SCIENTIFIC AMERICAN

JUNE 1995
$3.95

The world's strongest magnets.

Is complexity a sham?

Found: 2,000-year-old blueprint.

***Picky wildflowers** choose which pollen to accept.*

SCIENTIFIC AMERICAN

June 1995 Volume 272 Number 6

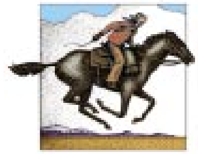
Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 1995 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in 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. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). 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; fax: (212) 355-0408 or send E-mail to SCAinquiry@aol.com. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.**

DEPARTMENTS

THE COVER painting portrays a mountain delphinium (*Delphinium nelsonii*), which possesses a rudimentary system of kin recognition. These wildflowers can chemically distinguish pollen of related plants, which tend to live near one another, from pollen of unrelated flowers. Natural selection has favored many other plants and animals that can identify kin, either with genetic labels or with environmental clues (see "Kin Recognition," by David W. Pfennig and Paul W. Sherman, page 98). Painting by Rosemary Volpe.

SCIENTIFIC AMERICAN®
Established 1845

PRINTED IN U.S.A.



LETTERS TO THE EDITORS

On the Road Again

"Improving Automotive Efficiency," by John DeCicco and Marc Ross [SCIENTIFIC AMERICAN, December 1994], reminds me of those paeans to the mainframe computer heard in the 1970s. Tomorrow's power train is likely to be a hybrid with electric motor wheels. The energy will be provided by a humble 20- to 40-horsepower, constant-speed internal-combustion engine connected to a generator, assisted by batteries that will ensure good dynamic performance. The big frontal engine and mechanical drivetrain will go the way of the dinosaur. It is clear that electric current and magnetic fields are much more amenable to computer control than are the current gears and hydraulics.

MARCEL COTÉ
Montreal, Quebec

I disparage the proposed use of public funding or legislation toward improving the fuel economy of the consumer automobile. The weight efficiency of mass-transit vehicles such as buses and trolleys overwhelms that of automobiles. In addition, mass transit delivers lower maintenance costs per passenger mile, less pollution and traffic reduction. Rail systems can also handle freight more expediently and cleanly than the swarms of trucks now wrecking our highway system. Restoring mass transit certainly will not cure pollution, traffic, escalating highway costs and so on, but government subsidy of individual commuting will abate these problems even less.

MICHAEL W. SHAFFER
Salem, S.C.

Preserving Bits

In the article "Ensuring the Longevity of Digital Documents" [SCIENTIFIC AMERICAN, January], Jeff Rothenberg too quickly dismisses standards as excessively restrictive solutions to the problems he describes. Standards can be designed to be open-ended and to provide for continuing evolution of technology while remaining compatible with existing applications; the FORTRAN programming language is one such example. The very existence of a standard makes it more likely that a number of computer systems will continue to support it well into the future. File-system standards in particular would probably be supported indefinitely, even after being superseded by newer ones.

TED TOAL
Nevada City, Calif.

Rothenberg has severely underestimated the physical lifetimes of digital magnetic tape. A chart in his article indicates a lifetime of only one to two years. Experience indicates, however, that physical lifetimes for digital magnetic tape are at least 10 to 20 years. Properly cared for reel-to-reel nine-track computer tapes recorded in the 1970s can still be played back. Given that digital-recording technologies can be supplanted by a newer format every five to 10 years, the bigger problem facing archivists is the lifetime of the technology, not the lifetime of the medium. Of course, media life expectancies are like miles-per-gallon ratings on cars—"your actual mileage may vary" according to storage conditions.

JOHN W. C. VAN BOGART
National Media Laboratory, 3M
St. Paul, Minn.

Rothenberg replies:

Open-ended standards by themselves do not solve the problem of digital longevity. They cannot bridge major paradigm shifts, such as those between hierarchical and object-oriented databases or between linear, textual documents and distributed hypermedia. And given the infant state of information technology, it is premature to attempt to design long-lived standards. File systems, which Toal cites as examples of such standards, serve better as counterexamples: several generations of file systems have already come and gone.

As to the question of tape longevity, Van Bogart himself points out the difference between theoretical lifetime and what is achievable in common practice. The experiences of many data administrators indicate that the reliable lifetime of tape under realistic conditions is often quite short. The crucial point is that the weakest link in the chain—whether it is the physical lifetime of the medium, the tenure of its format or the availability of required software or hardware—is what limits the longevity of digital documents.

Solid Old

Thomas T. Perls theorizes in his article "The Oldest Old" [SCIENTIFIC AMERICAN, January] that the oldest old constitute a select group of "the strong"—"the weak" having been weeded out earlier by death. He then concludes that as medical technology extends longevity, the old may be stronger than we expect and hence less of a burden than their increasing numbers would indicate. I would have concluded the opposite: that artificially restocking the "oldest old" population with some of the weak (by medically knocking out certain selection factors) would produce a population sicklier than expected because it would be more vulnerable to other factors.

JEROME W. RIESE
Appleton, Wis.

Royal Rubrics

If Brian G. Marsden wants to avoid controversial names for celestial bodies ["Science and the Citizen," "The Astronomical Naming Game," by Corey S. Powell; SCIENTIFIC AMERICAN, December 1994] but wants to use the name Elizabeth I for Elizabeth Tudor, he has obviously never been to Scotland. There Elizabeth Tudor is referred to as Elizabeth of England, to distinguish her from the present queen, Elizabeth (the first) of Britain. Because the Union of the Crowns intervened in 1603, there was never an Elizabeth II of England, and Elizabeth II of Britain has yet to put in an appearance. When mailboxes in Scotland first appeared with the notation "E II R," they were bombed in protest, so they now read just "ER."

DONALD M. GRAHAM
Vancouver, B.C.

Letters selected for publication may be edited for length and clarity. Unsolicited manuscripts and correspondence will not be returned or acknowledged unless accompanied by a stamped, self-addressed envelope.

50 AND 100 YEARS AGO

JUNE 1945

If you insist on eating arsenic but wish to avoid being fatally poisoned by it, perhaps you can save yourself by taking highly toxic fluorine compounds at the same time. A report of recent experiments with rats shows that drinking water containing potassium fluoride before and after the rats took arsenic trioxide mixed with sugar prevented the death of the animals. While the finding is interesting and may shed light on the action of arsenic taken into the human system, the method is hardly to be recommended for first aid since an overdose of fluoride, familiar as ant poison, kills one quite as dead as arsenic does."

"An electronic guardian for food storage space is the Sterilamp ultra-violet tube, which resembles a slenderized fluorescent lamp and kills bacteria and mold either on food surfaces or afloat in air. In addition to bactericidal rays, the lamps produce in the air a small amount of ozone, a colorless gas used commercially for purification of water and foods. Ozone scatters rapidly from the lamps and circulates throughout the storage space, acting as an additional weapon to halt mold growth."

"Less than two years ago producers were frantically trying to enlarge their production of penicillin to meet the tremendous demand for this wonder-working substance for the Armed Forces. Now, that goal has been exceeded far enough to allow placing penicillin in civilian drug stores and hospitals. This powerful drug, hailed as the most valuable development of modern medicine during the war period, is now available to any physician needing it."

"Those tightly organized minorities whose purpose it is to obtain special privileges at the expense of the public are exerting pressures that are pushing our democratic government out of alignment and causing it to function with impaired efficiency, according to Stuart Chase, author of *Democracy Under Pressure.* The author is not content with only describing the evil—he has a number of cogent proposals for curbing the pressure groups and lobbies."

SCIENTIFIC AMERICAN

JUNE 1895

The days of ballot box stuffing and other modes of cheating at elections appear to be numbered. Inventive genius has provided machinery that will not lie and will not allow deception at the polls. As soon as the voter has recovered from the shock of the sudden and rather awful imprisonment in a chamber of steel, he is able to realize what is expected of him. Inside the voting machine, names of the candidates of the democratic party are printed upon a yellow background, candidates of the republican ticket upon a red background, and prohibition candidates upon a blue background. To the right of each name is a little knob which he must press in order to register his vote—the machine does the rest."

"The cause of baldness in man is said by Dr. Leslie Phillips to be the fact that he cuts his hair. He says: 'In men the hair is habitually cut short from childhood, while in women the converse is usually true.' Dr. Phillips warns the 'new woman' against wearing her hair short. Almost every theory has some defect, and we might ask Dr. Phillips why men who clip their beards or shave for a long time do not get bald on their chins?"

"The great Chupaderos meteorite, which was discovered broken in two immense pieces in the year 1581, may now be seen at the portal of the National School of Mines, in the city of Mexico. The form of the two pieces leaves no room for doubt that they were originally parts of one great meteorite weighing more than 27 tons. The two sections were found 800 feet apart, at a point 900 miles from the city of Mexico. More than three centuries later, in 1893, the pieces were carried to that city and placed in their present position."

"If we examine certain small crustacea under the microscope, we shall be much surprised at their odd aspect. We shall mention a few examples that are illustrated here. The Calocalanus pavo, which is quite common in the Mediterranean, exhibits a transparent body, and at the extremity of the abdomen, carries eight golden yellow symmetrically arranged feathers. The Copilia vitrea is an odd little animal, every leg of whose transparent body is provided with a rich fan of microscopic feathers of a brick red. The Pontellina plumata presents a multitude of silken hairs that ornament its legs and abdomen. These hairs, often arranged as feathers, contrast by their bright orange color with the blue body of the animal and its transparent limbs. There is no doubt that these numerous hairs singularly facilitate the aerial excursions of the crustacean and uphold it after it has once risen into the air by a leap."

SCIENCE AND THE CITIZEN

Persistently Toxic

The Union Carbide accident in Bhopal continues to harm

More than 10 years after a sparkling mist of methyl isocyanate, or MIC, floated onto the sleeping Indian city of Bhopal, its impact has yet to dissipate. Four fifths of the hundreds of thousands of survivors of the disaster have not received any compensation. And they are still sick. Because MIC has no antidote, victims were treated only for burning eyes and choking lungs, while the poison lodged itself in vital tissues. Now a bewildering array of ailments—respiratory, ophthalmic, intestinal, reproductive and neurological—is belying early hopes that the repercussions would be short-lived.

The accident occurred in December 1984 when water leaked into an MIC tank at a Union Carbide pesticide plant, setting off a runaway reaction that released much heat and vaporized between 30 and 40 tons of the gas. The cloud settled over 30 square miles, exposing up to 600,000 people. Choking and burning, inhabitants woke up, tried to run but succumbed on the streets. Many died in fields, forests and towns far from Bhopal. Although the official body count was 2,500, unofficial estimates—based on the sale of shrouds and cremation wood—start at 7,000.

Although the route by which MIC invades the body has now been elucidated by animal studies, little has been reliably learned of the toxin's effects on the people of Bhopal. "For a disaster of this magnitude," observes Ramana Dhara of the Agency for Toxic Substances and Disease Registry in Atlanta, "almost nothing has been published." For political reasons, extensive surveys conducted by the Indian Council of Medical Research (ICMR) are unavailable through official channels. Nor are any of their findings being used to aid the survivors. "In all these years, no medical protocol has been developed for treating the victims," states Rosalie Bertell of Canada's International Institute of Concern for Public Health, who recently led a team of doctors in an independent survey.

The unpublished ICMR studies—which followed 80,000 people over nine years—indicate that at least 50,000 people are still suffering. Despite migration, the occurrence of ailments in the affected area was 27 percent in 1989, as opposed to 18 percent in control subjects. And a 10-month study from 1989 to 1990 found that the mortality rate was 16.7 percent higher in the severely exposed region. This number and current newspaper accounts suggest that one person is dying every two days from effects of the gas.

From the outset the lack of information on MIC has hindered the treatment of survivors. Although MIC has been commonly used, its toxicity had been little studied; some scientists who tried to work with it said it was too dangerous. (It is more potent than phosgene, which was used as a poison gas during World War I.) In 1963 and 1970 Union Carbide commissioned animal studies on MIC but until 1987 treated the results as proprietary.

RAGU RAI Magnum

***CHILD'S BURIAL** is one of the lingering images from Bhopal. Health effects linger as well.*

Thus, those who tried to help the injured in Bhopal had no clue as to what they were up against. Ajay Khare, an ophthalmologist, had visited the MIC unit before the accident: "I knew only the name, not the properties." Sufferers who went to the factory's clinic were told that the gas was not, in fact, toxic: washing the eyes and drinking water would take care of the problem. Indeed, because MIC reacts readily with water, a wet cloth placed over the face would have stopped it from penetrating. But since no one knew to do this, the gas broke down cells in the lung walls, inducing respiratory failure.

Fifteen weeks later researchers documented that 38 percent of 261 subjects living within two kilometers of the plant had burning eyes, 19 percent had diminished vision and 6.5 percent had corneal opacities. Breathing problems were also widespread. Three months after the accident the Industrial Toxicology Research Center, based in Lucknow, reported that in a group of 1,279 persons from surrounding areas, 39 percent had some form of respiratory impairment. An unpublished ICMR study from 1989 found that 93.4 percent of 1,601 gas-affected children suffered from cough. New cases of asthma keep showing up as the population ages.

Many patients were also afflicted with pain in the gastrointestinal tract, liver and kidneys. And women had additional complications. Daya R. Varma of McGill University found that in a sample of 865 women who were pregnant at the time of the accident, 43 percent miscarried; 14 percent of the babies born died within a month. Another survey reported that the two surviving infants out of 38 pregnancies had limb deformities, spina bifida and heart disease. Of 198 women residing within 10 kilometers of the plant, 100 had abnormal uterine bleeding. Because reproductive disorders are so commonplace, young women who were exposed to the gas are assumed to be infertile, and now no one will marry them.

One ICMR survey has also indicated that MIC attacks the neuromuscular system: 72 percent of exposed persons had muscular weakness, compared with 0.2 percent of controls. Neurological disorders, such as intermittent loss of memory, have recently been observed by Thomas J. Callender of the Environmental Occupational Medical Research Institute in Lafayette, La. Just as debilitating are the psychological illnesses. Those living in the worst-affected areas were laborers, who became unable to earn a living. Anxiety and depression have set in, along with posttraumatic stress and pathological grief, including guilt at not having been able to protect loved ones.

Although they clearly document devastation, many of the surveys have weak epidemiology and do not lead to firm conclusions about MIC's effects. Controversy remains even about just what gases escaped on that ill-fated night. Some physicians, such as Heeresh Chandra of Gandhi Medical College, became convinced that they were seeing symptoms of cyanide poisoning. At 350 degrees Celsius or higher, MIC starts to break down to hydrogen cyanide (HCN), so presence of that gas, along with others, was not inconceivable.

Two days after the explosion, scientists from India's Air Pollution Control Board found cyanide near the MIC tank at the factory; in addition, Max Daunderer of the Munich Institute of Toxicology detected cyanide in patients' blood. Union Carbide denied any possibility of cyanide poisoning—perhaps, as some doctors suggest, because the toxicity of HCN was well documented, as opposed to the unknown MIC, and would have expanded the scope of legal claims. The state government ordered doctors to stop administering the harmless cyanide antidote, sodium thiosulfate. When physicians at one clinic, observing that the antidote was relieving the symptoms of many of their patients, persisted in using it, police arrested them and their paramedics, keeping eight in jail for two weeks and confiscating their supplies.

That controversy may never be resolved, but releasing the government data could still help the survivors. The Indian government is disbursing $470 million that it obtained from Union Carbide, for an estimated 3,000 dead and 52,000 injured. But more than 16,000 claims for death and 600,000 for injury have been filed with the tribunals in Bhopal. Eighty percent of the death cases and 30 percent of the injury ones have so far been heard; the remainder could take another five years. Half the claims are being rejected, mostly for want of documentation. Court-ordered medical tests measured only pulmonary and ocular impairment; that other ailments are related to the gas is very hard to prove. Vibhuti Jha, a local advocate, does not quite get how it works: "The government represented the victims. Now it says to them, prove your injury to me."

Perhaps the least of the survivors' worries is that toxic breakdown products from the explosion might still be around. Despite contradictory clues, the National Environmental Engineering Research Institute in India concluded in 1989 that there was no contamination. But in the same year the Citizens Environmental Laboratory in Boston found dichlorobenzenes in drinking water from near the plant. The laboratory is testing more recent samples of soil and water. Meanwhile residue in the MIC tank continues to be highly dangerous.

Despite the medical and legal morass, the survivors seem to retain faith in the human spirit, weaving legends around the accident. "One factory worker," relates taxi driver Firoze Muhammed, "gave his life staying there to plug the leak. Else many more people would have died." But others insist that Shakeel did not die; he still lives, somewhere in America. *—Madhusree Mukerjee*

This is the first of a two-part article on the aftermath of Bhopal.

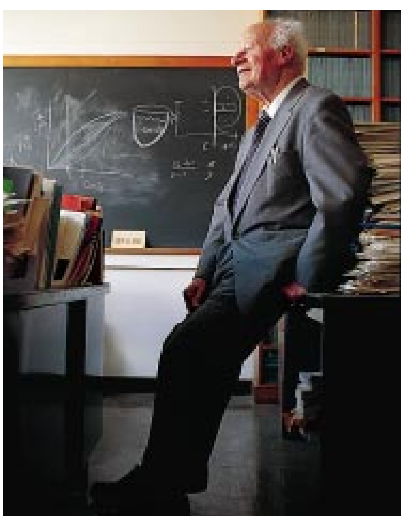
FIELD NOTES

Ambivalent Anniversaries

"We're Off to See the Wizard" loses its lilt as it peals from the 10-ton bells in Cornell University's clock tower. Yet somehow the incongruity of the tune of an Oz medley and its tone when hammered on bronze fits this April Fool's afternoon. Top physicists have converged in Ithaca, N.Y., for a symposium honoring Hans A. Bethe's 60th anniversary on Cornell's faculty.

At 88, Bethe is a living legend in theoretical physics, and the event can hardly avoid an awkward feeling of pre-posthumous commemoration as speakers recall his influence on giants he has outlived: Niels Bohr, J. Robert Oppenheimer, Richard P. Feynman. One presenter flips up slides of blackboards scribbled on by Feynman, architect of quantum electrodynamics and one-time Bethe subordinate, just before his demise. Below "Should I accept BBC interview?" and next to "Learn how to solve every problem that has been solved" are "To learn: Bethe *Ansatz"* and three questions about this important mathematical technique, which Bethe invented when he was 25.

I try, and fail, to avoid the implication of mortality when asking Bethe what he would like to be most remembered for. "Powering the stars," he replies, unfazed. His explanation of the stellar fusion cycle is, after all, why he was awarded the Nobel Prize in 1967. History might grant his wish. But in 1995 minds are focused more on a 50th anniversary than a 60th, and 50 years ago Bethe was leading the Manhattan Project's theoretical physics division as it prepared to test the first nuclear weapon. No surprise, then, that many of those paying homage to Bethe feel compelled to revisit the memory of the atomic bombs dropped on Japan.


ROBERT PROCHNOW

COMPLEX LEGACY of Hans A. Bethe includes explaining how stars burn—and how to build an atomic bomb.

Leading the agenda is Silvan S. Schweber, Bethe's biographer, who recalls "a kind of magic" at Los Alamos that made "everyone feel whole." He does not sully his account with talk of weapons. Freeman J. Dyson appeals more directly to the lessons of history, arguing that had the U.S. invaded rather than bombed Japan, "the results would likely have been as disastrous a failure as Hitler's and Napoleon's invasions of Russia."

As for Bethe, he has faced the postwar consequences of his work by arguing frequently, emphatically and, for the most part, bootlessly for arms reduction and against missile defense schemes. "Both sides should eliminate all but a few hundred warheads," he reiterates to me. One can always hope, I think, as the last heavy metal strains of "Somewhere over the Rainbow" are carried off by the chill breeze. *—W. Wayt Gibbs*

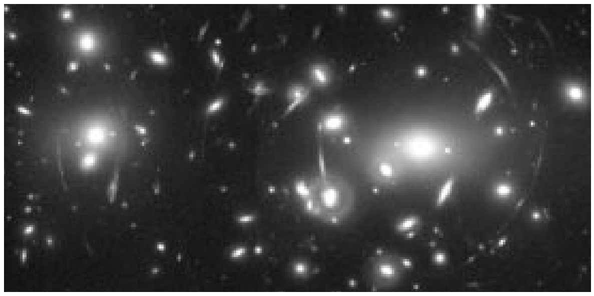
Crisis? What Crisis?

Reports of cosmology's demise have been greatly exaggerated

"Unraveling Universe," warns *Time* magazine. "Crisis in the Cosmos," declares the cover of *Discover.* What in the heavens is going on? Is modern cosmology coming apart at the seams? "The science is a lot more stable than you'd guess from the popular press," reflects Allan R. Sandage of Carnegie Observatories in Pasadena, Calif., one of the central players in the current drama. "I think the damn big bang model works."

The recent hoopla centers on a study led by Wendy L. Freedman, also at Carnegie, that has raised the perennial question of the age of the universe. Astronomers estimate the age by measuring the distance to faraway galaxies and the speed at which they appear to be receding from the earth. From that information, they can calculate the rate at which the universe is expanding—known as the Hubble constant—and then try to infer how much time has elapsed since the big bang.

Using the *Hubble Space Telescope,* Freedman and her colleagues are attempting to compile accurate distances to a number of relatively nearby galaxies. That effort just produced a new measurement of the distance to M100, located in the Virgo Cluster. The M100 observation indicates a high value for the Hubble constant, which, if correct, implies a cosmic age as little as eight billion years. Yet other data indicate that certain stars are at least 14 billion years old, a seeming paradox.



SPACE TELESCOPE SCIENCE INSTITUTE

GRAVITATIONAL LENS may reveal the universe's age.

But measuring the distance to galaxies millions of light-years away is a job fraught with potential errors—as Freedman herself quickly acknowledges. Virginia L. Trimble of the University of California at Irvine points out that the distance to the Virgo Cluster remains uncertain because astronomers do not know where M100 is relative to the cluster's center; subtle compositional differences might also make the stars in M100 behave slightly unlike the ones in our Milky Way. Even now, however, Freedman is ready to assert that "the best values of the Hubble constant are high."

Others are not so sure. High values make for exciting stories, but many studies still run at odds with the new results. For more than 25 years Sandage has vehemently argued for a much lower Hubble constant, which could imply a universe as old as 20 billion years. Sandage, too, draws on the latest data from the *Hubble Space Telescope* to back up his claims, and he is not alone. Bradley E. Schaefer of Yale University compiled recently published estimates of the Hubble constant and found as much support for low values as for high ones. "There ain't no consensus out there," he concludes with a laugh.

And even if astronomers ultimately nail down the local Hubble constant, Trimble explains, "there is no one-to-one connection between the Hubble constant and the age of the universe"—it depends strongly on the (unknown) density of the universe and on which version of the big bang theory one looks at. Moreover, the local rate of expansion may not reflect what has happened at other places and other times. P. James E. Peebles of Princeton University takes the seeming discrepancies merely as a sign that "we need to be a little elastic in our thinking."

Much of the air of crisis is the scientists' own doing. "There's a certain natural overexuberance—you wouldn't go after these very difficult observations unless you were very excited about the problem," Peebles says. "There is also a practical reason people want to find an answer now—it's called funding." And then there is the related matter of publicity. "The popular press is responding to the loudest noise, and anything the *Hubble Space Telescope* does makes a great bang," Schaefer remarks wryly.

Answers rarely come instantly, however. Emilio E. Falco of Harvard University is confident that studies of gravitational lenses will lead to a meaningful measurement of the large-scale expansion of the universe, "but we may have to wait a number of years."

Veteran cosmologists accept such time frames. "I hope I live long enough to see it resolved," Trimble says. "But I'm not optimistic." —*Corey S. Powell*

Flying in the Face of Tradition

Avian evolution may have been anything but gradual

For the past 150 years birds have been considered an evolutionary exception. The dinosaurs may have died off 65 million years ago along with early mammals and other organisms, but according to most scholars, birds survived the Cretaceous cataclysms. Now, however, the classic timeline for avian evolution—an uninterrupted 150 million years—appears to be on its own way to extinction. With it goes the premise that vast time is needed to produce such diverse creatures as hummingbirds, penguins and ostriches.

"You're basically talking about five to 10 million years for every type of modern bird to evolve," says Alan Feduccia, an ornithologist at the University of North Carolina who is renowned as the cardinal advocate of the unpopular theory that birds did not derive from dinosaurs. This shorter timescale suggests a "major revolution in our thinking about how evolution occurs," he adds.

Feduccia's argument, proposed in *Science* earlier this year, posits that avian evolution is analogous to that of mammals. In his view, early proliferation of bird species during the Mesozoic was followed by massive extinctions at the Cretaceous-Tertiary boundary: only a few survivors slipped through the keyhole into the Tertiary. Then, in two spurts of evolutionary fervor, all modern birds arose from the lucky few. Within 10 million years a first phyletic surge had replenished the void left by the Cretaceous extinctions with the avian orders that exist today. The second phase filled the skies with trilling, twittering and cheeping, as the age of songbirds, or passerines, dawned.

"The old model was one of sluggish, gradualistic evolution, with all the modern bird orders appearing back in the Mesozoic and then oozing into the present. It makes no sense," Feduccia asserts. The proverbial canary in a coal mine illustrates his point: "Birds are the first environmental indicator of a disaster. If there's a catastrophe at the end of the Cretaceous, birds are going to be the first thing completely knocked out."

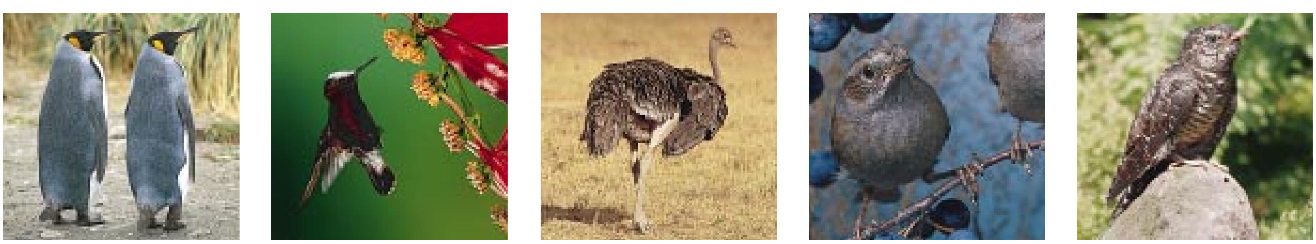
Feduccia says the theory has been gestating for years, but it was not until he was writing a book on the origin and evolution of birds (to be published by Yale University Press next year) that the picture came together. "In a flash it occurred to me that everybody has been wrong just because of tradition," he recalls. "The real question is: Why wouldn't bird evolution parallel mammal evolution? I think the beauty of this is the fact that it all of a sudden makes sense."

The idea of rapid morphological development among mammals is not so old itself. Only last year did Philip D. Gingerich, a paleontologist at the University of Michigan, describe the 10-million-year evolution of whales from land mammals. Gingerich's study, in turn, encouraged Feduccia to take on the intellectual heirs of Charles Darwin and Thomas Huxley. Indeed, until very recently, the fossil record had not provided a reason to believe in anything but lengthy, incremental evolution for birds. The several ancient-looking modern species—such as ostriches and loons—could be explained by Huxley's idea that they were the few Cretaceous survivors.

The foundation of this theory began to shake in 1981, when fossils of "opposite birds" were unearthed in China and, then, all over the Northern Hemisphere. Described by British Natural History Museum paleontologist C. A. Walker, these creatures—whose tarsal bones fused downward instead of upward like those of modern birds—predominated in the Mesozoic. Their inverted morphology needed explanation.

University of Kansas paleontologist Larry Martin provided one. He proposed that not only did opposite birds rule the roost in the Cretaceous, but they were different from all current birds. In Martin's view, the group became extinct at the end of the Mesozoic, and today's birds derived from what he calls "an infinitesimally small group of shorebird-like animals." More evidence for this revised time frame came from the Hawaiian Islands, where flightless ducks have evolved in less than four million years—a radically brief time in the old model.

If Feduccia and the others are correct, many additional ornithological assumptions may be subverted. Modern flightless birds in South America and Africa, for instance, are assumed to have an 80-million-year-old common ancestor in the southern supercontinent Gondwana. Such a history may have to be revised, and phylogeneticists using geologic timescales will have to recalibrate their molecular clocks. "If this new idea is correct," Feduccia says, "no modern orders can be attributed to drifting continents." —*Christina Stock*



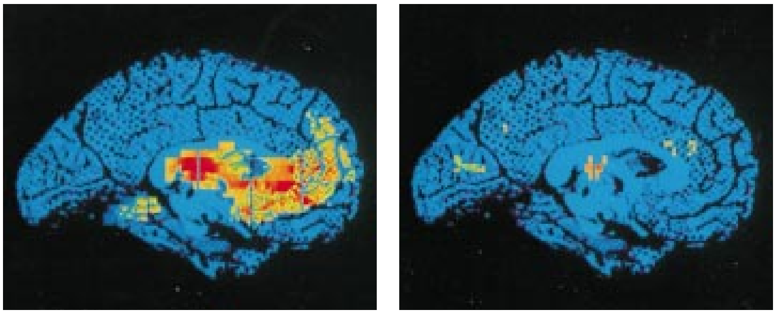
JOHNNY JOHNSON *Animals Animals*; MICHAEL FOGDEN *Animals Animals*; DAVID M. BARRON *Animals Animals*; HANS REINHARD *Bruce Coleman Inc.*; HANS REINHARD *Bruce Coleman Inc.*

BIRDS OF VARIOUS FEATHERS may have evolved very rapidly, according to a new theory.

Depression's Double Standard

Clues emerge as to why women have higher rates of depression

Mental health workers have long noticed a preponderance of women among the clinically depressed. Until recently, though, it was unclear whether more women than men were ill or, instead, whether more women sought help. In fact, a mounting collection of studies has confirmed that major depression is twice as common among women as it is among men. "This is one of the most consistent findings we have ever had," says Myrna M. Weissman of Columbia University. Women also seem more susceptible to milder melancholia and to seasonal affective disorder (SAD).



MARK GEORGE *NIMH*

***PET SCANS reveal that during sadness women's brains* (left) *become more metabolically active than men's* (right).**

Scientists searching for explanations are challenged by the fact that a variety of cues prompt depression in different people. Sorting out which factors might have a greater influence on women has not proved easy. Both sexes stand an equal chance of inheriting major depression, so genes are most likely not to blame. Yet hormones and sleep cycles—which differ dramatically between the sexes—can alter mood. Also, many workers have proposed that social discrimination might put women under more stress, thereby doubly disposing them to depression.

In 1990 Weissman and Gerald L. Klerman of Cornell University convened an international group to examine mood disorders. In the 10 nations reviewed so far, the team has found that among generations reaching maturity after 1945, depression seems to be on the rise and occurs at a younger age. Although overall incidence varies regionally, "everywhere the rates of depression among women are about twice as high as they are among men," Weissman says. In contrast, lifetime rates for manic-depressive illness do not differ according to sex or culture.

Meanwhile neurologists and endocrinologists suggest women may well have a biological bent for depression. Mark S. George and his colleagues at the National Institute of Mental Health (NIMH) recently studied which regions of the brain have increased blood flow during periods of sadness. They asked 10 men and 10 women to feel sad while they took a positron emission tomographic (PET) scan. The participants then judged how much sentiment they had mustered. George found that men and women deemed themselves equally sad, but "the brain activity of the two groups looked very different." Both sexes had equally activated the left prefrontal cortex, but the women showed blood flows in the anterior limbic system that were eight times greater. He has since compared feelings of anger, anxiety and happiness, finding no discrepancies anywhere near as large. Most significant, the regions of the brain activated during sadness are two that malfunction during clinical depression. George speculates that hyperactivity of the anterior limbic system in women experiencing sadness could, over time, exhaust that region and lead to the hypoactivity seen there during clinical depression. If he is right, the theory would explain the gender gap, at least in part.

Others at the NIMH have more to add. "There are hints of gender differences in both responses to seasonal patterns and to day and night, or sleep patterns," says Ellen Leibenluft. "Either might put women at a greater risk for depression." Thomas Wehr, also at the NIMH, has found that during the winter, women increase their nightly production of melatonin, a hormone whose levels are governed by the circadian pacemaker; women produce less melatonin during summer nights. Nocturnal secretions of melatonin in men are unchanging.

Another intriguing find is that without time cues such as daylight, women seem more prone to sleep excessively. (Patients who sleep a great deal during depression are, in fact, those who most often respond to light therapy, Leibenluft says.) Further, sleep and activity cycles are governed by the estrus cycle. Some conjecture that testosterone, which promotes activity, protects men against depression, whereas estrogen may lengthen the sleep phase in women. Gonadal steroids clearly regulate circadian rhythms in animals, and Leibenluft plans to see if they hold similar sway in humans.

George, too, plans to consider the effects of estrogen on brain activation levels during bouts of sadness. Epidemiological data indicate that hormones could play an important role in the onset of depression. Equal numbers of boys and girls experience depression before puberty, but shortly thereafter the rate among girls doubles.

The fact that many depressed patients are women of childbearing age must be considered in research efforts, Leibenluft emphasizes. She notes that although most psychotropic drugs are given to women (75 percent by some estimates), there is little information on how the menstrual cycle might influence the efficacy of these medications. Moreover, no one knows how menopause might alter the course of a mood disorder or its treatment. Because one in five American women has a history of depression, many of those who are going through menopause could be affected—especially as they often pursue estrogen replacement therapy, sometimes on top of an antidepressant regime. Says Leibenluft: "It is remarkable how little work has been done on this subject." *—Kristin Leutwyler*

Dinosaurs in the Halls

LOUIS PSIHOYOS *Matrix*

Although they have been gone for some 65 million years, the dinosaurs at the American Museum of Natural History in New York City are getting ready to shake off their dust and try some fresh moves. The new dinosaur halls, which open this month, incorporate reconstructed skeletons, renovated rooms and sleek, glass-paneled displays. The changes are part of an effort "to show the museum as full of life—not a dead, didactic place," says Ralph Appelbaum, the designer hired to oversee the remodeling.

Along with its "dusty, dingy reputation," the museum has also discarded the familiar scheme in which fossils were organized along a linear timeline, notes Lowell Dingus, the exhibit's project director. The dinosaurs are now grouped by kinship, so that strolling through the halls is "like walking along the very bushy evolutionary tree," he explains.

Some of the extinct lizards themselves have also undergone stunning changes. The *Apatosaurus* (formerly known as *Brontosaurus*) has a new head, four additional neck vertebrae and a dynamic tail-in-the-air pose, reflecting the latest understanding about dinosaur anatomy and locomotion. The *Tyrannosaurus rex* skeleton "has been completely taken apart and rebuilt," Dingus reports. It now hunkers down "in a stalking pose," gunning straight for unsuspecting visitors as they enter the hall.

Notably absent are life-size dinosaur reconstructions. Even the classic Charles R. Knight paintings sit discreetly in the background. Instead the museum emphasizes the fossils themselves: 85 percent of the material on view is real, not casts or replicas. The evolutionary configuration is intended to bring visitors in contact with the cladistics research going on behind the scenes at the museum, according to co-curator Mark A. Norell. "Lots of museum directors say, 'I need virtual reality,'" Appelbaum says excitedly. "My God, you've got actual reality—when did that go out of style?" *—Corey S. Powell*

Death by Analysis

Science by fiat could hurt the environment

One of the aims of the Republican majority in the House of Representatives, as detailed in its Contract with America, is to end government that is "too big, too intrusive, and too easy with the public's money." Yet several bills working their way through Congress could impede the government's environmental policy by requiring massive new scientific analyses.

"Current proposals before Congress would create a procedural nightmare and endless litigation that would hamstring effective administration of our environmental laws and effectively roll back environmental protection across the board," says Russell E. Train, who was administrator of the Environmental Protection Agency under presidents Richard Nixon and Gerald Ford.

One of the bills, the Job Creation and Wage Enhancement Act, has already passed in the House. It requires agencies to base decisions about protecting health and the environment on assessments of benefits, risks and the cost-effectiveness of the action. If the act becomes law, EPA administrator Carol M. Browner estimates that complying would require 980 new government employees and an additional $220 million a year—money that is unavailable.

Beyond the costs lie the legal ramifications of such a change in approach.


JOHN CHIASSION *Gamma Liaison*

TOXICS sit in a New Jersey warehouse awaiting new regulations.

Many programs would be vulnerable to legal challenge, because they employ a different regulatory tack. In the Clean Air Act revisions of 1990, for example, Congress mandated the use of the best economically feasible technology to minimize emissions of acid rain–causing pollutants from power stations. It chose that formula because cost-benefit analysis was impractical.

Proponents of Congress's new strategy argue that current environmental regulations err by excessive caution and that they are based on political expediency. But the Union of Concerned Scientists replies that the Republican bills stretch cost-benefit analysis beyond its capabilities. Nicholas A. Ashford, a professor of technology and policy at the Massachusetts Institute of Technology, states that in order to use the cost-benefit approach, as the bills require, agencies would need equally reliable assessments of different hazards—such as the risk of death in an automobile accident caused by less effective asbestos-free brakes versus the risk of cancer caused by asbestos. Agencies would also require a formula to compare such different consequences as the higher cost of lead-free gasoline versus the intellectual impairment of children through lead poisoning. "What is the value of that loss to society?" Ashford asks.

The bills also encourage risk managers to "split the difference" when there are different theories about an unmeasurable hazard, notes Adam M. Finkel of the Occupational Safety and Health Administration. Regulations now assume that a low exposure to a carcinogen produces a proportionately low risk. Another theory assumes that very low doses are harmless. As there is no easy way to tell which assumption is right, the new proposals favor using an average that could underestimate or overestimate risks.

As well as requiring complex analyses, the bills stipulate that agencies' actions be subject to review by peers, including representatives from regulated industries, and by the courts. Although judicial vetting is in principle limited to procedural matters, courts "have very little institutional capacity to deal with risk assessment questions," argues William S. Pease of the University of California at Berkeley; the bills represent "an invitation to litigation."

