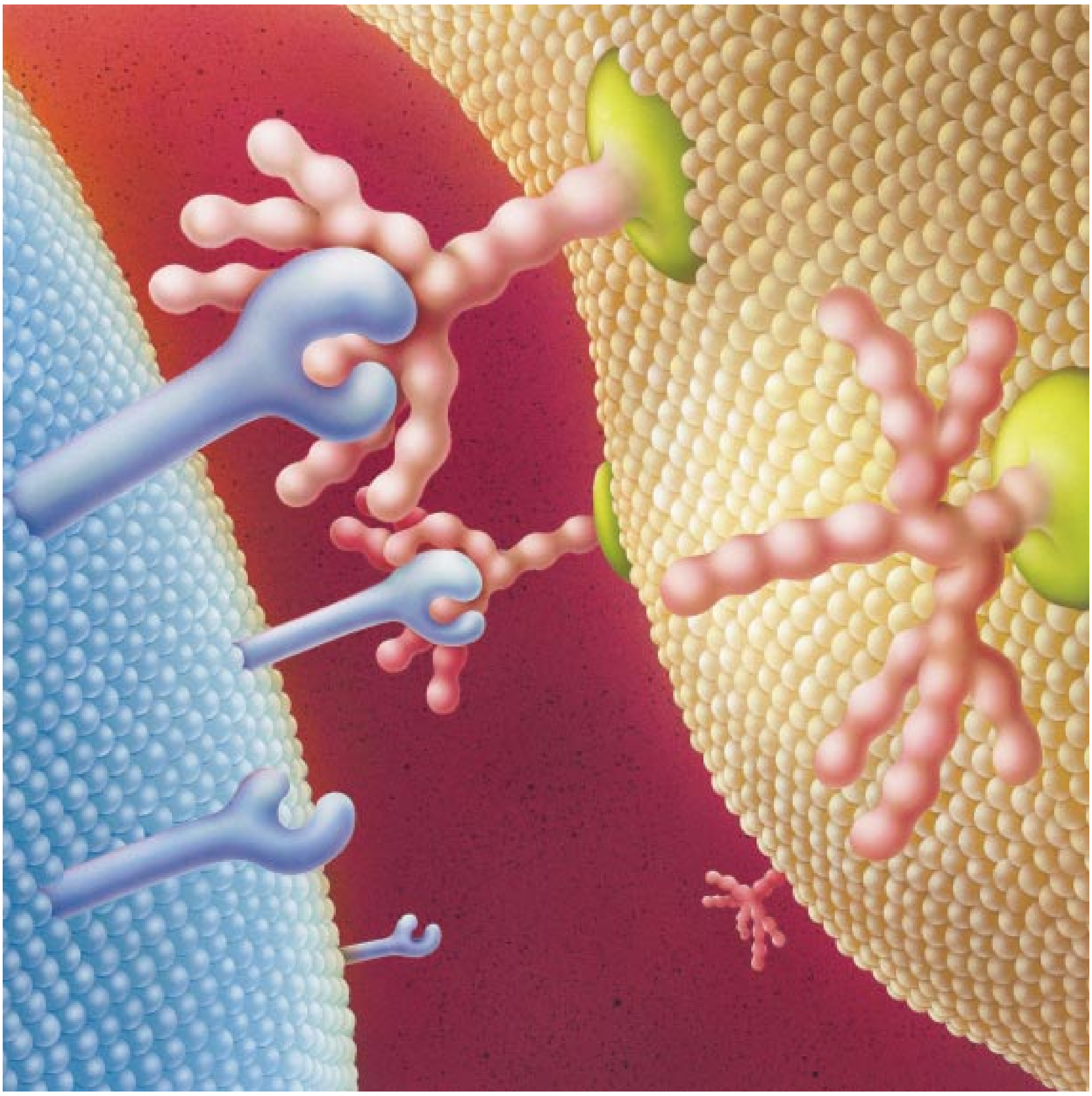
# SCIENTIFIC AMERICAN

JANUARY 1993
$3.95

*The turbulent birth of the Milky Way.*

*Lemurs: a glimpse at our evolutionary past.*

*From quantum dots to designer atoms.*



***Sticky sugars:*** *carbohydrates mediate many cellular interactions, such as infection and inflammation.*

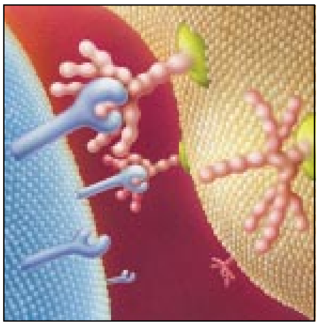
SCIENTIFIC AMERICAN

January 1993 Volume 268 Number 1

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1993 by Scientific American, Inc. All rights reserved. Printed in the U.S.A. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Second-class postage paid at New York, N.Y., and at additional mailing offices. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.



# DEPARTMENTS



THE COVER painting depicts selective adhesion between two cells. This attachment is mediated by the carbohydrates in a branching molecule (*pink*) that extends from an endothelial cell. A complementary molecule on a lymphocyte called an L-selectin (*blue*) binds specifically to a subunit in the carbohydrate, thereby tethering the cells together. Carbohydrates determine many interactions between cells, including infection (see "Carbohydrates in Cell Recognition," by Nathan Sharon and Halina Lis, page 82).

## THE ILLUSTRATIONS

Cover painting by Tomo Narashima

| Page | Source |
|---|---|
| 64–65 | Larry Lipsky/Bruce Coleman, Inc. |
| 66–67 | Joe LeMonnier (*top*), Jana Brenning (*bottom*) |
| 68 | Jana Brenning |
| 69 | Alan E. Strong, U.S. Naval Academy |
| 70 | Barbara E. Brown |
| 72–73 | Alfred Kamajian |
| 74 | Michael J. Bolte, Lick Observatory; Johnny Johnson |
| 75 | Johnny Johnson (*top*), Canada-France-Hawaii Telescope Corporation (*bottom*) |
| 76 | Sidney van den Bergh (*top*), Johnny Johnson (*bottom*) |
| 77 | Alfred Kamajian (*top*), Anglo-Australian Telescope Board (*bottom*) |
| 78 | Anglo-Australian Telescope Board |
| 82–83 | Tomo Narashima |
| 84 | Jared Schneidman/JSD |
| 85 | Courtesy of Kazuhiko Fujita, Juntendo University School of Medicine, Tokyo |
| 86 | Kathleen Katims/JSD |
| 87 | Courtesy of Steven Rosen, University of California, San Francisco |
| 88 | Kathleen Katims/JSD (*top*), courtesy of Kazuhiko Fujita, Juntendo University School of Medicine, Tokyo (*bottom*) |
| 89 | Photo Researchers, Inc. |
| 90–91 | Wayne Fields |
| 92 | Johnny Johnson |
| 93 | Wayne Fields (*left*), Samuel A. Bowring (*right*) |
| 94–96 | Ian Worpole |
| 110 | Joe LeMonnier |
| 111 | David Haring, Duke University Primate Center |
| 112 | Frans Lanting/Minden Pictures (*top* and *middle*), David Haring (*bottom*) |
| 113 | David Haring (*top left* and *bottom*), Frans Lanting (*top right* and *middle*) |
| 114–115 | Patricia J. Wynne |
| 116 | Joe LeMonnier |
| 117 | Frans Lanting (*left* and *right*) |
| 119 | Robert Prochnow |
| 120–121 | Michael Goodman |
| 122 | Mark A. Reed |
| 123 | Daniel E. Prober, Yale University |
| 124–125 | Courtesy of Peter M. Milner |
| 126 | Gabor Kiss |
| 127 | Eric Mose |
| 128–129 | Courtesy of Samuel M. Feldman |
| 130–131 | *left to right:* Robert A. Blanchette, University of Minnesota; Todd A. Burnes and Robert A. Blanchette, University of Minnesota; M. P. Kahl, Bruce Coleman, Inc.; R. J. Erwin, Photo Researchers, Inc.; Murrae Haynes |
| 132–133 | Ian Worpole |
| 134 | Jason Goltz |
| 135 | Murrae Haynes |
| 138 | Andrew Freeberg |
| 140 | Murrae Haynes |
| 150 | Andrew Christie |
| 151–152 | Johnny Johnson |


# SCIENTIFIC AMERICAN®

*Established 1845*





**SCIENTIFIC AMERICAN, INC.**
415 Madison Avenue
New York, NY 10017
(212) 754-0550

# LETTERS TO THE EDITORS

## Thinking about Mind

In your special issue on "Mind and Brain" [SCIENTIFIC AMERICAN, September 1992], each article provided more food for thought than my neural networks could process in a score of readings. A veritable mental feast!

Indeed, if the issue was a banquet, then Jonathan Miller's stimulating essay "Trouble in Mind" was a fine brandy at the end of a good meal. To me, Miller has always embodied the questioning mind turned inward on itself.

THOMAS SALES
Somerset, N.J.

The splendid "Mind and Brain" issue seems to end on an unduly negative note. Miller forecasts that we will never fully understand the connection between brain and consciousness. That assumption appears to overlook that consciousness is routinely interrupted by general anesthetics. The loss of consciousness under anesthesia and the later recovery of it can, in principle, surely be elucidated as thoroughly as any other drug-induced changes.

B. RAYMOND FINK
Department of Anesthesiology
University of Washington School
of Medicine

## Sex on the Brain

In her otherwise well-balanced review ["Sex Differences in the Brain," SCIENTIFIC AMERICAN, September 1992], Doreen Kimura perpetuates some longstanding myths. The traditional view that the female pattern of neural organization occurs by default from the lack of exposure to masculinizing levels of testosterone and estradiol should finally be put to rest.

Considerable evidence has accumulated during the past 20 years that feminization of neural structure and function is an active process. Numerous studies in rodents have demonstrated that feminization depends on levels of estrogen that are too low to elicit masculinization. My own studies have shown that in the absence of testosterone, the removal of endogenous estrogen dramatically reduces the outgrowth of neuronal processes.

The dogma that estrogen-binding plasma proteins, such as alpha-fetoprotein (AFP), "protect" the female brain from masculinization is erroneous. In rodents, AFP is far more likely to act as a reservoir for estrogen, which may be used to initiate the growth of axons and dendrites. Estrogen may therefore regulate sexual differentiation in both male and female brains.

C. DOMINIQUE TORAN-ALLERAND
Department of Anatomy
and Cell Biology
Columbia University

Kimura contends that many of the skill differences between men and women are mediated by brain organization. Yet two of her examples can be explained by simple physical distinctions. Some experiments have shown that performance differences that favor women in pegboard tasks disappear when the larger finger size of a man is factored out.

Men are reported to be better than women at dart throwing and other target-directed motor skills. It has been consistently demonstrated that both timing and spatial errors decrease in ballistic motor tasks as force approaches maximum. The greater strength of men should grant them an advantage in such tasks. Perhaps sex differences in ballistic motor tasks found in prepubertal children, where strength is similar between the sexes, are influenced by socialization.

JOHN S. RAGLIN
Department of Kinesiology
Indiana University

***Kimura replies:***

Toran-Allerand makes a valid point, which for brevity I had to omit from my article. Moreover, it is still problematic whether the evidence for such an influence on the organization of reproductive behavior is compelling.

Raglin suggests that the sex differences in motor behavior are reducible to physical differences. Even if the physical differences were decisive, one would expect neural parallels to them. The strength differences among three-year-old children are minimal, yet boys have shown superior accuracy in a targeting task. Other data also demonstrate performance differences between the sexes and between homosexual and heterosexual men that cannot be attributed to differences in size. It seems reasonable to conclude that over and above considerations of size, speed and strength, women's brains are endowed with better digital control and that men's brains are better endowed for targeting external stimuli.

## Genes on the Menu

The first half of Deborah Erickson's article "Hot Potato" ["Science and Business," SCIENTIFIC AMERICAN, September 1992], about new biotech-derived food, is overly negative.

Yeasts have been used to brew beer for 8,000 years, and farmers were crossbreeding livestock long before Gregor Mendel and his experiments. For decades, genes have been transferred from one species to another and even from one genus to another. These "genetically engineered" plants are the very same oats, rice, currants, potatoes, tomatoes, wheat and corn that we now buy at the local supermarket or farm stand. The techniques of "new biotechnology" speed up the process and target with greater precision the kinds of genetic improvement we have long conducted with other methods.

Contrary to the assertions of the neo-Luddites, the recently announced policy of the Food and Drug Administration for the regulation of new plant varieties is based on solid scientific principles. The bottom line is that the FDA will not tolerate unsafe foods, and our policy reflects this commitment.

HENRY I. MILLER
Director, Office of Biotechnology
Food and Drug Administration

ERRATUM

The graph on page 122 of the November 1992 issue illustrates the budget of Sematech between 1988 (not 1982) and 1992.

*Because of the volume of mail, letters to the editor cannot be acknowledged. Letters selected for publication may be edited for length and clarity.*

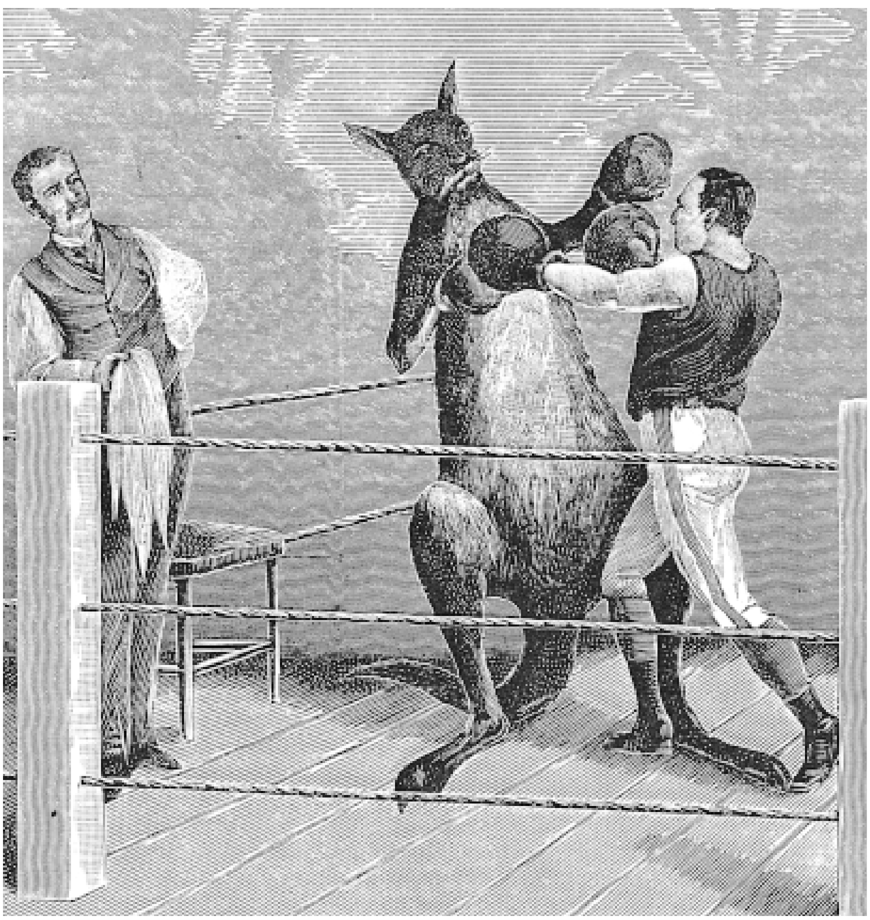
# 50 AND 100 YEARS AGO

JANUARY 1943

"Formerly, if an enemy submarine lay quietly on the bottom of the sea to avoid detection, the business of 'putting the finger' on a sub became more difficult and less accurate in its results. In the present conflict, the principle of sound reflection under water, long applied to larger merchant and war ships to maintain a continuous graphical record of the ocean's floor beneath the cruising ship, is being adapted to search out silent submersibles that endeavor to 'play possum' far beneath the waves. The exact extent to which echo-sounding devices are utilized and their scientific and mechanical constituency are among those things which cannot now be told."

"In a degenerate mass of gas, when the velocities of the moving electrons begin to become comparable with that of light, the law connecting pressure and density changes. Chandrasekhar has shown that, when this is taken into account, a star of small mass (less than twice the Sun's) will settle down into a permanent state with a degenerate core, as a white dwarf, and finally as a 'black dwarf,' cold on the surface; but a large mass (ten times the Sun's or more) should continue to contract without limit. It is natural to suppose that something would ultimately happen to end this process, and it may well be that the contracting star blows up, ejects enough matter to leave a residue small enough to form a degenerate core, and then develops successively into a blue, a white, and a black dwarf. At the Paris Conference of 1939, Chandrasekhar suggested that some catastrophic change of this sort might be responsible for a super-nova."

"The requirements for carotene (provitamin A), ascorbic acid (vitamin C), and iron can readily be met by eating moderate quantities of dried grass. In the case of calcium and the vitamin B complex factors, between four and six ounces need be eaten, amounts so large as to be undertaken only by an enthusiast. Undoubtedly the wisest and safest recommendation is to use dried grass, if at all, in small amounts and finely ground, either as an added ingredient in common foods such as bread, or as a supplement to the diet in the form of tablets, which should be prescribed only on advice of a physician."



***The kangaroo as a prizefighter***

JANUARY 1893

"To set off this piece of fireworks it is not necessary to be a pyrotechnist. Provide yourself simply with a blowpipe or even a clay tobacco pipe. Take a few sheets of thin tinfoil, such as is used as a wrapping for chocolate, and cut them into strips of a width of about an inch. Then present each slip to the flame of the blowpipe, when the metal will ignite and fall in incandescent globules, which will rebound and run over the table on which you operate and travel a considerable distance. When the flame is strong and the tinfoil burns briskly, the globules are very abundant and then present the aspect of a bouquet of fireworks in miniature. By such combination of a metal with the oxygen of the air, the tinfoil is converted into a white oxide. It was by studying the increase in weight exhibited by tin heated in contact with air that John Rey, a chemist of the seventeenth century, succeeded in understanding the fixation of the air upon metals.—*La Nature*."

"The way in which the natural kangaroo spars in the bush, his birthplace, is peculiar. He places his front paws gently—almost lovingly—upon the shoulders of his antagonist, and then proceeds to disembowel him with a sudden and energetic movement of one of his hind feet. From this ingenious method of practicing the noble art of self-defense the kangaroo at the Royal Aquarium has been weaned. The clever instructor of this ingenious marsupial has trained it to conduct a contest under the conditions known as the Marquis of Queensberry's rules."

## How Many Genes and Y

*Gene mappers find plenty, even in "junk" chromosomes*

When the idea of mapping and sequencing all the genes that make up a human being was first proposed, it seemed an undertaking tantamount to putting a man on the moon. The massive international effort was expected by some to continue for 15 years or more. But after only two years, the Human Genome Project is proceeding more rapidly than most biologists had dared predict. "We are two or three years ahead of schedule," says Daniel Cohen of the Center for the Study of Human Polymorphism (CEPH) in Paris. "I believe it will be possible to have a very good map of the genome by the end of 1993. Probably the sequence of the genome will be finished by the end of the century."

Sketchy though they are, the latest genetic road maps are already obliging geneticists to reappraise their theories about the functions of some human chromosomes. Meanwhile parallel work on simpler organisms, such as the much studied roundworm *Caenorhabditis elegans,* is revealing that they have unexpectedly large numbers of genes. As a result, some investigators are speculating that the human genome may turn out to be far larger than the 100,000 or so genes it is believed to contain.

Norton D. Zinder of the Rockefeller University, a former co-leader of the project who now advises the National Institutes of Health on its effort, believes many of the recent discoveries could not have been made without a comprehensive gene-sequencing effort. "There are real data coming in, and it proves that we are going in the direction we should be," he says.

The genome project involves developing three increasingly detailed maps of the DNA in cells. The first is a genetic linkage map, which shows the relative distances between markers on a chromosome. The second is a physical map, which locates similar genetic landmarks but specifies the actual number of nucleotide bases, or DNA subunits, between them. The ultimate map is the ordered sequence of bases in a chromosome that describes the genes and the proteins they make.


STANLEY ROWIN

***DRAWING A MAP of the Y chromosome was the task undertaken by Adrienne Hilton and her molecular geneticist colleagues at the Whitehead Institute at M.I.T.***

In early October, through a colossal combined effort by the NIH and CEPH, genetic linkage maps for 23 of the 24 types of human chromosomes were compiled and published. Simultaneously, physical maps for two of the chromosomes were released: chromosome 21, which was mapped by Cohen and his colleagues, and the Y chromosome—for which there was not a linkage map—by David C. Page, Simon Foote, Douglas Vollrath and Adrienne Hilton of the Whitehead Institute at the Massachusetts Institute of Technology.

Cohen and Page both used essentially the same techniques to map the chromosomes. Through a process called sequence-tagged site mapping, they established the order of small marker sequences on the chromosomes. They then chopped the chromosomes into pieces of DNA about a million bases long and spliced the pieces into yeast DNA to produce artificial chromosomes, which could be measured conveniently. By looking for the markers on the artificial chromosomes, the researchers deduced how to fit them together, like pieces of a puzzle. Some segments of the human chromosomes are missing from these maps, but they are not believed to contain any genes.

Because chromosome 21 has been associated with Down's syndrome, some forms of Alzheimer's disease and other disorders, the clearer picture of its genetic contents is expected to have great medical relevance. In the short run, however, the Y chromosome may benefit most from the new map because it is the least typical of the human chromosomes and in many ways the least understood. "We're trying to make this chromosome respectable," Page says.

The Y chromosome, according to Page, has often been regarded as "basically a junkyard" containing no more than a few genes related to spermatogenesis and other functions peculiar to males. "Many people refer to the Y as a maleness chromosome," he says. "I think that is much too narrow a cubbyhole to fit this chromosome into."

One piece of evidence on his side is the discovery by his mapping team that 25 percent of the studied Y regions are homologous, or highly similar, to parts of the X chromosome. On the X, several genes essential to both sexes are found in these areas. Other studies have also found similarities in gene sequence on the two chromosomes. "I'm sure that's just the tip of the iceberg," Page adds enthusiastically.

One important implication of those similarities is that a classic tenet of genetics—that males have only one copy of all the genes on the X—is wrong. Consequently, Page argues, the work on the Y chromosome sequence "forces us to rethink not only the functions of the Y but also of the X." How important the genes on the Y chromosome are remains to be seen, but Page contends that "history is on the side of people predicting an ever widening array of functions for this chromosome."

The final stage of the project, sequencing the individual genes, has not yet begun. But related efforts in other species are well under way. In the spring of 1992 Robert Waterston of the Washington University School of Medicine and John Sulston of the Medical Research Council in Cambridge, England, and their colleagues published the sequence of more than 120,000 bases in the DNA of *C. elegans.* That represented only a tenth of a percent of the total genome, but the pace of sequencing is accelerating: Waterston reports that they have now sequenced about one million bases and expect to finish another two million bases within a year.

Meanwhile a European consortium of 145 scientists has been sequencing chromosomes of the common yeast *Saccharomyces cerevisiae.* Last May the group published the complete sequence of chromosome III. According to Stephen G. Oliver of the University of Manchester Institute for Science and Technology, who served as DNA coordinator on the project, yeast chromosome XI is now about two thirds finished, and chromosome II is about half done; extensive work has also been done on chromosomes I and VI.

Perhaps the most surprising observation about the newly sequenced genes is

## Endangered Genes

Can you name the male and female leads of the *Human Genome Project?* They star in *Gray's Anatomy* and have white skin, urban homes and composite ancestry. Still can't place them? They are John and Jane Doe.

So much for ethnic diversity. The ethnocentric bias of the genome project has riled an international group of anthropologists who hope a more extensive catalogue of human genes will allow them to reconstruct human evolution. For the past two years, they have been planning a parallel initiative called the Human Genome Diversity Project. Their goal is to sample the genes of aboriginal peoples before these peoples die out or assimilate.

A quick survey of the most endangered groups should take about five years and cost about $23 million, says Luigi L. Cavalli-Sforza of Stanford University. Those who map the genes of John and Jane Doe will never miss that paltry sum, although they may gain a substantial return on the investment. If the sample turns up genetic adaptations to disease, for example, workers may use the knowledge to develop new therapies.

The project began two years ago, when five geneticists published a manifesto challenging the ethnocentricity of the genome project in the journal *Genomics.* Others quickly jumped on the bandwagon because two of the authors commanded such respect in the field's main camps: those who study populations and those who study individual gene lineages. The first approach was championed by Cavalli-Sforza, the second by the late Allan C. Wilson of the University of California at Berkeley.

Proponents of the two approaches worked out their differences at a workshop held last summer at Stanford. Cavalli-Sforza argued for intensive sampling, the only way to get the statistical depth he needs to look at gene frequencies in different populations. But to obtain enough specimens in each sample, Cavalli-Sforza conceded, he would have to make do with relatively few samples. He therefore wanted to study ethnic groups whose linguistic distinctiveness suggests they are of ancient descent.

Wilson's disciples favored a more extensive survey. Because they study the lineages of individual genes, they could broaden the coverage at the expense of sample size. In their most controversial work, they surveyed a few hundred individuals to build a genealogy tracing all humans to an African matriarch who lived some 200,000 years ago.

The two schools clashed on a practical matter as well. Cavalli-Sforza's group wanted to preserve specimens by immortalizing cells, a procedure that requires rushing fresh blood to the laboratory before the white cells die. Wilson wanted to facilitate a broad survey by letting ethnographers put the blood on ice, so that they could go on collecting for weeks. They could then deposit their trove in repositories from which future generations could draw repeatedly, using the new techniques of DNA amplification.

The workshop compromised: ethnographers would concentrate on distinct ethnic groups, as Cavalli-Sforza wanted, but they would spread their resources over a greater number of groups, as Wilson's team wanted. They also agreed to immortalize only a fraction of the specimens. They projected a sample of about 400 groups.

A second workshop chose the groups at Pennsylvania State University over the Halloween weekend. Anthropologists, linguists and geneticists divided into teams specializing in each region save Europe, which has its own project under way. Eyes glazed as specialists struggled to fill out forms assigning priorities to tribes and pointing out problems ethnographers might face. Watch out for guerrillas and coca smugglers, said the South America group. Survey the hundreds of Polynesian populations at a few central labor exchanges, suggested the Pacific group. Refuse to report HIV-positive cases to governments on grounds of medical confidentiality, counseled the Africa group.

All were concerned about the language they—and reporters—might use to describe their work. "You can talk of 'tribes' in Africa but not in this country," said one participant. Others worried that labeling a group as "endangered" would offend the majority group in their country.

The third workshop, to be held in Washington early this year, and the fourth, to be held in Sardinia next fall, will discuss the logistics of reaching all points on the globe, the techniques for collecting and analyzing materials and the ethical problems in exploiting native peoples for their genes. Some groups find anthropomorphic sampling so repugnant that they refuse access to the dead as well as to the living. —*Philip E. Ross* their staggering number. If the sequence analyzed by Sulston and Waterston is representative, *C. elegans* may have 15,000 genes—three times more than was once believed. Researchers had thought yeast chromosome III contained only about 34 genes, but the Europeans found evidence for 182.

Most of the genes seem to have escaped detection previously because mutations in them did not have noticeable effects. Some biologists have therefore speculated that many of the genes are redundant or unnecessary. That notion has its critics, however. As Page asserts, "We don't have any idea of how many genes it ought to take to perform functions." He points out that nobody has yet shown what happens if combinations of these seemingly redundant genes are knocked out. "How deep is the redundancy?" he asks.

Oliver suggests that the seemingly redundant genes may be important only during brief periods of an organism's life—and possibly not at all under standard laboratory conditions. "One may need to make the organism jump through rather specific physiological hoops before only gene X and not gene X′ will work," he says.

When researchers sequence the human genome, it is difficult to predict whether they will find more than the 100,000 genes they now expect. Zinder, for one, thinks they may. Cohen, who believes the current figure is roughly correct, says it is far harder to recognize genes in humans because human genes are more extensively subdivided and separated than are those of yeast and roundworms. Page maintains that gene estimates in all organisms have been creeping up for years. Just a few years ago, he notes, most geneticists claimed that humans had only 30,000 genes.

Investigators are also discovering that many of the genes sequenced so far in *C. elegans* and yeast are extremely similar to ones found in other organisms, from mammals to bacteria. Waterston believes the strong similarities between roundworm enzymes and mammalian enzymes show they are serving almost the same function. Nevertheless, he adds, "whether they're working on the same substrate or not is another matter."

Some of the shared genes are incongruous: yeast, for example, carries a gene for a protein that enables bacteria to fix nitrogen into biological compounds, even though yeast does not have that ability. To Oliver, the presence of that gene in yeast suggests "we don't understand in any deep way the function of the protein in the nitrogen-fixing bacteria." It is likely, he thinks, that all organisms use that gene somehow; its application to the fixation of nitrogen is just particularly noticeable.

Such a gene may therefore turn up in humans as well, once the sequencers are ready. That time may come soon, because the mapping stage may not last much longer. While Page and others are continuing to make physical maps of individual chromosomes, Cohen is boldly pursuing a complementary approach: he is mapping all the chromosomes at once. "We believe the approach used for chromosomes 21 and Y is far too tedious and expensive," he explains.

Instead of using the sequence-tagged site markers, Cohen's group is using older DNA "fingerprinting" technology to flag distinctive sequences on all the chromosomes simultaneously. Cohen reports that they have currently mapped about 70 percent of the entire human genome this way, and he expects to have the complete genome mapped at low resolution by this February. That physical map will have fairly little detail but can serve as a "backbone," Cohen says, for further physical maps based on the new technology. "With the two, you get a synergy," he remarks.

That synergy should greatly speed up the mapping. But because no one has experience with sequencing large chunks of human DNA, Page is more cautious than Cohen about projecting an end to the genome project by 2001. "I don't think that's unreasonable, but mostly I don't think it's unreasonable because it's still eight years away—I mean, in eight years, we could do almost anything, right?" —*John Rennie*

## Pitohui!

*The colorful bird looks better than it tastes*

The pitohui sounds like what it is: something to spit out. The skin, feathers and some organs of this orange-and-black New Guinean bird contain a potent poison. Although other species—including certain snakes, insects and frogs—were known to produce toxins as deterrents, it was generally thought that birds did not. So the discovery of the same device in the pitohui has ruffled some notions of avian defensive strategies and coloration.

The pitohui's defense mechanism was noticed two years ago by John P. Dumbacher, a graduate student in ecology and evolution at the University of Chicago. He felt numbness and burning in his mouth when he licked his hands after handling the hooded pitohui, referred to in New Guinea as a "rubbish" bird because of the taste of its skin. Dumbacher and his colleagues recently reported in *Science* that three species of the genus *Pitohui*—the hooded, the variable and the rusty—produce a noxious chemical, which they identified in 1992.

The poison, homobatrachotoxin, turns out to be identical to that of a South American poison-dart frog, which also has aposematic, or warning, coloring of orange and black. "I was very surprised," says John W. Daly, a chemist at the National Institutes of Health who analyzed the frog toxin in the 1970s and that of the pitohuis last year. "There certainly has been a specific evolutionary ability to accumulate this toxin. I would like to say 'make' it, but we do not know if it is from the diet."

Although the pitohui is the first poisonous bird to be reported in the literature, there have been anecdotal reports of bad-tasting birds. Some experts anticipated the finding. "I am not at all surprised," comments Lincoln P. Brower, an ornithologist at the University of Florida, "especially given the fact that some insects are poisonous and that birds behave like those insects: they are conspicuous and brightly colored."

Conventional ornithological wisdom holds that bright plumage among birds exists to facilitate courtship and mating. But the colorful feathers of the pitohui could serve as a warning to predators. The fact that male and female pitohuis share the same palette reinforces this assumption. And just as there are nonpoisonous mimics of the poisonous monarch butterfly, there are nonpoisonous mimics of the pitohui.

Taste tests of birds conducted in the 1940s and 1950s support the possibility that color and palatableness are inversely linked. Hugh B. Cott, a zoologist at the University of Cambridge, observed that hornets in Africa avoided certain bird carcasses yet ate others. Cott then did the "hornet test" on a series of birds. The results encouraged him to conduct his own gourmet, double-blind trials. To carry them out, a colleague's wife prepared repasts of 200 bird species. After Cott's feasts, diners agreed with what might be called Cott's rule: the blander the bird looks, the better it tastes. Birds that had cryptic coloring—that is, those that blended in with the background—tasted best.

"Conversely, there is some evidence that numbers of highly conspicuous birds belonging to many different orders are definitely unfit for the table:

W. PECKOVER VIREO

***POISON PLUMAGE** recently discovered in the hooded pitohui suggests that other brightly colored birds may also use toxins to repel predators.*

sheld-duck, crocodile bird, magpie, and swallows being examples," Cott wrote in *Nature* in 1945. Nevertheless, Jared M. Diamond of the University of California at Los Angeles recently reported in *Nature* that his field assistants' meal of pitohui produced no "untoward effects."

Gustatory recommendations aside, the issue of coloration may be twofold: vivid markings could be selected for by natural or sexual selection, or both. "The two forces are not mutually exclusive," Brower suggests. Sexual selection could have favored brightly colored males and cryptic females at first. But if the birds then acquired a toxin from their diet and if it successfully protected against predators, natural selection for bright coloration in both sexes would occur.

In addition, Stanley A. Temple, a wildlife ecologist at the University of Wisconsin, has found a correlation between diets rich in fruit and aposematic coloring in birds. He suggests that such meals may allow the birds to sequester chemicals that could form toxins. (Cott did not relate birds' diets to their tastiness.) Temple says he, too, has found a poisonous bird, the pink pigeon of Mauritius, but has not yet published his research. The pigeon apparently derives a toxic alkaloid from its diet. "The species is probably in existence today because of its defense mechanism," Temple notes. "The dodo and others were exterminated."

Dumbacher and Daly and their colleagues plan to study the pitohui's potential predators to see if they are repelled by the toxin. They will also examine pitohuis to determine how they avoid poisoning themselves. Homobatrachotoxin works by opening up ion channels, causing cells to be infused with sodium. But "there are a number of creatures that are resistant to their own poison," Daly says. In such species, ion channels do not respond to their toxin. Daly expects to see the same kind of mechanism in pitohuis.

Daly also hopes to determine if pitohuis can synthesize the chemical or if their metabolism produces it as a byproduct. Poison-dart frogs living in captivity do not make the stuff of their deadly bolus, apparently because they lack something that their rain-forest diet of leaf litter normally supplies.

The search is also on for more poisoned plumage. Although Dumbacher says he will confine himself to pitohuis for now, he does admit a numbing, burning curiosity: "I've thought of licking other birds." *—Marguerite Holloway*

## Crunching Epsilon

*Cryptography may be the key to checking enormous proofs*

It began with a method for keeping spies honest and may end up verifying the notorious four-color map theorem. The technique, known as a holographic proof, makes it possible to achieve a high degree of confidence that a set of logical assertions (such as a theorem and the reasoning involved in its proof) is internally consistent by checking only a tiny fraction of the set's statements.

Testing a mere 300 lines of a 100,000-line proof could reduce the probability of an undetected error to less than one divided by the number of particles in the universe, asserts mathematician Leonid Levin of Boston University. Some mathematical proofs already run upward of 10,000 pages, and no one can possibly comprehend them in their entirety, much less certify their reasoning.

Furthermore, the same technique could in theory be used to check the output of complex computer programs. Rigorous proof that a program does what its designers intended is infeasible by means of conventional methods unless the program is only a few pages long. In addition, even if the software is correct, there is no guarantee that the hardware has not suffered a random glitch. A holographic check would test the program and its execution simultaneously.

The "zero knowledge" proof, which helped to set the stage for the holographic variety, was developed as part of cryptographic protocols for verifying facts without revealing them. Cryptographers have shown that one can answer a series of random mathematical queries about some hidden fact in such a way that its secret remains hidden, but anyone who does not actually have the knowledge has only an infinitesimal chance of answering the queries consistently.

In a holographic proof, instead of answering a series of random queries, the prover in effect writes down answers to all possible queries in a book, and the verifier samples the answers randomly, looking for inconsistencies, says Lance Fortnow, a mathematician at the University of Chicago. The effect is the same.

The key to the trick is a technique called arithmetization. According to Fortnow, once a proof has been stated in strict logical form, one turns it into a polynomial expression of many variables by (more or less) substituting addition operations for every "or" in the proof and multiplication for every "and." The holographic proof then consists of a series of equations giving both the polynomial and its value for different combinations of the values of its variables.

Checking is simply a matter of making sure the calculated value of the polynomial at any point matches that asserted in the proof. Only a small number of points need be checked, Levin explains, because it is very difficult to construct two polynomials of low degree that are equal at some points yet different at others.

Indeed, for a single-variable polynomial of degree 10 (of the form $Ax^{10} + Bx^9 + Cx^8\ldots$), a mere 11 values are sufficient to specify its shape precisely. As long as the proof does not contain too many logical "and" statements strung together, its polynomial degree will be low. The number of tests required to check it will be trivial compared with the size of the proof, Levin says.

Unfortunately, one reason the number of tests will be so small is that the method causes the proof itself to become enormous. First, instead of the shorthand mathematicians usually use, Levin notes, every step and every rule of inference must be rigorously defined. Mario Szegedy of AT&T Bell Laboratories estimated it might take 500 lines just to prove the Pythagorean theorem, in contrast to the dozen or so now considered adequate.

Worse yet, the arithmetized version of the proof will be even longer. If the formal version contains *N* lines, the arithmetized one will contain *K* times *N* raised to the power of one plus epsilon. *K* is very large, and the various mathematicians disagree on the size of epsilon. Szegedy pegs it conservatively near one, in which case the arithmetized version of a 10,000-line proof could run upward of 100 million lines. Fortnow marks epsilon at one half, Levin nearer one third. Indeed, Levin says, epsilon can be reduced as close to zero as desired but only at the cost of increasing *K*—perhaps to a point that would swamp any improvements in the exponent.

Bringing *K* and epsilon down to levels that might make the holographic technique practical—either for mathematical proofs or for checking computer programs—will take "a lot of hard work," Fortnow says. Indeed, he suggests, it will probably require the invention of one or two mathematical tricks for doing holographic transformation and perhaps the same number of fundamental insights. "It's not clear that it will be possible," he asserts.

Even in its present unwieldy form, however, the holographic proof technique has its uses. Szegedy and his colleague Carsten Lund and others have employed a variant of its principles to prove lower bounds on the difficulty of solving certain hard problems in computer science. And Szegedy is hoping to perform a holographic verification of the four-color theorem, whose brute-force computer enumeration of all relevant maps has troubled certainty-minded mathematicians for most of the two decades since it was completed. (To this day, no one knows if there might have been some mistake in the algorithm.)

Of course, for purists the idea of a proof that is only probably verified presents its own problem. But Szegedy is not one of those. "People are becoming satisfied with probabilistic methods," he says, noting that probabilistic techniques are also being used to factor large prime numbers. "You might believe that mathematics is part of nature or that it is what humans use to their advantage. If you take the practical rather than the idealistic approach, you save a lot of headaches." —*Paul Wallich*

## A Gene for Hypertension

When a greasy burger or a handful of salted peanuts sends someone's blood pressure soaring, there may be more to the clinical picture than the hazards of gobbling on the run. Research teams in the U.S. and France have found that for some people, a particular gene seems to increase the likelihood of acquiring a form of hypertension—specifically, the kind involved in salt retention. Although physicians have known for some time that heredity plays a strong role in the illness, the gene, one of perhaps several that are thought to be implicated, represents the first direct, supporting evidence.

The studies, conducted by the University of Utah's Howard Hughes Medical Center and the French research institute INSERM, looked at hypertensive siblings and compared them with unrelated people who had normal blood pressure. The researchers found that the hypertensive siblings (those with blood pressure exceeding 140 over 90) tended to have the same kind of variation in the gene that encodes a protein called angiotensinogen.

Angiotensinogen works in conjunction with renin, an enzyme produced by the kidneys, to form angiotensin, which raises blood pressure by constricting blood vessels and, perhaps more important, changing the body's balance of sodium and water. Jean-Marc Lalouel, who headed the Utah group, speculates that the gene variation could lead to a small increase in circulating angiotensinogen. By the time an individual reaches middle age, the overproduction makes the body sensitive to sodium. The retention of sodium causes the volume of blood to expand. To compensate for the additional fluid fed to the body's tissues, the arterioles constrict. As a result, blood pressure rises. "A small increase in angiotensinogen may act over the long term," he says.

Like many discoveries concerning the genetics of disease, the ability to identify the angiotensinogen gene would be useful in screening for susceptible individuals. The new finding could be especially relevant to black Americans, in whom the condition develops two to six times more frequently than in white Americans, according to various estimates. Furthermore, most hypertensive American blacks have the salt-sensitive form of the disease. Preliminary findings from other studies indicate that such individuals also display the related variations in the angiotensinogen gene.

The finding could breathe new life into a hypothesis offered a few years ago by Clarence Grim of the Charles R. Drew University of Medicine and Science in Los Angeles. Grim has proposed a highly controversial idea of why black Americans have high blood pressures. Grim noted that indigenous populations in sub-Saharan Africa have a very low incidence of hypertension, unlike those in the Western Hemisphere.

Grim used historical accounts of the slave trade to explain the difference. The cause of death for most slaves was diarrhea, and the sweating, vomiting and lack of drinking water on the grueling trans-Atlantic journey in the poorly ventilated cargo holds contributed to dehydration. Those who could retain water—with body salt—survived. "Death during the slave trade may well have focused on the ability to conserve salt," he says.

According to Grim, the death rates, which ranged from 30 to 50 percent, were sufficient to select for a gene. He cites several studies of black populations that, while not expressly proving the theory, are at least consistent with it. Given a candidate gene, he will look for variations in African and American populations.

Although Lalouel thinks "clearly there is something that has to do with salt handling in blacks," he does not buy Grim's hypothesis. He believes most of the observed blood pressure differences seen in African and American populations arise from environmental factors. The food of most Africans living in aboriginal conditions, he points out, is low in sodium; he predicts that changing their diets to resemble those of the more industrialized nations would raise their blood pressures. Avoiding fast foods is still good advice for anyone. —*Philip Yam*

## Anything Goes

*Why two sexes are better than 13*

Suppose a United Nations peacekeeping force disarms one of two opposing sides in an international hot spot and then directs the still battle-ready adversary to annihilate its now helpless enemies. Although this may be a draconian solution, a total war in which both sides could potentially be wiped out is avoided.

University of Oxford researcher Laurence D. Hurst offers this metaphor to answer a question that has intrigued geneticists for many decades. The gametes from sexually reproducing organisms fuse into a cell that contains one set of chromosomes from the nucleus of each parent. Yet there are other parties to the event: the strands or bundles of extranuclear DNA carried in the mitochondria of animals or the chloroplasts and mitochondria of plants. Unless one set of this material is eliminated, the growing daughter cell will be imperiled by withering exchanges of enzymes and other chemical ordnance as the two sides struggle for dominance.

How is the prospect of conflict to be prevented? Hurst thinks the nuclear DNA plays the role of the U.N. force. It eliminates one set of contestants. This, Hurst believes, may go a long way toward explaining why there are separate sexes and why it is often so difficult to find a date on a Saturday night. Hurst contends that it is easier to keep the cytoplasmic peace when there are only two sexes. Imagine, he says, sorting out such a conflict if any pairing among four, eight or 13 sexes could produce a new individual.

Hurst is a leading proponent of a school of biologists who conceive of evolution as more than just a competition to determine which organism adapts best to its environment, the classic Darwinian interpretation. Rather these advocates of what is called intragenomic conflict assert that much of evolutionary history may be explained by a kind of genetic Hegelian dialectic in which groups of genes within an organism engage in a constant game of one-upmanship with each other.

Early advocates of these theories, such as William D. Hamilton, a professor of evolutionary biology at Oxford, believe intracellular conflicts—and the way they get resolved—may help answer other major evolutionary questions. One of these conundrums may be the very beginnings of sex itself.

In a recent article in the *Proceedings of the Royal Society of London,* Hurst and Hamilton (who was Hurst's former graduate adviser at Oxford) focus on the war zone of the cytoplasm. Whereas the nuclear genes have created a comfortable demarche, the mixing of cytoplasmic genes when gametes come together is potentially a microscopic Bosnia—"a tragedy of the common cytoplasm," Hurst calls it.

A war in which mitochondria from different gametes battle to the death endangers the very existence of the combined cell, or zygote. Successful sex requires that the DNA in the nucleus find some means of suppressing conflict in order to preserve the organism. The ultimate evolutionary winners are cellular unions in which the nuclear command post in one of the two gametes issues strict orders on the fate of mitochondria or chloroplasts.

A sperm, for example, sheds its mitochondria before entering the egg. Mushrooms avoid potentially lethal cytoplasmic skirmishes altogether by forgoing cell fusion—which, Hurst says, distinguishes them from organisms that have developed discrete sexes.

Instead of merging two gametes, they simply transfer nuclei to one another through the process of conjugation, creating a small opening between the two cells. Thus, they can mate with any other member of their species, except a few individuals that have incompatibility markers, indicating that they are genetically too close.

To provide evidence for the existence of cytoplasmic dueling, Hurst and Hamilton sorted through the literature and identified a strange miscellany of ciliates, fungi and slime molds. In *Chlamydomonas* algae, for example, one chloroplast is inherited from the female and one from the male—or more precisely a "+" and "–" mating type, since this alga has not developed the size differences (a small sperm and a large egg) that are characteristic of other sex cells. In the merged cells the two cytoplasms begin to attack each other with deadly enzymes. But the + type, roughly analogous to the female, gets eaten away more slowly than does the – type (the quasi-male) and so prevails.

Hurst's prize organism is a primitive slime mold that a group of Japanese researchers reported on in 1987 in the *Journal of General Microbiology.* The slime mold, *Physarum polycephalum,* appears to have 13 sexes, each of which can mate, or permanently fuse cells, with any other sex except its own. If humans had 13 sexes, and a person could mate with anyone but one's own sex, there would be no more lonely nights. "It would be gorgeous beyond belief—the first person you meet you could mate with," Hurst says.

But, Hurst notes, the slime mold pays an intracellular price for its bliss. To curb cytoplasmic conflict, the slime molds have a rigid hierarchy that sets out which sexes can inherit the mitochondria of others. Moreover, this elaborate bookkeeping system is subject to cheating. In his own journeys through the literature, Hurst found documentation of a renegade mitochondrion that refused to respect the pecking order. Avoiding this complexity is probably why evolution favors two sexes, the Oxford researchers argue. Indeed, the slime molds may find themselves hard-pressed in centuries to come as sexual chaos reigns. "Multiple sex types might be expected to collapse to binary types," they write.

Although amused, not everyone is convinced of this explanation. "It's what I would call advocacy science," comments Brian Charlesworth, a researcher in population genetics at the University of Chicago, who has proposed a different theory for how cytoplasmic genes are inherited. "You try to make a model and then find something later that supports it." Charlesworth also charges that Hurst and Hamilton fail to explain adequately in their paper the origins of sexual dimorphism: why the male and female sex cells take on different sizes and shapes, something that Hurst asserts is a distinct evolutionary issue.

Charlesworth also raises questions about whether Hurst and Hamilton's several dozen citations constitute a broad enough inspection of the literature to make sweeping claims about the evolution of separate sexes. "Neither of these guys works in experimental genetics, which is involved with these phenomena," Charlesworth remarks. "They are very much armchair theorists. It is difficult to evaluate these data without hands-on experience."

Hurst acknowledges the need for experiments to confirm his ideas. He is working with Rolf Hoekstra, a biologist at the Agricultural University of Wageningen in the Netherlands, who supplied the evolutionary model used by Hurst and Hamilton in making their predictions. Hoekstra is trying to determine whether a highly inbred species of fungus, *Aspergillus nidulans,* whose cytoplasmic genes would be identical from cell to cell, has any need for separate sexes. That work may provide a clue as to why opposites attract—or pluses and minuses, if you prefer. —*Gary Stix*

JASON GOLTZ

***LAWRENCE KRAUSS of Yale University has abandoned the telescope in favor of the particle accelerator to search for the missing mass in the universe.***

## MACHOs or WIMPs?

*Astronomers stalk the invisible cosmic majority*

Almost 60 years ago pioneering cosmologist Fritz Zwicky made the shocking claim that much of the mass of the universe is "missing." Within the past decade, improved observations have transformed Zwicky's assertion into the accepted wisdom. The rate at which galaxies rotate and the manner in which they sail about in clusters and superclusters indicate that as much as 99 percent of the cosmos consists of an invisible component, known as dark matter. Theorists have identified three major possible dark components of the universe: MACHOs, WIMPs and the more prosaic neutrinos. A bevy of new experiments using computerized telescopes, particle accelerators and neutrino detectors may finally pin down the true nature of this mysterious matter. As Kim Griest of the University of California at Berkeley puts it: "Dark matter's time has come."

One of the most fruitful places to search for dark matter is in the outer halos of galaxies. Studies of disk galaxies show that their outer regions rotate much faster than would be expected from just the visible stars and gas they contain. Large amounts of unseen matter must be present to create an extra gravitational tug.

Many astronomers have speculated that the mass in the outer parts of galaxies may be hidden in such nonluminous bodies as free-ranging planets, burned-out stars, brown dwarfs (starlike objects too small to shine) and black holes. Griest has whimsically coined the term "MACHOs"—massive compact halo objects—for this class of dark matter candidates. Charles Alcock of Lawrence Livermore National Laboratory, working with Griest and several others, has recently embarked on an ambitious search for MACHOs, as have groups of French and Polish astronomers.

MACHOs cannot be perceived directly, but if one were to pass between the earth and a more distant star, its gravity would slightly bend and amplify the star's light. Alcock and his collaborators are monitoring three million stars in the Magellanic Clouds for telltale signs of previously unperceived cosmic vagrants. The rate at which a star's brightness changes would reveal the intervening object's mass. A Jupiter-mass body would cause a star to brighten and dim over the course of a few days, whereas events associated with black holes could last well over a year.

Alcock's MACHO investigation will run for four years, but positive results could show up much sooner. "If we find MACHOs, we will have solved the observationally secure dark matter problem," Griest says. If, on the other hand, the various surveys come up empty-handed, astronomers will be forced to consider some of the more bizarre explanations for dark matter.

In fact, most cosmologists have already come to believe that unfamiliar forms of matter must be an important part of the puzzle. Theoretical models of the big bang imply that the density of ordinary, baryonic matter (protons and neutrons) cannot exceed one tenth of the critical density needed to halt the present cosmic expansion, otherwise the composition of the universe would be far different. But some studies of large-scale motions of galaxies, as well as the currently favored version of the big bang, require the universe to have the full critical density. Ninety percent of the universe must consist of exotic, as yet undetected, particles.

Lawrence Krauss of Yale University admits that "it sounds strange, but it is more conservative to assume nonbaryonic dark matter" than to abandon present models of cosmic genesis. Cosmologists have proposed two general kinds of exotic dark matter: cold dark matter, which would clump together readily, and hot dark matter, which would gather on far larger scales. Most cosmologists prefer cold dark matter because it seems better able to explain the known distribution of galaxies.

Current unified physics theories allow for the existence of a bewildering array of potential "cold" particles, including axions, magnetic monopoles and weakly interacting massive particles, or WIMPs. Krauss has done some housecleaning by analyzing recent particle physics experiments at the LEP collider in Switzerland and the Stanford Linear Collider in California. His work has eliminated a number of possible WIMPs and tightly constrained the potential properties of the others. Krauss is optimistic that the negative results will help guide the next round of searches for cold dark matter. Griest concurs. "I bet dark matter will be found in a particle accelerator, either at LEP or in the Superconducting Super Collider—if it is built," he says.

More direct searches for WIMPs are also under way. Assuming dark matter particles exist all around, they should occasionally collide with the nuclei of ordinary atoms, leaving a detectable trail of ionization. One major experiment to search for such ionization signals will

## Booby Prizes

Amid cries of "Excelsior!" and strains from the *Close Encounters of the Third Kind* theme, the oxymoronic Second First Annual Ig Nobel Prize Ceremony began. There was one problem, though: because the stage doors were locked, the presiding Swedish Meatball King and Queen had to knock for someone to let them in. At least a few of the evening's prize recipients probably wish no one had. The Ig Nobels, unlike their more prestigious counterparts, honor individuals whose achievements cannot or should not be reproduced.

The Ig Nobel Prize Ceremony is a new October tradition in bad taste and indifferent science co-sponsored by the Massachusetts Institute of Technology Museum and the *Journal of Irreproducible Results,* a compendium of ersatz experiments. The Ig Nobels are the brainchild of the journal's editor, Marc Abrahams, who hosted the festivities with deadpan earnestness.

Among the winners were half a dozen scientists from the Shiseido Research Center in Yokohama, who took the prize in medicine for their studies of the chemicals responsible for foot odor. Abrahams hailed their conclusion that "people who think they have foot odor do, and those who don't, don't."

Cecil Jacobson, a former physician described by Abrahams as a "relentlessly generous sperm donor," won the Ig Nobel Prize in Biology. This past March a Virginia jury convicted Jacobson of fraud in a case in which prosecutors claimed he had secretly impregnated female patients at his clinic with his own semen. Jacobson, who was sentenced to five years in prison but is free on bond pending an appeal, did not attend the ceremony.

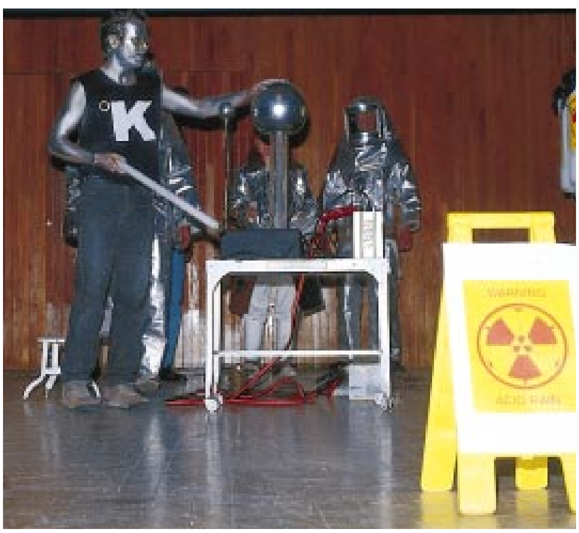
Prolificacy of a different kind brought the literature award to Yuri Struchkov of the Institute of Organoelemental Compounds in Moscow. As Abrahams explained, Struchkov published 948 scientific papers between 1981 and 1990—an average of one every 3.9 days.

