans. It seems to me that we have more to worry about than simply having our egos crushed.

Hubel: What do anthropologists have to say about greed being the cause of extinction of the predecessors of *Homo*

sapiens? Would greed have been as powerful a force in the early stages of our evolution?
Sepkoski: Greed may be another word for competition, which has been proposed to explain the disappearance of Neanderthals and certain other hominids. Neanderthals had larger brains than *Homo sapiens* does, and a lot of Neanderthal genes may still exist in Europe. A Neanderthal could walk down the street today and cause no comment. The australopithecenes disappeared in Africa about the time *Homo habilis* started to become common. Perhaps their disappearance was due to competition between the two, but perhaps it was due to some change in the ecology.

Audience: Why are we looking for signals originating someplace out there when there are reports of signals from extraterrestrial beings right here?
Drake: I assume you are referring to UFOs. The evidence for UFO sightings that have been studied in detail simply falls apart. Of course, not all the reports have been studied in detail, but those that have can be attributed to natural phenomena or to hoaxes. It would be nice if the intelligent beings came to us—it would make life simple—but I see no evidence that they have.

Let me comment on our failure so far to detect other intelligent life in the universe. The silence we have heard is not in any way significant. We simply have not looked long enough and hard enough, have not explored a large enough chunk of the cosmic haystack. We can speculate that they are watching us to see if we are worth talking to, but an even more likely speculation is the existence of an ethic that says there is no free lunch in the galaxy. If we want to join the community of advanced civilizations, we should have to work as hard as they do. So they would send a signal that can be detected only if we put as much effort into receiving as they do into transmission. They are not going to serve up wondrous things on a silver platter to a new civilization. We must earn access to their information.

Audience: What is the opinion of the panel about research at reputable institutions in areas beyond the boundaries of pragmatic science, for example, at Princeton on engineering anomalies and at Duke on paranormal phenomena?
Wald: I know the people at Princeton, and I like them very much. They are serious and well intentioned, but I am not familiar with the data on the anomalies they are investigating [see *Margins of Reality: The Role of Consciousness in the Physical World* by Robert G. Jahn and Brenda J. Dunne, Harcourt Brace Jovanovich, Inc., 1987]. What most interests me is the very concept of a system of communication that we don't have to pay the telephone company for—a universal mind or a collective mentality. I think that the attempts to study such means of communication are too mechanical, though. What goes on in a good mathematician's head is closer to the answer, and that isn't going to start or stop machinery.
Hubel: People, especially people with little scientific education, can come up with some pretty silly explanations for natural phenomena. But I feel that trying to disprove such explanations is not a sensible strategy. Let them die of attrition as facts accumulate. That usually works, and it certainly saves time and money. I think Duke's venture into the paranormal brought it great discredit scientifically. As for astrology and flying saucers and such, I put them under the heading of things I wouldn't believe even if they were true.
Drake: I want to add a comment here. Most people don't understand statistics and probability, and they don't recognize that although an event may be very

improbable, it does eventually happen if enough opportunities for it to happen exist. So the fact that a friend calls you on the telephone at the exact instant you are thinking about the friend is not evidence for telepathic communication. And I repeat that many experiments refute the idea. We certainly don't see any evidence for telepathic communication between bridge partners, for example. By the way, the ESP project is now disassociated from Duke.

Audience: Dr. Drake, you said yesterday that the rate of generation per galaxy of planets with intelligent life was about one per year. Did you include in your derivation of that rate the possibility that large mass extinctions may be necessary for evolution of intelligent life?
Drake: No, I didn't include that factor, which is rather speculative. But if mass extinctions are somehow involved in evolution of intelligence, we have no reason to believe that the processes that cause them on the earth would not also be operative out there. Clouds of comets, for example, should be present in solar systems other than our own. The rates of mass extinctions would undoubtedly vary from place to place, but that would not significantly change the rate of generation of civilizations.

Bitensky: Despite the richnesss of our discussion, I must point out that we can continue it only a little longer.

Audience: My question is addressed to Dr. Sepkoski. How significant is the difference between the periodicity of extinctions and that of magnetic reversals or of the comet impacts predicted by the Nemesis scenario?
Sepkoski: The difference is real—greater than the uncertainties in the data.
Audience: Then what is the cause of the extinctions?
Sepkoski: I don't know. Impacts are certainly involved in some cases, and so are climatic changes. Maybe the thing to do at this point is to throw in all the data that might be relevant and carry out a huge analysis of variance.
Bitensky: Are you willing to entertain a multiplicity of causes of extinctions?
Sepkoski: Yes. The nonperiodic mass extinction of large mammals that occurred about 10,000 years ago is clear evidence for a multiplicity of causes.

Audience: I believe Einstein is credited with saying that imagination is more important than knowledge. Would any of the panel care to comment on the process of imagination or the enhancement of that process?
Wald: The degree to which we program our children is fantastic. A child is a wonderful thing, and it lives in the whole universe. It does everything—it dances, it sings, it paints pictures, it makes objects. Then comes the point, in our culture at the age of eight or so, at which the family, the school, the whole of society say to a child that it is time to stop playing and to learn how to work. The child is put on a track and brainwashed with questions like "Why sing? You aren't going to be a singer, are you?" and "Why paint a picture? You aren't going to be an artist, are you?" Putting a child on a track is satisfying because it implies the child is going somewhere—there are stations and a schedule. But the track prevents the child from going anywhere else. Einstein and Bohr, the greatest persons I have ever known, were also the most childlike in the sense of being eager to explore just everything. Something terribly traumatic has happened to all of us, as evidenced by our lack of memory of early childhood. Very few of us can remember much more than occasional snapshots of our lives before school age. At a conference in India on consciousness, the first I was ever exposed to, there was much talk about superconsciousness, the idea of using more than what is said to be a small fraction of our brains, and about reforming education to foster superconsciousness. When my turn came to speak, I said that I thought what they were reaching for lay not ahead of them but behind them—in their childhoods. ■