They also require agencies to compare regulated risks with other well-known hazards. But according to Paul Slovic, a past president of the Society for Risk Analysis, today's risk assessments cannot account for dimensions of risk that are important to the public. For instance, many people are willing to expose themselves to the chance of death in an automobile accident; the same people may justifiably object to a smaller risk of cancer caused by pollution in their drinking water. "The legislation being proposed is naive with respect to the complexities and limitations of risk assessment" and is likely to create "anger and distrust," Slovic states.

One showdown over Congress's enthusiasm for cost-benefit analysis seems likely to be on the floor of the Senate. Senator J. Bennett Johnston of Louisiana, who has previously encouraged the EPA to make greater use of such techniques, has urged colleagues to oppose the principal Senate bill, which was introduced by Majority Leader Robert Dole of Kansas. The suggested legislation, Johnston wrote, "has gone too far." President Bill Clinton, for his part, has threatened to veto acts that would weaken environmental protection. *—Tim Beardsley*

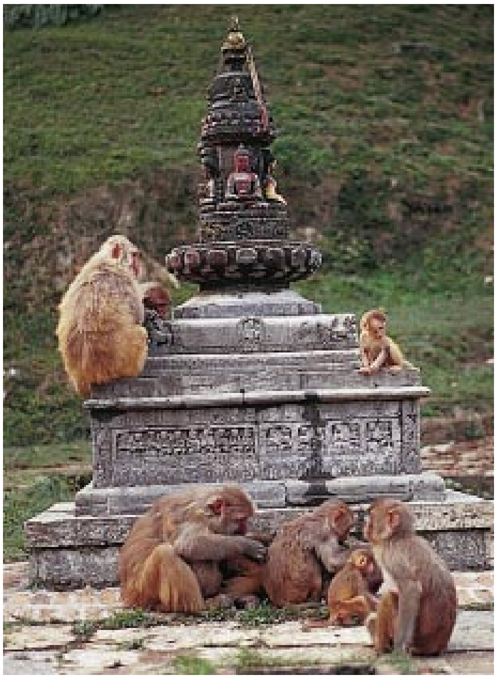
The Importance of Being Sneaky

Dominance may not be key to mating of rhesus macaques

"Power is the great aphrodisiac," Henry Kissinger once boasted to a newspaper. For cabinet officials and baboons, it may very well be, but for rhesus and possibly other macaques, novelty is beginning to look like an even stronger lure.

DNA fingerprinting and other data show that low-ranking males in a free-ranging troop of rhesus macaques have considerable reproductive success—so much, in fact, that researchers are having difficulty reconciling their results with the traditional view of paternity as a simple perquisite of high rank. The finding and related observations suggest that procreation of some wild macaques is tied up with strategies and patterns of emigration, rank and female preference. The news "makes social organization more complex and more interesting," says Irwin S. Bernstein of the University of Georgia at Athens.

Evidence that so-called sneaky matings can be a significant source of progeny is not new. With primates, however, the experimental record is inconsistent, with some studies—of, say, baboons—showing a strong correlation between high rank and reproductive success and others revealing little or no relation. As a group, macaques are rather diverse: with long-tailed macaques, a strong correlation has been seen; with rhesus and Japanese macaques, rank seems to be less of a factor.



M. N. BOULTON *Bruce Coleman Inc.*

RHESUS MACAQUES frolic at a Buddhist temple in Kathmandu, Nepal. Newcomer males may have a reproductive advantage.

In the past, these results were not considered definitive, because they used captive animals and inconclusive techniques. But in recent years, zoologists have begun using genetic techniques on wild animals—the study at the Caribbean Primate Research Center in Puerto Rico is the first to use such methods in free-ranging rhesus macaques. Behaviorists from the University of Puerto Rico joined forces with geneticists from several German institutes to follow a 100-member social group on Cayo Santiago, an island off the southeastern coast of Puerto Rico, as well as a captive group of 150.

In most regards, dominance appears to be a boon. High-ranking males invariably win contests for limited necessities and luxuries. They are groomed more often and have largely unrestricted access to females in estrus, who may or may not choose to mate with them or even suffer their presence. In contrast, a low-ranking male does not mate with a female in the presence of a higher male, who would disrupt such a coupling.

In the early 1980s John D. Berard of the University of Puerto Rico observed that high-ranking males on Cayo Santiago often forsook their spoils by emigrating to a different social group. There the male was consigned to the bottom of the hierarchy; upward mobility was a matter of outlasting males of greater status and forming allies. Being at the bottom of the heap is costly: on Cayo Santiago, Berard found that 21 percent of males emigrating from their birth groups died within one year. The deaths were attributable to many factors, including aggression from other males.

Subsequent observations began to show why males would take such risks. Those who had been with a group more than two or three years tended to have less mating success. In general, Berard says, the males did fairly well their first year in a new group, experienced peak sexual activity in their second year and then had diminishing opportunities beginning with the third year—just as they began moving up the social ladder. In a troop with, say, 100 or more members, a male rhesus typically would not begin moving up in rank before his third or fourth year.

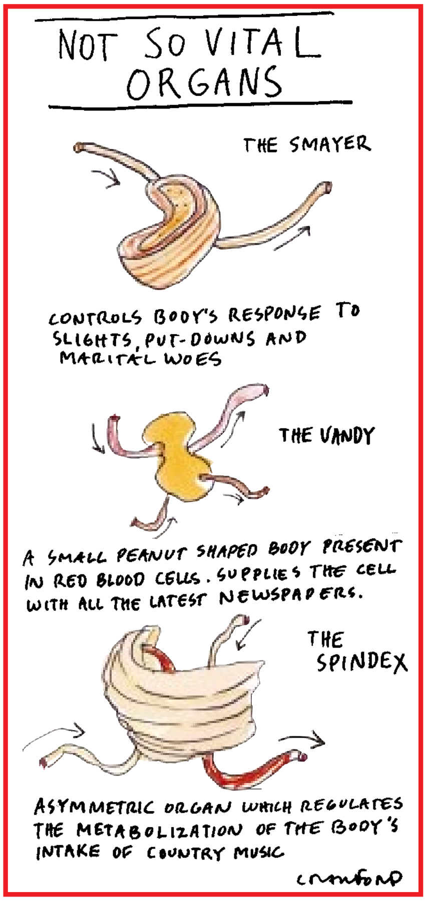
"The game is the low-ranking males hide, and the females try to get to them. But the high-ranking males follow the females and try to keep them from mating with the low-ranking males," Berard describes. This sneaky mating with low-ranking males is necessarily brief, usually less than 15 minutes. On the other hand, on a good day a high-ranking male might copulate half a dozen or more times with a female.

Nevertheless, preliminary DNA evidence shows the quick matings can be effective. In one study, covering a year of reproduction in social group "S"—one of the six on Cayo Santiago—two high-ranking males managed to sire two offspring apiece. But a male from a different group managed to impregnate two females from group S, and nine other males of mostly middle and low ranks had one offspring apiece. Previous work occasionally had similar findings, but without the details on the circumstances of sneaky mating and on diminishing mating opportunities among high-ranking males.

Given the long history of inconsistent results in studies of macaque reproduction, the findings have not exactly won over the primate research community. "What we need to do is identify under what social and demographic circumstances" low-ranking males are able to sire many offspring, says David Glenn Smith of the University of California at Davis. Various factors such as the size of a social group, the ratio of females to males and the age of male procreators may all be important, he believes.

The persistence of traditional, simpler theories of rank and reproductive success may be partly explained by their compatibility with results from the many captive groups, according to Fred B. Bercovitch of the University of Puerto Rico. In the confines of even a relatively large compound, the scarcity of hiding places appears seriously to cramp the style of low-ranking male macaques.

The mating sociology proposed by Berard seems to have an obvious evolutionary advantage—as well as suggesting more interesting roles for males and females. Males must weigh the reproductive opportunities of being an attractive new face against the dangers and stresses attendant to transferring and having low rank. By striving to mate with new males, females may be responsible for the promotion of genetic diversity. Berard notes that such diversity would be important for rhesus macaques, which occupy many different habitats in an area stretching across Asia. In addition, the strategy would tend to reduce inbreeding. "After five or six years, a high-ranking male probably has daughters in the group," Berard says. They are "better off mating with newcomers. It points out why familiarity breeds contempt." —*Glenn Zorpette*



MICHAEL CRAWFORD

Better Late Than Never

A chemical attack in Tokyo adds urgency to arms-control efforts

It is hard to imagine how a crime that left 11 people dead, thousands injured and countless others terrified could have a silver lining. Nevertheless, arms-control advocates hope the nerve-gas assault on commuters in Tokyo this past March could spur the U.S. and other nations to implement—at long last—an international ban on chemical weapons.

The Chemical Weapons Convention (CWC), which would prohibit possession and use of such weapons, has been under discussion for 26 years. It has been signed by 159 nations—but only 27 have ratified it. By far the most significant holdouts are the U.S. and Russia, which accumulated vast stockpiles of nerve gas—including sarin, the agent thought to have been used in Tokyo—during the cold war.

Barbara Hatch Rosenberg of the State University of New York at Purchase, an arms-control expert for the Federation of American Scientists, expects that when the U.S. ratifies the treaty, Russia will, too, since it will then obtain aid from the U.S. for the destruction of its chemical arsenal. Dozens of other countries will quickly follow suit, Rosenberg believes, thus satisfying the minimal requirement of 65 ratifiers for the pact to go into force.

President George Bush signed the treaty just before he left office, and the Clinton administration submitted the agreement to the Senate in late 1993. But the Senate failed to ratify it last year—less for ideological reasons than because the pact was not given high priority, according to Owen A. Kean of the Chemical Manufacturers Association, which has vigorously supported the convention. "As tragic as this event in Tokyo is, it may serve to get [the Senate's] attention," Kean says.

SYGMA

MASKED WORKERS clean up a subway in the aftermath of the recent nerve-gas attack in Tokyo.

Indeed, days after the incident, Lori Esposito Murray, the chief U.S. official responsible for the CWC, asserted that the ban could help thwart attacks not only by nations but also by religious or political terrorists. The pact would compel chemical manufacturers to record sales of potential ingredients, or precursors, of chemical weapons—making it more difficult for terrorists to obtain precursors covertly and easier for police to track down those who do.

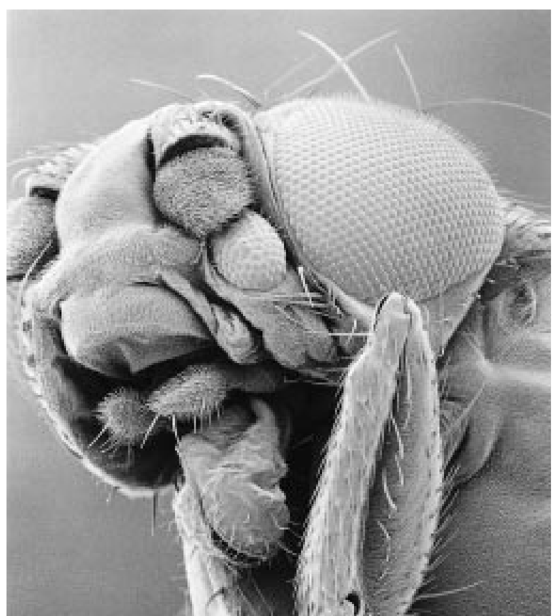
Matthew S. Meselson, a biochemist at Harvard University who is an authority on chemical and biological weapons, hopes successful implementation of the CWC would also improve the prospects for stiffer prohibitions against biological weapons. After all, he notes, a biological agent such as anthrax is more lethal by weight than the deadliest nerve gas. The 1972 Biological Weapons Convention has been both signed and ratified by 136 nations, including the U.S. and Russia, but the accord contains no verification provisions. Members of the treaty are now considering measures to improve "transparency," notably short-notice inspections of suspicious sites.

Frank J. Gaffney, Jr., a Pentagon official under President Ronald Reagan, argues that the attack in Tokyo demonstrates the "futility" of arms control; the U.S. should defend itself by maintaining a potent chemical arsenal and developing more effective technological countermeasures. Michael L. Moodie, an arms-control official during the Bush administration, concedes that arms-control agreements do not represent a "silver bullet," but neither do purely military measures. Even if international laws provide only a small measure of added security, Moodie maintains, they are worthwhile: "Clearly, we are going to need all the tools we have available." *—John Horgan*

The Eyes Have It

One cannot blame reporters for joking about the sci-fi movie *The Fly* or eyes in the back of heads. Nevertheless, these eye-studded fruit flies represent an advance in the search for a "master gene" controlling the generation of eyes. Researchers at the University of Basel in Switzerland focused on the *eyeless* gene, so-called because *Drosophila* with mutant versions of the gene often lack eyes. When *eyeless* was activated in regions of the fly embryo destined to become legs, antennae or other body parts, those sites spawned eyes.

The similarity of *eyeless* to eye-related genes in mice and even humans suggests that, contrary to current belief, all animal eyes may stem from a common evolutionary root. Could scientists grow poly-eyed mammals? "You could find thousands of reasons why it couldn't work," says Georg Halder, one of the Basel team. "But you could also have found thousands of reasons why it wouldn't work in *Drosophila*." *—John Horgan*



Rio Redux

Surprise! Promises of the Earth Summit are still unmet

For a fleeting moment in 1992, there was something close to an international consensus that humankind's ravaging of the earth's fauna and flora, together with the threat of global warming, justified better stewardship of the planet. In Rio de Janeiro dozens of nations signed conventions on climatic change and biodiversity and agreed to a lofty set of principles known as Agenda 21.

That was then. Fast-forward to 1995. Just as St. Augustine prayed for chastity—"but not yet!"—parties at the climate convention meeting in Berlin in April expressed an earnest desire do something about releases of greenhouse gases, chiefly carbon dioxide—but not yet. Only two or three developed nations have any real chance of keeping their emissions in 2000 to the levels of a decade earlier, the target vaguely endorsed at the Earth Summit. Data already show that the U.S. and Europe will probably go 6 percent over that goal.

Conflicting interests in Berlin ensured that the best that could be achieved was an agreement to talk soon. Oil-producing nations blocked agreement on procedures for voting. The Global Climate Coalition, an organization supported by fossil-fuel-burning industries, talked up the uncertainty of global-warming predictions. Poor countries were unwilling to accept limits that might imperil their economic growth; rich countries were unwilling to bear the burden of acting alone. One of the few successes was the acceptance of "joint implementation," which will allow wealthy nations to exceed targets if they support projects to reduce production of greenhouse gases in poorer countries.

Maurice F. Strong, the chairman of Ontario Hydro and a prominent big-business supporter of sustainable development, admitted during the Berlin meeting that "there is no question that there has been a recession of political will" since Rio. Total governmental development assistance decreased by 7.2 percent between 1992 and 1993.

The biodiversity convention, like the climate convention, is hobbled by a lack of consensus on voting procedures. The Washington, D.C.–based Global Environment Facility—a fund that was designated as the interim source of finance for projects under the biodiversity and climate change conventions—has only $2 billion to last until 1997. Other Agenda 21 goals, including negotiations to protect forests and plans to stabilize population, remain elusive.

Sustainable development, the pithy maxim that was on everyone's lips at Rio, is still, well, a pithy maxim. For all the talk, evidence of major decisions promoting sustainability is hard to find. The United Nations's Commission on Sustainable Development has produced "more words and wind than action," according to Gordon Shepherd of the International Fund for Wildlife, headquartered in Switzerland.

Nevertheless, the spell of lethargy could end soon. Global surface air temperatures are up: 1994 tied for the fifth warmest year in more than 100 years. If the unusual warmth of the 1980s returns, political temperatures might start to rise, too. —*Tim Beardsley*

THE ANALYTICAL ECONOMIST

Yesterday the Peso, Tomorrow the Dollar?

Last December, when the Mexican central bank ran out of money to support the price of the peso on foreign-exchange markets, the currency lost half its value rapidly. Since January, the dollar has been sliding less precipitously but inexorably down in relation to the yen and the mark. The detailed mechanisms behind the two declines are different, economists say, but together they help delineate the forces that drive currency trading.

In the long run, says Jeffrey A. Frankel of the Institute for International Economics, the relative values of dollars, marks, yen, rubles, pesos, pounds or forints depend on PPP—"purchasing power parity": prices rise or fall so that a dollar's worth of marks will buy about the same amount of goods in Germany that a buck buys in the U.S. There are a lot of problems with measuring the cost of a representative sample of goods in each country, and you need between 100 and 200 years of data to see the effect properly, Frankel says, but PPP is still the touchstone economists use.

For somewhat shorter periods, perhaps about 10 years, differing rates of inflation govern the movement of exchange rates, says Andres Velasco of Harvard University. If one country's prices rise by 5 percent a year, and those of another by 15, the first country's currency should appreciate by 10 percent a year (the difference) against the second. Differences in interest rates among countries will modify this equation somewhat: when the U.S. Treasury was paying high rates in the early 1980s, for example, everyone bought dollars because their net return was still high.

None of that, however, seems to make much difference in the day-to-day operation of foreign-exchange markets. Although many analysts blamed the start of the dollar's recent decline on international worries about the deficit, it is not as if everyone suddenly woke up in early 1995 and realized that $5 trillion of debt was a problem. And, as Velasco notes, the value of a dollar or a mark to a trader is what the next trader is willing to pay for it. Being right about the "fundamental" value of a currency is of little use if no one else is willing to buy or sell at that price.

Indeed, among the players who have learned this expensive lesson most often are central banks, which often intervene to control the price of their nation's currency. In Mexico the central bank maintained the price of the peso within a narrow

D. BUSQUETS-SORDO *Gamma Liaison*

PSYCHOLOGICAL CALCULATIONS may play as vital a role as financial ones in Mexican and other currency exchanges.

band for nearly four years by buying pesos and selling dollars at the official exchange rate—even though Mexican inflation was eroding its value. Eventually the bank ran out of dollars to sell, and the peso collapsed. Kathryn Dominguez of Harvard says that when central banks fight market forces to maintain a fixed exchange rate, they almost always lose. She points out that speculators who sense that a bank is running out of reserves can attack, forcing the bank to buy up more of its national currency and so deplete its remaining funds. They will then score a large profit after the ensuing devaluation.

Central banks whose governments are not wedded to a fixed exchange rate, in contrast, can both make money and have a significant influence on exchange rates. The U.S. has profited to the tune of several billion dollars from Federal Reserve trading during the past decade, according to Dominguez.

The key to such politically and financially profitable transactions, she says, is not so much in the detailed pattern of trades—the billion dollars or so that the Fed may put in play is no larger than what a large multinational corporation might stake—as in the information that government actions convey to other market participants. When the Fed moves to shore up the dollar, it may accompany its purchases with public predictions that U.S. interest rates might soon rise, thus making dollars a more attractive investment. Many observers have traced recent failures to halt the dollar's decline to a belief that the Fed will not raise rates because of the potential damage to the U.S. economy.

So what is next, and how much does it matter? Dominguez and her colleagues are fairly sanguine about effects on the U.S.: the falling dollar aids exporters, and imports from Japan and Germany are a small enough part of the economy that price increases for foreign goods will probably not cause significant inflation. In Mexico the peso crash helped to push up some consumer prices by 25 percent or more.

Paradoxically, the countries with stronger currencies may be hurt more as their exports dry up. Eventually exchange rates should settle to a new equilibrium, but, as Velasco points out, there is always another shock waiting to push them out of line.—*Paul Wallich*

TECHNOLOGY AND BUSINESS

Suck It to Me

Pneumatic tubes make a comeback

New York University professor Ken Phillips gives his class a trick question on examinations for a course on the history of technology. Students have to pick the fastest form for transmitting digital data: among the choices are a high-speed fiber-optic connection, a microwave radio link or a pneumatic tube. Invariably, Phillips's students think the pneumatic tube to be a ludicrous answer.

AIRBORNE EXPRESS by pneumatic tube delivers tools and repair parts at the Denver International Airport.

GEOFFREY WHEELER *Black Star*

They are wrong. What they forget is that some technologies still prove more efficient when electrons remain firmly attached to the odd bundle of protons and neutrons. A pneumatic tube can, in seconds, send over short distances anything from a vial of blood or a corned beef sandwich to a few 40-billion-byte data storage tapes. The 17th-century throwback is a kind of information superhighway of the corporeal. True, most department store clerks, equipped with the electronic cash register, no longer send money to a bookkeeper by tube. Nor does mail travel across Paris—or the Brooklyn Bridge, for that matter—by air pipe.

But in a small triumph for a reality that has yet to go virtual, the tube is back. Manufacturers have seen sales of their products rise in recent years. At about $100 million, U.S. sales have doubled during the past decade, according to TransLogic, a Denver-based company that is the largest domestic producer. (Market researchers ignore pneumatic tubes; they prefer to keep books on multimedia personal computers.)

Tubes have also helped automate one of the fastest-growing U.S. industries. They increasingly serve as the means by which blood or urine samples, spinal fluids and other specimens get from an emergency room or intensive care unit to laboratories. This form of conveyance is quicker and less costly than having a human porter do the same job. Kaiser-Permanente Medical Center in Los Angeles has a $3-million system of 15 miles of tubing that it uses to transport medical records throughout its 11-building complex.

Besides selling to hospitals, tube companies have installed their systems at airports for shuttling aircraft parts to hangar repair stations and in steel mills to move molten samples to a testing department.

This renaissance has occurred because of a marriage of the ancient with the new. Canisters, called carriers, that rocket through air-blown pipes can now be tracked along each leg of a journey with optical sensors that relay to a computer the whereabouts of a parcel. This precision tracking means the location of a carrier that gets stuck can be pinpointed within a network, making delivery more reliable.

Modern tube systems dispense with the 19th-century hand-operated bellows that created suction to move a carrier along. Yet the physical principles remain the same. Propulsion results from a combination of blowing and sucking. A motorized fan either directs a draught behind a carrier, or it removes air, creating a vacuum that sucks the container through the ducts. Air volume and pressure can be varied to control the carrier's speed, allowing it to be eased to a soft landing. Blood products might otherwise hemolyze, or rupture, because of exposure to high gravity forces through turns or on ejection from the tube. Tube networks are also equipped with railroad-track-like switches that route a carrier among different segments of the network.

The basic concept predates the industrial revolution by about a century. In the 1660s the Royal Society of London received a paper for a "double pneumatic pump," and a prototype was eventually constructed. One of the first working tube systems arrived in the 1850s. It was then that the Electric and International Telegraph Company built a 220-yard tube to relay paper telegrams to the London stock exchange—a Victorian version of electronic mail.

In the U.S., tubes became the means of delivering cash from a clerk to a store bookkeeper in five-and-dime stores. The bookkeeper would receive the money and send back the change. Tubes replaced "cash children," 10-, 11- and 12-year-olds who were exploited to perform this task.

The most elaborate plan for these air-driven guided missiles originated with a former editor and publisher of this magazine, Alfred Ely Beach. He secretly built New York City's first subway in 1870, a pneumatically propelled train that ran in a block-long, nine-foot-diameter tunnel below Broadway, near City Hall. But Beach made a fatal error by failing to enlist the support of then reigning Tammany Hall captain William "Boss" Tweed. When belatedly told of the tunnel train under Broadway, Tweed quashed Beach's ambitious intent to expand his subway citywide.

A variation on Beach's idea reemerged in the mid-1960s, when L. K. Edwards, an engineer for Lockheed Missiles and Space Company, took an extended leave of absence to become president of Tube Transit, Inc. Writing for *Scientific American* in August of 1965, Edwards proposed building two evacuated pipes from Washington to Boston that would let "tube trains" travel between the cities at 500 miles per hour, making the trip into a 90-minute journey.

Even today there is something about these burrowlike wall cavities that appeals to the rodent in everyone. A pneumatic tube became an important prop in last year's film revival of the radio classic *The Shadow.*

And a Japanese company has actually used a pneumatic tube to build a prototype of a better mousetrap. Ikari Corporation lines inner building walls with tubes that contain holes for the vermin to climb into. When a sensor detects the body heat of a furry little creature, a shutter closes over the holes and a plastic ball, blown through the tube by air jet, carries the mouse or rat through the tube and deposits it in a freezer.

What goes around comes around. But N.Y.U. professor Phillips may soon no longer be able to pop his trick question. Telephone companies are beginning to put in place fiber-optic networks that can carry the equivalent of many encyclopedias' worth of textual data in a mere second. Phillips does not fret, however: "Light waves still can't deliver a corned beef sandwich." *—Gary Stix*

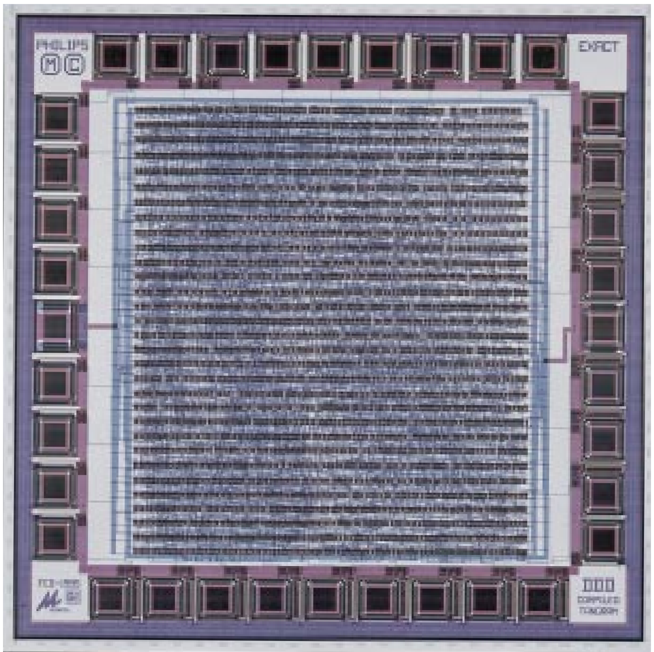
Turning Back the Clock

Reviving a challenger to the modern microchip

In computer design the obsolescence of technology is often planned. Resurrection comes as more of a surprise. The recent resurgence of an idea retired long ago—so-called asynchronous processors—thus amazes and confuses some mainstream computer designers. One has even offered a $1,000 wager that the research will fail.

The revivalists have called that bet. These researchers argue that microprocessors can work better when severed from the clock crystals that control their pulse in most contemporary computers. With working prototypes now in hand, they assert that clockless processors will soon become cheaper, more reliable, more energy efficient and easier to design than chips based on today's prevailing technology.

Nearly all microprocessors have stepped to the beat of fast-ticking clocks since the late 1950s, when the switch from vacuum tubes and relays to transistors and integrated circuits threatened to overwhelm engineers with complexity. Oscillators helped subdue design through discipline, marshaling the myriad components on a chip to march in lockstep rather than race about haphazardly.



PHILIPS RESEARCH LABORATORIES

***PROTOTYPE** error detector cuts power use 80 percent by running without a clock. But is it in sync with commercial reality?*

But simplification has its price. The clock can run no faster than the slowest part of the system without causing errors. That limitation forces designers to reduce performance by large safety margins and to optimize laggard circuits painstakingly by hand, increasing development time. More important, the faster, larger and more dense chips become, the more power and time they take to deliver the clock signal to the farthest reaches of the microprocessor.

In contrast, components on an asynchronous chip, like sprinters in a relay, run as fast as they are able and only when they are needed. In principle, that should improve performance a bit and reduce power consumption a lot. In practice, the first benefit may be illusory. All the time saved by fleet-footed components seems barely to balance the extra time needed to keep them from trampling one another as they pass batons of data.

But the second benefit—energy efficiency—appears real. Kees van Berkel of Philips Research Laboratories in the Netherlands plans to demonstrate this month a clockless error-correction chip for a digital compact cassette player. The chip uses only one fifth the power of its synchronous counterpart. Still, it is a relatively simple design.

Stephen B. Furber of the University of Manchester in England was more ambitious when he built an asynchronous version of the ARM6 chip that runs Apple's Newton. Although tests revealed that the clockless prototype ran slower and burned more electricity than the original, Furber was not discouraged. Simulations show that the next version, which should be finished later this year, "is going to be three times faster than the first and significantly better on power consumption," he says.

Despite its relative simplicity, Philips's chip does illustrate another potential advantage to asynchronous technology. "The design is expressed in a high-level programming language," van Berkel explains. "A so-called silicon compiler then translates the program automatically into circuits." He boasts that his compiler has created six chips, all of which worked perfectly the first time.

"In the long run," observes Alain J. Martin, a computer scientist at the California Institute of Technology who built the first asynchronous microprocessor in 1989, "the main advantage may be in the ease of design." Martin says his tools, which Caltech plans to license, can mathematically prove that a design is correct before it is built, avoiding embarrassments such as the division bug in Intel's Pentium processor.

By liberating the chip from a fixed clock speed, adds Robert F. Sproull, a vice president at Sun Microsystems Laboratories, "you can make one piece of a system go a little faster without redesigning the whole thing." The ability to evolve a commercial processor one small section at a time "would be a tremendous win in terms of time to market," he notes. Sproull is currently working with Turing laureate Ivan E. Sutherland, a fellow at Sun, on a novel asynchronous processor intended to rival the performance of Sun's SPARC chips. Martin is completing the design of a clockless version of Silicon Graphics's MIPS processor, with the less ambitious goal of reducing power consumption.

Researchers may find asynchronous plans a tough sell until they can convince pragmatists such as Gordon Bell of their commercial advantages. Bell, who invented Digital's VAX computer and led its R&D division for 23 years, howls with derision when he hears optimistic predictions for clockless computers. "I'll bet $1,000 that there won't be a fully asynchronous processor commercially available by April Fool's Day, 2000," he challenges.

Three researchers have rushed to pick up Bell's gauntlet: Furber (who points out that Advanced RISC Machines has already purchased all rights to his work), Richard Lyon of Apple's Advanced Technology Group and Takashi Nanya of the Tokyo Institute of Technology. But some see another reason underlying Bell's bravura. "Billions of dollars in tools and training have been invested in synchronous design," Sproull states. "It will be easier to adapt those than to start over with a whole new paradigm."

Yet HaL Computers has quietly replaced the division unit in its processor with an asynchronous version that is four times as fast. Martin says Intel has inquired about designing a small part of the Pentium to run off the clock. Although a grand revolution in chip design seems unlikely, a slow subversion might be inevitable. —*W. Wayt Gibbs*

Tell Us Another Story, Please, Bill

Microsoft's Hollywood forays may not have a happy ending

In the movie of the deal, it would be hard to cast the role of Bill Gates, leading his most recent venture into Hollywood. The basic plot is clear. Software magnate Gates invests tens of millions in superstudio DreamWorks—created by superegos David Geffen, Steven Spielberg and Jeffrey Katzenberg—in order to create a new genre of interactive entertainment. Gates seems clearly the outsider here, the kid. But the role could be played any number of ways. It could have the enthusiasm of, say, Mickey Rooney in *Babes on Broadway.* ("[It's all] just awesome to me," Gates gushed at the deal's announcement.) As comedy, it could take on some of the deadpan anomie of John Turturro in *Barton Fink.* Or Gates could be an empire builder with the overweening drive of young citizen Charles Foster Kane (played by a young, thin Orson Welles).

Any of these could work. It depends on point of view, on how the characters fit in with the tone and structure of the tale. Which is appropriate, really, because determining point of view is also the greatest challenge facing the deal itself—and every undertaking in new, interactive media. In joining with DreamWorks, Gates has allied himself with some of the most spellbinding storytellers and glittering stars that traditional media have to offer. The question, however, is whether they will still seem as spellbinding from the vantage of the media Gates would create.

In pure business terms, the logic of the agreement is compelling. Microsoft plays from a position of strength in all its other new markets. To establish the Microsoft Network, it is leveraging its dominant position in desktop-computer operating systems. To speed its move into electronic commerce, Microsoft bought Intuit, the leading maker of home-finance software. So as Microsoft moves into multimedia, it is natural that the company should try to capture the high ground in that domain as well.

DreamWorks, for its part, sits atop the ramparts of Hollywood. It can deliver brand-name stars who draw millions into movie theaters and will presumably provide similar appeal to interactive entertainment. But the business logic of the contract with Gates assumes that Hollywood's talents will dominate the new media as surely as they do the old. And that is a big assumption. At its heart lies the challenge of control—particularly of control over point of view.

Hollywood has raised control to an art form. Final cut, the size of the dressing room, or the hors d'oeuvres available for on-set snacking—the point of power in Hollywood is to use it to exercise control. Yet the point of interactive entertainment is precisely to cede that hard-earned control to some viewer in a T-shirt, slurping a beer, who just might decide that, say, *E.T.* would be a lot more interesting if the men in white coats got to dissect the alien after all.

Such issues of mastery over new media are not new to Gates. When he was licensing the digital reproduction rights to paintings now hanging in Britain's National Gallery (as well as other museums), one of the issues raised by curators was artistic integrity. They did not want future art lovers changing the colors of a Holbein to match the decor of their living rooms; they wanted to preserve Holbein's vision. But in Hollywood, issues of integrity also involve commercial considerations.

A star's brand name is the sum of his or her performances. Part of the promise implicitly made by Arnold Schwarzenegger as an action hero is that he will usually win in the end, however overwhelming the odds. But in an interactive world, the odds may prevail—with potentially damaging consequences for Schwarzenegger's reputation and his entertainment value.

More fundamentally, interactivity challenges the suspension of disbelief that gives Hollywood dramas their impact. Movies appeal to those seeking escapism because, by definition, their characters inhabit a world that is not our own. Yet there is no escape in interactivity; it injects our characters and our world into the plot—which shapes the process of narrative.

Imagine yourself as Cathy in *Wuthering Heights* (say, the 1939 black-and-white version with Laurence Olivier and Merle Oberon). You are sitting by the fire while storms crash upon the moors. Do you: (a) rush out into the wild night to meet your notoriously unreliable and unstable lover? (b) darn another sock? or (c) call together a group of friends to discuss the difficulties of maintaining self-esteem in a society with so little regard for women's rights?

One way or another, the act of choosing changes the story. And therein lies the rub for the DreamWorks team. The brands created by existing stars can probably sell escapist interactivity without emotional involvement—like the brain candy that is the Sega MegaDrive version of *The Lion King.* But once emotions get involved, the rules of the game change. In the new world, DreamWorks's strengths could look like weaknesses, leaving the big interactive prizes to some upstart, unencumbered by a dynasty or by tradition. Perhaps Orson Welles as Kane should play Gates in this movie after all. *—John Browning*

JERRY OHLINGER'S MOVIE MATERIAL STORE, INC.

EYEING THE VIEWER with suspicion may be the next drama for characters in the soon-to-be interactive movies planned by Microsoft's Bill Gates and the Hollywood dream team of Steven Spielberg, David Geffen and Jeffrey Katzenberg.

Have a Heart

Tissue-engineered valves may offer a transplant alternative

The lub-dub, lub-dub of a beating heart happens because the two sets of valves that keep blood flowing the right way close, one after another. Every year some 60,000 people do not hear this reassuring rhythm and must have faulty valves replaced. Unfortunately, the three available types of makeshift substitutes all introduce foreign tissue into the body. Yet recent work in tissue engineering indicates that patients could eventually receive valves grown from their own cells.

The first attempts at forming such homegrown organs were led by Joseph Vacanti of Children's Hospital in Boston and Robert S. Langer of the Massachusetts Institute of Technology. In the past several years the two groups have worked with liver, cartilage, bone, breast and certain muscle tissues. New findings, reported at the American Chemical Society meeting in April, just expanded the lineup to include heart valves.

Engineered valves have several advantages. Most patients today receive either a mechanical valve or one transplanted from a pig. The first kind are long-lasting, but the body can rebel against the intruding object. Animal valves are chemically treated to prevent an immune response, but the process kills cells and weakens the structures, reducing their durability. Human transplants are used successfully, but donor organs are scarce. With these obstacles in mind, Christopher Breuer, also at Children's Hospital, collaborated with Vacanti to develop a tissue-engineered valve.

The researchers have completed seven transplant operations in lambs, the standard test animal for valve surgery. They began by removing a one-centimeter-long section of the animal's blood vessel tissue, similar to that found in heart valves. The cells were grown in culture, purified, then placed on a biodegradable polymer, polyglycolic acid.

Breuer describes the polymer as a "jungle gym" on which the tissue can replicate and form the leaflets that make up a heart valve. As the cells reproduce, the polymer degrades, leaving behind only leaflet cells. How the tissue forms the correct structure remains mysterious. "Nature does a lot of work for us," Breuer says. After about 10 weeks, the leaflets are ready to be implanted.

Preliminary results have led the team to conclude that the transplants do function properly in the body. For example, using the technique known as cardiac catheterization, the group observed close to normal blood flow in the transplanted leaflet after one week. The researchers will now follow the lambs' survival rates, to determine whether the valves function well over the long term.

Although the workers will not speculate about when these transplants might be available, Vacanti calls the results "very important because they fit into the general story" of tissue engineering. The company supporting the study, Advanced Tissue Sciences, is in the final stages of exploring another tissue-engineered device made from skin cells. Researchers expect to complete Food and Drug Administration tests this year and begin marketing the treatment for burns and diabetic foot ulcers in 1997.

Until that time, the company has to worry about FDA protocol and cash flow: in 1994 Advanced Tissue Sciences lost $22.8 million, which it attributes to the high cost of human clinical trials. The future of tissue-engineered heart valves appears to depend on whether they can withstand economic as well as blood pressure. —*Sasha Nemecek*

Computing with Fire

Hollywood directors like to exploit our fascination with flames—preferably with minimal expense or risk. One way to reduce both would be to enlist a computer to create flames from a mathematical model of fire. But special-effects programmers, accustomed to stretching textured surfaces (such as dinosaur skins) over solid frames, have found fire's diaphanous dance and liquid spread devilishly difficult to model. So some experimenters have traded surfaces for clouds of tiny particles moving according to the physical laws of combustion. These models capture fire's behavior, but they tend to blur into glowing pointillistic mists.