For identifying the cause of the mysterious circles of flattened crops that appeared in many British fields, David Chorley and Doug Bower of the U.K. won the physics prize. Many hypotheses, ranging from odd meteorologic phenomena to UFOs, have been advanced, but their explanation has simplicity on its side: they claim to have made the circles themselves with boards and pieces of string.

The French youth group Eclaireurs de France, a name meaning "those who light the way," won special recognition for its singular contributions to archaeology. While on an antigraffiti campaign near the village of Bruniquel, the eager youths erased 15,000-year-old paintings from the walls of the Mayrieres Cave.

Daryl Gates, the former Los Angeles police chief, garnered the peace prize "for his uniquely compelling methods of bringing people together." Accepting the award for Gates was Stan Goldberg of a Harvard Square camera store, who confirmed that "Daryl Gates has done more for the video camera industry than any other individual."

STANLEY ROWIN




***WEIRD SCIENCE prevails at the Ig Nobels.***

The final prize of the evening, for Ig Nobel accomplishments in art, went to Jim Knowlton, whose poster "Penises of the Animal Kingdom" shows the relative sizes and shapes of phalluses from humans, pigs, whales and other species. The National Endowment for the Arts was named as a co-recipient for allegedly encouraging Knowlton "to extend his work in the form of a pop-up book."

At the end of the evening, with the stage doors still locked, the king and queen had to shuffle out the side exit. Next year, along with a new list of "ignitaries," perhaps they will bring a key.

*—Shawna Vogel and J. Rennie*

begin later this year in an unused tunnel at the Stanford High Energy Physics Laboratory. The experiment will use chunks of germanium cooled nearly to absolute zero, which act as hypersensitive energy detectors.

David O. Caldwell of the University of California at Santa Barbara, a participant in the project, notes that such direct dark matter searches "are not sensitive to a particular candidate." Positive findings can be tested by looking for the expected annual variation in ionization energy as the solar system moves through the sea of dark matter.

Even WIMPs may not be able completely to solve the dark matter problem. During the past year, the *Cosmic Background Explorer* satellite has made sensitive microwave maps of the sky that underscore a serious defect in cold dark matter models: they can accurately simulate the structure of the universe on very large scales or on the scales of individual galaxies, but not both.

One way to fix the models is to mix in a smidgen of hot dark matter with the cold dark matter. Such cosmic combination can explain the structure of the universe on all scales, but it requires the existence of a hot dark matter particle. Even Marc Davis of Berkeley, who recently published such a mixed model, admits that "a few years ago I would have called this abhorrent—in fact, I did call it abhorrent."

Cosmologists looking for a hot dark matter particle can at least point to a candidate that actually exists: the neutrino. Physicists had long assumed neutrinos to be massless. Unexpected results from experiments designed to detect neutrinos emitted by the sun have begun to suggest otherwise. Preliminary findings from two new neutrino detectors, SAGE in Russia and GALLEX in Italy, seem to bolster theories that neutrinos do indeed possess a small mass. But nobody yet knows if neutrinos are massive enough to have played a significant role in the evolution of galaxies.

Finally, John A. Bahcall of the Institute for Advanced Study in Princeton, N.J., cautions that the dark matter problem may be a sign that some fundamental aspect of physics, such as the theory of gravity, demands revision. And many assumptions about dark matter depend on the essential validity of big bang cosmology. "Dark matter is the fundamental problem that astronomers and physicists share," Bahcall says. The outcome of the current searches will test not only the cosmological orthodoxy but scientists' ability to deduce the nature of a universe that is mostly inaccessible to their gaze. —*Corey S. Powell*

# Finding the Good in the Bad

As a feminist in a family with Victorian mores and as a Jew and freethinker in Mussolini's Italy, Rita Levi-Montalcini has encountered various forms of oppression many times in her life. Yet the neurobiologist, whose tenacity and preciseness are immediately apparent in her light, steel-blue eyes and elegant black-and-white attire, embraces the forces that shaped her. "If I had not been discriminated against or had not suffered persecution, I would never have received the Nobel Prize," she declares.

Poised on the edge of a couch in her apartment in Rome that she shares with her twin sister, Paola, Levi-Montalcini recalls the long, determined struggle that culminated in joining the small group of women Nobelists in 1986. She won the prize for elucidating a substance essential to the survival of nerve cells. Her discovery of nerve growth factor led to a new understanding of the development and differentiation of the nervous system. Today it and other similar factors are the subject of intense investigation because of their potential to revive damaged neurons, especially those harmed in such diseases as Alzheimer's.

The journey from Turin, where she was born in 1909, to this serene and impeccable Roman living room laden with plants and with the etchings and sculptures of Paola, a well-known artist, tested Levi-Montalcini's mettle from her earliest years. "It was a very patriarchal society, and I simply resented, from early childhood, that women were reared in such a way that everything was decided by the man," she proclaims. Initially, she wanted to be a philosopher but soon decided she was not logically minded enough. When her governess, to whom she was devoted, died of cancer, she chose to become a doctor. There only remained the small matter of getting her father, an engineer, to grant permission and of making up for the time she had lost in a girls' high school, where graduation led to marriage, not to the university. That "annoyed me so much that I decided to never do as my mother did. And it was a very good decision—at that time, I could never have done anything in particular if I had married." Levi-Montalcini pauses, leans forward and asks intensely, "Are you married?" She sighs with relief at the answer. "Good," she says, smiling.

AP Worldwide Photos

*NOBEL LAUREATE Rita Levi-Montalcini conducted neurobiological research as bombs fell on her town during World War II.*

After she received her father's grudging consent, Levi-Montalcini studied for the entrance examination and then enrolled in the Turin School of Medicine at the age of 21. Drawn to a famous, eccentric teacher, Giuseppe Levi, she decided to become an intern at the Institute of Anatomy. There Levi-Montalcini became adept at histology, in particular at staining nerve cells.

Since Levi was curious about aspects of the nervous system, he assigned his student a Herculean labor: to figure out how the convolutions of the human brain are formed. In addition to the overwhelming undertaking of finding human fetuses in a country where abortion was illegal, "the assignment was an impossible task to give your student or an established scientist," Levi-Montalcini explains, her voice hardening. "It was a really stupid question, which I couldn't solve and no one could solve."

She abandoned the project—after a series of unpleasant forays for subject matter—and with Levi's permission began to study the development of the nervous system in chick embryos. Several years later she was forced to stop that work as well. Mussolini had declared his dictatorship by 1925 and since then anti-Semitism had grown in Italy. By 1936, hostility was openly apparent, and in 1939, Levi-Montalcini withdrew from the university, worried about the safety of her non-Jewish colleagues who would be taking a risk by letting her study.

Levi-Montalcini accepted an invitation to conduct her research at a neurological institute in Belgium. But, fearing for her family, she soon returned to Turin—just before Mussolini and Hitler forged their alliance. Undeterred, Levi-Montalcini continued her research: "I immediately found a way to establish a laboratory in my bedroom." In the years that followed, bombs fell repeatedly, and again and again she would lug her microscope and slides to safety in the basement.

In spite of the hardship—or perhaps, as Levi-Montalcini sees it, because of the adversity—it was during this time that she laid the groundwork for her later investigation of nerve growth factor. "You never know what is good, what is bad in life," she muses. "I mean, in my case, it was my good chance." Levi-Montalcini and her family left Turin in 1942 for the surrounding hills and successfully survived the war in hiding. By convincing farmers that she needed eggs for her children (whom she did not have), Levi-Montalcini studied how embryonic nerve tissue differentiates into specialized types. The prevailing theory, developed by renowned biologist Viktor Hamburger of Washington University, held that the differentiation, or specialization, of nerve cells depends in large part on their destination. In his experiments, Hamburger removed developing limbs in chick embryos to see how such excision would affect the later growth and differentiation of the nerve cells destined for that region of the embryo.

Hamburger observed that the centers of embryonic nerve cells near and in the developing spinal column—where the cells start their journey out to other tissues—were much smaller when he excised the limb buds. He suggested that some inductive or organizing factor, probably contained in the limb, could no longer call out to the nerve cells. Therefore, they neither specialized nor grew away from the developing spinal cord into the region of the absent limb.

After conducting experiments directed at the same question, Levi-Montalcini reached a different conclusion. She found that fewer nerve cells grew into the area where the limb bud had been eliminated, but she proposed that some kind of nutrient, important for the survival of nerve cells and normally produced by the limb, was missing. Her theory differed from Hamburger's view because Levi-Montalcini proposed that nerve cell differentiation did take place despite the removal of the limb but that the cells soon died because they did not receive some sustaining, trophic factor. The limb did not contribute to differentiation, that is, it did not contain an organizing factor; rather it produced something that nourished already specialized nerve cells.

A paper of hers on this topic was published in a Belgian journal and was read by Hamburger, who invited her to St. Louis in 1946. Hamburger wanted to work with Levi-Montalcini on the problem of nerve cell differentiation—and, indeed, later came to agree with her interpretation. Although she initially accepted a semester-long research position at Washington, Levi-Montalcini remained until 1961. She is now professor emeritus at Washington but spends most of her time in her native country.

Levi-Montalcini recalls being unsure of the future of her research after she arrived in the U.S. One afternoon, a series of observations, as well as the presentation of a challenge, gave her a renewed sense of purpose. At that time, neurobiologists thought differences in the number and function of various nerve cells were mostly the consequence of proliferative processes.

But Levi-Montalcini was about to discover that the developing nervous system, at least in parts, uses a strategy different from the one previously assumed. She had prepared a series of tissue slides of chick embryo spinal cords in different stages of development. By looking at the succession of slides, she was able to observe the migration of nerve cells early in development to their final positions alongside the spinal column. There, for the first time, she saw the later elimination, or pruning back, of some of them. "I put on a Bach cantata because I was so terribly happy. I had realized that there was still so much to be discovered," says Levi-Montalcini, her delight vividly clear.

*"I simply resented…that women were reared in such a way that everything was decided by the man."*

Over the next several years, Levi-Montalcini focused on searching for the mysterious trophic factor that she had intuited during the war. A former student of Hamburger's had fortuitously noticed that a certain mouse tumor cell line—called sarcoma 180—caused more nerve cells to grow. When Levi-Montalcini incorporated the tumor cells into developing chicks, she observed the same effect. Something in the tumor caused the differentiation of the nerve cells to accelerate; it also caused the creation of excessive numbers of nerve fibers.

Levi-Montalcini started trying to isolate the trophic factor and began to collaborate with biochemist Stanley Cohen, then at Washington and now at the Vanderbilt University School of Medicine. They found that the partially purified factor contained both protein and nucleic acid. By adding enzymes from snake venom—which breaks down these compounds—in hopes of determining which component contained the biological activity, the two discovered that the venom itself contained the factor.

This finding (described in detail in her autobiography, *In Praise of Imperfection*) led to the realization that nerve growth factor is produced in salivary glands in mice, providing a new, easy source for studies of the material. By designing an antiserum, Levi-Montalcini and Cohen were able to chart the role of the factor. It became clear that it is essential to the differentiation and health of nerve cells.

In 1986 Levi-Montalcini and Cohen shared the Nobel Prize for this achievement. When the phone rang in Rome with the news, she was pages from the end of Agatha Christie's *Evil under the Sun*. "At the moment that I was finding out about the criminal, they told me that I was awarded the Nobel," she laughs, getting up to retrieve the book from the hallway. She points to a handwritten note on the second-to-last page—befitting a neuroscientist, her edition has a skull on the cover—where she had marked "call from Stockholm" and the time. "So I was very happy about it, but I wanted much more to know the end of the story," she admits.

Although she says her popularity interferes with her life, Levi-Montalcini has used the Nobel to extend her work into areas that concern her. She is president of the Italian Multiple Sclerosis Association and is a member of the Pontifical Academy of Sciences; she was the first woman to be elected to the academy. "I can do things that are very, very important, which I would never have been able to do if I did not receive it," she says. "It has given me the possibility of helping a lot of people." And she helps whomever she can. The phone rings incessantly in her apartment. "People ask for medical help," she explains, after answering each call and graciously talking with the parents or other relatives of someone ill. "But sometimes there is nothing to do."

In addition, Levi-Montalcini and her sister recently started their own project: a foundation that will provide mentors, counseling and grants to teenagers deciding what field, whether it be art or science, to enter. For several hours every week, she receives young students in her laboratory at the Institute of Neurobiology at the National Research Council in Rome and talks with them about their interests and her experiments. "The only way to help is to give young people a chance for the future. Because we cannot fight the Mafia, we cannot fight corruption without giving an alternative to young people," she says.

Levi-Montalcini's research at the institute, which she founded in the 1960s, has also taken a new turn. She is studying the role of nerve growth factor in the immune and endocrine systems. "The neotrophic factor was just the tip of the iceberg," she notes. "So even now I am doing something entirely different. Just in the same spirit as when I was a young person. And this is very pleasing to me," she says, laughing. "I mean, at my old age, I could have no more capacity. And I believe I still have plenty." —*Marguerite Holloway*

SCIENTIFIC AMERICAN

# Coral Bleaching

*Environmental stresses can cause irreparable harm to coral reefs. Unusually high seawater temperatures may be a principal culprit*

by Barbara E. Brown and John C. Ogden

Late summer of 1987 seemed typical for that time of year in the Virgin Islands. Huge, flat-bottomed cumulus clouds moved westward on light trade winds. Calm seas were rarely disturbed by squalls sweeping into the northeastern Caribbean Sea from the Atlantic Ocean. The only suggestion that something might be amiss was the water, which, though not systematically measured, seemed unusually warm to people swimming near the shallow coral reefs.

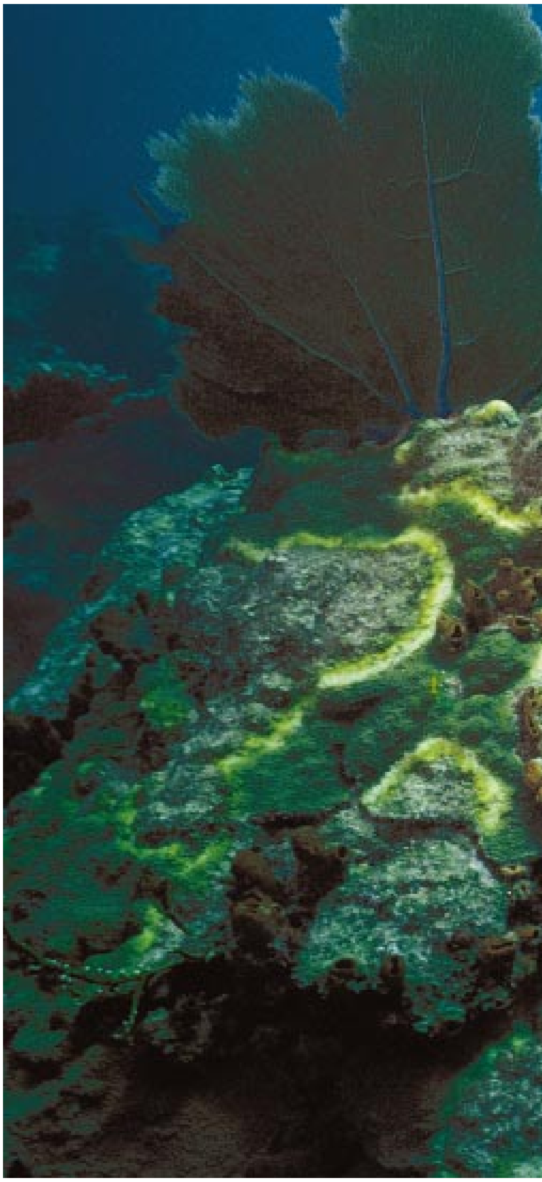
Something atypical had indeed occurred. The normally golden-brown, green, pink and gray corals, sea whips and sponges had become pure white. In some cases, entire reefs were so dazzlingly white that they could be seen from a considerable distance. In other areas, pale corals punctuated the reef surface while unbleached corals of the same species grew as neighbors.

The phenomenon, which can be lethal to coral, was not confined to the Virgin Islands. Observers at numerous marine laboratories in the Caribbean noted the same whitening. Nor was it the first occurrence of such bleaching. In 1982 and 1983, after the atmospheric and oceanographic disturbance called El Niño/Southern Oscillation (ENSO), corals in certain areas of the Florida Keys whitened and died, and off the coast of Panama mortality reached 50 percent. But it was only between 1987 and 1988, also an ENSO year, that reports of extensive bleaching became widespread. They have increased in frequency ever since.

The association of coral bleaching with ENSO, which ushers in warm water, and with water temperatures two to three degrees Celsius above normal has led some scientists to suggest that the bleaching is a manifestation of global warming. Others point out that coral reefs have been studied only for a few decades—too short a time to permit generalized conclusions about a poorly understood event.

Nevertheless, coral reefs around the world are suffering bouts of bleaching from which many do not recover. Although several factors can cause the process—including disease, excess shade, increased ultraviolet radiation, sedimentation, pollution and changes in salinity—the episodes of the past decade have consistently been correlated with abnormally high seawater temperatures. Understanding the complex pro-

BARBARA E. BROWN and JOHN C. OGDEN work on international issues of coral reef ecology and management. Brown is director of the Centre for Tropical Coastal Management and is reader in tropical marine biology at the University of Newcastle upon Tyne in the U.K. She is also a founder of the International Society for Reef Studies. Ogden is the director of the Florida Institute of Oceanography and professor of biology at the University of South Florida. He has served as a member of the faculty and as the director of the West Indies Laboratory in St. Croix, where he began his work on coral reefs.



**BOULDER CORAL has bleached only in parts—the rest remains healthy for now. Although several factors cause potentially lethal whitening, recent bouts have been consistently correlated with higher than average seawater temperatures.**

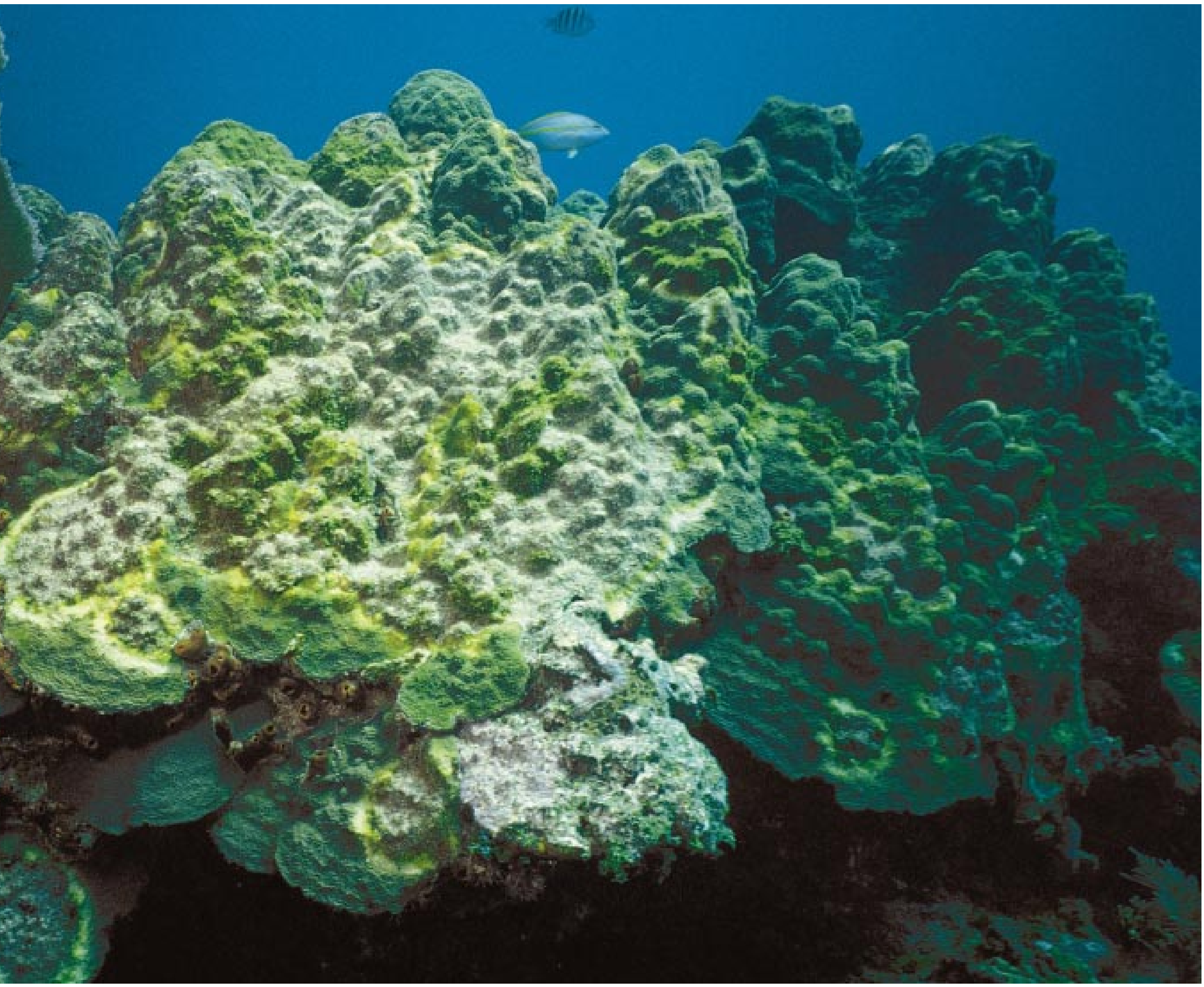
cess of bleaching can help pinpoint, and perhaps eventually deter, this threat to the ecology of the reefs.

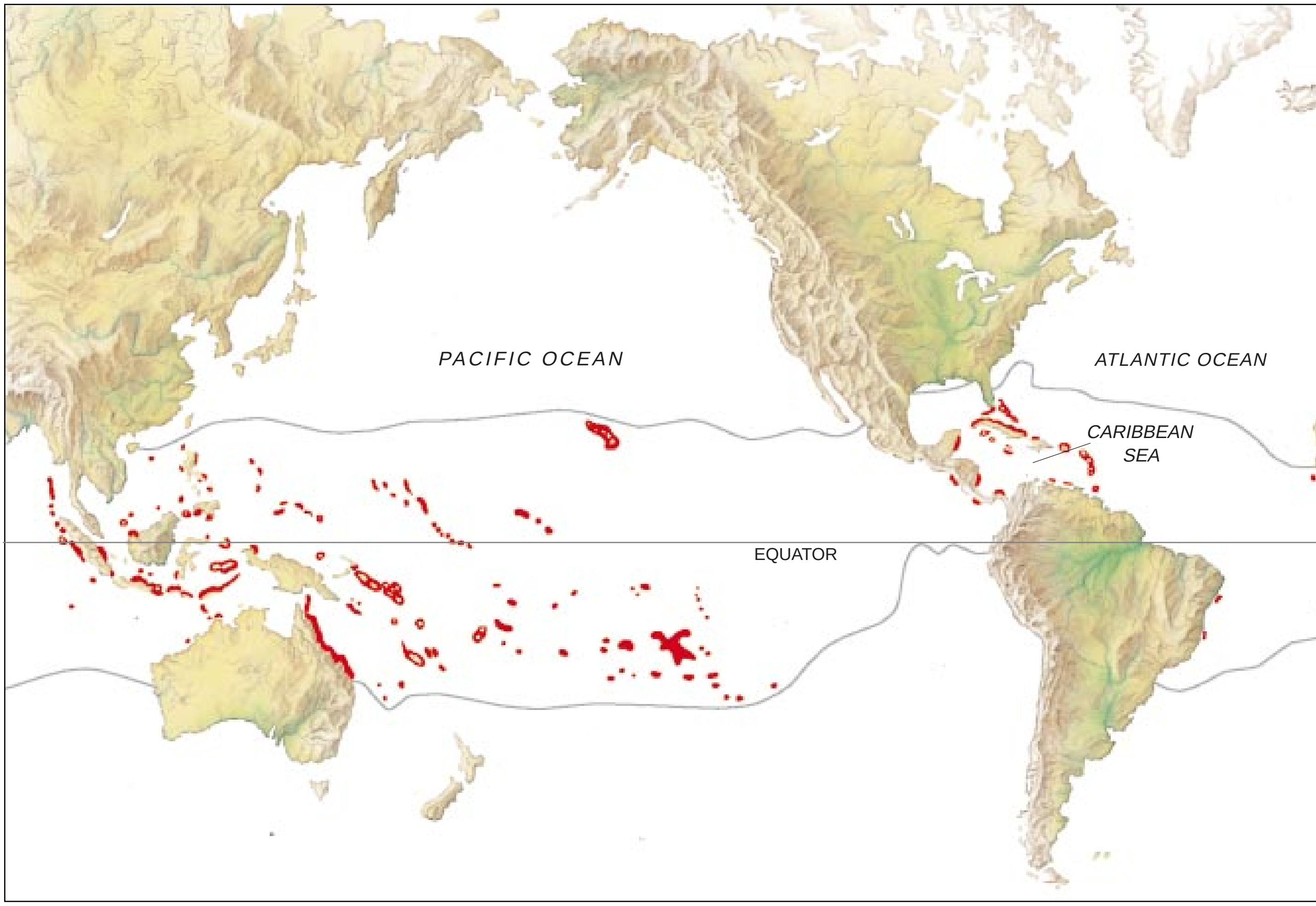
Tropical, shallow-water ecosystems, coral reefs are found around the world in the latitudes that generally fall between the southern tip of Florida and mid-Australia. They rank among the most biologically productive of all marine ecosystems. Because they harbor a vast array of animals and plants, coral reefs are often compared to tropical rain forests. Reefs also support life on land in several ways. They form and maintain the physical foundation for thousands of islands. By building a wall along the coast, they serve as a barrier against oceanic waves. And they sustain the fisheries and tourist diving industries that help to maintain the economies of many countries in the Caribbean and Pacific.

Although corals seem almost architectural in structure—some weigh many tons and stand between five and 10 meters high—they are composed of animals. Thousands of tiny creatures form enormous colonies: indeed, nearly 60 percent of the 220 living genera of corals do so. Each colony is made up of many individual coral animals, called polyps. Each polyp is essentially a hollow cylinder, closed at the base and interconnected to its neighbors by the gut cavity. The polyps have one or more rings of tentacles surrounding a central mouth. In this way, corals resemble sea anemones with skeletons. The soft external tissues of the polyps overlie a hard structure of calcium carbonate.

Many of the splendid colors of corals come from their symbionts, creatures that live in a mutually dependent relation with the coral. Symbiotic algae called zooxanthellae reside in the often transparent cells of the polyps. There are between one and two million algae cells per square centimeter of coral tissue. Through photosynthesis the algae produce carbon compounds, which help to nourish the coral—some species receive 60 percent of their food from their algae. Algal photosynthesis also accelerates the growth of the coral skeleton by causing more calcium carbonate to be produced. The corals provide algae with nutrients, such as nitrogen and phosphorus, essential for growth, as well as with housing. The association enables algae to obtain compounds that are scarce in the nutrient-poor waters of the tropics (where warm surface waters overlie and lock in cold, nutrient-rich waters—except in restricted areas of upwelling).

When corals bleach, the delicate balance among symbionts is destroyed.






**CORAL REEFS *(red)* thrive around the world in tropical, shallow waters—those areas falling between the lines. They are the most biologically productive of all marine ecosystems. Coral reefs also support life on land by providing a barrier**

The corals lose algae, leaving their tissues so colorless that only the white, calcium carbonate skeleton is apparent. Other organisms such as anemones, sea whips and sponges—all of which harbor algae in their tissue—can also whiten in this fashion. Some of this loss is routine. A healthy coral or anemone continuously releases algae, but in very low numbers. Under natural conditions, less than 0.1 percent of the algae in a coral is lost during processes of regulation and replacement. When subject to adverse changes, such as temperature increases, however, the corals release increased numbers of algae. For example, transferring coral from a reef to a laboratory can cause a fivefold elevation in the numbers of algae expelled.

The mechanism of algal release is not fully understood. Even defining bleaching remains tricky. The current definition has its basis in laboratory measurements of the loss of algae and the reduction in algal pigments. The laboratory approach, however, is rarely, if ever, applied in the field. There judgment must rely on the naked eye's ability to detect loss of coloration. Although such methods may be reliable for instances of severe bleaching, a determination that pale colonies are bleached can be extremely arbitrary, given the natural variability of pigmentation.

In some cases, normal coral undergoing an adaptive behavioral response can look bleached. In 1989, while working at the Phuket Marine Biological Center in Thailand, one of us (Brown) observed that some intertidal coral species—those that thrive in shallow water and are exposed to the air at low tide—

## Some of the Species That Thrive on Coral Reefs

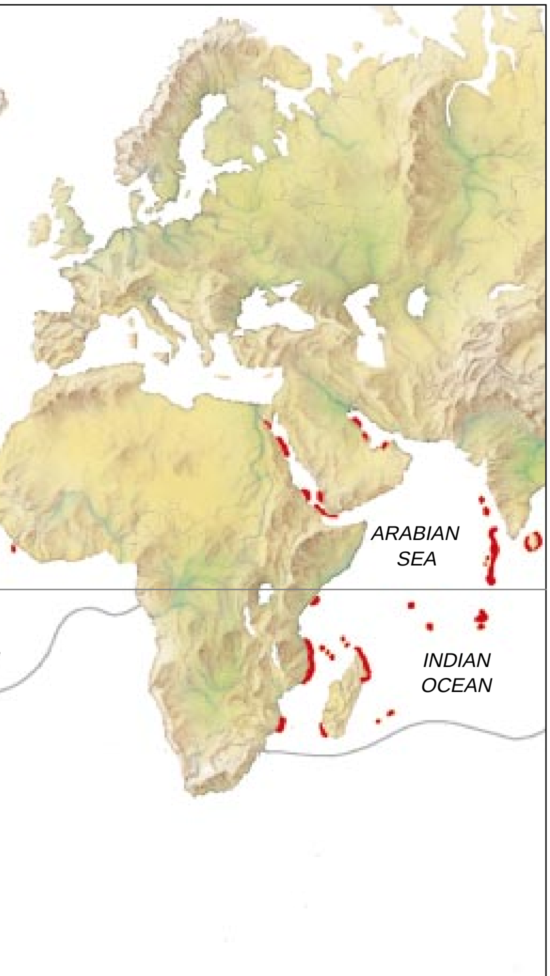
SPINY LOBSTER

QUEEN CONCH

LOGGERHEAD SPONGE

SEA ANEMONE

BUTTERFLY FISH




**against oceanic waves and by forming the foundation for thousands of islands.**

appear completely white during low spring tides. It became clear that these corals are able to pull back their external tissues, leaving their skeletons exposed; they do not lose their algae. This behavior should perhaps be more accurately described as blanching, a response that may reduce desiccation during exposure to air.

Despite the absence of an unassailable definition of the bleaching process, several mechanisms have been proposed that may be at work. In 1928 Sir Maurice Yonge and A. G. Nicholls, who participated in an expedition to the Great Barrier Reef, were among the first to describe coral bleaching. They suggested that algae migrated through coral tissue in response to environmental stress, before being released into the gut and ultimately expelled through the mouth. The precise trigger for the release and the stimulus causing the algae to be so conveyed were unknown then and remain largely unknown today.

One of the several theories proposed by Leonard Muscatine of the University of California at Los Angeles is that stressed coral polyps provide fewer nutrients to the algae. According to this theory, the algae would not necessarily be directly affected by, say, high temperature, but the metabolism of the coral would be lowered. Supplies of carbon dioxide, nitrogen and phosphorus would become insufficient, and this rationing would in turn cause the algae to abandon their residence.

In addition, Muscatine, R. Grant Steen and Ove Hoegh-Guldberg, also at U.C.L.A., studied the response of anemones and corals to changes in temperature, light and salinity. They described the release of algae from the tissues into the gut cavity and hypothesized that the coral was actually losing animal tissue along with the algal cells. Work by Suharsono at the University of Newcastle upon Tyne supported this idea. He showed that anemones exposed to warmer temperatures in the laboratory lose their own cells and algal cells during bleaching. Thus, host tissue thinned significantly, perhaps reducing the space available for the algae.

The direct release of algae into the gut, however, may be a mechanism of algal loss that results only from the extreme shocks invoked in a laboratory. It is not yet clear that algae behave the same way in corals in situ. Under natural conditions, it is quite likely that algae are released by a variety of mechanisms. All experimental work carried out on bleaching has involved exposure to extreme temperature changes—that is, increases of six degrees C or more over a period of 16 to 72 hours. In nature the temperature increases that induce bleaching are much smaller, about two degrees C, and may occur over several months.

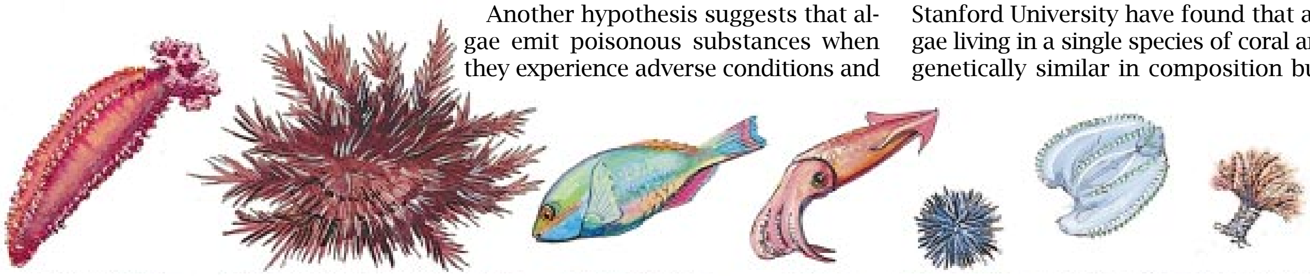
Another hypothesis suggests that algae emit poisonous substances when they experience adverse conditions and that these toxins may deleteriously affect the host. Algae produce oxygen compounds, called superoxide radicals, in concentrations that can damage the coral. (Molecular oxygen is relatively unreactive, but it can be chemically altered to form the superoxide radical.) An enzyme, superoxide dismutase, in the coral detoxifies the radicals.

But Michael P. Lesser and his colleagues at the University of Maine noted that in certain cases oxygen toxicity could lead to bleaching. Although Lesser and his team were unable to measure the oxygen radicals directly, they followed the production of superoxide dismutase. They found that exposure to elevated temperatures and to increased ultraviolet radiation independently spurred enzymatic activity. The researchers concluded that oxygen toxicity could be responsible for bleaching because harmful oxygen radicals were exported from the damaged algae to the animal host.

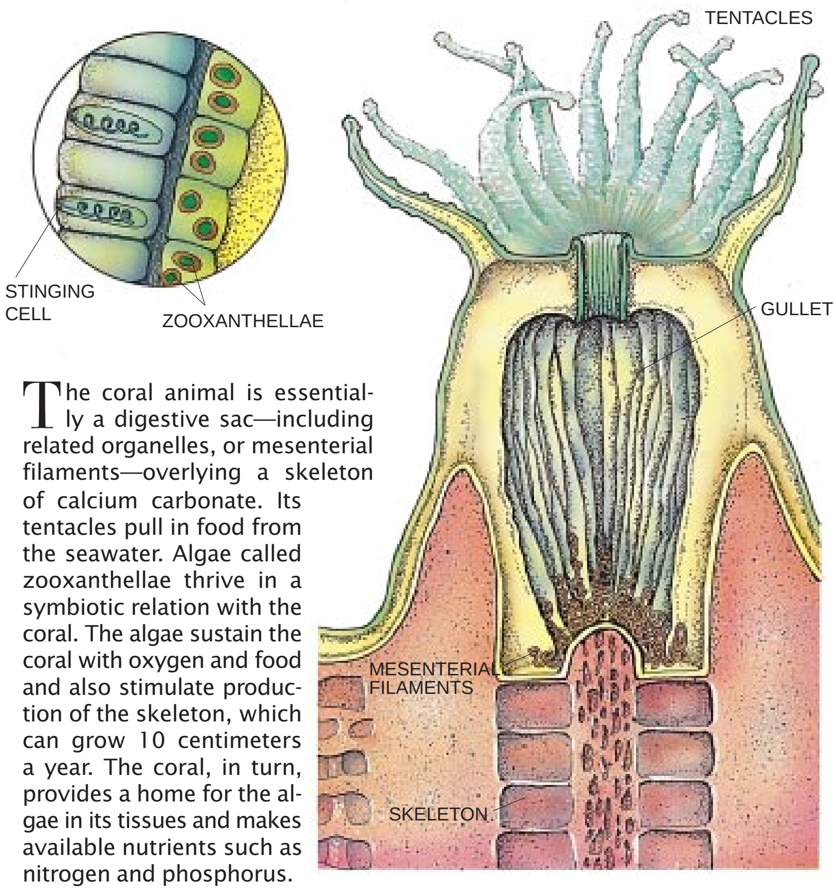
Other biochemical changes may take place as well. David Miller of the University of Leeds and students from Brown's laboratory suggest that alterations in gene expression occur as a response to deleterious environmental changes. These changes may involve the synthesis of heat-shock proteins—compounds found in all living systems subject to adversity that serve to protect cells temporarily from heat damage. Miller determined that these proteins are enhanced in anemones undergoing heat shock. Furthermore, in anemones that tolerate temperature increases, the presence of proteins appears to be correlated with reduced bleaching during heat shock.

Genetic variability also plays an important role in bleaching. Environmental factors may affect species of algae or coral in different ways. Of course, predicting the ability of corals and their algae to adapt to increases in seawater temperature or global climatic change may be possible by identifying the types of corals or algae at highest risk.

When working together at the University of California at Santa Barbara, Robert K. Trench and Rudolf J. Blank showed that different corals act as hosts to varied strains of algae. Subsequently, Rob Rowan and Dennis A. Powers of Stanford University have found that algae living in a single species of coral are genetically similar in composition but



## Anatomy of a Coral Polyp




The coral animal is essentially a digestive sac—including related organelles, or mesenterial filaments—overlying a skeleton of calcium carbonate. Its tentacles pull in food from the seawater. Algae called zooxanthellae thrive in a symbiotic relation with the coral. The algae sustain the coral with oxygen and food and also stimulate production of the skeleton, which can grow 10 centimeters a year. The coral, in turn, provides a home for the algae in its tissues and makes available nutrients such as nitrogen and phosphorus.

genetically different from algae in other coral species. Certain algae may prove to be particularly sensitive to temperature and may have varying temperature tolerances. If so, Rowan and Powers's findings would help explain why related, but not identical, corals exposed to warmer temperatures frequently show different susceptibilities to bleaching.

Alternatively, the variability may lie with the coral instead of the algae. Studies of several species of coral have indicated that genetically dissimilar strains exist within a species of coral. Such strains may have different environmental tolerances, which could account for the observation that one colony of a particular species appears to have been bleached while a nearby member of the same species has not been.

The reports of bleaching in the Caribbean in the 1980s seem to be related most consistently to elevated sea temperatures. Coral reefs normally thrive between 25 and 29 degrees C, depending on their location. When patterns depicting coral diversity are plotted on a globe, it is apparent that diversity declines as the reefs get farther away from two centers—one in the Indo-Pacific and the other in the Caribbean. The outlines of a map marking plummeting diversity coincide with the contours of lower seawater temperatures.

The narrow temperature range for healthy coral is very close to its upper lethal temperature: an increase of one to two degrees above the usual summer maximum can be deadly. Paul Jokiel and Stephen Coles of the University of Hawaii have shown that bleaching and coral mortality are not induced by the shock of rapidly fluctuating temperatures but are a response to prevailing high temperatures and to significant deviations above or below the mean.

Many times during a 10-month period in 1982–1983, an unusually severe ENSO warmed the waters of the eastern Pacific three to four degrees C over the seasonal average. Peter W. Glynn and his colleagues at the University of Miami tracked the event and the subsequent developments in that region. As a result of elevated temperatures, coral reefs underwent bleaching. Between 70 and 90 percent of the corals in Panama and Costa Rica perished several weeks later; more than 95 percent of the corals in the Galápagos were destroyed.

Glynn and Luis D'Croz of the University of Panama also linked coral mortality with high temperatures in a series of laboratory experiments that duplicated field conditions during ENSO. The major reef-building coral of the eastern Pacific, *Pocillopora damicornis,* took the same amount of time to die in the laboratory at 32 degrees C as it did in the field, indicating that the experiments had replicated the natural condition. Glynn and D'Croz also suggested that the temperature disproportionately affected types of coral that normally experienced seasonal upwelling of deep, cool water in the Gulf of Panama.

Evidence for the 1987 warming of the seawater in the Caribbean is not as definitive. Donald K. Atwood and his colleagues at the Atlantic Oceanographic and Meteorological Laboratory in Miami examined the National Oceanographic Data Center's sea-surface temperature records from 1932 to the present and found no discernible long-term increases in seawater temperature in the Caribbean. The monthly mean sea-surface temperature did not exceed 30.2 degrees C in any of the regions examined—in other words, the water remained well below the 32 degrees C required to induce bleaching in Glynn's laboratory experiments.

Atwood also examined the maps from the National Climate Data Center of the National Oceanic and Atmospheric Administration (NOAA). These records provide average monthly sea-surface temperature and track anomalies derived from satellite data that are validated by measurements taken from ships. The maps indicate that in 1987 the surface of the Caribbean was generally less than 30 degrees C. Other groups examined similar temperature records and concluded that the temperatures of some sectors of the Caribbean reached 31 degrees C or more during 1990, another year of bleaching.

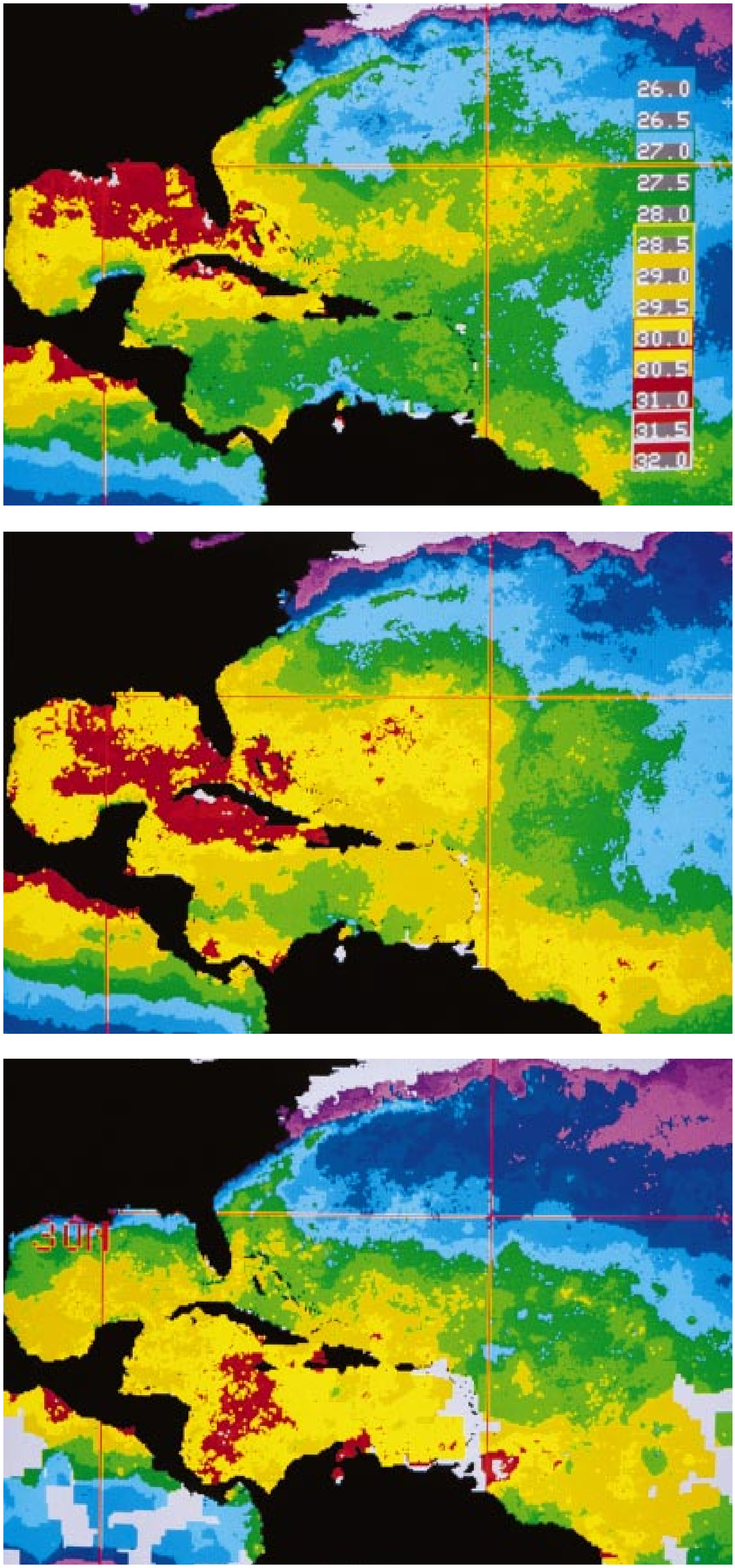
The records, of course, are subject to interpretation based on the geographic scale of the satellite measurements and the integration of these data with in situ measurements. Unfortunately, there are no long-term temperature records taken at the small geographic scale needed to clarify the cause of damage to corals.

The 1987 reports of coral bleaching coincided with escalating concern about global warming. It was not surprising, therefore, that some scientists and other observers reached the conclusion that coral reefs served as the canary in the coal mine—the first indication of an increase in global ocean temperatures. Although it appears that elevated local seawater temperatures caused bleaching, linking this effect to global warming cannot be conclusive at this time. With the support of the National Science Foundation, NOAA and the Environmental Protection Agency, reef scientists and climatologists convened in Miami in June 1991 to discuss coral reefs and global climatic change. The workshop determined that reports of coral bleaching were indicative of threats to the ecosystem and that bleaching did appear to be associated with local temperature increases. But the paucity of knowledge about the physiological response of corals to stress and temperature, the inadequacy of seawater temperature records and the lack of standardized protocol for field studies made it impossible to decide whether bleaching reflects global climatic change in the ocean.

Several international monitoring efforts are now in progress or are planned so that the appropriate data can be gathered. For example, the Caribbean Coastal Marine Productivity Program, a cooperative research network of more than 20 Caribbean marine research institutions in 15 countries, was founded in 1990 and began systematic observations of coral reefs in 1992. Other networks of marine laboratories have been proposed in the central and western Pacific Ocean.

Whatever its cause, bleaching has important implications for the community structure, growth and accretion of coral reefs. Developing countries are particularly dependent on coral reefs for food resources and have made heavy investments in reef-related tourism. Bleaching, added to the accumulated toll taken by pollution and overfishing, may seriously burden the future economies of many nations.

The death of coral over such a wide geographic range in the eastern Pacific during the 1982–1983 ENSO had severe biological repercussions. Before the widespread bleaching, Glynn and his colleagues noticed that large fields of *Pocillopora* served to protect more massive coral species from the coral-feeding crown-of-thorns sea star (*Acanthaster planci*). The starfish did not venture across the dense coral stands, because




**SEAWATER TEMPERATURE fluctuations in the Caribbean Sea in 1990 were tracked by satellite. Temperatures reached between 31 and 32 degrees Celsius (*red*) in certain areas during the months of August (*top*), September (*middle*) and October (*bottom*). Such unusually warm water is believed to cause coral reefs to bleach.**

**INTERTIDAL CORALS in Thailand "blanch" when exposed to the air. Unlike bleaching, this phenomenon appears to be adaptive: the coral polyps retract their soft tissues during low tide, leaving the calcium carbonate skeleton exposed. When water washes over again, the polyps and tentacles expand to cover the skeleton.**

*Pocillopora* repelled it with the stinging cells of its tentacles. In addition, several species of symbiotic shrimp and crab in the *Pocillopora* attacked the sea stars, driving them away. As a result of warmer water, however, *Pocillopora* suffered higher mortality and lower fecundity, and large corals were consequently open to attack by the sea star. The predatory crustaceans were also affected. Because they normally feed on the lipid-rich mucus produced by the *Pocillopora* coral, a decline in the quantity and lipid content of the mucus brought about by the thermal stress triggered a decrease in the crustacean population.

The massive reduction in coral cover on the reefs of Panama and the Galápagos in 1982 and 1983 also restricted the range of one species of hydrocoral called *Millepora* and caused the apparent extinction of another species of the same genus. Glynn and W. H. de Weerdt, now at the University of Amsterdam Zoological Museum, speculate that these species of corals were most severely affected because of their limited range and extreme sensitivity to increases in temperature. The disturbance also caused a nearly complete interruption in the long-term accumulation of calcium carbonate on the reefs of the region.

The fact that healthy reefs flourished in the eastern Pacific Ocean before 1982 indicates that an event of this magnitude is rare. Glynn estimated the age of two species of corals that were killed or heavily damaged in the Galápagos by multiplying their radius by their annual growth rate of approximately one centimeter. He concluded that an ENSO like that of 1982–1983 had not occurred in the Galápagos for at least 200 years, possibly 400. The estimate is similar for corals in Panama. Interestingly, even at their most healthy, the coral reefs of the eastern Pacific are less well developed than those of the Caribbean. Their comparatively meager development may be partly explained by the relatively frequent high- and low-temperature disturbances over thousands of years.

The coral frameworks of the reefs of Panama and the Galápagos have changed dramatically as a result of the bleaching. Large areas of dead coral have become colonized by benthic algae, which in turn support increased populations of herbivores, particularly sea urchins. Sea urchins are grazers; they scrape the coral rock surface of the reef as they feed, contributing to the erosion of the reef structure.

Glynn and Ian Macintyre of the Smithsonian Institution and Gerard M. Wellington of the University of Houston have estimated the rates of calcium carbonate accretion and erosion on the reef. The rates of erosion after the 1982–1983 ENSO attributable to sea urchins alone are greater than the rates of accumulation on the healthy reefs before 1983. This finding suggests that without recovery of coral populations, these reefs will soon be reduced to carbonate sediments. Because the grazers erode the reef surface, they may also interfere with the recruitment of new coral colonies, prolonging, or even preventing, their recovery.

Coral became bleached at many other places in the Indo-Pacific during the 1982–1983 ENSO, including the Society Islands, the Great Barrier Reef, the western Indian Ocean and Indonesia. Brown and Suharsono noted widespread bleaching and loss of as much as 80 or 90 percent of the coral cover on the shallow coral reefs of the Thousand Islands in the Java Sea. The corals most affected were on the shallow reef tops. Five years later coral cover was only 50 percent of its former level.

The extent of bleaching, environmental tolerances and the life-history characteristics of the dominant corals determine whether a reef recovers from the loss of most of its living coral. So do the nature and timing of other disturbances, such as predation and grazing. When bleaching is severe or prolonged, the coral may die. If the bleaching episode is short, the coral can rebuild its algal population and continue to live, but biological processes such as growth and reproduction may be impaired.

Because we are only now forming networks of sites that will conduct cooperative observations, the extent of coral reef damage brought about by bleaching has not been globally assessed. In 1987 Ernest H. Williams, Jr., of the University of Puerto Rico collected reports of bleaching from nearly every tropical ocean region. But until we have an adequate definition of coral bleaching in the field and have standardized our observations, the global impact of coral bleaching will remain a mystery.

If the temperature increase of one or two degrees C, predicted by the Intergovernmental Panel on Climate Change, does take place over the next 50 years in the tropical latitudes, the consequences for coral reefs could be disastrous. Unlike the miners with the canary, we cannot yet link bleaching to a clear cause. But that does not mean we should ignore the coral's message.


FURTHER READING

GLOBAL ECOLOGICAL CONSEQUENCES OF THE EL NIÑO SOUTHERN OSCILLATION 1982–83. Edited by Peter W. Glynn. Elsevier, Amsterdam, 1989.

CORAL BLEACHING. Edited by Barbara E. Brown. Special Issue of *Coral Reefs,* Vol. 8, No. 4; April 1990.

WORKSHOP ON CORAL BLEACHING, CORAL REEF ECOSYSTEMS AND GLOBAL CHANGE: REPORT OF PROCEEDINGS. Organized by Christopher F. D'Elia, Robert W. Buddemeier and Stephen V. Smith. Maryland Sea Grant College, 1991.

# How the Milky Way Formed

*Its halo and disk suggest that the collapse of a gas cloud, stellar explosions and the capture of galactic fragments may have all played a role*

by Sidney van den Bergh and James E. Hesser

Attempts to reconstruct how the Milky Way formed and began to evolve resemble an archaeological investigation of an ancient civilization buried below the bustling center of an ever changing modern city. From excavations of foundations, some pottery shards and a few bones, we must infer how our ancestors were born, how they grew old and died and how they may have helped create the living culture above. Like archaeologists, astronomers, too, look at small, disparate clues to determine how our galaxy and others like it were born about a billion years after the big bang and took on their current shapes. The clues consist of the ages of stars and stellar clusters, their distribution and their chemistry—all deduced by looking at such features as color and luminosity. The shapes and physical properties of other galaxies can also provide insight concerning the formation of our own.

The evidence suggests that our galaxy, the Milky Way, came into being as a consequence of the collapse of a vast gas cloud. Yet that cannot be the whole story. Recent observations have forced workers who support the hypothesis of a simple, rapid collapse to modify their idea in important ways. This new information has led other researchers to postulate that several gas cloud fragments merged to create the protogalactic Milky Way, which then collapsed. Other variations on these themes are vigorously maintained. Investigators of virtually all persuasions recognize that the births of stars and supernovae have helped shape the Milky Way. Indeed, the formation and explosion of stars are at this moment further altering the galaxy's structure and influencing its ultimate fate.

Much of the stellar archaeological information that astronomers rely on to decipher the evolution of our galaxy resides in two regions of the Milky Way: the halo and the disk. The halo is a slowly rotating, spherical region that surrounds all the other parts of the galaxy. The stars and star clusters in it are old. The rapidly rotating, equatorial region constitutes the disk, which consists of young stars and stars of intermediate age, as well as interstellar gas and dust. Embedded in the disk are the sweepingly curved arms that are characteristic of spiral galaxies such as the Milky Way. Among the middle-aged stars is our sun, which is located about 25,000 light-years from the galactic center. (When you view the night sky, the galactic center lies in the direction of Sagittarius.) The sun completes an orbit around the center in approximately 200 million years.

That the sun is part of the Milky Way was discovered less than 70 years ago. At the time, Bertil Lindblad of Sweden and the late Jan H. Oort of the Nether-

SIDNEY VAN DEN BERGH and JAMES E. HESSER both work at Dominion Astrophysical Observatory, National Research Council of Canada, in Victoria, British Columbia. Van den Bergh has a longtime interest in the classification and evolution of galaxies and in problems related to the age and size of the universe. He received his undergraduate degree from Princeton University and a doctorate in astronomy from the University of Göttingen. Hesser's current interests focus on the ages and compositions of globular star clusters, which are among the oldest constituents of the galaxy. He received his B.A. from the University of Kansas and his Ph.D. in atomic and molecular physics from Princeton.


**MILKY WAY COMPONENTS include the tenuous halo, the central bulge and a highly flattened disk that contains the spiral arms. The nucleus is obscured by the stars and gas clouds of the central bulge. Stars in the bulge and halo tend to be old; disk stars such as the sun are young or middle-aged.**

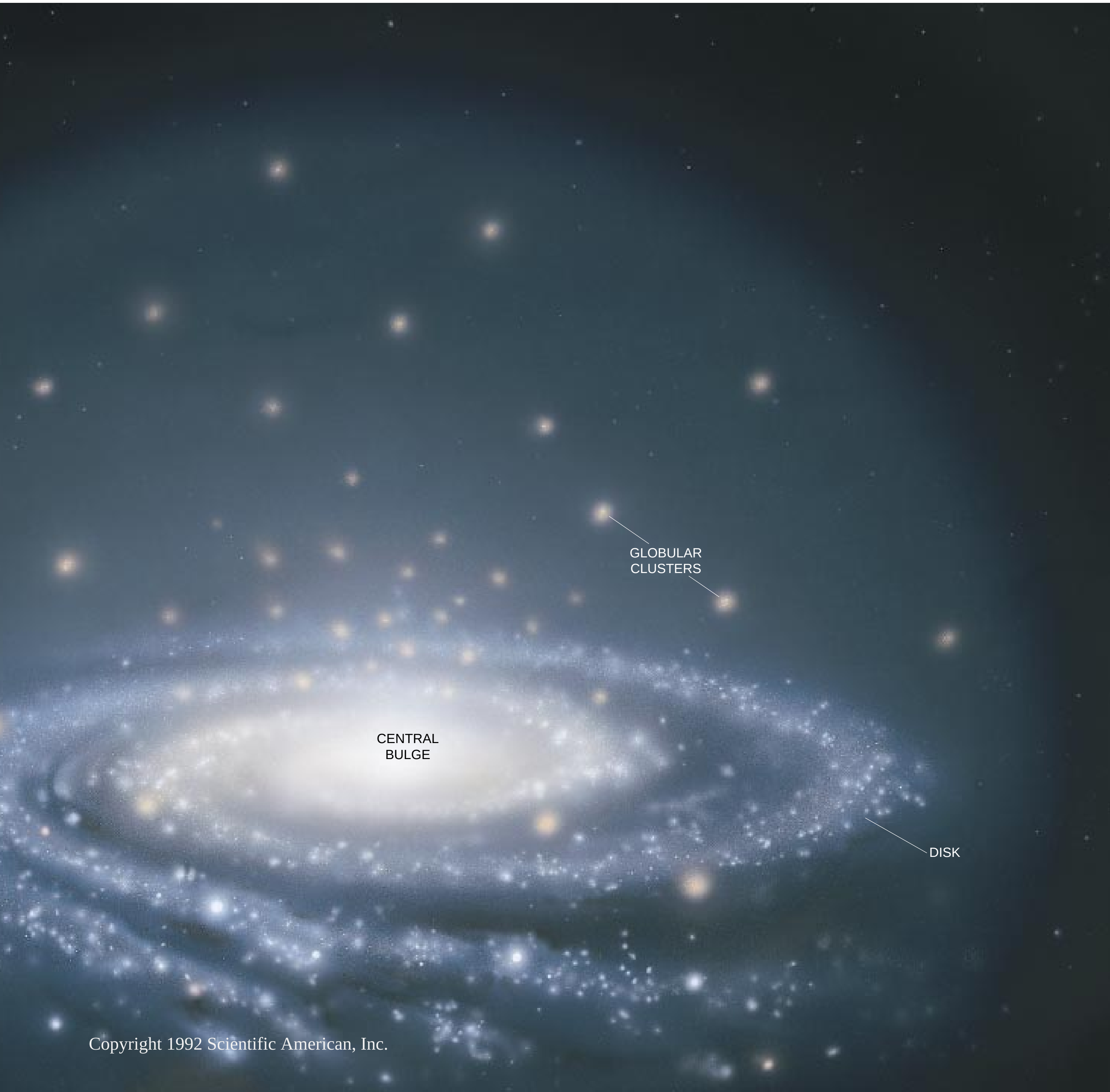
lands hypothesized that the Milky Way system is a flattened, differentially rotating galaxy. A few years later John S. Plaskett and Joseph A. Pearce of Dominion Astrophysical Observatory accumulated three decades' worth of data on stellar motions that confirmed the Lindblad-Oort picture.

In addition to a disk and a halo, the Milky Way contains two other subsystems: a central bulge, which consists primarily of old stars, and, within the bulge, a nucleus. Little is known about the nucleus because the dense gas clouds in the central bulge obscure it. The nuclei of some spiral galaxies, including the Milky Way, may contain a large black hole. A black hole in the nucleus of our galaxy, however, would not be as massive as those that seem to act as the powerful cores of quasars.

All four components of the Milky Way appear to be embedded in a large, dark corona of invisible material. In most spiral galaxies the mass of this invisible corona exceeds by an order of magnitude that of all the galaxy's visible gas and stars. Investigators are intensely debating what the constituents of this dark matter might be.

The clues to how the Milky Way developed lie in its components. Perhaps the only widely accepted idea is that the central bulge formed first, through the collapse of a gas cloud. The central bulge, after all, contains mostly massive, old stars. But determining when and how the disk and halo formed is more problematic.

In 1958 Oort proposed a model according to which the population of stars forming in the halo flattened into a



thick disk, which then evolved into a thin one. Meanwhile further condensation of stars from the hydrogen left over in the halo replenished that structure. Other astronomers prefer a picture in which these populations are discrete and do not fade into one another. In particular, V. G. Berman and A. A. Suchkov of the Rostov State University in Russia have indicated how the disk and halo could have developed as separate entities.

These workers suggest a hiatus between star formation in the halo and that in the disk. According to their model, a strong wind propelled by supernova explosions interrupted star formation in the disk for a few billion years. In doing so, the wind would have ejected a significant fraction of the mass of the protogalaxy into intergalactic space. Such a process seems to have prevailed in the Large Magellanic Cloud, one of the Milky Way's small satellite galaxies. There an almost 10-billion-year interlude appears to separate the initial burst of creation of conglomerations of old stars called globular clusters and the more recent epoch of star formation in the disk. Other findings lend additional weight to the notion of distinct galactic components. The nearby spiral M33 contains a halo but no nuclear bulge. This characteristic indicates that a halo is not just an extension of the interior feature, as many thought until recently.

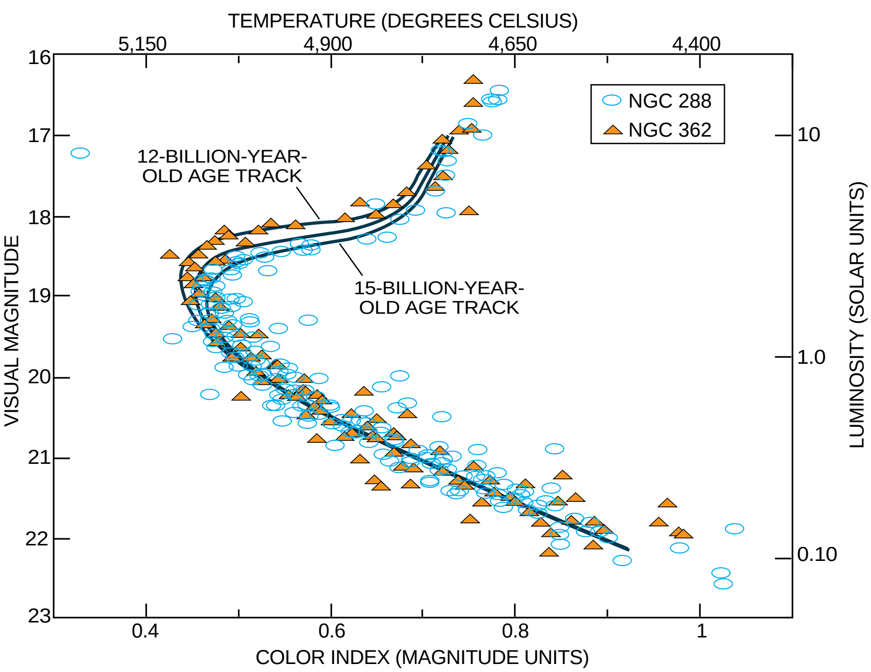
In 1962 a model emerged that served as a paradigm for most investigators. According to its developers—Olin J. Eggen, now at the National Optical Astronomical Observatories, Donald Lynden-Bell of the University of Cambridge and Allan R. Sandage of the Carnegie Institution—the Milky Way formed when a large, rotating gas cloud collapsed rapidly, in about a few hundred million years. As the cloud fell inward on itself, the protogalaxy began to rotate more quickly; the rotation created the spiral arms we see today. At first, the cloud consisted entirely of hydrogen and helium atoms, which were forged during the hot, dense initial stages of the big bang. Over time the protogalaxy started to form massive, short-lived stars. These stars modified the composition of galactic matter, so that the subsequent generations of stars, including our sun, contain significant amounts of elements heavier than helium.

Although the model gained wide acceptance, observations made during the past three decades have uncovered a number of problems with it. In the first place, investigators found that many of the oldest stars and star clusters in the galactic halo move in retrograde orbits—that is, they revolve around the galactic center in a direction opposite to that of most other stars. Such orbits suggest that the protogalaxy was quite clumpy and turbulent or that it captured sizable gaseous fragments whose matter was moving in different directions. Second, more refined dynamic models show that the protogalaxy would not have collapsed as smoothly as predicted by the simple model; instead the densest parts would have fallen inward much faster than more rarefied regions.

Third, the time scale of galaxy formation may have been longer than that deduced by Eggen and his colleagues. Exploding supernovae, plasma winds pouring from massive, short-lived stars and energy from an active galactic nucleus are all possible factors. The galaxy may also have subsequently rejuvenated itself by absorbing large inflows of pristine intergalactic gas and by capturing small, gas-rich satellite galaxies.

Several investigators have attempted to develop scenarios consistent with the findings. In 1977 Alar Toomre of the Massachusetts Institute of Technology postulated that most galaxies form from the merger of several large pieces rather than from the collapse of a single gas cloud. Once merged in this way, according to Toomre, the gas cloud collapsed and evolved into the Milky Way now seen. Leonard Searle of the Carnegie Institution and Robert J. Zinn of Yale University have suggested a somewhat different picture, in which many small bits and pieces coalesced. In the scenarios proposed by Toomre and by Searle and Zinn, the ancestral fragments may have evolved in chemically unique ways. If stars began to shine and supernovae started to explode in different fragments at different times, then each ancestral fragment would have its own chemical signature. Recent work by one of us (van den Bergh) indicates that such differences do indeed appear among the halo populations.

Discussion of the history of galactic evolution did not advance significantly beyond this point until the 1980s. At that time, workers became able to record more precisely than ever before extremely faint images. This ability is critically important because the physical theories of stellar energy production—and hence the lifetimes and ages of stars—are most secure for so-called main-sequence stars. Such stars burn hydrogen in their cores; in general, the more massive the star, the more quickly it completes its main-sequence life. Unfortunately, this fact




**COLOR-LUMINOSITY DIAGRAMS can be used to determine stellar ages. The one above compares the plots of stars in globular clusters NGC 288 and NGC 362 with age tracks (*black lines*) generated by stellar evolution models. The color index, expressed in magnitude units, is a measure of the intensity of blue wavelengths minus visual ones. In general, the brighter the star, the lower the color index; the trend reverses for stars brighter than about visual magnitude 19. The plots suggest the clusters differ in age by about three billion years. The temperature (inversely related to the color index) and luminosity have been set to equal those of NGC 288.**

means that within the halo the only remaining main-sequence stars are the extremely faint ones. The largest, most luminous ones, which have burned past their main-sequence stage, became invisible long ago. Clusters are generally used to determine age. They are crucial because their distances from the earth can be determined much more accurately than can those of individual stars.

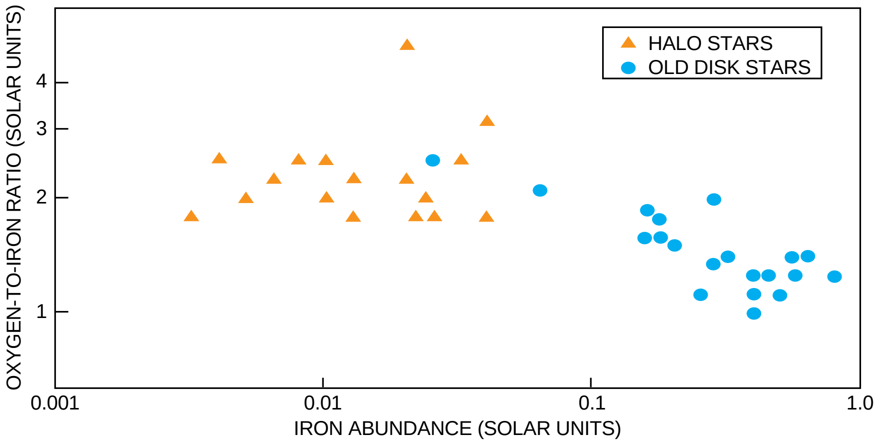
The technology responsible for opening the study of extremely faint halo stars is the charge-coupled device (CCD). This highly sensitive detector produces images electronically by converting light intensity into current. CCDs are far superior in most respects to photographic emulsions, although extremely sophisticated software, such as that developed by Peter B. Stetson of Dominion Astrophysical Observatory, is required to take full advantage of them. So used, the charge-coupled device has yielded a tenfold increase in the precision of measurement of color and luminosity of the faint stars in globular clusters.

Among the most important results of the CCD work done so far are more precise age estimates. Relative age data based on these new techniques have revealed that clusters whose chemistries suggest they were the first to be created after the big bang have the same age to within 500 million years of one another. The ages of other clusters, however, exhibit a greater spread.

The ages measured have helped researchers determine how long it took for the galactic halo to form. For instance, Michael J. Bolte, now at Lick Observatory, carefully measured the colors and luminosities of individual stars in the globular clusters NGC 288 and NGC 362 [*see illustration on opposite page*]. Comparison between these data and stellar evolutionary calculations shows that NGC 288 is approximately 15 billion years old and that NGC 362 is only about 12 billion years in age. This difference is greater than the uncertainties in the measurements. The observed age range indicates that the collapse of the outer halo is likely to have taken an order of magnitude longer than the amount of time first envisaged in the simple, rapid collapse model of Eggen, Lynden-Bell and Sandage.

Of course, it is possible that more than one model for the formation of the galaxy is correct. The Eggen–Lynden-Bell–Sandage scenario may apply to the dense bulge and inner halo. The more rarefied outer parts of the galaxy may have developed by the merger of fragments, along the lines theorized by Toomre or by Searle and Zinn. If so, then the clusters in the inner halo would have formed before those in the more tenuous outer regions. The process would account for some of the age differences found for the globular clusters. More precise modeling may have to await the improved image quality that modifications to the *Hubble Space Telescope* cameras will afford.

Knowing the age of the halo is, however, insufficient to ascertain a detailed formation scenario. Investigators need to know the age of the disk as well and then to compare that age with the halo's age. Whereas globular clusters are useful in determining the age of the halo, another type of celestial body—very faint white dwarf stars—can be used to determine the age of the disk. The absence of white dwarfs in the galactic disk near the sun sets a lower limit on the disk's age. White dwarfs, which are no longer producing radiant energy, take a long time to cool, so their

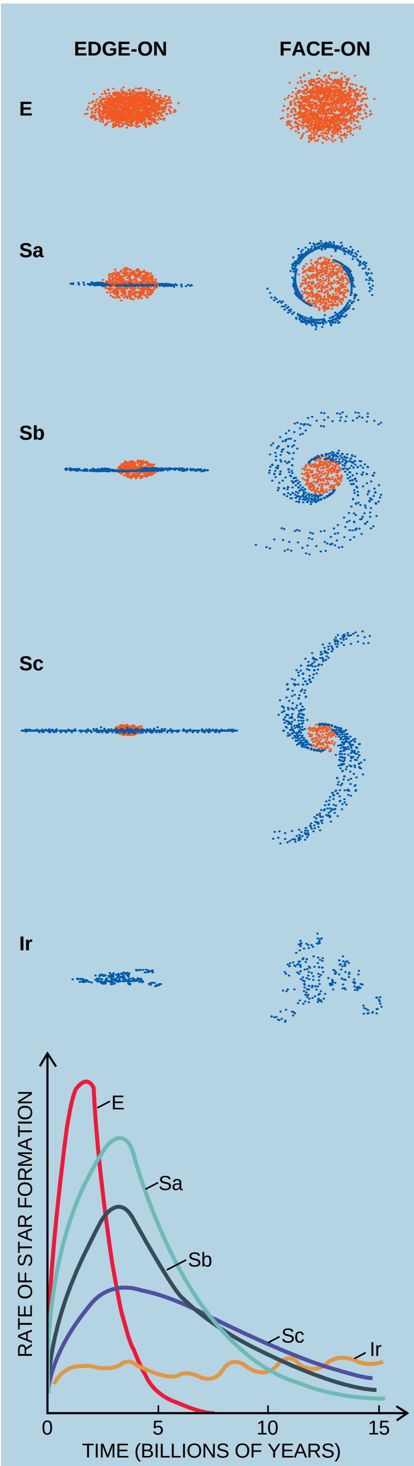



**OXYGEN-TO-IRON RATIOS as a function of metallicity (abundance of iron) for halo and old disk stars indicate different formation histories. The high ratios in metal-deficient halo stars suggest that those stars incorporated the oxygen synthesized in supernovae of types Ib, Ic and II. Type Ia supernovae seem to have contributed material only to the disk stars. Beatriz Barbuy and Marcia Erdelyi-Mendes of the University of São Paulo made the measurements.**

**GLOBULAR CLUSTERS, such as Messier 5 above, appear to be among the oldest objects known. They offer invaluable insight into the halo's formation some 15 billion years ago. The 100,000 or so stars exhibit similar abundances of heavy elements, implying that the gas cloud from which each arose was chemically homogeneous.**

absence means that the population in the disk is fairly young—less than about 10 billion years. This value is significantly less than the ages of clusters in the halo and is thus consistent with the notion that the bulk of the galactic disk developed after the halo.

It is, however, not yet clear if there is a real gap between the time when formation of the galactic halo ended and when creation of the old thick disk began. To estimate the duration of such a transitional period between halo and disk, investigators have compared the ages of the oldest stars in the disk with those of the youngest ones in the halo. The oldest known star clusters in the galactic disk, NGC 188 and NGC 6791, have ages of nearly eight billion years, according to Pierre Demarque and David B. Guenther of Yale and Elizabeth M. Green of the University of Arizona. Stetson and his colleagues and Roberto Buonanno of the Astronomical Observatory in Rome and his co-workers examined globular clusters in the halo population. They found the youngest globulars—Palomar 12 and Ruprecht 106—to be about 11 billion years old. If the few billion years' difference between the disk objects and the young globulars is real, then young globulars may be the missing links between the disk and halo populations of the galaxy.




**MORPHOLOGICAL CLASSIFICATION of galaxies (*top*) ranges from ellipticals (E) to spirals (subdivided into categories Sa, Sb and Sc) and irregulars (Ir). The history of star formation varies according to morphology (*bottom*). In elliptical galaxies, stars developed in an initial burst. Star formation in spirals was less vigorous but continues today. In most irregular galaxies the birthrate of stars has probably remained constant.**

At present, unfortunately, the relative ages of only a few globular clusters have been precisely estimated. As long as this is the case, one can argue that the Milky Way could have tidally captured Palomar 12 and Ruprecht 106 from the Magellanic Clouds. This scenario, proposed by Douglas N. C. Lin of the University of California at Santa Cruz and Harvey B. Richer of the University of British Columbia, would obviate the need for a long collapse time. Furthermore, the apparent age gap between disk and halo might be illusory. Undetected systematic errors may lurk in the age-dating processes. Moreover, gravitational interactions with massive interstellar clouds may have disrupted the oldest disk clusters, leaving behind only younger ones.

Determining the relative ages of the halo and disk reveals much about the sequence of the formation of the galaxy. On the other hand, it leaves open the question of how old the entire galaxy actually is. The answer would provide some absolute framework by which the sequence of formation events can be discerned. Most astronomers who study star clusters favor an age of some 15 to 17 billion years for the oldest clusters (and hence the galaxy).

Confidence that those absolute age values are realistic comes from the measured abundance of radioactive isotopes in meteorites. The ratios of thorium 232 to uranium 235, of uranium 235 to uranium 238 or of uranium 238 to plutonium 244 act as chronometers. According to these isotopes, the galaxy is between 10 and 20 billion years old. Although ages determined by such isotope ratios are believed to be less accurate than those achieved by comparing stellar observations and models, the consistency of the numbers is encouraging.

Looking at the shapes of other galaxies alleviates to some extent the uncertainty of interpreting the galaxy's evolution. Specifically, the study of other galaxies presents a perspective that is unavailable to us as residents of the Milky Way—an external view. We can also compare information from other galaxies to see if the processes that created the Milky Way are unique.

The most immediate observation one can make is that galaxies come in several shapes. In 1925 Edwin P. Hubble found that luminous galaxies could be arranged in a linear sequence according to whether they are elliptical, spiral or irregular [*see top illustration on this page*]. From an evolutionary point of view, elliptical galaxies are the most advanced. They have used up all (or almost all) of their gas to generate stars, which probably range in age from 10 to 15 billion years. Unlike spiral galaxies, ellipticals lack disk structures. The main differences between spiral and irregular galaxies is that irregulars have neither spiral arms nor compact nuclei.

The morphological types of galaxies can be understood in terms of the speed with which gas was used to create stars. Determining the rate of gas depletion would corroborate estimates of the Milky Way's age and history. Star formation in elliptical galaxies appears to have started off rapidly and efficiently some 15 billion years ago and then declined sharply. In most irregular galaxies the birth of stars has taken place much more slowly and at a more nearly constant rate. Thus, a significant fraction of their primordial gas still remains.

The rate of star formation in spirals seems to represent a compromise between that in ellipticals and that in irregulars. Star formation in spirals began less rapidly than it did in ellipticals but continues to the present day.

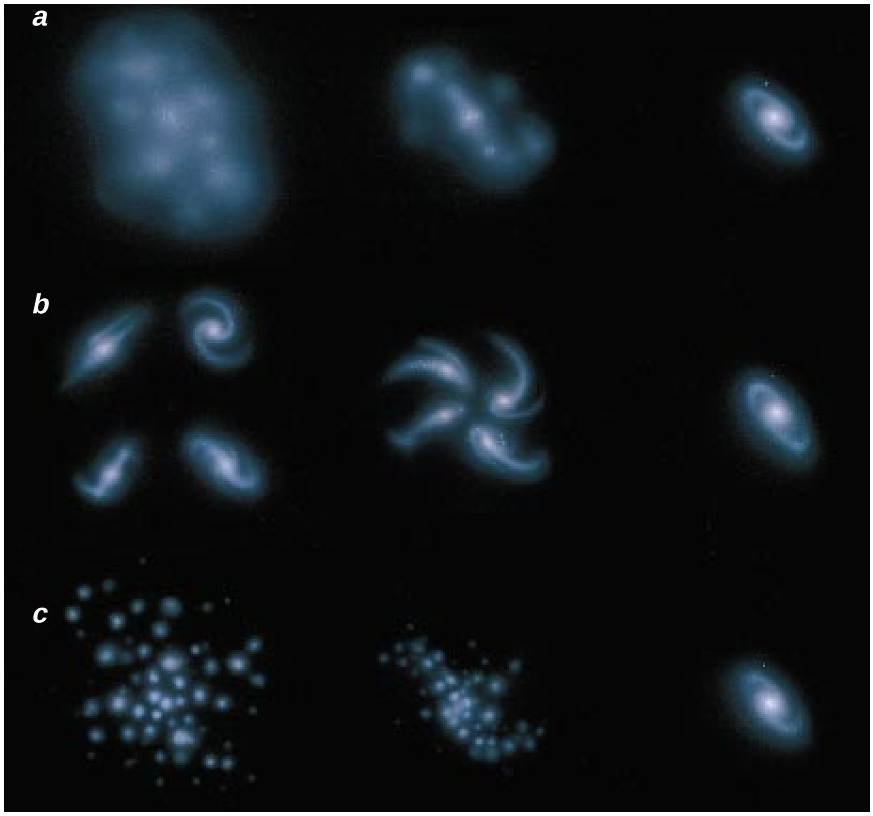
Spirals are further subdivided into categories Sa, Sb and Sc. The subdivisions refer to the relative size of the nuclear bulges and the degree to which the spiral arms coil. Objects of type Sa have the largest nuclear bulges and the most tightly coiled arms. Such spirals also contain some neutral hydrogen gas and a sprinkling of young blue stars. Sb spirals have relatively large populations of young blue stars in their spiral arms. The central bulge, containing old red stars, is less prominent than is the central bulge in spirals of type Sa. Finally, in Sc spirals the light comes mainly from the young blue stars in the spiral arms; the bulge population is inconspicuous or absent. The Milky Way is probably intermediate between types Sb and Sc.

Information from other spirals seems consistent with the data obtained for the Milky Way. Like those in our galaxy, the stars in the central bulges of other spirals arose early. The dense inner regions of gas must have collapsed first. As a result, most of the primordial gas initially present near the centers has turned into stars or has been ejected by supernova-driven winds.

There is an additional kind of evidence on which to build our understanding of how the Milky Way came into existence: the chemical composition of stars. This information helps to pinpoint the relative ages of stellar populations. According to stellar models, the chemistry of a star depends on when it formed. The chemical differences exist because first-generation stars began to "pollute" the protogalaxy with elements heavier than helium. Such so-called heavy elements, or "metals," as astronomers refer to them, were created in the interiors of stars or during supernova explosions. Examining the makeup of stars can provide stellar evolutionary histories that corroborate or challenge age estimates.

Different types of stars and supernovae produce different relative abundances of these metals. Researchers believe that most "iron-peak" elements (those closest to iron in the periodic table) in the galaxy were made in supernovae of type Ia. The progenitors of such supernovae are thought to be pairs of stars, each of which has a mass a few times that of the sun. Other heavy elements—the bulk of oxygen, neon, magnesium, silicon and calcium, among others—originated in supernovae that evolved from single or binaries of massive, short-lived stars. Such stars have initial masses of 10 to 100 solar masses and violently end their lives as supernovae of type Ib, Ic or II.

Stars that subsequently formed incorporated some of these heavy elements. For instance, approximately 1 to 2 percent of the mass of the sun consists of elements other than hydrogen or helium. Stars in nuclear bulges generally harbor proportionally more heavy elements than do stars in the outer disks and halos. The abundance




**MODELS OF GALAXY FORMATION fall into three general categories. In the Eggen–Lynden-Bell–Sandage model, the Milky Way formed by the rapid collapse of a single gaseous protogalaxy (*a*). In the Toomre model, several large aggregates of gas merged (*b*). The Searle-Zinn picture is similar to the Toomre model except that the ancestral fragments consisted of much smaller but more numerous pieces (*c*).**

**MESSIER 83 is a typical type Sc spiral galaxy. The Milky Way probably has a similar appearance, although its arms may be somewhat more tightly coiled.**

**LARGE MAGELLANIC CLOUD is one of the Milky Way's two largest satellite galaxies. Slowly spiraling into the Milky Way, the cloud will briefly rejuvenate our galaxy at some time in the distant future.**

of heavy elements decreases gradually by a factor of 0.8 for every kiloparsec (3,300 light-years) from the center to the edge of the Milky Way disk. Some 70 percent of the 150 or so known globular clusters in the Milky Way exhibit an average metal content of about one twentieth that of the sun. The remainder shows a mean of about one third that of the sun.

Detailed studies of stellar abundances reveal that the ratio of oxygen to iron-peak elements is larger in halo stars than it is in metal-rich disk stars [*see upper illustration on page 75*]. This difference suggests the production of heavy elements during the halo phase of galactic evolution was dominated by supernovae of types Ib, Ic and II. It is puzzling that iron-producing type Ia supernovae, some of which are believed to have resulted from progenitor stars with lifetimes as short as a few hundred million years, did not contribute more to the chemical mixture from which halo stars and some globular clusters formed. This failure would seem to imply that the halo collapsed very rapidly—before supernovae of type Ia could contribute their iron to the halo gas.

That idea, however, conflicts with the four-billion-year age spread observed among galactic globular clusters, which implies that the halo collapsed slowly. Perhaps supernova-driven galactic winds swept the iron-rich ejecta from type Ia supernovae into intergalactic space. Such preferential removal of the ejecta of type Ia supernovae might have occurred if supernovae of types Ib, Ic and II exploded primarily in dense gas clouds. Most of type Ia supernovae then must have detonated in less dense regions, which are more easily swept out by the galactic wind.

Despite the quantity of data, information about metal content has proved insufficient to settle the controversy concerning the time scale of disk and halo formation. Sandage and his colleague Gary A. Fouts of Santa Monica College find evidence for a rather monolithic collapse. On the other hand, John E. Norris and his collaborators at the Australian National Observatory, among others, argue for a significant decoupling between the formation of halo and disk. They also posit a more chaotic creation of the galaxy, similar to that envisaged by Searle and Zinn.

Such differences in interpretation often reflect nearly unavoidable effects arising from the way in which particular samples of stars are selected for study. For example, some stars exhibit chemical compositions similar to those of "genuine" halo stars, yet they have kinematics that would associate them with one of the subcomponents of the disk. As vital as it is, chemical information alone does not resolve ambiguities about the formation of the galactic halo and disk. "Cats and dogs may have the same age and metallicity, but they are still cats and dogs" is the way Bernard Pagel of the Nordic Institute for Theoretical Physics in Copenhagen puts it.

As well as telling us about the past history of our galaxy, the disk and halo also provide insight into the Milky Way's probable future evolution. One can easily calculate that almost all of the existing gas will be consumed in a few billion years. This estimate is based on the rate of star formation in the disks of other spirals and on the assumption that the birth of stars will continue at its present speed. Once the gas has been depleted, no more stars will form, and the disks of spirals will then fade. Eventually the galaxy will consist of nothing more than white dwarfs and black holes encapsulated by the hypothesized dark matter corona.

Several sources of evidence exist for such an evolutionary scenario. In 1978 Harvey R. Butcher of the Kapteyn Laboratory in the Netherlands and Augustus Oemler, Jr., of Yale found that dense clusters of galaxies located about six billion light-years away still contained numerous spiral galaxies. Such spirals are, however, rare or absent in nearby clusters of galaxies. This observation shows that the disks of most spirals in dense clusters must have faded to invisibility during the past six billion years. Even more direct evidence for the swift evolution of galaxies comes from the observation of so-called blue galaxies. These galaxies are rapidly generating large stars. Such blue galaxies seem to be less common now than they were only a few billion years ago.

Of course, the life of spiral galaxies can be extended. Copious infall of hydrogen from intergalactic space might replenish the gas supply. Such infall can occur if a large gas cloud or another galaxy with a substantial gas reservoir is nearby. Indeed, the Magellanic Clouds will eventually plummet into the Milky Way, briefly rejuvenating our galaxy. Yet the Milky Way will not escape its ultimate fate. Like people and civilizations, stars and galaxies leave behind only artifacts in an evolving, ever dynamic universe.

FURTHER READING

GALACTIC ASTRONOMY: STRUCTURE AND KINEMATICS. Dimitri Mihalas and James Binney. W. H. Freeman and Company, 1981.

THE MILKY WAY AS A GALAXY. Gerard Gilmore, Ivan R. King and Pieter C. van der Kruit. University Science Books, Mill Valley, Calif., 1990.

THE FORMATION AND EVOLUTION OF STAR CLUSTERS. Edited by Kenneth Janes. Astronomical Society of the Pacific, 1991.

THE STELLAR POPULATIONS OF GALAXIES. Edited by B. Barbuy and A. Renzini. Kluwer Academic Publishers, Dordrecht, Holland, 1992.

# Carbohydrates in Cell Recognition

*Telltale surface sugars enable cells to identify and interact with one another. New drugs aimed at those carbohydrates could stop infection and inflammation*

by Nathan Sharon and Halina Lis

In 1952 Aaron Moscona of the University of Chicago separated the cells of a chick embryo by incubating them in an enzyme solution and swirling them gently. The cells did not remain apart; they coalesced into a new aggregate. Moreover, Moscona saw that when retinal cells and liver cells were allowed to coalesce in this way, the retinal cells always migrated to the inner part of the cellular mass. Three years later Philip L. Townes and Johannes Holtfreter of the University of Rochester performed a similar experiment with cells from amphibian embryos, which re-sorted themselves into tissue layers like those from which they had come.

Those experiments and countless other observations testify to the keen ability of cells to recognize one another and to respond accordingly. Sperm, for example, can distinguish eggs of their own species from those of others, and they will bind only with the former. Some bacteria settle preferentially in the intestinal or urinary tract; others fancy different organs.

It is not surprising, then, that decoding the language of cellular interactions holds profound interest for researchers in many areas of biology and medicine. Although we still do not understand the chemical basis of most cell-recognition phenomena, clear explanations for some have emerged during the past decade. Proteins, which mediate most of the chemical reactions inside living organisms, appear on the cell surface as well, and they certainly play a part.

Yet the accumulating evidence also suggests that in many cases carbohydrates (frequently referred to as sugars) are the primary markers for cell recognition. Discoveries about the involvement of specific sugars in recognition will have practical applications to the prevention and treatment of a variety of ailments, including cancer.

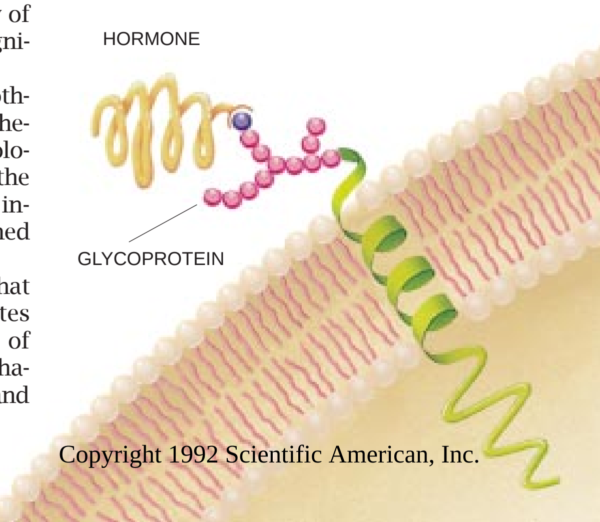
Biologists generally accept that cells recognize one another through pairs of complementary structures on their surfaces: a structure on one cell carries encoded biological information that the structure on the other cell can decipher. That idea represents an extension of the lock-and-key hypothesis formulated in 1897 by Emil Fischer, the noted German chemist. He used it to explain the specificity of interactions between enzymes and their substrates. Pioneering immunologist Paul Ehrlich extended it in 1900 to account for the highly specific reactions of the immune system, and in 1914 Frank Rattray Lillie of the University of Chicago invoked it to describe recognition between sperm and eggs.

By the 1920s the lock-and-key hypothesis had become one of the central theoretical assumptions of cellular biology. Yet for many years thereafter, the nature and identity of the molecules involved in cellular recognition remained a complete mystery.

To most biologists, the idea that the molecules might be carbohydrates seemed farfetched. That large class of compounds consists of monosaccharides (simple sugars such as glucose and fructose) and of oligosaccharides and polysaccharides, which are composed of linked monosaccharides. Until the late 1960s, carbohydrates were thought to serve only as energy sources (in the forms of monosaccharides and storage molecules such as the polysaccharide starch) and as structural materials (the polysaccharides cellulose in plants and chitin in the exoskeletons of insects). The two other major classes of biological materials—nucleic acids, which carry genetic information, and proteins—were obviously far more versatile. By comparison, carbohydrates looked like dull, second-class citizens.

Interest in carbohydrates was further discouraged by the extraordinary complexity of their structures. In contrast to the nucleotides in nucleic acids and the amino acids in proteins, which can interconnect in only one way, the monosaccharide units in oligosaccharides and polysaccharides can attach to one another at multiple points. Two identical monosaccharides can bond to form 11 different disaccharides, whereas two amino acids can make only one dipeptide. Even a small number of monosaccharides can create a staggering diversity of compounds, including many with branching structures. Four different nucleotides can make only 24 distinct tet-

NATHAN SHARON and HALINA LIS have been members of the biophysics department of the Weizmann Institute of Science in Rehovot, Israel, for more than 30 years. During most of that time, they have collaborated closely on the study of complex carbohydrates and lectins, proteins that bind selectively to carbohydrates. In addition to their many joint scientific papers, they have written several widely cited review articles on those topics, as well as the recent book *Lectins* (Chapman & Hall, London). This is Sharon's fifth article for *Scientific American*.



ranucleotides, but four different monosaccharides can make 35,560 unique tetrasaccharides.

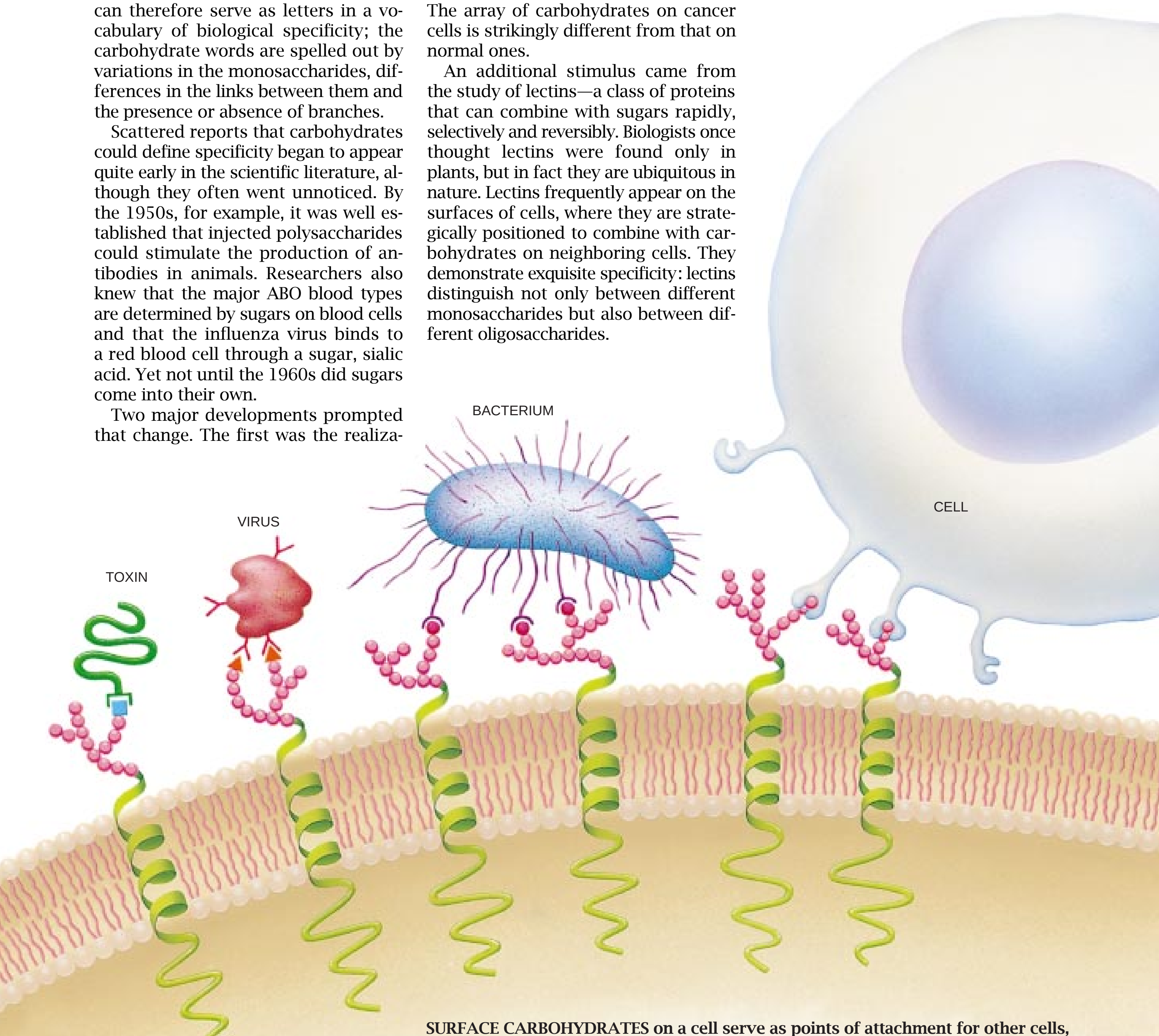
This potential for structural diversity is the bane of the carbohydrate chemist, but it is a boon to cells: it makes sugar polymers superbly effective carriers of information. Carbohydrates can carry much more information per unit weight than do either nucleic acids or proteins. Monosaccharides can therefore serve as letters in a vocabulary of biological specificity; the carbohydrate words are spelled out by variations in the monosaccharides, differences in the links between them and the presence or absence of branches.

Scattered reports that carbohydrates could define specificity began to appear quite early in the scientific literature, although they often went unnoticed. By the 1950s, for example, it was well established that injected polysaccharides could stimulate the production of antibodies in animals. Researchers also knew that the major ABO blood types are determined by sugars on blood cells and that the influenza virus binds to a red blood cell through a sugar, sialic acid. Yet not until the 1960s did sugars come into their own.

Two major developments prompted that change. The first was the realization that all cells carry a sugar coat. This coat consists for the most part of glycoproteins and glycolipids, two types of complex carbohydrates in which sugars are linked to proteins and lipids (fats), respectively. Several thousands of glycoprotein and glycolipid structures have been identified, and their number grows almost daily. This diversity is surely significant: the repertoire of surface structures on a cell changes characteristically as it develops, differentiates or sickens. The array of carbohydrates on cancer cells is strikingly different from that on normal ones.

An additional stimulus came from the study of lectins—a class of proteins that can combine with sugars rapidly, selectively and reversibly. Biologists once thought lectins were found only in plants, but in fact they are ubiquitous in nature. Lectins frequently appear on the surfaces of cells, where they are strategically positioned to combine with carbohydrates on neighboring cells. They demonstrate exquisite specificity: lectins distinguish not only between different monosaccharides but also between different oligosaccharides.

A landmark discovery about the role of lectin-carbohydrate interactions in cell recognition came from the work of G. Gilbert Ashwell of the National Institutes of Health and Anatol Morell of the Albert Einstein College of Medicine. In 1968 they enzymatically removed a few sialic acid molecules from certain




**SURFACE CARBOHYDRATES** on a cell serve as points of attachment for other cells, infectious bacteria, viruses, toxins, hormones and many other molecules. In this way, carbohydrates mediate the migration of cells during embryo development, the process of infection and other phenomena. Compounds consisting of carbohydrates that are chemically linked to proteins are called glycoproteins; those in which the carbohydrates are linked to fats are glycolipids.

blood plasma glycoproteins, then injected the glycoproteins into rabbits. Ordinarily, such molecules would persist in the animals' circulation for some time, but the sialic acid–deficient molecules quickly disappeared.

Ashwell and Morell found that the glycoproteins ended up in the liver. The removal of the sialic acids had unmasked galactose in the glycoproteins, and the exposed galactoses had attached to a lectin on the liver cells. Subsequently, the researchers learned that if they removed both the sialic acids and the uncovered galactoses from the glycoproteins, the rate at which the molecules were eliminated from the blood returned to normal. From those results, Ashwell and Morell concluded that carbohydrate side chains on proteins may serve as markers for identifying which ones should be removed from the circulation and eventually degraded.

Like the surface carbohydrates, the surface lectins go through changes that coincide with a cell's physiological and pathological states. For instance, in 1981 Reuben Lotan and Abraham Raz of the Weizmann Institute of Science showed that tumor cells from mice and humans carry a surface lectin not found on normal cells. They and other researchers later proved that this lectin is involved in the development of metastases.

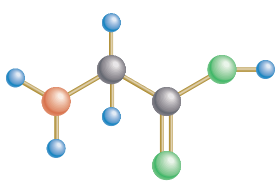
A striking recent illustration of the role of surface sugars and the molecules that bind to them comes from studies of embryo formation by Senitiroh Hakomori of the Fred Hutchinson Cancer Research Center in Seattle and by Ten Feizi of the Clinical Research Center in Harrow, England. Working with mouse embryos, they have shown that as a fertilized egg divides, the carbohydrate structures on the resulting embryonic cells change in characteristic ways. One of the carbohydrates is a trisaccharide known both as stage-specific embryonic antigen 1 (SSEA-1) and as Lewis$^{x}$ (Le$^{x}$). It appears at the eight- to 16-cell stage, just as the embryo compacts from a group of loose cells into a smooth ball.

Hakomori's group has shown that a soluble compound carrying multiple units of the same trisaccharide inhibits the compaction process and disrupts embryogenesis. Closely related but structurally different carbohydrates have no effect. Thus, the Le$^{x}$ trisaccharide appears to play a part in compaction.

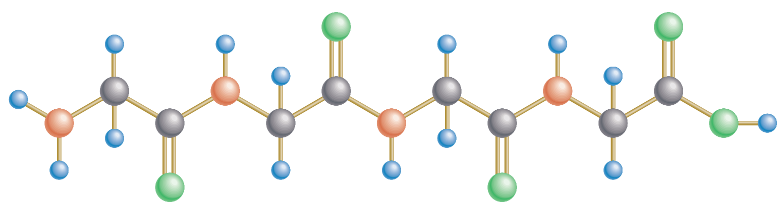
Adhesive carbohydrates are therefore essential to embryonic development. As research continues, their role in that process will become more detailed. Today the two best-understood phenomena of that type are microbial adhesion to host cells and the adhesion

## The Complexity of Carbohydrate Structures

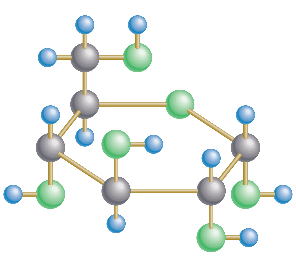
Carbohydrates, nucleic acids and proteins all carry biological information in their structures. Yet carbohydrates offer the highest capacity for carrying information because they have the greatest potential for structural variety. Their component molecules, monosaccharides, can interconnect at several points to form a wide variety of branched or linear structures; in the example below, the branching carbohydrate is only one of many possible structures that can be made from four identical glucose molecules. The amino acids in proteins as well as the nucleotides in nucleic acids can form only linear assemblies, which restricts their diversity. The peptide (protein fragment) shown here is the only one possible made from four molecules of the amino acid glycine.



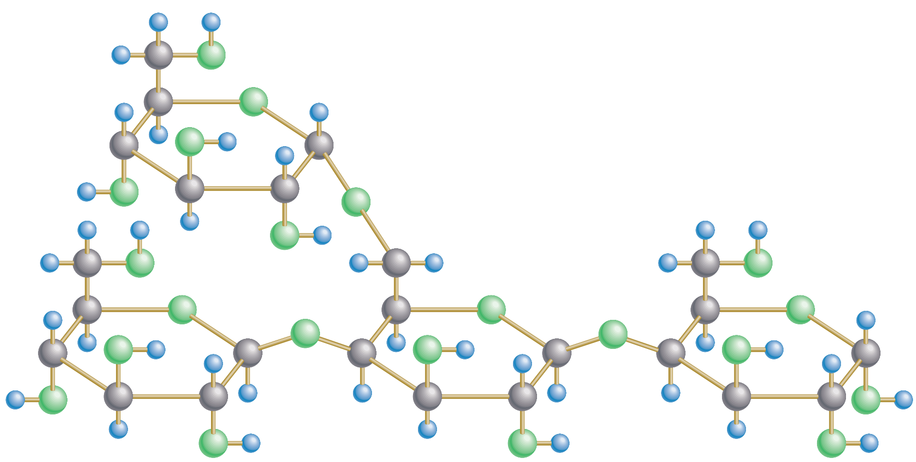
AMINO ACID (GLYCINE)



PEPTIDE (TETRAGLYCINE)



MONOSACCHARIDE (GLUCOSE)



OLIGOSACCHARIDE (BRANCHED TETRAGLUCOSE)

CARBON OXYGEN NITROGEN HYDROGEN

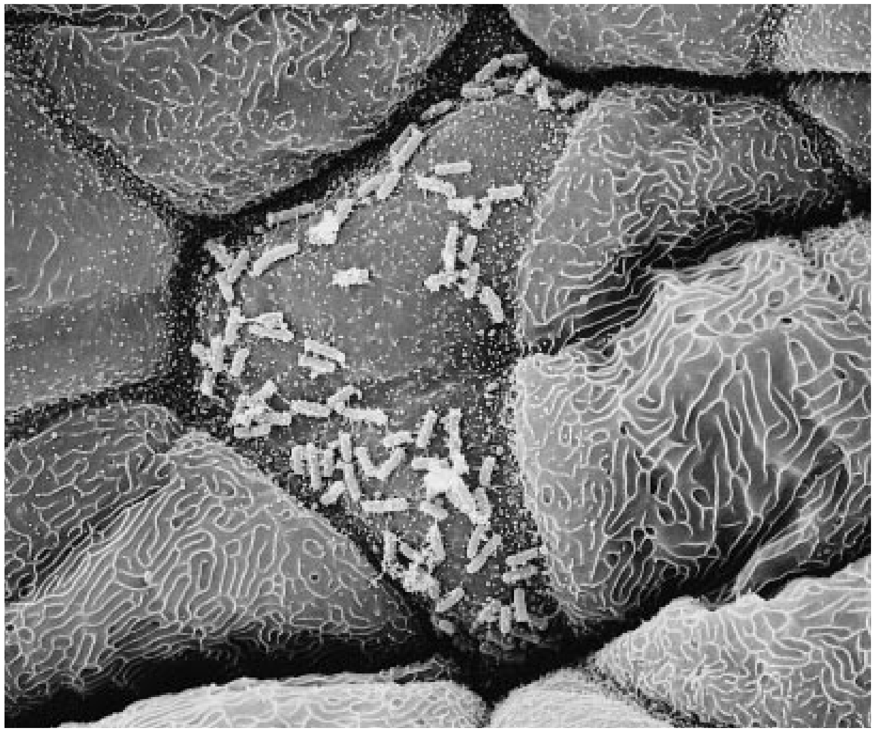
of white blood cells to blood vessels. The more thoroughly characterized of these interactions is microbial adhesion, which has been studied for nearly two decades and serves as a model for other forms of carbohydrate-mediated cell recognition.

To cause disease, viruses, bacteria or protozoa must be able to stick to at least one tissue surface in a susceptible host. Infectious agents lacking that ability are swept away from potential sites of infection by the body's normal cleansing mechanisms. Microorganisms in the upper respiratory tract, for example, may be swallowed and eventually destroyed by stomach acid; those in the urinary tract may be flushed out in the urine.

The first clues about the mechanism of bacterial adhesion sprang from a series of pioneering studies by J. P. Duguid of Ninewells Hospital Medical School in Dundee that began in the 1950s. Duguid demonstrated that many strains of *Escherichia coli* (a bacterial denizen of the intestines that can also colonize other tissues) and related bacteria adhered to cells from the epithelial lining of tissues and to erythrocytes, or red blood cells. In the presence of sticky bacteria, the erythrocytes would clump together—a phenomenon called hemagglutination. (Researchers still routinely use hemagglutination as a simple test for the adhesion of bacteria to animal cells.) To learn how the bacteria bound to the cells, Duguid exposed them to a wide range of compounds. He found that only the monosaccharide mannose and very similar sugars could inhibit hemagglutination.

Duguid also made the important observation that the bacterial strains responsible for mannose-sensitive hemagglutination had submicroscopic, hairlike appendages on their surfaces. These structures were five to 10 nanometers in diameter and several hundreds of nanometers long. He called them fimbriae, from the Latin word for fringe. Almost simultaneously, Charles C. Brinton, Jr., of the University of Pittsburgh described the same structures and named them pili, from the Latin word for hairs. Both terms are still in use.

Later, starting around 1970, Ronald J. Gibbons of the Forsyth Dental Center in Boston and his colleagues began reporting on the selective adhesion of bacteria to niches within the oral cavity. Gibbons observed that *Actinomyces naeslundii* colonizes both the epithelial surfaces of infants without teeth and the teeth of children and adults. Conversely, the related bacterium *A. viscosus* does not appear in the mouth until the teeth erupt from the gums; it exhibits a preference for teeth rather than oral epithelial surfaces.



**BACTERIA ADHERE to tissues selectively. Hairlike protrusions called fimbriae on the bacteria bind exclusively to certain surface carbohydrates. These interactions determine which tissues are susceptible to bacterial invasion. Rod-shaped *Escherichia coli* bacteria are shown here on tissue from the urinary tract.**

Today it is clear that the tissue specificity of bacterial adhesion is a general phenomenon. For example, *E. coli,* the most common cause of urinary tract infections, is abundant in tissues surrounding the ducts that connect the kidneys and the bladder, yet it is seldom found in the upper respiratory tract. In contrast, group A streptococci, which colonize only the upper respiratory tract and skin, rarely cause urinary tract infections.

Bacterial adhesion varies not only between tissues but also between species and sometimes between individuals of the same species, depending on their age, genetic makeup and health. In the early 1970s R. Sellwood and Richard A. Gibbons and their colleagues at the Institute for Research on Animal Diseases in Compton, England, studied the infectivity of the K88 strain of *E. coli*. Because those bacteria cause diarrhea in piglets, they are a costly nuisance for farmers. Gibbons's group found that the K88 bacteria adhered to the intestinal cells of susceptible piglets but not to those of adult pigs or of humans, which the bacteria cannot infect. Bacterial mutants that had lost the ability to bind to intestinal cells proved unable to infect the animals.

Moreover, as Gibbons's work showed, some piglets had a genetic resistance to K88 bacteria: even potentially virulent bacteria could not bind to cells from their intestines. By selecting genetically immune piglets for breeding, farmers were able to obtain K88-resistant progeny.