The problem intrigued Christopher H. Perry and Rosalind W. Picard of the Media Lab at the Massachusetts Institute of Technology, and they invented an alternative technique that composes flames from virtual sparks. These shaded, translucent polygons rise from the point of ignition, shrinking until they wink out. Perry holds open the shutter of his software camera so that the journeys of just a few fat sparks trace a convincing flame. His model can set any object ablaze with a fire that spreads naturally and chars its fuel. Hollywood is already profiting by the innovation: Perry is now working on a fiery television commercial for Universal Studios's theme park. —*W. Wayt Gibbs*

ED BOCK *The Stock Market; computer manipulation courtesy of Rhythm & Hues, Inc.*

Setting a Standard

A British project produces a test for Alzheimer's disease

Every day for several months, my mother relived afresh the pain of learning that my father had died. When does a lapse of memory cease to be a trivial failure of our brain's capabilities and portend something far more serious? *I can remember when I was a sort of person.* At what stage does the loss of sharp recall rob us of speech, cut our links with the rest of humanity?

These are urgent questions for our aging society: estimates suggest there are at least four million with Alzheimer's disease in the U.S.; that number could grow fourfold over the next 30 years. For many, the questions are crucial because several reversible conditions—such as vitamin B_{12} deficiency, thyroid disorders and some forms of depression—resemble Alzheimer's.

A project called OPTIMA, undertaken at the Radcliffe Infirmary in Oxford, England, may now be poised to bring precision to the identification of the disease. Ever since Alois Alzheimer first described the condition in 1906, clinical diagnosis has depended on a psychiatric evaluation of the patient. Not surprisingly, there is disagreement over diagnoses, resulting in a failure to agree in up to a third of all cases. Even after death, when an autopsy can be done, pathologists debate the defining criteria of Alzheimer-type brain changes.

Researchers with the OPTIMA project, however, claim they can identify the condition in nearly all cases—and long before the patient dies. The study of more than 350 people, both healthy as well as those with memory deficits, began in 1988. Each subject spends one day a year at the hospital for clinical assessment and brain scanning; every six months his or her memory and cognitive skills are assessed. Of the 115 who have died, 110 have been autopsied.

The bottom line is that OPTIMA has demonstrated a way in which diagnostic accuracy appears to be improved from 65 to 97 percent and has simultaneously offered a mechanism for making a robust physical measurement of the disease's advance. These results emerged from sets of sequential brain scans. The site of most disturbance in Alzheimer's is the limbic system—a brain region critically involved with emotion, motivation and memory. An ordinary computed tomographic (CT) scan failed to reveal sufficient detail of pathology in this system, but when they angled the scan at 20 degrees along the plane of the limbic system, project leaders David Smith and Kim A. Jobst found a far better picture—and their first major insight.

They found that over a period of years, the size of the limbic system in Alzheimer's patients diminished catastrophically—thinning by as much as 15 percent a year, 10 times the rate seen in healthy people. When it was seen in apparently unaffected individuals, this thinning was predictive. Using the CT scan alone increased specificity to 93 percent. As these findings became clearer, the group began to look to other types of imaging.

Creating images of a brain at work has long been a dream of neuroscientists. During recent years, positron emission tomographic images of regions involved in reading or performing mathematical tasks have become icons of popular culture. The images show regions that are metabolically active or not. Using a similar, but more widely available imaging system—SPET (single photon emission tomography)—the scientists were able to make additional images from their subject's brains. These scans revealed which areas were working and which were "switched off."

Consistently in the Alzheimer's patients, the areas involved in language skills as well as visual and spatial skills appear to be less active. What was seen in the CT and SPET images was confirmed in the autopsies. Combining the results of both scans produces a diagnosis with a false positive rate of only 3 percent: the team seems to have arrived at a technique that can diagnose Alzheimer's disease at least five years before death. Jobst and his colleagues say they now want their methods to be tested by other groups.

The significance of the work lies partly in its sheer scale. OPTIMA has a unique database—one that may be critically useful when chemists and biologists find agents that might slow down or even reverse the progressive brain degeneration seen in this dehumanizing disease. *—David Paterson*

Employment Blues: Nothing to Do with Being Green

As the battle between jobs and the environment rages, at least one economist says he has reason to call a truce. Eban S. Goodstein of Skidmore College and the Economic Policy Institute in Washington, D.C., recently published his study tracking the number of jobs lost because of environmental legislation. Using U.S. Department of Labor statistics from 1987 through 1990, Goodstein found that for that period an average of only 0.1 percent of all larger-scale layoffs nationwide were the result of environmental regulations, such as the Clean Air Act—according to employers' own estimates. Changes in a company's ownership, in contrast, accounted for almost 35 times the number of jobs being terminated. *—Sasha Nemecek*

PROFILE: RUTH HUBBARD

Turning the Inside Out

Ruth Hubbard has not always stood outside science looking in. She was once on the inside looking around, a biologist exploring the visual system of frogs, cattle and squid. But as she sits in the sun-filled, green-hued closed porch of her home in Woods Hole, Mass., and revisits her past, it seems unlikely that she could have become anything other than a scrutinizer of science, an advocate and writer. Hubbard, initially famous for her biochemical forays, is now principally known for her work on women's health, on the position of women in academia and research and, more recently, on the tangled ethics of gene therapy and genetic testing.

Just as being a member of the club appears to have given Hubbard insight into the scientific establishment, so has being a woman and Jewish seemingly given her an outsider's perspective. From somewhere between those two worlds, Hubbard lobs what several of her colleagues describe as stinging criticism about how science fits, or does not fit, into society.

Hubbard was born in Vienna to two doctors. Her family was among the first to leave Austria several months after the Nazi invasion in 1938, when Hubbard was 14. With their possessions but no money, they settled in Boston, where her father reestablished his practice. Hubbard pursued medicine as an undergraduate at Radcliffe College, because "everyone around me was a doctor."

For a brief time she considered studying philosophy and physics. "No one said, 'Do not go into physics,' but in the physics course I took there were 350 men and I and another woman. So you know, there are messages," she says laughing, drawing one foot up under her. Hubbard moves and speaks with a certain languor, even though her voice is studied and strong and her opinions unmistakable. Her writing has the same steady directness—which has made her many books widely accessible. "Mind you, I am not pretending that I was astute enough to pick up the cues in an overt form," she continues. "If you had asked me why there were so few women in this physics course, I guess I would have said because they were not good enough. I did not have any feminist consciousness about these matters." That awareness was to come later.

Her college years coincided with World War II, and Hubbard wanted to do something for the Allied effort. She went to work on infrared vision with George Wald at Harvard University until she moved down to Tennessee for a short time, where her first husband, a G.I., had been stationed. She remembers their Chattanooga sojourn as bizarre. Soldiers began returning home, and there, in the Deep South, "people were wondering if there was going to be civil war now that black men had been taught to shoot white men."

Hubbard soon returned to the more familiar terrain of the Northeast, and in 1946 went back to Radcliffe to earn a doctorate in biology. She continued to work in Wald's laboratory, investigating vision. Specifically, Hubbard studied the architecture of visual pigments such as rhodopsin, a molecule that responds to light. "Everyone knew that vitamin A was involved, but we found that it came in different shapes and that only one of those can be used to form rhodopsin," she explains. "Then we found that what light, in fact, does is change the shape of visual pigments, and that initiates all the changes that lead to electrical charges" and, ultimately, to neurotransmission. (Wald—whom Hubbard later married—received the Nobel Prize in Medicine in 1967 for the laboratory's work on vision.)

Although Hubbard says she loved the research, her interests began to shift toward "the whole issue of social relevance that was part of the Vietnam War: What were we doing and why, and what good was it for anybody anyway?" She was further disconcerted by one, pivotal aspect of her studies. "At that point I was working with squid, and I think squid are the most beautiful animals in the world. And it just began to bother me. I began to have the feeling that nothing I could find out was worth killing another squid."

At about this time—after almost 20 years of vision-related experimentation—Hubbard recalls, "The women's movement sort of hit me over the head." In the late 1960s, she was asked to give a talk at an American Association for the Advancement of Science meeting about being a female scientist. For added material, she interviewed other female scientists and discovered that there were more than subtle similarities in their experience: "I don't know whether any of us, until that moment, realized that we were all accomplished and were all recognized in our fields but that none of us had real jobs." Every woman in what became an ongoing, informal discussion group was "off-ladder," that is, an associate, a lecturer, an assistant.

Hubbard simultaneously joined an organization at Harvard that petitioned the university to examine the status of faculty women. As a result, the university began, albeit slowly, to offer women tenure-track positions. In 1973 Hubbard became the first woman to receive tenure in the sciences at Harvard. She continued to lecture on photochemistry but soon added courses on health and women's issues—a combination she taught until she became professor emerita five years ago.

Despite the public attention that has been given to hurdles faced by women professionally, Hubbard claims the situation is not much improved. According to the National Science Foundation's most recent estimate, only 18.8 percent of employed U.S. scientists and engineers are women; although opinions vary greatly, many female researchers say they experience direct discrimination from male colleagues or more subtle dissuasion from fields such as physics.

Nor has the workplace been reconfigured so as to accommodate both sexes. "The focus is on this as a woman's problem rather than as a problem for society, which has a very limited view of the participation of men in family life. I mean it is always phrased in terms of how are women going to be able to structure work and family, as though this were not an issue for men at all," Hubbard observes.

Hubbard says she felt no compunction about devoting herself to her work—and raising two children. "I grew up in central Europe, in Austria, in a society in which women of my class were expected to be professionals and entrust the care of their children to other women. So I did not have any major emotional barriers to cross over what I was doing to my children," she states.

Hubbard has criticized science for excluding women and for being structured around a view of society determined only by European and American men. This notion has been explored by several feminist thinkers who question the culture of science—looking at how girls and boys are socialized differently from the outset, at the positions that men reach in the hierarchy as opposed to women. Their writings—Hubbard's *The Politics of Women's Biology,* among them—have argued that feminist theory enriches science by raising questions about point of view or bias. The resulting dialogue, according to Hubbard, can only make science more egalitarian.

This vision of equality is central to Hubbard's current preoccupation with genetics and molecular biology. She cautions that society is undergoing "genomania," oversimplifying science and assigning every trait, including behaviors, a genetic cause. Hubbard reviews the implications of this trend in *Exploding the Gene Myth,* which she coauthored with her son, Elijah Wald, and which was published two years ago, just before the genes-as-destiny argument of *The Bell Curve* swept the nation again. (As for the late Richard J. Herrnstein, who was also at Harvard, "he's saying the same things he was saying in the 1970s," Hubbard comments. "It's just warmed-over racism.")

RICARDO AZOURY *Saba*

RUTH HUBBARD scrutinizes science in society.

Hubbard argues that the search to identify all genes, including those for diseases—the $3-billion travail of the Human Genome Project—will necessitate ethical choices that society and particularly scientists are not confronting with enough energy. These quandaries include potential abuse by insurers, who may deny coverage because of genetic conditions, as well as the ability to craft custom-made embryos. "I suspect I came to this by virtue of being a Hitler refugee and being interested in eugenics and then in the revival of eugenics and the race and IQ debate in the 1960s and early 1970s," she muses.

Further, Hubbard maintains that science is being presented in terms that obscure the truth. By saying, for instance, that a gene has been found for breast cancer, researchers obfuscate interactions between many biological factors, including genes, and the environment. As she painstakingly explains, a gene is only a piece of DNA—it rarely represents a one-to-one link to a disease. Knowledge may not be helpful if a patient cannot do anything with it. "You are telling somebody that you have a greater than average tendency to get cancer, and somehow it is assumed that this prediction in itself somehow functions as prevention," she says.

The choices made as a result of such knowledge—however incomplete it may be—may revisit eugenics. One of her favorite examples is that of Huntington's disease: What if parents decide not to bring to term a fetus that tested positive for the disease, which often strikes late in life? Society would have been denied a talent such as Woody Guthrie, Hubbard warns. "I have gone out on a limb on this by saying that most people in our culture are very judgmental about women who terminate a pregnancy because of sex. How different is that from terminating a pregnancy because of Down syndrome?" she asks. And when people use economic justification: "It is very much like what the Nazis were doing when they decided it was cheaper to kill mentally ill people and retarded children than to care for them."

Some of Hubbard's fears about ethical considerations may be borne out. Although the Working Group on the Ethical, Legal and Social Implications of the Human Genome Project can grapple with complex issues in meetings, it has no enforcement powers, no means of establishing policy.

Hubbard, meanwhile, refuses to wait and see. And on her porch—which seems simultaneously indoors and out amid the overgrown, knotted wetland that surrounds the house—there is room to move. —*Marguerite Holloway*

Debt and the Environment

Loans cause great human hardship, but their connection to ecological troubles is hard to prove

by David Pearce, Neil Adger, David Maddison and Dominic Moran

BHUDEV BHAGAT *UNEP, courtesy of Select; computer manipulation by Laurie Grace*

Some environmentalists claim that loans to developing nations have led to a spiral of debt and environmental degradation. According to their argument, the domestic policy changes that countries make to generate cash for loan payments—often under duress from the International Monetary Fund or the World Bank—hasten the depletion of natural resources, increase pollution and harm the poor, who may be uprooted in ways that cause further environmental damage. Many critics also contend that lenders should write off the loans because the money, in any case, is effectively irretrievable and because relieving countries of repayment obligations will encourage "sustainable development."

Although economic theory does not automatically render repayment incompatible with full employment, steady prices, economic growth or an equitable distribution of income, in reality these goals have suffered. As a result, most debtor nations continue to rely on outside funds, even though additional loans only make their predicament sharper. Whether the environment has also been harmed directly is less clear. There is scant empirical evidence to suggest that the connection between debt and the environment is significant. Indeed, in some cases, the fiscal discipline imposed by debt may rein in environmentally harmful spending.

The debt crisis has its origins in the oil-price shock of 1973, when energy prices roughly doubled in a matter of months. Commercial banks, flush with deposits from oil producers, were eager to lend money to developing countries, especially as they took it as an article of faith that nations always repay their debts. The borrowers, meanwhile, were glad to see money plentifully available at low interest rates. In 1979, when oil prices doubled again, industrial nations raised interest rates to slow their economies and thus reduce inflation. This action spurred a global recession that stifled demand for the raw materials developing nations were producing.

As interest rates rose, debtor nations faced higher payments on their outstanding loans but had less income with which to pay. Many found themselves unable to meet their current obligations, much less get new loans. Repayment became an overriding policy objective, affecting both government and private spending, because only wide-ranging changes in developing economies could generate the needed hard currency. A large fraction of many countries' earnings continues to be earmarked for the repayment of debt.

Exports at Any Cost

The most commonly held view linking debt with environmental degradation is known as the exports promotion hypothesis. To earn foreign exchange with which to repay international debts, a country must divert resources away from production of domestic goods to sectors generating commodities for export. According to this theory, production of goods for export causes more environmental degradation than does production of goods for domestic consumption, and so debt repayment harms the environment. There is no a priori reason to expect such a difference, but some environmentalists contend that it still does occur. They point to the possibility, for example, that countries will raze their forests for tropical timbers or to open up land for cash crops.

Nevertheless, statistical analysis of data from many developing countries suggests that national income and commodity prices have just as much influence on levels of exports as debt does. Raymond Gullison of Princeton University and Elizabeth C. Losos, now at the Smithsonian Tropical Research Institute, examined the effect of debt on timber exports from Bolivia, Brazil, Chile, Costa Rica, Paraguay, Peru, Colombia, Ecuador and Mexico but found only minimal correlations overall. Furthermore, in Paraguay, the only country on which debt apparently did have a significant effect, increased debt was associated with reduced production.

More recently James R. Kahn of the University of Tennessee and Judith A. MacDonald of Lehigh University found more concrete evidence of a correlation between debt and deforestation, although they also found that country-specific factors played a strong role. They estimate that reducing a country's debt by $1 billion might cut annual deforestation by between 51 and 930 square kilometers. Brazil currently clears more than 25,000 square kilometers a year and Indonesia more than 6,000.

Deforestation is only one aspect of environmental degradation in developing countries. Unfortunately, the impact of indebtedness on other environmental indicators such as pollution, biodiversity or depletion of other resources has not been tested.

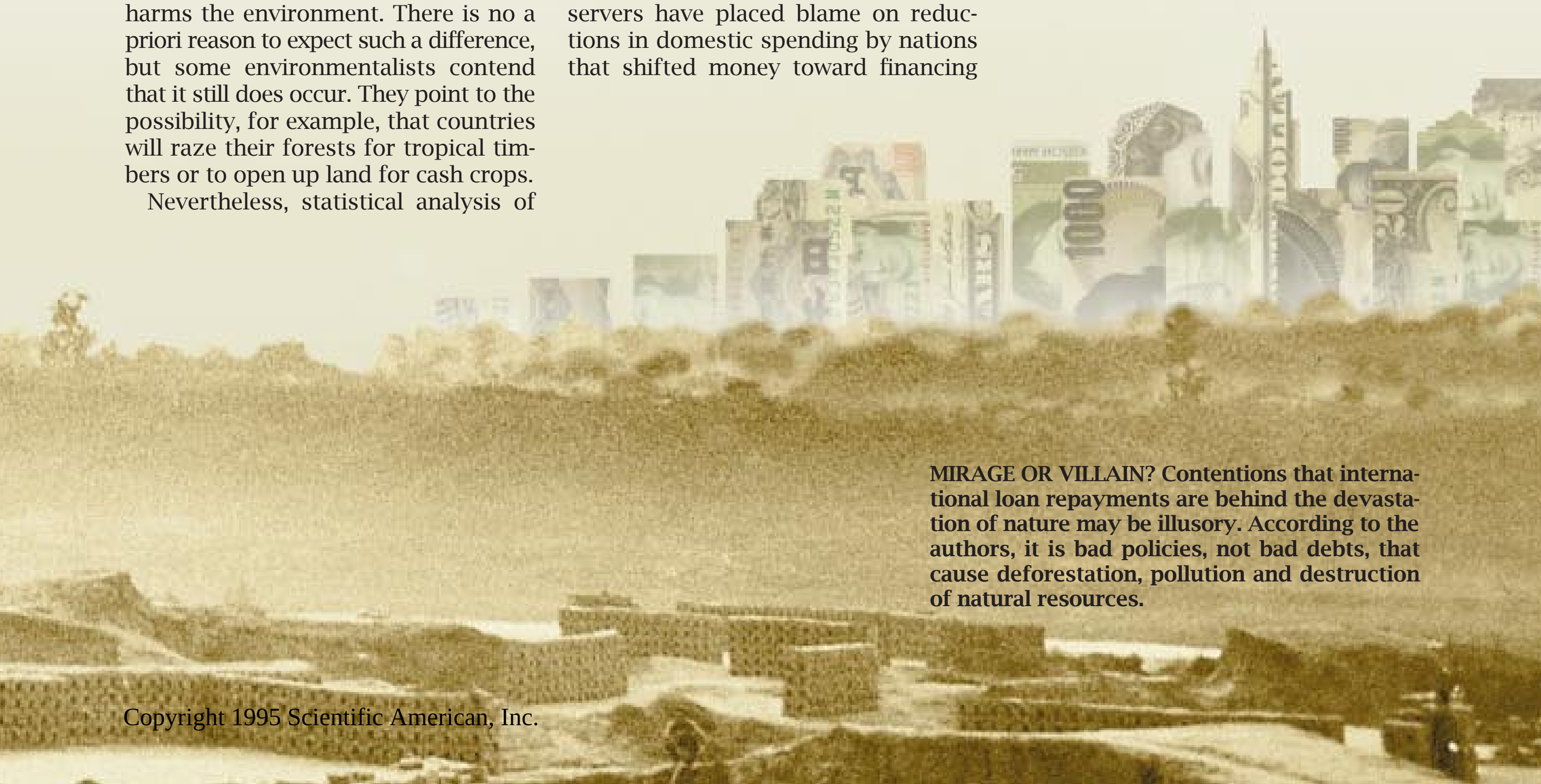
Reductions in Domestic Spending

Evidence for or against other mechanisms by which debt repayment might damage the environment is largely anecdotal and speculative. Some observers have placed blame on reductions in domestic spending by nations that shifted money toward financing their debt. In sub-Saharan Africa, outlays for health, education and other public services decreased by more than 40 percent during the 1980s, in parallel with a sharp rise in money spent on repayment of debt. Yet the effects on the environment remain unclear.

Cuts in government spending may fall on measures designed specifically to enhance the environment, such as schemes to improve water quality and sanitation. On the other hand, some cuts may cancel large capital projects, including the construction of dams and roads, that have often been criticized for causing environmental devastation far in excess of any financial returns they may bring. Elimination of road-building programs in the Brazilian Amazon, for instance, may have helped curtail deforestation. From a theoretical point of view, then, reductions in government spending can act to either improve or degrade the environment.

The same uncertainty emerges when one looks specifically at the effect of debt on spending for environmental protection. Some economists assert that environmental quality should increase with national income. (Their claim, called the Kuznets curve effect, grows out of the observation by Simon Kuznets, who won the Nobel Prize in Economics in 1971, that richer nations have more equitable income distributions.) Conversely, a nation faced with a huge debt is likely to divert money from the environment to more pressing problems.

But does the relaxing of environmental standards inevitably increase pollution? Perhaps not. Many debtor nations



MIRAGE OR VILLAIN? Contentions that international loan repayments are behind the devastation of nature may be illusory. According to the authors, it is bad policies, not bad debts, that cause deforestation, pollution and destruction of natural resources.

have probably always spent little on protecting the environment, and so decreases in spending could have a minimal impact. Moreover, a country repaying debt might not be able to afford high standards for pollution control, but it also might be unable to afford goods whose production damages the environment.

It is conceivable, however, that absent or unenforced environmental regulations or their lax enforcement might also make for the establishment of “pollution havens”—a situation that many claim has occurred in the Maquiladora export-processing zone of northern Mexico. Supposedly, U.S. companies have been attracted to this area because of the lower environmental standards. Yet a 1992 study by Gene M. Grossman and Alan B. Krueger of Princeton found little statistical support for this claim. They argue that low wages and easy access to U.S. markets spurred investment. Thus, the net effect that reduced domestic spending has on pollution and resource degradation is not clear.

In addition to causing ecologically unsound export drives or budget cuts, excessive debt can potentially exacerbate local practices that already put the environment at risk. Diversion of money to debt service has triggered massive unemployment in many countries, sometimes prompting poor people to migrate in search either of work or land on which to grow food to live. Marginal lands and fisheries whose ownership was indeterminate have often attracted migrants, until the topsoil or fish have been depleted.

Overexploitation would have come about even without these evils, however; the inefficiency results from the lack of established rules for access to land, water and other resources. Many indebted countries seem to be characterized by the hallmarks of shared resources, a high level of migratory subsistence farming, overgrazing and declining yields. Ill-defined or nonexistent ownership has often precipitated the exhaustion of land. The enforcement of property rights would in many cases have prevented overuse by giving people incentives for proper husbandry.

DAVID PEARCE, NEIL ADGER, DAVID MADDISON and DOMINIC MORAN investigate the interplay between economic forces and the environment. Pearce is director of the Center for Social and Economic Research on the Global Environment (CSERGE) at University College London and professor of economics at the university. He has headed several international working groups on economics and the environment. Among his hobbies are porcelain collecting and birdwatching. Adger is a senior research associate at CSERGE at the University of East Anglia. Maddison, a research fellow at CSERGE, teaches at University College and conducts research on the socioeconomic effects of climatic change. Moran, a research associate at CSERGE, investigates the valuation of environmental assets and the economics of biodiversity.

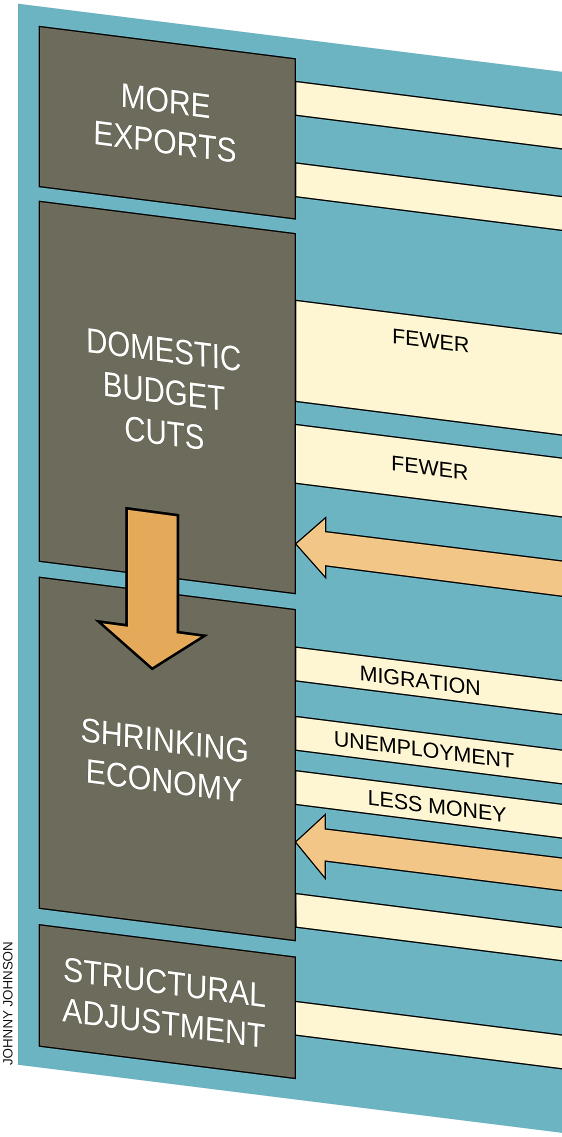
Structural Adjustment

Nations that face increasing poverty are often forced to try to secure additional loans. As a condition of attaining this money, heavily indebted countries have often had to make “structural adjustments” to their economies: eliminating subsidies, removing tariffs and privatizing government-owned enterprises. These reforms aim to help them grow out of indebtedness by removing glaring economic inefficiencies.

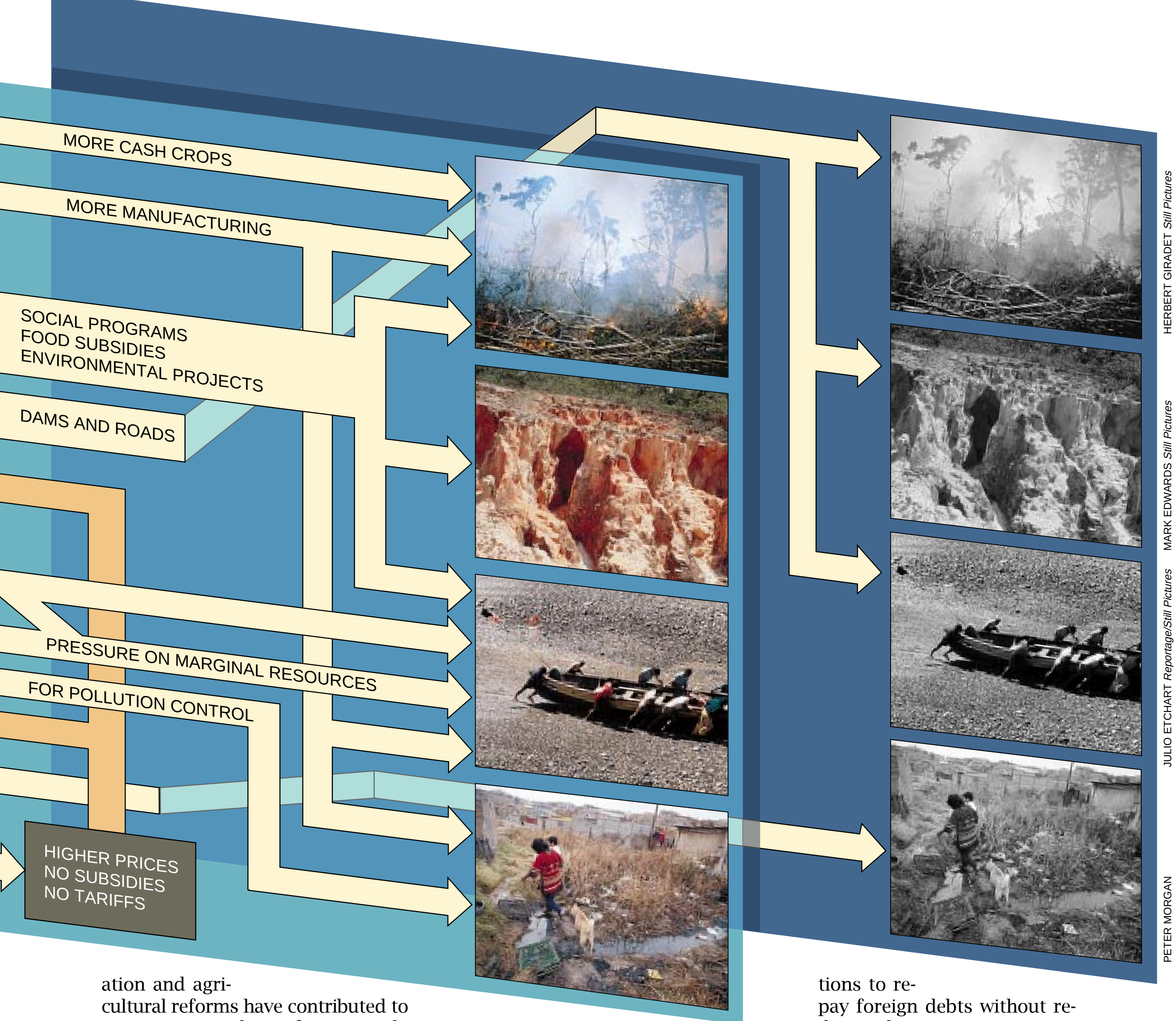
Some observers contend that these “conditionality” programs let governments push ahead with policy changes that were previously impossible, by putting multilateral institutions in the position of political lightning rods. Others think conditionality makes all objectives, including environmental ones, subordinate to debt repayment. Systematic cuts in such vital services as education, health and food (which would typically be higher on a government’s list of priorities than the environment) give some weight to the latter view. If it is correct, one would expect to find environmental repercussions in countries that have received significant structural adjustment loans.

The evidence is equivocal. Case studies of conditionality programs of the International Monetary Fund in Mexico, Ivory Coast and Thailand, sponsored by the World Wildlife Fund International, suggest that structural adjustment programs have on balance benefited the environment. In Thailand the removal of indirect irrigation subsidies has helped reduce waterlogging and salinization. Yet in Malawi, which has undertaken four International Monetary Fund restructuring packages and six World Bank structural adjustment loans since 1979, the Overseas Development Institute found many more negative results than positive ones. In the Philippines, a study by the World Resources Institute found that structural adjustments undertaken in the 1980s encouraged overexploitation of natural resources and resulted in increased emissions of pollutants, concentrated pollution and congestion in urban areas.

Whether the consequences for the environment are good or bad seems to depend a great deal on the particular provisions of the loan agreement and on the individual circumstances of a country or even a region within a country. In Ivory Coast, controls on food prices and subsidies for fertilizers and pesticides have been minimized. These changes could cause the abandonment of some environmentally harmful farming practices, but lower yields may also bring about cultivation of remaining forestland. In Malawi, currency devalu-



PURPORTED LINKS between debt and environmental damage (*front panel*) are uncertain. Efforts to increase exports could lead to deforestation and pollution (*top*), for example, but statistical analyses have yet to uncover strong correlations. Government spending cuts may also have good environmental effects (*rear panel*) by stopping destructive trends, depending on the programs that are cut. The same is true of the economic contraction that generally accompanies repayment. Evidence on the ultimate consequences of various market-freeing measures that go by the name of “structural adjustment” is also mixed.



ation and agricultural reforms have contributed to an increase in tobacco farming and created incentives for planting such crops as cotton and hybrid maize, which tend to be grown in a manner that promotes erosion. Malawi's case cannot be taken entirely at face value, however: large-scale migration of refugees from Mozambique has placed an unprecedented strain on the country's resources, unrelated to structural reform.

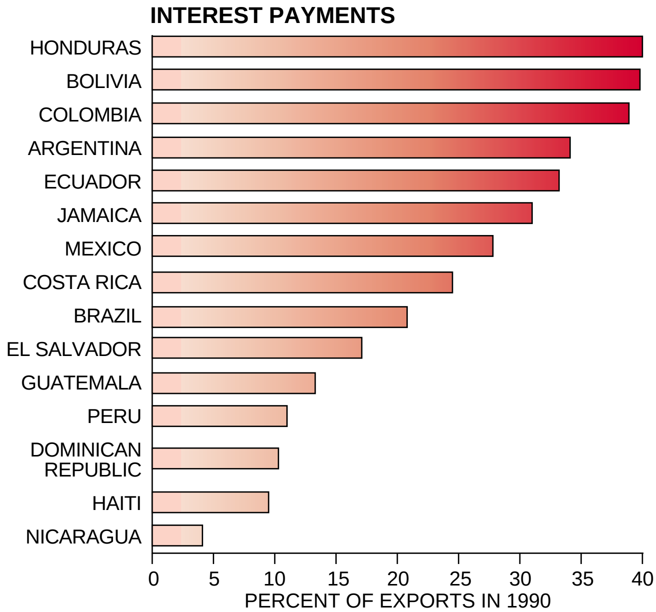
In other parts of the world, structural adjustment appears to have helped the environment. Agricultural subsidies, which adjustment programs curtail, have played a significant role in deforestation and the destruction of soil in areas subject to erosion. Until the late 1980s, for example, Brazil gave tax credits for cutting down forests, and it subsidized loans for crops and livestock development. These government incentives typically covered more than two thirds of the cost of cattle ranches, which reportedly accounted for 72 percent of all deforestation in the Brazilian Amazon up to 1980. Toward the end of the 1980s, after the government abandoned many of the clearance subsidies, deforestation slowed. Other evidence from Brazilian Amazonia also supports the idea that public expenditure on infrastructure, such as on roads, is linked to migration onto marginal land that results in deforestation, erosion and other deleterious effects.

Reduction of subsidies for energy use could also have a salutary effect on the environments of developing nations. Anwar M. Shah and Bjorn K. Larsen of the World Bank have calculated that total world energy subsidies amounted to $230 billion in 1990. Eliminating them could cut emissions of carbon dioxide (the main greenhouse gas) by 9.5 percent and improve prospects for economic growth by freeing the money for other uses. Indeed, the developing world has many resources that could be used more efficiently, thereby enabling nations to repay foreign debts without reducing domestic consumption.

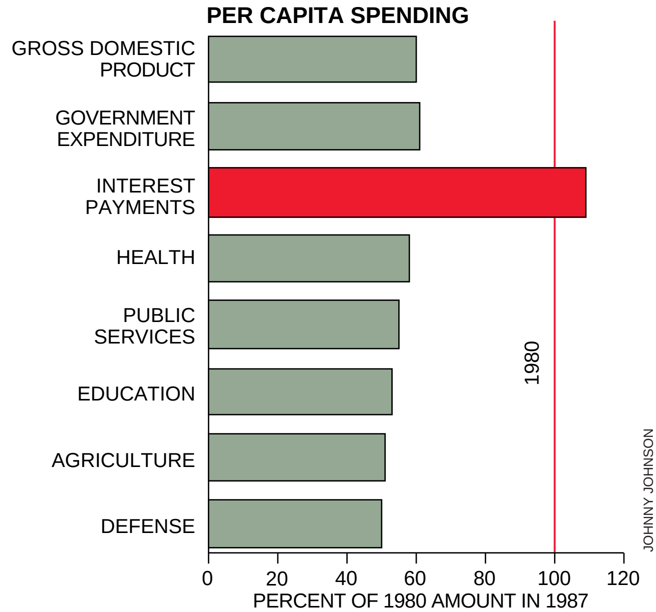
Structural adjustment programs can potentially prove beneficial in another way as well. Recent structural loans have been made on the condition that governments clarify land-ownership questions. Such arrangements should reduce somewhat the environmental degradation brought about by shared use of land that seems to belong to anybody and to nobody in particular.

Debt Forgiveness

At best, then, evidence on the extent to which environmental degradation in the developing world can be attributed to debt repayment remains inconclusive. Nevertheless, given the fact that debt repayment is largely to blame for the drastic reductions in per capita spending on social programs in these countries, cannot a strong case be made in favor of debt forgiveness? In secondary markets, debt of developing coun-



SOURCE: The World Development Report 1992, *World Bank, Oxford University Press*



SOURCE: Africa's Recovery in the 1990's, from Stagnation and Adjustment to Human Development, *by G. A. Cornia et al., St. Martin's Press, 1993*

SCOPE OF DEBT PROBLEMS can be seen in data from Latin America (*left*) and Africa (*right*). Interest payments on Latin American debt consume as much as 40 percent of nations' earnings, leaving less money to pay for needed imports. Nations in sub-Saharan Africa have slashed all government expenditures except for debt, leaving their citizens considerably worse off.

tries often changes hands at a fraction of its contractual value. This steep discount is an indication that the commercial banks expect that the loans will never be repaid in full.

Banks do not simply write off the uncollectible debt, for two simple reasons. First is the problem of moral hazard: forgiveness may encourage countries to get into more debt in the expectation that it, too, would be forgiven. Such profligacy would obviously jeopardize other assets held by the banks. The second reason is uncertainty. There is a slight possibility that unexpected favorable developments will eventually enable developing countries to repay more of their debts, and so it is not in the interest of any bank to deprive itself of the opportunity of benefiting from any windfall to the borrower.

Conversely, a reduction of part of the contractual debt that the debtor is not expected to repay anyway is of little value. It neither reduces current cash requirements nor makes it easier to get new loans. When Bolivia spent $34 million to buy back $308 million in bonds in 1988, the price of the remaining bonds rose from six to 11 cents on the dollar. As a result, the real value of outstanding debt declined from $40.2 million ($670 million at six cents on the dollar) to $39.8 million (362 million at 11 cents), less than $400,000.