The gonorrhea organism, *Neisseria gonorrhoeae,* serves as another example of species and tissue specificity. It adheres to human cells of the genital and oral epithelia but not to cells from other organs or other animal species. That fact explains why humans are the exclusive host for *N. gonorrhoeae* and why other animals do not contract gonorrhea.

A strong impetus to the study of bacterial adhesion was a proposal made in 1977 by Itzhak Ofek of Tel Aviv University, David Mirelman of our department at the Weizmann Institute and one of us (Sharon). We suggested that bacterial adhesion is mediated by surface lectins on bacteria that bind to complementary sugars on host cells. That idea has proved to be generally valid. Work in many laboratories has shown that bacteria produce lectins specific for certain carbohydrates and that the bacteria depend on those lectins for adhering to a host's tissue as the first step in the process of infection.

Bacterial lectins have already been the focus of much study, although far more remains to be done. The best-characterized lectins are the type 1 fimbriae of *E. coli,* which bind preferentially to surface glycoproteins containing mannose. Other research on *E. coli* during the past decade, primarily by Catharina Svanborg-Edén and her colleagues while working at the University of Göteborg, has described in detail the P fimbriae. Those fimbriae interact specifically with the P blood-group substance, an extremely common glycolipid containing the disaccharide galabiose. Research groups led by Karl-Anders Karlsson of the University of Göteborg and Victor Ginsburg of the NIH have mapped the specificities of lectins from a wide range of other bacterial species and strains.

These studies have shown that bacteria do not bind solely to the ends of surface carbohydrates—they can also sometimes bind to sugars located within the structure. Furthermore, different bacteria may bind to different parts of the same carbohydrate. Occasionally, only one face of an oligosaccharide may be exposed on a particular cell, and as a result the cell will bind bacteria of one kind and not other ones. The ability of cell-surface sugars to serve as attachment sites therefore depends not only on the presence of these sugars but also on their accessibility and their mode of presentation.

Considerable experimental evidence now greatly strengthens the conclusion that the binding of bacteria to host cell-surface sugars initiates infection. For example, uroepithelial cells from those rare individuals who lack the P blood-group substance do not bind to P-fimbriated *E. coli*. Such individuals are much less susceptible to infections from those bacteria than the rest of the population is. Experiments have shown, however, that the bacteria will bind if the epithelial cells are first coated with a synthetic glycolipid containing galabiose.

Similarly, intestinal cells from piglets that are resistant to the diarrhea-causing K88 *E. coli* lack the large carbohydrate to which the bacteria bind. Although the exact structure of this carbohydrate has not yet been elucidated, it is known to be present in susceptible piglets but absent in adult pigs. This explains why the bacteria were unable to attach to and colonize the intestines of the adult pigs, while they caused infection in the young piglets.

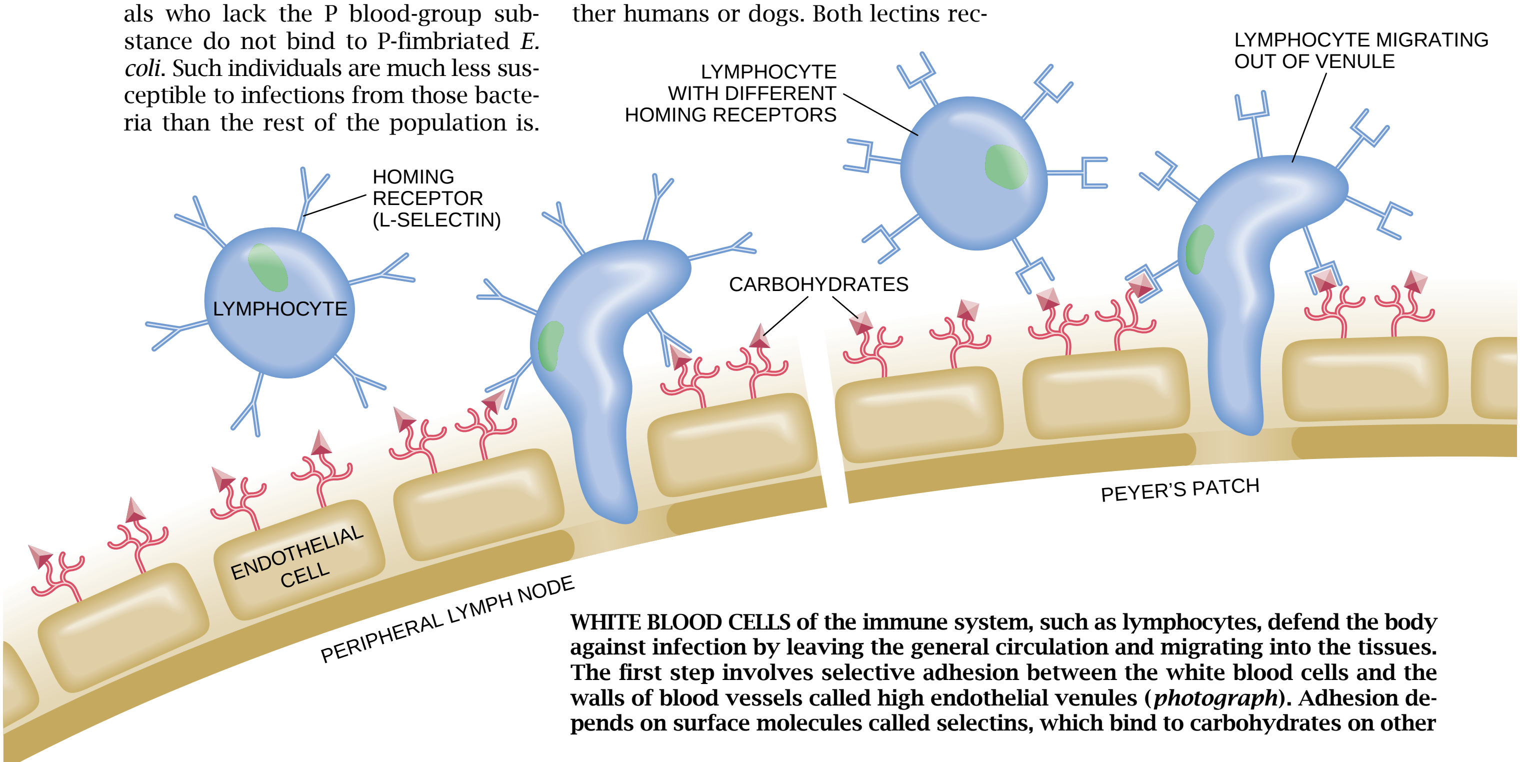
Another interesting case is that of the K99 strain of *E. coli*. Like the K88 strain, K99 causes diarrhea in farm animals but not in humans. It is less specific than K88, however, because it infects young calves and lambs as well as piglets. The K99 bacteria bind specifically to an unusual glycolipid that contains *N*-glycoloylneuraminic acid (a special type of sialic acid) linked to lactosylceramide. This glycolipid, which is present in piglets, calves and lambs, is absent from the cells of adult pigs and humans, which instead contain *N*-acetylneuraminic acid, a nonbinding analogue of sialic acid. Here a small difference between two highly similar sugars—the replacement of an acetyl group by a glycoloyl group—is readily detected by the bacteria and explains the host range of infection by the organism.

Further confirmation of the above conclusions was recently obtained in experiments on two fimbrial lectins from *E. coli* that infect the urinary tracts of either humans or dogs. Both lectins recognize galabiose, yet one binds only to the human uroepithelial cells and the other only to canine cells; the galabiose-bearing glycolipids on the surface of the cells are presented in subtly different ways. Those lectin-binding patterns accord with the host specificities of the *E. coli* strains.

Because bacterial adhesion is so critical to infection, medical researchers are seriously considering the use of sugars for prevention and treatment. Sugars that selectively inhibited adhesion could act as molecular decoys, intercepting pathogenic bacteria before they reached their tissue targets. Urinary tract infections have been the focus of particular attention because they are second only to respiratory infections in frequency.

In collaboration with Ofek, Moshe Aronson of Tel Aviv University and Mirelman, we performed the first study along those lines in 1979. We injected a mannose-specific strain of *E. coli* into the urinary bladder of mice. In some animals, we also injected methyl alpha-mannoside, a sugar that in the test tube inhibited bacterial adhesion to epithelial cells. The presence of the sugar reduced the colonization of the urinary tract by bacteria.

Svanborg-Edén has performed analogous experiments with P-fimbriated *E. coli* that infect the kidneys of mice. She incubated the bacteria in solutions of globotetraose, a sugar found in the glycolipid of kidney cells. When she subsequently injected those bacteria into




**WHITE BLOOD CELLS of the immune system, such as lymphocytes, defend the body against infection by leaving the general circulation and migrating into the tissues. The first step involves selective adhesion between the white blood cells and the walls of blood vessels called high endothelial venules (*photograph*). Adhesion depends on surface molecules called selectins, which bind to carbohydrates on other**

mice, they persisted in the kidneys for less time than untreated bacteria did. James A. Roberts of Tulane University obtained similar results in experiments on monkeys: incubation of P-fimbriated *E. coli* with a galabioselike sugar significantly delayed the onset of urinary tract infections.

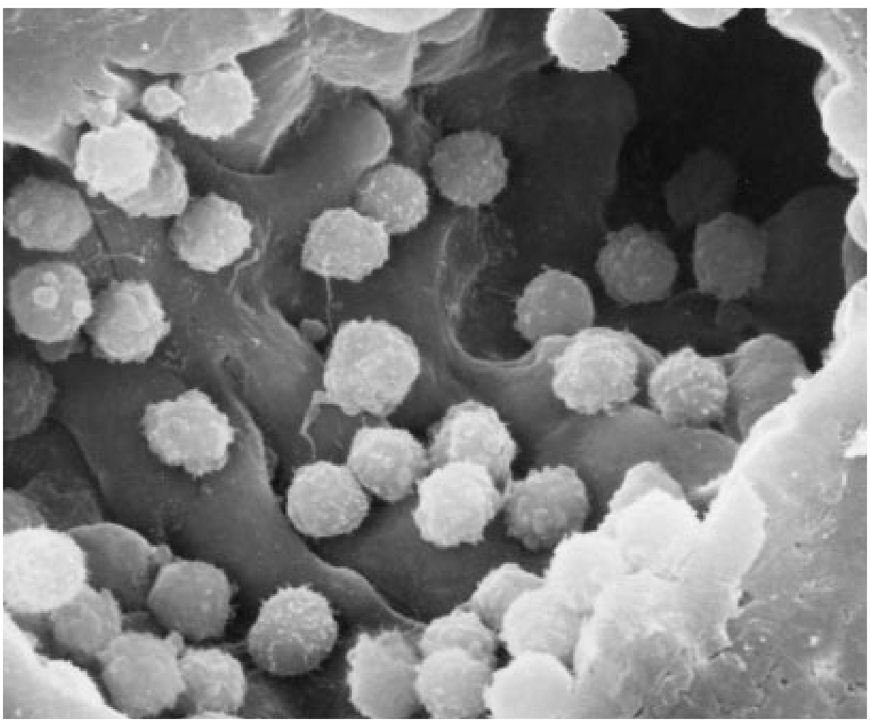
Glycopeptides can also interfere with the binding of bacteria to host tissues. In 1990 Michelle Mouricout of the University of Limoges in France and her co-workers showed that injections of glycopeptides taken from the blood plasma of cows can protect newborn calves from lethal doses of *E. coli*. The glycopeptides, which contain sugars for which the bacteria have affinities, decrease the adhesion of the bacteria to the intestines of treated animals.

Indeed, to interfere with bacterial adhesion, one need not even use a carbohydrate—any agent that competitively binds to either the bacterial lectin or the host cell's surface carbohydrate will do. For example, Edwin H. Beachey and his colleagues at the Veterans Administration Medical Center and the University of Tennessee at Memphis have used antibodies against mannose to prevent certain mannose-specific *E. coli* from infecting mice. The antibodies bind to mannose on the cells, thereby blocking the sites of bacterial attachment.

Those successful experiments make a clear case for antiadhesive therapies against microbial diseases. The application of this approach in humans is now the subject of intense research. Further studies of the sugars on host cells and of bacterial lectins should lead to the design of better adhesion inhibitors. One point about the approach is certain: because different infectious agents—even different bacteria within the same strain—can have a wide variety of carbohydrate specificities, a cocktail of inhibitors will undoubtedly be necessary to prevent or treat the diseases.

Carbohydrate-directed interactions between cells are not restricted to pathological phenomena; they are also crucially important to the healthy operation of the immune system. The immune system has many parts, but its most important soldiers are the cells called leukocytes. This group includes an array of diverse white blood cells—lymphocytes, monocytes and neutrophils—that act jointly to eliminate bacteria and other intruders and to mediate the inflammation response in injured tissues. All these cells circulate in the blood, but they accomplish their major functions in the extravascular spaces.

The picture emerging from research is that the inner lining of blood vessels, called the endothelium, actively snares white blood cells and guides them to where they are needed. This process requires an exquisitely regulated recognition between the circulating leukocytes and the endothelial cells.



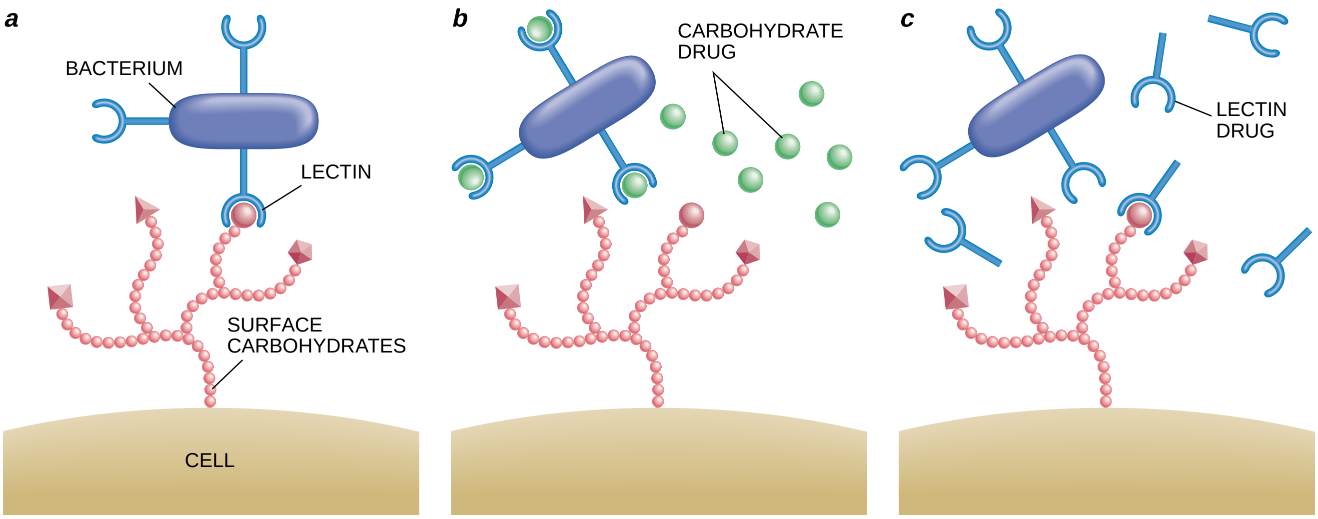
**cells. The L-selectins, or homing receptors, on lymphocytes determine the endothelial cells to which a lymphocyte will stick: for example, some adhere only in peripheral lymph nodes or to the Peyer's patches in the intestines. After a lymphocyte has attached to the endothelium, it can migrate out of the blood vessel.**

Such recognition seems to be mediated by a family of structurally related lectins. Because this field of research is so new and because different laboratories often identify the same adhesion molecules simultaneously, the nomenclature is still in a somewhat chaotic state. Most researchers refer to these molecules as selectins because they mediate the selective contact between cells. Another name in vogue is LEC-CAMs, an acronym for leukocyte-cell (or lectin) adhesion molecules.

The selectins are highly asymmetric composite proteins, with an unusual mosaic architecture. They consist of three types of functional domains: one domain anchors the selectin in the cell membrane, and a second makes up most of the body of the molecule. The third domain, located at the extracellular tip of the molecule, structurally resembles animal lectins that work only in the presence of calcium ions. The binding of carbohydrate ligands to that domain is central to the function of selectins in interactions between cells.

About 10 years ago Eugene C. Butcher and Irving L. Weissman of Stanford University laid the foundation for our current understanding of how selectins (which were then unknown) direct lymphocyte traffic. Lymphocytes are unique among leukocytes in that they continuously patrol the body in search of foreign antigens (immunologically significant molecules) from bacteria, viruses and the like. For that purpose, lymphocytes leave the blood vessels and migrate through the lymph nodes, the tonsils, the adenoids, the Peyer's patches in the intestines or other secondary lymphoid organs. Different lymphocytes migrate selectively, or home, toward particular organs. To exit from the bloodstream, lymphocytes must first bind to specialized submicroscopic blood vessels less than 30 microns in diameter, known as high endothelial venules.

Using an assay technique developed by Hugh B. Stamper, Jr., and Judith J. Woodruff of the State University of New York in Brooklyn, Butcher and Weissman observed that the homing specificity of mouse lymphocytes is dictated by their selective interaction with the high endothelial venules in their targeted organs. Butcher and Weissman then developed a monoclonal antibody, MEL-14, that bound only to mouse lymphocytes that went to the peripheral lymph nodes. On slices of tissue, the antibody blocked the attachment of the lymphocytes to high endothelial venules




**BLOCKING BACTERIAL ATTACHMENT is one strategy for combating infections. As a prelude to infection, bacterial surface proteins called lectins attach to surface carbohydrates on susceptible host cells (*a*). Drugs containing similar carbohydrates could prevent the attachment by binding to the lectins (*b*). Alternatively, drugs consisting of lectinlike molecules could have the same effect by innocuously occupying the binding sites on the carbohydrates (*c*).**

from those tissues but not from other lymph organs. When injected into mice, MEL-14 inhibited the migration of lymphocytes into the peripheral lymph nodes.

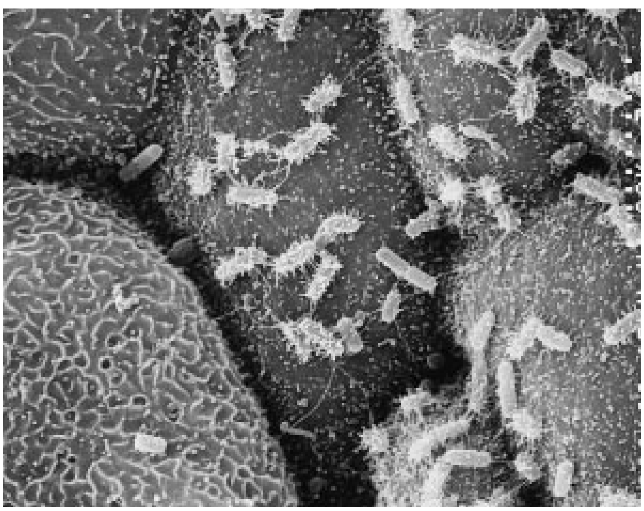
Butcher and Weissman went on to show that their antibody bound on the lymphocyte membrane to a single glycoprotein, now known as L-selectin. Because that glycoprotein is responsible for the specific binding of the lymphocytes to the high endothelial venules, it is also known as the homing receptor.

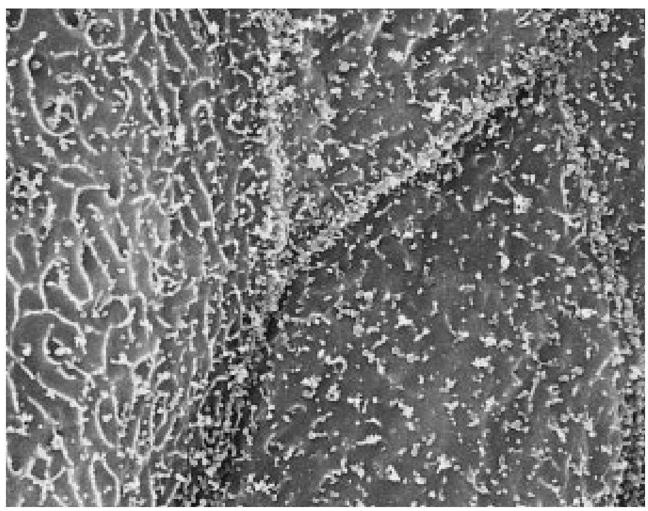
If high endothelial venules from the lymph nodes are exposed to solutions of L-selectin, lymphocytes cannot bind to them: the L-selectin molecules occupy all the potential attachment sites on the endothelial cells. Conversely, as Steven D. Rosen of the University of California at San Francisco has shown, certain small sugars and larger polysaccharides can also block interactions between lymphocytes and endothelial venules. In those cases, the sugars are binding to the L-selectin.

In 1989 separate experiments by Weissman and by Laurence A. Lasky of Genentech in South San Francisco in collaboration with Rosen proved conclusively that the homing receptor mediates the adhesion of lymphocytes to endothelial cells. The structure of the endothelial carbohydrate to which it binds is still unknown.

In contrast to the homing receptor, the two other known selectins are found mainly on endothelial cells, and then only when they are actively attracting leukocytes. One of these, E-selectin (ELAM-1), was discovered in 1987 by Michael P. Bevilacqua of Harvard Medical School. The third member of the group, P-selectin (previously known as GMP-140 and PADGEM), was independently discovered a couple of years later by Rodger P. McEver of the Oklahoma Medical Research Foundation and by Bruce and Barbara Furie of the Tufts University School of Medicine.

Research has clarified how tissues use selectins to steer white blood cells where they are needed. When a tissue is infected, it defensively secretes proteins called cytokines, such as interleukin-1 and tumor necrosis factor. The cytokines stimulate endothelial cells in the venules to express P- and E-selectins on their surfaces. Passing white blood cells adhere to these protruding mole-





**SELECTIVE EFFECTS of carbohydrates on bacteria are illustrated in these photographs. These *E. coli* have a lectin for the P glycolipid. Bacteria incubated in the sugar mannose can still cling to epithelial tissue (*left*). A constituent of the P glycolipid binds to the bacteria's lectin and prevents adhesion (*right*).**

cules because their carbohydrate coat contains complementary structures. Once attached to the wall of a venule, a leukocyte can leave the bloodstream by squeezing between adjacent endothelial cells.

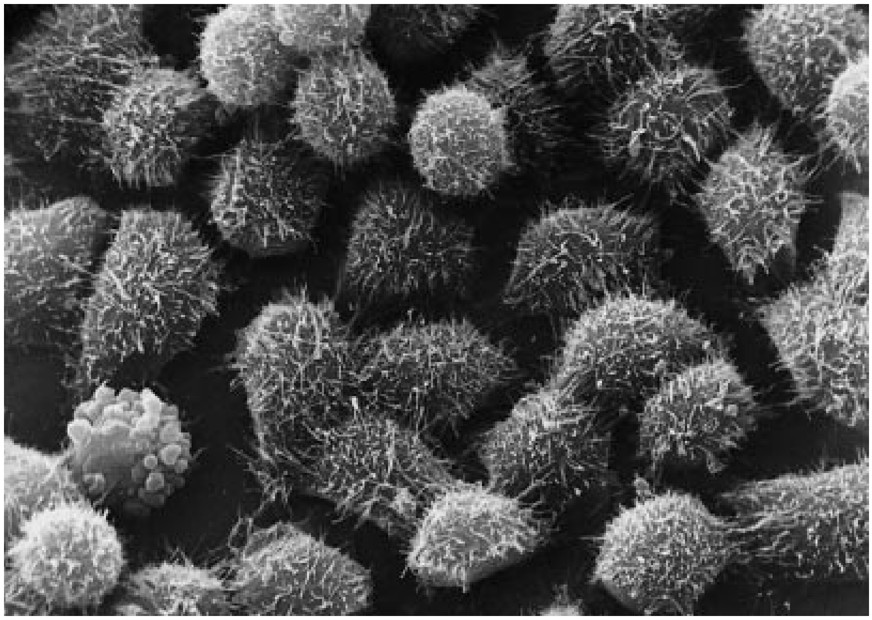
These two selectins appear on endothelial cells at different times and recruit different types of white blood cells. Endothelial cells have an internal stockpile of P-selectin that they can mobilize to their surface within minutes after an infection begins. P-selectin can therefore draw leukocytes that act during the earliest phases of the immunologic defense. In contrast, endothelial cells synthesize E-selectin only when it is required, so it takes longer to appear. That selectin seems to be most important about four hours after the start of an infection, after which it gradually fades away.

The mechanism that helps leukocytes to breech the endothelial barrier is indispensable to the fulfillment of their infection-fighting duties. Yet acting inappropriately, that same mechanism also allows leukocytes to accumulate in tissues where they do not belong, thereby causing tissue damage, swelling and pain.

The inflammation of rheumatoid arthritis, for instance, occurs when white blood cells enter the joints and release protein-chopping enzymes, oxygen radicals and other toxic factors. Another example is reperfusion injury, a disorder that occurs after the flow of blood is temporarily cut off from a tissue, such as during a heart attack. When the blood flow resumes, the white blood cells destroy tissues damaged by lack of oxygen.

The development of pharmaceutical reagents that would inhibit adverse inflammatory reactions holds major interest for the academic, clinical and industrial sectors. In theory, any drug that interferes with the adhesion of white blood cells to the endothelium, and consequently with their exit from the blood vessel, should be anti-inflammatory. The key to developing such drugs is the shape of the binding regions of the selectin molecules and the shape of the carbohydrates that fit into them. Work on determining those shapes is proceeding at a ferocious pace. In parallel research, intensive attempts are being made to synthesize carbohydrate inhibitors of the P- and E-selectins.

For an antiadhesive therapy to be successful, the drugs must simultaneously accomplish two seemingly incompatible ends. On the one hand, they must stop white blood cells from leaving the bloodstream inappropriately; on the other hand, they must still allow the cells to go where they are needed. Those goals may be achievable because the specificities of adhesion molecules vary in different tissues. One can envision, for example, a drug that keeps white cells from entering the joints but not other parts of the body.



**CANCER CELLS have unusual carbohydrates on their surface, which may account for many of their invasive properties. Drugs that interfere with the adhesiveness of abnormal cells may someday be used in cancer therapies.**

Aside from their involvement in inflammation, cell-adhesion molecules may play a role in other diseases, such as the spread of cancer cells from the main tumor throughout the body. For example, the carbohydrate recognized by E-selectin is expressed on cells from diverse tumors, including some cancers. Bevilacqua has recently reported that at least one type of human cancer cell binds specifically to E-selectin expressed on activated endothelium. Perhaps to promote their own metastasis, some malignant cells recruit the adhesion molecules that are part of the body's defenses.

If so, antiadhesive drugs may also turn out to be antimetastatic. One hopeful sign in that direction recently came from Hakomori's group, which was studying highly metastatic mouse melanoma cells that carry a lectin for lactose, the sugar in milk. The researchers found that by exposing the melanoma cells to compounds containing lactose before injecting them into mice, they could reduce the metastatic spread of the cells almost by half.

Although the importance of carbohydrates in cell recognition is immense, other modes of recognition that rely on a peptide language do exist. Some forms of attachment, for instance, involve surface proteins called integrins and complementary peptides. The existence of more than one system for binding activities lends greater flexibility to a cell's repertoire of interactions.

Biomedical researchers are still striving for a better understanding of the sugar structures on cell surfaces and of the specificities that lectins have for those structures. As they learn more, they will be in a better position to design highly selective, extremely powerful inhibitors of cell interactions. The day may not be far off when antiadhesive drugs, possibly in the form of pills that are both sugar-coated and sugar-loaded, will be used to prevent and treat infections, inflammations, the consequences of heart attacks and perhaps even cancer.


FURTHER READING

LECTINS AS CELL RECOGNITION MOLECULES. N. Sharon and H. Lis in *Science,* Vol. 246, pages 227–234; October 13, 1989.

GLYCOBIOLOGY: A GROWING FIELD FOR DRUG DESIGN. Karl-Anders Karlsson in *Trends in Pharmacological Sciences,* Vol. 12, No. 7, pages 265–272; July 1991.

CARBOHYDRATES AND GLYCOCONJUGATES: UPWARDLY MOBILE SUGARS GAIN STATUS AS INFORMATION-BEARING MOLECULES. K. Drickamer and J. Carver in *Current Opinion in Structural Biology,* Vol. 2, No. 5, pages 653–654; October 1992.

# The Earliest History of the Earth

*Radioactive dating techniques have illuminated vast stretches of geologic history, bringing the most ancient eras of the earth's evolution into view*

by Derek York

Radioactive dating provides a powerful way to measure geologic time. It has revealed the overall pace of the earth's evolution and has enabled researchers to calculate that our planet is about four and a half billion years old. Yet the earliest stages of terrestrial history—during which the earth acquired its large iron core and light, mobile continents—have eluded easy investigation because of the many processes that act to reset the radioactive clock. As continents drift across the surface, the ocean floor between them recycles into the hot interior. Where continents clash, they raise folded mountains. Hot underlying material invades the continental rocks and can break through, unleashing lava that covers the surface. Erosion planes off the mountains and sweeps sediments into ocean trenches where they, too, return to the mantle.

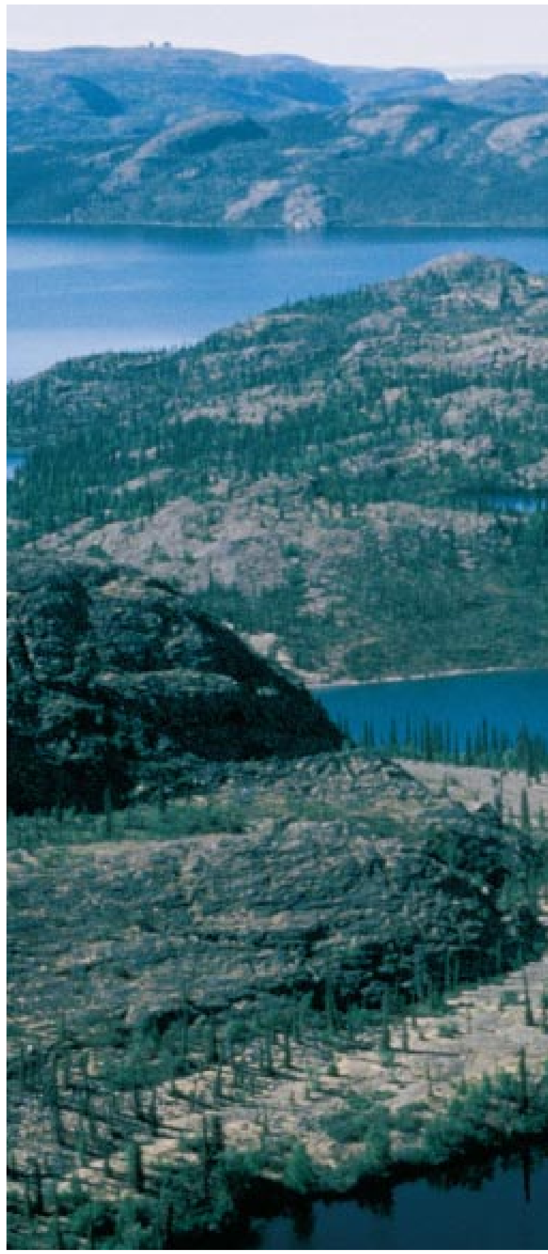
Earth scientists are now using increasingly sophisticated techniques to pry meaningful stories about the earliest events in the earth's history from previously taciturn rocks. Examinations of ancient minerals are revealing when the first continents appeared and how extensive they were. Workers are also uncovering evidence that plate tectonics has operated throughout most of the earth's history much the same as it does now, contrary to some theoretical expectations. The recent discoveries are filling in the long-mysterious details of the formative years when our planet acquired many of its most characteristic traits.

DEREK YORK has spent more than three decades refining the tools of geochronology in order to investigate the earth's distant past. York received his D.Phil. from the University of Oxford in 1960, at which point he joined the department of physics at the University of Toronto. Last year he was appointed chairman of the department. York has also been extensively involved in the popularization of science. He has frequently contributed news stories to *The Globe and Mail* and was a guiding force behind the recent film *Chaos, Science and the Unexpected,* produced by the Canadian Broadcasting Corporation.

To search for clues about the earth's youthful nature, geophysicists make use of an assortment of radioactive dating methods. These methods vary in their strengths and weaknesses, but they all rely on determining the relative abundances of a radioactive isotope and the subsequent isotope, or daughter nucleus, into which it decays. Every radioactive isotope eventually produces a final, stable decay product. Knowing the rate at which the nuclear transformation occurs (which can be measured to high precision in the laboratory) allows one to infer how long the decay products have been collecting in a rock. That information, taken with other evidence, reveals much about geologic history.

In the ongoing search for the oldest continental remnants, researchers primarily examine isotopes of uranium. Uranium ultimately decays into lead, so the relevant dating technique is called the uranium-lead method. That approach greatly benefits from the fact that samples of uranium and lead large enough to analyze can usually be extracted from zircon crystals. Such crystals are very commonly found in granitic and metamorphic rocks, as well as in some volcanic rocks and in sedimentary material derived from any of those rocks. Zircons also resist heat and weathering strongly, so they may survive intact in rocks that have experienced one or more metamorphic episodes.

A potential problem with the uranium-lead dating method is that rocks exposed to tremendous heat and pressure may lose a significant amount of their



**ACASTA GNEISS, a group of metamorphic rocks in northern Canada, is the**

lead, thereby resetting the radioactive clock. In 1956 George W. Wetherill of the Carnegie Institution of Washington showed a way to circumvent the difficulty. His procedure depends on the fact that there exist two radioactive isotopes of uranium, uranium 238 and uranium 235. Each form of uranium follows its own decay path: uranium 238 breaks down into lead 206, uranium 235 into lead 207. Therefore, for any uranium-bearing mineral, researchers can derive an age estimate from two sources.

Wetherill measured the two uranium-lead abundance ratios in a large number of samples and plotted them against each other. On such a plot, samples that have never been disturbed, and hence that are perfect clocks, would lie along a continuous curve that Wetherill called a concordia curve. (The curve simply reflects the fact that both uranium 235 and uranium 238 decay at a steady, predictable rate.)

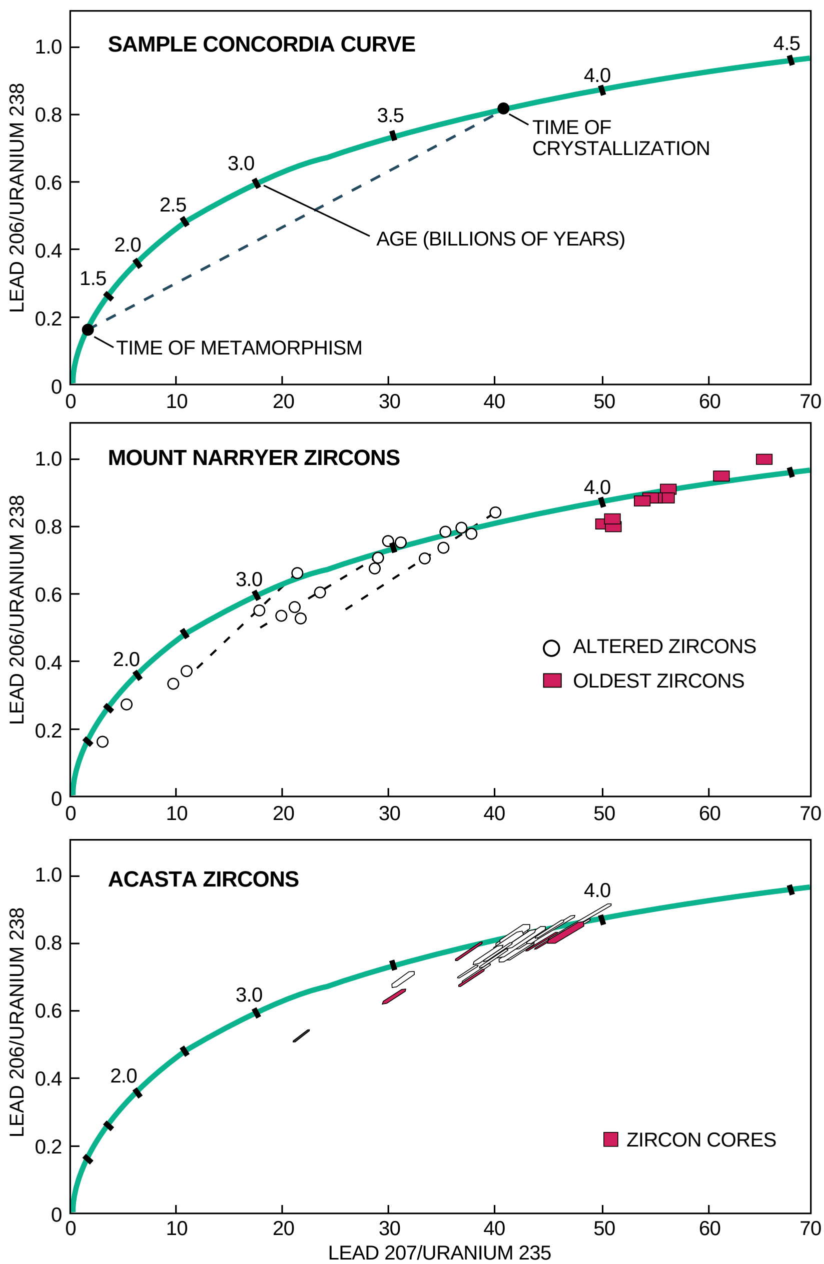
Wetherill then made the remarkable discovery that by plotting abundance ratios he could determine the age of a group of rock samples (all of the same age), even if much of their lead had leaked out during a metamorphic episode. His method works because lead 206 and lead 207 are chemically identical, and equal fractions of the two isotopes would have escaped from the rocks. When the abundance ratios of uranium to lead in the rocks are measured and plotted, the data points associated with the various samples fall on a straight line lying below the concordia curve. The end points of that line intersect the concordia curve at locations corresponding to the time of crystallization and to the time of metamorphism [*see box on next page*].

Applying the uranium-lead method to zircons can be difficult because zircon crystals frequently have a layered structure in which the original core is wrapped in subsequent mineral coatings. In the 1970s Thomas E. Krogh of the Royal Ontario Museum in Toronto demonstrated how to abrade zircons to isolate their cores; he also showed that

**oldest known intact piece of the earth's surface. Radioactive dating indicates that the Acasta gneiss is nearly four billion years old, proving that some continental material existed only a few hundred million years after the earth's formation.**

## How Uranium-Lead Dating Works

Natural minerals such as zircons contain two isotopes of radioactive uranium: uranium 235, which decays into lead 207, and uranium 238, which decays into slightly lighter lead 206. In undisturbed zircons, plots comparing the two uranium-lead abundance ratios fall on a curve known as the concordia curve; their locations on the curve indicate the age of each sample. The uranium-lead data for zircons that underwent a metamorphic episode, thereby losing some of their lead, will fall along a line intersecting the concordia curve at two points (*top*). The upper point represents the original time when the rocks crystallized; the lower point indicates the time of metamorphism. Zircons from Mount Narryer in Australia all show some signs of disturbance (*middle*). Judging from how they fall on the concordia diagram, the oldest Mount Narryer zircons seem to be 4.1 to 4.2 billion years old; the others form three families 3.1, 3.3 and 3.75 billion years old. Based on similar reasoning, zircons from the Acasta gneiss appear to have crystallized about 3.96 billion years ago (*bottom*).




the uranium-lead ratios of the cores frequently fall on Wetherill's concordia curve. Krogh concluded that the inner parts of the zircons had not been chemically altered and so still recorded the time when the mineral first crystallized.

During the 1980s, William Compston and Steven Clement of the Australian National University in Canberra took zircon dating one step further. Instead of analyzing the entire zircon core all at once, as researchers had previously done, Compston and Clement sought to study the composition—and hence the age—of the zircon at many spots. The researchers constructed a device that could blast a sample with a tightly focused (only 25 microns wide) beam of ionized, or electrically charged, oxygen atoms. Compston and Clement called their instrument SHRIMP, an acronym for Super High-Resolution Ion Micro-Probe. They trained SHRIMP at one spot on the inside of a zircon crystal that had been cut in half. The ions vaporize atoms of uranium and lead from that spot; the atoms then pass through a mass spectrometer that separates them by mass and counts them.

In 1983 SHRIMP began to supply noteworthy new information about the age of the earth's crust. Derek O. Froude of the Australian National University, working with Compston and others, began probing single crystals of zircon in quartzite, or metamorphosed sandstone, at Mount Narryer in Western Australia. Earlier work had shown that this region contains rocks about 3.6 billion years old. Froude's group performed detailed analysis on 20 zircon crystals from one rock specimen. Four of the crystals yielded lead-to-uranium data points indicative of an age of 4.1 to 4.2 billion years. Previously, the oldest known pieces of terrestrial material were rocks from southwest Greenland, which Stephen Moorbath of the University of Oxford and his co-workers had found to be 3.8 billion years old. The other 16 zircons furnished isotopic ratios that clustered around three lines that intersected the concordia curve at ages of about 3.75, 3.3 and 3.1 billion years, respectively.

Froude inferred that the 4.1- to 4.2-billion-year-old zircons and the 3.75-billion-year-old specimens formed long before they were incorporated into the surrounding sedimentary rocks. Somehow the zircons eroded from their parent rocks and found their way into sediments that later, under enormous heat and pressure, became the Mount Narryer quartzite. The younger, 3.3- and 3.1-billion-year-old zircons probably started growing during that period of metamorphism. Because zircons are found overwhelmingly in continental rather than oceanic rocks, the Australians' finding strongly suggested that at least some continental material existed more than four billion years ago. Unfortunately, the zircons seem to be the only surviving relics of those ancient rocks.

In 1989, Samuel A. Bowring, then at Washington University, along with Ian S. Williams of the Australian National University and Compston, proved the existence of intact rocks nearly as old as the Australian zircons. The researchers trained SHRIMP on zircons from the Acasta gneiss, a small patch of metamorphic rock southeast of Great Bear Lake in the Northwest Territories. Bowring and his colleagues had earlier used Krogh's abrasion technique to determine that some zircons in the Acasta gneiss are at least 3.8 billion years old. Bowring suspected that buried inside the individual zircons lay evidence of even older ages; such information possibly could be extracted using SHRIMP. He therefore flew to Canberra carrying zircons from two Acasta rock samples.

The researchers probed a total of 82 spots on 53 of the zircons. When displayed on a Wetherill chart, the uranium-lead ratios from the two samples formed fan-shaped plots lying near the concordia curve. One plot fell between 3.6 and 3.96 billion years, the other between 3.8 and 3.96 billion years. Bowring and his co-workers concluded that the oldest zircons recorded the original crystallization age of the rocks. The spread in the data probably indicates that the zircons underwent at least two metamorphic episodes, one during the first few hundred million years after crystallization and another about two billion years ago.

If that interpretation is correct, then the Acasta gneiss is the metamorphosed remains of the oldest known intact, solid rock on the earth's surface. Geologists have identified other rocks nearly as old in Greenland, Labrador and Western Australia. Lance P. Black of the Bureau of Mineral Resources in Canberra and others, also using SHRIMP, recently reported finding 3.87-billion-year-old zircons in Antarctica as well.

These results leave little doubt that at least small patches of continental rock existed on the earth's surface during the first 700 million years of its history. They also underscore the incredible scarcity of crustal rocks more than four billion years old. Plate tectonics alone may not suffice to explain why such rocks are so rare. Perhaps the earth had an extensive primeval crust that was destroyed and remixed into the interior by the impact of giant meteorites, the leftovers from planetary formation. Vigorous convection, driven by the great internal heat of the newborn planet, might have aided that disruption by tearing apart blocks of continental rocks and by pulling continental sediments into the hot depths.

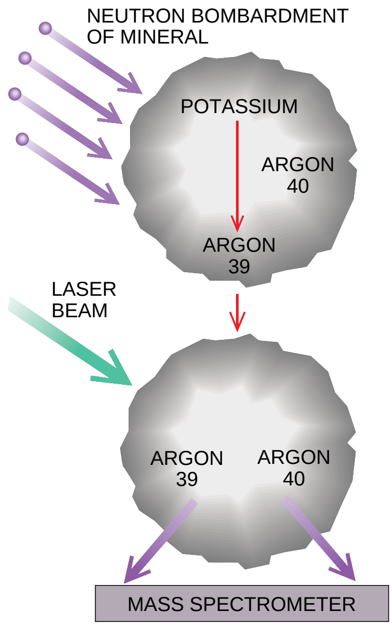
On the other hand, evidence is mounting, based on geochemical arguments, that the total amount of continental crust before about four billion years ago was minuscule. Studies of the relative abundances of isotopes of neodymium, strontium and lead in continental and oceanic crust imply that no more than trivial quantities of continental crust existed before then. About 3.8 billion years ago the earth's mantle began to separate into lighter and denser components, releasing the raw material from which continental blocks formed. The continents seem to have continued growing rapidly until roughly 2.5 billion years ago.

What were the earth's internal dynamics like during that era of swiftly expanding continents? My group in Toronto, in collaboration with Alfred Kröner of Gutenberg University in Mainz and Michael O. McWilliams of Stanford University, has been trying to address that question. We do so by assessing the extent of early continental drift. Geophysicists can trace recent continental motions by means of the magnetic record preserved in the ocean floor. Ocean floors survive only about 200 million years before they sink back into the earth's mantle at oceanic margins, such as along the Pacific trench off Asia.

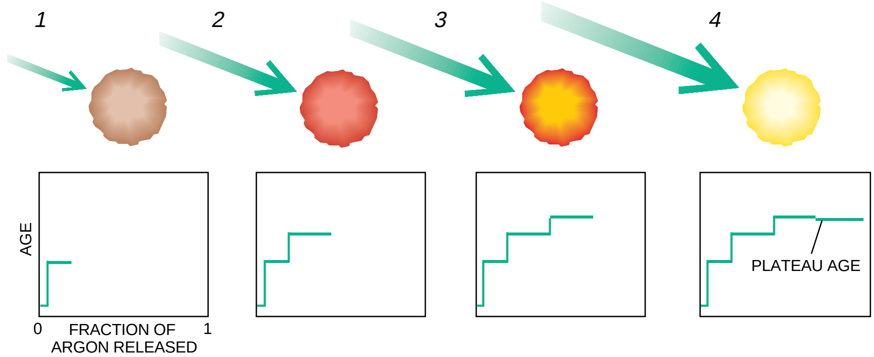
Determining the motions of the continents more than two billion years ago demands extending the tools of geochronology to include measurements

**SEARCH FOR THE OLDEST ROCKS led Samuel A. Bowring (*left*), now at the Massachusetts Institute of Technology, to the Acasta gneiss. Bowring and his collaborators collected zircons from those rocks (*right*) and polished them to reveal their inner structure. The researchers then measured the lead-to-uranium ratios at various spots in the zircons to find the oldest parts. The shallow pits were produced by beams of ions used to vaporize bits of the zircon crystals for analysis.**

## Using a Laser to Illuminate the Past




Laser step-heating offers an efficient way to measure both the age and the thermal history of a mineral sample. In this approach, the age is measured by the degree to which radioactive potassium has decayed into argon 40. In the first step, the sample is bombarded by neutrons, which transform some of the potassium into argon 39 (*left*). An intense laser beam heats the sample and releases both isotopes of argon; a mass spectrometer measures the exact abundances of the isotopes. As the laser heat increases, more and more argon is




of a rock's internal magnetism. When lavas erupt or when granites form in the outer layers of the earth, iron oxides in the rock become magnetized in the direction of the earth's magnetic field at that site (the iron oxides act as tiny bar magnets that point toward the north pole). Measuring the direction of the field frozen into the rock reveals how far it was from the magnetic pole when it cooled. By studying the magnetism of rocks of varying ages, all from the same general site, one can, in principle, learn how much the continent has moved toward or away from the pole over time.

Unfortunately, if a rock is heated above a critical temperature, it loses its original magnetic direction and becomes remagnetized the next time it cools. The new direction may be totally different from the original if the continent has drifted significantly in latitude during the intervening years. If the rock was not severely reheated, however, some of the original magnetic orientation remains. In that case, geophysicists can extract from the rock a record of two ancient pole positions: one at the time of crystallization or initial cooling, the other at the time of metamorphic heating. Because all known Precambrian rocks have suffered some heating episode, it is critical to uncover a rock's thermal history to decode its true magnetic record.

A valuable radioactive dating method known as the potassium-argon technique can sometimes permit researchers both to determine the age of magnetized rocks and to assess whether (and how much) they have been heated during their lifetime. Potassium 40, a rare isotope of potassium, decays to produce argon 40, a heavy version of the unreactive gas argon; the half-life of potassium 40 is 1.3 billion years. From the accumulation of argon 40 in potassium-bearing minerals, one can determine how long ago the mineral solidified.

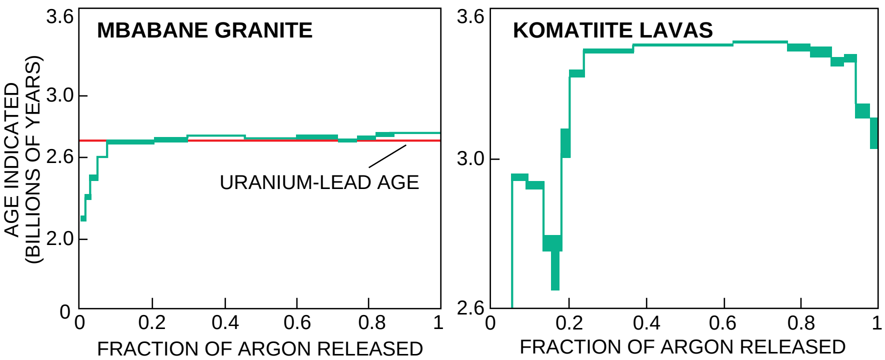
Potassium-argon dating has assisted geochronologists in outlining the time scale of biological evolution over the past 500 million years. It has proved less appropriate for probing the earth's more distant history because argon tends to leak from minerals during times of metamorphic heating. The uranium-lead method, along with a somewhat less widely used approach involving isotopes of rubidium and strontium, is more effective for revealing crystallization ages of the oldest rocks.

The ease with which the potassium-argon clock can be disturbed offers a compensating advantage: it enables researchers to unravel a rock's thermal history. That information in turn makes it possible to make sense of the rock's magnetic history. Geochronologists have studied many different kinds of minerals to determine how readily they release entrapped argon when heated. The common mineral hornblende has been shown to be quite resistant to loss of argon. Usually, only severe heating (above about 500 degrees Celsius) suffices to let some of the argon escape. The minerals muscovite and biotite, two forms of mica, are somewhat less resistant to heat; they can be disturbed by temperatures in the range of 250 to 350 degrees C. At the other extreme, the mineral feldspar shows signs of argon loss below 200 degrees C.

Workers searching for evidence of continental drift during the first half of the earth's history have concentrated much of their attention on a series of remarkably well preserved rocks in the Barberton Mountain Land greenstone belt, located on the border between South Africa and Swaziland. The rocks are part of the Kaapvaal Craton, a section of stable, extremely old continental crust. The greenstone belt consists of numerous volcanic rocks buried underneath later sediments. Younger, granitic rocks have forced their way into the older greenstone layers.

My co-workers and I have looked for signs of tectonic drift more than two billion years ago by carrying out detailed magnetic analyses and by conducting both potassium-argon and uranium-lead zircon dating of two samples of granite from the Kaapvaal Craton. In Mainz, Kröner applied the uranium-lead method to single crystals of zircon from lavas in the craton and showed that they crystallized about 3.5 billion years ago, in broad agreement with earlier measurements made at the University of Cambridge. He also found that granites from the Nelshoogte region of the Kaapvaal Craton crystallized about 3.2 billion years ago; granites from the nearby Mbabane region first cooled 2.69 billion years ago.

Paul W. Layer, Margarita Lopez-Martinez and I performed potassium-argon dating in my laboratory at the Universiforced out (*steps 1–4, center*); at each step, an age is derived from the ratio of argon 40 to argon 39. Low ages at cool temperatures indicate that some argon leaked out during the sample's history; the plateau age at hotter temperatures should reflect the sample's true age. At the right, laser step-heating of four hornblende grains from the Mbabane granite in Swaziland shows they are 2.7 billion years old, in close agreement with the age derived from uranium-lead measurements. Analysis of komatiite lavas from the same region reveals that they first cooled about 3.5 billion years ago.




ty of Toronto to examine the thermal history of the rocks as the first step in interpreting their magnetic record. We used a clever variant of the potassium-argon method, known as argon-argon dating, that Craig M. Merrihue of the University of California at Berkeley had originally proposed. In Merrihue's approach, one irradiates a rock sample with neutrons from a nuclear reactor, transmuting some of the potassium into argon 39. The sample is then melted in a vacuum chamber, freeing both the argon 39 and argon 40. Instead of making separate measurements of the potassium and argon abundances in the sample, a researcher need make only a single observation of the two argon isotopes to derive the age of a sample.

Merrihue realized that his technique could also provide vital information about the thermal history of the rock being studied. He suggested that rather than melting the irradiated mineral all at once, one could heat the sample through a series of steps and measure the ratio of argon 40 to argon 39 (and deduce an age corresponding to that ratio) at each temperature. If every step gives the same age, one can reasonably assume that the sample has suffered little thermal disturbance during its history. If the measured ages begin low at cool laboratory temperatures and climb to a plateau of greater ages at hotter ones, then some of the argon 40 must have been lost from the mineral during one or more heating events. The age of the plateau takes into account all the argon 40 that is released only under fierce heat and so should render a good indication of the time when the rock first cooled and the potassium-argon clock switched on.

At Toronto, my collaborators and I applied the step-heating method to the Nelshootge and Mbabane granites, as well as to lavas related to the volcanic rock whose age had been measured by Kröner using the zircon method. We also used another innovation: a laser-heating technique I developed at Toronto in collaboration with Chris M. Hall, now at the University of Michigan, and Yotaro Yanase. A similar technique had been pioneered in the 1970s by George H. Megrue of the Smithsonian Institution. He showed that it is possible to measure the argon-argon age of a mineral by vaporizing part of a sample with a pulsed laser. His method did not gain wide acceptance at the time, because he could not modulate the laser's power to heat the specimen in neat, steplike increments.