The same problem afflicts other debt-reduction mechanisms, such as debt-for-nature swaps. Until 1992, 17 countries had participated in such swaps: donors spent $16 million to retire nearly $100 million of debts in developing countries in return for the establishment of national parks and other environmental improvements. Although the swaps do preserve some environmentally vulnerable regions, the large nominal reduction barely touches nations' real burdens. Indeed, they may even have increased expected repayments.

No Easy Path

It seems, then, that excessive debt inevitably causes radical restructuring of a nation's economy. Because economic policies play a crucial part in determining how natural resources are used, the environment is bound to be affected. Yet it is very difficult to predict whether any particular change will cause harm or prevent it.

Such links as exist between indebtedness and damage to the environment stem largely from structural adjustment programs and their requirement that money be reallocated from government subsidies and other spending to repayment. Even then, most environmental degradation in the developing world probably has causes other than the servicing of debt. Structural adjustment is more justly criticized on humanitarian grounds than on environmental ones. Instead of alleviating unemployment and equitably redistributing income, price reforms—particularly elimination of subsidies for food and fuel—have fallen most heavily on the poor.

Given that the connection between debt repayment and environmental degradation is tenuous at best, attempting to improve the environment by debt forgiveness would probably be futile. The most effective way to confront pollution, deforestation and similar problems in debtor nations is to establish individual ownership of resources that are currently open to all, to end environmentally damaging subsidies, to institute market-based pollution-control mechanisms (in which those who produce toxic substances pay for their effects) and to make direct payments where necessary to preserve environmental assets of global significance. Undoubtedly, the flow of funds from the South to the North causes poverty, malnourishment, ill health and lack of educational opportunity. These consequences make a compelling case for debt relief. But such aid is not a panacea for environmental degradation.

FURTHER READING

A FATE WORSE THAN DEBT. Susan George. Penguin Books, 1989.

THE ENVIRONMENTAL EFFECTS OF STABILIZATION AND STRUCTURAL ADJUSTMENT PROGRAMMES: THE PHILIPPINES CASE. W. Cruz and R. Repetto. World Resources Institute, 1992.

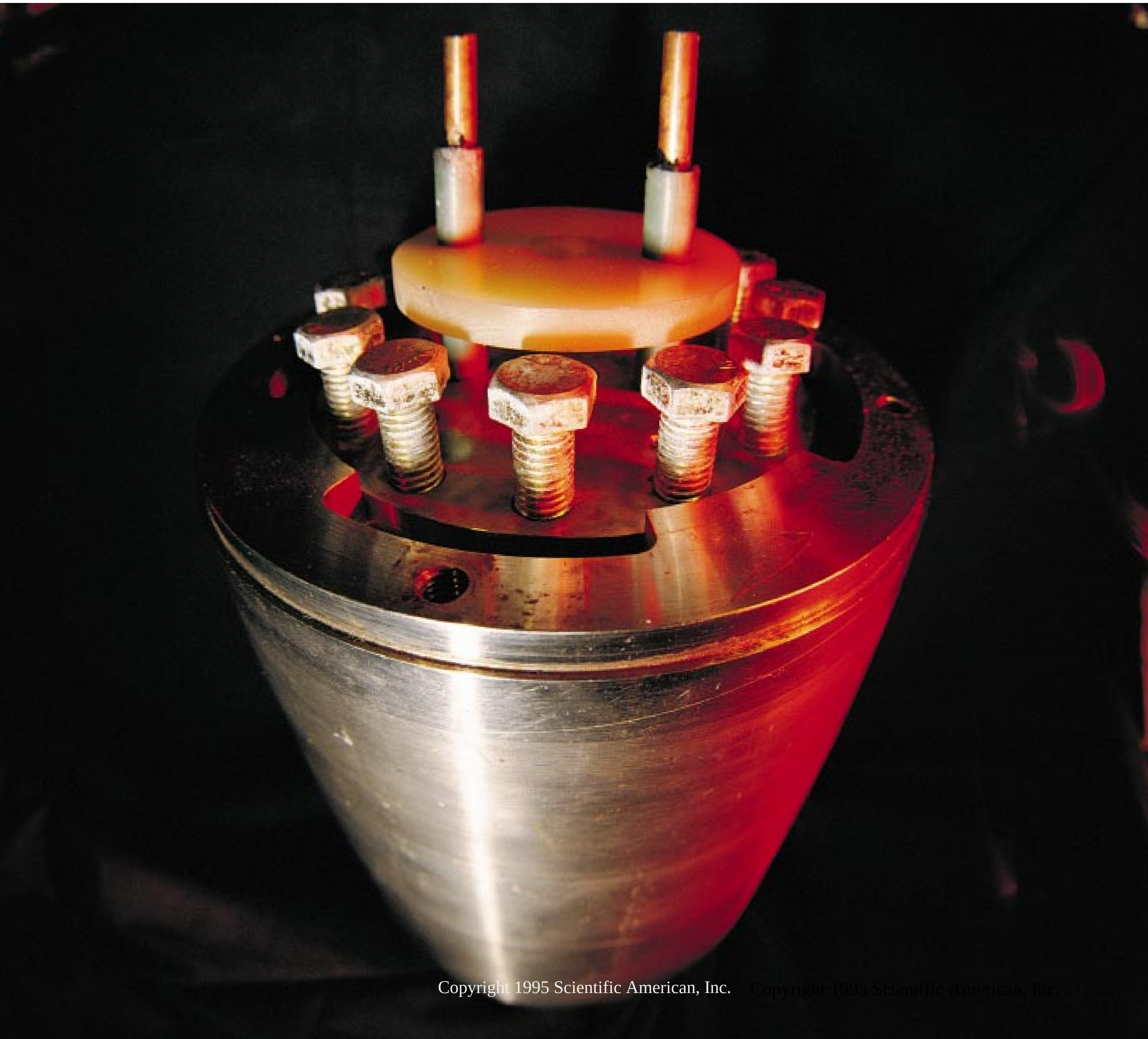
STRUCTURAL ADJUSTMENT AND THE ENVIRONMENT. D. Reed. World Wide Fund for Nature/Westview Press, 1992.

AFRICA'S RECOVERY IN THE 1990'S, FROM STAGNATION AND ADJUSTMENT TO HUMAN DEVELOPMENT. G. A. Cornia, R. van der Hoeven and T. Mkandawire. St. Martin's Press, 1993.

Building World-Record Magnets

Packing the energy equivalent of a stick of dynamite, powerful electromagnets around the globe compete to advance our knowledge of materials science and physics

by Greg Boebinger, Al Passner and Joze Bevk



Copyright 1995 Scientific American, Inc.

December 3, 1992, began routinely enough. For several months we had been studying how an intense magnetic field nullifies superconductivity—the complete absence of electrical resistance in certain materials. Our high-strength electromagnet, designed and built 11 months earlier, had generated thousands of magnetic-field pulses. Each had produced a field more than a million times stronger than the earth's and had concentrated a burst of energy comparable to an exploding stick of dynamite into a volume the size of a fist.

As was our custom, we submerged the electromagnet in liquid nitrogen, to reduce the electrical resistance in its coiled wires. The experimental sample, one of the early high-temperature superconductors, was positioned in the center of the magnet. We shut and locked the door to the steel bunker enclosing the magnet, its energy supply and all our data-collection equipment. An arming and charging sequence energized the power supply to 7,600 volts. One of us pressed the "FIRE" button, and the routine ended abruptly.

A gunshotlike sound and icy streams of supercooled nitrogen blasting out of the vents of the bunker were sure signs of a catastrophic magnet failure.

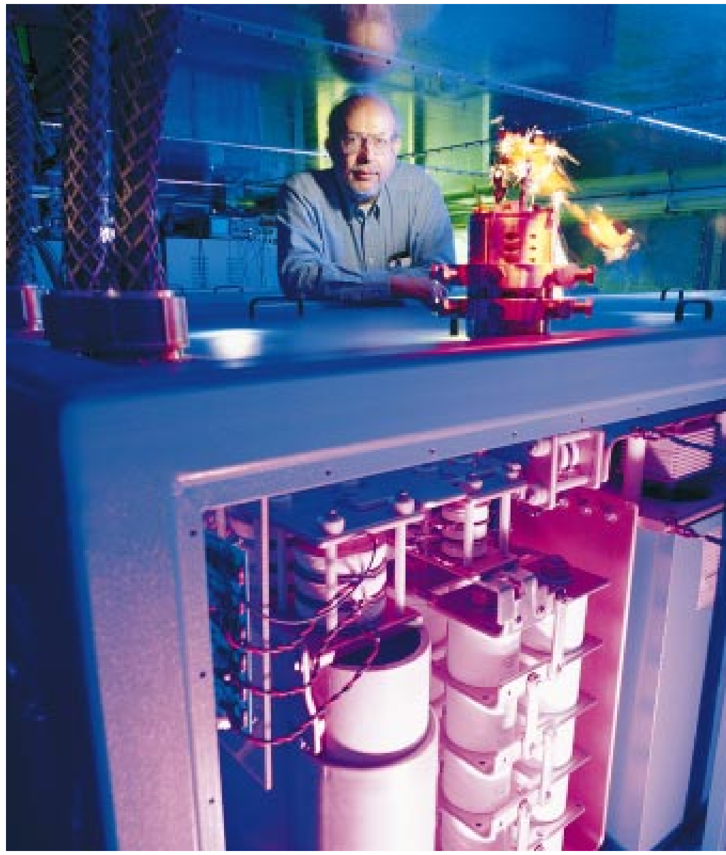
Once we regained our composure, we entered the bunker to survey the damage. All things considered, it could have been worse. In 1988, while testing our first pulsed-magnet designs with a power supply and laboratory graciously made available by a colleague in Belgium, we repaid our host's hospitality with a magnet failure that hurled hunks of steel violently around the room. This time the steel casing around the magnet had remained intact, although mechanical forces generated by the field had snapped the eight steel bolts holding the magnet in place. These forces then lifted and flipped the 60-pound magnet, which was a little larger than a gallon-size paint can, destroying some nearby equipment and bending a half-inch-thick aluminum floor plate. We never did find the sample we had been studying.

Why do we, and our colleagues in a couple of dozen laboratories around the world, put ourselves through such trouble? The race for stronger electromagnets is a significant challenge. It motivates the development of novel materials, conductors and insulators alike, which extend the limits in strength, ductility and electrical performance. Furthermore, the creation of these magnets is actually the means, not the goal, of research in many laboratories. The generation of extremely strong magnetic fields has many purposes, ranging from developing more powerful permanent magnets to exploring the complex behavior of electrons in advanced materials—each a new manifestation of electromagnetism itself.

Permanent magnets are a key component of many electric motors and audio speakers, and improvements to these magnets have permitted miniaturization and portability of these products. Electric motors come in many types, but they all derive mechanical motion from the interactions of electric currents and magnetic fields. Stronger permanent magnets lead to smaller, stronger and lighter-weight motors, which are particularly important for applications demanding battery power and portability, such as mobile robots and laptop-computer disk drives. The controlled interplay between permanent and electromagnets is also critical to audio speakers; about 10 years ago the sudden appearance of small, lightweight and high-fidelity headphones for small personal stereo systems resulted from the commercial development of the more powerful samarium-cobalt magnets.

Such permanent magnets are known as hard magnetic materials, and new or experimental examples of them are routinely tested with very strong pulsed electromagnets. This kind of testing, in which the materials are exposed to the pulsed fields, tells researchers how intensely and tenaciously the novel materials can be magnetized. The term "hard" distinguishes them from "soft" magnetic materials, which easily change their magnetization and have found wide application in such products as cassette tapes, computer hard disks and floppy disks.

Powerful magnets also have more esoteric roles, such as the levitation and propulsion of high-speed trains and the launching of projectiles using pulsed magnetic fields. In experimental nuclear fusion reactors, strong pulsed magnets contain the plasma undergoing fusion because the plasma is too hot to be held by any solid vessel. Some of the most fascinating applications of pulsed electromagnets involve their use as experimental tools for performing sensitive physics experiments in a forbidding environment. To describe them adequately, however, requires a little background and historical perspective.



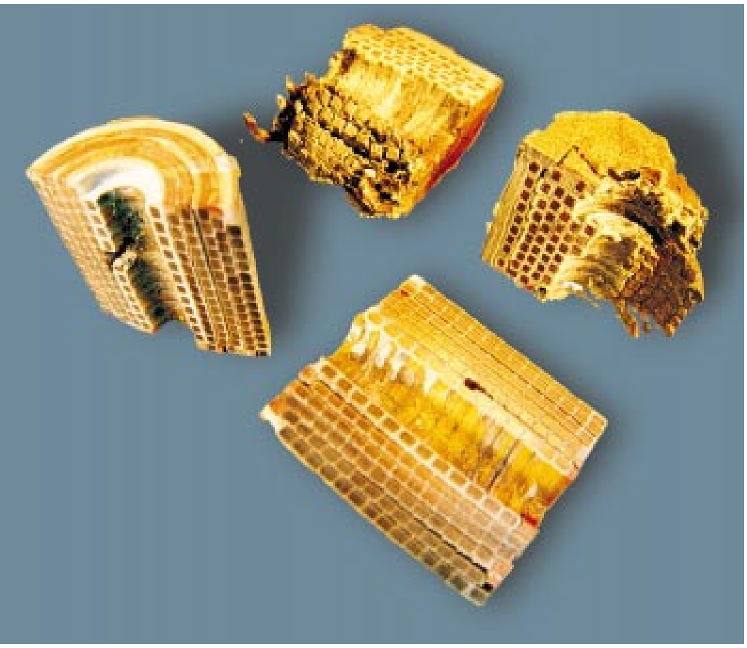
CHIP SIMONS

HIGH-ENERGY CAPACITORS fill this room and surround researcher Al Passner. The 24 capacitors (one of which is the vertical bluish box visible at the lower right) store the half a million watt-seconds of energy that are injected into an electromagnet in a single pulse lasting less than a tenth of a second. The electromagnet to the right of Passner is dwarfed by its energy supply. When pulsed, it is in a liquid nitrogen bath in a room above the one shown here. The magnet is connected to the capacitors through the heavy black cables (*upper left*). The electromagnet pictured above is similar to the one that exploded memorably in Belgium in 1988.

PULSED ELECTROMAGNET consists of a coil of high-strength wire inside a housing of reinforcing steel. This magnet generates fields approaching three quarters of a million gauss in a one-cubic-centimeter volume at its core. The bolts allow restraining pressure to be applied to the coil; the two copper rods are the electrical contacts to the coil itself.

CHIP SIMONS

Like gravity, magnetism is a part of everyday experience, as close by for many people as their refrigerator. Iron-based permanent magnets hold in place their minigalleries of photographs, recipes and children's artwork. Another familiar magnet is the needle of a compass, which is constantly aligned by the earth's own weak magnetic field.



CHIP SIMONS

TEST-COIL FRAGMENTS are from small, six-layer coils used to try out new conductors and analyze electromagnet failures. The two fragments at the lower left were cut neatly by a saw; the others were first torn open violently by the magnetic forces in a catastrophic failure.

Permanent magnets are a macroscopic manifestation of the minuscule magnetic field that accompanies each electron—its "spin." No one has established that an electron is actually spinning; the prosaic terminology acknowledges that if the electron were a small sphere of negative electrical charge, it would have to rotate to generate its observed magnetic field. Although all materials contain about 10^{24} electrons per cubic centimeter, in most materials the electron spins point in random directions, and the magnetic fields therefore cancel one another. In permanent magnets, on the other hand, the electron spins are aligned—typically 1 to 10 percent of them—within small regions called magnetic domains. Each domain acts as a single microscopic permanent magnet, established by the fields of the many individual electrons. With the right materials, appropriate processing and some luck, the fields of each of these domains can be made to align and resist change, creating forceful permanent magnets. The route to stronger permanent magnets lies in aligning as many of the electron spins as possible.

Although awareness of magnetism goes back more than 2,000 years to the Greeks and Romans, the realization that electricity and magnetism are two components of a single force—electromagnetism—is only about 175 years old. In 1821 the French physicist André-Marie Ampère established that magnetism is caused by electrical charges in motion. He recognized that a coil of wire carrying an electric current produces a magnetic field that emerges from one end of the coil and enters the other, just like the magnetic field of a permanent magnet. In an electromagnet of this kind, the field is strongest in the center of the coil.

By 1825 Ampère and Michael Faraday, the English physicist, had separately investigated the mechanical forces experienced by current-carrying wires in a magnetic field. These forces are exploited in all electric motors, from the huge one in a locomotive to the tiny one that turns the hands of a quartz watch. They also destroyed our magnet on December 3, 1992.

Not only do moving electrons give rise to magnetic fields, but, as Ampère and Faraday first discovered, magnetic fields exert forces on moving electrical charges. In most samples we study, these are the electrons, which in metals move about freely and in insulators travel in confined orbits centered on a given atomic nucleus. Magnetic fields interact with both these types of orbital electron motion. Moreover, an external magnetic field causes the spins of the electrons to align. Thus, magnetic fields interact with the electrons' orbital motion and spin in a material under study.

These principles make strong magnets very useful as experimental tools. Superconductivity, for example, results from the pairing of electrons with spins aligned in opposite directions. These pairs of electrons, held together by a certain binding energy, travel through the superconductor without encountering resistance. A sufficiently high magnetic field can inject enough energy to sever this binding, destroying the superconductivity. This phenomenon is helpful to researchers, who often wish to study the low-temperature behavior of these materials in the absence of superconductivity. Because many of the high-temperature superconductors remain superconducting even in very intense magnetic fields, pulsed magnets, with their extremely strong fields, offer the only opportunity to perform these types of experiments.

The more intense the applied magnetic field, the more energetic is its probing of electronic behavior. Certain electronic phenomena can be induced only by extraordinarily intense magnetic fields. At these intensities, furthermore, the effects on electrical behavior can be dramatic—the suppression of superconductivity or conversion of an insulator into a metal, for example. Such a transformation can be quite sudden, once the energy of the probe becomes comparable to, or resonant with, some specific characteristic energy in the material being probed, such as the binding energy in the superconductor.

Another example occurs in semiconductor physics. As the dimensions of devices approach 0.1 micron, the electrons within them become so confined that they can no longer carry an arbitrary amount of energy. Like the electrons in an atom, they are restricted to discrete energy levels that can be shifted or split into multiple levels by a magnetic field. Researchers measure the gaps between these energy levels by applying intense magnetic fields and observing their effect on some physical characteristic of the material, oftentimes electrical resistance or light absorption. A sudden increase in absorption would indicate that a resonance has been achieved between the changing gaps separating electron energy levels and the energy of the optical probe. Out of magnetic-field experiments has come better understanding of the behavior of small semiconducting devices, as well as, on occasion, discovery of an entirely new magnetic-field phenomenon.

Pushing the Limits

So how intense is a strong magnetic field? The unit of magnetic-field strength is the gauss, and as a benchmark, the strength of the earth's magnetic field is about one-half gauss. Iron-based refrigerator magnets, of the kind that hold up recipes, are a few hundred gauss. The most powerful permanent magnets, such as the samarium-cobalt or neodymium-iron-boron magnets, have fields of 3,000 to 4,000 gauss, and several of them could easily lift the entire refrigerator.

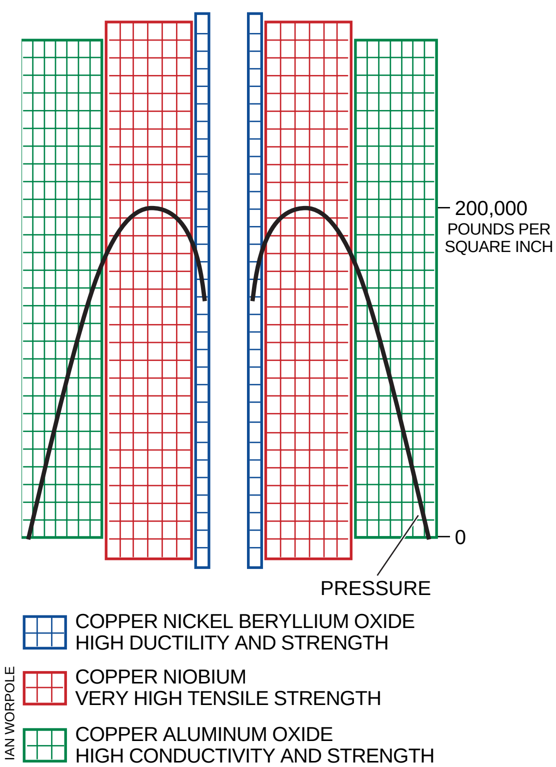
Although the future will surely bring improvements in permanent magnets, their ultimate strength is probably limited to around 30,000 gauss, simply because there is a limit to the density of electrons whose spins can be aligned. Making a permanent magnet physically larger increases primarily the extent, rather than the intensity, of the field. Thus, the most intense fields are produced by electromagnets, whose magnetism is a simple consequence of moving charges.

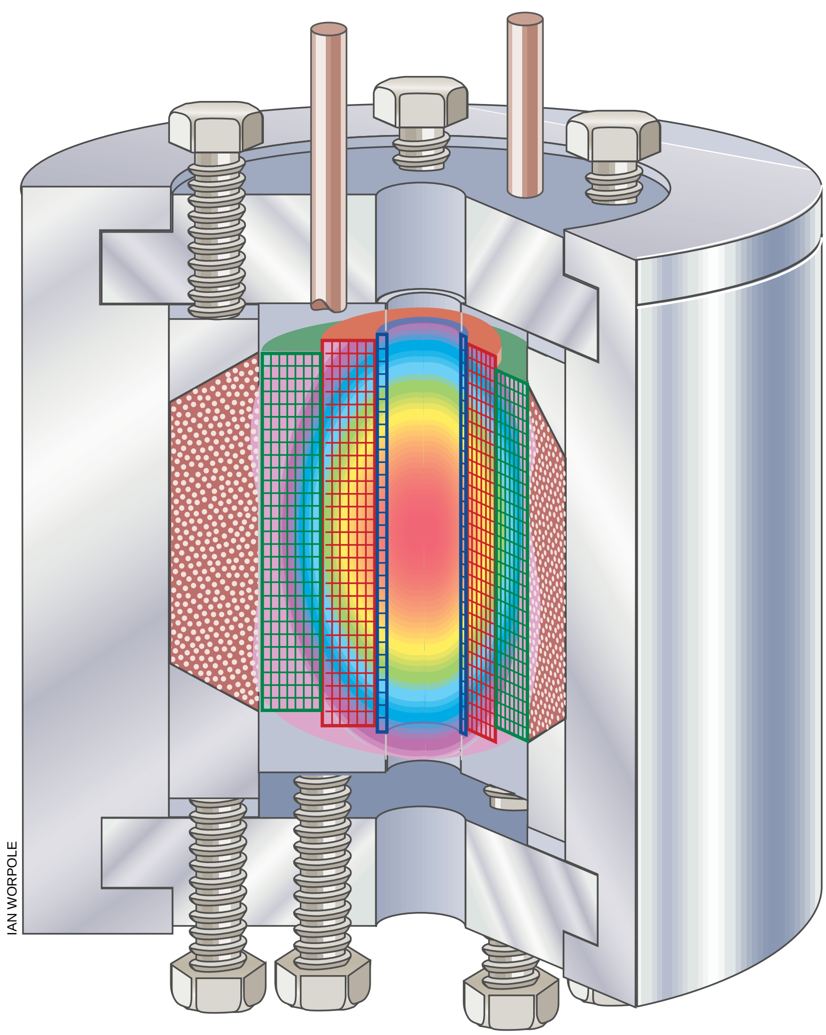
In our laboratory we have achieved magnetic fields of 730,000 gauss. (To put this effort in humbling perspective: nature provides a magnetic field of 10^{12} gauss at the surface of neutron stars.) Generating fields of this intensity in the laboratory requires an electric current pulse in an electromagnet that exceeds, albeit momentarily, the amount of current flowing through 15,000 100-watt lightbulbs. Because the magnetic forces on current-carrying wires are proportional to the product of the electric current and the strength of the magnetic field, our wires are subjected to an explosive pressure exceeding 200,000 pounds per square inch (14,000 kilograms per square centimeter)—more than 35 times the pressure on the ocean floor under four kilometers of water. Despite these challenges, the magnets generally survive thousands of pulses, although the occasional explosive failure is not without a certain pyrotechnic charm.

In principle, the field that an electromagnet can generate is unlimited: infinite current would produce an infinitely intense magnetic field. In practice, nature is not so accommodating. As the fields exceed half a million gauss, the forces imposed on the current-carrying wire surpass the tensile strength of hardened copper. Conductor strength usually limits the achievable field, making magnet building a hotbed of development and testing of new, high-strength conductors.

The mechanical force generated by the magnetic field is not the only obstacle, either. Tremendous heat is generated in the wire by the electric current required to produce the field. All materials, other than superconductors, resist the flow of current. This resistance converts some electrical energy into heat, which, in one of our magnets, can exceed 10 million watts—enough to melt its five kilograms of copper in less than a second.

World-record electromagnet designs take different approaches to the problems of heat and stress. They are divided into two broad classes: DC magnets (driven by direct, or nonoscillating, current) and pulsed magnets (energized by a short pulse of current). Pulsed magnets, which can be further divided into destructive and nondestructive designs, seek to avoid the problem of ex-





MAGNETIC-FIELD INTENSITY, depicted as a rainbow (*right*), is highest in the red region at the center of the electromagnet's wire coil, where experimental samples are placed. The trapezoidal cross sections surrounding the coil contain small beads of zirconium oxide, which support the coil while allowing liquid nitrogen to cool it. Finally, the entire structure is contained in a steel vessel and pressurized by tightening the bolts. Stress on the magnet wire (*plotted, above, on a cross section of the coil*) is highest in the middle layers of the coil, where the strongest conductor, copper niobium wire (*shown in red*), is used.

Toward a Megagauss

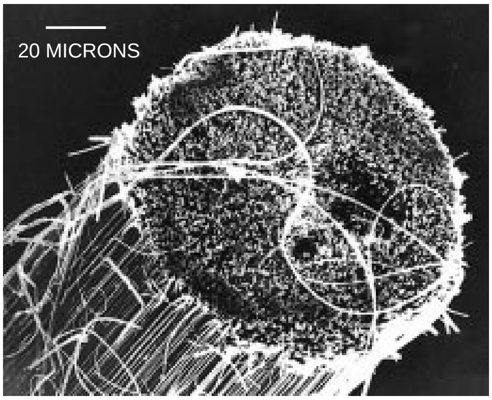
Generation of ever higher pulsed magnetic fields raises a few obvious questions. How high can we go? What are the material limitations, and is there enough room for improvement to justify further development? Answers to these questions depend entirely on materials science—specifically, how materials deform and ultimately fail under stress, how they can be made stronger and how mechanical and electrical properties interact.

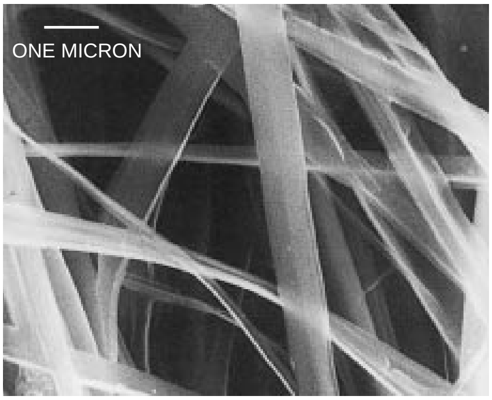
Consider a copper wire being pulled apart. It should break when the stress exceeds the product of the strength of an individual atomic bond and the number of bonds per unit area. This calculation yields a theoretical strength for copper of 350,000 pounds per square inch. Yet we know that copper deforms and breaks at much lower stresses. Indeed, pure metals often start deforming at less than 1 percent of their theoretical strength.

What accounts for such puzzling behavior? In a word, dislocations. Discovered some 60 years ago, dislocations are long arrays, or rows, of defects in which the atoms are not arranged in perfect, orderly fashion as elsewhere in the crystalline lattice [*see illustration at far right*]. Plastic deformation of all crystalline solids, including metals, is linked to the motion of these dislocations through the lattice. It is important to note that not all atomic bonds in a plane need be broken simultaneously for dislocation motion to occur. Because of the dislocations, the material can deform through the breaking and re-forming of a single row of atomic bonds along the dislocation line, in a process analogous to moving a big, heavy carpet by propagating a ripple from one edge of the carpet to the other. This kind of deformation requires much less energy than does the shearing of a dislocation-free crystal.

Most successful attempts to increase the mechanical strength of crystalline materials have centered on impeding the motion of dislocations, by embedding immobile defects in the crystalline lattice. We can introduce such defects into the crystal in various ways: by mixing in a second element to form an alloy, by introducing precipitates, such as small aluminum oxide or beryllium oxide particles, or even by generating such a tangle of dislocations that they interfere with one another's motion. In general, strengthening becomes more efficient as the density of the obstacles increases.

Unfortunately, for our purposes, an increasing density of these obstacles interferes not only with the motion of dislocations but also with the motion of electrons, causing them to scatter and conduct electricity less efficiently. A great density of defects also inhibits its ductility, which requires substantial movement of dislocations. Thus, in general, the greater the increase in strength, the greater the loss in conductivity and ductility.





TANGLE of ultrafine niobium filaments embedded in copper increases strength 10-fold.

For some time now, researchers have realized that a weak material, such as epoxy, can be reinforced by embedding tough filaments, such as fibers of glass, graphite or boron, to form a composite material. These old-style composites derive their strength from the filaments. In recent years a new family of high-strength composites has appeared, the "in-situ formed" conducting composites. These new materials feature much more closely spaced and much finer filaments [*see photographs above*]. These tiny ribbons are under a millionth of a centimeter in thickness (more than 5,000 times thinner than a human hair). Because dislocation motion becomes exceedingly difficult with these fine filaments, the material begins to approach its theoretical maximum strength. In fact, these new composites can become even stronger than their reinforcing filaments. Moreover, a relatively small percentage of filaments can achieve this spectacular increase in strength. For example, in the super-tough copper-niobium conductor, the niobium filaments occupy only 18 percent of the volume. Thus, the desirable high conductivity of copper is retained.

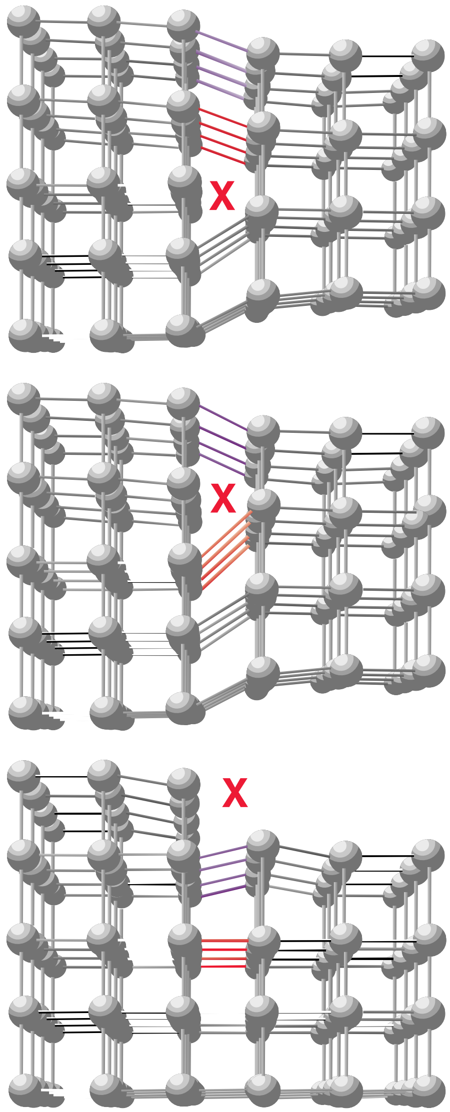
We achieve the optimum combination of strength, ductility and conductivity in these composites by varying the size and spacing of filaments in the material. Test samples have approached 320,000 pounds per square inch—twice the strength of the copper-niobium wire used in our magnets. Such improvements in wire performance suggest that peak magnetic fields achievable with nondestructive pulsed magnets might approach one million gauss.

---

cessive heat by limiting the duration of the magnetic-field pulse to under a second. The destructive pulsed magnets also sidestep the stress problem. They are designed for a single pulse, which they never survive intact. The pulse lasts only a few microseconds before a mechanical shock wave, moving at nearly the speed of sound, obliterates the magnet.

Because of their high cost, the strongest DC magnets are confined to a handful of the world's nationally funded magnet laboratories. They are energized by electric power substations of a size that might supply a small town. Impressive plumbing circulates deionized, highly pressurized water through the magnets, to keep them cool. They can operate for many hours, and the electric bill alone can exceed $1,000 an hour. Recently the new National High Magnetic Field Laboratory in Tallahassee attained 300,000 gauss, a record for a conventional (resistive) DC electromagnet.

Superconducting electromagnets avoid the heating problem. To sustain the superconductivity, researchers typically operate these magnets at a temperature of 4.2 kelvins (–269 degrees Celsius), achieved by submerging the magnet in liquid helium. The primary drawback of superconducting magnets is that their field, if sufficiently intense, will interfere with the superconductivity of their own wire. For this reason,



EDGE DISLOCATION, denoted by the red X, arises from an extra plane of atoms in part of the crystal. In the sequence shown, the dislocation moves through the material as the chemical bonds shown in red and purple are broken and re-formed, one at a time. Eventually, the entire left half of the crystal is upwardly displaced. Repeated millions of times, this common process leads to mechanical failure—the separation of the two halves of the crystal.

superconducting magnets have so far been limited to about 200,000 gauss.

In several laboratories, however, they have been combined with resistive DC magnets to create a hybrid design, with the resistive element inside the large superconducting one. A hybrid magnet at the Francis Bitter National Magnet Laboratory at the Massachusetts Institute of Technology holds the record—385,000 gauss—for all types of DC magnets.

Still higher magnetic fields are generated by pulsed magnets. One destructive design achieved nearly 10 million gauss with the help of high explosives, which symmetrically compressed the magnetic field into an exceedingly small volume around the sample (much the same kind of implosion sets up the fast fission reaction in an atomic bomb). With this approach, the entire apparatus and sample are vaporized, preferably in a remote area. Perhaps it is no surprise that the two government laboratories specializing in this design are the Los Alamos National Laboratory in New Mexico and its counterpart, in Arzamas, Russia.

An alternative destructive magnet design, developed at the Megagauss Laboratory in Tokyo, is particularly convenient because the sample is rarely damaged when the magnet explodes. The design is elegantly simple: the coil is a single loop of copper, which generates a magnetic field of 1.5 million gauss in the few microseconds before it is blasted open. The main challenge in protecting the apparatus is catching the shrapnel, a task entrusted to carefully positioned plywood.

Less spectacular though they may be in normal operation, nondestructive pulsed magnets have considerable advantages. They make possible a much wider assortment of scientific experiments, because pulse duration is increased up to 10,000 times, to a range of 10 to 100 milliseconds. Also, they offer more control over the pulse shape and shorter intervals between pulses, as little as 20 minutes. All these characteristics make it easier to determine the integrity of the experimental data acquired during the pulse. Finally, the magnets often involve a simpler laboratory infrastructure and lower operating costs, both manageable by a small research group. The price paid for these benefits comes in confronting the extreme stresses exerted on the magnet by its field.

Handling the Stress

There are several different ways of designing an electromagnet to survive pulse after jarring pulse. Our approach relies on a careful arrangement of different types of wire within a single magnet. The ideal conductor wire in an electromagnet would have strength, ductility and conductivity in abundance; unfortunately, conductor strength invariably comes at the expense of conductivity and ductility. So wires with different properties are mixed and matched to fit conditions inside the magnet.

In our designs, the wire is wound into a coil of 14 concentric layers, each of which has about 30 turns of wire [*see illustration on page 61*]. The intensity of the magnetic field is greatest in the center of the magnet and decreases more or less linearly through the 14 layers of the magnet. The greatest stresses, on the other hand, occur in the fourth layer from the center, because the stress, it turns out, is proportional to the product of the local magnetic field, the electric current density in the wire and the radius of the layer of wire.

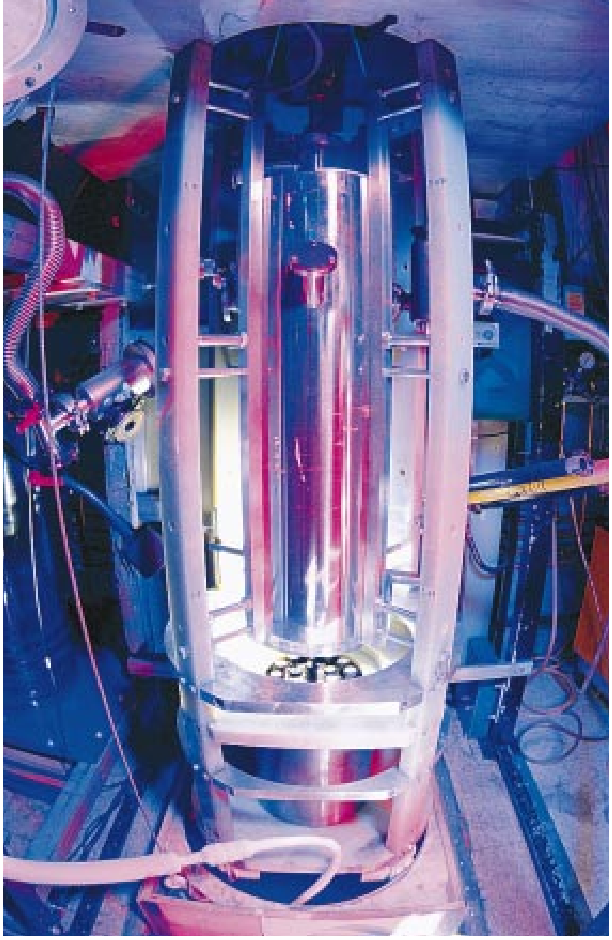
We build our magnets with three different copper-based conductors. For the middle layers, which are most highly stressed, our choice is a copper-niobium composite wire, the strongest conductor available. It can withstand as many as 165,000 pounds per square inch—more than many steels, which in any case are too brittle and not nearly conductive enough for this application. First used at M.I.T., this wire derives its strength from a dense distribution of microscopic niobium filaments embedded in a copper host. The way these filaments endow the wire with such remarkable strength is an interesting story in itself [*see box on opposite page*].

In the first layer the tight bending of the small coil during winding demands a much more ductile material than copper niobium; our choice is a mixture of copper, nickel and beryllium oxide. The seven largest-diameter, outermost layers experience the lowest stresses, both during the winding and pulsing of the coil. They also dominate the overall electrical resistance, simply because they require much greater lengths of wire. So they are wound with wire made of a weaker material, copper aluminum oxide, which is more conductive, available in large quantities and still three times stronger than ordinary copper.

Such an arrangement enables the coils to withstand considerable pressure. To achieve the highest magnetic fields, pulsed-magnet coils must distribute the stress effectively. When these forces exceed a certain point, permanent deformation of a material results. If the heavily stressed conductor has sufficient ductility, it can "lean" against the next layer for support. We design this capability into our magnets, with the help of a computer model developed with Phil Snyder of Princeton University. Indeed, materials science challenges aside, the most difficult problems in designing, modeling and constructing world-record magnets involve distributing stress effectively.

An alternative design strategy has been pursued by our colleagues at the pulsed-magnet laboratory in Leuven, Belgium. They use only a single type of conductor within a magnet. Each layer of conductor is mechanically reinforced by a thick supporting layer of strong glass fibers. This design separates somewhat the tasks of generating the magnetic field and withstanding the extreme stresses. Magnets of this type recently achieved peak magnetic fields comparable to our own.

Our magnets are also mechanically reinforced, but only around the outermost layer of conductor. The entire coil winding is rigidly supported in a reservoir of small (one millimeter in diameter) beads of stabilized zirconium oxide. These beads, made of a high-strength cousin of the well-known costume-jewelry diamond substitute, are among the strongest known insulating materials. The beads, pressurized within a steel housing, provide support rigid enough to help keep a small, localized failure—a simple break in the coil wire, for example—from triggering a more catastrophic chain of events. In many cases, it is a simple matter of unwinding a magnet to discover precisely where and how it failed. Indeed, our December 3, 1992, explosion was the last of its kind. Magnet failures these days are usually silent and leave the experimental apparatus undamaged in the center of the magnet.


CHIP SIMONS

RIGID SUPPORT STRUCTURE secures the electromagnet—within the steel cylinder at the bottom—and connects it to capacitors in a room below. The long, shiny cylinder that runs almost the length of the structure is a cryogenic container that cools the experimental sample and guides it into the magnet.

Unleashing the Pulse

Even with all this reinforcement, operation of a nondestructive pulsed magnet is an involved procedure. Before unleashing the pulse, the magnet is cooled to 77 kelvins by submersion in liquid nitrogen. This critical step increases the tensile strength of the conductors by about 15 percent, while decreasing the resistivity of the conductors by two to four times. Lower resistivity means less energy lost as heat. Even so, the temperature of the magnet increases by about 200 kelvins during a pulse, reaching room temperature in about 0.01 second (pulsing a magnet at room temperature would melt the entire coil almost instantly). After a pulse, a wait of about 20 minutes is needed while the violently boiling liquid nitrogen reservoir settles down and cools the magnet for the next pulse.

Sometime in the next decade, nondestructive magnetic-field pulses of a million gauss or more will probably be achieved. When this milestone is reached, who knows what new physical phenomena might be observed. Past increases in magnetic fields have led to completely unanticipated discoveries. Experiments on magnetic materials might bring the next revolution in stronger permanent magnets. Experiments in superconductivity might verify an exciting and counterintuitive theoretical prediction: superconductivity in some materials, once destroyed by an intense magnetic field, might actually be reestablished by still higher fields.

Stronger conductors that could achieve a megagauss would also have some intriguing applications of their own. They might dramatically improve the performance of plasma fusion reactors because power production is greatly enhanced by increased magnetic fields. More intense pulsed fields might also increase the velocity of projectiles fired by electromagnetic means. This technology, developed at Sandia National Laboratories in New Mexico, might one day compete with rockets for sending satellites into space. For intense magnetic fields, it would seem, not even the sky is the limit.

Further Reading

HIDDEN ATTRACTION: THE HISTORY AND MYSTERY OF MAGNETISM. Gerrit L. Verschuur. Oxford University Press, 1993.

PERMANENT MAGNET MATERIALS AND THEIR DESIGN. Peter Campbell, Cambridge University Press, 1994.

RESISTANCE IN HIGH-TEMPERATURE SUPERCONDUCTORS. David J. Bishop, Peter L. Gammel and David A. Huse in *Scientific American,* Vol. 268, No. 2, pages 48–55; February 1993.

ARTIFICIAL ATOMS. Marc A. Kastner in *Physics Today,* Vol. 46, No. 1, pages 24–31; January 1993.

SEVENTY-TWO TESLA NON-DESTRUCTIVE PULSED MAGNETIC FIELDS AT AT&T BELL LABORATORIES. G. S. Boebinger, A. Passner and J. Bevk in *Physica B,* Vol. 201, pages 560–564; July–August 1994.

The Authors

GREG BOEBINGER, AL PASSNER and JOZE BEVK are researchers at AT&T Bell Laboratories in Murray Hill, N.J. Boebinger and Passner collaborate on all aspects of the design, construction, use and occasional disintegration of pulsed electromagnets. In recent years Boebinger has used magnetic fields to study a variety of materials, including semiconducting microstructures and organic and high-temperature superconductors. His avocations include traveling, reading history and telling humorous stories that are mostly true. Passner has worked on optical bistability in solids and on a magnetic antimatter "bottle" for trapping a positron plasma. He is involved in civic and charitable activities. Bevk brings expertise in the physics and materials science of ultrafine filamentary composites, having performed pioneering work on high-strength conductors while on the faculty at Harvard University in the 1970s. His current research interests are silicon-based transistors smaller than one quarter of a micron.

Hookworm Infection

It retards growth and intellectual development in millions of children yet is largely ignored by researchers. New findings suggest excellent possibilities for a vaccine

by Peter J. Hotez and David I. Pritchard

If everyone in the world were surveyed today, an estimated one billion individuals—roughly a fifth of the planet's inhabitants—would be found to harbor hookworms in their small intestine. These parasites engage in an insidious form of thievery. With their sharp teeth, the half-inch-long bandits grasp the surface and subsurface layers of the intestinal wall and extract blood. Each one empties a fraction of a teaspoon from the circulation every day, but when 20, 100 or even 1,000 worms drain this much blood simultaneously (in the last case, drawing almost a cup of blood), the consequences can be profound.

DAVID SCHARF

Blood delivers iron, protein and other nutrients to tissues. If the host cannot replace the lost substances quickly enough (as often is true of children, women of childbearing age and anyone who is malnourished), the result may be iron deficiency anemia and protein malnutrition, the hallmarks of passage from mild infection to outright disease. Together anemia and protein malnutrition, which occur in up to 25 percent of infected individuals, can lead to extreme lethargy and weakness. What is worse, when children are chronically infected by many worms, the lack of iron and protein can cause severe retardation of growth and can impair behavioral, cognitive and motor development, sometimes irreversibly. Occasionally hookworm disease may even be fatal, especially to infants.

Hookworm disease, which is most prevalent in the developing nations of the tropics, can be treated. Yet in many places where the condition is common, the medicines—parasite-killing agents and sometimes iron supplements—may be unavailable or hard to get. For this reason and others there is a pressing need for vaccines that can prevent hookworms from establishing thriving populations in the gut. Sadly, however, hookworm disease has been ignored by most of the biomedical research community for the past 25 years. The main reasons are straightforward: funding for investigation of disorders that predominantly affect the Third World is scarce, and maintaining the worms in the laboratory is difficult. As a result, study of hookworm infection has not benefited from the revolution in biotechnology that has led to impressive advances in the understanding and treatment of other human disorders.

We are now trying to reverse that trend. Our laboratories, along with a few others, have begun applying modern molecular approaches to exploring the two main genera of hookworms, *Ancylostoma* and *Necator.* During the past few years, this effort has made it possible to identify a range of hookworm proteins that can potentially serve as preventive vaccines. In a welcome and unexpected turn of events, many of these proteins also seem to hold promise as drugs for cardiovascular and immunologic disorders common in industrial countries.

PETER J. HOTEZ and DAVID I. PRITCHARD have been investigating the molecular aspects of hookworm infection for most of their scientific careers. Hotez is associate professor of pediatrics and epidemiology and director of the Medical Helminthology Laboratory at the Yale University School of Medicine. He graduated from the combined M.D. and Ph.D. program at Cornell and Rockefeller universities before completing his residency training at Massachusetts General Hospital. Hotez joined the faculty of Yale in 1989. Pritchard is a reader in life sciences at the University of Nottingham in England. He earned his bachelor's degree in zoology from the University of Wales and his Ph.D. in immunology from the University of Birmingham. After a spell in the pharmaceutical industry, he moved to Nottingham in 1981.

A Remarkable Route to Infection

There was a time when the U.S. and other nations put hookworm disease higher on the research agenda. The medical community became well aware of its potential seriousness in 1880, when an epidemic of what was called miners' anemia struck Italian laborers building the Saint Gotthard railway tunnel in the Swiss Alps. *A. duodenale,* one of the two species most responsible for hookworm disease in humans, was at fault in that case. By 1902 the second species, *N. americanus,* had been identified as well, and many details of how hookworms are passed from one individual to the next had been worked out. Shortly thereafter hookworm infection by this second species was found to be rampant in the southeastern U.S. In response, oil baron John D. Rockefeller founded the Rockefeller Sanitary Commission, to try to eradicate the disorder from the region. He also supported a campaign to control hookworm infection overseas. Those efforts led to introduction of new treatments and laid important groundwork for the progress being made today.

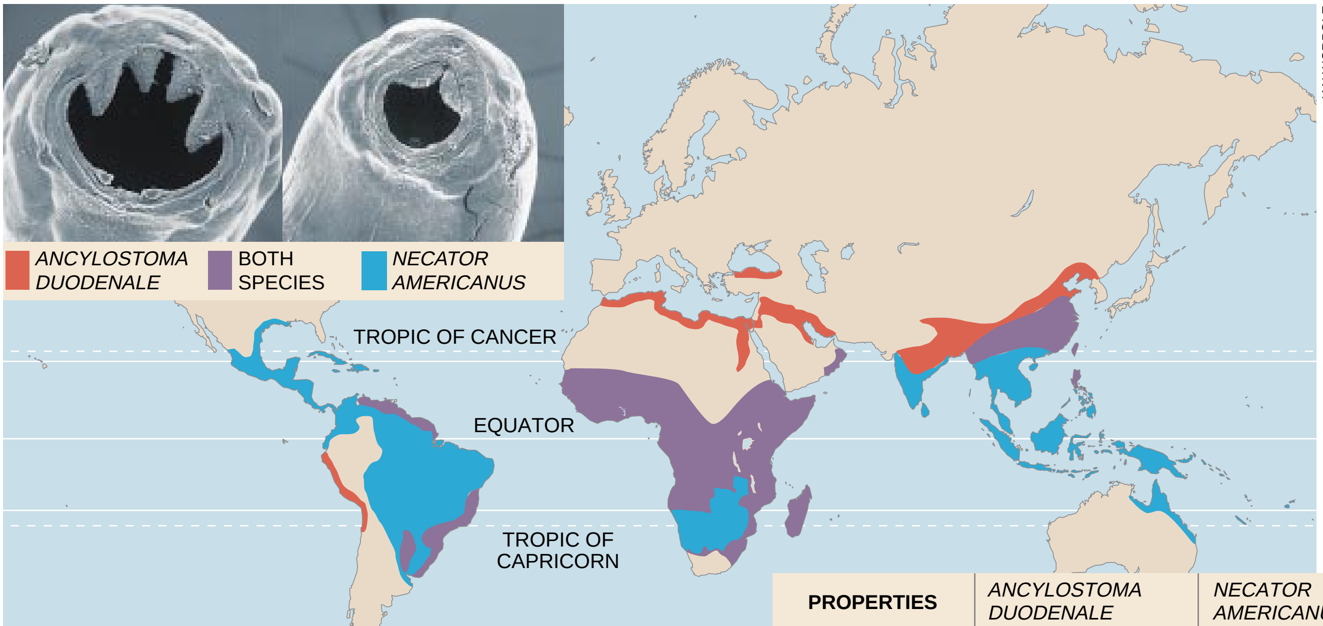
Even before Rockefeller became involved, however, researchers had made the fascinating discovery that the route to human infection is rather circuitous and requires eggs and young larvae to spend time in the soil, outside the human host. Every day the adult female worm releases thousands of fertilized eggs that pass out of the host's body with feces. To

DAVID SCHARF

HOOKWORMS *Ancylostoma duodenale* (*right*) and *Necator americanus* (*opposite page*) are the main species affecting humans. Hookworms draw nourishment by latching onto the wall of the small intestine and drawing blood.

DAVID I. PRITCHARD

IAN WORPOLE



PROPERTIES	*ANCYLOSTOMA DUODENALE*	*NECATOR AMERICANUS*
Size of adult (female is slightly larger than male)	0.3 to 0.5 inch	0.2 to 0.4 inch
Number of eggs produced daily per worm	10,000 to 30,000	5,000 to 10,000
Estimated amount of blood removed daily per worm	0.04 teaspoon (greater draw makes this species more virulent)	0.004 teaspoon
Mouthparts	Sharp teeth	Cutting plates
Natural adult lifespan	1 year	3 to 5 years
Ability to produce infection if ingested	Yes	No
Ability of larvae to lie dormant in host	Yes	No

LISA BURNETT

TROPICS HARBOR most *A. duodenale* and *N. americanus* hookworms in the world (*map*); red and blue colors indicate one of these species predominates in a region, but the other species might also be present in spots. Salient features of the worms are listed in the table. Collectively these organisms perform the equivalent of draining all the blood from some 1.5 million people every day.

survive and grow, the nascent larvae need warm, moist and well-aerated soil shielded from sunlight. Those conditions arise most often in rural areas of the tropics, especially places where crops provide shade or are cultivated under taller trees. Hence, most heavy infections (involving 100 worms or more) occur in people who live where coconuts, cocoa, coffee beans, tea, sugarcane, sweet potatoes or mulberry trees are grown.

After eggs have been in the soil for a day, microscopic larvae emerge. They feed on organic debris and bacteria before undergoing two molts to the infective stage. The infective larvae do not eat. Instead they move up to the soil surface to make contact with a host. In this pursuit, they may exhibit what is called questing behavior: they climb up a blade of grass or a particle of soil, stand on their tail and sway.

Typically the barely visible larvae infiltrate the body by burrowing into the skin of the legs or feet, although oral ingestion of *A. duodenale* larvae can result in infection as well. Travel through the skin induces an apparently inefficient inflammatory response, in which white blood cells accumulate in the affected tissue. For whatever reason, the cells generally do not eliminate the worms. Meanwhile, though, the inflammation may give rise to an intense, burning itch. Indeed, a notorious outbreak of "ground itch," as the reaction is called, occurred in northern India in 1978, after a group of men began a popular game, *kabaddi,* that involves tackling opponents to the ground. The players were unaware that the muddy field was contaminated with infective hookworm larvae until horrible itching all over their bodies forced them to run off in a desperate search for relief.

Some investigators suspect that passage to the small intestine is an easy matter for larvae that are ingested—they may migrate directly to their final destination. Larvae that cross into the body through the skin, however, must work rather hard to bore through the epidermis to the underlying dermis and into small vessels of the blood or lymphatic systems. Shortly after penetrating these vessels, they are carried by the venous circulation to the right side of the heart and into the lungs. There they break out of the circulation into the airways and are coughed up and swallowed. In the small intestine the larvae, which may have begun to mature in the skin, finally develop into adult males and females. Now they begin to mate and feed. Females generally begin releasing eggs within about two months after arriving in the body as larvae.

Curiously, this continuous cycle is sometimes interrupted midstream. Gerhard A. Schad and his colleagues at the University of Pennsylvania have shown that *A. duodenale* larvae can remain dormant for months in the body before reaching the intestine. No one knows exactly where they rest or how they get to their hideout, but Keun Tae Lee, Maurice Dale Little and the late Paul C. Beaver and their colleagues at Tulane University have identified hookworm larvae in muscle fibers of animals and humans. Hence, muscle tissue may be one repository of dormant larvae.

From the worm's point of view, entering a dormant state can make good survival sense. If larvae mature promptly during the dry seasons of the year, females may release eggs onto inhospitable soil. But if the larvae can somehow time their arrival to the intestine so that egg production coincides with the wet season, the eggs will have a much greater chance of thriving. It seems that larvae may well possess such a sensing capacity. Working in West Bengal, India, in the late 1960s, Schad's group determined that infective *A. duodenale* larvae remain developmentally arrested in tissues during the hot and dry months of the year and then progress to egg-laying adults just before the monsoon season. Thus, the eggs are deposited, as is needed, on moist soil.

The period of dormancy may be good for hookworms, but it can be disruptive to the host. Larvae may not be killed by drugs that eliminate adult hookworms from the intestine. Thus, patients who are treated with seeming success can fall ill again months later, even if they manage to avoid additional exposure.

Also troubling is circumstantial evidence that arrested *A. duodenale* larvae can enter the milk of nursing mothers, thereby causing severe infection in newborns. Part of that evidence is the fact that transmission of this kind occurs in dogs. In addition, Yu Sen-hai and Shen Wei-xia of the Institute of Parasitic Diseases in Shanghai recently identified many human newborns who had severe *A. duodenale* infections. The possibility for transmission through breast milk is particularly disturbing because neonates with severe infections can die from the resulting blood loss.

The Trouble with Therapy

Knowing the life cycle of hookworms, the two of us independently set out to identify the molecules enabling them to burrow through the skin, mature and survive in the human body. We reasoned that if we could isolate those molecules, we might be able to enlist some of them as vaccines. One of us (Hotez) focused on the genus *Ancylostoma,* which includes several species; the other (Pritchard) focused on *Necator,* which consists mainly of *N. americanus.* Hotez has worked mostly with *A. caninum* as a representative of its genus because the species is easier to maintain in the laboratory than is *A. duodenale.* As might be expected from the name, *A. caninum* primarily infects dogs, but it can also affect humans, as was shown recently by Paul Prociv of the University of Queensland in Australia and John Croese of Townsville General Hospital, also in Australia.

We were especially interested in creating vaccines, largely because, as we have mentioned, gaining access to therapy is problematic in many regions of the world and because dormant larvae can generate new infections months after treatment is completed. These difficulties are compounded by the fact that humans afflicted with an initial hookworm infection apparently do not acquire the strong protective immunity against future infections that benefits people who have once contracted, say, chickenpox. Consequently, people who live in areas where hookworms are prevalent will be reexposed and reinfected continually and will need repeated courses of drug therapy—an impossibility for most victims. Moreover, frequent use of the drugs is not an ideal solution to the hookworm problem. Investigators at the Institute of Parasitic Diseases in Shanghai, who have studied a great many patients, have now discovered that the medications are not always as effective as had been thought.

The best approach to prevention, of course, is improving sanitation to the point where transmission cannot occur. This measure is thought to have been largely responsible for eradicating human hookworm disease from the southeastern U.S. Yet it is unlikely that sanitation systems will be constructed in most of the developing world any time soon. Nor is it likely that people living in the hot tropics will always wear protective clothing in hopes of reducing exposure. At the moment, then, the most logical alternative is the broad administration of a vaccine designed to boost host defenses.

We can imagine two different types of vaccines. In the classic, anti-infection approach, molecules or dead or defective whole organisms are administered to evoke an immune response targeted to one or more molecules appearing on the surface of the organism. When the approach is successful, induction of the response enables the body to eliminate the organism promptly any time it enters the body.

Even though natural hookworm infection does not seem to induce much protective immunity, vaccination might still confer protection. More than 30 years ago Thomas A. Miller, then at the University of Glasgow, produced immunity in dogs by inoculating them with live but radiation-damaged *A. caninum* larvae in the infective stage. A human vaccine consisting of living larvae would probably fail to gain governmental approval. Nevertheless, Miller's success might well be reproduced in people by administering genetically engineered larval proteins or protein fragments. If the vaccine worked, it would prevent larvae from surviving or maturing in the body.

We are also exploring new kinds of vaccines, which we call antidisease vaccines. Because hookworms do not multiply in the host, anything that reduces the amount of blood they extract should minimize the damage they do. We suspect that if we can identify molecules that are secreted by hookworms but that are not made by the human body, delivery of those molecules might elicit an immune response against them. Although the response might not kill the worms directly (because the released substances diffuse away from the parasites), it could neutralize secretions that enable the worm to mature, to draw blood or to evade attack by other parts of the immune system.

As part of our effort to identify molecules that could be useful in vaccines, we have examined the biochemical

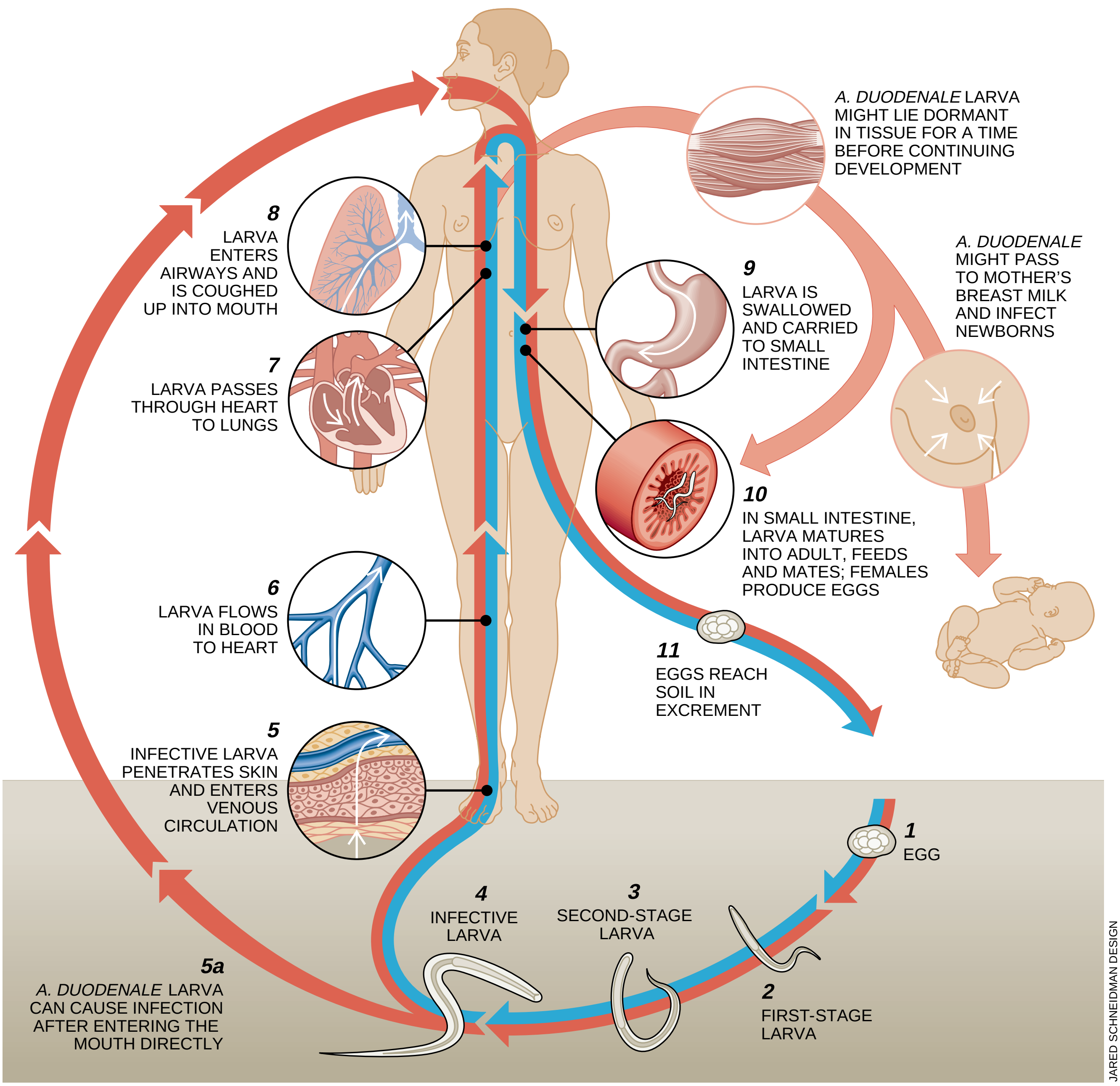
ROCKEFELLER ARCHIVE CENTER

CHILD FROM INDIA featured in a 1919 report by the Rockefeller Foundation Health Board was thin and frail from chronic hookworm disease before treatment (*left*) but gained weight and muscle mass after therapy eliminated the parasites from his intestines (*right*). Unfortunately, certain potential effects of chronic infection, such as compromised intellectual development, are sometimes irreversible, and reinfection is common.

events that occur after a hookworm fastens to the intestinal wall. Having attached, the parasite pumps its muscular pharynx and thereby sucks in a mouthful, or plug, of the mucous membrane and underlying material. At the same time, it secretes strong enzymes capable of breaking down the tissue.

Some of these enzymes have been identified by our laboratories and by that of Paul J. Brindley of the Queensland Institute of Medical Research in Brisbane, Australia. They include proteases, which degrade host proteins, as well as a hyaluronidase, a compound that breaks down hyaluronic acid and other structural components of the intestine. The combination of this mechanical and chemical assault not only releases potentially nourishing components of the wall for the worm, it also ruptures small blood vessels in the plug. The worm swallows much of the liberated blood, but a lot is wasted as well.

The human host responds to this invasion by mounting a two-pronged defense. Part of the defense aims to coagulate blood at the site of vascular damage. Coagulation can prevent further loss and inhibit feeding by hookworms. The other part of the defense is immunologic. It involves the activation of white blood cells (leukocytes), including neutrophils and eosinophils. The leukocytes attempt to kill the parasites by generating free radicals and lipid hy-



PARTS OF THE LIFE CYCLES of *A. duodenale* (*red arrows*) and *N. americanus* (*blue arrows*) are alike. The cycles can be said to begin with the initial development of larvae in the ground (*1–4*). Infective larvae generally enter the body through the skin (*5*) and follow a circuitous route to the small intestine (*6–9*). There they mature into adults and mate to produce fertilized eggs that are transported to the ground in excrement (*10* and *11*). *A. duodenale* larvae can, however, sometimes establish infection in the small intestine after entering the body orally (*5a*). Some evidence suggests *A. duodenale* larvae can take worrisome detours as well (*light red arrows at right*). They may lie dormant in muscle tissue before heading to the intestine; they may also enter the mammary glands of pregnant women and be passed to newborns through breast milk.

droperoxides, both of which are highly reactive and destructive.

The immunologic defense also involves the dispatching of antibody molecules, which can act in concert with the leukocytes. Among the antibodies are those of the immunoglobulin E (IgE) type—the kind that play a role in allergic responses. Studies that Pritchard and his colleagues recently completed in Papua New Guinea indicate IgE-mediated reactions can be quite strong in some patients (possibly those most prone to allergies) and seem able to diminish feeding by the worms; nevertheless, they apparently do not kill many worms outright.

Hookworms as Chemical Factories

The worms persist because they have evolved strategies to thwart the vascular and immunologic defenses. For instance, the parasites interfere with at least two components of the process leading to clot formation. When a small blood vessel is damaged, vascular cells release a sugary protein, known as tissue factor, that combines with a circulating protein, factor VII. The resulting complex triggers a cascade of events that culminates in activation of a third molecule—an enzyme called factor X. Activated factor X (known as factor Xa), in turn, converts the protein prothrombin into the enzyme thrombin, which converts another protein, fibrinogen, into insoluble strands of fibrin. Fibrin then collects as a mesh on the vessel wall and constitutes the essence of a clot. The mesh usually entraps platelets and other substances as well.

Investigators have known for almost a century that hookworms produce at least one anticlotting agent, but its nature was a mystery. We have found that hookworms produce a substance that prevents aggregation of platelets in the test tube. We and others have also established that hookworms secrete a protein that operates early in the clotting pathway and blocks the activity of factor Xa; it thereby impedes all subsequent steps in the pathway. That protein has now been isolated from *A. caninum* by Michael Cappello in Hotez's laboratory, in collaboration with George P. Vlasuk of CORVAS International in San Diego, Calif. The molecule, called AcAP (*A. caninum* anticoagulant peptide), is a rather small, highly potent protein never before seen in nature—properties that, as will be seen, mean it might be useful as a drug for diseases unrelated to hookworm infection.

The hookworm's strategy for evading immune destruction similarly depends on the release of chemicals. Investigators at CORVAS have isolated one such substance, called neutrophil inhibitory factor, from the dog hookworm. It inhibits the activity of neutrophils and eosinophils, in part by preventing them from releasing strongly oxidative chemicals.


DAVID SCHARF

A. CEYLANICUM* INFECTS people in some parts of Asia but is less prevalent worldwide than are *A. duodenale* and *N. americanus.

Pritchard and his colleague Peter M. Brophy have shown that *N. americanus* has evolved a related defense. It secretes antioxidant enzymes (such as superoxide dismutase and glutathione-*S*-transferase) that neutralize free radicals and the like. For added protection, *N. americanus* produces an enzyme that degrades the neurotransmitter acetylcholine, which is emitted by neurons in the intestinal wall. Were the neurotransmitter to act freely, it would help activate leukocytes against the parasite.

Finally, we find that the chemicals used to break up the tissue in the ingested plug also facilitate evasion of immune destruction. One of the proteases made by *N. americanus* dissolves antibodies, and the hyaluronidase made mainly by species of *Ancylostoma* promotes the local spread of the various secretions. This spreading ensures that hookworms are surrounded by a protective chemical shield.

In many respects, then, the adult hookworm in the gut functions as a self-contained pharmaceutical manufacturer, producing molecules designed to inhibit blood clotting and the immunologic responses of the host. The adult, though, is not the only one to produce interesting chemicals; so do larvae. John M. Hawdon in the Hotez laboratory has demonstrated, for instance, that infective *A. caninum* larvae elaborate at least two proteins: a protease that presumably helps the organisms to bore through the skin, and another protein of unknown function. This second protein, called ASP (*Ancylostoma* secreted polypeptide), may participate in maturation to the feeding stage.

Many Vaccine Candidates

Discovery that hookworms generate so many proteins during their life cycle suggests exciting ideas for vaccine development. For example, Hawdon and Brian F. Jones in Hotez's group have recently cloned the gene for ASP and determined the amino acid sequence of the protein it encodes. One part of the protein turns out to be very similar to a segment of a protein found in the venom of stinging insects. This venom protein elicits a strong response by the human immune system. Moreover, antibodies responsive to the venom constituent also respond to ASP when they are mixed with it in a test tube.

These results provide grounds for thinking that ASP might serve as a vaccine able to elicit an antibody response that would destroy infective hookworm larvae or impede their maturation. Large quantities are now available through genetic engineering. Armed with that supply, Hotez will soon collaborate with investigators at the Institute of Parasitic Diseases in Shanghai to test it as a vaccine in experimental animals. The other hookworm products mentioned earlier—notably AcAP and other anticlotting factors, the antioxidants and neutrophil inhibitory factor—are also being considered for testing as vaccines.

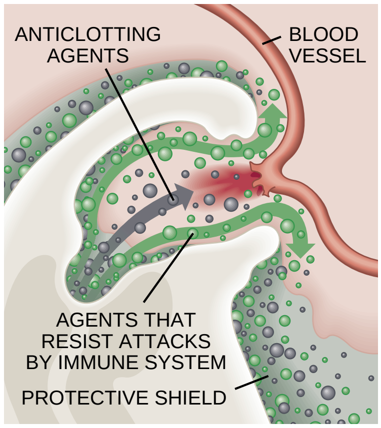
Studies by Edward A. Munn and his colleagues at the Babraham Institute in England have pointed us to still another molecule meriting further evaluation. They have demonstrated that a vaccine consisting of an enzyme made by a stomach worm (not a hookworm) of sheep can protect sheep from reinfection by that parasite. Pritchard and his colleagues have shown that *N. americanus* makes a similar enzyme, which suggests that injection of the hookworm enzyme might elicit a protective response against hookworm infection in people.

Vaccine possibilities may emerge as well from investigations focused on the genetic determinants of susceptibility to hookworm infection. Several years ago Schad and Roy M. Anderson, who is now at the University of Oxford, noted that people who initially acquire a heavy infection seem to reacquire heavy infections, whereas lightly infected individuals reacquire light infections. There are some indications that these patterns are controlled by the genes of the hosts. If so, identifying the genes that influence susceptibility would provide additional clues to preventing severe infections. Pritchard's work in Papua New Guinea suggests that genes involved in allergy would particularly be worth pursuing.

Such prospects are tantalizing, but a serious obstacle to vaccine production remains. Pharmaceutical manufacturers are loath to invest in vaccines for the Third World because lack of a wealthy

ADULT HOOKWORM in the small intestine (*micrograph* and *detail*) sucks in superficial layers of the intestinal wall and breaks up tissues and blood vessels to obtain nourishment. It ensures itself a continuing supply of blood and protects itself from immune attack by secreting a raft of chemicals (*spheres in detail*), some of which are listed in the table. Adult and larval secretions may one day serve as vaccines for hookworm disease or as treatments for cardiovascular and immune disorders. The worm in the micrograph was found on autopsy in an infant who died of an undetected hookworm infection.

WAYNE M. MEYERS *Armed Forces Institute of Pathology*



JARED SCHNEIDMAN DESIGN

	CHEMICAL	FUNCTION
CHEMICALS THAT BLOCK CLOTTING OF BLOOD	Anticoagulant peptide (in *Ancylostoma*)	Inhibits activity of an enzyme crucial to blood clotting
	Inhibitor of platelet aggregation (in *Ancylostoma, Necator*)	Prevents platelets from facilitating clot formation
CHEMICALS THAT DEFLECT IMMUNE ATTACK	Superoxide dismutase (in *Necator*)	Antioxidant; neutralizes destructive, oxidizing agents secreted by immune cells
	Glutathione-*S*-transferase (in *Necator*)	Antioxidant
	Enzyme that degrades antibodies (in *Necator*)	Prevents antibodies from participating in immune response
	Acetylcholinesterase (in *Necator*)	Breaks down acetylcholine, which can induce immune cells to attack hookworms
	Neutrophil inhibitory factor (in *Ancylostoma*)	Prevents immune cells from secreting oxidizing agents
	Hyaluronidase (in *Ancylostoma*)	Facilitates local spreading of other hookworm secretions

LISA BURNETT

market means the companies will not recoup their costs [see "Better Than a Cure," by Tim Beardsley; SCIENTIFIC AMERICAN, January]. Unless significant amounts of money become available, development of hookworm vaccines may well languish.

Ironically, the hookworm proteins that have been identified so far may find initial application not as vaccines but as therapies for disorders not related to hookworm infection. As a case in point, the anticlotting agent AcAP might prove useful in the treatment of heart disease.

Recall that AcAP interferes with the activity of factor Xa, a crucial player in the cascade leading to blood clotting in the intestine. Factor Xa also seems to be important to formation of the blood clots that cause heart attacks. These facts, combined with the uniqueness of AcAP's structure, imply the protein might have use as an alternative to existing anticoagulant drugs in some patients with cardiovascular disease. AcAP might be beneficial, for instance, in preventing the reclogging of arteries opened by angioplasty (the expansion of a "balloon" in an artery to clear the passageway).

Neutrophil inhibitory factor and other immune-blocking chemicals also hold promise for treating diseases common in the industrial nations. Studies are under way to determine whether they should be developed as therapies for autoimmune diseases, transplant rejection, and asthma or other allergies.

Remembering the Children

The possibility that molecules found in hookworms could have diverse applications is pleasing, but therapeutic applications for the industrial world are not our first priority. We are still driven by the hope that we will find a way to eradicate hookworm disease.

In 1911 C. L. Pridgen, a physician in Columbus County, North Carolina, described the heartbreaking condition of a 15-year-old child afflicted with hookworms. In the Second Annual Report of the Rockefeller Sanitary Commission, he wrote:

> The boy was a typical subject. His skin had a cadaverous appearance.... He seemed to take no notice of the crowd... and submitted himself to a physical examination with the most profound indifference. When the examination was completed he at once turned, without waiting to hear the diagnosis, and went to a log near by and sat down, apparently very fatigued.... [The father] said that the boy had always been of no account...would lie on the porch like a dog, day after day apathetic. He could not be aroused to either work or play.

With treatment the youngster's condition improved dramatically. Pridgen quotes the father as later reporting, "When I left home, that boy was following a plow and yelling at the mule in a way to let you know he was just glad to be living."

Chances are good, however, that the child later became infected and lethargic again. Had he lived in the Third World, chances are also good he would not have received treatment every time it was needed. We hope someday to produce a simple vaccine that will prevent such profound lethargy—and worse—in millions of children around the world.

FURTHER READING

HOOKWORM DISEASE: CURRENT STATUS AND NEW DIRECTIONS. Edited by G. A. Schad and K. S. Warren. Taylor and Francis, 1990.

IMMUNITY TO HELMINTHS: READY TO TIP THE BIOCHEMICAL BALANCE. P. M. Brophy and D. I. Pritchard in *Parasitology Today,* Vol. 8, No. 12, pages 419–422; 1992.

PARASITIC DISEASES, third edition. D. Despommier, R. Gwadz and P. J. Hotez. Springer-Verlag, 1995.