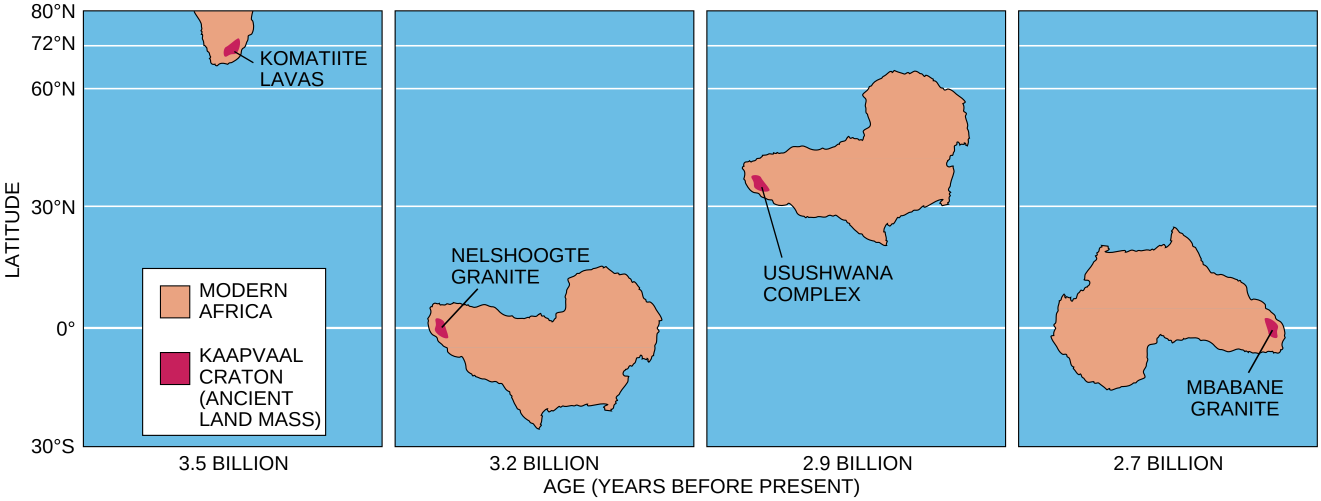
By modifying Megrue's technique, my colleagues and I succeeded in using a continuously beamed laser to produce a precise age spectrum from a single grain of rock. We begin by bathing a bit of mineral in a low-power laser beam for 30 seconds and analyzing the released argon by means of a mass spectrometer. Then we increase the power of the laser beam in a series of increments and derive an age at each step. When the mineral fuses, it has reached its crystallization temperature, and the age spectrum is complete [*see box above*].

The combination of laser heating and argon-argon analysis is supplying impressively accurate information about the age and the thermal histories of ancient rocks. That information has enabled us to reconstruct an unprecedentedly well documented account of continental drift as it occurred billions of years ago. We began by considering the history of the Mbabane granite. Layer and McWilliams had discovered that this rock contains a magnetic image of the pole, indicating that the rock solidified somewhere near the earth's equator. Further analysis showed that the Mbabane granite acquired its magnetic orientation when the rock cooled from around 600 to 500 degrees C.

Laser argon-argon dating of four hornblende crystals in the granite, conducted by my group at Toronto, yielded an age spectrum displaying a broad plateau of 2.69 billion years, essentially identical with the uranium-lead age of the surrounding zircons derived by Kröner. The close agreement of the two dating techniques implies that the hornblendes, and hence the whole granite, had never been heated to temperatures above 500 degrees C after they first cooled. We concluded that the magnetic pole position detected in the granite was recorded when the rocks initially cooled and solidified some 2.69 billion years ago, at which time the Mbabane granite lay at the equator.

Layer and his colleagues had already deduced that another group of igneous rocks now located 12 kilometers from the Mbabane granite sat a little more than 30 degrees from the equator 2.875 billion years ago. If those two formations have always been close neighbors, then that area of the Kaapvaal Craton must have drifted through 30 degrees of latitude between 2.875 and 2.69 billion years ago. Such movement implies a drift rate of about 1.5 centimeters a year, comparable to the speed at which North America has drifted away from the Mid-Atlantic Ridge during the past 100 million years. We naturally wondered if plate tectonics occurred at the same pace during even earlier times.

Kröner had shown that the Nelshootge granite is 3.21 billion years old. Laser analysis of individual hornblende crystals in the granite suggested that the rock acquired the magnetization early in its history, at least 3.18 billion years ago. When we examined the orientation of the magnetic field in the granite, we found that it, too, seemed to have formed 90 degrees away from the pole, that is, on the equator. But the direction of the North Pole recorded in the Nelshootge granite lay many degrees away from the pole position frozen into the Mbabane granite, so the Kaapvaal




**ANCIENT CONTINENTAL DRIFT is recorded by magnetic fields frozen into four kinds of rocks found in the Kaapvaal Craton, a region of stable, extremely old continental crust (the African continent, shown for reference, did not then exist in its modern form). The magnetized rocks act as compasses that reveal the craton's orientation and latitude at the time when the rocks crystallized. The rate at which the Kaapvaal Craton wandered broadly resembles the present pace of plate tectonics.**

Craton must have rotated considerably during the intervening years.

Finally, we moved on to examine the oldest Barberton rock in our collection, fragments of magnesium-rich lavas known as komatiites. Zircons in volcanic rocks related to those lavas date from almost 3.5 billion years ago. David J. Dunlop and Chris J. Hale of the University of Toronto recovered a weak magnetic field frozen into the komatiites, indicating that when they first cooled, the lavas were only about 18 degrees from the pole—nowhere near the equator. The significance of this finding was unclear, however. The lavas evidently have been exposed to intense heat and pressure subsequent to their formation; they preserve few if any of their original minerals, so the magnetization observed by Dunlop and Hale could be much younger than 3.5 billion years.

To clarify the matter, Lopez-Martinez, then a graduate student in my laboratory, carried out an argon-argon measurement to determine the age of the komatiite samples. Much to her surprise—and to mine—her measured ages were indistinguishable from those that Kröner had found by studying zircons in the lavas. Zircons date the initial crystallization of the lavas, whereas argon-argon analysis measures the age of tremolite, a mineral that forms during metamorphism. The only way to make sense of Lopez-Martinez's result is to conclude that the metamorphic event that created the tremolite must have occurred almost immediately after the lavas first erupted. If so, the pole location in the Barberton komatiites is a genuine recording of the location of the rock nearly 3.5 billion years ago.

Pulling together the above evidence yields an 800-million-year history of the drift of the Kaapvaal Craton, beginning 3.5 billion years ago. At that time, the craton was near the pole. By about 3.18 billion years ago it had wandered to the equator. The craton then moved more than 3,000 kilometers poleward, so that 2.875 billion years ago it lay at least 30 degrees from the equator. By 2.69 billion years ago the craton had drifted back to the equator, but its orientation was significantly different than it had been 490 million years earlier.

Our work implies that the Kaapvaal Craton was drifting about as fast as do modern continents since at least 3.5 billion years ago. Presumably, other continental fragments were behaving the same way. I should emphasize that despite the many advances in methodology and instrumentation, studies of magnetic traces still leave considerable room for error. Nevertheless, uranium-lead and argon-argon dating, when combined with paleomagnetic studies, have very much bolstered understanding of the earth's dynamic early history.

The success of our investigation of the Barberton komatiites led us to wonder if the very oldest known rocks, the Acasta gneiss, carry a snapshot of their position relative to the magnetic pole 3.96 billion years ago. Much to our regret, they seem not to. In my laboratory, Hall performed preliminary argon-argon dating of hornblende crystals from one of Bowring's samples. His results show that the rocks' argon-argon clocks were reset by a pulse of tectonic activity 1.8 billion years ago. Any 3.96-billion-year-old magnetic traces evidently would have been completely erased.

Yet hope persists. Perhaps further measurements of rocks from the same region in northern Canada may yield a small fragment that escaped severe reheating during mountain-building events. The magnetic record in such a rock would be invaluable for decoding the earth's evolution. Evidence that the earth possessed a significant magnetic field at that early date would be a telling sign that it had already developed a sizable metallic core—a key step in the transformation of an infant planet into a well-ordered, complex world.

FURTHER READING

COOLING HISTORIES FROM $^{40}$AR/$^{39}$AR AGE SPECTRA: IMPLICATIONS FOR PRECAMBRIAN PLATE TECTONICS. D. York in *Annual Review of Earth and Planetary Sciences,* Vol. 12, pages 383–409, 1984.

THE DERIVATION OF $^{40}$AR/$^{39}$AR AGE SPECTRA OF SINGLE GRAINS OF HORNBLENDE AND BIOTITE BY LASER STEP-HEATING. P. W. Layer, C. M. Hall and D. York in *Geophysical Research Letters,* Vol. 14, No. 7, pages 757–760; July 1987.

ELEMENTS OF THE ARCHEAN THERMAL HISTORY AND APPARENT POLAR WANDER OF THE EASTERN KAAPVAAL CRATON, SWAZILAND, FROM SINGLE GRAIN DATING AND PALEOMAGNETISM. P. W. Layer, A. Kröner, M. McWilliams and D. York in *Earth and Planetary Science Letters,* Vol. 93, No. 1, pages 23–34; May 1989.

CRUST FORMATION AND PLATE MOTION IN THE EARLY ARCHEAN. A. Kröner and P. W. Layer in *Science,* Vol. 256, pages 1405–1411; June 5, 1992.

# Madagascar's Lemurs

*These primates can tell us a great deal about our own evolutionary past. But many species are already extinct, and the habitats of those that remain are shrinking fast*

by Ian Tattersall

From dense rain forests to broad coastal plains, from deciduous thickets to desert, Madagascar offers an extraordinary range of environments. These habitats harbor an equally extraordinary primate fauna that shows more clearly than any other what our own primate ancestors were like early in the Age of Mammals, approximately 50 million years ago.

These animals are the lemurs, Madagascar's dominant mammals. Why they are there is not entirely clear. The story used to be that the 1,000-mile-long Madagascan minicontinent had simply preserved (in somewhat impoverished fashion) an archaic fauna that was marooned on it when Madagascar separated from the African mainland and drifted out to sea. In many ways the island's living "lower" primates more closely resemble the primates of the Eocene epoch (about 57 to 35 million years ago) than they do the "higher" primates dominating the tropical continents today. That fact was taken to imply that the separation had occurred in the Eocene or thereabouts.

But we now know that Madagascar began its journey away from Africa as much as 165 million years ago, when the dinosaurs ruled and the only mammals were tiny and vaguely shrewlike. Moreover, the island appears to have reached its present separation of roughly 250 miles from Africa some tens of millions of years before the great diversification of the mammals and thus well before today's familiar groups such as the primates, bats and rodents came into existence.

IAN TATTERSALL is a paleontologist and primatologist who has worked extensively on the living and subfossil lemurs of Madagascar as well as on a range of problems in early primate and human evolution. Curator and chairman in the department of anthropology of the American Museum of Natural History, Tattersall is especially interested in the integration of evolutionary theory with the fossil record. He has also served as curator for several major exhibitions at the museum, the most recent of these being the Hall of Human Biology and Evolution, scheduled to open in early 1993.

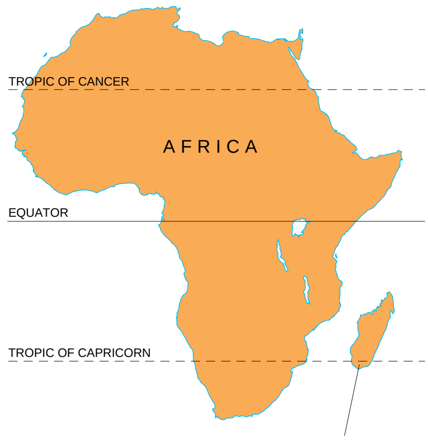
For land-bound latecomers such as primates, then, the only possible means of access to Madagascar was by "rafting": floating across the Mozambique Channel from Africa on matted tangles of vegetation. Arriving thus, the ancestral lemurs would have found an incredible wealth of ecological opportunities on an island that is only slightly smaller than Texas and topographically, climatically and ecologically much more diverse.

The island of Madagascar looks rather like a giant left footprint in the sea, its long axis oriented more or less north-south. It extends from within 12 degrees of the equator all the way to the southern subtropical zone. Its eastern side presents a steep escarpment to the prevailing easterly winds; here heavy rainfall year-round supports a dense growth of rain forest. To the west, a rugged central plateau slopes more gently to broad coastal plains that become drier toward the south. Moist forests in the northwest give way to deciduous forests and thickets. These in turn yield in the far south to an extraordinary desert-adapted flora in which as many as 98 percent of the species are unique. Add to these major regions a host of local microclimates and secondary physiographic features, and you have an unparalleled assortment of environments for forest-living mammals to exploit.

No one knows what the range of habitats was in Madagascar at the remote time when primates first arrived or what lived in those habitats; the fossil record is simply lacking. What is certain, though, is that the primates flourished and that by the time human beings arrived under 2,000 years ago the island was home to at least 45 lemur species. These primates ranged in body size from the two-ounce mouse lemur, *Microcebus,* to the 400-pound *Archaeoindris,* a match for a large gorilla.

Each of the lemur species was a lower primate, belonging (with the bush babies, pottos and lorises) to the primate suborder Strepsirhini. We, on the other hand, are higher primates, belonging (with the monkeys and apes) to the suborder Anthropoidea. (There is no consensus about which group the tiny and enigmatic tarsier of Southeast Asia belongs to.) The distinction between lower and higher primates is in fact a rather archaic concept that is now going out of fashion, but it is a convenient one to use here.

The higher primates appeared on the evolutionary scene much later than did the lower primates, from one of which they presumably evolved toward the end of the Eocene. Madagascar's lemurs




**COQUEREL'S SIFAKAS, a variety of lemur, rest in a forest in northwest Madagascar (*opposite page*). These lower primates evolved from ancestral lemurs that may have traveled on rafts of vegetation across the Mozambique Channel long after Madagascar separated from Africa. The 1,000-mile-long island lies about 250 miles off the coast (*above*).**

RUFOUS MOUSE LEMUR *Microcebus rufus*

RINGTAILED LEMURS *Lemur catta*

GOLDEN BAMBOO LEMUR *Hapalemur aureus*

VERREAUX'S SIFAKA *Propithecus verreauxi verreauxi*

and their continental cousins have much in common with Eocene forms, retaining a suite of physical characteristics that have been lost among higher primates. But only in Madagascar do we still find lower primates that are diurnal, or active during the day. Virtually all modern higher primates are diurnal, and if anything in paleontology is certain, it is that all anthropoids are descended from a diurnal common ancestor. So if we wish to find analogies to our remote Eocene ancestors, we must turn to the primates of Madagascar.

Modern lower and higher primates are distinguished from one another by a number of structural features, most significantly in their nervous systems and sense organs. The lower primates have much smaller brains relative to body size than do higher primates. They also differ in the development of the association areas, which govern the transfer of information between the various brain centers.

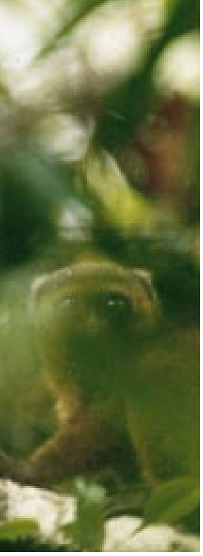
The balance between the senses of vision and smell also differs. Although the eyes of lower primates are quite forward-facing, the left and right visual fields do not overlap as much as they do in higher primates. This arrangement limits depth perception to the central part of the field of view. And whereas it is natural that the nocturnal lower primates lack color-sensitive cone cells in their retinas, what little is known about visual discrimination in the diurnal lemurs suggests that their color vision is at best limited.

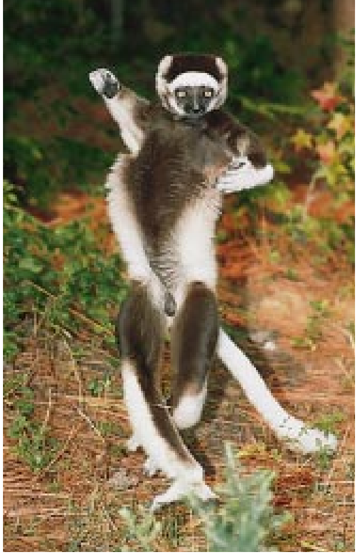
As for the sense of smell, the lower primates have roomier nasal cavities than do higher primates, with more complex internal structures. Living lower primates retain the primitive mammalian rhinarium, or "wet nose." It is part of a system for the transfer of particles to the nasal cavity, where they are analyzed by an organ that is, at most, vestigial in most higher primates.

Many lower primates (but only a few South American monkeys among the higher primates) have scent glands that exude substances used to "mark" the environment. This process is important in communication between individuals. Visual cues are less important; the fac-

GOLDEN CROWNED SIFAKA
*Propithecus tattersalli*

RED RUFFED LEMURS
*Varecia variegata rubra*





VERREAUX SIFAKA, FORSYTH MAJOR'S VARIANT *Propithecus verreauxi verreauxi*

CROWNED LEMUR, FEMALE* *Eulemur coronatus*
*Sex is indicated if male and female coloring differs

es of lower primates lack the musculature needed to produce the complex expressions through which higher primates convey their states of mind.

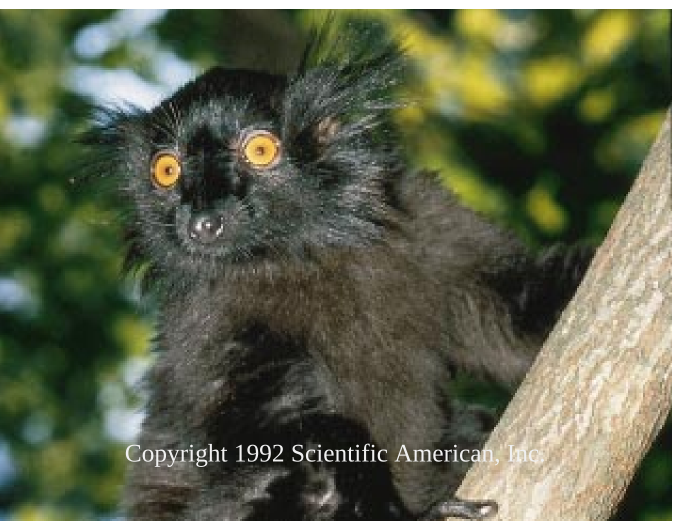
Living primates in general have lost the primitive claws that enabled the first pre-Eocene primates to scale trees without the ability to grasp. In their stead is a thumb, shifted away from the other digits, that is at least to some extent opposable to them, and sensitive tactile pads on the fingertips, which are backed by flat nails. This fundamental change has consequences that extend far beyond locomotion as well as into manipulative abilities. But whereas higher primates generally manipulate objects using the thumb in opposition to the other digits, lemurs tend to pick things up with the whole hand. An item held in this way is then more likely to be sniffed rather than inspected visually and turned in the fingers for further examination.

Both the lower and the higher primates are, of course, extremely diverse groups, whose members represent myriad variations on the themes outlined above. Nevertheless, the lower primates are obviously more primitive than the higher primates (with respect to the body systems discussed) in the sense that they more closely resemble the ancestor from which both modern groups emerged. Although fossils are limited to bones and teeth, they reveal clearly that in the Eocene primate brains were even smaller in relation to body size than are those typical of modern lemurs and that the visual sense had not achieved the total domination over olfaction we see in today's higher primates. Eocene primate hands and feet were certainly capable of grasping but could probably manipulate objects no more precisely than lemurs do at present. In other words, as functioning organisms, Eocene primates were probably not too different from today's lemurs. They were certainly close enough so that by studying the lives of lemurs we can glimpse something of the Eocene behavioral potential from which our vaunted human capacities ultimately arose.

Not until the 1960s did information about the behavior of lemurs begin to be available from field studies in Madagas-

BLACK LEMUR, FEMALE *Eulemur macaco macaco*

BLACK LEMUR MALE
*Eulemur macaco macaco*





MONGOOSE LEMUR, FEMALE AND INFANT *Eulemur mongoz*

BLUE-EYED LEMUR, FEMALE AND INFANT *Eulemur macaco flavifrons*

car. The results of those studies have been increasingly at odds with earlier anticipations. For example, as long as it was possible to think of "the lemur" as a general ancestral model, the expectation, if any, was simple: lemurs would show a set of relatively stereotypical behavior patterns, similar to those exhibited by early primates and from which the higher primates had managed to emancipate themselves. Yet perhaps the most remarkable revelation of the field studies is how diverse the lemurs are in their ways of life.

Of course, certain predictions from anatomy were indeed borne out. Olfactory marking, using urine and feces as well as the secretions of specialized glands, has been shown to be a significant component of lemur behavior. To judge by its wide occurrence among mammals, such marking is an ancient behavior, and few would question its importance in communication among Eocene primates.

Lemurs also turn out, as expected, to have a tendency to explore their environments with their noses—to choose ripe fruit, for example, by smell rather than by appearance. Undoubtedly, then, olfaction is of the greatest importance to members of all five living families of lemurs, as it was to their Eocene precursors, and much more so than it is to most higher primates. But does the reverse apply to vision? Anyone who has watched a sifaka lemur hurtle through the forest would conclude that the animal is hardly handicapped by its lack of a higher primate's visual system.

Lemurs, in fact, turn out to be as diverse as or even more so than the higher primates in most aspects of their behavior and ecology. Their diets, for instance, consist of much the same items as those consumed by the higher primates: fruit, flowers, leaves, buds and insects. But feeding on nectar, which is unusual in higher primates, is being increasingly noted. As a whole, lemurs tend to be dietary generalists, but some at least are highly specialized—for example, the bamboo lemurs. Many of them concentrate not just on bamboo but on particular parts of bamboo plants. Indeed, individuals of the species *Hapalemur aureus,* weighing only a couple of pounds each, daily eat shoots containing enough cyanide to kill half a dozen humans.

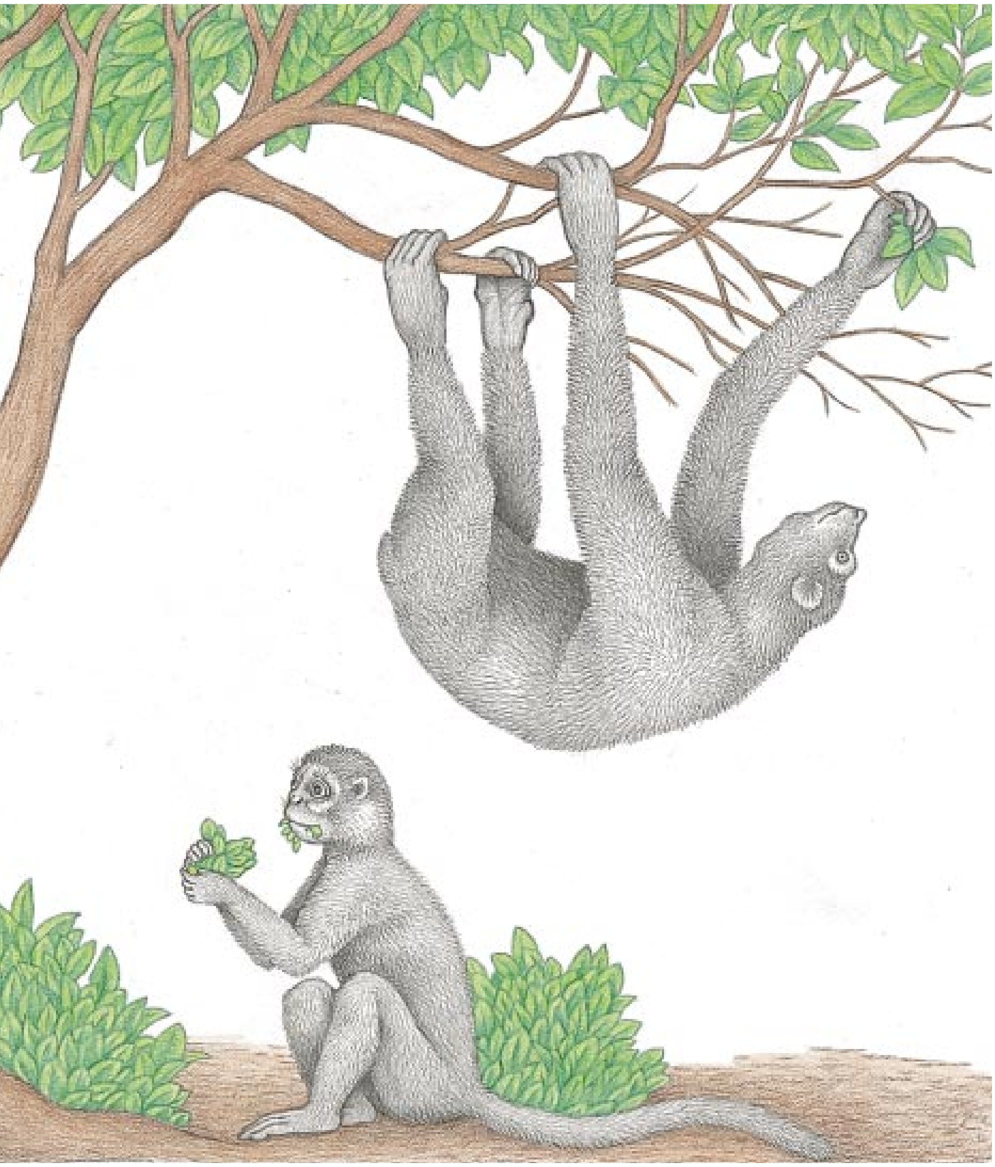
Similarly, lemurs occupy nearly all of the vegetational habitats exploited by higher primates, from rain forest to arid brush. Again, whereas almost all higher primates are diurnal, some lemurs are nocturnal, others diurnal and yet others "cathemeral," spreading their activity fairly evenly between daylight and darkness.

**EXTINCT LEMURS include *Archaeolemur* (*two are shown at bottom*), *Megaladapis* (*upper left*) and *Palaeopropithecus* (*upper right*). *Archaeolemur* was about the size of a female baboon and was adapted to life on the ground. In contrast, *Megaladapis,* weighing as much as 170 pounds, was arboreal and seems to have had adaptations similar to those of the Australian koala. *Palaeopropithecus,* which**

The range of types of social organization among lemurs is enormous. Some lead more or less solitary lives, in which small female ranges are overlapped by larger male ones. In certain species, adult pairs rear immature offspring, whereas in others, small groups consist of just a few adults of each sex. And still other species live in fluid groups or in more stable larger units that contain a couple of dozen individuals or more.

Within these broad categories, there are still further variations. What is perhaps most surprising is that even within the same species, substantial variation in social organization may be found from place to place. What seems to be most significant is that even though lemur brains tend to be smaller than those of higher primates, at least some lemurs display the complex sociality we usually associate with the higher primates.

The picture that is therefore beginning to emerge of the behavioral variety of the lemurs has important implications for the understanding of our own Eocene ancestors. Despite the retention of such primitive behavioral traits as olfactory marking, the wide variety suggests that from the beginning, well before the increase in brain



**weighed at most 130 pounds, was a slothlike arboreal dweller. These reconstructions—done with the advice of Laurie R. Godfrey of the University of Massachusetts at Amherst—are based on fossils that have been unearthed since the late 19th century. Excavations have provided mostly incomplete skeletons of at least 15 species of extinct lemurs, belonging to some eight genera.**

size that we associate with the higher primates today, primates showed a behavioral flexibility and adaptability that belie the inference most people would draw from the description of these early forms as "primitive." It seems to me that this is the essential evolutionary heritage of our order, something far more important than any of the anatomic characteristics to which we are wont to draw attention. As higher primates ourselves, we tend to look at the lemurs and ask why they did not evolve further in our direction. But in doing this, we are missing the critical point that, as a whole, the lemurs have actually inherited from their Eocene ancestors, as we have from ours, the most significant primate characteristic of all.

Looking at this picture from another perspective, it is certainly clear that the lemurs had no need of higher primate physical characteristics to make full use of the varied ecological opportunities that Madagascar offered them. Perhaps this point is emphasized most dramatically by considering the full range of lemur species that existed on Madagascar when humans first arrived there. The present range is impressive enough: at the small end is the tiny two-ounce mouse lemur and at the large end, the 15-pound babakoto. But in its pristine state, the island was home to a spectrum of primate types that equals, if it does not surpass, the entire variety achieved by anthropoids in the rest of the world.

Late in the 19th century, excavations at sites in the central plateau of Madagascar began bringing to light the subfossil, or partially fossilized, remains of large-bodied extinct lemurs that were clearly not of great age. We know of at least 15 species of subfossil lemurs, belonging to eight or possibly more different genera. All of them are larger than any surviving lemurs. The same pattern of large body size is found among the nonprimate members of the subfossil fauna, including the "elephant bird," *Aepyornis maximus* (the largest bird that ever lived, possibly weighing almost half a ton), the pygmy hippopotamus and a giant tortoise.

A long series of studies has revealed an astonishing array of locomotor and positional behaviors among the subfossil lemurs. Even by themselves, the living lemurs show a great variety of such behaviors, ranging from the rapid, scurrying quadrupedal motion of the tiny *Microcebus* to the spectacular leaping of the long-legged indrid lemurs (sifakas, babakotos and the like). It cannot really be said that any of the living lemurs is excluded from any part of the forest environment by its anatomic locomotor specializations, but it is generally true that most species avoid spending much time on the ground. The only notable exception is the ringtailed lemur, *Lemur catta*.

Among the extinct lemurs, on the other hand, was a group that was clearly adapted for life on the ground. This is the family Archaeolemuridae, containing the two medium-sized genera *Archaeolemur* and *Hadropithecus*. Laurie R. Godfrey of the University of Massachusetts at Amherst (the source for all the estimates of subfossil body weight presented here) has estimated that the various archaeolemurid species weighed between 35 and 55 pounds.

These lemurs were rather powerfully built, short-legged relatives of the indrids, with highly specialized teeth. Clifford J. Jolly of New York University has compared them, respectively, with two African higher primates, the common and gelada baboons. The common baboon is an extremely adaptable form, at home in woodlands and deciduous forests as well as in the savanna habitats in which it is so familiar. The gelada baboon is specifically adapted to the treeless Ethiopian highlands, and virtually all its food is derived from terrestrial sources. Both the dentition and what is known of the body skeleton of *Hadropithecus* suggest strongly that this lemur had similar dietary and habitat preferences.

Another extinct indrid relative showed

a totally different set of adaptations. For analogues, one must look well beyond the primates. *Palaeopropithecus* comprised at least two species with a weight range of perhaps 90 to 130 pounds. Ross MacPhee, my colleague at the American Museum of Natural History, has analyzed a substantially complete skeleton recovered in northern Madagascar a few years ago and concluded it was a generally slow-moving and somewhat slothlike arboreal hanger, built for flexibility rather than strength. Its even larger relative *Archaeoindris* (probably more than 400 pounds) is poorly known, but Martine Vuillaume-Randriamanantena of the University of Madagascar believes it was probably a terrestrial quadruped somewhat like the extinct ground sloths of the New World. Both of these forms had specializations of the skull, particularly of the nose area, that are unmatched among living primates.

We also have to look outside our own order to find an analogue to the best-known subfossil lemur, *Megaladapis*. The three species, ranging in weight from about 90 to 170 pounds, have been described by Alan C. Walker of the Johns Hopkins School of Medicine as closest in locomotion to the marsupial koala of Australia. Like the koala, these lemurs would have climbed slowly, presumably preferring vertical supports, and they had limited leaping capabilities. A number of specializations of the skull may have compensated for such locomotor limitations. They would have allowed the animal to feed in a large radius from a single sitting position.

The best known of the sites that have yielded extinct lemurs is Ampasambazimba in Madagascar's central highlands. The 14 primate species (including both extinct and surviving lemurs) whose bones have been recovered there compare favorably in abundance with the species at any other place in the world where primates are found. Yet Ampasambazimba is in the middle of what is now an essentially treeless plateau. How does it boast this rich and heterogeneous forest-living fauna?

Studies dating from early in this century seemed to have the answer. Before the advent of humans, the argument goes, Madagascar was essentially fully forested. The fact that forest had survived only in patches was attributed to the propensity of the early settlers to burn off vast areas of forest to provide grazing for cattle and land for agriculture. Indeed, the process is still only too evident.

Following this argument, loss of habitat must have been at least a major influence in the disappearance of Madagascar's large birds and mammals. There was also a selective aspect of this extinction: the lemurs known to have become extinct were large bodied and thus of the greatest interest and vulnerability to hunters. Presumably, they also reproduced more slowly than the smaller forms that have survived. The case for a combination of direct and indirect human activity as the agent of extinction seems compelling.

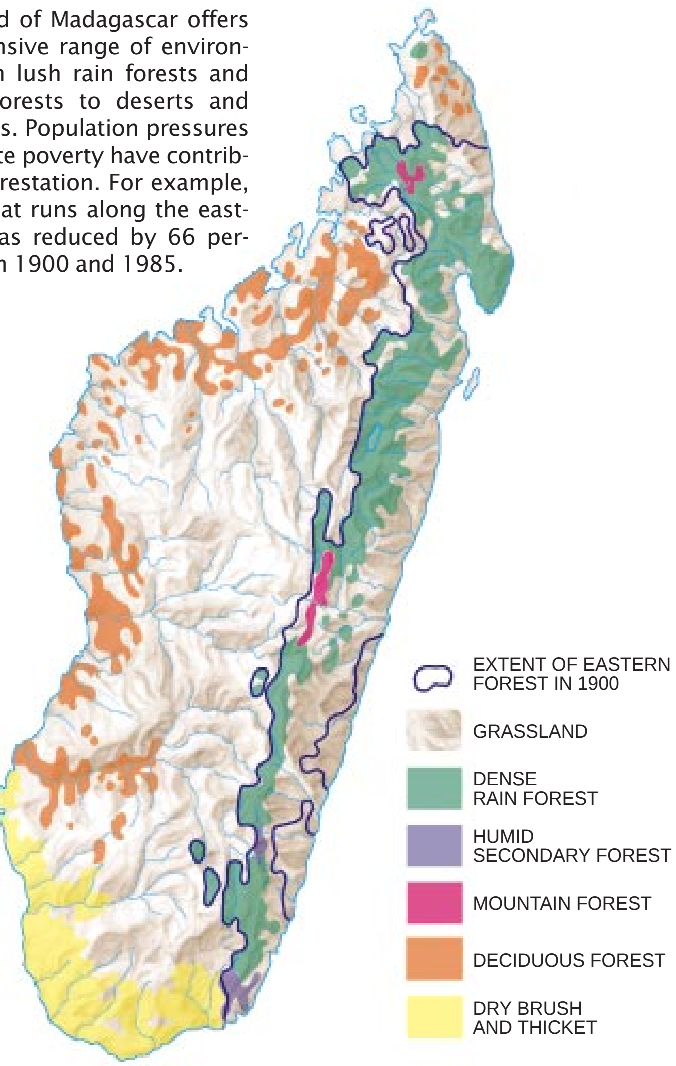
Climatic change has also been promoted as an agent of extinction, initially because many subfossil sites consist of dried-up lakes or marshes. As a total explanation of the extinctions, drying was never completely convincing. Still, interest in a possible climatic effect has been revived recently by the demonstration that some of the grasslands of Madagascar's center are of long (and certainly prehuman) standing.

Analyses of lake cores by David A. Burney of Fordham University have underscored the fact that, like every other region of the earth, Madagascar has undergone climatic fluctuations over the past few thousand years. At the end of the last ice age some 10,000 years ago, Madagascar's forests were apparently just beginning to recover from a period of contraction. It is therefore hardly surprising that the central highlands were not fully forested by the time Madagascar's first people arrived.

Climatic stress and the consequent reduction and redistribution of forests thus constitute one potential factor in the disappearance of the subfossil fauna of Madagascar. But it is clear

## Madagascar's Habitats

The island of Madagascar offers an extensive range of environments—from lush rain forests and deciduous forests to deserts and coastal plains. Population pressures and desperate poverty have contributed to deforestation. For example, the forest that runs along the eastern coast was reduced by 66 percent between 1900 and 1985.




SOURCES: Glen M. Green, Robert W. Sussman, H. Humbert and G. Cours Darne

**BURNED LAND is seeded for agriculture by Malagasy farmers (*left*). The slash-and-burn technique began when settlers first arrived on Madagascar under 2,000 years ago. Extensive deforestation for logging has also led to erosion on many slopes (*right*). The loss of habitat resulting from these two practices has contributed to the disappearance of large birds and mammals.**

that in that land as elsewhere, periodic disturbances of this kind have occurred throughout time. The ancestral lineages of the extinct lemurs obviously survived these earlier vicissitudes, and there is no reason to believe that on their own the most recent (rather mild) round of climate changes should have had a fatal effect on several lineages simultaneously. This is especially true when we consider that most subfossil lemurs were probably ecological generalists. *Megaladapis, Palaeopropithecus* and *Archaeolemur,* among others, lived in environments ranging from humid to arid and were clearly highly adaptable in terms of choice of habitat. Hence, something other than simply another cycle of disturbance of natural habitat must be invoked to explain the disappearance of the giant lemurs. Only one truly novel factor presents itself: *Homo sapiens.*

To talk of the "extinction event" of the large-bodied lemurs implies a process that has ended. That is not so. The extinct and living lemurs form part of a single process that is continuing. The smaller, fleeter lemurs have survived so far, but they are themselves under increasing pressure from an expanding human population. The toll of hunting increases as human numbers and the use of more sophisticated weapons do. Another important consideration is the population movements that tend to erode the local beliefs that, in certain places, have traditionally protected various lemur species.

More worrying yet is destruction of habitat, principally by slash-and-burn agriculture but also by the cutting of trees for fuel and by commercial logging. The largest continuous forest tract in Madagascar lies along the island's humid eastern escarpment. By analyzing historical documents and satellite images, Glen M. Green and Robert W. Sussman of Washington University have shown that 66 percent of the area covered by this forest at the turn of the century had been denuded by 1985. They estimate that within another 35 years only the steepest slopes of the escarpment will still bear trees. In the flatter western and southern regions of the island, the rate of disappearance of forest is probably faster.

Such pressures have existed for many decades. In the 1920s the colonial authorities in Madagascar set up one of the world's first systems of natural reserves. But as one of the world's poorest countries, Madagascar, despite its government's genuine concern, cannot afford to police these reserves adequately or, in some cases, at all. Fortunately, in recent years the island has attracted the attention of the international conservation community. The country figured recently in one of the first "debt for nature" swaps, in which foreign debt is reduced in exchange for the protection of natural areas.

It is, of course, the real and immediate needs of desperately poor local communities that are being met by most of the forest destruction. In many cases, large-scale agreements on conservation have yet to devolve from the higher governmental sphere to actual projects on the ground. But as they do, one can hope for the stabilization and perhaps even for the long-term improvement of Madagascar's environmental situation.

Meanwhile the lemur populations continue to dwindle. It is tragic to see any part of the world's biodiversity disappear, but the tragedy is particularly acute in the case of Madagascar's lemurs, which still have so much to teach us about our own past.

FURTHER READING

A WORLD LIKE OUR OWN: MAN AND NATURE IN MADAGASCAR. Alison Jolly. Yale University Press, 1980.

THE PRIMATES OF MADAGASCAR. Ian Tattersall. Columbia University Press, 1982.

ENVIRONMENT, EXTINCTION, AND HOLOCENE VERTEBRATE LOCALITIES IN SOUTHERN MADAGASCAR. R.D.E. MacPhee in *National Geographic Research,* Vol. 2, No. 4, pages 441–455; Autumn 1986.

LEMURS LOST AND FOUND. Patricia C. Wright in *Natural History,* Vol. 97, No. 7, pages 56–60; July 1988.

MADAGASCAR'S LEMURS: ON THE EDGE OF SURVIVAL. Alison Jolly in *National Geographic,* Vol. 174, No. 2, pages 132–161; August 1988.

MADAGASCAR: A NATURAL HISTORY. Ken Preston-Mafham. Facts on File, 1991.

# Quantum Dots

*Nanotechnologists can now confine electrons to pointlike structures. Such "designer atoms" may lead to new electronic and optical devices*

by Mark A. Reed

During the past few years, research in semiconductors has taken on, quite literally, new dimensions. Their numbers are two, one and zero. Electrons in recently developed devices can be confined to planes, lines or mathematical points—quantum dots.

Microchip manufacturers have developed a toolbox of nanofabrication technologies capable of creating structures almost atom by atom. These techniques have opened up a new realm of fundamental physics and chemistry as workers make and study artificial analogues of atoms, molecules and crystals. Experimenters are no longer limited by the atomic shapes, sizes and charge distributions available in nature.

In addition to the exciting science they portend, quantum dots promise properties that could be harnessed for a range of electronic and optical applications. Arrays of densely packed dots could form a substrate for computers of unprecedented power; indeed, Norman Margolus of the Massachusetts Institute of Technology has coined the term "crayonium." Dots could also constitute materials capable of absorbing and emitting light at whatever set of wavelengths their designers specify or could even serve as the basis of semiconductor lasers more efficient and precisely tuned than any now in existence.

Planes, lines and dots are mathematical constructs. They have no physical extent. How is it possible to make them in a real, three-dimensional material? The answer lies in quantum mechanics and Heisenberg's uncertainty principle. The position of an object (an electron, for instance) and its momentum cannot both be known to arbitrary precision. As an electron is more closely confined, its momentum must be more uncertain. This wider range of momenta translates to a higher average energy. If an electron were confined in an infinitely thin layer, its energy would also be infinite.

In general, the energy of electrons in a semiconductor is limited by their temperature and by the properties of the material. When the electrons are confined in a thin enough layer, however, the requirements of the uncertainty principle in effect override other considerations. As long as the electrons do not have enough energy to break out of confinement, they become effectively two-dimensional.

This locution is not just an approximation. Electrons confined in a plane have no freedom of motion in the third dimension. Those confined in a quantum wire are free in only one dimension, and those confined in a quantum dot are not free in any dimension. For common semiconductors, the length scale for a free conduction electron is about 100 angstroms. (One angstrom is $10^{-10}$ meter, approximately the radius of a hydrogen atom.) An electron inside a cube of semiconducting material 100 angstroms on a side is essentially confined to a point.

Engineering of less than three-dimensional semiconductors began in earnest during the early 1970s, when groups at AT&T Bell Laboratories and IBM made the first two-dimensional "quantum wells." These structures, made by thin-film epitaxial techniques that build up a semiconductor one atomic layer at a time, are thin regions of semiconducting material (usually gallium arsenide and related compounds) that attract electrons. The energy of electrons residing within the well is lower than the energy of those residing elsewhere, and so the electrons flow in, just as water runs downhill to fill up a deep well.

It is possible to create not only quantum wells but also quantum barriers—two-dimensional "hills" of material that repel electrons. In combination, the wells and barriers can be used to build complex structures that previously existed only as examples in quantum mechanics textbooks [see "Diminishing Dimensions," by Elizabeth Corcoran; SCIENTIFIC AMERICAN, November 1990].

Quantum wells have now become commonplace. They are the basis of the laser diodes found in compact-disc players and the sensitive microwave receivers that pull in signals from a satellite dish. In the meantime, researchers have learned how to confine electrons not simply in a plane but in a point.

The first hints that zero-dimensional quantum confinement was possible came in the early 1980s, when A. I. Ekimov and his colleagues at the Ioffe Physical-Technical Institute in St. Petersburg noticed unusual optical spectra from samples of glass containing the semiconductors cadmium sulfide or cadmium selenide. The samples had been subjected to high temperature; Ekimov suggested tentatively that the heating had caused nanocrystallites of the semiconductor to precipitate in the glass and that quantum confinement of electrons in these crystallites caused the unusual optical behavior.

To understand this chain of reasoning, imagine an electron trapped in a box. Quantum mechanics states that the electron has wave properties, like the ripples on water or the vibrations of a violin string. Just as a violin string is tied down at both ends, so the electron wave is bounded by the walls of the box. The wavelength of the string's vibrations (or the electron's) must fit within those confines [*see illustration on page 121*].

In the case of the violin string, the point at which it is tied down changes

MARK A. REED studies mesoscopic semiconductor systems—those whose behavior shows a mixture of classical and quantum mechanical traits. After receiving his doctorate in solid-state physics from Syracuse University in 1983, Reed joined Texas Instruments. In 1987 his team there made the first lithographically defined quantum dots. Since 1990 he has been a professor of electrical engineering at Yale University.

as the violinist's finger slides up the fingerboard. The length of the allowed waveform shortens, and the frequency of the string's vibrations increases, as does that of all its harmonic overtones. If the size of an electron's confining box is made smaller, the electron's lowest energy level (the analogue of the fundamental pitch of the violin) will increase. For semiconductor nanocrystallites, the fundamental "pitch" is the threshold energy for optical absorption, and the harmonic overtones correspond to new absorption features at higher energies.

How small must a nanocrystallite be for this phenomenon to be visible? In a vacuum the effects of confinement would begin to appear when the electron was trapped in a volume about 10 angstroms across. That size implies an electron wavelength of 20 angstroms and therefore an energy of about one fortieth of an electron volt.

Here semiconductor physics comes to the aid of the nanotechnologist. The wavelength of an electron depends on its energy and its mass. For a given wavelength, the smaller the mass, the larger the energy and the easier it is to observe the energy shift that confinement causes. The electrostatic potentials of the atoms in the crystalline lattice superimpose to provide a medium in which electron waves propagate with less inertia than they do in free space. The "effective mass" of the electron is thus less than its actual mass. In gallium arsenide the effective mass is about 7 percent of what it would be in a vacuum, and in silicon it is 14 percent. As a result, quantum confinement in semiconductors occurs in volumes roughly 100 angstroms across.

The optical absorption threshold for nanocrystallites of this size shifts to higher energies—away from the red end of the spectrum—as the crystallite becomes smaller. This effect appears most elegantly in cadmium selenide clusters; the progression from deep red to orange to yellow as the diameter of the

**QUANTUM CONFINEMENT is responsible for the colors of cadmium selenide crystallites, each a few nanometers across, synthesized by Michael L. Steigerwald of AT&T Bell Laboratories. Electrons within the tiny specks of semiconductor scatter photons whose energy is less than a minimum determined by the size of the crystallite and absorb those whose energy is higher. The largest crystallites can absorb lower-energy photons and so appear red, whereas the smallest absorb only higher-energy quanta and so appear yellow.**

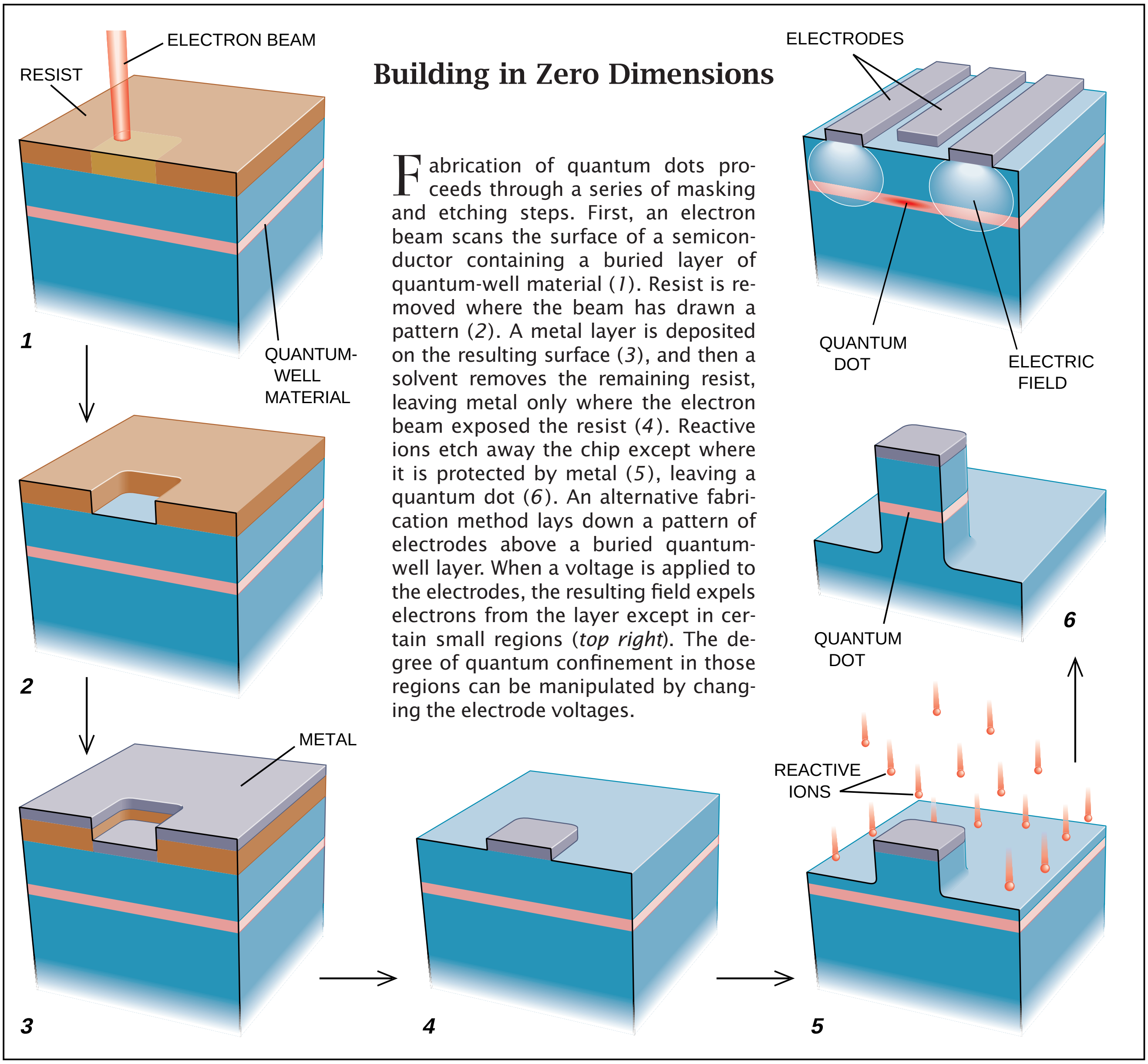
cluster declines can be clearly seen by the naked eye. (An intriguing and as yet unresolved question is what happens when the crystallite is so small—less than 10 angstroms across—that the effective mass concept, derived from bulk solids, no longer makes sense. Quantum dots that small have yet to be made.)

Ekimov's hypothesis turned out to be true, but it took years of work by groups at Corning Glass, IBM, City College of New York and elsewhere to sort out the correct glass preparation techniques and convincingly demonstrate quantum confinement. Meanwhile Louis E. Brus and his co-workers at Bell Labs were making colloidal suspensions of nanocrystallites by precipitation from solutions containing the elements that make up semiconductors.

Such crystallites grow by the addition of individual ions until the supply is either depleted or removed. Consequently, by arresting the precipitation after a certain time, Brus and his colleagues could control the size of the precipitate in a range between 15 and about 500 angstroms. Sizes within a single batch varied by no more than 15 percent. Just as in the case of the glass-encased nanocrystallites, a dramatic shift of the fundamental absorption energy to higher energies suggested quantum confinement.

Workers in many laboratories worldwide have built on this approach. For example, A. Paul Alivisatos and his colleagues at the University of California at Berkeley have extended the range of elements from which crystallites can be made. In addition to the II-VI compounds (such as cadmium from the second column of the periodic table and selenium from the sixth), they have also precipitated III-V compounds such as gallium arsenide. Michael L. Steigerwald of Bell Labs and many others have employed an organic "soap bubble" wrapping known as a reverse micelle to stabilize the surface of the tiny semiconductor crystals. Groups at the University of California at Santa Barbara, the University of Toronto and elsewhere are stuffing clusters of atoms into the nanometer-scale cavities of zeolites, a technique that confers the advantage of precise dimensional control.

Encasing nanocrystals inside another material could significantly improve their quantum performance. The tiny specks of semiconductor have a very large surface-to-volume ratio, and sur-




## Building in Zero Dimensions

Fabrication of quantum dots proceeds through a series of masking and etching steps. First, an electron beam scans the surface of a semiconductor containing a buried layer of quantum-well material (*1*). Resist is removed where the beam has drawn a pattern (*2*). A metal layer is deposited on the resulting surface (*3*), and then a solvent removes the remaining resist, leaving metal only where the electron beam exposed the resist (*4*). Reactive ions etch away the chip except where it is protected by metal (*5*), leaving a quantum dot (*6*). An alternative fabrication method lays down a pattern of electrodes above a buried quantum-well layer. When a voltage is applied to the electrodes, the resulting field expels electrons from the layer except in certain small regions (*top right*). The degree of quantum confinement in those regions can be manipulated by changing the electrode voltages.

faces in general are marked by atoms with dangling chemical bonds. These improperly terminated bonds can act as dampers, absorbing the energy of electrons vibrating in higher-energy (shorter-wavelength) modes. As a result, many nanocrystallites do not show the dramatic harmonic series of energy levels that would be expected from a quantum dot.

The inherent difficulties of making quantum dots from clusters of atoms led researchers in the mid-1980s to look for other fabrication schemes. My co-workers and I at Texas Instruments in Dallas made the first lithographic quantum dots in 1987. We cut slabs of quantum-well material into pillars by means of advanced etching techniques similar to those used in the fabrication of state-of-the-art integrated circuits.

Making pillars 100 angstroms wide requires electron-beam lithography instead of the optical techniques used to make most chips. An electron beam scans the semiconductor surface, which has been coated with a thin polymer layer called a resist. (Similar effects can also be achieved by means of x-rays or ion beams.) A series of process steps replaces the resist with a thin layer of metal in areas where the beam was scanned at high intensity. A shower of reactive gas then etches away the unprotected quantum-well material, leaving the pillars behind. Using this technique, pillars or other features as small as 1,000 angstroms across can be quite easily constructed. But the process becomes increasingly difficult as the scale falls to about 100 angstroms, the limit of the best-known resist.

Above and below the quantum-well material in these pillars lie ultrathin insulating layers called tunnel barriers, followed by conductive contacts. The insulators confine electrons in the well for a very long time, but eventually the electrons can quantum-mechanically tunnel in and out, carrying a small current that can serve to probe the internal energy states of the well. Whenever the voltage across the well matches the energy of one of its resonant states, current flow increases. If the diameter of the pillar is very small, its current-voltage spectrum displays the harmonic series of peaks that marks quantum confinement. Indeed, by making only a single pillar that is isolated from its surroundings, one can calculate the properties of a single quantum dot, a task that is hard to imagine carrying out with nanocrystallites.