MOLECULAR PATHOBIOLOGY OF HOOKWORM INFECTION. P. J. Hotez, J. Hawdon, M. Cappello, B. Jones and D. I. Pritchard in *Infectious Agents and Disease,* Vol. 4, No. 2; June 1995 (in press).

The Arithmetics of Mutual Help

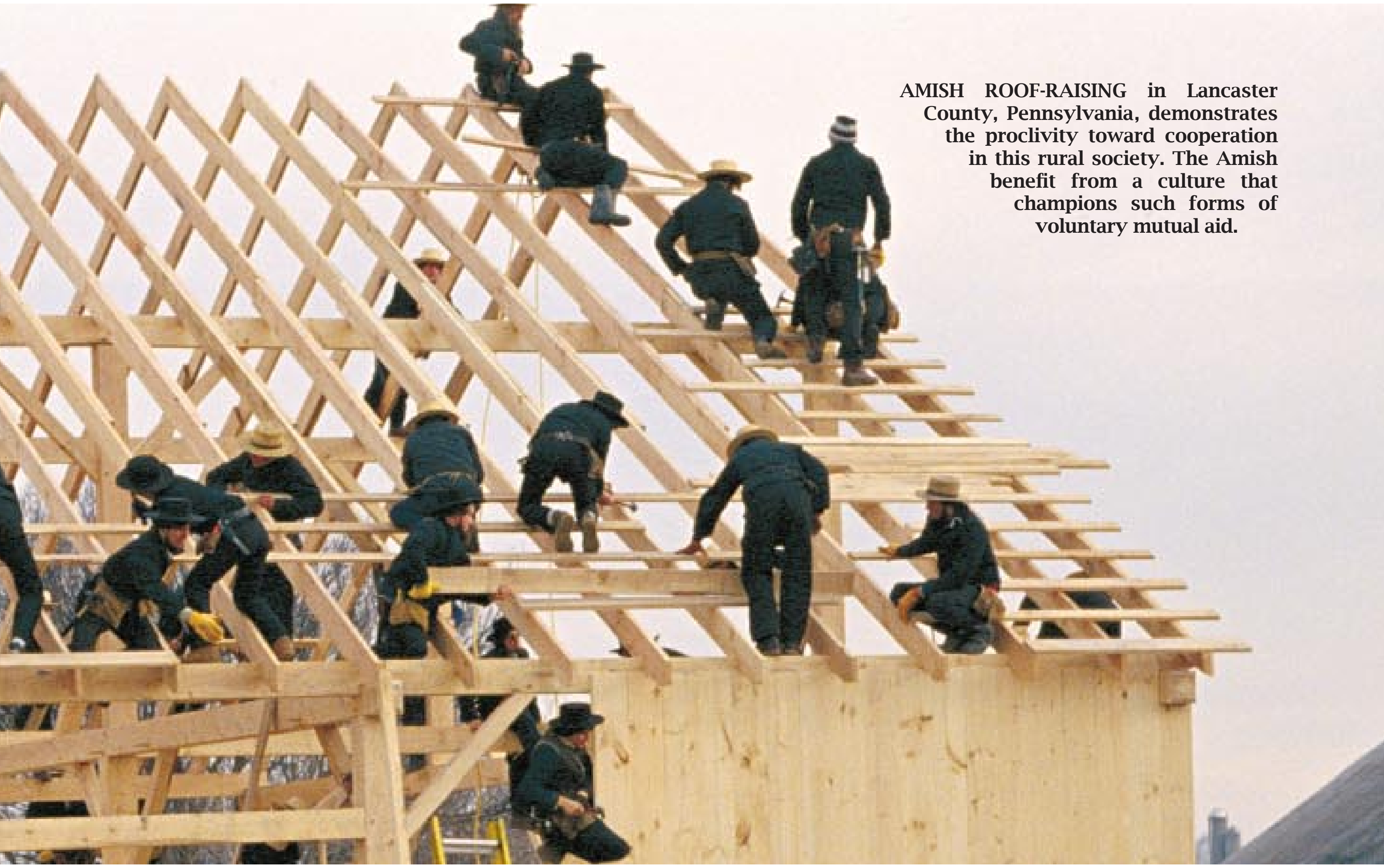
Computer experiments show how cooperation rather than exploitation can dominate in the Darwinian struggle for survival

by Martin A. Nowak, Robert M. May and Karl Sigmund

The principle of give and take pervades our society. It is older than commerce and trade. All members of a household, for example, are engaged in a ceaseless, mostly unconscious bartering of services and goods. Economists have become increasingly fascinated by these exchanges. So have biologists, who have documented many comparable instances in groups of chimpanzees and other primates. Charles Darwin himself was well aware of the role of cooperation in human evolution. In *Descent of Man* he wrote that "the small strength and speed of man, his want of natural weapons, & c., are more than counterbalanced by his...social qualities, which lead him to give and receive aid from his fellow-men."

Obviously, this is a far cry from the savage human existence that the philosopher Thomas Hobbes described as "solitary, poor, nasty, brutish, and short." Nevertheless, a number of Darwin's early followers emphasized the ferocious aspects of the "struggle for survival" to such an extent that the Russian prince Kropotkin felt compelled to write a book to refute them. In *Mutual Aid,* hailed by the London *Times* as "possibly the most important book of the year" (1902), he drew a vast fresco of cooperation acting among Siberian herds, Polynesian islanders and medieval guilds. Kropotkin was a famous ideologue of anarchism, but his dabbling in natural history was no mere hobby: for someone bent on getting rid of the State, it was essential to show that human cooperation was not imposed from an iron-fisted authority but had its origins rooted in natural conditions.

In a way, his arguments have succeeded far beyond what Kropotkin could ever have foreseen. A wealth of studies



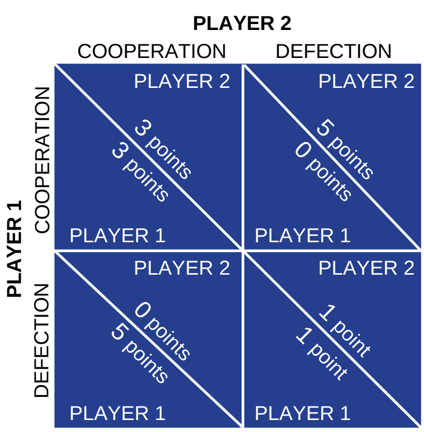
AMISH ROOF-RAISING in Lancaster County, Pennsylvania, demonstrates the proclivity toward cooperation in this rural society. The Amish benefit from a culture that champions such forms of voluntary mutual aid.

RUNK/SCHOENBERGER *Grant Heilman Photography, Inc.*

in anthropology and primatology point to the overwhelming role of reciprocal help in early hominid societies. Textbooks on animal behavior are filled with examples of mutual aid: grooming, feeding, teaching, warning, helping in fights and joint hunting. In ecology, symbiotic associations are increasingly seen as fundamental. Biologists find examples of cooperation at the level of cells, organelles and even prebiotic molecules.

But at the same time, the ubiquity of cooperation seems to have become ever more paradoxical. The Russian anarchist had failed to see how threatened it is by exploitation. What prevents mutualists from turning into parasites? Why should anyone share in a common effort rather than cheat the others? Natural selection puts a premium on individual reproductive success. How can this mechanism shape behavior that is altruistic in the sense that it benefits others at the expense of one's own progeny?

There are two main approaches to this question that go under the headings of kin selection and reciprocal aid. These concepts are not mutually exclusive, but they are sharply distinct. Kin selection is rooted in genetics [see "Kin Recognition," by David W. Pfennig and Paul W. Sherman, page 98]. If a gene helps in promoting the reproductive success of close relatives of its bearer, it helps in promoting copies of itself. Within a family, a good turn is its own reward. But a good turn to an unrelated fellow being has to be returned in order to pay off. Reciprocal aid—the trading of altruistic acts in which benefit exceeds cost—is essentially an economic exchange. It works less directly than kin selection and is therefore more vulnerable to abuse.



LISA BURNETT

VARIABLE PAYOFF applies when one, both or neither player opts to cooperate. Such point assignments generate the classic conundrum of game theory known as the Prisoner's Dilemma.

Two parties can strike a mutually profitable bargain, but each could gain still more by withholding its contribution. In modern society an enormous apparatus of law and enforcement makes the temptation to cheat resistible. But how can reciprocal altruism work in the absence of those authoritarian institutions so despised by Kropotkin's anarchists? This difficult question is best answered by first considering simple, idealized systems.

The Prisoner's Dilemma

To demonstrate the conundrum, Robert L. Trivers, a sociobiologist (and, fittingly, a former lawyer), now at the University of California at Santa Cruz, borrowed a metaphor from game theory known as the Prisoner's Dilemma. As originally conceived in the early 1950s, each of two prisoners is asked whether the other committed a crime; their level of punishment depends on whether one, both or neither indicates the other's guilt. This situation can be viewed as a simple game. The two players engaged in it have only to decide whether they wish to cooperate with each other or not. In one illustration of the Prisoner's Dilemma, if both choose to cooperate, they get a reward of three points each. If both defect (by not cooperating), they get only one point each. But if one player defects and the other cooperates, the defector receives five points, whereas the player who chose to cooperate receives nothing.

Will they cooperate? If the first player defects, the second who cooperates will end up with nothing. Clearly, the second player ought to have defected. In fact, even if the first player cooperates, the second should defect, because this combination gives five points instead of three. No matter what the first player does, the second's best option is to defect. But the first player is in exactly the same position. Hence, both players will choose to defect and receive only one point each. Why didn't they cooperate?

The prisoners' decisions highlight the difference between what is best from an individual's point of view and from that of a collective. This conflict endangers almost every form of cooperation, including trade and mutual aid. The reward for mutual cooperation is higher than the punishment for mutual defection, but a one-sided defection yields a temptation greater than the reward, leaving the exploited cooperator with a loser's payoff that is even worse than the punishment. This ranking—from temptation through reward and punishment down to the loser's payoff—implies that the best move is always to defect, irrespective of the opposing player's move. The logic leads inexorably to mutual defection.

Most people feel uneasy with this conclusion. They do often cooperate, in fact, motivated by feelings of solidarity or selflessness. In business dealings, defection is also relatively rare, perhaps from the pressure of society. Yet such concerns should not affect a game that encapsulates life in a strictly Darwinian sense, where every form of payoff (be it calories, mates or safety from predators) is ultimately converted into a single currency: offspring.

Virtual Tournaments

One can conceive a thought experiment in which an entire population consists of programmed players. Each of these automata is firmly wedded to a fixed strategy and will either always cooperate or always defect. They engage in a round-robin tournament of the Prisoner's Dilemma. For each contestant, the total payoff will depend on the other players encountered and therefore on the composition of the population. A defector will, however, always achieve more than a cooperator would earn in its stead. At the end of the imaginary tournament, the players reproduce, creating progeny of their own kind (defectors or cooperators). The next generation will, again, engage in a round-robin competition and get paid in offspring, and so on. In this caricature of biological evolution, where the payoff is number of offspring and strategies are inherited, the outcome is obvious: defectors will steadily increase from one generation to the next and will eventually swamp the population.

There are several ways to escape from this fate. In many societies the same two individuals interact not just once but frequently. Each participant will think twice about defecting if this move makes the other player defect on the next occasion. So the strategy for the repeated game can change in response to what happened in previous rounds.

In contrast to a single instance of the Prisoner's Dilemma, where it is always better to defect, countless strategies for the repeated version exist, and none serves as a best reply against all opponents. If the opposite player, for instance, decides always to cooperate, then you will do best by always defecting. But if your adversary decides to cooperate until you defect and then never to cooperate again, you will be careful not to spoil your partnership: the temptation to cheat in one round and grab five points instead of three will be more than offset by the expected loss in the subsequent rounds where you cannot hope to earn more than one point.

The absence of a best choice is crucial. There is no hard-and-fast recipe for playing the repeated Prisoner's Dilemma. Success will depend on the other player's strategy, which one does not know beforehand. A strategy that does well in certain environments can fail miserably in others.

In the late 1970s the political scientist Robert Axelrod, at the University of Michigan, conducted round-robin tournaments of the repeated Prisoner's Dilemma on his computer. The contestants—programs submitted by colleagues—were quite sophisticated, but it turned out that the simplest entry ultimately won. This strategy is aptly called Tit-for-Tat. It starts with a cooperative response and then always repeats the opposing player's previous move.

Remarkably, a player applying Tit-for-Tat is never

PRIOR PLAY			STRATEGIES			
Last move	Opponent's move	Outcome	Always Cooperate	Always Defect	Tit-for-Tat	Pavlov
C	C	"REWARD"	C	D	C	C
C	D	"LOSER'S PAYOFF"	C	D	D	D
D	C	"TEMPTATION"	C	D	C	D
D	D	"PUNISHMENT"	C	D	D	C

COOPERATION DEFECTION

ALWAYS COOPERATE	C	C	C	C	C	C	C
TIT-FOR-TAT	C	C	C	D (orange)	C	C	C
ALWAYS DEFECT	D	D	D	D	D	D	D
TIT-FOR-TAT	D	D	D	C (orange)	D	D	D
ALWAYS COOPERATE	C	C	C	C	C	C	C
PAVLOV	C	C	C	D (orange)	D	D	D
ALWAYS DEFECT	D	D	D	D	D	D	D
PAVLOV	C	D	C	C (orange)	D	C	D
TIT-FOR-TAT	C	C	C	C	D	C	D
TIT-FOR-TAT	C	C	C	D (orange)	C	D	C
PAVLOV	C	C	C	C	D	C	C
PAVLOV	C	C	C	D (orange)	D	C	C

TIME →

LISA BURNETT

REACTIVE STRATEGIES for the repeated Prisoner's Dilemma can depend on the outcome of the previous round. Four key strategies of the 16 possible alternatives are shown (*top*). Repeated rounds of the Prisoner's Dilemma reveal persistent patterns of cooperation (*blue*) and defection (*red*) when selected strategies are paired off during successive rounds (*bottom*). An established sequence may recover or alter after an isolated mistaken play (*orange*).

COMMON VAMPIRE BATS frequently engage in acts of mutual help. A bat that feeds successfully on blood from horses or cattle will share its nourishment with an unfed companion by regurgitating a portion of its stomach contents.

ahead at any stage of the repeated game, being always last to defect. The Tit-for-Tat player can nonetheless win the whole tournament, because the Prisoner's Dilemma is not a zero-sum game: it is possible to make points without taking them away from others. By its transparency Tit-for-Tat frequently persuades opponents that it pays to cooperate. In Axelrod's tournaments the Tit-for-Tat strategy (entered by the game theorist Anatol Rapoport) elicited many rewarding rounds of cooperation, whereas other players, among themselves, were apt to get bogged down in long runs of defection.

By winning the round-robin tournament, Tit-for-Tat obtained more representatives among the next generation than did other strategies. Moreover, those players who had cooperated tended also to receive more offspring than those who had not. With each generation Tit-for-Tat shaped a more congenial environment. The strategies that ruthlessly exploited cooperators succeeded only in depleting their own resources.

Unpredictable Adversaries

We recently performed computer simulations with an extended set of strategies that base their next move on the result of the previous round rather than just the opponent's previous move (as does Tit-for-Tat). A strategy based on prior outcome must determine the response for each of four eventualities: temptation, reward, punishment or loss. Two possible responses for each of four prior outcomes give 16 possible types of players.

We further allowed for "stochastic" strategies that respond to the four possible outcomes by changing only their statistical propensity to cooperate. Such strategies are not obliged to respond always in the same way to a given outcome. One form of stochastic player might, for example, cooperate 90 percent of the time after experiencing the reward. Such uncertainty simulates the inevitable mistakes that occur during real interactions.

The addition of stochastic responses resulted in a huge array of possibilities. Our computer searched for the most successful of these players by simulating the forces of natural selection, adding to every hundredth generation some small amount of a new, randomly selected stochastic strategy. We followed many such mutation-selection rounds for millions of generations, not because the emergence of cooperation needed so many iterations but because this span allowed us to test a very large number of possible strategies.

In spite of the rich diversity displayed in these chronicles, they led us invariably to some simple, clear results. The first is that the average payoff in the population can change suddenly. Indeed, the behavior we found is a showpiece for punctuated equilibria in biological evolution. Most of the time, either almost all members of the population cooperate, or almost all defect. The transitions between these two regimes are usually rare and abrupt, taking just a few generations. We found that later in the run, quiescent periods tended to last longer. And there was a definite trend toward cooperation. The longer the system was allowed to evolve, the greater the likelihood for a cooperative regime to blossom. But the threat of a sudden collapse always remained.

Cooperative populations are sometimes dominated by a strategy called Generous Tit-for-Tat, a variant that on random occasions will offer cooperation in response to defection. But much more frequently an altogether different strategy, named Pavlov by the mathematicians David P. Kraines of Duke Uni-

MARTIN A. NOWAK, ROBERT M. MAY and KARL SIGMUND have experienced sundry examples of cooperation and competition in their varied careers. Nowak is Wellcome Trust Senior Research Fellow and Fellow of Keble College at the University of Oxford, where he works in the department of zoology. May is Royal Society Research Professor at the University of Oxford and at Imperial College, London. He is also chairman of the trustees of the Natural History Museum. Sigmund is professor at the Institute of Mathematics at the University of Vienna and also holds a position at the International Institute for Applied Systems Analysis in Laxenburg, Austria. All three work on mathematical models to address a wide range of problems in evolutionary biology. Both May and Sigmund have written several books; Nowak claims only to have read several books.

versity and Vivian Kraines of Meredith College in Raleigh, N.C., reigns paramount. After experiencing a reward for mutual cooperation, a Pavlov player repeats the former cooperative move. After getting away with unilateral defection, it similarly repeats its last move. But after being punished for mutual defection, Pavlov switches to cooperation. And after getting a loser's payoff for unilaterally cooperating, it reacts by defecting. In short, the Pavlov rule says to stick to the former move if it earned a high payoff (reward or temptation) but to switch if the last move brought a low return (punishment or loser's payoff).

This principle of "win-stay, lose-shift" seems to work well in many other situations. In animal psychology it is viewed as fundamental: a rat is ready to repeat an action that brings reward, whereas it will tend to drop behavior that has painful consequences. The same crude application of carrot-and-stick underlies most attempts of bringing up children.

Within the repeated Prisoner's Dilemma game, retaliation after having been exploited is usually seen as evidence for behavior resembling Tit-for-Tat, but it holds as well for Pavlov players. A society of Pavlov strategists is very stable against errors. A mistaken defection between two of its members leads to one round of mutual defection and then back to cooperation. But faced with a player who does not retaliate, a Pavlov player keeps defecting. This behavior makes it difficult for players who always cooperate to gain a foothold in the community. In contrast, a society of Generous Tit-for-Tat players does not discriminate against unconditional cooperators. This beneficence is costly, in the long term, because players who do not retaliate can drift into the population and ultimately undermine cooperation by allowing exploiters to thrive.

Although Pavlov is a good strategy to prevent exploiters from invading a cooperative society, it fares poorly among noncooperators. Against persistent defectors, for instance, it tries every second round to resume cooperation. In Axelrod's tournaments, Pavlov would have ranked close to the bottom. Pavlov's advantages show only after sterner, unyielding strategies such as Tit-for-Tat have paved the way by steering evolution away from defection.

Innate Cooperation

One can safely conclude that the emergence and persistence of cooperative behavior are not at all unlikely, provided the participants meet repeatedly, recognize one another and remember the outcomes of past encounters. These circumstances may seem familiar from daily life in the home or office, but among the larger world of living things, such requirements demand a high degree of sophistication. And yet we observe cooperation even among simple organisms that do not possess such abilities. Furthermore, the strategies discussed will work only if benefits from future encounters are not significantly discounted as compared with present gains. Again this expectation may be reasonable for many of the activities humans conduct, but for most simpler organisms delayed payoffs in the form of future reproductive success may count for little: if life is short and unpredictable, there is scant evolutionary pressure to make long-term investments.

But what of the creatures, such as many invertebrates, that seem to exhibit forms of reciprocal cooperation, even though they often cannot recognize individual players or remember their actions? Or what if future payoffs are heavily discounted? How can altruistic arrangements be established and maintained in these circumstances? One possible solution is that these players find a fixed set of fellow contestants and make sure the game is played largely with them. In general, this selectivity will be hard to attain. But there is one circumstance in which it is not only easy, it is automatic. If the players occupy fixed sites, and if they interact only with close neighbors, there will be no need to recognize and remember, because the other players are fixed by the geometry. Whereas in many of our simulations players always encounter a representative sample of the population, we have also looked specifically at scenarios in which every player interacts only with a few neighbors on a two-dimensional grid. Such "spatial games" are very recent. They give an altogether new twist to the Prisoner's Dilemma.

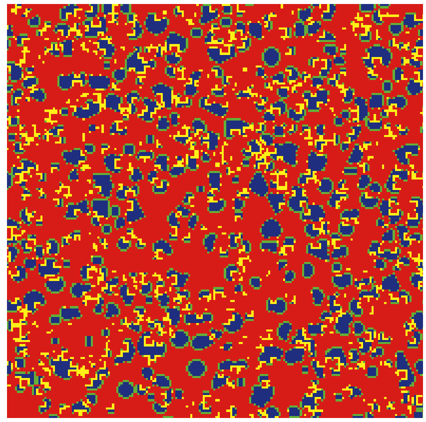
Fixed in Flatland

It should come as no surprise that cooperation is easier to maintain in a sedentary population: defectors can thrive in an anonymous crowd, but mutual aid is frequent among neighbors. That concept is clear enough. But in many cases, territorially structured interactions promote cooperation, even if no follow-up encounter is expected. This result favors cooperation even for the seemingly hopeless single round of the Prisoner's Dilemma.

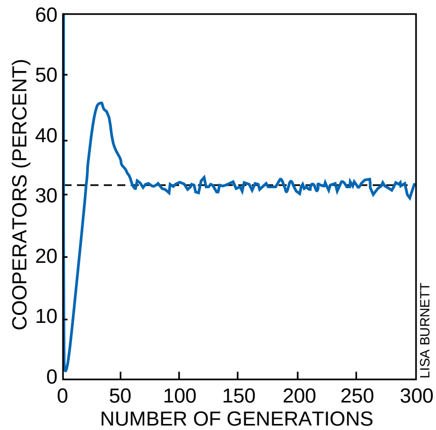
Consider a spatially constrained version of the tournament, with each member of the population sitting on a square of an extended chessboard. Each player is either a pure cooperator or a pure defector and interacts only with the eight immediate neighbors, playing one round of the Prisoner's Dilemma with each. In the next generation the square is inherited by whoever totaled the most points.

A lone cooperator will be exploited by the surrounding defectors and succumb. But four cooperators in a block can conceivably hold their own, because each interacts with three cooperators; a defector, as an outsider, can reach and exploit at most two. If the bonus for cheating is not too large, clusters of cooperators will grow. Conversely, lone defectors will always do well, because they will be surrounded by exploitable cooperators. But by spreading, defectors surround themselves with their like and so diminish their own returns.

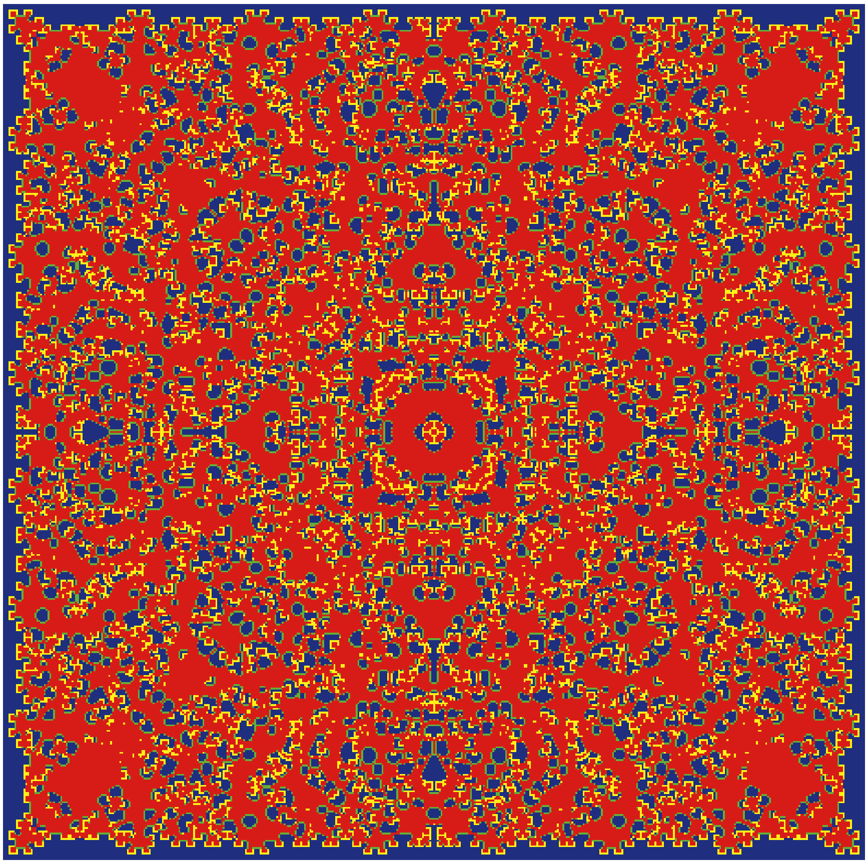
The actual evolution of such spatial systems depends on the payoff values. It is certainly possible that cooperators are wiped off the board. But we frequently find variously shifting mosaics, with both strategies being maintained. Mixtures of pure cooperators and pure defectors can coexist indefinitely, in fluctuating proportions, but the long-



MARTIN A. NOWAK



SPATIAL GAMES of the Prisoner's Dilemma display the evolution of a grid of players, each of which interacts only with opponents in eight adjacent squares. The portion of the population composed of cooperators gradually



achieves a stable value after many generations of play (*bottom left*). In one snapshot taken after 50 generations (*top left*), each blue grid element contains a cooperator that was a cooperator in the previous round; green shows a cooperator that was a defector; red represents a defector that was a defector; yellow indicates a defector that was a cooperator. When the initial conditions are symmetrical, the spatial game can develop a pattern resembling a Persian carpet (*above*).

term average composition of the population is predictable. This conclusion is remarkably robust. In its essentials, it holds true for other choices of grid patterns and even for irregular or random arrays. The important requirement is that each player should not interact with too many neighbors.

The straightforward rules of these spatial games define dynamics of dazzling complexity. They allow for patterns that wander across the board, periodically resuming their former shape. They can also display motifs that grow without limit. Some of these features look like the results of John Horton Conway's game Life [see "Mathematical Games," by Martin Gardner; SCIENTIFIC AMERICAN, October 1970], a scheme to construct evolving spatial patterns using simple rules to mimic regenerating organisms. It may well be that the results generated by any one of our spatial versions of the Prisoner's Dilemma—be they irregular patterns or symmetrical Persian carpets—are intrinsically unpredictable and chaotic in the sense that no algorithm can possibly predict what will occur. Perhaps more clever mathematicians could devise a way to determine the future patterns. We are satisfied to watch the arabesques unfold [see "The Amateur Scientist," by Alun L. Lloyd, page 110].

That's Life

Throughout the evolutionary history of life, cooperation among smaller units led to the emergence of more complex structures, as, for example, the emergence of multicellular creatures from single-celled organisms. In this sense, cooperation becomes as essential for evolution as is competition.

Spatial structures in particular act to protect diversity. They allow cooperators and defectors to exist side by side. In a different but related context, similar spatial patterns allow populations of hosts and parasites, or prey and predators, to survive together, despite the inherent instability of their interactions.

Such cooperative strategies may have been crucial for prebiotic evolution, which many researchers believe may have taken place on surfaces rather than in well-stirred solutions. Catalyzing the replication of a molecule constitutes a form of mutual help; hence, a chain of catalysts, with each link feeding back on itself, would be the earliest instance of mutual aid [see "The Origin of Genetic Information," by Manfred Eigen, William Gardiner, Peter Schuster and Ruthild Winkler-Oswatitsch; SCIENTIFIC AMERICAN, April 1981].

Cooperative chemical reactions would have been vulnerable to "cheating" molecular mutants that took more catalytic aid than they gave. Such difficulties were thought to undercut many ideas about prebiotic evolution based on cooperative chains. But Maarten C. Boerlijst and Pauline Hogeweg of Utrecht University have recently demonstrated with computer simulations that self-generated spatial structures akin to those we devised can hamper the spread of destructive parasitic molecules.

Our models, crude as they are, illustrate how cooperation might arise and be maintained in real biological systems. Sophisticated creatures may be drawn to follow strategies that encourage cooperation because of repeated interactions among individuals who can recognize and remember one another. But in simpler organisms, cooperation persists, perhaps by virtue of self-organized spatial structures generated by interactions with immediate neighbors in some fixed spatial array. In the course of evolution, there appears to have been ample opportunity for cooperation to have assisted everything from humans to molecules. In a sense, cooperation could be older than life itself.

FURTHER READING

THE EVOLUTION OF COOPERATION. Robert Axelrod. Basic Books, 1985.

TIT-FOR-TAT IN HETEROGENEOUS POPULATIONS. Martin A. Nowak and Karl Sigmund in *Nature,* Vol. 355, No. 6357, pages 250–253; January 16, 1992.

EVOLUTIONARY GAMES AND SPATIAL CHAOS. Martin A. Nowak and Robert M. May in *Nature,* Vol. 359, No. 6398, pages 826–829; October 29, 1992.

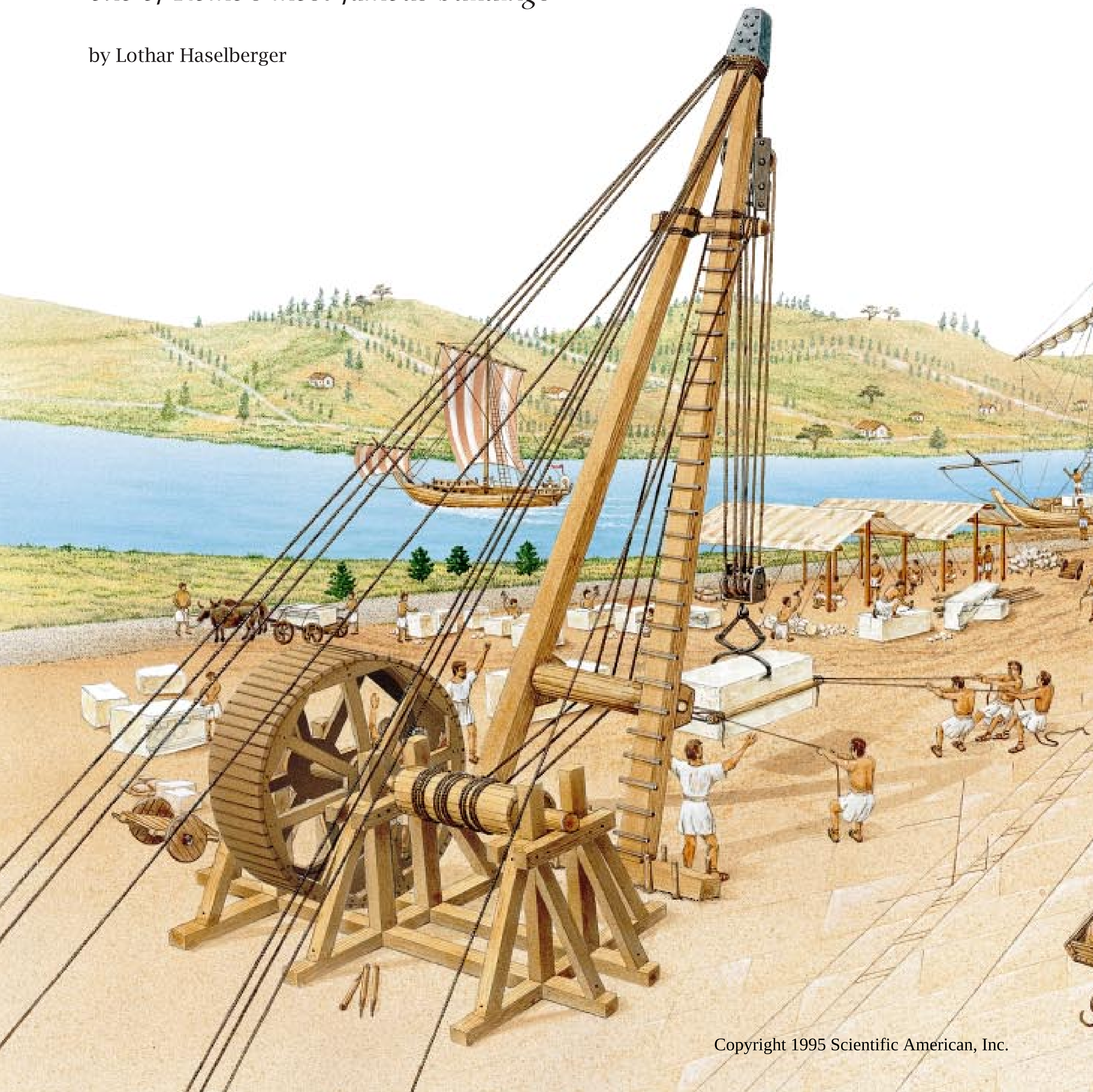
A STRATEGY OF WIN-STAY, LOSE-SHIFT THAT OUTPERFORMS TIT-FOR-TAT IN THE PRISONER'S DILEMMA GAME. Martin A. Nowak and Karl Sigmund in *Nature,* Vol. 364, No. 6432, pages 56–58; July 1, 1993.

GAMES OF LIFE: EXPLORATIONS IN ECOLOGY, EVOLUTION, AND BEHAVIOUR. Karl Sigmund. Penguin, 1995.

Deciphering a Roman Blueprint

Scholarly detective work reveals the secret of a full-size drawing chiseled into an ancient pavement. The "blueprint" describes one of Rome's most famous buildings

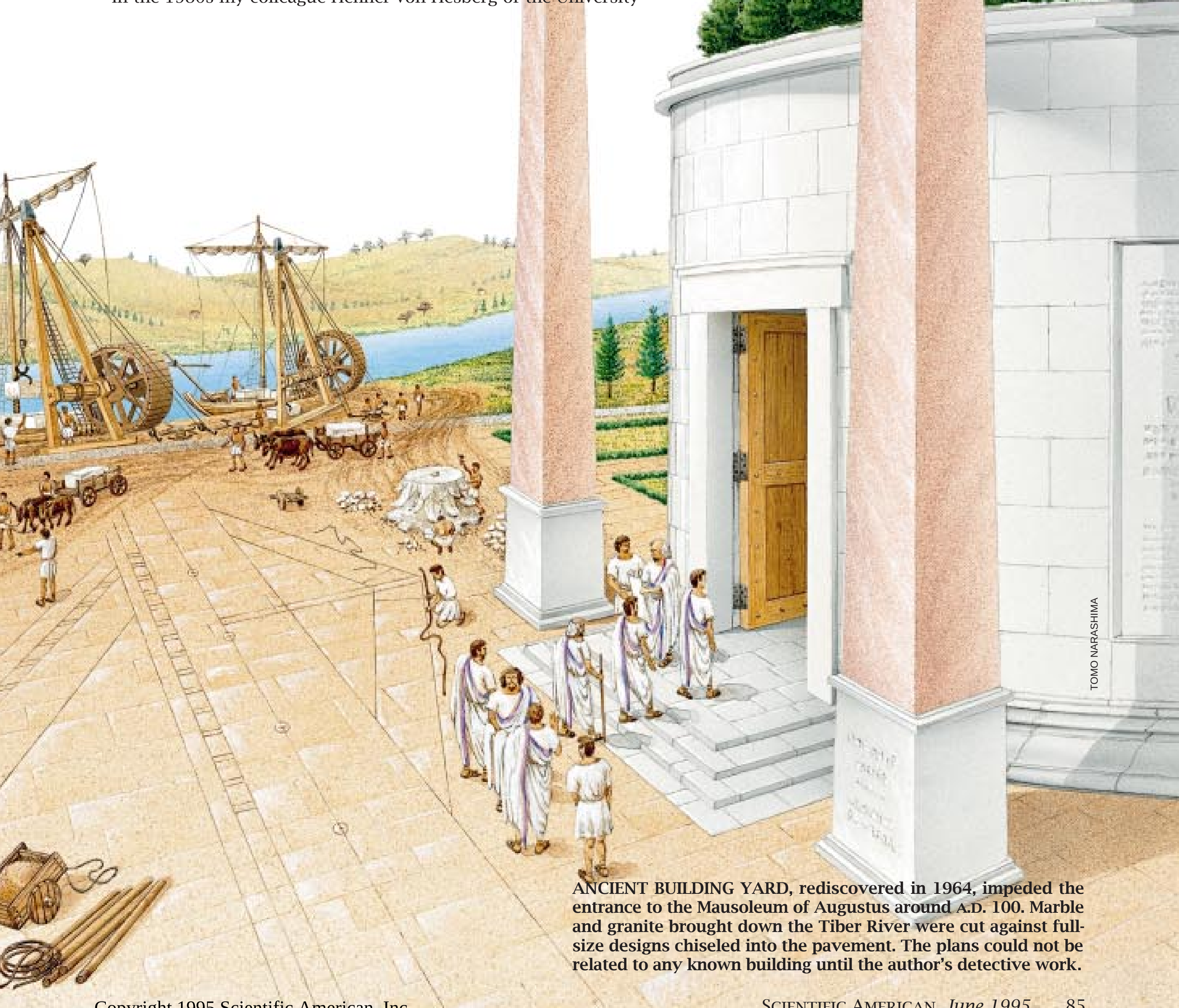
by Lothar Haselberger



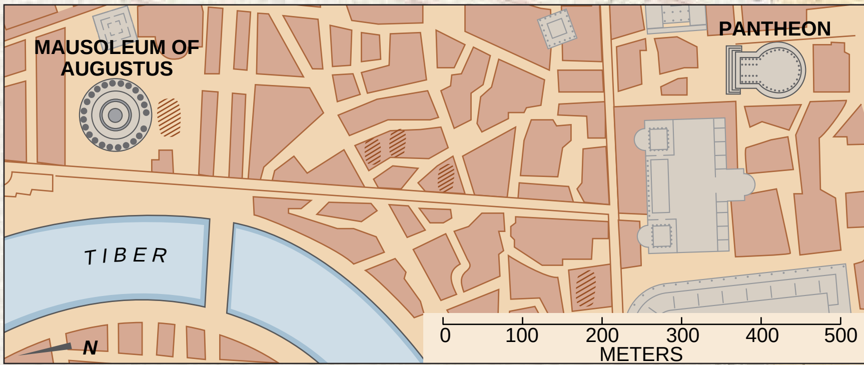
In the 1930s, hoping to energize Italy with memories of its glorious past, Benito Mussolini ordered extensive and hasty archaeological excavations in Rome. Over the centuries, annual floods of the Tiber River had deposited soil all over the city. In particular, the mausoleum erected by Augustus around A.D. 30 to demonstrate his power had acquired a layer of earth many meters deep outside its walls. Digging through six or seven meters of it, the workers hit a pavement made of massive blocks of rock.

When World War II started, the discovery was forgotten. But in 1964 the Roman Service of Antiquities observed that chiseled into this pavement were full-size architectural plans for "a pediment and an inverted capital." The courtyard had apparently served as a workshop for an ancient building project. Used as a drawing board, the surface was inscribed with a series of designs of new monuments, so that the plans are superimposed in no discernible order.

In the 1980s my colleague Henner von Hesberg of the University

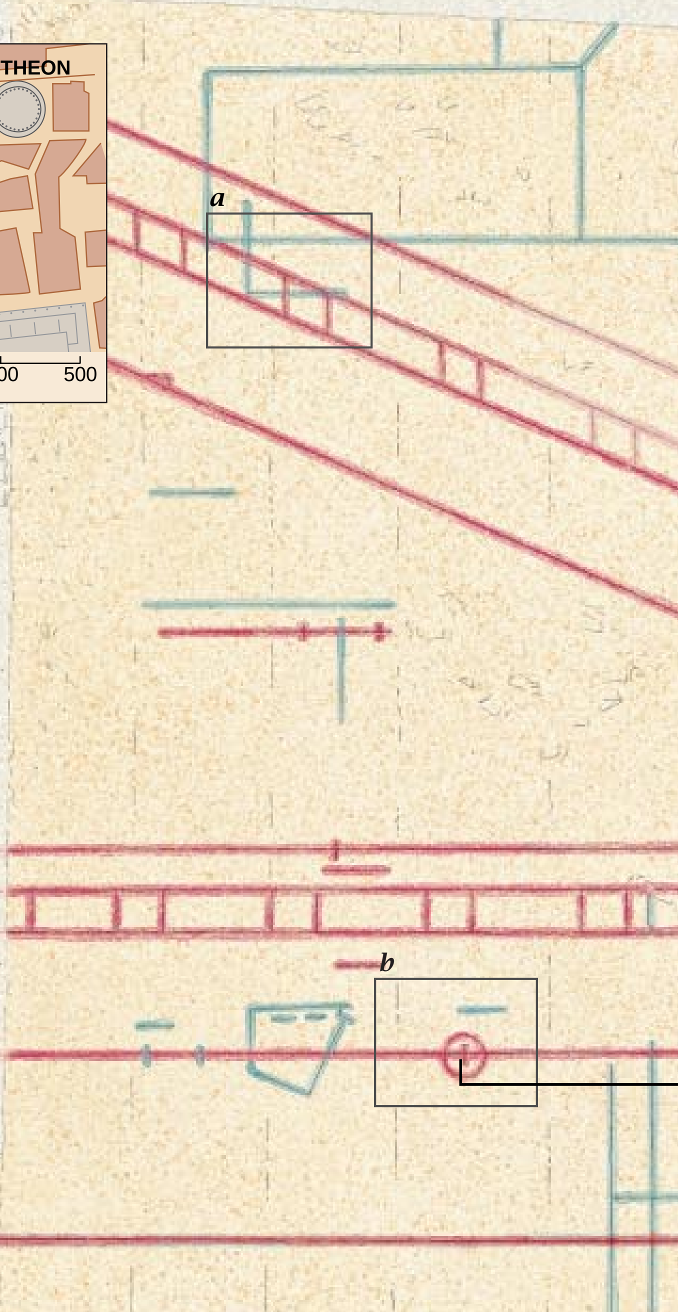


ANCIENT BUILDING YARD, rediscovered in 1964, impeded the entrance to the Mausoleum of Augustus around A.D. 100. Marble and granite brought down the Tiber River were cut against full-size designs chiseled into the pavement. The plans could not be related to any known building until the author's detective work.



IAN WORPOLE

MAP OF ROME shows the Tiber River flowing past the Mausoleum of Augustus and several marble workshops (*hatched areas*). The Pantheon lies 800 meters south of the mausoleum. These monuments originally marked the city's northern outskirts, known as Campo Marzio.



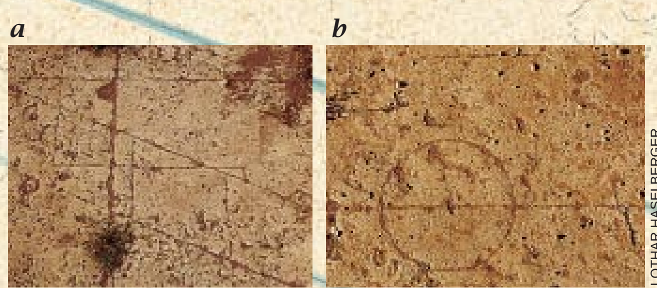
of Cologne brought the drawings to my attention. Not one but two of the incised outlines were of pediments, which crown the entrances to monuments, and von Hesberg hoped that from these I would be able to reconstruct a possible front porch that might have been built for the mausoleum. But when I sketched the drawings for myself, I became quite despondent. Neither of the pediments could be related to the mausoleum: they were simply too big.

The topmost step at the mausoleum's entrance is about six meters wide. But the "elevation," or architectural drawing as seen from the front, for the smaller pediment spans 17 meters. Intersecting with this drawing are the curved lines of a Corinthian capital (which, placed on a column, would support a roof). The uppermost molding, or abacus, of this capital is 2.8 meters wide—far surpassing the abaci of most extant buildings. The second elevation is drawn using the same baseline as the first but on its opposite side. The pitch of its triangular corner, at 24 degrees, is rather high: the Romans usually kept close to the Greek norm of about 14 degrees. The fragment that is exposed shows the pediment to be at least 18 meters wide and probably much wider.

The sheer size of the drawings makes it evident that they were laid out at full scale. The workmen probably used them to build the structures they represent, measuring blocks of marble directly against their lines. Blueprints of this kind are known from the Temple of Apollo at Didyma (in modern Turkey) where they cover the unfinished inner walls [see "The Construction Plans for the Temple of Apollo at Didyma," by Lothar Haselberger; SCIENTIFIC AMERICAN, December 1985]. A fine 17th-century example occurs in Rome itself. The plans for Francesco Borromini's bell towers—which once embellished the Pantheon but, reviled as "donkey's ears," were removed in the late 19th century—are detailed on top of the cornice slabs encircling the Pantheon's dome.

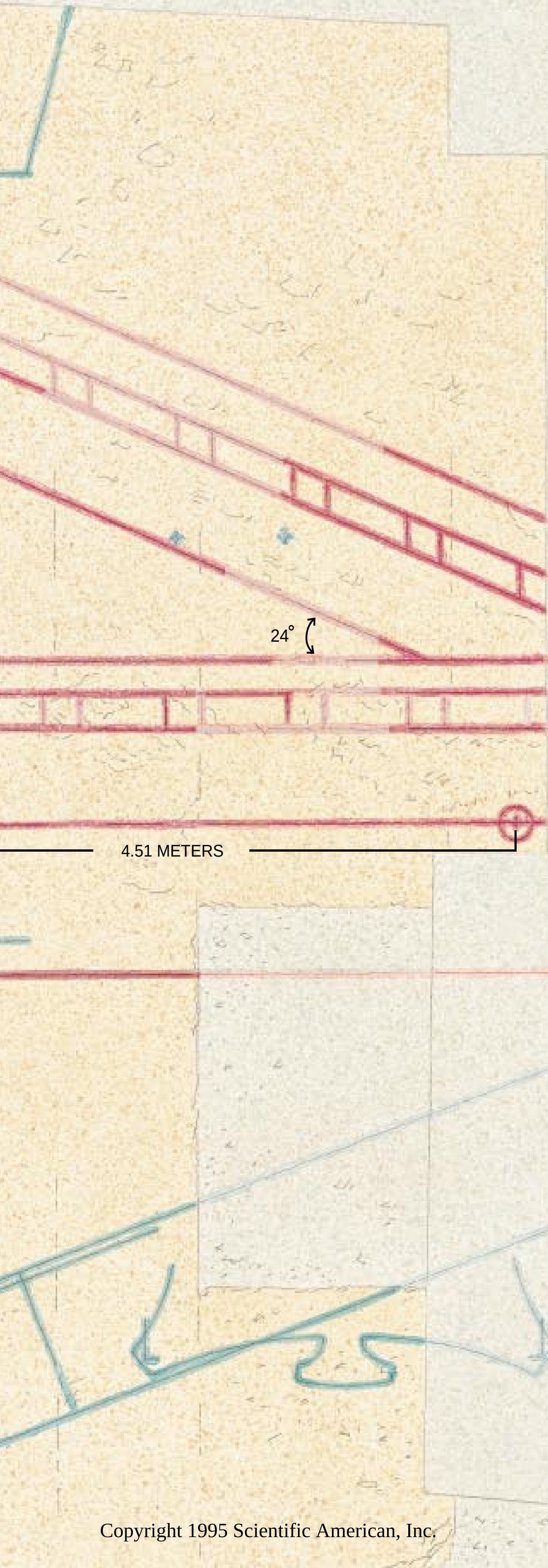
But what monumental buildings did the blueprints at the mausoleum's entrance describe? Trying to find a match seemed like hunting for the proverbial needle in a haystack. No existing nearby structure was big enough. And looking at the wider cityscape, how many parts of ancient Rome were not once ornamented by pediments of imperial size? I do not believe in the oft-used procedure of trying to connect every new finding to the few structures known from classical times. Such results give a seemingly well-rounded impression but for the most part remain speculative.

As it turned out, my pessimism was



LOTHAR HASELBERGER

PATTERNS cut into the mausoleum's pavement turn out to be the plans (*red* and *blue*) of two pediments. These facades, drawn above and below the same horizontal baselines, have been extrapolated (*lighter lines*) wherever they are obscured by earth. Details of the larger drawing (*red*) correspond to various edges of the pediment (*parallel lines in a*), the borders of its consoles (*vertical hatches*) and the centers of its supporting columns (*circled mark in b*). The smaller pediment, the curved capital (*right*) and the trapezoid (*center*) do not correspond to any known structures.



LOTHAR HASELBERGER holds the Morris Russell Williams and Josephine Chidsey Williams Chair in Roman Architecture at the University of Pennsylvania. Trained at the Technical University of Munich and at Harvard University, Haselberger works primarily on the Temple of Apollo at Didyma, Turkey. He is the academic director for its restoration, a project undertaken by the German Archaeological Institute in Berlin.

misplaced. One thing was already clear at the time. Although ancient—as evinced by the six meters of earth under which they lay—the blueprints were from a time later than that of Augustus. The pavement abuts the staircase to the mausoleum, lying at the same level as the lowest visible marble step. Because a stair is not normally built flush with the ground, I saw that the inscribed pavement could not be the original one. Indeed, the excavation reports from 1938 revealed that the wall base of the mausoleum lies about 1.7 meters underneath this pavement.

As the Tiber unendingly deposited its soil, the mausoleum must have acquired at least three pavements—the original one, still buried; the "drawing board," 1.7 meters above; and a third, of modern cobblestone, that was dug up in the 1930s. The inscribed courtyard has the same level as other pavements built in this area, Campo Marzio, around A.D. 100.

This clearly post-Augustan date raised another question. The mausoleum, surrounded by carefully designed parks and ornamented with obelisks celebrating the Divine Augustus, was a grandiose dynastic monument. If not Augustus as the patron of his tomb, who could have taken the liberty to use the area in front as a building yard? There can be no doubt that this was imperial property. One of the later emperors himself, or somebody with his consent, must have assigned it to the design office.

That is where I left the problem in 1985. In the spring of 1992 I gave a seminar on the Pantheon, a temple that has been called the pinnacle of Roman architecture. One night that summer, as I walked home along a sylvan creek, my thoughts turned to the plans that had been gathering dust in my drawers. Nonsensical as it seemed, the designs certainly resembled the Pantheon's facade. "I should make a detailed comparison," I told my wife over dinner, "even if just for fun."

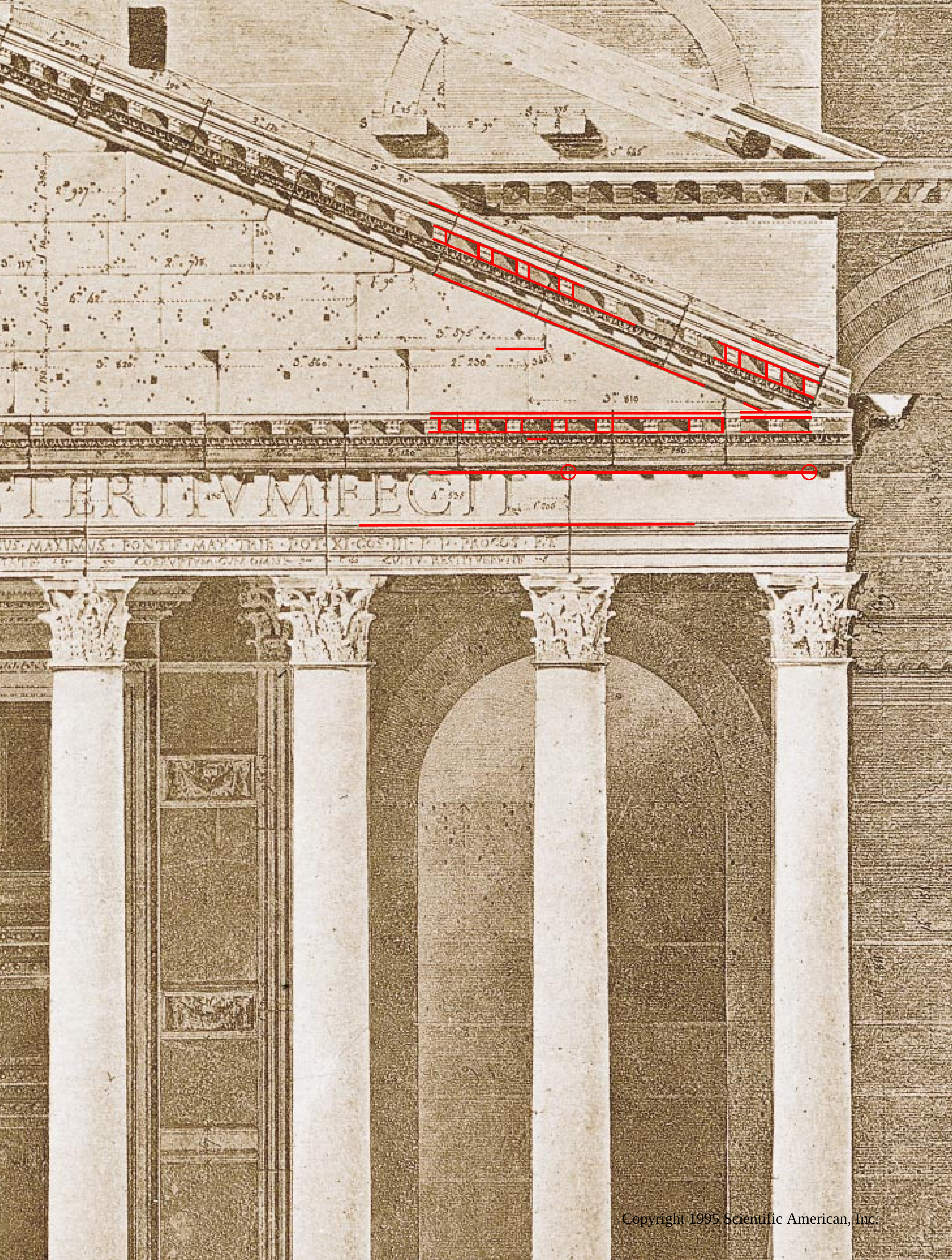
But try as I might, I could not find an accurate drawing of the Pantheon's pediment. After some effort I located a measured drawing sketched in 1813 by the French architect Achille Leclère. But it did not list the measurements with sufficient precision, so I searched further. Slowly, from a British volume of 1821, an American study of 1924 and a Danish work of 1968, I gleaned data on diverse aspects of the facade. When I finally pieced together the basic dimensions, I was thrilled. It showed an amazing match with the larger elevation.

The Danish study lists the unusually steep pitch of the pediment at "ca. 23 degrees"; the working drawing, we have noted, has 24 degrees. The consoles, or brackets, holding up the top and bottom cornices, as well as the height of a horizontal frieze, fit the blueprint to within one centimeter. (The average distance between the consoles is about 82 centimeters; the elevation has 81 centimeters.) The distances between the massive columns average, from center to center, 4.52 meters; the drawing shows 4.51 meters. The positions of the two columns nearest the corner coincide with two circled marks in the elevation. Even a horizontal line where two rows of blocks meet is indicated in the working drawing.

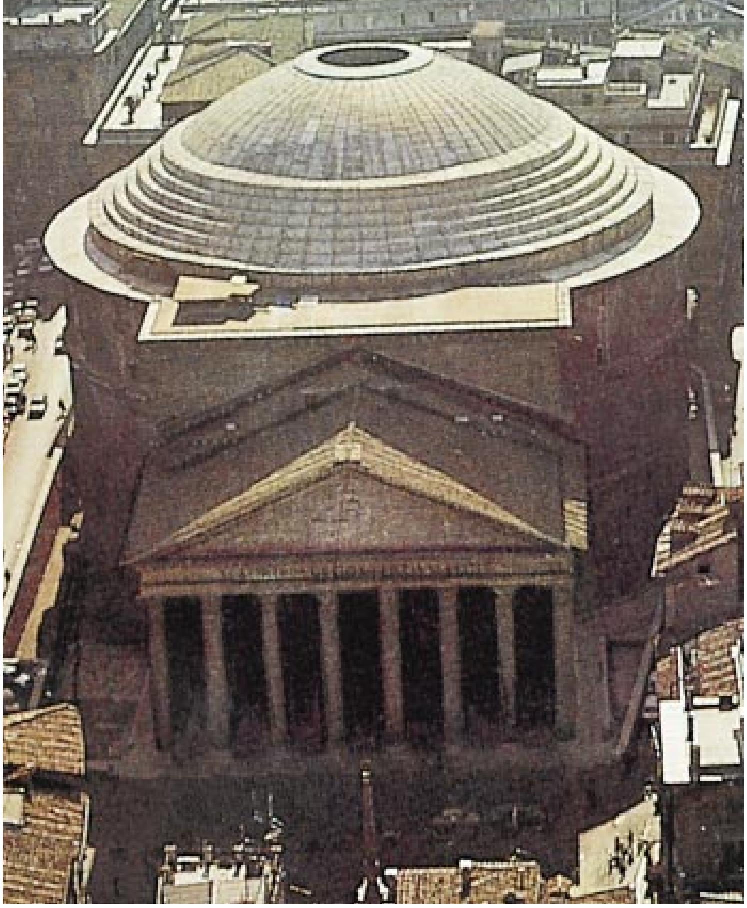
The plan was clearly used to construct the imposing facade of the Pantheon. The Pantheon was commissioned either by the emperor Trajan or his successor, Hadrian, who finished it around A.D. 120. (Of an earlier Pantheon at the same site, only some foundation blocks have been found so far.) Dedicated to Christianity in A.D. 609, the Pantheon remains a living temple, where thousands worship and Italy's greatest citizens—including the painter Raphael—are buried.

The imperial building yard on which the Pantheon's blocks were sliced—the mausoleum's entrance—lies 800 meters from the Pantheon's site. This inconvenient distance must have been offset by the presence of the Tiber River, along which the marble arrived by boat. The blocks were apparently offloaded, cut and then carted to the construc-

TOMO NARASHIMA


TERTIVM FECIT

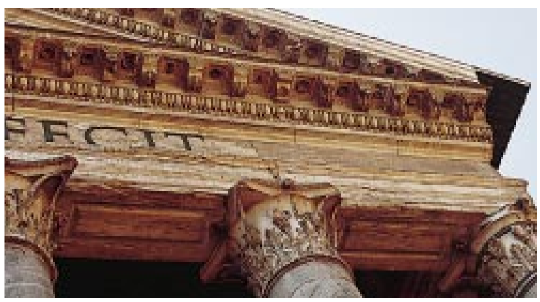
JEFF FADELLIN (*photographed at Fisher Fine Arts Library, University of Pennsylvania*)



LOTHAR HASELBERGER

DRAWING of the Pantheon (*far left*) by the architect Achille Leclère gives the most precise dimensions measured to date for its facade. The larger elevation from the mausoleum's pavement (*red*) is superimposed on this image from 1813. The steep pitch of the pediment matches the blueprint to within a degree; the rims of the cornices, the positions of the columns and consoles and even the edge of a line of blocks in the pediment fit to within a centimeter. To check whether the elevation fits even more precisely with the Pantheon itself (*left*), researchers will have to remeasure the facade (*below*).

tion site. Transporting rocks that weighed tons, it must also be said, posed little problem to Roman technology. The many obelisks that adorn Rome—single blocks of granite up to 33 meters high—came from Egypt. The Pantheon is ornamented with precious and colorful stones that came from distant corners of the empire and even form a catalogue of its extent.



LOTHAR HASELBERGER

In addition to the logistics of Roman architecture, the elevation reveals its aesthetics. The three elementary dimensions of the Pantheon's facade—diameter of the columns (1.495 meters plus or minus two centimeters); "clear distance," or space between the columns (3.02 meters plus or minus three centimeters); and height of the columns (14.14 meters plus or minus one centimeter)—are in the ratio of 1 to 2 to 9.5.

This formula is among those described by Hermogenes, one of the most celebrated architects of the Hellenistic age, as belonging to an ideal facade. (His ideas are brought to us by Vitruvius, the engineer and theoretician from the first century B.C.) For the first time in the exploration of Roman architecture, a numerical recipe for beauty, as known from a textual source, can be tied to an existing monument. Each time they visit the Mausoleum of Augustus, tourists and scholars trample on this record of the Pantheon's exquisite proportions.

FURTHER READING

THE ROTUNDA IN ROME—A STUDY OF HADRIAN'S PANTHEON. Kjeld de Fine Licht. Jutland Archaeological Society Publications VIII. Nordisk Forlag, Copenhagen, 1968.

EIN GIEBELRISS DER VORHALLE DES PANTHEON: DIE WERKRISSE VOR DEM AUGUSTUSMAUSOLEUM. Lothar Haselberger in *Mitteilungen des Deutschen Archaeologischen Instituts, Roemische Abteilung,* Vol. 101. Verlag Philipp von Zabern, Mainz, 1994.

Halo Nuclei

Nuclei having excess neutrons or protons teeter on the edges of nuclear stability, known as drip lines. Under this stress, some develop a halo

by Sam M. Austin and George F. Bertsch

For the past 50 years physicists have pictured the atomic nucleus—made of protons and neutrons—as a liquid drop that has a well-defined surface. But this is not always so. Researchers at a handful of laboratories have now witnessed an entirely novel structure: in certain nuclei, some of the constituent neutrons or protons will venture beyond the drop's surface and form a misty cloud, or halo, in much the same way that electrons form clouds around nuclei and make atoms. Not surprisingly, these extended nuclei behave very differently from ordinary ones. Normal nuclei are difficult to excite or break apart, but halo nuclei are fragile objects. They are larger than normal nuclei and interact with them more easily as well. In fact, the halo is a quantum phenomenon that does not obey the laws of classical physics. Thus, halo nuclei may well yield fresh insight into one of the central mysteries of physics, namely, that of nuclear binding.

Indeed, physicists have long puzzled over the possible combinations of neutrons and protons, or nucleons, that will stay together as a nucleus. This balance depends in rather subtle ways on how many neutrons and protons are involved and the forces acting among them. All nucleons attract one another, but only protons and neutrons can bind to each other in pairs, called deuterons. As a result, only those nuclei that contain roughly equal numbers of neutrons and protons are stable enough to occur naturally on the earth.

Nuclei having unequal numbers of neutrons and protons exist as well, but their lifetimes are limited. Although they are bound—meaning it takes energy to remove one of their nucleons—they are not stable. Beta radioactivity can change them into a more stable species by transforming some of their neutrons into protons, or vice versa. Some of these transitions take place within milliseconds and others only after millions of years. But in general, if the nuclei are displayed on a graph so that the number of protons lies along one axis and the number of neutrons lies along the other, those farther away from the diagonal have shorter lifetimes [*see illustration on opposite page*].

At a certain distance from this diagonal—both above it and below it—the nuclei break up just as quickly as they form. No truly bound nuclei can exist beyond these borders, termed drip lines. The most exotic nuclei are those that lie just within the drip lines, on the edges of nuclear stability. Such extreme systems appear only in far more hostile environments than our own. They result from those reactions that synthesized the heavy elements in the universe and now power stellar explosions in novae, supernovae and x-ray bursters. Astrophysicists think that nuclei along the lower drip line are found in the crust of neutron stars.

SAM M. AUSTIN and GEORGE F. BERTSCH have been long-term colleagues at Michigan State University, studying nuclear physics from the complementary points of view as experimentalist and theorist, respectively. Austin obtained his Ph.D. from the University of Wisconsin in 1960 and has been at Michigan State since 1965, where he is now Distinguished Professor of Physics. Bertsch received his Ph.D. from Princeton University in 1965 and is now on the physics faculty at the University of Washington.

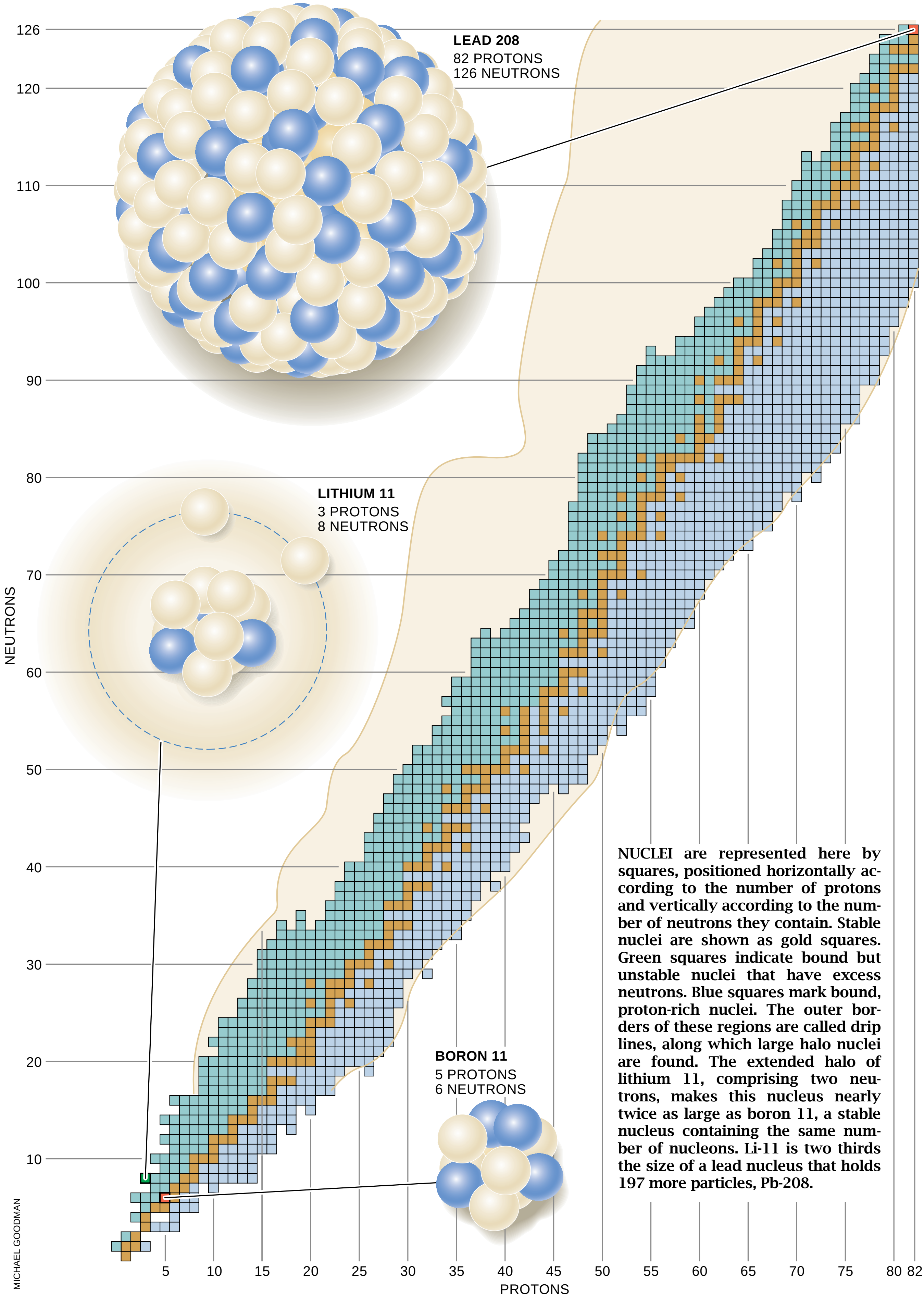
Early Evidence of Neutron Halos

Until a decade ago, physicists had few means for studying such nuclei. Then Isao Tanihata and his collaborators at Lawrence Berkeley Laboratory developed a technique for observing how unstable nuclei interact with other nuclei. This method has led to the discovery of halos in a variety of nuclei. To date, the most scrutinized halo nucleus is an isotope of lithium, Li-11, which has three protons and eight neutrons. Analyses of its fleeting structure have revealed a great deal about the surprising nature of halos in general.

Workers at Lawrence Berkeley Laboratory first discovered Li-11 in 1966, but not until more than a decade later did its unusual structure become evident. In 1985 Tanihata tried to measure its size. He collided ordinary nuclei at high energies to produce a beam of unstable isotopes in a process called projectile fragmentation. Next he placed a carbon foil in the beam. He then counted how many of the beam nuclei survived passage through the foil. This number reflects how likely was the chance of their interacting with nuclei in the target foil. Physicists express this probability by a measure called the cross section. Tanihata found that Li-11 nuclei had particularly large cross sections. The explanation that emerged was that the nuclei bore halos. Two neutrons in the Li-11 nucleus were bound so weakly that they roamed well beyond the core, where they were easily stripped away by the target.

It was an astonishing find. According to the laws of classical physics, a bound particle must stay within range of the core's forces. But in quantum mechanics, a remarkable effect called tunneling makes halos possible. To visualize this phenomenon, imagine a skateboarder in a trough-shaped arena [*see illustration on page 92*]. His total energy limits the distance he travels: the more energy he has, the higher he will go. He cannot rise any higher than the amount of energy he puts into his movement. In quantum mechanics the confinement is not so strict; even a lazy skateboarder will occasionally pop out of the arena. The amount of time he can spend there is limited, and it is related by Heisenberg's uncertainty principle to the extra energy he would need to get out. The lower the energy cost, the longer he can stay outside.

For an object as large as a person on a skateboard, the probability that tunneling will happen is unimaginably small, but on atomic and nuclear scales



NUCLEI are represented here by squares, positioned horizontally according to the number of protons and vertically according to the number of neutrons they contain. Stable nuclei are shown as gold squares. Green squares indicate bound but unstable nuclei that have excess neutrons. Blue squares mark bound, proton-rich nuclei. The outer borders of these regions are called drip lines, along which large halo nuclei are found. The extended halo of lithium 11, comprising two neutrons, makes this nucleus nearly twice as large as boron 11, a stable nucleus containing the same number of nucleons. Li-11 is two thirds the size of a lead nucleus that holds 197 more particles, Pb-208.

the effect can be significant. As Tanihata observed, the effect in Li-11 is dramatic. The last two neutrons are bound by only a few hundred thousand electron volts, more than an order of magnitude smaller than normal binding energy. Consequently, these neutrons need very little energy to move away from the nucleus. They can remain there a relatively long time, spreading out and forming a tenuous halo. Indeed, compared in size with other nuclei, the average distance of Li-11's halo from its center measures about five femtometers, or more than double the normal distance for a nucleus of its mass [*see illustration on preceding page*].

Further work revealed that the Li-11 nucleus was highly unusual in other ways. The isotope Li-10, which would contain one fewer neutron, is unbound, meaning that its three protons and seven neutrons will not hold together as a nucleus. If one neutron is taken away from Li-11, a second neutron will come out immediately as well, leaving behind Li-9. Thus, Li-9 and the two neutrons are bound as a three-body system that comes apart if any one particle is taken away. For this reason, Mikhail Zhukov of the University of Göteborg in Sweden called Li-11 a Borromean nucleus; it resembles the heraldic symbol of the Italian princes of Borromeo. Their crest has three rings interlocked in such a way that if any one ring is removed the other two separate. There are half a dozen other known examples of Borromean nuclei.

Characterizing Halos

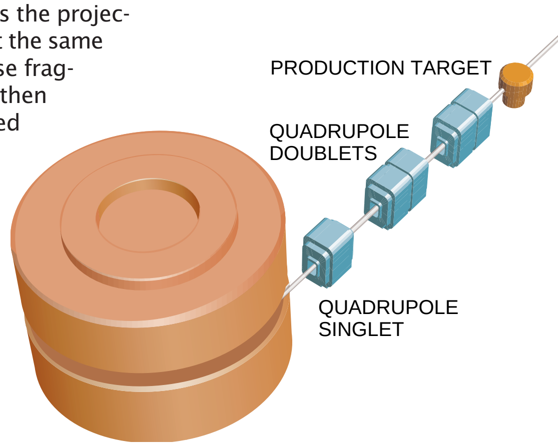
Rainer Neugart and his co-workers at CERN, the European laboratory for particle physics near Geneva, investigated the interaction between Li-11's three components (the two halo neutrons and the Li-9 core), testing specifically whether the halo had any effect on the core. They measured the isotope's magnetic and electrical properties in a clever way and found they matched those of the Li-9 nucleus [*see top box on pages 94 and 95*]. Because the halo neutrons carry no charge—and as a pair they have no spin or magnetic moment—this result supported the notion that the Li-9 core and the two-neutron halo are nearly independent objects.

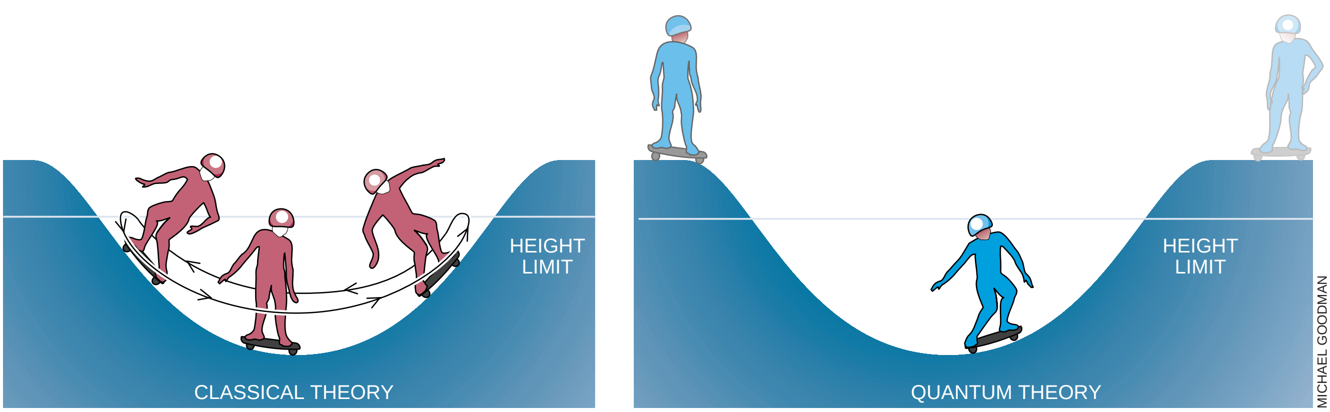
Given this information, experimentalists next hoped to learn how the individual nucleons in Li-11 nuclei were arranged. To find out, Toshio Kobayashi and his collaborators at Lawrence Berkeley Laboratory examined the momentum distributions of Li-11 nuclei. Their motion under the laws of quantum mechanics satisfies yet another relation that is part of Heisenberg's uncertainty principle. It states that particles cannot have a precise momentum but will have a range of momenta depending on how they are distributed in space, as reflected by their wave function. The more spread out and smooth the wave function, the more definite a particle's momentum. So if a halo spans a large distance and a target breaks it away from its core, the momenta of the separated neutrons deviate little from their initial

Making Exotic Nuclei

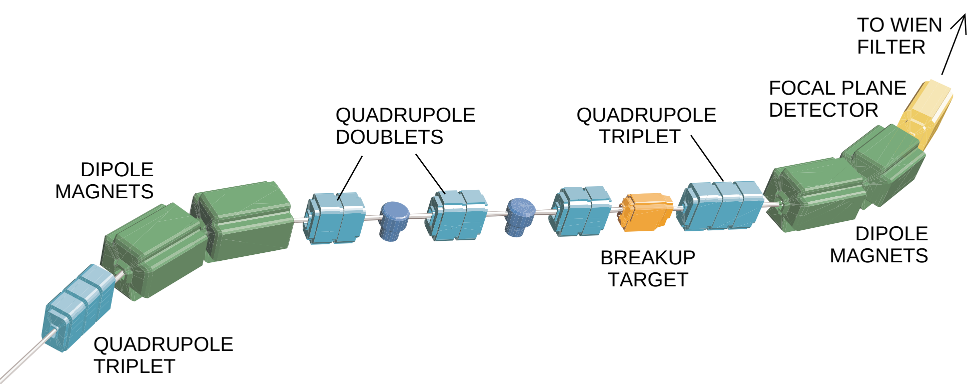
During the past decade, experimentalists have developed two fundamentally different approaches for studying halo nuclei. Some examine the fragments of target nuclei, whereas others analyze the fragments of projectiles bombarding a production target. In the first strategy, the interesting isotopes must be extracted from the target material. If an element is volatile, its isotopes will diffuse out of the target when it is heated. These isotopes can then be ionized and separated. This technique is called ISOL (isotope separation on line). The lifetime of Li-11 was first measured using ISOLDE, the ISOL-type laboratory at CERN. New facilities are currently under construction at Oak Ridge National Laboratory in Tennessee and at several other laboratories around the world.