Moreover, the lithographic fabrication process naturally clads and protects the quantum dot from surface effects, at least on two faces. The top and bottom of the dot are single crystal interfaces made by advanced epitaxy and are essentially perfect. Because the pillar is electrically conductive, the surface bonds of the semiconductor we used create a positive charge with respect to the internal core of the pillar. This charge repels electrons from the surface into the quantum-confined interior; the region from which the electrons have departed forms an insulating sheath around the pillar, protecting the sides of the dot. A 1,000-angstrom pillar would thus contain a 100-angstrom dot.

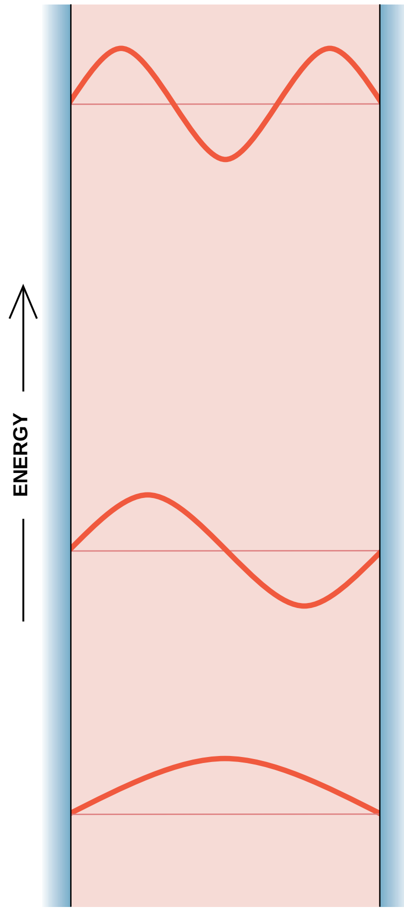
The realization of a quantum dot depends on an insulating sheath of the correct thickness, which depends in turn on the size of the etched pillar. When we first tried to make quantum dots, no one knew what the correct sizes might be, and many attempts ended in failure. But early on the morning of August 20, 1987, as I was preparing to give a talk at a conference on quantum-well devices, my colleagues called to say they had successfully measured a quantum dot. I rushed to the hotel fax machine just in time to see the data print out, showing a rich harmonic series of electron energy levels. An hour later I had rewritten the finale of my talk, the hotel staff had transformed the fax into a viewgraph, and I delivered the news.

Subsequent measurements confirmed that dots of different sizes produced different harmonic spectra, a clear signature of quantum confinement. Since then, groups at CNET in France, NTT in Japan, the University of Cambridge, the State University of New York at Stony Brook and Princeton University have also employed this fabrication technique. Pierre Gueret and his co-workers at IBM Zürich have even made a "squeezable" dot by placing an electronic gate around the dot in a tour de force of fabrication technology. Increasing the electric potential on the gate reduces the size of the dot and increases the fundamental energy and harmonics in its spectrum.

The success of these electrical measurements on lithographically defined quantum dots, in contrast to the relative difficulty of optical measurements on dots made from atomic clusters, has underscored the importance of controlling damaging surface effects. Groups at IBM, AT&T, the universities of Hamburg and Munich, Delft University of Technology in the Netherlands, Philips, Cambridge, the Max Planck Institute for Solid State Physics in Stuttgart and M.I.T. have managed to eliminate surface effects entirely. They make quantum dots by placing tiny gate electrodes on top of a buried layer that confines electrons in two dimensions. The top electrodes squeeze the electrons into quantum-confined "islands."

One advantage of this approach is that it is possible to put as many or as few electrons in the dot as desired, simply by varying the squeezing voltage. The result is what might be called a designer atom: the confining potential acts as an attractive nucleus, and the valency (the number of electrons) is determined by the external gate voltage.

In natural atoms, confinement of electrons is caused by the radially directed electrostatic force of the nucleus, and the electron wave functions are radially




**ELECTRON IN A BOX is constrained to have a quantum wave function that fits evenly within its borders. The lowest energy level corresponds to a standing wave with a single antinode, the next lowest to a wave with two, and so on. Electron energy is inversely proportional to the square of the wavelength, and so the energy levels rise rapidly. This harmonic series of energy levels is the signature of a quantum dot.**

symmetric. In these quantum dots the shape of the gate electrodes controls the size, shape and symmetry of the confining potential, and so "wave-function engineers" may eventually be able to study atomic physics previously inaccessible in nature, such as the wave functions or electrons in square or rectangular atoms.

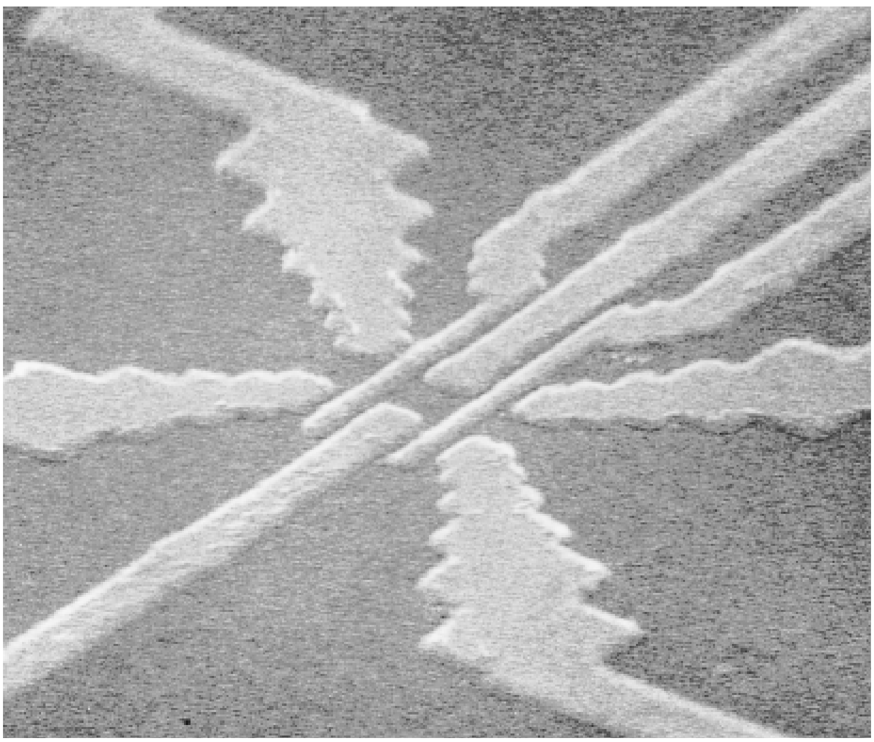
A group at the Stuttgart Max Planck Institute, as well as teams at IBM and AT&T, has made large, periodic arrays of dots by fabricating a gridlike gate electrode—the nanostructure equivalent of a window screen. Voltage applied to the grid forms a regular lattice of quantum confinement in the underlying material. The size and the number of electrons in each dot can be controlled, as can the height and thickness of the barrier between the dots. Regular peaks appear in the optical absorption spectra of these structures. This surprising phenomenon testifies to the precision with which the arrays—some containing more than a million dots—have been made, as any variation in size would smear out the harmonic spectra. Ray C. Ashoori and Horst L. Störmer of AT&T recently measured the capacitance of individual dots and demonstrated that it is possible to capture a single electron in each one. Electrons can then be added one at a time, in a digital fashion.

These results open up the possibility of making a planar artificial lattice in which virtually all the properties of the constituent "atoms" can be controlled. Just as individual quantum dots display energy levels analogous to those of atoms, an artificial lattice would possess an energy band structure analogous to that of a crystalline semiconductor. It could be used to study many questions in quantum physics and might also form the basis for a superfast electronic oscillator.

No one, however, has yet made a planar artificial lattice and unambiguously demonstrated its band structure. Success will require not only exacting precision in fabricating the electrode grid but also heroic control of defects in the underlying quantum-well material. In natural semiconductor lattices, engineers can rely on the fact that all silicon atoms, for instance, are identical, but in an artificial lattice, they will have to impose this uniformity by craft.

An intriguing twist in this genre is the "antidot" lattice. If the voltage on the grid is reversed, the islands that attracted electrons now repel them. Electrons are forced to reside in the intervening space, bouncing off the antidots as they move through the array in what is probably the smallest "pinball machine" ever made.

Another variant of the grid technique, demonstrated by Kathleen Kash and her co-workers at Bell Communications Research (Bellcore), uses compressive stress in place of electrodes to impose quantum confinement. The Bellcore team lays down a strained layer (a layer of material whose atomic lattice spacing differs from that of the substrate below it) on top of the quantum-well material, compressing it laterally. The workers then etch a pattern into the strained layer; wherever the layer is etched away, the compressive stress is relieved. The resulting tiny variations in atomic spacing within the quantum-well layer cause changes in electron energy levels that can form quantum dots.



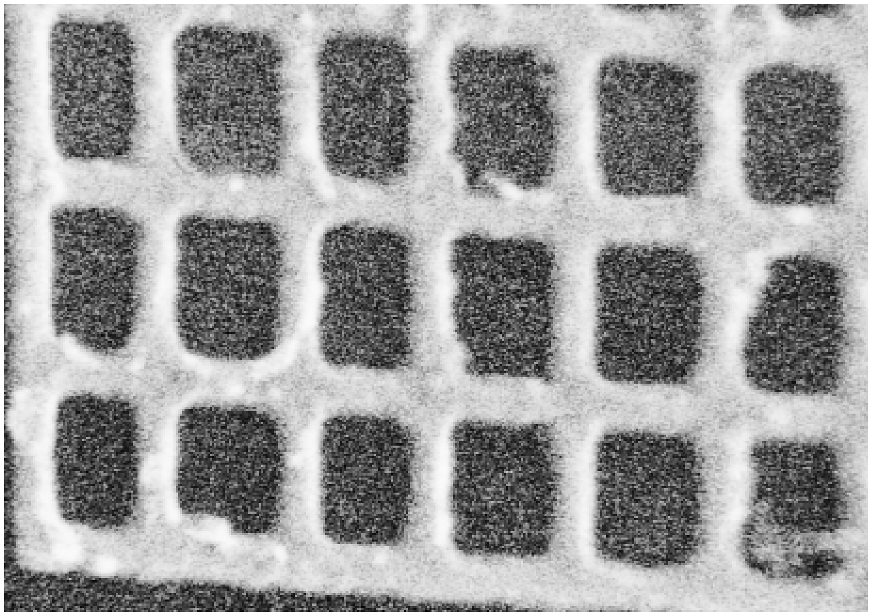
**ADJUSTABLE QUANTUM DOT lies buried at the intersection of electrodes in the above micrograph. The four interior electrodes "squeeze" electrons in the buried quantum-well layer into the dot. The outer electrodes serve as contacts for electrons to tunnel in or out of the dot; tunneling rates increase when the electron energies match the dot's energy levels. Those levels, in turn, can be controlled by changing the voltage on the inner electrodes.**

Electrostatic squeezing produces dots whose quantum confinement can be controlled more easily than can that of dots produced by other methods. Until recently, it was impossible to make electrodes small enough for tunnel-barrier contacts to a single electrostatically squeezed dot. During the past three years, a number of groups have succeeded in this task, producing lateral contacts to the dot that consist of electrostatically controllable tunnel barriers.

This structure gives the researcher control of many of the variables that define a dot, including size, number of electrons and transparency of the confining barriers. Such systems are ideal for testing textbook quantum mechanics problems, such as the properties of zero-dimensional states or the probability of electrons tunneling through barriers. By stringing two dots together to form an artificial molecule, one can investigate coupling between the states of adjoining quantum dots. And, as Leo P. Kouwenhoven of Delft University has demonstrated, it is even possible to string many dots together in a pearl-necklace fashion to generate an artificial one-dimensional crystal and to watch how the energy band structure of a crystal forms.

The Delft and M.I.T. groups have discovered that the energy levels of these small dots are determined not only by quantum mechanical rules based on size but also by the quantization of the electron charge [see "Single Electronics," by Konstantin K. Likharev and Tord Claeson; SCIENTIFIC AMERICAN, June 1992]. The energy level of a dot depends in part on its capacitance and the amount of charge contained within it, and the amount of charge must of course be a multiple of *e*.

The coexistence of these two kinds of quantization causes a complex interplay of effects. To understand which will be most important, one needs to know not just the wavelength and effective mass of the electron in the dot but also its electrical capacitance. If a dot is made from a metallic particle, it has many more conduction electrons than does a semiconductor; furthermore, the wavelength of the conduction electrons is only a few angstroms. As a result, in a 100-angstrom metallic dot, charge quantization exerts a much stronger effect, relatively speaking, than size quantization. The capacitance of the metal dot, however, is not so different from that of a semiconductor dot of the same size, and in the semiconductor the energies of the two effects may be approximately the same.



**ELECTRODE GRID creates a lattice of quantum dots in the material underneath it. Such a lattice is, in effect, a crystalline layer made of artificial atoms whose energy levels can be controlled precisely. Arrays of quantum dots aid in the study of fundamental physics and also may eventually be made into novel electronic or optical devices.**

The development of quantum dots is a culmination of 20 years of work, during which researchers have learned how to tailor electronic materials. Before the 1970s, research in solid-state science was confined to materials provided by nature. The refinement of ultrathin-layer epitaxy during that decade gave researchers the tools to fabricate the two-dimensional structures that dominate technology today. Extensions of that technology have now led to exploration of the one- and zero-dimensional domains. Before these discoveries can be applied on a commercial scale, however, a new generation of fabrication techniques must be developed.

The most challenging obstacle is to achieve essentially perfect control over the size and purity of these nanostructures. The "top-down" approach to fabrication—carving, dicing or squeezing semiconductors—may not be sufficient without revolutionary advances in materials and nanofabrication. Current prototype devices are large (although the active region of the device is quantum-sized, the electrodes and contact pads take up enormous space), and they operate only at very low temperature.

Moreover, these devices are made by electron-beam lithography, a fabrication technology that cannot be used to make large numbers of the complex circuits crucial for economic success. New lithographic tools that permit three-dimensional atomic-scale control, such as structured epitaxial growth or self-organizing molecular assembly, are needed. Indeed, it may be necessary to develop novel materials and synthesis techniques that blend conventional semiconductor technology with alternative approaches. Workers at the Fujitsu Laboratories in Japan, for example, have made quantum wires and dots from organic polymers. The location of conducting atoms within the polymer molecules is fixed, so this method offers much finer control than is possible with electron-beam lithography. If "bottom-up" assembly of quantum devices proves feasible, the methods now used to produce quantum dots will seem akin to making the books in the Library of Congress by whittling away at a large block of wood.

The most important challenge that researchers face, however, is not learning how to build quantum-confinement devices in quantity; instead it is to design useful circuits that exploit their potential. Although the technological size limits to quantum devices are in theory significantly smaller than those projected for silicon, the viability of quantum circuits will be determined by how well they compete in the marketplace against the coming decade's developments in conventional silicon technology.

Just as transistors found uses far beyond their initial role in radio receivers, so the ultimate application of quantum devices could be quite tangential to the tasks of digital computation and communications for which they have thus far been developed. If engineers can fabricate lattices containing millions or billions of quantum dots, specifying the shape and size of each one, they will be able to make any electronic or optical material of which they can conceive. Emission, absorption and lasing spectra could be precisely tailored, and a single slab of material could even be designed to contain a myriad tiny computers whose interconnections and internal architecture would change to match each new problem posed to them.

Even beyond the practical applications of quantum devices and the new intellectual territory they offer experimental physicists, quantum dots are exciting to researchers. The ability to manipulate matter on an atomic scale and create unique materials and devices with custom-designed properties has universal appeal. It marks a triumph of human ingenuity and imagination over the natural rules by which materials are formed.

FURTHER READING


TRENDS IN MATERIALS: DIMINISHING DIMENSIONS. Elizabeth Corcoran in *Scientific American,* Vol. 263, No. 5, pages 122–131; November 1990.

ENGINEERING A SMALL WORLD: FROM ATOMIC MANIPULATION TO MICROFABRICATION. Special Section in *Science,* Vol. 254, pages 1300–1342; November 29, 1991.

QUANTUM SEMICONDUCTOR STRUCTURES: FUNDAMENTALS AND APPLICATIONS. Claude Weisbuch and Borge Vinter. Academic Press, 1991.

NANOSTRUCTURES AND MESOSCOPIC SYSTEMS. Edited by Wiley P. Kirk and Mark A. Reed. Academic Press, 1992.

# The Mind and Donald O. Hebb

*By rooting behavior in ideas, and ideas in the brain, Hebb laid the groundwork for modern neuroscience. His theory prefigured computer models of neural networks*

by Peter M. Milner

Donald O. Hebb, one of the most influential psychologists of his time, began his adult life intending to be a novelist. Deciding that his calling required an understanding of psychology, he embarked on a course that led him into two decades of research. His studies culminated in 1949 with the publication of *The Organization of Behavior,* a keystone of modern neuroscience.

The monograph broke new ground by positing neural structures, called cell assemblies, which were formed through the action of what is now called the Hebb synapse. The cell-assembly theory guided Hebb's landmark experiments on the influence of early environment on adult intelligence. It foreshadowed neural network theory, an active line of research in artificial intelligence.

Hebb's book came at the right time because it flew in the face of behaviorism just as that school was losing its dominance. The behaviorists denounced explanations of behavior by association of ideas (which they called mentalism) and by the action of neurons (which they called physiologizing). But many psychologists had grown weary of the artificial theories these strictures had engendered, and they were captivated by Hebb's project and his engaging literary style. The book became a classic, and Hebb became a household word (at least in psychologists' households).

Hebb never claimed that his 1949 theory was firmly grounded in physiology. His model gave workers something to look for, and later, as knowledge of the brain grew, it became possible to frame his ideas in more realistic neural terms. None of this subsequent research has invalidated Hebb's basic hypothesis. Indeed, its influence appears in many areas of current research.

Hebb was born in Chester, a small fishing and boat-building town in Nova Scotia. His parents were physicians, and his two brothers and his sister followed in their parents' footsteps. But Donald demonstrated his independence early by studying English in preparation for a career as a writer, graduating in 1925 from Dalhousie University in Halifax. To earn his living while gestating his first novel, he taught school in his hometown. A year later he set out to see life, going west to work an eight-horse team on prairie farms. Then, failing to get a job as a deckhand on a freighter to China, he returned east and got a job as a laborer in Quebec.

In 1927 an aspiring novelist not only had to know life but also the works of Sigmund Freud. This was Hebb's introduction to psychology. He was sufficiently intrigued to apply to the psychology department of McGill University, where he was accepted in 1928 as a part-time graduate student. Again he supported himself by teaching and, again, what started out as a temporary interest verged on becoming a career. After one year he was made principal of an elementary school in a working-class district of Montreal. He was determined to make learning enjoyable, taking care to prevent schoolwork from being used as a punishment, instead sending miscreants out of class to play in the school yard. Hebb became absorbed in his educational experiments and seriously considered remaining in the profession. Two developments dissuaded him. He came down with a tubercular hip that confined him to bed for a year and left him with a slight limp. Then his bride of 18 months was killed in an automobile accident. He therefore decided to leave Montreal.

While confined to bed, Hebb wrote a master's thesis that involved him in the nature-nurture controversy. The thesis attempted to explain spinal reflexes as the result of Pavlovian conditioning in the fetus. He subsequently buried all references to this essay both because he changed his mind about its content and because he came to oppose psychological research that lacked an experimental foundation.

One of his examiners was Boris P. Babkin, a physiologist who had worked with Pavlov in St. Petersburg. He recommended that Hebb get some experience in the laboratory and arranged for him

PETER M. MILNER is professor emeritus of psychology at McGill University. He received a B.Sc. in engineering from the University of Leeds in the U.K. in 1942 and was immediately recruited for radar research by the novelist C. P. Snow. "My first serious encounter with psychology followed soon afterward," he says, "while working on the interface between radar displays and the human visuomotor system." He later went to Canada to work on nuclear energy. In 1948 he left physics to take a Ph.D. in physiological psychology at McGill under the supervision of Donald O. Hebb. Milner is co-discoverer, with the late James Olds, of the reward mechanism in the brain and has written extensively on the system and its implications for theories of reinforcement and learning. He has also written a textbook on physiological psychology as well as papers on attention, visual recognition, memory and the mind-body problem.

**DONALD O. HEBB made his name as a theoretician but was equally distinguished as a teacher. Here he appears in a seminar held in the late 1960s.**

to work with another Russian emigré, Leonid Andreyev. Hebb conditioned dogs and became less impressed with Pavlovian techniques. After much soul-searching as to whether he should continue in psychology, he decided in 1934 to burn his boats, borrow money and go to Chicago to continue his doctoral research under Karl S. Lashley.

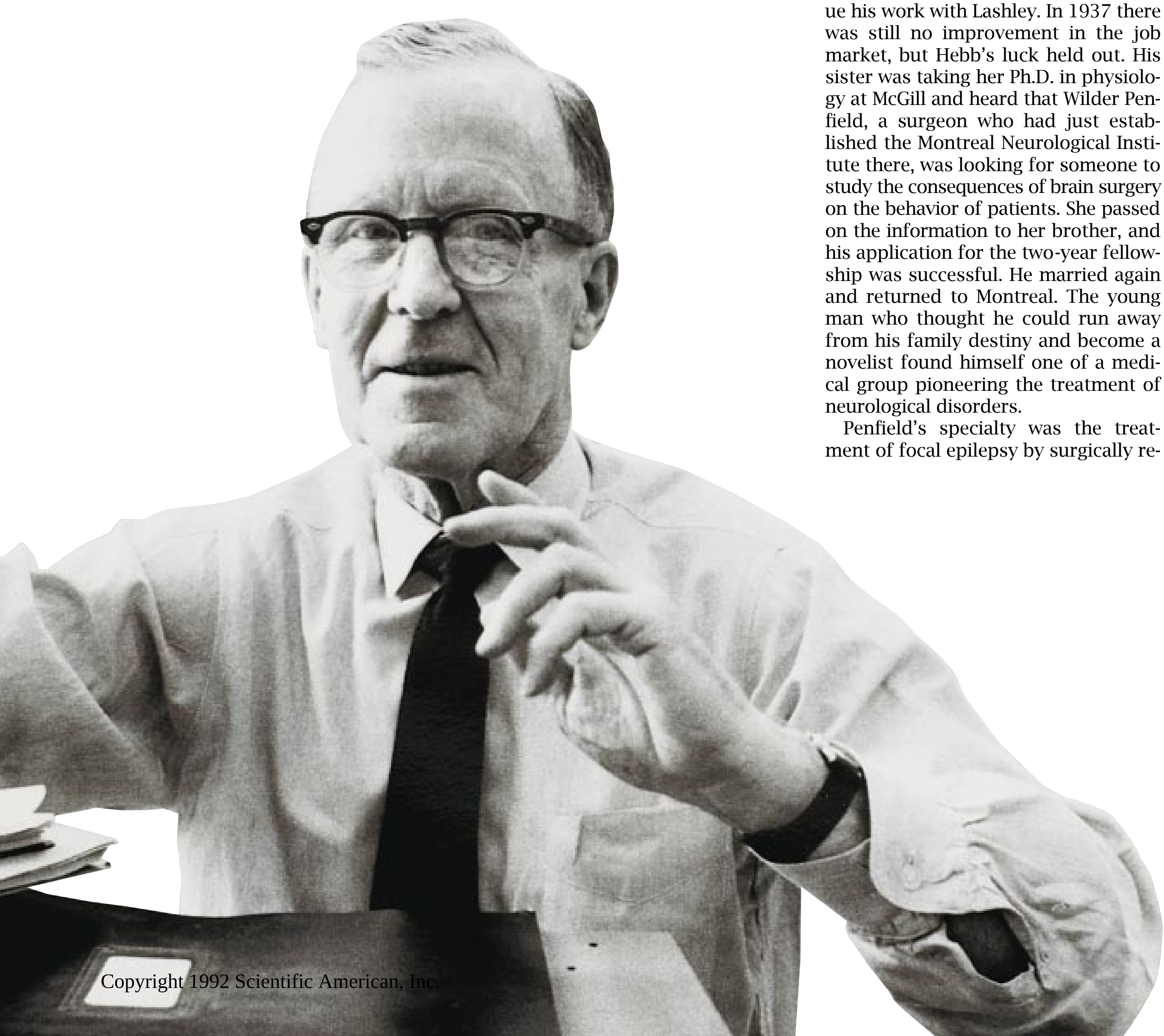
The elder scientist was to exert a profound influence on Hebb's approach, above all in his emphasis on physiology. Lashley had never doubted that to understand behavior one must first understand the brain. As a lab boy in 1910, he had salvaged slides of a frog brain from the trash heap and tried to find in the neural connections some clue to frog behavior. Lashley performed experiments to detect memory traces in the brain, inventing techniques for making brain lesions and measuring their location and extent. By around 1930 he had become convinced that memories could not be stored in a single region of the brain but must be spread throughout. In 1934, when Hebb went to Chicago, Lashley was concentrating on the study of vision.

A year later Lashley was offered a professorship at Harvard University and managed to take Hebb along. Hebb had to start his research from scratch, and having only enough money for one more year, he sought an experiment that could support a thesis no matter how it came out. He contrived to adapt his interest in the nature-nurture question to Lashley's vision project by investigating the effects of early experience on the development of vision in the rat.

Contrary to the empiricist ideas of his master's thesis, Hebb found that rats reared in complete darkness could distinguish the size and brightness of patterns as accurately as rats reared normally. This finding indicated that the organization of the visual system was innate and independent of environmental cues, a view coinciding with that of the Gestalt school, to which Lashley was sympathetic [see "The Legacy of Gestalt Psychology," by Irvin Rock and Stephen Palmer; SCIENTIFIC AMERICAN, December 1990]. What Hebb did not notice, although the results were included in a paper he published at the time, was that the dark-reared rats took much longer than normal rats to learn to distinguish vertical from horizontal lines. Only many years later, after he had again changed his ideas about the relative importance of innate and learned mechanisms, did he appreciate the significance of this result.

Hebb received his Ph.D. from Harvard in the middle of the Depression, when there were no jobs in physiological psychology to be had. He therefore stayed on for a year as a teaching assistant, a post that enabled him to continue his work with Lashley. In 1937 there was still no improvement in the job market, but Hebb's luck held out. His sister was taking her Ph.D. in physiology at McGill and heard that Wilder Penfield, a surgeon who had just established the Montreal Neurological Institute there, was looking for someone to study the consequences of brain surgery on the behavior of patients. She passed on the information to her brother, and his application for the two-year fellowship was successful. He married again and returned to Montreal. The young man who thought he could run away from his family destiny and become a novelist found himself one of a medical group pioneering the treatment of neurological disorders.

Penfield's specialty was the treatment of focal epilepsy by surgically re-



moving scarred areas of the cerebral cortex. He was acutely aware that he was operating on the organ of the mind and that a false move could deprive his patient of speech, intelligent behavior or even consciousness. Although Penfield was not a psychologist, his work exposed him to the relation between the mind and the nervous system. This experience no doubt influenced his decision to appoint psychologists to his team and explained the close interest he took in their findings.

Hebb's main responsibility was to study the nature and extent of any intellectual changes in patients consequent to cortical excisions. Such research was not new: it began after World War I with the psychometric testing of soldiers who had suffered penetrating head wounds and continued later in patients with brain tumors. In many cases, the lesions produced significant intellectual loss, but their locus and extent were difficult to determine. In contrast, surgical removals are more precisely defined, and epileptic scars do not cause the widespread damage that bullets or tumors do.

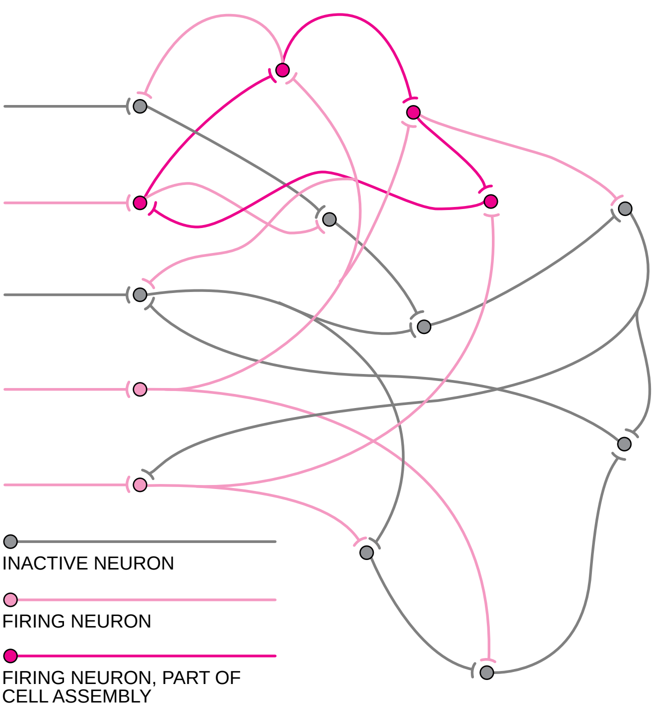
Hebb soon faced a peculiar problem. Psychologists then regarded the frontal lobes of the cerebral cortex as the seat of human intelligence, on the grounds that this region is relatively much larger than the corresponding areas in less intelligent animals. Yet Hebb was not able to detect intellectual loss in patients whose frontal lobes had been destroyed by accident or surgical necessity. This seeming lack of effect impressed Hebb deeply and inspired his quest for a theory of the brain and intelligent behavior.

Although his observations set him off on fruitful lines of inquiry, later work showed that Hebb had relied too heavily on standard intelligence tests. Brenda Milner, one of his students, who continued the work he had begun on Penfield's patients, found that frontal-lobe lesions often make it difficult for the patient to relinquish a behavior that has ceased to be appropriate. Although they may not be detected by intelligence tests, personality changes after frontal-lobe damage can profoundly affect the patient's life.

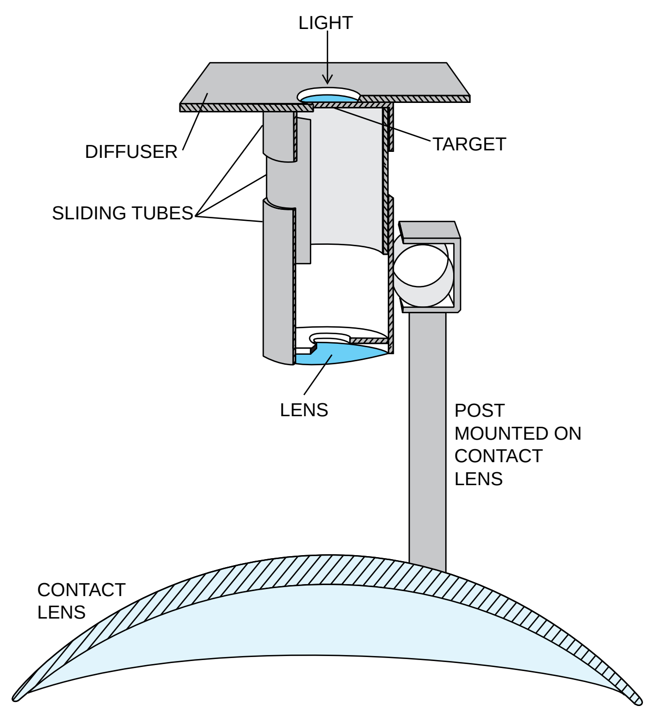
At the end of his fellowship at the neurological institute, Hebb finally found a permanent job at Queen's University in Kingston, Ontario. There, despite his heavy teaching load, he kept up work on the problem of intelligence. Together with a student, Kenneth Williams, he developed a variable-path rat maze as an analogue to human intelligence tests. The Hebb-Williams maze was widely used for the next quarter century. But Hebb was proudest of a theoretical paper in which he proposed that adult intelligence was crucially influenced by experience during infancy, basing his argument on the results of his research at the Montreal Neurological Institute. The paper was virtually ignored at the time, although it is now accepted almost as a commonplace, having been embodied in such preschool enrichment programs as Head Start. But the concept was too advanced for its time: in 1940 most psychologists practically defined intelligence as an innate characteristic.

To reconcile his studies of childhood influences with the apparent harmlessness of frontal-lobe lesions, Hebb hypothesized that the region's main function was not to think but rather to facilitate the tremendous acquisition of knowledge during the first few years of life. Experiments to determine the relative effects of early and late brain lesions did not always support this idea, but it provided a stepping-stone to Hebb's later theories.

In 1942 Lashley became the director of the Yerkes Laboratories of Primate Biology in Florida, and he invited Hebb to join his research team to study chimpanzee behavior. Hebb jumped at the chance of doing full-time research with



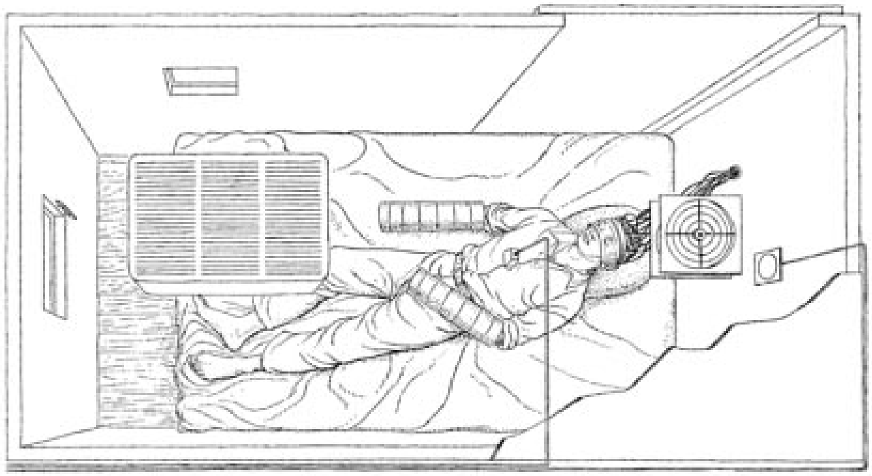

**HYPOTHETICAL CELL ASSEMBLY begins with parallel fibers connecting input from the retina to corresponding points in the primary visual cortex. These neurons, in turn, connect to the "association" cortex. Converging input fires cells and activates closed loops (*dark red*). Synaptic changes ensue that enable the loop to fire with little input, producing output that represents to the brain what the eye has seen.**




**RETINAL FATIGUE supports the cell-assembly theory by causing images to fade in a peculiar fashion. The apparatus fixes an image on receptors until their signal decays. Then lines drop out, one or two at a time, until the figure is gone. Hebb argued that each line was represented by a neuronal feedback loop. When the retinal signal falls below the critical value, the loop stops oscillating, and the line disappears.**

Lashley again, although he was not at first very enthusiastic about working with chimpanzees. Lashley's intention was to develop tests of learning and problem solving for the animals, while Hebb would study their personalities and emotional characteristics. Then they would start a program to determine how brain lesions affected a range of variables.

The chimpanzees proved more difficult to train than Lashley had imagined. The delays meant that no brain operations were carried out during Hebb's tenure at Yerkes. Nevertheless, he was fascinated by his observations of chimpanzees and said he learned more about human personality in his five years of watching chimpanzees than at any other time since his own first five years of life. The apes manifested distinct personalities and a sense of fun that tended toward slapstick. Hebb and the other members of the staff derived a more cerebral amusement from the verbal contortions of orthodox behaviorist visitors as they attempted to describe the animals' practical jokes and broad clowning without resorting to "mentalistic" language.



**ISOLATION EXPERIMENT carried the study of sensory deprivation beyond the realm of individual cell assemblies. Cuffs prevented touch, a plastic shield disrupted pattern vision and a U-shaped foam cushion attenuated sounds not masked by the air conditioner in the ceiling. EEG electrodes recorded the subject's brain waves, and a microphone enabled him to report his experiences. The volunteers' ability to think deteriorated, and some of them even started to hallucinate.**

Hebb's long and close observation of the many chimpanzees in the primate laboratory taught him that experience was not the only factor in the development of personality, including pathological manifestations such as phobias. He showed, for example, that young chimpanzees, born in the laboratory and known never to have seen a snake before, are frightened the first time they are shown one. Chimpanzees are also frightened of models of chimpanzee or human heads or other isolated body parts or of familiar caretakers wearing unusual clothing. Moreover, Hebb was one of the first to observe the social behavior of captive porpoises and to suggest that it implied a level of intelligence comparable to that of the apes. His observations may have influenced his later conclusion that level of play provides a good index of intelligence.

Lashley's interest in the ways the brain categorizes perceptions into knowledge about the world rekindled Hebb's curiosity about concepts and thinking. The problem can be rephrased as a question: How does the brain learn to lump one triangle, car or dog with another even though no two triangles, cars or dogs produce the same pattern of stimulation on sensory receptors?

The turning point came when Hebb read about the work of Rafael Lorente de Nó, a neurophysiologist at the Rockefeller Institute for Medical Research, who had discovered neural loops, or feedback paths, in the brain. Up to that point, all psychological theories, whether physiological or not, assumed that information passed through the organism along a one-way track, like food through the digestive system. Hebb recognized that Lorente's looping paths were just what he needed to develop a more realistic theory of the mind.

Feedback was not entirely new in learning theory. Almost all models assumed that the output of the organism influences the input in some way, for instance, by enabling the animal to receive a reinforcing stimulus. Unfortunately, feedback proceeding in this way, through a single path, would operate slowly and often unreliably. But with millions of internally connected feedback paths, it would clearly be possible to establish internal models of the environment that might predict the effects of possible responses without having to move a muscle.

Hebb's specialization in vision led him to concentrate his early neural theories on that system. Knowing that the point-to-point projection from the retina to the cortex does not extend beyond the primary visual cortex, he assumed that the neural relays projected into the surrounding cortex in random directions, thus scrambling the retinal pattern [see "The Visual Image in Mind and Brain," by Semir Zeki; SCIENTIFIC AMERICAN, September 1992]. Such an arrangement could recombine signals from different parts of the image—that is, they could converge on the same target neuron, causing it to fire. The resulting impulses could then return to the earlier neurons in the path, closing the feedback loops.

Repeated activation of any given loop might then strengthen that loop in the following way. If the axon of an "input" neuron is near enough to excite a target neuron, and if it persistently takes part in firing the target neuron, some growth process takes place in one or both cells to increase the efficiency of the input neuron's stimulation. Synapses that behave according to this postulate became known as Hebb synapses—somewhat to Hebb's amusement, it may be said, because this postulate is one of the few aspects of the theory he did not consider completely original. Something like it had been proposed by many psychologists, including Freud in his early years as a neurobiologist.

Nevertheless, Hebb's postulate was the most clear and formal statement, although in 1949 it was pure speculation. Since then, however, studies of single neurons have confirmed that synaptic strengths do change in some neurons in accordance with the postulate. Hebb may also have been correct about the mechanism of permanent change. A former student of his, Aryeh Routtenberg of Northwestern University, has recently pointed out that a protein associated with neuronal growth is produced when neurons are stimulated in ways that increase synaptic strength.

Hebb assumed that most of the synapses in the cortical lattice are initially too weak to fire spontaneously. To fire, they would require the converging of stimulation from a number of active neurons. Some neurons in the lattice receive converging inputs and thus fire when a particular pattern of neurons in the sensory cortex is fired by a stimulus. Some of the activated neurons have

**HEBB'S INFLUENCE propagated as much through his disciples as through his publications. Here, in a graduate seminar from the early 1950s, Hebb appears at the far right, the author in the foreground. The participants went on to pioneer the new field of physiological psychology.**

BEN DOAN *Dalhousie University* — HELEN MAHUT *Northeastern University* — PETER MILNER *McGill University* — JIM OLDS *Caltech* (DECEASED)

synaptic connections with one another, which are also strengthened whenever the stimulus is presented. Eventually the connections between the simultaneously firing neurons in the lattice become strong enough for them to continue firing one another in the absence of input from the stimulus, creating an internal representation of the stimulus, called a "cell assembly" by Hebb.

The concept of the cell assembly, in my view, was Hebb's greatest contribution to psychological theory, not to mention philosophy. It revived the 19th-century psychologists' attempt to explain behavior in terms of the association of ideas, a project that the behaviorists had derailed by arguing that "ideas" were no more real than the notion of little men inside the head. By so arguing, the behaviorists maintained that ideas, and thus mentalism, had no place in scientific psychology.

Unfortunately, few seemed to notice that the behaviorists replaced ideas with equally insubstantial constructs with misleading names, such as "stimuli" and "responses." These were not real events or chains of events but attributes that became associated with one another in some imaginary black box that scientists were forbidden to refer to as the brain. Hebb put a stop to this charade by showing, in principle at least, that ideas could have just as firm a physical basis as muscle movements. They could consist of learned patterns of neuronal firing in the brain, initially driven by sensory input but eventually acquiring autonomous status.

In its original form the neural theory was undoubtedly too simple to have worked. A major problem was that the cell assembly did not incorporate inhibition, because contemporary science did not recognize it. Sir John C. Eccles, a very influential neurophysiologist at the Australian National University in Canberra, was still vigorously denying the existence of inhibitory synapses. Moreover, many important connections of the neocortex had not yet been discovered, and the functional significance of the diversity of cortical neurons was only hinted at.

Without inhibiting factors, however, learning would strengthen synaptic connections until all neurons fired continuously, making the system useless. This effect was observed in computer models of the cell assembly, called conceptors, constructed in the 1950s by Nathaniel Rochester and his colleagues at the IBM research laboratory in Poughkeepsie, N.Y. Hebb himself seems never to have set finger to a computer to test his idea that random nerve nets could organize themselves to store and retrieve information. But such so-called neural nets later inspired many computer models, from the perceptron to parallel distributed processing, and have even found applications in industry.

By the time *The Organization of Behavior* reached publication, Hebb was back in Montreal as chairman of McGill's psychology department. Ten years later, when he stepped down as chairman, he had forged one of the strongest departments in North America. He found it easier to build what he wanted because the department was almost nonexistent when he began, and he turned out to be adept at campus politics and soon discovered how to use his growing reputation to apply pressure where it would do the most good. It is perhaps significant that he was also one of the best chess players at the university.

Most of Hebb's research at McGill was related to his cell-assembly theory. Experiments to obtain direct physiological evidence for the theory were far beyond the scope of contemporary methodology. (They still are.) Instead he tested behavioral predictions of the theory. He tried, for instance, to strengthen his earlier conclusions on the influence of rearing on adult intelligence. Most of the results supported his theory that animals raised in an enriched, or more complex, environment would, in later life, outperform animals raised in bare cages.

There was one embarrassing exception. Litters of pure-bred Scotties were split, and half the pups were reared as pets in the homes of members of the staff and half were reared in cages in the laboratory. Hebb was not fortunate in the choice of his puppy, Henry. It was congenitally incapable of finding its way around, invariably got lost as soon as it was out of sight of the house and had to be recovered from the dog pound on several occasions. Naturally, Henry turned out to be near the bottom of the class when, as a full-grown dog, it was tested in a maze.

In a related series of experiments, Hebb investigated the effect of impoverished sensory input on the behavior of adults, including human volunteers [see "The Pathology of Boredom," by

G. ROLFE MORRISON *McMaster University* — UNIDENTIFIED — SETH SHARPLESS *University of Colorado, Boulder* — DONALD O. HEBB *McGill University* (DECEASED)

Woodburn Heron; Scientific American, January 1957]. Students were paid generously to undergo severe sensory deprivation for as long as they could stand it (none lasted even a week). Their ability to think began to deteriorate, and some of them even started to hallucinate. The Korean War was then in progress, and many workers attempted to use such isolation experiments to understand and combat the "brainwashing" techniques employed by the Chinese.

Hebb also pursued his old idea that early brain injury should be more damaging than injury in an adult. But the results were rendered uncertain by several factors, the most important being the capacity of the young brain to reorganize itself. For example, if an infant sustains an injury in an area of the left hemisphere that is important for speech in the adult, the right hemisphere takes over this function, and speech is not seriously impaired. But if an adult sustains damage in the same area, the result may be a permanent loss of language skills.

Because of such problems with the study of cognition, Hebb came to believe that the best evidence for the cell assembly came from experiments on retinal fading. Images of simple figures were projected onto the eye by a very small lens system attached to a contact lens, ensuring that the image always fell on the same place. As the receptor cells become fatigued, the image fades and disappears, but not all at once. Usually entire lines disappear suddenly, one or two at a time, until the entire figure is gone. Hebb explained the phenomenon by saying that each line is represented by neuronal activity circulating in a closed loop. The activity, once started, continues even after the input from the retina has decayed to a low value because of feedback around the loop. But at some critical value the reverberation stops abruptly, and the line disappears. These experiments do not provide conclusive evidence for the cell assembly as Hebb envisaged it. Yet even if Hebb's version should turn out to be incorrect, it would not diminish the value of his idea that some neural activity continues to symbolize an object even after the object has stopped stimulating the sense organs.

Had *The Organization of Behavior* consisted only of the chapters in which Hebb criticizes current approaches and elaborates his cell-assembly theory, it is likely that few people would have read it. The book's appeal lies in its second half, in which Hebb discusses emotion, motivation, mental illness and the intelligence of humans and other species in the light of his theory. These essays are refreshingly forthright. On mental health, for example, Hebb wrote: "We still need an Ajax to stand up and defy the lightning and ask, What is the evidence? when some authority informs the public that believing in Santa Claus is bad for children, that comic books lead to psychological degeneracy, that asthma is due to a hidden mental illness."

Hebb built his department and his field by capturing the interest and imagination of the best students at an early stage. He taught the introductory course himself, making it immensely popular—at one point it numbered 1,500 students, about half the yearly undergraduate enrollment. Many future professors of psychology found their calling in these lectures. Like most of what Hebb did, his course was unique; no textbook at the time came close to including the material and ideas he dealt with, so he wrote his own. The first edition of *A Textbook of Psychology* appeared in 1958. In contrast to the majority of introductory texts of the day, it had more ideas than pictures.

Hebb also gave a graduate seminar that was attended by every psychology graduate student at McGill over a period of 30 years. It was famous not only for its stimulating discourse but also for Hebb's ever-present stopwatch and the slips of paper on which he noted incorrect pronunciations and other errors of presentation. It was Hebb's ambition never to have a McGill student overrun his or her allotted time at a meeting, and on the whole he was successful. McGill honored Hebb in 1970 by naming him chancellor; he became the only faculty member ever appointed to that position.

In 1977 Hebb retired to his birthplace in Nova Scotia, where he completed his last book, *Essay on Mind.* He was appointed an honorary professor of psychology at his alma mater, Dalhousie, and regularly participated in colloquia there until his death, at 81, in 1985.

FURTHER READING


THE ORGANIZATION OF BEHAVIOR: A NEUROPSYCHOLOGICAL THEORY. D. O. Hebb. John Wiley, 1949.

ESSAY ON MIND. D. O. Hebb. Lawrence Erlbaum Associates, 1980.

PARALLEL LEARNING IN BRAINS AND MACHINES. G. Ferry in *New Scientist,* Vol. 109, No. 1499, pages 36–38; March 13, 1986.

TEXTBOOK OF PSYCHOLOGY. Fourth edition. D. O. Hebb. Lawrence Erlbaum Associates, 1987.

MIND AND BRAIN. Special Issue of *Scientific American*, Vol. 267, No. 3; September 1992.

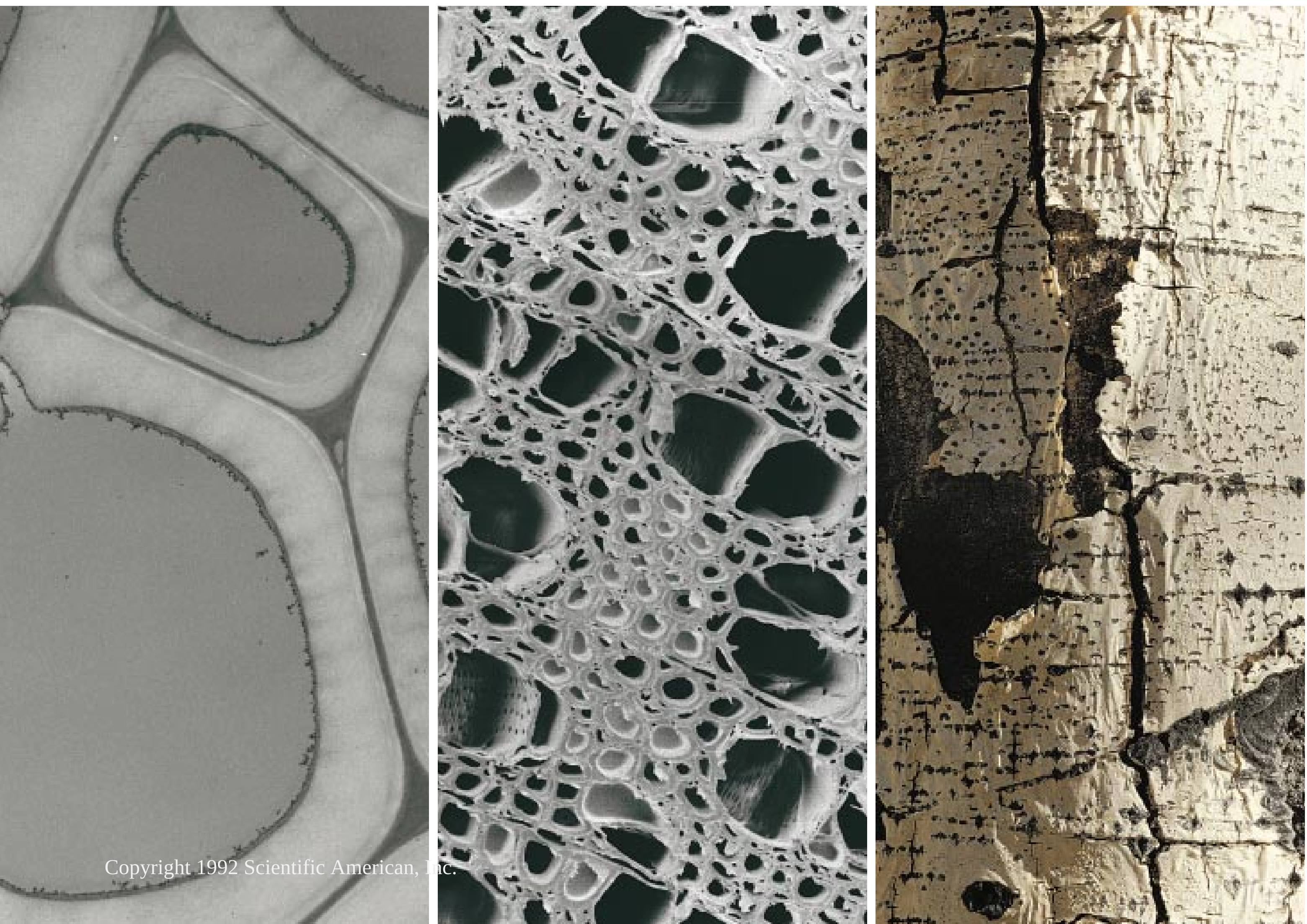
# ADAPTING TO COMPLEXITY

by Russell Ruthen, *staff writer*

On an autumn afternoon John Holland strolls among the pines and aspens that blanket, like a green and golden patchwork quilt, the mountains just north of Santa Fe, N.M. What impresses Holland about the scene is not just its beauty but also its complexity. The aspens compete with the pines for light as well as for water and nutrients from the rocky soil. As each tree grows and adapts, it alters the supply of resources available to its neighbors, and in doing so, it changes its own chances for survival.

Holland and his colleagues at the Santa Fe Institute in the valley below are searching for a unified theory that would explain the dynamics of all living systems, be they groves of trees, colonies of bacteria, communities of animals or societies of people. All are systems of many agents, each of which interacts with its neighbors and, most important, can adapt to change.

The loose group of economists, biologists, physicists, computer scientists and researchers from many other disciplines who have assembled at the institute hope to discern the principles that underlie all such complex adaptive systems. In the process, they have devised new theoretical frameworks, mathematical tools and computer simulations that attempt to capture the essence of complex adaptive systems. And they are gain-

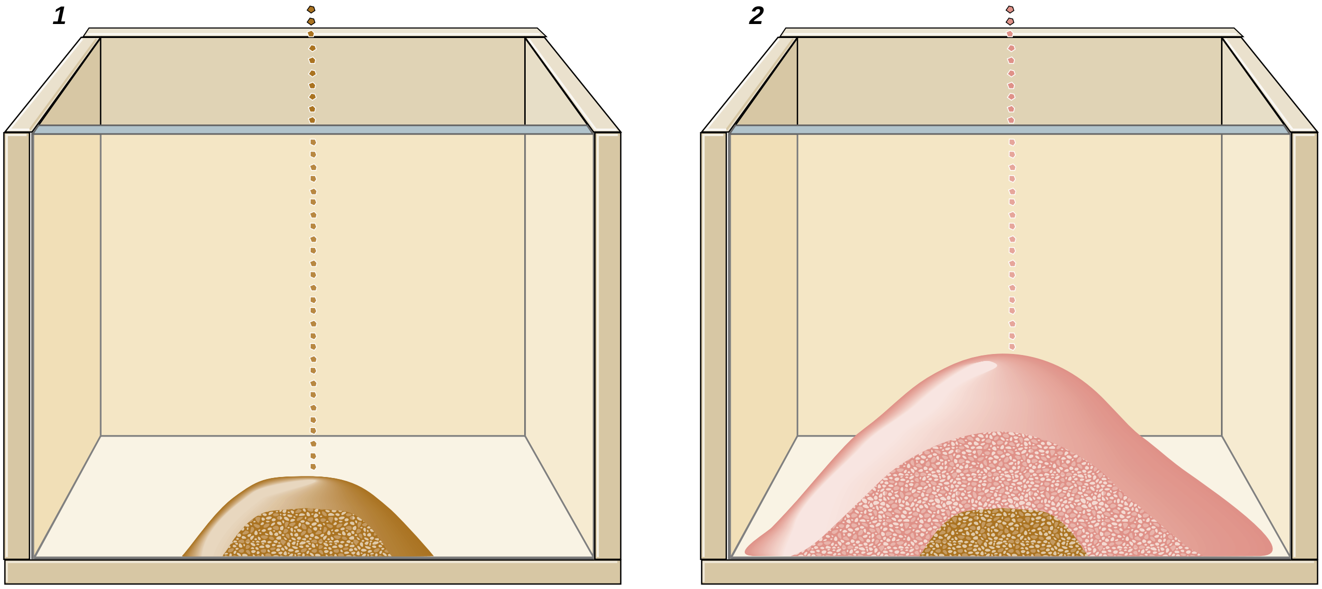


**From a primeval sea of organic molecules arose plants, animals, global ecosystems, intelligent beings, international organizations. What drives the natural world toward complexity?**

ing a better understanding of what complexity is, why it emerges naturally and why living systems seem to evolve toward a boundary between order and randomness.

Indeed, the Santa Fe group is trying to answer the question of evolution in the broadest sense of the word. "As these various systems organize themselves and learn and remember, evolve and adapt, persist and eventually disintegrate and disappear, what common patterns and fundamental principles, if any, shape their remarkable behavior?" wonders George A. Cowan, one of the founders of the Santa Fe Institute. Holland, a computer scientist at the University of Michigan who helped pioneer the field, has no illusions about the speculative nature of the enterprise. When asked what kind of complex adaptive systems scientists really understand, he just chuckles. "The really safe answer is none at all."

*QUAKING ASPENS—and all living things—display complexity at many levels of structure as the numerous components of the system grow and adapt. Leaves, branches and roots arise from the interaction of cells and their genetic material. Forests and ecosystems emerge from the interplay between the trees and their environment. To predict the behavior of such adaptive systems, researchers are consolidating the theories of evolution, the laws of dynamics and the power of computer simulation.*




**SAND PILE illustrates the theory of self-organized criticality. When grains of sand fall on a flat pile, they quickly find a resting spot (*1*). As grains continue to fall, the pile grows to a critical state in which a falling grain can trigger an avalanche**

To be sure, this generation of scientists is not the first to attempt to understand the behavior of these systems. In the 18th century Newton and his contemporaries hoped they could predict the behavior of any complex system by identifying all the parts and studying all the interactions. But that hope soon vanished.

In the 19th century Sadi Carnot and his peers realized it was impractical, if not impossible, to describe every interaction, because most physical systems contained zillions of parts. They discovered, however, that they could predict the statistical behavior of a system as long as the parts were identical and the interactions were weak. These statistical predictions became the laws of thermodynamics; they could explain, for example, the increase in temperature and pressure when gas molecules are heated in a container. Yet thermodynamics, Carnot knew, did not provide a complete description of the most complex systems, particularly those in which the interactions are not weak, as they would be if the gas molecules were strongly attracted to one another.

### From Order to Chaos

During the past 100 years, Henri Poincaré and his intellectual disciples realized that if a system consisted of a few parts that interacted strongly, it could exhibit unpredictable behavior. They invented chaos theory. For example, if three planets orbit around one another influenced only by the force of gravity, it is impossible to predict the motions for a long period of time, even if the positions and velocities of the planets are known with great accuracy.

Beginning in the 1970s, physicists flirted with the idea that chaos theory could account for the behavior of complex systems. Although chaos theory provided many mathematical tools useful for the study of complexity, it did not capture the wide range of dynamics exhibited by complex systems. "The explosive development of computer hardware and the development of new mathematical concepts," Cowan says, "helped raise the level of interest in complexity beyond explorations of chaos."

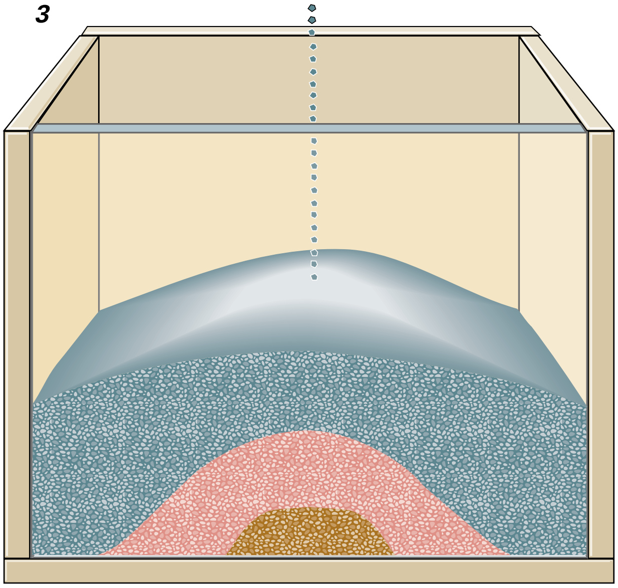
In fact, many years before the theory of chaos became popular, a small group of scientists, including Herbert Simon, Ilya Prigogine, Herman Haaken and others, had been searching for a theory of complexity, the general principles for a system whose many parts interact to produce complex behaviors. These workers had some success in describing complexity in physical systems.

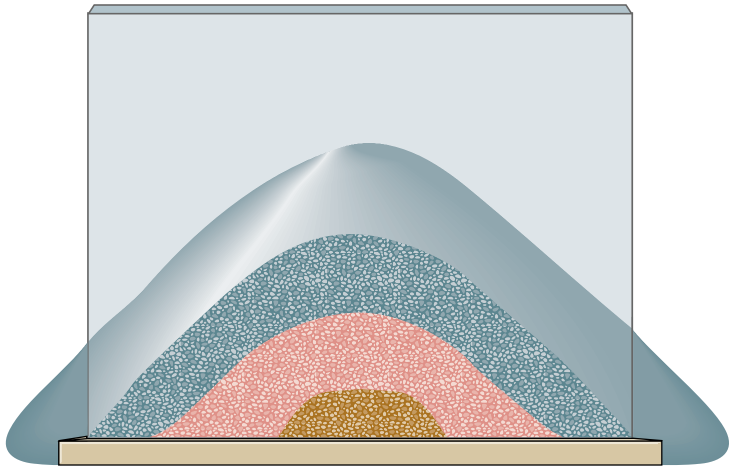
When Cowan began directing the Santa Fe Institute eight years ago "to nurture research on complexity," investigators there began to wonder if they could apply the emerging theory of complexity to adaptive systems, such as cells, organisms and economies. Adaptive systems, like other complex systems, consist "of many relatively independent parts that are highly interconnected and interactive," Cowan explains. "The systems of interest are not in thermodynamic equilibrium—they metabolize, absorbing energy from an external source and dumping back waste."

An adaptive agent can be anything from a single-cell organism to human society; it must be capable of forming and changing strategies. For instance, an aspen will orient its leaves to capture the most sunlight, a strategy that is encoded in its genes, and can change through mutation or recombination of its genetic material. Complex systems can "learn and adapt to changing conditions in the environment with which they interface," Cowan says.

An adaptive system, then, consists of many agents, each interacting with others that have the same or different strategies. The system's behavior obviously depends on the sophistication of the strategies and the mechanism of change. Yet adaptive agents can develop good solutions to extraordinarily difficult problems. Holland is intrigued at the way a group of independent companies somehow manage to supply food for the seven million people in New York City. The companies constantly replenish the restaurants and supermarkets, which do not keep more than a few days' worth of food in reserve, without creating shortages or surpluses. "From the point of view of physics, it is a miracle that happens without any control mechanism other than sheer capitalism," Holland exclaims.

To analyze the behavior of such systems, researchers must rely heavily on computer simulation because the equations that describe the forces underlying complex processes prove too difficult to solve. Holland, for one, simulates the process by which economic and biological agents adapt [see "Genetic Algorithms," by John H. Holland; SCIENTIFIC AMERICAN, July 1992]. "I can caricature a process and get a model




**of any size (*2*). Because of the avalanches, the slope of the pile remains constant. If some external force alters the shape of the pile but is then removed—a box, in this case (*3*)—the pile will return to the critical state (*4*).**

that is almost certainly overly simple but captures features in the same sense that a political cartoon captures something," he comments.

### Echoing Evolution

Four years ago Holland began work on a simulation of an ecosystem in which digital organisms try to survive and reproduce. He calls it "Echo," a play on the words "ecosystem" and "echo." All organisms are endowed with "chromosomes" that encode offensive and defensive strategies. Each organism wanders through an artificial environment, searching for resources and absorbing them into an internal reservoir. If one organism encounters another, the two fight, and the winner consumes the loser, acquiring all the resources contained in the loser. If an organism obtains enough resources, it produces offspring whose chromosomes may contain mutations. New species arise with ever more elaborate strategies for offense and defense.

Holland was surprised that such a simple simulation could generate speciation and arms races. Yet it was missing a key feature of evolutionary processes: cooperation. Organisms, or for that matter companies and nations, learn to cooperate even as they may compete aggressively with one another. Holland thinks he has a possible answer, which was inspired by the work of Robert Axelrod on a problem known as the Prisoner's Dilemma. In this problem, the police take two criminals into custody to extract a confession. If both prisoners confess, they both go to jail. If one confesses and the other does not, the confessor is granted immunity and is set free; the one who keeps silent goes to jail. If both prisoners keep silent, they are set free, although they risk prosecution if new incriminating evidence is found.

Clearly, the two prisoners would benefit by cooperating. Yet both will be tempted to seek immunity and fear that the other will confess. This logic leads them both to confess. Similarly, two organisms may benefit from cooperation, but one may receive the greatest payoff if it can obtain the resources of another without sacrificing its own. The upshot is the organisms may fight.

If adaptive agents are confronted many times with a situation similar to the Prisoner's Dilemma, they develop strategies that include both combat and cooperation. One of the best strategies is known as tit for tat: an agent will cooperate the first time it encounters an opponent and thereafter mimics the previous response of its opponent. If agent A tries to cooperate and agent B attacks, agent A will fight on the next round. But if agents A and C cooperate initially, they will continue to do so.

This dynamic was just what Holland needed for his digital ecosystem. He made three major modifications. First, each organism can choose to fight for resources or trade for them. Second, each is assigned a tag, which is analogous to a molecular marker on the membrane of a cell. Last, each organism is allowed to develop rules such as "fight only those organisms that carry a red tag." If strategies and tags are initially assigned at random, some organisms, by luck, are more inclined than others to use a strategy resembling tit for tat. Those that play such strategies are then more likely to survive and reproduce, thereby perpetuating the existence of particular tags.

The changes had an immediate effect. "What started out as a completely random juxtaposition of tags and strategies begins to develop in a very organized way," Holland declares. Eventually, the organisms "learn" to associate tags with certain strategies. In the end, the digital ecosystem displays both speciation and cooperation—and such deviations as mimicry and lying.

The apparent ability of Holland's simulations to reflect nature have piqued the interest of a small group of economists, including W. Brian Arthur of Stanford University and Nobel laureate Kenneth J. Arrow, who believe economies can also be regarded as complex adaptive systems. "We must view the economy as an evolving, changing system," Arthur remarks.

Such activities as international trade, high-technology business and the emergence of new companies exhibit truly complex dynamics, Arthur argues, because they involve both negative and positive feedback mechanisms. Negative feedback maintains the balance between supply and demand. Positive feedback, on the other hand, is destabilizing. If a firm gets ahead, it gains further advantage so that a clever strategy or luck can decide what company becomes a leader in a field [see "Positive Feedbacks in the Economy," by W. Brian Arthur; SCIENTIFIC AMERICAN, February 1990].

Yet a mixture of positive and negative

feedback presents two challenging problems for economists. First, it means that any given economic system can evolve down many possible paths. Economists must therefore figure out how a system ends up on one particular path. Second, a firm that is choosing a product strategy must try to guess which strategies other firms might use, knowing that small changes in its own strategy might alter the direction of others. Thus, economists must decide how firms behave in such complicated environments.

To solve these problems, Arthur and his cohorts are using "Holland-like" computer simulations. From their perspective, the economy consists of many firms (or consumers), each continually forming hypotheses about the environment they face and the strategies of their rivals. As a firm adds to its knowledge, it strengthens its belief in some hypotheses while discarding others. It chooses a strategy from these hypotheses, thereby influencing the environment in which other firms form their hypotheses. "The resulting dynamics are so complicated," Arthur says, "they can only be studied on a computer."

Arthur is bemused by the idea in conventional economic theory that each agent arrives at the absolute best strategy given the circumstances. The theory means, he quips, that "an economic agent can solve in an instant a problem that a theoretical economist might have to think hard about to formulate over a period of two years." Arthur hopes the study of complex adaptive systems will lead to a much better understanding of how economic agents form strategies.

Yet computer simulations go only so far. Although they mimic some of the attributes of complex adaptive systems, they provide little help to researchers who want to examine the complexity of existing systems. Some of the best ideas for defining and measuring complexity emerged from the work of Claude E. Shannon, Andrei N. Kolmogorov and Gregory J. Chaitin during the 1950s and 1960s. They proposed that the degree of complexity is related to the size of the smallest description of a system's behavior. This theory led to a proposed technique for measuring the so-called algorithmic complexity of a process.

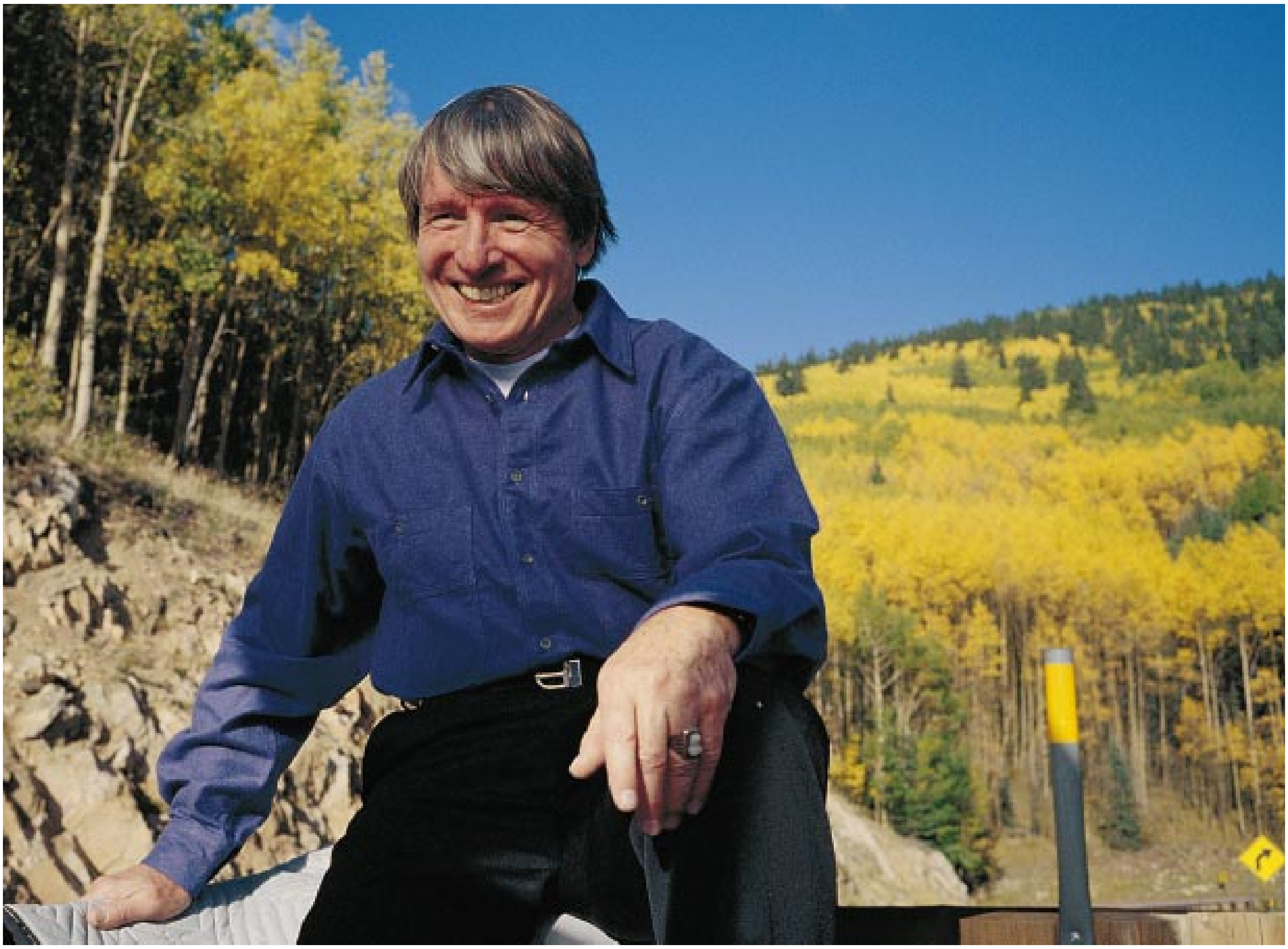
### Computing Complexity

In theory, the complexity of two systems could be compared by writing two computer programs that are the shortest capable of reproducing the original data. One data set, for example, might be a daily count of the number of leaves on an aspen during many seasons, and the other set might be the number of needles on a pine. The algorithmic complexity of the data is then related to the number of instructions in the computer program. The program with the fewest instructions would describe the less complex system.

The hitch is that algorithmic complexity is usually impossible to compute, notes James P. Crutchfield of the University of California at Berkeley. To measure algorithmic complexity accurately, a programmer would need to use a theoretical type of computer known as a universal Turing machine. Proposed by Alan M. Turing in the 1930s, this machine could perform any computation, given enough time. A universal Turing machine would have infinite memory, and a program could access or change any part of its memory. Such a machine would preserve every detail of the data on aspen leaves or pine needles. Indeed, the computer could simulate any physical process, although the programmer might literally have to wait from now to eternity to find out the results.

Other problems stand in the way of algorithmic complexity's becoming a practical tool for studying complex systems. For one thing, the technique is much more sensitive to randomness than order. A universal Turing machine would require many more instructions to reproduce a series of random events than it would take to simulate a periodic process. "If you are working with a com-

**NEW YORK CITY PARADOX: How is it possible for a group of independent companies to supply food for the millions of people in the city when most restaurants and supermarkets keep only a few days' worth of food in reserve? More generally, how is it possible for adaptive agents to find good solutions to complicated problems?**



**JOHN H. HOLLAND, a computer guru at the University of Michigan, created a digital ecosystem "to facilitate gedanken experiments on complex adaptive systems. It offers a way," he says, "to build intuitions about typical interactions and the emergence of structure in such systems." The ecosystem exhibits such lifelike properties as speciation and cooperation.**

pletely deterministic Turing machine," Crutchfield says, "you have to introduce many computational steps to generate random numbers." Moreover, there is no general method for finding the minimal program that simulates a given set of data. Even though a program reproduces the data, a researcher still cannot be sure it is the shortest one—or, if the data are sufficiently complex, whether the program is even close to the shortest one.

As a result, workers continue to search for new methods. In the late 1980s Crutchfield and his colleagues proposed a scheme based on the work of the linguist Noam Chomsky and others who developed a system for classifying different kinds of computers. The universal Turing machines fall in the most powerful "model" class. Other kinds of computers are categorized according to the size and structure of their memories as well as other parameters. For example, one class includes those computers whose memory is infinite but is organized as a stack.

To measure complexity with this hierarchy of model classes, a researcher would first choose a model class, then attempt to find the shortest program that reproduces the data. Given the constraints of the class, the investigator is forced to approximate the data, thereby preserving some essential features and discarding some random detail. If the program cannot reproduce the data, the exercise is repeated on a model class with more power. If the program does reproduce the data, the complexity is then related to the number of instructions and the number of constraints associated with the model class.

Crutchfield's ideas may also be useful to investigators searching for ways to simulate complex processes. If a programmer attempts to reproduce the data set from a complex process but chooses a model class that is too simple, the simulation will generate only a rough approximation of the original data. To improve the simulation, the programmer must select a new model class from an infinite number of choices. Crutchfield has found that by comparing a series of increasingly better approximations to the original data, scientists can find clues as to what choice of model class would be an improvement. It can help discover what kind of computer model is necessary to reproduce some process. "There is a way to look at the data," Crutchfield remarks, "and in the data themselves is information about how to innovate to a new model class."

Yet all the attempts to define, measure and model complexity do not address the fundamental question: Why is there so much complexity in the world? Why do some systems, both adaptive and nonadaptive, seem to evolve away from the extremes of complete order and complete randomness?

### Affinity for Criticality

Some researchers, most notably Per Bak, a senior researcher at Brookhaven National Laboratory, believe that one of those principles is a phenomenon known as self-organized criticality. During the late 1980s, Bak and his collaborators found a class of systems that appear to evolve toward complexity. The clearest example of these systems is a pile of sand [see "Self-Organized Criticality," by Per Bak and Kan Chen; SCIENTIFIC AMERICAN, January 1991].

If grains of sand are dropped onto the center of a pile, the system can exhibit three types of behavior: subcritical, critical or supercritical. If the sand pile is flat—the subcritical state—falling grains quickly come to rest and may trigger small avalanches of sand on the surface. As grains continue to fall, the pile grows and eventually achieves a critical state. In this state, adding grains of sand triggers avalanches of any size, from one grain to the entire surface of the pile. As more grains are added and more avalanches occur, the pile will continue to grow, even as its slope remains constant. If the sand pile is then shaped so that the slope is steeper than the critical value—the supercritical state—adding a few grains of sand will trigger an enormous avalanche. The pile, once again, returns to the critical state.

The sand pile increases in complexity, growing out of the subcritical state, or orderly phase, and collapsing when placed in the supercritical state, or random phase. As long as energy is added to the system—that is, grains of sand are dropped—the system will remain in the critical state, and its dynamics will never settle. "No matter what you do to the system—you can perturb as much as you like—it will always return to the critical state," Bak asserts.

In recent years scientists have gathered evidence that suggests self-organized criticality plays a role in many complex systems, from chemistry to geology and perhaps biology. "Since Darwin, we have come to think of organisms as tinkered-together contraptions and selection as the sole source of order," explains Stuart A. Kauffman, a theoretical biologist at the University of Pennsylvania. "Yet Darwin could not have begun to suspect the power of self-organization. We must seek our principles of adaptation in complex systems anew" [see "Antichaos and Adaptation," by Stuart A. Kauffman; SCIENTIFIC AMERICAN, August 1991].

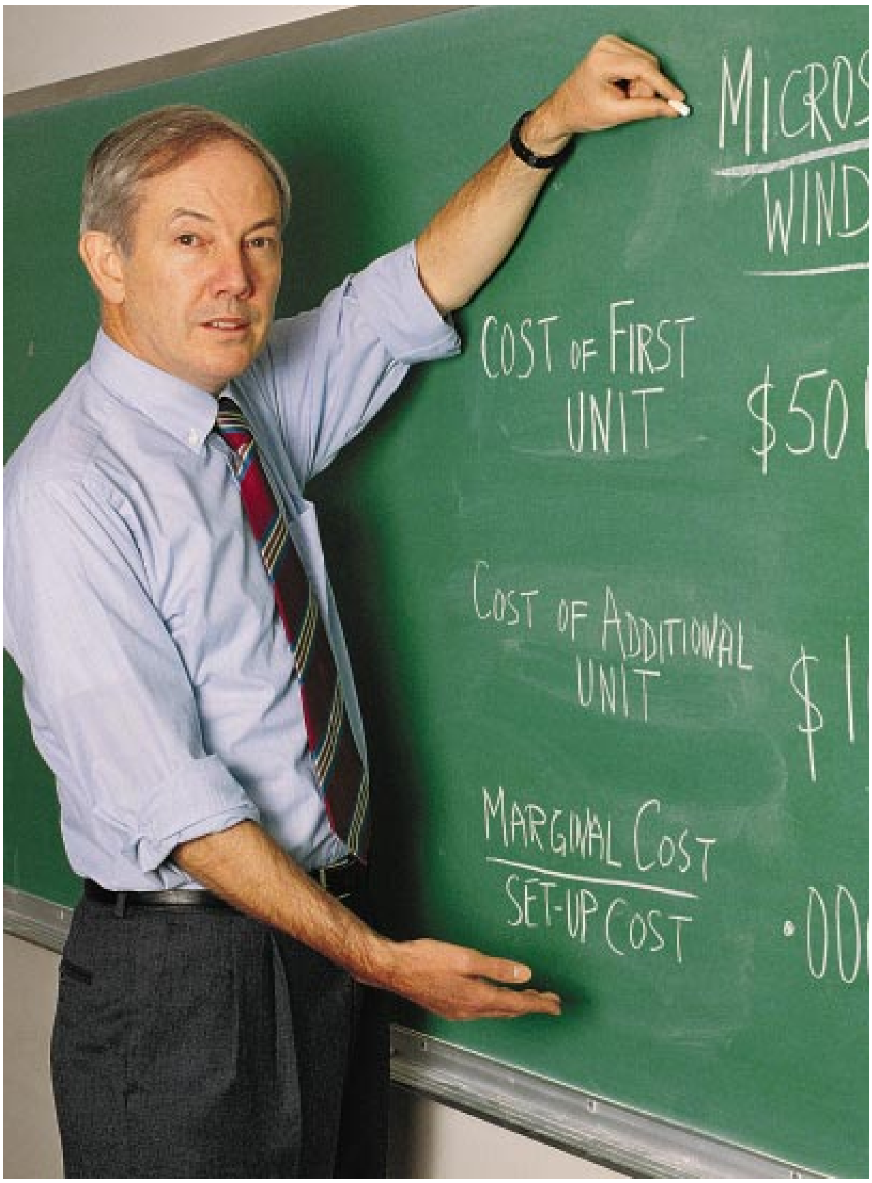
### The Edge of Chaos

The theory of self-organization resonated strongly with the work of Kauffman, Bernard Derrida of the Saclay Research Centre, Christopher G. Langton of Los Alamos National Laboratory and Norman Packard of the University of Illinois. In 1985 Derrida showed analytically that a system goes through a phase transition from order to randomness if the strength of the interaction between interconnected agents is gradually increased. A few years later Langton demonstrated that a system can perform the most sophisticated computations at the boundary between order and randomness. Meanwhile Packard began creating computer simulations showing that complex adaptive systems tend to evolve toward the boundary through the process of natural selection.

Working from these ideas, Kauffman and others have proposed that organisms change how strongly they interact with others in such a way that they reach the boundary between order and randomness, thereby maximizing the average fitness of the organisms. "This candidate principle," Kauffman warns, "is just that, a candidate principle."

To illustrate why the principle is plausible, Kauffman invokes a variety of biological systems—a pond harboring frogs and flys, for example. In the putative pond the frogs and flies have access to limited resources, and they have survival strategies that are encoded in their genetic material and nervous systems. Each frog or fly can change its strategy in an attempt to maximize its fitness, which can arguably be defined as the chance that it will survive long enough to reproduce.

At any instant in time, a frog or a fly will have a strategy that yields a particular fitness level. For each organism, the set of all possible strategies can be represented by a surface. Each point on




**W. BRIAN ARTHUR of Stanford University views the economy as a complex adaptive system. "A satisfactory theorizing in economics cannot proceed," he says, "without facing and resolving the key issues of learning, adaptation and cognition."**

**STUART A. KAUFFMAN is searching for the principles of adaptation in complex systems. Do all living systems—bacteria, humans, societies—evolve toward a boundary between order and randomness?**

the surface represents a different strategy; the height of each point indicates the fitness level of a particular strategy. As the organism changes its strategy, it moves to new points on the surface. Yet since an organism can only change its strategy to a certain degree, it does not make large jumps across the surface. It might be advantageous, for instance, for a frog to learn to fly, but it is not going to suddenly sprout wings.

The dynamics of the frog-fly system depends greatly on how strongly the two species interact. If the frogs do not find the flies tasty at all, then the fitness of the frog population will be independent of the fitness of the flies. The flies can attempt to maximize their fitness without worrying about the frogs, and vice versa. For each organism, there are a few strategies that would improve fitness, and there are many strategies that would make things worse. The fitness surface of each organism, therefore, has a few well-defined bumps (high fitness) and a large, flat valley (low fitness). The frog-fly system is in an ordered, static, subcritical state.

The dynamics of the system become much more complicated, however, if the fitness surfaces of the organisms are coupled, that is, if the frogs depend on the flies as a food source. As the flies develop new strategies, they will alter the resources available to the frogs, thereby influencing the fitness of the frogs. A change in strategy thus means a change in the contours of the surface. Consequently, the dynamics depend on just how strongly the flies and frogs interact.

If the coupling is very strong, on the one hand, any slight change in strategy is likely to change the character of the whole fitness surface. No frog or fly can adjust to the rapidly changing system. So the fitness surface of the average organism will have many short hills and will always be changing. The frog-fly system is in a random, dynamic, supercritical state.

If the coupling is neither too strong nor too weak, on the other hand, the dynamics become truly complex. In this case, the frogs may develop a successful strategy that works for a long time before the flies find a new strategy causing a major change in the fitness surface. The fitness surface of each organism will change over time, displaying many bumps that vary greatly in size. The frog-fly system is in a complex, dynamic, critical state.

"It turns out," Kauffman claims, "that in a wide variety of coupled systems the highest mean fitness is at the phase transition between order and chaos." By selecting an appropriate strategy, organisms tune their coupling to their environment to whatever value suits them best, Kauffman asserts. And if they adjust the coupling to their own advantage, he believes, they will reach the boundary between order and randomness—the regime of peak average fitness. "The bold hypothesis is that complex adaptive systems adapt to and on the edge of chaos," Kauffman declares. "It now begins to appear that systems in the complex regime can carry out and coordinate the most complex behavior, can adapt most readily and can build the most useful models of their environments."

Researchers at Santa Fe and elsewhere are just beginning to think about ways in which this framework and other new insights into complex adaptive systems can be proved. But Kauffman is confident that more robust models and further experiments will support a view of evolution that bridges living and nonliving systems. "Every attempt to find something that is being maximized in evolution has always met with failure," Kauffman observes. "Yet I have this feeling that there is something very general going on about how far from equilibrium systems have organized themselves. I don't know what that something is yet. But I can taste it."

FURTHER READING


EXPLORING COMPLEXITY: AN INTRODUCTION. Grégoire Nicolis and Ilya Prigogine. W. H. Freeman and Company, 1989.

COMPLEXITY: THE EMERGING SCIENCE AT THE EDGE OF ORDER AND CHAOS. M. Mitchell Waldrop. Simon & Schuster, 1992.

COMPLEXITY: SCIENCE ON THE EDGE OF CHAOS. Roger Lewin. Macmillan, 1992.

## Back to Roots

*Drug companies forage for new treatments*

***TRADITIONAL MEDICINALS bring new ideas for Shaman Pharmaceuticals, a San Carlos, Calif., firm that interviews healers in Latin America, Africa and Asia.***

From cocaine to quinine, about one quarter of U.S. prescription drugs contain at least one compound derived from plants. Yet in recent years plants lost their cachet at the big pharmaceutical firms as new ideas came from microbes or variants gleaned from huge data bases of synthetic chemicals. Now drug companies are emphasizing their roots. Botanists are once again avidly scouring the world's fields and forests to locate plant sources for new drugs. "Synthetics haven't proven to be the panacea," observes Mark J. Plotkin of Conservation International, a Washington, D.C.–based group that is working to stop destruction of tropical forests. "There's still no cure for AIDS or the common cold."

The lure is twofold: a "green" public relations boon and the prospect of discovering the next taxol, a treatment for ovarian cancer that was originally extracted from the bark of the relatively rare Pacific yew tree. Botanists estimate that 10 percent or less of the more than 250,000 flowering plant species have been surveyed for pharmacological activity. But the task of finding an active molecule in a haystack is being eased by devices that can quickly search through tens of thousands of samples.

One molecule in 10,000 may get to market through random screening. Automated methods that rely on enzymes and other chemical tags can enable a company to test as many as 150,000 samples annually, hundreds of times more than was possible by testing chemicals in laboratory animals. "Once you get to that level, you're bound to find something," says Alan L. Harvey, director of the Strathclyde Institute for Drug Research at the University of Strathclyde in Scotland.

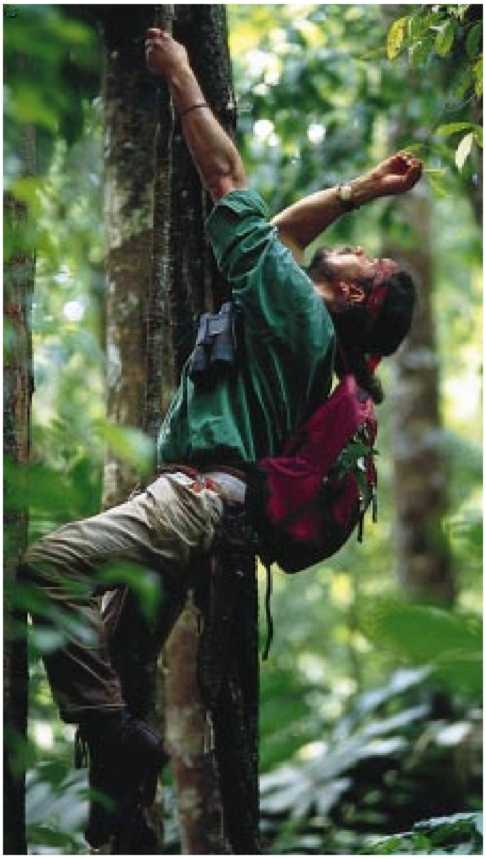
Among those looking is Monsanto, which has contracted with the Missouri Botanical Garden to supply the company with several thousand plants from both the U.S. and tropical countries. Merck & Co. has an arrangement with the New York Botanical Garden's Institute of Economic Botany for receiving plants from around the world. Biotics, a small start-up company in England based at the University of Sussex, embodies another approach—that of a "science broker." It has supplied Glaxo and SmithKline Beecham and others with samples and chemical extracts from plants gathered from the tropics.

Many of the plant-screening programs are so recent that the leads they have gathered remain shrouded in the secrecy that cloaks drug development programs. But the wraps have come off a few of the results. SmithKline, for example, is pushing a cancer treatment called topotecan through clinical trials. An enzymatic assay revealed this compound, which comes from a Chinese tree, *Camptotheca accuminata*. The active principal is derived from a chemical extracted from a tree that the National Cancer Institute (NCI) tagged for potential anticancer activity about 20 years ago.

The NCI has been a locus for searching for new drugs from plants. It mounted what may be the world's most extensive plant-testing program in 1986, replacing a 20-year screening project launched in 1960. Although the earlier effort (which relied on mice as screening agents) produced taxol, it was scrapped in 1980 because it had not turned up enough leads. The effectiveness of taxol against ovarian cancer was not detected until the late 1980s.

So far the NCI's revived effort has collected 23,000 samples from 7,000 species in tropical areas and has identified three compounds that seem to work against AIDS in a laboratory dish. All three compounds have entered preclinical testing for toxicity.

One "hit" is an extract from a creeping vine that inhabits the canopies of rain forests in Cameroon. This member of the *Ancistrocladus* genus inhibits the replication of the AIDS virus. The plant was found in an attempt to search across the broadest possible taxonomic spectrum. It had no known medical use and still lacks a species name. "There are very few people on the planet who have ever seen this plant," says James S. Miller, a botanist from the Missouri Botanical Garden. More recently a University of Illinois team working for the




***COSTA RICAN FLORA attracts botanists who conduct basic research and who hope to find plants that may yield new drugs.***

NCI discovered a Malaysian tree, *Calophyllum lanigerum,* with potent anti-AIDS properties. Calanolide A, a compound from the tree, seems to work against an AZT-resistant form of the AIDS virus.

Another potential anti-AIDS drug was discovered in a plant in Samoa. Paul A. Cox, a Brigham Young University professor of botany, collected *Homalanthus nutans* on the recommendation of women healers there. The NCI extracted prostratin from the plant, a chemical that seems to protect immune cells from being destroyed by the AIDS virus. In Samoa, *H. nutans* was used as a treatment for yellow fever and other viral illnesses. "You have accumulation of knowledge in the culture that resembles human bioassay data," Cox says.

Indeed, several entrepreneurial companies are turning to traditional medicine to narrow their search for new drugs. Their idea, like Cox's, is to test materials that have shown pharmacological activity for hundreds of years and that are probably relatively nontoxic. Chemex Pharmaceuticals in Fort Lee, N.J., recently won approval from the Food and Drug Administration to market a treatment for precancerous skin lesions. Called Actinex, the drug is derived from the creosote bush, which has a long history of medicinal use. Two companies founded within the past five years—Xenova in Slough, England, and Pharmagenesis in Palo Alto, Calif.—are combing through the traditional Chinese pharmacopoeia for promising leads.

Another company, Shaman Pharmaceuticals, was founded by a former venture capitalist, Lisa A. Conte, who wanted to combine drug development with an effort to preserve rain-forest flora. Conte's firm, based in San Carlos, Calif., sends teams of botanists and physicians to Latin America, Africa and Asia to search for plants that work against viruses, fungi or diabetes or that may become sedatives or analgesics.

Plants sent back to Shaman for testing must have been used in three geographically distinct areas. Conte claims that of the 200 that have passed this preliminary probe, more than half show activity against a targeted malady, as compared with less than 1 percent for mass screening.

In only 16 months Shaman has moved an antiviral drug for childhood flu caused by the respiratory syncytial virus from rural healers and urban botanicas in South and Central America into FDA testing in humans. By showing medical photographs to healers in Peru and Ecuador, Shaman's botanists also turned up an antiviral agent that works against drug-resistant herpes infections. "What Shaman is doing is using thousands of years of human clinical trials," says William L. Current, a senior research scientist with Eli Lilly's infectious disease group. Last October Lilly made a $4-million equity investment in Shaman and also struck an agreement to collaborate with the firm for four years in developing drugs that work against fungal diseases.

Gone are the days when a company could create drugs from plants that originate in developing countries without negotiating to pay royalties, as Lilly did during the 1960s with the rosy periwinkle. Extracts from the periwinkle produce vincristine and vinblastine, drugs used primarily against childhood leukemia and Hodgkin's disease, respectively. Shaman has pledged to pass up endangered plants and is committed to furnishing royalties from drug revenues to both the government and the native communities from which a plant is harvested, a policy that Lilly supports. Shaman will make these payments through the Healing Forest Conservancy, a foundation the company set up to promote rain-forest conservation.

More unusual is Merck's agreement to pay $1 million over a two-year period directly to Costa Rica's National Institute of Biodiversity (INBio) for collecting plants, insects and microbes. It also pledged to pay royalties for any leads that turn into pharmaceuticals. A portion of both the initial expenditure and any eventual royalties will help preserve wild areas in Costa Rica. Merck's collaboration with a nonprofit institute in a country whose gross national product is less than the drugmaker's annual revenues has generated intense interest in other nations between the tropics of Cancer and Capricorn.

INBio, which signed the accord with Merck in 1991, is now helping Indonesia to set up a similar institute devoted to biodiversity. Some public-interest groups have criticized the Merck-INBio agreement because of what they perceive to be a lack of accountability. "A nongovernmental organization doesn't have the right to sell genetic material to the rest of the world," says Jason W. Clay of Cultural Survival, an organization in Cambridge, Mass., that tries to find uses for indigenous materials, such as nuts for Ben & Jerry's Rainforest Crunch ice cream. INBio emphasizes that it was established three years ago by a governmental decree. A Merck official pointed out that others are not precluded from looking for natural samples in Costa Rica.

A five-year program set up last June by the National Institutes of Health, the National Science Foundation and the U.S. Agency for International Development aims to promote biodiversity and establish new economic ties with the developing world. The program, with $1.5 million in annual funding, will set up consortia of universities, nonprofit institutes and industry to catalogue plants and other organisms, with the goal of isolating compounds that have pharmaceutical value.

Some environmentalists are rooting for the drug companies. "It is my hope and my expectation that people are going to find something pretty soon that is marketable," says Michael J. Balick, director of the Institute of Economic Botany at the New York Botanical Garden. This, he thinks, may lead to a "greater appreciation of the true value of these resources." Even if that happens, time may be running out for the rain forest. According to Cornell University professor of biology Thomas Eisner, "the looking for new chemicals is going slower than the rate at which species are becoming extinct." —*Gary Stix*

# National Conundrums

*Finding new work for the national weapons labs*

The new administration has vowed to divert about $7 billion a year from armaments research to civilian technologies. To do so, it must confront an urgent issue: What to do with the national laboratories?

Now that the cold war has thawed, America's premier weapons laboratories—namely, Los Alamos, Lawrence Livermore and Sandia—seem to have lost their principal mission. Those three laboratories, which are part of the Department of Energy (DOE), command a collective annual budget in excess of $3.5 billion and employ more than 23,000 people. "We've got to change," declares Siegfried S. Hecker, director of Los Alamos. "If we don't have a legitimate mission in the world, we should die."

The plight of the weapons laboratories is spelled out most clearly in dollars and political pledges. Almost half of the labs' budgets are devoted to nuclear weapons development and maintenance. In many cases, the organizations have parallel programs. Yet international accords aim to quash the evolution of new nuclear arms. Last year Congress called for a nine-month moratorium on nuclear testing, which began in October. After that ban expires, the U.S. will be permitted 15 tests (to evaluate "safety" issues); an unlimited test ban is slated to go into effect as of October 1996.

Few believe the laboratories will wrap up their work on nuclear weapons altogether. National security will always remain a primary objective, confirms Sidney D. Drell, deputy director of the Stanford Linear Accelerator Center and recently appointed chairman of a council charged with articulating a strategic role for the science and technology programs of Livermore and Los Alamos. But rather than emphasizing design and testing, the research groups will step up work on technologies for monitoring nuclear proliferation, ensuring the safety of stockpiled weapons and disposing of obsolete arms. "There are still many thousands of nuclear weapons around," Hecker adds. "You don't just throw them away."

At the same time, as the battle cry of "international competitiveness" echoes throughout U.S. industry, Drell and others see the labs devoting more resources to civilian technologies. A recent report from the private-sector Council on Competitiveness recommended that between 10 and 20 percent of the budgets of the DOE (and other) labs should be devoted to technology transfer. "The federal labs have a new customer—U.S. industry—and need to develop customer-driven technology transfer programs to service its needs," the council stated.

On the face of it, the laboratories seem well equipped to lend a hand on the competitiveness front. Traditionally, the U.S. nuclear weapons labs have employed some of America's sharpest scientific elite. These researchers have done groundbreaking work across a range of fields, including semiconductors and lasers, advanced materials, high-speed computation and networking. "You can tick off five or six areas where the labs are superb," Hecker asserts, "and then you just have to find the right way of linking them with industry."

Finding those kinds of links has proved to be a headache. Beginning with the Stevenson-Wydler Technology Innovation Act of 1980, Congress tried to unlock the doors of federal labs for industry. Nine years and several acts later the government gave the DOE labs another push toward the private sector through the National Competitiveness Technology Transfer Act of 1989.

Some progress has been made at the federal laboratories outside the DOE. The National Institute of Standards and Technology, for instance, is barely one third the size of any one of the weapons groups but has more than 240 cooperative research and development agreements (CRADAs) with industry. Even divisions of the Department of Defense, such as the Army Research Laboratory in Fort Monmouth, N.J., have won praise for securing deals with industry. But as of November, the three DOE labs had signed only 118 CRADAs. "The DOE and the weapons labs have been very slow on CRADAs," observes Julie Fox Gorte, a senior associate with the Office of Technology Assessment.

Instead the three weapons groups—along with the DOE's other six "multipurpose" labs—have become entangled in a mesh of bureaucratic trip wires. "Do you want to hear about the legal or the cultural 'demotivators' first?" asks John R. DeMember, the federal laboratory liaison for Digital Equipment Corporation, who participated in creating a model CRADA between an association of computer makers and the DOE.

DeMember insists he is optimistic about the possibilities of working with the laboratories. But he is blunt about the stumbling blocks. The government researchers have seldom had to meet the strict time and cost budgets faced by their private-sector colleagues, DeMember notes. Lab scientists have traditionally shied away from product-oriented work even as company researchers have nurtured a "not invented here" bias. As a result, "the process of getting the lab researchers and my researchers together has been painful," DeMember says.

Legal issues have been even more vexing. Rather than allowing laboratory directors to sign off on CRADAs, all such contracts have required the DOE's approval. And there is a collection of regulations that can be CRADA "showstoppers," DeMember notes. For instance, the DOE has been reluctant to exempt from the Freedom of Information Act any information the laboratories receive from industrial partners—meaning that the material is publicly available even if considered proprietary by the company. "Because the DOE labs are lawyer-driven,

MARK POULSEN *Sandia National Laboratories*

***CAMPAIGN PLEDGE: Bill Clinton vowed to help the weapons labs turn to civilian technologies such as the environmentally benign process for making printed-circuit boards showcased by Sandia National Laboratory head Al Narath.***

they're more focused on what could go wrong than on taking chances," he adds.

A December 1990 report from the General Accounting Office reinforced those tendencies. The audit criticized the laboratories for excessive use of discretionary funding. "That report was really not fair!" Hecker declares. Los Alamos endured the brunt of the attacks. Hecker maintains that both the former Atomic Energy Commission and the DOE had long encouraged Los Alamos to fund a wide range of scientific projects so that the weapons labs could attract top scientific talent. "We have no excuses, no apologies. The work funded was first-rate stuff," Hecker insists.

Yet the media attention spurred the DOE to increase the rules and regulations governing the laboratories. In addition, the contracts between the DOE and independent "management and operating" organizations that run the facilities have become far more detailed.

For instance, the University of California has overseen Livermore and Los Alamos since they were created. When the contracts came up for renewal in 1992, the documentation ballooned from a gentleman's agreement of a few tens of pages to an extensively negotiated contract of several hundred. "In the past, the standard the labs were measured against was technical excellence," says Robert W. Kuckuck, who has recently been appointed special assistant for laboratory administration in the Office of the President at the University of California. "Today it's very much about the details of how you get the job done."

Experts nonetheless point to signs that the laboratories will soon find a smoother path to industrial partnerships. Kuckuck, who worked as a physicist at Livermore for 29 years, is determined to use his new office to comb out the snarl of requirements facing the labs. DeMember feels some of the most recent CRADAs signed by the DOE resolve many of the legal problems and will serve as pivotal models for public- and private-sector cooperation. "I've seen an enormous change in how both industry and the labs feel about cooperation," DeMember adds. As a result, "I'm very optimistic about future joint work—but it's still a hard process."

Many are looking to the Clinton administration to foster more collaboration with industry. "We are living in a world where brain power will determine the lives of all our people," Clinton told weapons workers during a campaign visit to Sandia in September. "I want to make sure that each of you has a chance to contribute your full potential to the welfare and to the future of your country." *—Elizabeth Corcoran*

## Soft Lego

*How software designers hope to make programs reusable*

In the laboratory where about 140 Microsoft engineers are laboring on a project code-named Cairo, their mascot—a plywood cutout of a camel—leans patiently against the corridor wall. Strapped to its side is a package of Tylenol. "The camel was sick for a while," quips James Allchin, Microsoft's vice president of advanced systems and manager of Cairo.

Along with a pack of other companies, Microsoft is racing to create a very different kind of software, an operating system based on "object-oriented programming." The camel is ready testimony to the headaches and long hours of the effort. But the payoff promises to be enormous. "We have to do something to get to where every person can construct his or her own program," Allchin says. "For people to become more productive in the information age, there can be no canned programs."

At present, writing computer programs is mostly a handicraft in which even mundane sections demand time and painstaking work. Software manufacturers such as Microsoft legitimately worry about how to boost their own productivity. But making software is even more vexing for those who simply want to use it. At the top of that list are hardware manufacturers, who must increasingly lace their products with sophisticated programs and users, who find themselves tailoring code to accomplish specific tasks.

The solution is obvious: design software so that sections of code can be reused. Doing that, however, is fantastically hard. Reuse has been on research agendas since the 1960s, but only halting progress has been made. Now "the time is right, the basic technology is in place," asserts Daniel Sabbah, who manages software technology at the IBM Thomas J. Watson Research Center. "For the first time, we can look ahead and see a fundamental paradigm shift coming." Sheer momentum may help. "It's hard to find someone not doing reuse," says Norman M. Delisle, director of the software technology research laboratory at Tektronix in Beaverton, Ore. "It's part of survival in the '90s."

Object-oriented programming, a concept currently as trendy as artificial intelligence was in the 1980s, is a key tool being used to construct the brave new world of reusable software. Objects are chunks of data and related procedures and methods (which describe some way of manipulating the data). Like Lego blocks, such encapsulated components can be plugged together or pulled apart to make different systems, provided they all use the same "interfaces," or plugs. Objects boast a full complement of other features—some fit into "classes" with similar objects; others "inherit" characteristics. Engineers hotly debate the merits of all such traits.

In an operating system such as Cairo, objects promise to give users a connect-the-dots ability to get a job done. For instance, a person might see one "button" on a screen that represents a document—a sales report—and another button that represents a mailing list, Allchin explains. To send out the report, an individual need only draw a line between the two buttons with a mouse. But the user might also add a caveat, such as "if the sales figures are falling, send the report only to senior executives." The objects will check the numbers and route the document accordingly, Allchin says. What is more, the system will make use of objects wherever they are stored—on the user's desktop system or elsewhere in a network of computers.

Cairo is not scheduled to make a public debut until 1994. And engineers at Microsoft and another dozen or so companies working on competing projects continue to wrestle with weighty problems. The "semantics," or the precise descriptions of the nature of a specific object, are still as undeveloped as the speech of a baby. As a result, the objects designed by one company, say, Microsoft, may not be able to link together with objects from others, such as Borland International, NeXT Computer or Taligent (an IBM-Apple joint venture).

There are nonetheless a handful of tools in addition to objects that researchers hope will smooth the path to reusable software. IBM's Sabbah points to the usefulness of compilers (which translate software from one language to another) and to higher-level, more abstract, languages that let users express their tasks in the language of their business rather than in computerese.

In one project, IBM researchers wrote a "constraint language," which could optimize any set of equations to given parameters. With that language, the workers then devised an experimental stock-option trading program that recommended buy and sell options, given constraints such as a balanced portfolio of risks. Traders would need no special programming skills to run the system; it could, moreover, be applied with equal ease to optimizing the constraints that govern electrical circuit design.

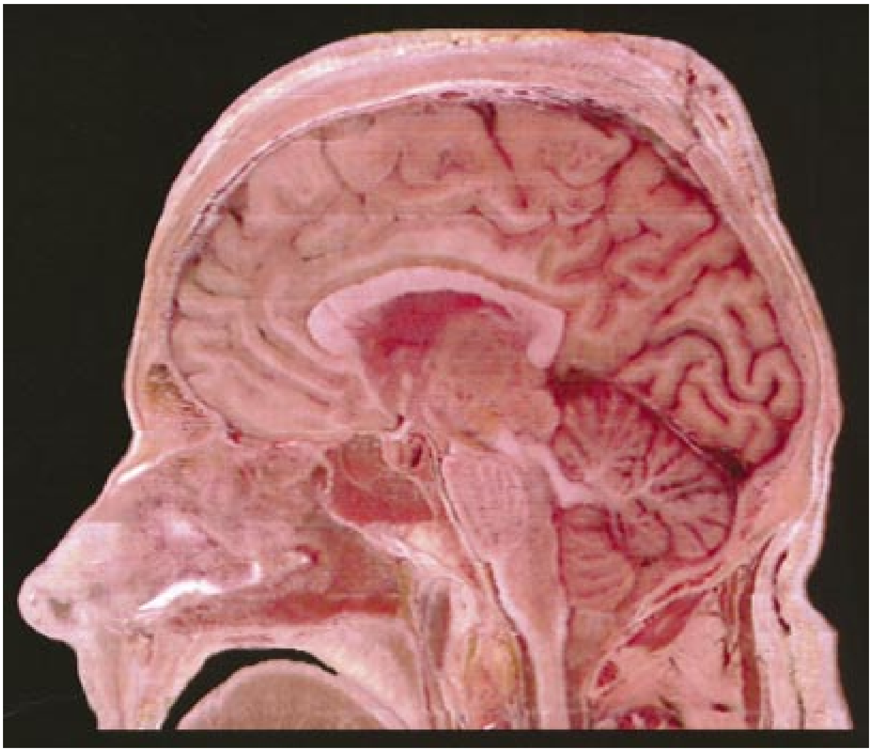
Also promising are "frameworks," or what David Garlan of Carnegie Mellon University calls "software architectures." Rather than plugging components together, designers first try to uncover—and then reuse—the skeleton shared by a collection of related products. Spreadsheets are a textbook example, Garlan points out. Even as users customize spreadsheet programs to suit their needs, they are "reusing" a huge portion of the prewritten code. Similarly, product manufacturers will find enormous opportunities to reuse code if they first identify a common software architecture for a family of products, Garlan suggests.

Such frameworks inevitably complement libraries of software components. At Hewlett Packard, for instance, Martin Griss has been forming a "flexible software factory" that aims to take advantage of both frameworks and component libraries. Even so, the biggest impediments to reusing software are more sociological than technical, he notes. "It's hard for a lab to change the balance of up-front design and downstream assembly," Griss says. Hewlett Packard is consequently trying several pilot projects that rely on multidisciplinary teams of engineers, designers, managers and even anthropologists.

More far-reaching ideas for reuse loom on the horizon. In many ways, all the techniques for reuse turn on inventing ways to talk about the work at hand. Objects describe a piece of code and procedures; frameworks describe common themes. At Microsoft, chief architect Charles Simonyi is thinking about software design from an even loftier plane. He wants to create a notation that engineers can add to code to describe what they are trying to do.

In this way, Simonyi hopes to capture a larger portion of the most creative efforts of designers—so that others can "reuse" the cleverness of the programmers, not just the code. To use a cooking metaphor, Simonyi's approach represents the difference between providing a chef with a packet of powdered sauce that needs only milk or a video of Jacques Pepin preparing the sauce. "If software artifacts are frozen human contributions, we want to look at the freezing process," he says. "It's the human thinking we want to reuse."

Simonyi expects he must put in another few years of effort before he can prove his theories of "intentional programming" can be applied in practice. But like earlier proposals for reusing software, Allchin of Microsoft observes, Simonyi's ideas are high risk—yet if successful, they also promise more than adequate returns. *—Elizabeth Corcoran*



***HEAD RECONSTRUCTION was created by combining 352 digitized photographs of tissue sections. The University of Colorado will now do the same for a whole body.***

## Habeas Corpus

*Seeking subjects to be a digital Adam and Eve*

A good man is hard to find. So, too, a good woman. Unlike the Marine Corps, the National Library of Medicine (NLM) does not need a few good specimens. Just one of each sex will do. The ideal candidates should be between the ages of 20 and 60 years, of medium height, not too thin, not too plump. Race or ethnic background is not an issue. But a prospect must meet more problematic criteria: he or she must have expired and been put into a cooler within the last 12 hours in preparation for being cut up by a team of medical specialists into more than 1,000 thin slices.

The NLM has embarked on an 18-month search for contestants to become the ideal male and female cadavers. Those chosen may become known as Adam and Eve. This search may mark the beginning of an era in which computers may rival cadavers as primary sources of information on anatomy.

By the end of 1993 the NLM wants to enter into a computer every millimeter of a male and a female cadaver via computed tomography (CT), magnetic resonance imaging (MRI) and standard photography. The result of this effort—the Visible Human Project—will be the most comprehensive digital record of an entire human body. "Before this, one university was doing the right knee while another was doing the left ear," says NLM director Donald A. B. Lindberg.

The main use for this computer-age *Gray's Anatomy* will, of course, be the study of anatomy. For the first time, a medical student will be able to "reverse dissect" a cadaver, reassembling it after having cut it apart on the screen. A researcher might also inject a mathematical model of abnormal cell growth into a computerized liver to grow a tumor. The data base may be used in designing prostheses or in assisting a surgeon to plan an operation better. A patient's CT image can be compared with Adam's and Eve's standard images for a "normal" human.

Before proceeding, NLM officials and outside experts debated for years the exact definition of normal but failed to achieve any consensus. Their loose classification now encompasses a male and a female of any race between the ages of 20 and 60 years, less than six feet in height and of average weight. "We decided that nothing can be considered average or normal, but we had to start somewhere," says Michael Ackerman, the NLM's associate director for specialized information services, which oversees the project. But even within this open definition, it has been difficult to find a body that has remained unscathed through the travails of daily living. "People die for a reason," Ackerman says.

The University of Colorado Health Sciences Center in Denver received the more than $700,000 contract (a word that is used advisedly among the project's staff) to find several male and female bodies. An NLM committee will choose one Adam and one Eve from the candidates the center screens.

Cutting was to have begun last July. But so far only five cadavers have been found, three males and two females. All but two of them still have shortcomings. "We've got two good males, but we're still not really happy with the females we have now," says Victor M. Spitzer, a professor of cellular and structural biology who is supervising the project at the university. The digital preservation of one of the two frozen males could begin within a few months, after the NLM review panel makes a choice. Spitzer is still searching in other states and explaining the purpose of the project at technical conferences.

A good candidate may be one who has died of a drug overdose, carbon monoxide poisoning or some other nondisfiguring cause of mortality and, of course, has agreed to donate his or her body to science. Colorado's Anatomical Visualization Laboratory will employ a milling machine that operates as if it were an automated carpenter's plane, shaving off a millimeter-thick slice from a section of the body. As each slice is removed, the newly exposed surface is photographed. The digitized photographs will be stored in a computer with MRI and CT images that are made before the cutting process is begun. Afterward, the real work will begin. Another contractor will finish the job by identifying and indexing each of the digital slices, a project that will take years more and will cost an estimated $10 million.

The final data base may consist of 50 gigabytes of digital information, enough, if left uncompressed, to fill more than 80 data-storing compact discs. Instead of distributing data on discs, the NLM may have the option to transfer the information over a proposed government-financed network that could accommodate transmission at the lightning speed of a gigabit per second, a project proposed by Vice President–elect Al Gore.

At some future date, Adam and Eve may have progeny, perhaps parents as well. The NLM would like to add imaged cadavers of children, the elderly and the sick to its storehouse. Meanwhile the search continues for a select few men and women who will be the recipients of 50 gigabytes of fame. —*Gary Stix*

## Geometry Acquisition

*Computed tomography is a boon to reverse engineering*

Working backward from a manufactured part to a blueprint description used to be a laborious process. Engineers would slice parts into pieces and take hundreds of measurements. Even so, they would often only come close to the precise shape and dimensions of the original. Now an industrial-strength version of the computed tomography (CT) machine that ferrets out brain tumors is making the job almost as easy as taking a medical x-ray.

Data on the interior and external dimensions of a manufactured part serve as a guide for constructing two- or three-dimensional computerized blueprints. There is no damage to the part nor distortions in spatial perspective that occur when a part is cut into pieces. The CT scan provides information about the contours of the part that may not have been on the manufacturer's original design drawing or for which there may have only been a wood or plaster model. The image also allows a drawing to be made of a part that was originally produced by a supplier—or a competitor.

At $1 million and higher, the price of many CT machines is far too steep to buy one just to spy on the competition. Indeed, the three main U.S.-based makers of industrial CT scanners sell fewer than 10 machines a year. The main reason for installing a machine is to measure dimensions of a finished part, to detect internal flaws or to construct computer-aided design (CAD) drawings for parts whose dimensional data are not yet in the computer.

General Motors' Power Train Division facility in Defiance, Ohio, for example, has been using a CT machine from Scientific Measurement Systems (SMS), an Austin, Tex., manufacturer, for several years to create design drawings of the metal castings for engine blocks, connecting rods and other parts.

Boeing turned to reverse engineering—or "geometry acquisition," as it is sometimes euphemistically called—to re-create a tail wheel for a 1940s Boeing 307 Stratoliner, which was being refurbished for the Smithsonian Institution. Engineers took several two-dimensional CT slices of the wheel. "It's as if you cut the part open and had it lying in front of you," says Richard H. Bossi, a research engineer for Boeing Defense & Space Group.

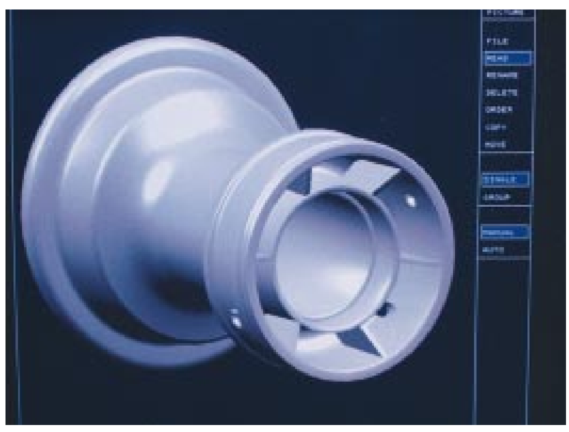
Next a design engineer extracted the relevant dimensional points from the cross section and used a CAD system to form a three-dimensional solid image. This drawing was used to produce a half-size replica of the wheel using a process called stereolithography. An ultraviolet laser, guided by the coordinates in the drawing, solidified layers of plastic to make the model.

Yet peeking at the competition still holds an undeniable allure. "GM could look at a Ford product and know all about it, or vice versa," asserts George A. Edwards, vice president of administration for SMS. Indeed, SMS has scanned a Toyota engine block for Ford Motor, though not for competitive reasons.

John C. Cooper, a senior engineer in Ford Manufacturing Development, says SMS imaged Toyota's aluminum cylinder heads because they resembled a new Ford design. The scan, Cooper says, enabled Ford engineers to determine the capabilities of the CT technology. "This was a typical example of the kind of part you might want to image," Cooper observes. "CT is the only thing I know of that can define a part's internal features without cutting it apart."

Cooper says the company might consider the purchase of a CT machine if the outlook for the automobile industry improves. "We'd be looking at our own parts or our competitors'," he says.

Still, CT is not yet a standard tool in the annual ritual of checking out the competition. But as these techniques are further automated and the price of the technology drops, reverse engineering is likely to find a variety of applications. And the temptation to acquire a little comparative geometry will certainly be one of them. —*Gary Stix*



***TAIL WHEEL for a vintage Boeing 307 Stratoliner was re-created using computed tomography to help determine the dimensions for a design drawing.***

## The Return on Infrastructure

The legions of any freshman administration march into Washington flush with armfuls of economic solutions. The previous generation had supply-side economics; Bill Clinton's team has infrastructure.

Infrastructure offers an intriguing policy platform. During the past decade, newspaper headlines have decried the collapse of bridges, the growth of potholes and the deterioration of public schools. Robert B. Reich, a muse for President-elect Clinton, lofted high the banner of infrastructure in his 1991 book, *The Work of Nations*. Reich writes, "Herein lies the new logic of economic nationalism: The skills of a nation's work force and the quality of its infrastructure are what makes it unique, and uniquely attractive, in the world economy."

No economist would deny that infrastructure investments, like family values, are important. But beyond that safe assertion rages one of the fiercest debates among economists today. At issue: Do investments in public infrastructure earn more, less or about the same return as do private investments? The answer suggests clearly different policies. Although economists argue that they lack enough evidence to clinch a consensus, politicians have been less shy. As a result, infrastructure has become "one of the more emotional issues economists get involved in," says Randall W. Eberts, an economist at the Federal Reserve Bank of Cleveland.

The Clinton clique has much in mind when they talk about infrastructure: not just improved roads but telecommunications and high-speed data networks and better schools. Yet as recently as a few years ago infrastructure was seldom mentioned in economic circles.

Economists might crunch through cost-and-benefit analyses of a construction project. But no one tried calculating the value of such investments for the economy at large. Even the classic "production function" describes growth as a function of changes in the amount of capital (contributed by industry) and labor used, with an added kick from new technology. Better roads and sewers were believed to contribute broadly to the "quality of life."

In 1989 David Alan Aschauer, who was then with the Federal Reserve Bank of Chicago, opened up a different perspective by arguing that infrastructure strongly bolstered private industry. He observed that U.S. productivity had declined in tandem with falling investments in what he called "core infrastructure," namely, transportation (roads, mass transit and airports), water and sewer systems and other publicly owned facilities. His studies suggested as much as 40 percent of the slowdown in productivity could be explained by declining public capital investments. Aschauer further estimated that every dollar invested in public infrastructure yielded four dollars in return.

The work stunned economists. "Aschauer did a great service" by identifying a role for infrastructure, declares Alicia H. Munnell, a senior vice president at the Federal Reserve Bank of Boston. "We should be embarrassed that we overlooked it." That said, many economists quarrel with the specific returns. "There's been a lot of controversy about those results," Aschauer concedes. "People say they are implausible."

*All agree infrastructure is important. But how much do government investments really pay off?*

Munnell has done several studies that support the proposition that infrastructure spending yields some kind of positive return. "But infrastructure isn't built to enhance the private-sector output," she notes. Logically, since the private sector purposefully invests to boost its output, its capital spending should earn at least as high a payoff as public infrastructure, she suggests.

Aschauer, now a professor at Bates College, nonetheless believes the economy reaps significant "network" effects from investing in infrastructure. A better bridge, for example, improves the business of private producers as diverse as computer makers, farmers and trucking fleets. By complementing private investment, such public infrastructure raises profits and stimulates private investment, Aschauer asserts.

Others respond that no economic theory explains why infrastructure alone should stimulate growth. "It's the *Field of Dreams* syndrome," Eberts says. "If we build it, consumers will come."

Statistical evidence is also muddled. In studies of regions and cities, Munnell has found positive correlations between public infrastructure and growth, but she hesitates to say such public investments contribute more than private ones might. Eberts believes he has found indications that during most of the 1900s, the mature cities of the North benefited more from public investments than did the newer, faster-growing urban areas of the South. "But the problem is that's all historical evidence," he complains. Showing that a new highway boosted growth in the past does not mean that building another road will have the same impact tomorrow.

Those who argue that more public construction will promote growth interpret the increased congestion on the roads and in the sky as a sign that demand has outstripped the supply of those facilities. Transportation economists, however, counter that such jams show that the fees charged for using these resources are just too low.

"I think infrastructure has what some would call nonlinear effects," suggests Charles R. Hulten, an economist at the University of Maryland. Sometimes new infrastructure is critical for growth, such as before the development of the highway system; other times, adding another road might just fill space.

Aschauer and others applaud Clinton's emphasis on information technologies in the mix of infrastructure components needed for the 21st century. "Dollar for dollar, I think the productivity benefits we can get from telecommunications are significant—higher than what we would receive from additional highways," he says. Even so, data that might illustrate the value of such high-tech infrastructure—funded by either the private or public sector—are as scarce as blue solid-state lasers.

Not that Aschauer plans to take much of a hand in building new policies in Washington. Although he helped spark the Democrats' infatuation with infrastructure, Aschauer is a greater fan of the policies of Alexander Hamilton than of Bill Clinton. "I'm pro-infrastructure," he says, "but I'm not enamored with more spending." *—Elizabeth Corcoran*

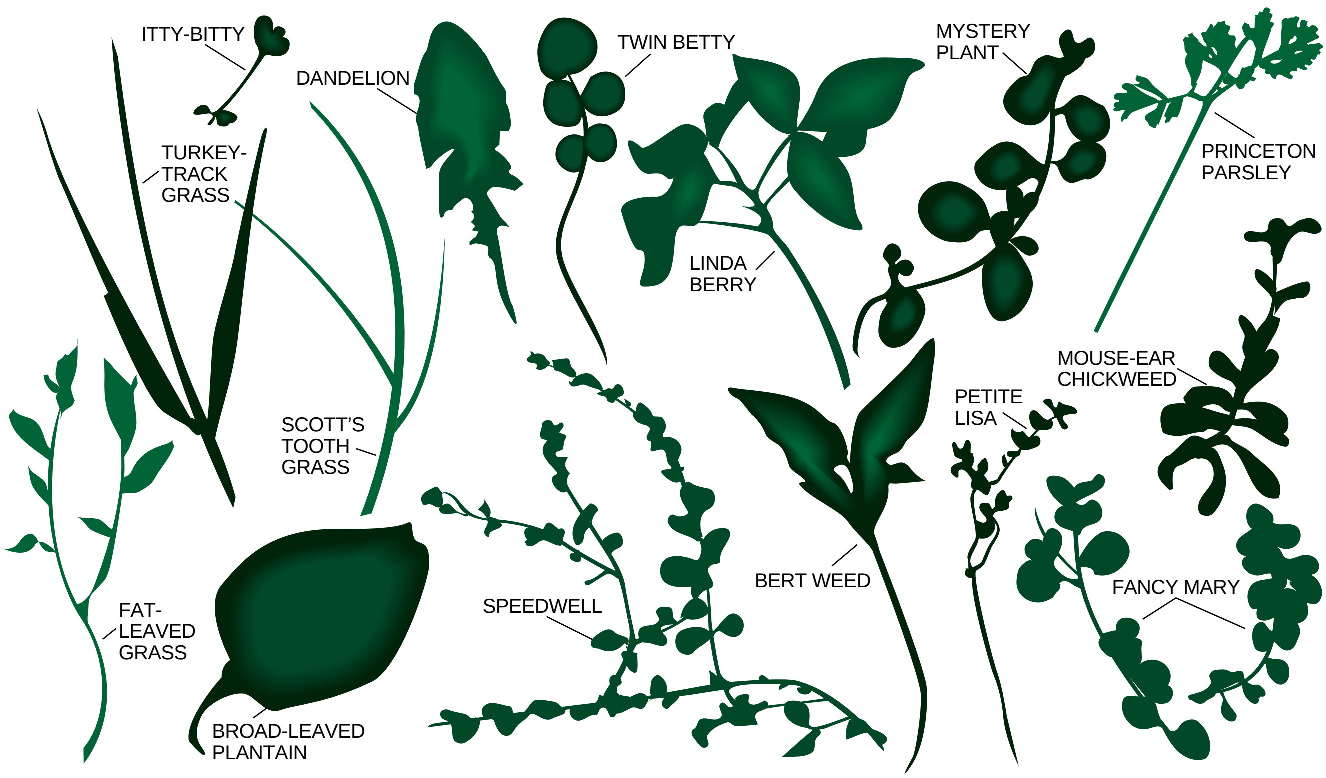
# Biodiversity in the Backyard

Systematic inventories of plots of woodlands and fields can be of practical use in planning how best to conserve wildlife in a given patch of land. These surveys show vividly how the number of species encountered in a plot varies with the amount of land inspected. They also help to provide a quantitative way to see how human activity affects local biological diversity. With such observations, conservationists, ecological planners and policymakers can estimate the smallest amount of land needed to preserve a percentage of the natural flora and fauna. Particularly useful in this regard is the relation between the diversity of woodland creatures and plants and the size of forest "islands" in an urban or suburban "sea." Such relations are technically referred to as species-area curves.

HENRY S. HORN constructs satirical artwork and sings blues, madrigals and choral music. He is also professor of ecology and evolutionary biology and director of the program in environmental studies at Princeton University. He thanks Bristol-Myers Squibb for sponsorship, Elizabeth Horn for help in designing and directing this project, the elementary school teachers for their participation and, last but not least, Robert May, William Bonini, Sheldon Judson, Patricia Matuszewski and Amy Wolman for constructive commentary and moral support.

Counting plant species within a lawn is an instructive analogue of such quantitative methods. (Tabulating things that crawl or fly is difficult and tends to lie beyond the amateur level.) I initially designed this project as an exercise for a summer course in mathematical geology and field biology for grade school teachers. The teachers have adapted it to be an exercise in exploration and classification for children in the lower grades. But even our preliminary analyses have been so informative that I plan to use the exercise in introductory data analysis and extrapolation for graduate students. The project can be done at any level of complexity, from childlike exploration to professional analysis. Although each level poses its own important questions about conservation, the basic issue that remains is how much land is needed to sustain species diversity.

The teachers and I selected a lawn behind a parking lot on the Princeton University campus. We worked in three teams of four people. One team started by staking string boundaries on the lawn in nested blocks. The blocks ranged in size from a meter square up to 16 by 16 meters. We set the boundaries for the largest area first. Because the ground bulged slightly (it made the sum of the four angles greater than 360 degrees), we fudged the plot into a square by making the diagonals equal in length. We divided this large square into four equal areas and then further subdivided one corner until the last blocks were one meter square [*see top illustration on opposite page*]. A tape measure and 3–4–5 right triangles came in handy.




***LEAF SHAPES provided the basis for distinguishing plant species. The "discoverer" named the plant, an honor that led to several idiosyncratic appellations. Only a few of the 34 species found are shown.***

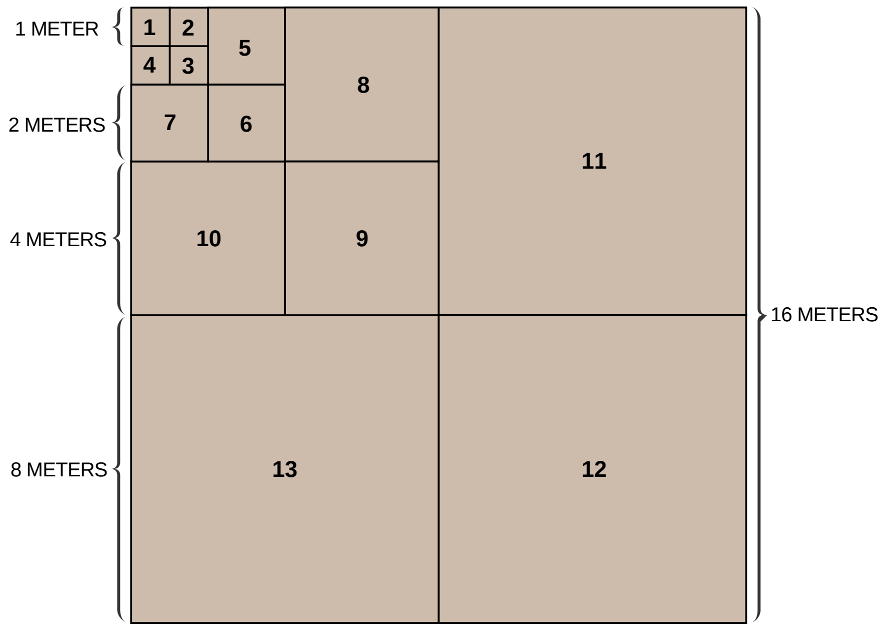
Of course, the area may be increased, or the smallest squares subdivided, depending on the number of species that appear during the investigation. A rough criterion for the right-size area is that a middle-aged and mildly myopic biologist can walk across it and count about 12 obviously different species. Such an area will yield about 30 to 40 species on closer examination.

For familiar plant species, we used the common name. A professional version of the activity would use a technical key to the flora. For our quantitative pattern and just for fun, we defined our own "species" by differences in the leaves. In effect, we were imitating the process by which the true species' names came about [see "How Many Species Inhabit the Earth?" by Robert M. May; SCIENTIFIC AMERICAN, October 1992].

We set up a "museum" of paper on which a "curator" wrote the name of each species found and taped a specimen next to it. While one crew set up the sampling boundaries, the other two explored the region for new species. Any specimen that showed novel features was taken back to the museum. The investigators compared the specimen with named species and assessed its novelty in consultation with the curator. If the specimen was truly new, it was added to the collection. The discoverer had the honor of naming it. Without thinking about it, we named species just as professional taxonomists do—as often for oneself or for a friend as for defining characteristics of the specimen, the habitat or related plants. Being amateurs, we could afford to be whimsical—hence, names such as "Hairy Harry" and "Itty-Bitty."

After completing the survey, we added the totals in each block. We also accumulated a running count of the numbers, starting with those in the smallest square and then adding those in subsequent blocks until we had included the entire plot. (A sample tally sheet appears at the right.) Even without technical analysis, the results provoke many interesting observations. Some species are common to nearly every block; others are rare. Some appear as lone or scattered individuals. Others are found in clumps of several individuals, although the clumps themselves are unique or scattered. Is there any pattern to which species are common and widespread, which are clumped, and which are rare and scattered?

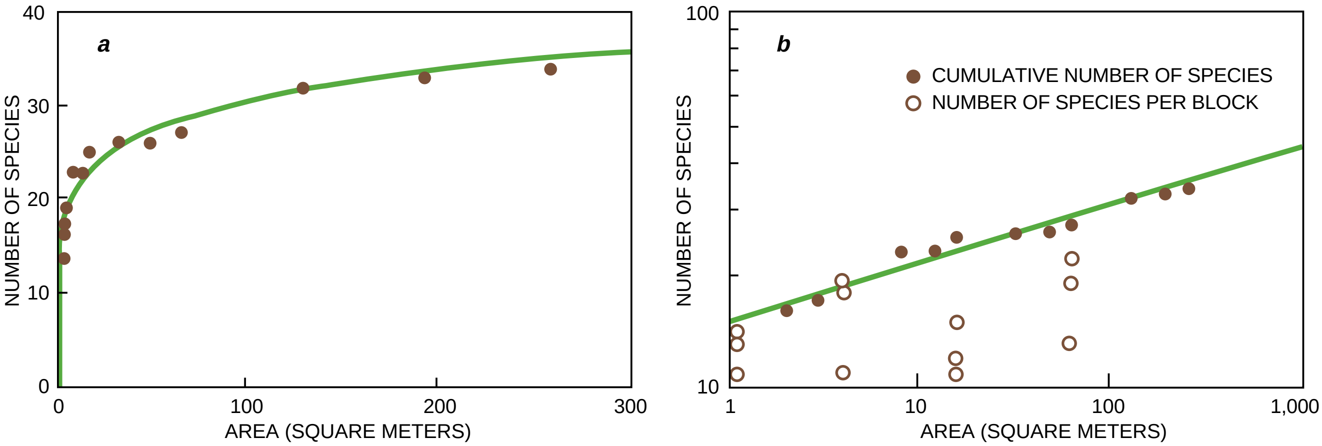
To explore for patterns, we plotted the number of species against the area surveyed in several ways. First, we graphed the cumulative numbers of species for each surveyed square, start-




*SAMPLING PLOTS consisted of 13 subdivisions of a 256-square-meter area. The plots were numbered in a clockwise spiral pattern.*

| SPECIES | BLOCK NUMBER | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 |
| TURKEYTRACK GRASS | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| SCOTT'S TOOTH GRASS | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| LITTLE BROAD-LEAVED GRASS | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| FAT-LEAVED GRASS | ✓ (red) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| WHITE CLOVER | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| ALSIKE CLOVER | ✓ (red) | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ |
| HOP CLOVER | | | | ✓ (red) | | | ✓ | | | ✓ | | | ✓ |
| YELLOW WOOD SORREL | ✓ (red) | | ✓ | ✓ | | ✓ | ✓ | ✓ | | | ✓ | | ✓ |
| MOUSE-EAR CHICKWEED | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ |
| JAMES'S 3-LEAF | | | | | | | ✓ (red) | | | ✓ | | | ✓ |
| SMOOTH-LEAVED BARBARA | | ✓ (red) | | ✓ | ✓ | | ✓ | | ✓ | ✓ | | | ✓ |
| HAIRY HARRY | | | | ✓ (red) | ✓ | | ✓ | | | | | | ✓ |
| BROAD-LEAVED PLANTAIN | ✓ (red) | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| FANCY MARY | ✓ (red) | ✓ | | | ✓ | | | | ✓ | | ✓ | ✓ | ✓ |
| DANDELION | ✓ (red) | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| ITTY-BITTY | | | | | ✓ (red) | | ✓ | | | | | | ✓ |
| WHITE ASH TREE | | | | | | | | ✓ (red) | | ✓ | | | |
| SPEEDWELL | | ✓ (red) | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ |
| BERT WEED | | | | | | | ✓ (red) | | | | | | ✓ |
| INDIAN STRAWBERRY | | | ✓ (red) | | | ✓ | | ✓ | | | ✓ | ✓ | ✓ |
| BROWN-TOP MUSHROOM | | | | | ✓ (red) | | | | | | | | |
| PETITE LISA | ✓ (red) | | | | ✓ | | | | | | | ✓ | |
| BOSS MOSS | ✓ (red) | ✓ | | ✓ | ✓ | | ✓ | | | ✓ | | | |
| FIELD SPEEDWELL | ✓ (red) | | | | | | | | | | | ✓ | |
| FUZZY CHICKWEED | | | | | | | | | | | ✓ (red) | ✓ | |
| SCARLET PIMPERNEL | | | | | | | | | | ✓ (red) | | | ✓ |
| TWIN BETTY | | | | | | | | | | | | ✓ (red) | |
| NARROW-LEAVED PLANTAIN | | | | | ✓ (red) | | | | | | | | |
| MYSTERY PLANT | | | | | ✓ (red) | | | | | | ✓ | | ✓ |
| NOVA TERRA SHARON | | | | | | | | | | | ✓ (red) | | |
| PRINCETON PARSLEY | | | | | | | | | | | ✓ (red) | | |
| POISON IVY | | | | | | | | | | | ✓ (red) | | |
| LINDA BERRY | | | | | | | | | | | | | ✓ (red) |
| ERNIE WEED | | | | | | | | | | | ✓ (red) | | |
| **TOTAL PER BLOCK** | 14 | 13 | 11 | 13 | 19 | 11 | 18 | 11 | 12 | 15 | 19 | 13 | 22 |
| **CUMULATIVE TOTAL** | 14 | 16 | 17 | 19 | 23 | 23 | 25 | 26 | 26 | 27 | 32 | 33 | 34 |
| **AREA OF BLOCK** | 1 | 1 | 1 | 1 | 4 | 4 | 4 | 16 | 16 | 16 | 64 | 64 | 64 |
| **CUMULATIVE AREA** | 1 | 2 | 3 | 4 | 8 | 12 | 16 | 32 | 48 | 64 | 128 | 192 | 256 |

*TALLY SHEET kept track of the species found. The red check marks denote the smallest block number in which the species was encountered. The cumulative total is the running count of the red checks. The areas are in square meters.*




***SPECIES-AREA CURVE shows that the cumulative number of plant species encountered increases with the area surveyed* (a)*. A logarithmic graph of the data reveals a straight line* (b)*, which provides the constants* c *and* z [*see box below*]*. The plots of the number of species per block, however, were inconsistent. Experimenter fatigue is a possible reason for the inaccuracy.***

ing with the most subdivided corner. This cumulative curve shows that 75 percent of our species are found in areas as small as 20 square meters [*see illustration above*].

To test the quantitative pattern we found against the traditional species-area equation [*see box below*], we plotted the same data on logarithmic axes. Some of our grade school teachers were wary of logarithms, but the sampling squares are already scaled multiplicatively by a factor of two in length, or a factor of four in area. A logarithmic scale is easy to construct for the number of species by marking fixed intervals on linear graph paper with 1, 2, 4, 8, 16 and so on. The bilogarithmic plot of our data is a straight line, which conforms to the theoretical generalization given by the species-area equation.

On the same graph, we plotted the surveys for each individual block. We expected the plots to show the same pattern as the cumulative data did, perhaps with a bit of variation and a slightly lower slope and species-intercept point. That is because the cumulative curve must rise continuously with increasing area. We discovered to our dismay that the pattern of the individual blocks was somewhat inconsistent.

Discussion suggested possible causes. One group admitted to being less than thorough in their surveys. They were more interested in the morphology of what they found than in the numbers. Several admitted to accumulating fatigue during the second hour of crawling around the larger plots. It is possible that lapses by one group or by a few individuals were compensated by others in the cumulative data, hence explaining the consistency of those data. It is also possible, however, that we underestimated the slope of the species-area curve for our lawn. In any case, the teachers were so impressed with the regularity of the cumulative data that they started an animated conversation about how to conduct more careful surveys the next time.

The discussion led to further questions. Can our results be safely extrapolated to areas larger than those sampled? How much area would be needed to preserve 50 percent, or even 90 percent, of the regional lawn species? How would the diversity of plants in real "islands" of lawn in a paved parking lot differ from marked-off samples of the same size in a continuous lawn? What insights does this analysis give into the planning of urban parks?

This exercise is just a conceptual metaphor for some far more practical uses of species-area curves. It is, however, a large empirical step toward making your own surveys of trees, shrubs, vines, wildflowers, ferns, mushrooms or vegetables in patches of various sizes. Then plot the number of species against area, think about the results and take your data to the next meeting of your local planning board.

## Deriving the Species-Area Curve

For many groups of organisms, the number of species encountered increases as the area increases. A suitable relation can be expressed as

$$S = c\,A^z$$

where $S$ is the total number of species observed in a surveyed area, $A$ is the area surveyed and $c$ and $z$ are constants fitted to the data.

Taking the logarithm of both sides gives

$$\log S = \log c + z \log A.$$

This equation is an empirical generalization. Many researchers are currently trying to pose theories that "predict" it. The reality of this equation can be tested, for a given region and group of organisms, by plotting surveys on logarithmic scales of both species and area to see if they conform to the generalization of a straight line. If they do, then the relation can be characterized by only two fitted parameters, $c$ and $z$.

As appropriate as this equation may be, the species-area curve is often more rhetorically convincing as an argument for conservation if the number of species and area are plotted linearly. Then it is clear that efforts to find as many species as possible have diminishing returns.

FURTHER READING

THE FRAGMENTED FOREST: ISLAND BIOGEOGRAPHY THEORY AND THE PRESERVATION OF BIOTIC DIVERSITY. Larry D. Harris. University of Chicago Press, 1984.

WEEDS. Alexander C. Martin. Western Publishing Company, 1987.

ECOLOGICAL DIVERSITY AND ITS MEASUREMENT. Anne E. Magurran. Princeton University Press, 1988.

NATURE RESERVES: ISLAND THEORY AND CONSERVATION PRACTICE. Craig L. Shafer. Smithsonian Institution Press, 1990.

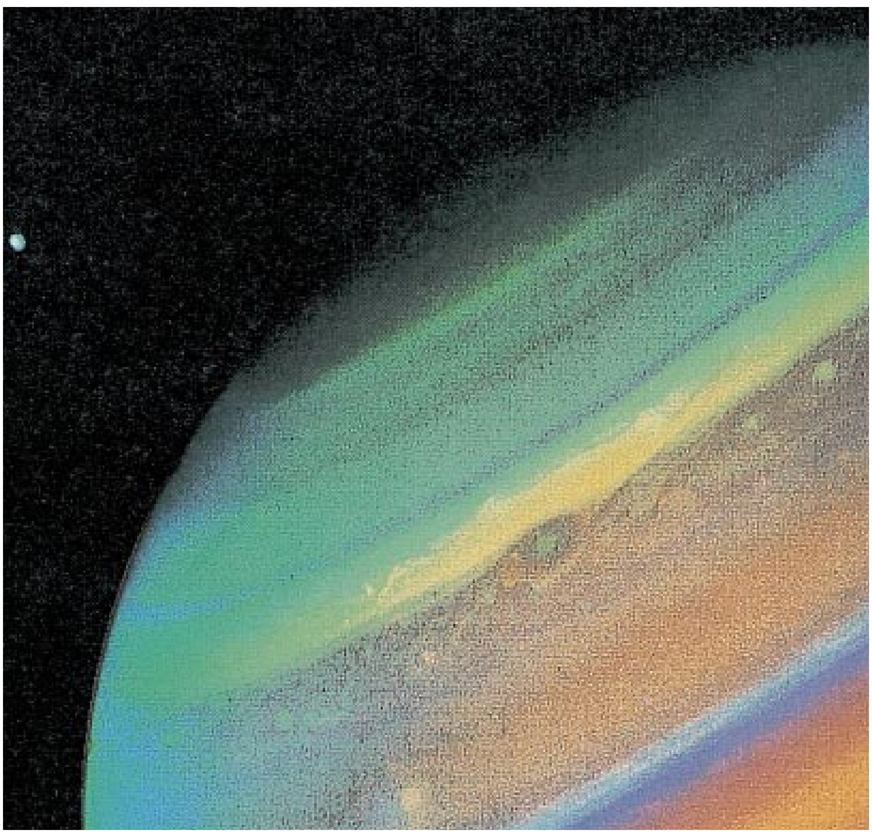
THE DIVERSITY OF LIFE. Edward O. Wilson. Belknap Press/Harvard University Press, 1992.

## Heavenly Objects

**SKY CALENDAR,** by the Abrams Planetarium (Michigan State University, East Lansing, MI 48824) ($6 per year, 12 monthly black-and-white sheets). **EPCOT SKY CALENDAR 1993,** by the Abrams Planetarium and Stephan Van Dam. Hyperion, 1992 (114 Fifth Ave., New York, NY 10011) ($12.95, wall calendar in color). **WANDERERS IN SPACE: EXPLORATION AND DISCOVERY IN THE SOLAR SYSTEM,** by Kenneth R. Lang and Charles A. Whitney. Cambridge University Press, 1991 (paperbound, $24.95). **IS ANYONE OUT THERE? THE SCIENTIFIC SEARCH FOR EXTRATERRESTRIAL INTELLIGENCE,** by Frank Drake and Dava Sobel. Delacorte Press, 1992 ($22).

The celestial action we all perceive with unaided eye is at its liveliest around sunset and sunrise, and it plays close to those decisive directions as they march seasonally along your horizon. Watch there every day you can, and follow the moon and really bright points up into the vault of the night sky, too.

The shepherd tending his flocks needed no schedule to draw his attention, but we who dwell indoors do. For some years now, the astronomer-artists at the Abrams Planetarium have made an extremely useful *aide memoire,* simple enough for tyros yet helpful to the experienced, including those who use binoculars or small telescopes. It has been noticed here on more than one January past. Their calendar offers a box for every day of the year, a month of boxes placed on a single page. Each box holds a small scaled drawing of just what sights to look for and when, clear weather assumed at the horizon (meant for viewers in middle northern latitudes). On the reverse of each page a neatly simplifed star map for the month urges your gaze up from the horizon and into the night hours. Text in the margins offers more data and much context. All you need to know is the time of day and the directions of east and west.



*WINDS ON SATURN form counterflowing eastward and westward patterns in the polar regions, as shown in this color-enhanced illustration from* **Wanderers in Space.**

Now the measured and beautiful dance unfolds with the year, linking happy watchers in a human tradition too old to estimate. A few 1993 events may draw you in: near dawn on 15 January, three luminaries will stand in a short straight line that aims right down at the still hidden sun; they are last-quarter moon, bright star Spica and brilliant Jupiter. If you miss that fine array, a quite similar line will form again before dawn on 21 November, with Mercury and Venus replacing the moon. Venus and bright star Regulus pose nearby at dawn during a couple of weeks in September, becoming on the 21st a mask of two unequal cat's eyes gleaming only a finger's breadth apart. A total eclipse of the moon will take place near midnight between 28 and 29 November, mid-totality coming everywhere about 1:26 A.M. Eastern time. The Pleiades and the Hyades clusters will provide a jeweled backdrop to the darkened and reddened moon.

This year a very attractive revision of the *Abrams Sky Calendar* has appeared under the aegis of Disney's EPCOT Center at Orlando. The original frugal offset pages, each evoking a single sheet straight out of the copier, are still available. But the little day boxes have metamorphosed into striking color through the work of a well-known New York cartographic designer. The material has been turned into a stapled calendar, each day box doubled in size and drawn against a colorful sky whose blues code for dawn, dusk or deep night. The rest of each month's double spread shows the month's star maps in color or large images from NASA, among them maps of the near and far sides of the moon, Jupiter's Great Red Spot or—most wonderful—a big showy *Voyager* close-up of Saturn, the rings curving past their own black shadow across the bright golden globe. That big calendar is just right on a wall that many passersby may consult; the loose pages provide as before an easy portability worthwhile for travelers and offer somewhat more fine-print detail, much simpler star maps for beginners and lower price.

*Wanderers* is a scrupulous yet imaginative introduction to what we know about our solar system and how we gained the knowledge. The two well-known astronomer-authors have avoided equations entirely—not one equal sign is to be found—but they have packed their book with data and graphs that summarize well what the probes have found. A number of works of art, mostly from our century (plus the Botticelli portrait of foam-born Aphrodite), help to extend rich factual images into apt metaphor. Particularly memorable is the circular poem-image by Annie Dillard mapping the wind regimes of the earth. The story of the moon's origin is given the latest and most plausible explanation: a glancing impact from a Mars-sized collider knocked a ring of matter out of both intruder and forming earth; that ring soon condensed into our outsized low-density moon. The book is at once an introductory text of stature and a thorough, serious and readable report for general readers, with much compact reference data. Chemical issues are less fully treated than one might wish, a limitation not uncommon among astronomers.

A few examples of argument and exhibit: an Ansel Adams photograph allows comparison between the dark mirror of the moon and the brighter granite of Half Dome, each lit by the same sun at the same distance; a photograph of auroral glow from space clarifies why dwellers in a band along the Canadian border west of the Great Lakes have the best auroral view anywhere outside of the polar regions; a chapter on asteroids, meteors and meteorites traces their lineage, even to a picture of icy Antarctic hills that hold a rich lode of the little fallen asteroid debris; *Voyager 2* discloses strange moons and stranger multiple rings around the planets beyond Saturn. So the chapters march in good order outward past every planet, to close with the trillion comets of the remote solar fringe. The last chapter orders us more in time than in space, examining the dimly lit birth of our solar system.

Is our system unique? It remains possible that in all the cosmos it is we alone who inhabit some wanderer of circumstellar space, although "many astronomers find it hard to believe." Radio astronomer and senior co-author of the fourth title above, Frank Drake has found it impossible to hold that particular belief ever since he was a talented first-year graduate student, when his discipline was growing up in America. He was the first person to muse effectively on how radio might open the way to purposeful signals between the stars. One night, all alone at the Harvard dish, he heard a "strikingly regular signal" above the stellar noise source that was his thesis topic. Its channel frequency was even offset to match the expected motion of the star cluster he was recording. The detection came to the eager student as a stunning miracle, one manifested only to him in all the world; it did not surprise him when soon his hair started to turn white! (He concedes that there is a family tendency to early graying.) But he investigated his miracle; when he pointed the dish away from the cluster, the signal continued loud and clear. Some quite earthly transmitter, probably a military source not bound by usual rules, had chosen a radio channel supposed to be held open for astronomy.

Within a few years the new Ph.D. was responsible for observations at the big new dish of the National Radio Astronomy Observatory at Green Bank, W.Va. Winning the consent of his famous and starchy Director, he quietly made ready the receiving equipment at hand to listen to a few nearby sunlike stars. Before his Project Ozma could begin in 1959, its rationale and possibilities became public knowledge. His own heady idea was in the air, but published first by a pair of Cornell physicists, outsiders to radio astronomy who had never heard of Drake or of Ozma (one of them was the present reviewer).

The entire book is as warm, candid and intimate as that anecdote. Drake is now fully established, a prematurely white-haired, fatherly, innovative leader of research, sometimes an observatory director himself. Drake's Ozma was followed by three decades of debate and pioneering receiver searches by individual astronomers in many lands. Together they act rather like one person who searches for a needle in a vast haystack by sorting a handful of hay each day as she passes by. But now an expert NASA team, starting with a share of time on two top American radio dishes, is listening systematically, greatly empowered by a modest investment in equipment that has applied the astonishing power of modern computing to listen to a multitude of channels all at once.

In this book, both names and ideas make news. The reader will encounter fascinating little scenes of turbulence and of concord, among such people as Carl Sagan, Timothy Leary (then in prison), Senator William Proxmire, Father Theodore Hesburgh and the late Dr. Iosif Shklovsky. An international cast of other men and women, from Karl F. Gauss and Nikola Tesla to contemporary radio astronomers and electronics wizards, has been entangled somehow or other in the same inquiry: Is anyone out there?

The great dish at Arecibo has just spent a couple of months in the initial search period with NASA's new gear. The grand find they dream of is "not a simple one to make. If there were once cockeyed optimists...there aren't any now.... We should be willing to sweat and crawl and wait." Yet we wait in hope; the only judgment that counts is the verdict after meticulous search.

## Earthly Creatures

**WALKER'S MAMMALS OF THE WORLD,** by Ronald M. Nowak. Fifth edition in two volumes. Johns Hopkins University Press, 1991 ($89.95).

Any mammalian browser will enjoy the diversity of one class of vertebrates laid across the 1,600 pages of this succinct, authoritative and most readable natural history. It lists genus after genus until all 1,100 recent genera have been described at least briefly. The particulars emphasize the nature of the living animal in its range and any relation it may have to our own abundant, capable, destructive yet hopeful species of the genus *Homo*. The length of discussion varies with interest and knowledge; the small cats earn an indulgent 20 pages with many a fine slinky photograph; the cheetah, a single famous genus and species, has one page and one photograph. Nine hundred genera are illustrated by clear black-and-white photographs of a living specimen of some typical species.

The originator of this much admired reference, Ernest P. Walker of the National Zoo, took 30 years to prepare the first edition of his book, written to interest and instruct the general reader and serve at the same time as a first resort for knowing professionals. He brought it out in 1964; this edition, like the fourth edition in 1983, was updated and improved by his able continuator, a Virginia mammalogist, very much in Walker's spirit. The apparatus of indexes and tables is easy to use, based on both the scientific and vernacular names. There are nearly 6,000 citations of the literature.

What can you find in this treasury? Look at random; soon a portrait stops the eye. This little creature is a narrow-footed marsupial "mouse" of northwestern Australia, with a tail that arcs upward like a rainbow, twice the length of its mousy body. Look more purposefully; you will find well treated all the seven species of rats commensal with humans, especially the black rat from Malaysia so abundant in the tropics and the temperate-zone "Norway" rat from north China. Try the largest of land carnivores, the brown bear we call the grizzly. It has a wide natural range, from the Alaskan islands all the way around the hemisphere to Eastern

# SCIENTIFIC AMERICAN

## CORRESPONDENCE

**Reprints** are available; to order, write Reprint Department, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111 or fax inquiries to (212) 355-0408.

**Back issues:** $6 each ($7 outside U.S.) prepaid. Most numbers available. Credit card (Mastercard/Visa) orders for two or more issues accepted. To order, fax (212) 355-0408.

**Index of articles since 1948** available in electronic format. Write SciDex™, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 355-0408.

**Photocopying rights** are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of SCIENTIFIC AMERICAN for the fee of $3.00 per copy of each article plus $0.50 per page. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 27 Congress Street, Salem, Mass. 01970. Specify CCC Reference Number ISSN 0036-8733/93. $3.00 + 0.50.

**Editorial correspondence** should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111. Manuscripts are submitted at the authors' risk and will not be returned unless accompanied by postage.

**Advertising correspondence** should be addressed to Advertising Manager, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 10017-1111, or fax (212) 754-1138.

**Subscription correspondence** should be addressed to Subscription Manager, SCIENTIFIC AMERICAN, P.O. Box 3187, Harlan, IA. 51537. The date of the last issue of your subscription appears on each month's mailing label. For change of address notify us at least four weeks in advance. Please send your old address (mailing label, if possible) and your new address.


Siberia, even including the mountains of North Africa.

No form is stranger than the pert little dwarf mongoose of East Africa. They dwell in harmonious groups of a few dozen, living among crevices and tree roots, seeking insects and small fruits by day. The groups are matriarchal families, ruled by a dominant female, the queen, with her consort. Only that pair produce offspring, and they suppress sexual activity among the working class, who form a hierarchy. The youngest members rank highest, mainly free to forage, though sometimes posted as group guards. Others nurse, tend and baby-sit the helpless young during a very long period of immaturity. When the queen dies, the group may split up. The analogy to social insects is plain; this mongoose genus closely resembles in behavior a strange social rodent of the region, the burrowing naked mole rats.

This is a book many libraries and many a reading family should keep on open shelves to help the puzzled and to tempt the curious.

## Step by Step

**THE STAIRCASE,** by John Templer, Vol. I: **HISTORY AND THEORIES,** Vol. II: **STUDIES OF HAZARDS, FALLS, AND SAFER DESIGN.** MIT Press, 1992 ($55 for the set; Vol. I, $27.50; Vol. II, $32.50).

The staircase is as treacherous as it is beautiful, at once "one of the most dangerous manufactured objects" and an art form with a glorious history. These two slim, beautifully made and illuminating volumes disclose these two sides of the topic. The Georgia Tech author, architect and engineer, who sees himself within the classical tradition, is persuasive that real understanding can come only by probing an issue from many directions.

Among us bipeds, steps are Neolithic in age or even older. They began as notched tree trunks leaning against a wall, a forked portion at the top to prevent rolling, and as ladders. Ladders can be designed for easy removal, a kind of moat against unwanted entrants: see Taos Pueblo. Typical ladders are steep, like the ship's companionway, so steep that you must face the ladder to go down it. At last, the oldest proper stair arrives, a simple, straight flight of steps, wide enough tread and low enough gradient to allow the user to face forward when going either up or down.

Up the designers climb, too, their staircases increasing in scale as they change in intent and surroundings, from old Delphi past Michelangelo, on to Le Corbusier and the Pompidou Center. Straight stairs may be walled in, open on one side, or open on both; some stairs patently invite and others daunt. Helical stairs, known earlier, became common only after the fall of Rome. Such structures save costly floor space and are defensible against forcible entry, but they require stonemasons of true skill.

The next evolutionary step is combination of the modules. Straight flights join flat landings. Dogleg stairs turn 180 degrees at a landing; two mirrored arcs of steps flow together at a balcony, often behind columns. Monk, Doge, Pope and King are given even grander stairs. These presage the baroque "staircases of honor," major integrating features of the whole edifice, built both out-of-doors and within. A century and more ago these left the aristocratic halls to engage pleasure-seeking crowds of burghers, until that theatrical staircase in the Paris Opera would give access to all public parts of the building, casting throngs of well-dressed spectators themselves as part of the spectacle. Its proud architect declared, "The Opera is the staircase." Romantic Victor Hugo thought the place "comparable to Notre Dame," but musical Claude Debussy held it rather like a Turkish bath.

In our age, elevators and escalators, and even ramps, begin to overshadow staircases, although public stairs remain visible, serving their old functions with newer designs; often even bare-boned fire stairs are boldly exposed.

The second volume looks squarely into the cruel face of stairs, to analyze the injurious accidents that take place there by the millions every year. "Much of the suffering is unnecessary." Like cars, soap and medicines, stairs are artifacts whose use entails some risk. One remedy is too little used: adequate research, both for accident prevention and for the reduction of injuries. It is as though only "a dozen small studies" of automobile safety had been made since 1900. "The automobile would still be the perilous machine it was at the turn of this century. The stair still is."

Epidemiology opens the study. The epidemic of falls is worldwide. Expect a misstep once in 2,000 stair uses, one that can be corrected before a fall, one minor accident among 30 such missteps, one hospitalization for each 50 minor injuries, and one fatality among somewhat fewer than 200 hospitalizations. The elderly are at particular risk from this plague. (This reviewer, elderly and somewhat infirm, will testify to long fear and avoidance of risky stairs.)

It is time that controls injury. The time scale of a fall is set by gravity and step size. The recovery from a misstep demands swift response by the nervous system. The reflex action to extend the foot takes enough time to fall two inches. Falls of less than that height are unretarded by muscular action and always lead to a jolt. Willed response comes too late to control the effect of encountering an unforeseen seven-inch step. The functional stretching reflex that might soften a fall also comes too late. Thus, "short, unexpected falls" often cause remarkable injuries; once control is lost the fall may continue, unless indeed handrails offer a second chance desperately reached for. (This reviewer regularly blesses well-designed and well-built handrails.)

Step geometry is a perpetual concern of architects. Alvar Aalto recalls a page in the *Inferno* where Dante says the worst thing there is that all the stairs have the wrong proportions. Modern building codes still adopt empirical proportions fixed in a rather idealized way by François Blondel in Paris in 1675. They were so clearly put that they have taken the force of law today in many places, although people's foot and leg dimensions have changed since, and Blondel's units are often even misconverted. Design tables are given here for better safety and comfort, the dimensions based on 20th-century work, not yet definitive.

Chapters on surfaces resistant to slip, on designing stairs safe for crowds, a full treatment of railings, and subtle studies of how people use stairs follow. Since walking gait is repetitive, precision is required; a "noise" in step dimensions of more than $^3/_{16}$ inch is a hazard. Lighting and the contrast and visibility of step edges, even distraction by the views a stair opens, are all sharply relevant to safe design. Transitions matter greatly; 70 percent of all stair accidents occur on the first three or on the last three steps of a flight. "Soft" stair design (but not too soft), using forgiving shapes and surfaces like up-to-date automobile dashboards, lies just ahead. A "quick design checklist" with much quantitative material ends the book.

The most interesting images in the book record human behavior on stairs over long periods, favored paths worn in stone as plainly as through grassy lawns. Erosion differs between ascent and descent. Coming down, the footstep rounds the step nosing; going up, it wears the treads concave. Study the old stone steps at Wells Cathedral or observe the paths of users in the New York subways: "Human behavior on stairs is largely consistent over time."

# The Life of a Black Scientist

I am a black professor at a major white university. As an immunologist, I am well known for my work, both in this country and abroad. I am odd in that there aren't many black scientists doing what I do. Achieving success in science is difficult, but my experiences have convinced me that it is harder for blacks, who must harbor greater ambition and develop a very thick skin.

How did I find my way here? A privileged background is certainly not the reason. I was born in 1936 in a rural community just outside Annapolis, Md. The state was then unabashedly segregated. The elementary school I attended was a two-room frame building that sat on blocks and lacked running water and central heating. I had only two teachers through those elementary years, and they both helped me appreciate the power of knowledge at a young age.

One incident stands out. I was in the second grade. My teacher, Mrs. Minor, sent me on an errand to the local store. I gave the clerk a dollar bill and received a handful of change, which I did not count. I returned with the item and the change for Mrs. Minor. Before I could turn around, she informed me that the clerk had short-changed me by a nickel. There was no punishment, but she did emphasize that it was the principle involved, and not the nickel, that was important. I was more than capable of performing the task given me, but I did not use that knowledge to ensure a fair transaction. The message to this seven-year-old was that he had better start using his brain. I have never forgotten that lesson.

As an elementary student I was oblivious to the unfairness of having poorer facilities than white students had. Perhaps this is not so surprising: our school reflected our standard of living. I was in my teens before our house had running water and central heating. My father worked as a laborer at the Naval Academy, my mother as a domestic. I am unusually lucky to have had parents with such vision and commitment: despite their modest income, they managed to set aside enough to pay a substantial part of my undergraduate education and that of my two brothers.

When I entered secondary school, I became quite aware of the inequities, but my own passion for learning and the dedication of my teachers enabled me to thrive. In that period I developed a passion for science that was nurtured by my teachers and admired by my classmates. My high school experiences left me feeling well educated and confident of succeeding at college. In those days graduates of my black high school continued their education at black institutions. For reasons I cannot clearly explain, I decided I would break tradition and attend a white university. This turned out to be the most significant educational decision I ever made, because it propelled me into the white academic and intellectual world, where I have remained.

After mulling through catalogues from a number of schools, I applied only to Ohio State University, where I was accepted without problem. During freshman orientation I discovered that about 5,900 of my 6,000 freshman classmates were white. At one session we were informed that only one in three of us would be around for graduation. I knew I would be among them. Indeed, I thrived in the competitive and somewhat impersonal environment at Ohio State and graduated in 1958 with a B.S. in microbiology. I had six or seven close black classmates. We stayed in the same dorm, had the same disciplined study habits and did well. We viewed our achievements not so much as black students but simply as students.

I chose to stay at Ohio State for graduate school to pursue research in microbiology. After many futile attempts to find a summer job in a laboratory, I turned to Matthew Dodd, from whom I had taken two courses in immunology. Although I received A's in both courses, I had initially avoided Dodd because his gruff manner intimidated me. He offered me a job, and the immunologic experiments in which I participated that summer spurred me to work with him toward a Ph.D. in immunology. It turned out I was not Dodd's first black student. Two others had preceded me.

Following graduation in 1962 and a year of postdoctoral work at Ohio State, I tried to find an academic position. I was singularly unsuccessful, and finally I accepted a research post with the Food and Drug Administration in Cincinnati. For the most part, I used the position for my own research interests and was among the first to show that staphylococcal enterotoxins have profound effects on the immune system. I was also one of the first researchers to study the role of interferons in immune regulation. The interferon studies were initiated in collaboration with Sam Baron at the National Institutes of Health.

Baron left the NIH for chairmanship of the department of microbiology at the University of Texas Medical Branch at Galveston and in 1975 invited me to join the department as an associate professor. Three years later I was promoted to full professor with tenure. As time passed, my program grew but so did personal strain between Baron and me. So in 1983 I accepted a visiting professorship at the University of Florida at Gainesville. In 1989 I was promoted to graduate research professor.

In addition to being successful in science, blacks who aspire to high academic position must cope with unique barriers. Here at the University of Florida, for example, I had broad faculty support for my appointment but little administrative backing. President Marshall Criser had to step in and insist. It seems that powerful forces must sometimes get involved in order to bring qualified black scholars into the academic community.

What role do the professional societies play in the life of a black scientist? Not a large one, from my experience. As an example, I have been an elected member of the American Association of Immunologists (AAI) since 1969. Until recently, the AAI has shown little or no interest in encouraging the participation of qualified blacks. Not playing a role in professional organizations is more significant than it sounds: it means fewer contacts in the various circles—the journals, the academic committees—that shape one's career.

I feel that my success comes from a combination of things. I discovered early that I could "cut it" cognitively in any kind of environment. I have an insatiable appetite for discovery that feeds and drives my love of research. I am ambitious and want to achieve. I am internally tough, which helps me function in an environment of people who do not much care for me. It is sad, but I feel that this is what it takes to succeed as a black scientist in a white intellectual environment. Thank God, it hasn't driven me crazy.

---

*HOWARD M. JOHNSON is a graduate research professor in the department of microbiology and cell science at the University of Florida at Gainesville.*