In the other tactic, the target breaks the projectile nuclei into fragments that move at the same high speed as did the projectile. Those fragments having unusual properties are then studied. Paradoxically, the high speed of the beam makes it easier to study many nuclei, especially short-lived isotopes such as Li-11 (its half-life is only nine milliseconds). Laboratories known by their abbreviations—including GANIL in Caen, France, the NSCL at Michigan State University, RIKEN near Tokyo and the GSI near Darmstadt, Germany—have built apparatus of this kind to work with radio-





CLASSICAL AND QUANTUM PHYSICS are contrasted in these pictures of a skateboarder. In the classical theory (*left*), the skateboarder reaches a definite height on each side of the arena, depending on how much energy he has. Under quantum laws, depicted on the right, he may move farther, past the limit imposed by his energy, as indicated by fainter figures.



active beams and to study unstable nuclei.

In 1990 Bradley M. Sherrill and his associates built the Michigan State fragment separator, called the A1200. It filters out exotic nuclei by subjecting the fragment beam to various forces (*left*). Dipole magnets bend the beam according to the momenta and charges of the beam nuclei; doublet and triplet quadruple magnets focus the beam.

The beam can also be sent through a thin slab, which slows the nuclei by different amounts depending on their velocities and charges. In addition, the beam can be diverted to a Wien filter, a device that produces perpendicular electrical and magnetic fields; only nuclei of a chosen velocity pass through the filter. Finally, it is sometimes possible to measure the time a nucleus takes to pass through the separator, giving yet another measurement of its velocity. Armed with all this information, researchers have identified the individual nuclei passing through and measured their velocities and momenta as well.

momenta. They will travel nearly straight forward and at nearly the same velocity.

Kobayashi and his team took a slightly indirect approach to infer the halo's momentum. They produced reactions in which the halo neutrons were stripped from Li-11 and then observed the Li-9 core that traveled forward. Because Li-11's initial momentum is fixed, the spread in the core momentum had to match the spread in the neutron momentum. Using this relation, the investigators found that the momentum distribution was exceedingly narrow, about one fifth of that measured during the breakup of normal nuclei.

Later experiments at GANIL in Caen, France, led by Alex C. Mueller, gauged the deflection of the neutrons themselves rather than the core. Under these conditions, the neutrons from halo nuclei went forward in a cone about two degrees wide, whereas neutrons from ordinary nuclei came out in a cone some 10 degrees wide. Unfortunately, it was somewhat difficult to interpret these experiments quantitatively because elastic forces from the target had also deflected the particles.

A team at Michigan State University, consisting of Bradley M. Sherrill, Nigel A. Orr and one of us (Austin), found a way around this limitation. Elastic forces deflect the particles mainly sideways and hardly change the momentum component parallel to the beam direction. We realized that the influence of the halo would be clearest if we could measure the spread in the parallel momentum, but the beam of Li-11 we were using already had a momentum spread 10 times larger than the effect to be measured. Fortunately, the fragment separator at Michigan State, the A1200, allows an experimenter to disperse the beams and focus the particles on spots according to how much their momenta has changed rather than on their ultimate momenta [*see box on this and opposite page*]. In this way, the separator can single out the changes in momenta caused by the breakup.

Using this so-called energy-loss mode of operation, the Michigan State workers obtained a resolution much smaller than the width of momentum distribution they wished to measure. A beam of Li-11 struck a variety of targets, ranging in mass from beryllium to uranium, placed near the center of the device one at a time. The Li-9 nuclei resulting from these breakups showed a narrow momentum distribution; this width was nearly independent of the target mass. Because nuclear interactions mediated the breakup for light targets, whereas electrical, or Coulomb, forces influenced the breakup for heavy targets, we concluded that the result was independent of the reaction mechanism and directly reflected the structure of the halo. These results indicated that the radius of the Li-11 halo was more than twice that of its core.

Models and Predictions

While these experiments were going on, theorists were trying to understand the unique behavior of Li-11. They faced two large obstacles—and still do. First, the forces between nucleons are not known accurately enough to predict the subtle binding properties of halo nuclei. Second, even if those forces were known, today's computers do not have the speed or memory needed to solve the equations of quantum mechanics for 11 interacting nucleons. Nevertheless, physicists have developed simpler models that exhibit the main physical attributes of halo nuclei.

One attribute they try to capture in their models is the role of pairing in many-nucleon systems. In general, the pairing interaction is the attraction between the least bound particles in a system; it can radically affect the properties of that system. In metals, for example, the pairing between electrons gives rise to superconductivity. The pairing interaction is also of fundamental importance in almost every aspect of nuclear structure. It determines which nuclei are stable, and its presence promotes fluidity in such shape-changing processes as nuclear fission. Pairing in a dilute neutron gas can influence the properties of neutron stars, which depend on whether the neutrons act as a superfluid. And, finally, pairing causes the Borromean behavior.

A wide range of useful models has been developed, based on very different assumptions about pairing. P. Gregers Hansen of Aarhus University in Denmark and Björn Jonson of the Chalmers University of Technology in Sweden proposed one simple model in 1988. They assumed that the pairing between the two last neutrons in Li-11 was so strong that these nucleons could be treated as a single particle, named the dineutron.

The motion of this particle in the field of the Li-9 core is a two-body problem, which is relatively easy to solve. In fact, if the binding is weak—such that the two particles have little chance to interact—the wave function can be looked up in a textbook. Using this approximation, Hansen and Jonson derived formulas for the size of the halo, for the breakup probability of the nucleus in the electric field of a highly charged target and for the energy of the dineutron after the breakup. In such a simple model, however, they could not calculate the binding energy of the halo.

While an undergraduate student at Michigan State, James Foxwell investigated another extreme model under the guidance of one of us (Bertsch). In contrast to the dineutron picture, Foxwell's model ignores the pairing between neutrons completely. It assumes that each of the last two neutrons is independently bound to the core. A two-body problem is then solved for one neutron at a time. Foxwell calculated breakup probabilities and the energy of the excited system. Like the Hansen-Jonson model, Foxwell's approach requires knowing the binding energy ahead of time. Interestingly, these two very different strategies produced similar predictions of the fragility of Li-11, differing by only a factor of two in estimating its cross section.

Recent Work on Halos

Since then, theorists have constructed more sophisticated models, explicitly incorporating the forces giving rise to pairing. Because the three-body problem in quantum mechanics is now amenable to numerical solution on large computers, it was practical to treat Li-11 as a three-particle system. Henning Esbensen of Argonne National Laboratory calculated the Li-11 wave function with a realistic description of the interaction between the neutrons and the Li-9 core and a more approximate treatment of the pairing force. His wave function showed that when the neutrons are far out in the halo they are likely to be very close together.

On the other hand, when the neutrons are near the core, they tend to stay farther apart. Thus, the actual quantum mechanics of pairing describes behavior within the limits of the two extreme models. The calculated cross section fell midway between the predictions of those models and agreed with experimentation. As is often the case in nuclear physics, quite different models can be valid, and their domains of validity can even overlap. The three-body model was also successful in predicting the momentum spread in the Li-11 breakup measured at the Michigan State facility.

Ian Thompson of the University of Surrey and his co-workers made similar calculations. This group used a more realistic force between the neutrons but treated the neutron-core interaction more approximately. It also found that Li-11 is Borromean and has a large halo. Such consistent results have given us confidence that we understand the pairing between neutrons in a low-density environment, such as might be found in the crusts of neutron stars.

Now that a new aspect of nuclear behavior has been discovered and studied, one naturally asks the question, Where do we go from here? Clearly, halos affect many nuclear reactions. For example, experimentalists plan to measure reactions between Li-11 and protons to determine the probability that a proton will pick up two neutrons and form tritium. The correlation between the neutrons directly influences this probability, since the two neutrons must be close together in order to combine with a bombarding proton. By analyzing such reactions, we will be able to obtain

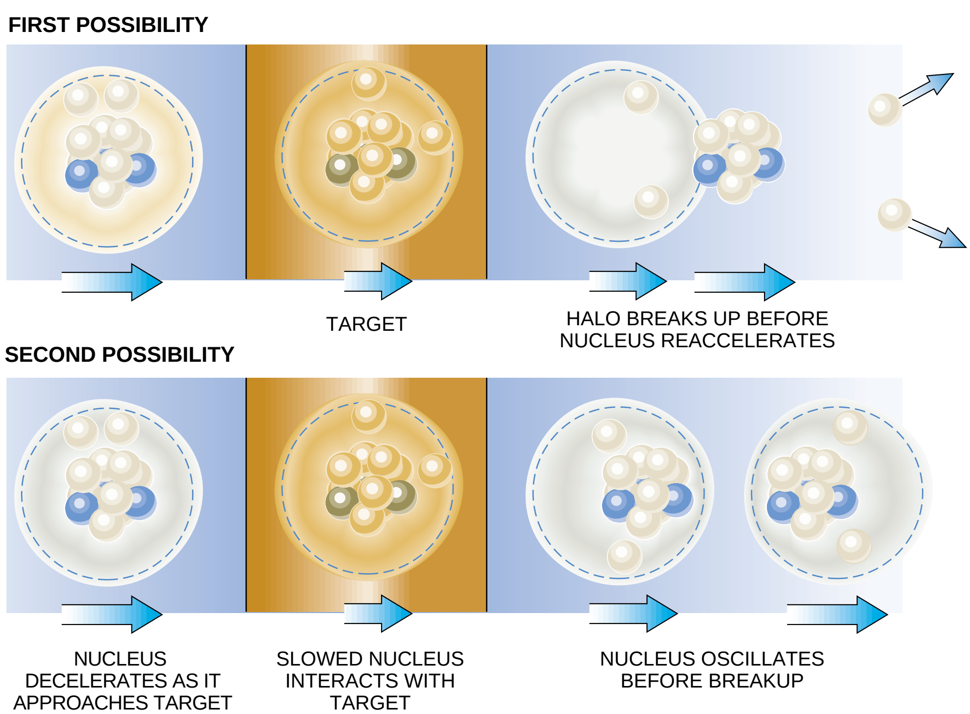
Measuring Electric and Magnetic Moments

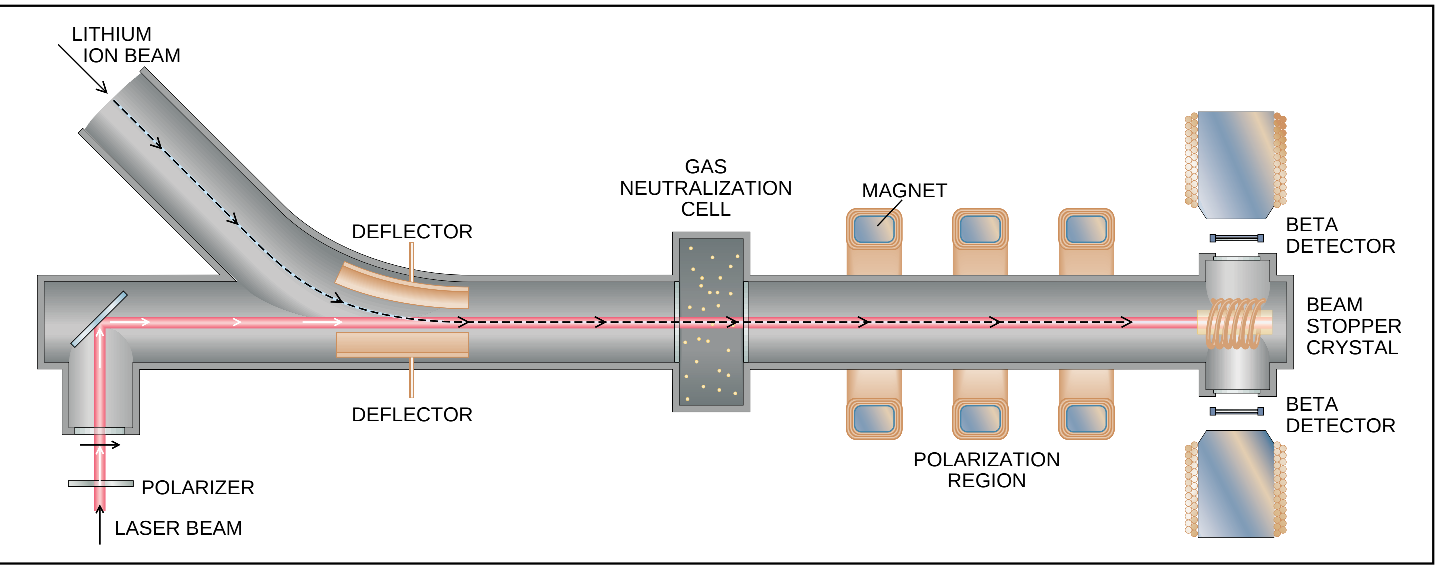
Rainer Neugart and his co-workers at CERN compared the magnetic and electrical properties of Li-11 and Li-9 in a unique apparatus. An electric field deflected ions from the ISOLDE separator down a beam pipe and through a gas, where they were neutralized. This beam was bathed in polarized laser light, aligning the spins of the atoms. Next, the atoms were stopped in a crystal. A magnetic field surrounding this part of the apparatus caused the spin axis to precess, changing its orientation. After a few milliseconds, the nuclei underwent beta decay, emitting electrons preferentially along the spin axis. From the emission directions of these decay electrons, the experimenters were able to deduce the electrical and magnetic properties of the nucleus.

How a Halo Is Lost

Aaron I. Galonsky and his collaborators at Michigan State University have investigated two contrasting pictures of how a nucleus loses its halo. In one picture, the halo neutrons are freed instantaneously when they interact with a target (*top*). In the other, the electric, or Coulomb, field generated by the target's charges sets the nucleus vibrating, with the charged core moving in one direction and the halo in another (*bottom*). To test these possibilities, Galonsky's group excited the Li-11 nucleus as gently as possible, passing the beam through a lead target, which is likely to produce Coulomb excitation. The researchers then measured the emission angles and energies of the two neutrons and Li-9 resulting from the breakup.

The energy absorbed was quite small and well defined. According to Heisenberg's uncertainty principle, therefore, the breakup of a vibrating





MICHAEL GOODMAN

a direct measure of these correlations.

Experiments by Karsten Riisager and his collaborators at CERN have shown that halo nuclei exhibit unique properties when they undergo beta decay. They observed the Borromean nucleus helium 6, which has two protons and four neutrons. When this nucleus undergoes beta decay, one of its halo neutrons may turn into a proton. Normally this proton would remain bound to the nucleus, but in He-6 it can combine with its partner neutron in the halo and escape as a deuteron.

More important, physicists would like to study the halos of heavier nuclei. Most work to date has focused on two nuclei, Li-11 and an isotope of beryllium, Be-11, both of which are fairly easy to produce and isolate. New facilities are being planned to make heavier systems. But scientists have already begun to use their current equipment to look for halo nuclei having masses of about 20. Some are now analyzing the Borromean nucleus Be-14. Workers at Michigan State have measured the momentum distributions for an isotope of carbon, C-19, which bears seven more neutrons than does the most stable form, C-12. And researchers at GANIL have discovered C-22, having yet three more neutrons.

Theorists are also beginning to investigate the properties of drip-line nuclei that have more than two halo nucleons. In such systems the many-particle aspects of pairing become especially significant. In Borromean nuclei, these halos may be vastly larger than those seen in Li-11. Physicist Vitaly Efimov of the University of Washington has predicted such a phenomenon. He showed that when the interaction between the particles in a three-body system is almost strong enough to bind them two at a time, the system may have many extended halo states, potentially an infinite number of them.

Finally, weakly bound protons may also give rise to nuclear halos. Perhaps the best example is an isotope of boron, B-8, which contains one very loosely bound proton. This proton is even less well bound than the neutrons in Li-11, and its halo is quite possibly aspherical. To determine the characteristics of B-8's halo, teams at several laboratories are measuring the nucleus's parallel momentum distribution.

Astrophysicists are particularly interested in the nucleus B-8 because in the sun it produces easily detected neutrinos. A serious anomaly has arisen because the observed number of neutrinos from the decay of B-8 in the sun is much less than predicted. Understanding the exact nature of this nucleus may well provide clues to this mystery. The study of nuclei near the drip lines will surely yield further surprises. But already halos have taught us quite a bit about what takes place at the outer limits of stability.

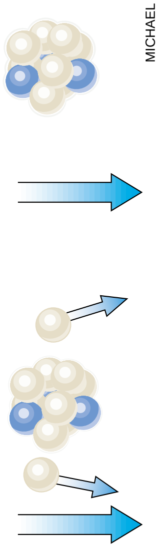
FURTHER READING

CAULDRONS IN THE COSMOS: NUCLEAR ASTROPHYSICS. Claus E. Rolfs and William S. Rodney. University of Chicago Press, 1988.

PHYSICS WITH RADIOACTIVE BEAMS. Richard N. Boyd and Isao Tanihata in *Physics Today,* Vol. 45, No. 6, pages 44–52; June 1, 1992.

NUCLEI AT THE LIMITS OF PARTICLE STABILITY. Alex C. Mueller and Bradley M. Sherrill in *Annual Review of Nuclear and Particle Science,* Vol. 43, pages 529–584; 1993.

NUCLEAR HALO STATES. K. Riisager in *Reviews of Modern Physics,* Vol. 66, No. 3, pages 1105–1116; July 1, 1994.


MICHAEL GOODMAN

system should take a relatively long time. But the workers found that Li-11 broke up quickly, as though the neutrons were freed instantaneously on collision. They inferred this fact from a seemingly unrelated observation: in comparing the velocities of the Li-9 core and the two neutrons, the neutrons moved more slowly. At first, it was puzzling that the neutrons should be slower if the breakup process was so gentle.

The explanation lay in the timing of the breakup itself. Because Li-11 is charged, it decelerates as it approaches the target's Coulomb field and then reaccelerates as it departs. In the oscillatory picture the breakup takes such a long time that the Li-9 nucleus would travel well beyond the target's Coulomb reach before it could occur. But in an instantaneous breakup, Li-9 is separated near the target, where it is subject to Coulomb reacceleration. In contrast, the uncharged neutrons are not affected by the electric field and so move more slowly, as observed. Hence, Galonsky's group concluded that the Li-11 nucleus had lost its halo when it passed the target.

Kin Recognition

Many organisms, from sea squirts to primates, can identify their relatives. Understanding how and why they do so has prompted new thinking about the evolution of social behavior

by David W. Pfennig and Paul W. Sherman

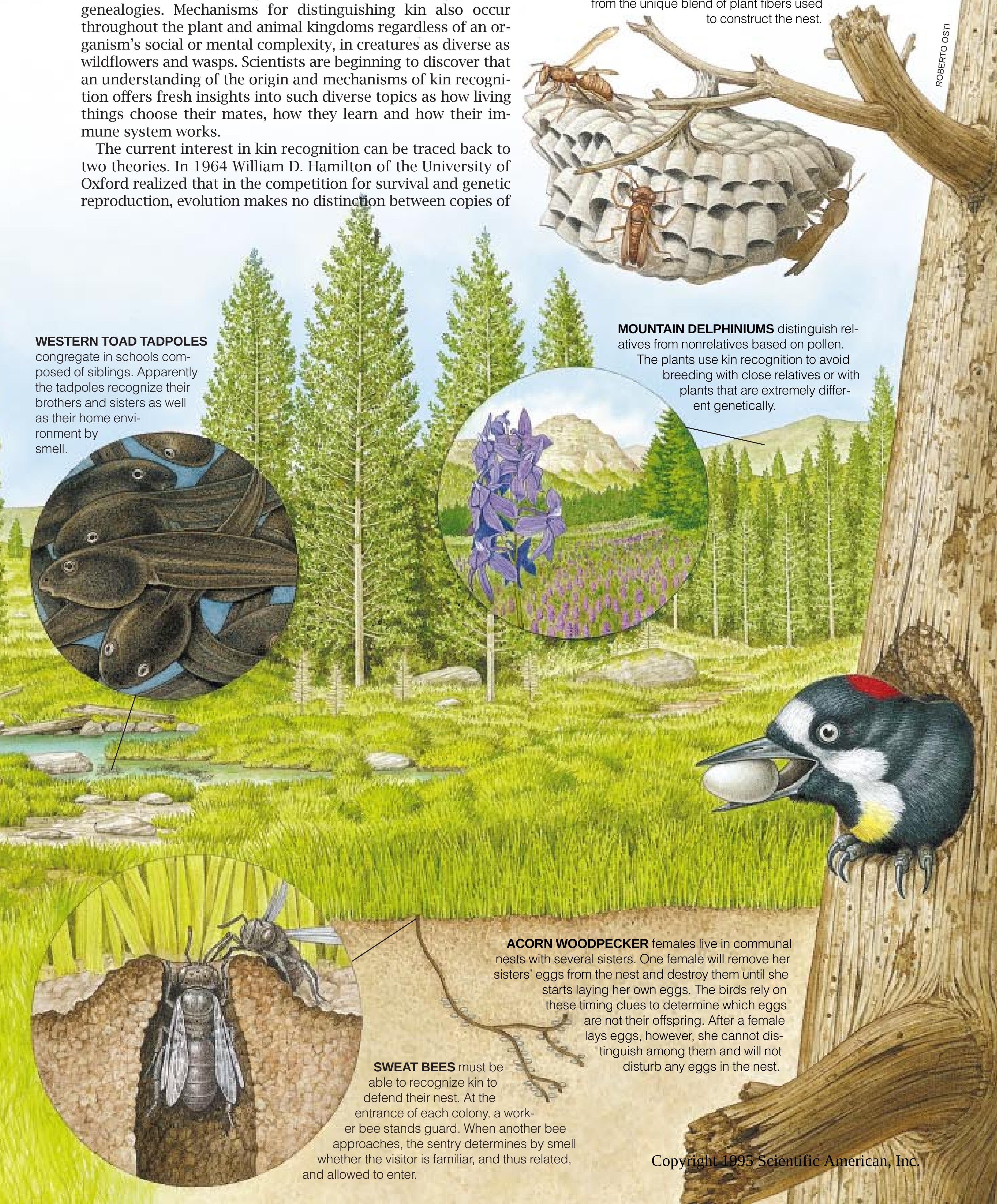
BANK SWALLOWS initially depend on location to identify their offspring. Parents remember where they have made their burrow and will feed any nestling they find there. Because the young birds generally remain in their parents' nest, adult swallows typically feed only their offspring. Once the chicks learn to fly, parents recognize their offspring's voices.

BELDING'S GROUND SQUIRRELS live in groups in which mothers, daughters and sisters cooperate extensively. By using odors, the squirrels can distinguish familiar nestmates, who are close kin, from nonnestmates. They can also discriminate between full sisters and half sisters.

WILDFLOWERS such as English plantains grow faster in the presence of kin than nonkin. The plants probably use chemical cues released by the roots to distinguish relatives.

Kinship is a basic organizing principle of all societies. Humans possess elaborate means by which to identify relatives, such as using surnames and maintaining detailed genealogies. Mechanisms for distinguishing kin also occur throughout the plant and animal kingdoms regardless of an organism's social or mental complexity, in creatures as diverse as wildflowers and wasps. Scientists are beginning to discover that an understanding of the origin and mechanisms of kin recognition offers fresh insights into such diverse topics as how living things choose their mates, how they learn and how their immune system works.

The current interest in kin recognition can be traced back to two theories. In 1964 William D. Hamilton of the University of Oxford realized that in the competition for survival and genetic reproduction, evolution makes no distinction between copies of



PAPER WASPS utilize odors to determine whether visiting wasps are related. All colony members have an identifying smell that results from the unique blend of plant fibers used to construct the nest.

WESTERN TOAD TADPOLES congregate in schools composed of siblings. Apparently the tadpoles recognize their brothers and sisters as well as their home environment by smell.

MOUNTAIN DELPHINIUMS distinguish relatives from nonrelatives based on pollen. The plants use kin recognition to avoid breeding with close relatives or with plants that are extremely different genetically.

ACORN WOODPECKER females live in communal nests with several sisters. One female will remove her sisters' eggs from the nest and destroy them until she starts laying her own eggs. The birds rely on these timing clues to determine which eggs are not their offspring. After a female lays eggs, however, she cannot distinguish among them and will not disturb any eggs in the nest.

SWEAT BEES must be able to recognize kin to defend their nest. At the entrance of each colony, a worker bee stands guard. When another bee approaches, the sentry determines by smell whether the visitor is familiar, and thus related, and allowed to enter.

alternative forms of genes, known as alleles, that are transmitted through direct descendants, such as offspring, and those propagated through nondescendant kin, such as siblings. Whereas the traditional view held that natural selection favored individuals that produced the greatest number of offspring, Hamilton shifted the emphasis to genes. He concluded that natural selection must favor organisms that help any relative, because by doing so they increase their total genetic representation.

Hamilton termed this idea inclusive fitness, because it includes both the genes an organism transmits through its offspring as well as copies of those genes it helps to propagate in reproductive relatives. Inclusive fitness theory can explain the evolution of nepotism, particularly in the unusual instances in which some members of certain species—ants, bees or naked mole rats, for example—have no offspring and exist only to nurture other relatives [see "Naked Mole Rats," by Paul W. Sherman, Jennifer U. M. Jarvis and Stanton H. Braude; SCIENTIFIC AMERICAN, August 1992].

A second explanation, optimal outbreeding theory, was developed in the early 1970s by Patrick Bateson of the University of Cambridge and William M. Shields of the SUNY College of Environmental Science and Forestry in Syracuse. Their hypothesis draws on the well-known fact that inbreeding between very close relatives, such as siblings, often causes offspring to display detrimental characteristics. All organisms possess a few deleterious alleles that are normally not expressed. The same rare versions of these genes are likely to be carried by close relatives. With close inbreeding, offspring can inherit such alleles from both parents, resulting in their harmful expression. Conversely, mating with individuals that are very different genetically can produce detrimental effects by breaking up gene combinations that produce favorable traits. Optimal outbreeding theory explains why many organisms prefer to mate with those to whom they are neither too closely nor too distantly related.


RICHARD K. GROSBERG

SEA SQUIRTS are marine animals that lack brains but can nonetheless identify their kin using chemical cues. Two organisms occasionally attempt to join together, an endeavor that is successful only if the two animals are related.

Two Forms of Recognition

More recent work has brought up additional ideas for why kin recognition takes place. But the evolutionary reasons for this ability are only part of the story, one to which we will return later. We turn first to the intriguing question of *how* organisms distinguish their relatives. In general, plants and animals use two mechanisms to identify kin. In some cases, physical features, known as phenotypes, allow individuals to recognize their relatives directly. Alternatively, kin can be identified indirectly without reference to phenotypes but by clues related to time or place.

Often organisms rely on a combination of direct and indirect techniques. For example, bank swallows (*Riparia riparia*), birds that nest in colonies on sandbanks, identify their young using both kinds of clues. John L. Hoogland of the University of Maryland and one of us (Sherman) found that bank swallow parents will feed any nestling that appears in their burrow. This behavior indicates that adult swallows recognize their young indirectly by learning the location of the burrow they have excavated. The flightless chicks usually remain in the burrow where they were born for three weeks after hatching, so during this time parents generally feed only their own young. After the chicks learn to fly, however, broods mix extensively, so parents must use direct clues to ensure that they continue to provide only for their own offspring. Michael D. Beecher and his colleagues at the University of Washington discovered that by the time bank swallow chicks are 20 days old, they have distinct vocal signatures that indicate to parents which young are their own.

To understand how these discriminations take place, researchers have divided the process of kin recognition into three components. Initially, a recognition cue is produced. Next, another individual perceives it. Finally, that individual interprets the cue and takes appropriate action. In indirect recognition the signal is external to the plant or animal; in direct recognition the label is produced by the organism itself. Communities of social animals, in which kin and nonkin frequently mix, are especially likely to use the direct method. Thus, scientists have become intrigued with the complex interplay of factors that takes place in the process of direct kin recognition.

A direct kin-recognition signal can be any physical characteristic that correlates reliably with relatedness. Such labels vary widely among species. Visual references are common among animals, such as primates, whose most prominent sense is sight. Organisms that must attract mates across a distance in the dark, such as frogs, use auditory signals. And, of course, chemical odors are important distinguishing labels for many animals.

In general, chemical markers convey information accurately while requiring less effort to produce than other signals, particularly sounds. An organism must expend a considerable amount of energy compressing air to create sound. In contrast, chemical labels often consist of a few molecules of a substance the body produces naturally during daily activities. Furthermore, a system is already in place to detect and decipher chemical substances: such signals are readily interpreted by the body's immune system. Some speculate that the physiological machinery used in kin recognition was borrowed from the immune system in the course of evolution.

Source of the Signals

Recognition labels differ not only according to which sense they draw on but also in their origin. These cues can reflect specific genetic traits; they can be acquired from the environment, or they can be a result of both. Studies of certain tunicates, or sea squirts, specifically *Botryllus schlosseri,* show that these marine animals use genetic labels to identify relatives. Tunicates lack a brain, thus proving that kin recognition does not depend on mental complexity.

Sea squirts begin life as planktonic larvae that eventually settle on a rock and multiply asexually to form an interconnected colony of structurally and genetically identical animals. Occasionally, two colonies will attempt to fuse; large organisms survive better than small ones, so combining with others is apparently beneficial. Richard K. Grosberg and James F. Quinn of the University of California at Davis discovered that the larvae settle near and merge with genetically similar organisms. If a tunicate attempts to join another, unrelated colony, the second tunicate emits toxic substances that repel the invader.

Grosberg and Quinn have also determined the area on the chromosomes that controls this recognition response. They noticed that larvae settle near others that carry the same allele in the location known as the histocompatibility complex. This region of the chromosome encodes for the chemicals that enable an organism to distinguish self from nonself as part of the immune system. The researchers also discovered that tunicates settle closer to nonrelatives that were bred in the laboratory to have the same version of the gene at this location in preference to establishing themselves near true kin that were bred to carry an alternative allele.

In nature, the chances of mistaking nonrelatives for kin are minuscule. For reasons that are not totally clear, the types of genes found at the histocompatibility complex are so variable across a species that if two organisms share the same allele there, they must have acquired it from a recent ancestor. So when one tunicate attempts to fuse its tissues with another, the immune system can recognize the encroaching tissue as being either foreign or similar—in other words, related or not—depending on the genetic makeup at the histocompatibility complex.

House mice (*Mus musculus*) also rely on the histocompatibility complex to identify kin. Because the genes there affect body odor, mice can depend on this trait to distinguish relatives. Just as was the case for tunicates, the genes in mice found at the histocompatibility complex are highly variable, but among family members the alleles tend to be the same. Therefore, individuals that smell alike are usually related. C. Jo Manning of the University of Nebraska and Wayne K. Potts and Edward K. Wakeland of the University of Florida observed that female mice tend to mate with males that smell different, apparently in order to avoid inbreeding. But they nest communally with females that smell similar, such as sisters, which helps to ensure the survival of nieces and nephews as well as offspring.

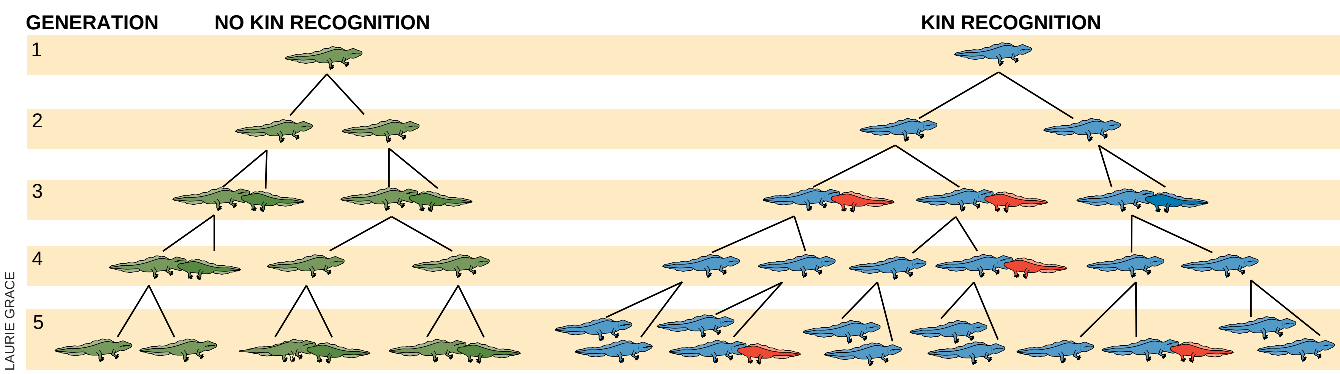
The Smell of Paper Wasps

In contrast to tunicates and mice, other organisms use labels acquired from their environment to recognize relatives. One of us (Pfennig) has studied such signals in certain paper wasps, specifically *Polistes fuscatus.* These common garden insects construct open comb nests composed of wafer-thin plant fibers. Colonies typically consist of a queen and her daughter workers.

Kin recognition is crucial because nests are frequently visited by other wasps with various intentions. In some cases, the visitors are homeless relatives whose nests were destroyed by predators, such as birds. In others, the intruding wasps come to steal eggs to feed the larvae in their own active colonies. Before allowing invaders on their nest, wasps must distinguish between orphaned kin, which will be helpers, and unrelated wasps, which are threats to the nest.

Paper wasps make this distinction directly using chemical odors. Pfennig, George J. Gamboa of Oakland University, Hudson K. Reeve and Janet Shellman-Reeve of Cornell University discovered that each wasp assimilates from its nest an odor specific to the insects that live there. This smell, which serves as the recognition cue, is locked into the wasp's epicuticle, or skin, before it hardens. Karl E. Espelie of the University of Georgia and his colleagues determined that the source of the smell is odoriferous hydrocarbons. These compounds are derived from the plant fibers that make up the nest paper as well as from secretions produced by the wasps that constructed the nest. Because each colony uses a unique mixture of plants in nest construction, family members often are more likely to share this environmentally acquired label than a genetic one. The mixing and recombination of genes that happen during sexual reproduction ensure that family members, though genetically similar, will not be identical.

Both genetic labels and environmentally acquired ones can lead to mistakes, however. Relying solely on signals picked up from the environment might cause acceptance errors, in which an individual mistakenly assists nonrelatives that live in similar surroundings. Such cheaters could then reap the rewards of misplaced beneficence without reciprocating and so become predominant in the population. Depending only on gene products also might cause an individual to accept nonrelatives that carry "outlaw alleles" that encode just the recognition trait. Again, the renegade alleles will spread throughout a population. Finally, relying on genetic cues increases the risk of committing rejection errors, in which relatives are mistakenly treated as nonkin because they do not, by chance, possess the recognition trait.



KIN RECOGNITION can help make one group of organisms more successful than others. In this example, each salamander produces two offspring (only one parent is shown), but not all of them survive, because these animals resort to cannibalism when faced with a food shortage. For instance, in the third generation, only half of the salamanders that cannot recognize kin (*green*) survive to reproduce; the others are eaten by siblings. But three out of four salamanders survive in the family that can identify relatives (*blue*) because half of them ate salamanders from another family (*red*). By the fifth generation, the family that is genetically disposed to distinguish kin predominates.

The likelihood that these types of mistakes will occur depends on the genetic makeup of the organisms involved as well as their surroundings. Organisms such as tunicates and mice minimize the chance that two nonrelatives will share similar genetic traits by exploiting regions of the chromosomes that are variable within a species but relatively constant in families. These genetic labels are most useful for organisms that inhabit a fairly uniform chemical environment, such as a rock where several colonies of tunicates might live. For organisms such as paper wasps that live in more diverse areas, environmentally acquired labels can provide more accurate clues.

Acting on a Cue

After a recognition cue has been produced, how do others use it to assess relatedness? As far as we know, these signals are always learned. Even the immune system must learn to recognize the self [see "How the Immune System Learns about Self," by Harald von Boehmer and Pawel Kisielow; SCIENTIFIC AMERICAN, October 1991]. Indeed, without learning how to make that distinction, the immune system would attack every tissue in the body.

Organisms learn labels from themselves, their relatives or their environment. Individuals form a template of these labels, much like the templates that are thought to be involved in birdsong learning. In most creatures the process of learning takes place early in life, when they are likely to be living among relatives. Memories of companions are durable, ensuring that throughout its life an organism can compare the remembered image with another individual's physical characteristics. In addition, many creatures update their templates from time to time, enabling them to recognize kin as their labels change with age, for instance.

To illustrate the role of learning in kin recognition, consider the part that the nest plays for paper wasps. In experiments done in the laboratory, wasps removed from their nest and nestmates later recognized nonrelatives as well as relatives as kin. Wasps isolated only from their nest but not from their nestmates still treated all wasps as kin. Furthermore, ones exposed to a nest other than their own learned to treat wasps emerging from that nest as their relatives. Only in the presence of their own nest did the insects learn the chemical signal that allows them to distinguish kin from nonkin.

In contrast to paper wasps, honeybees (*Apis mellifera*) can learn identification cues from their nestmates and from themselves. One reason for this difference between honeybees and paper wasps may be the mating patterns of the queens. Honeybee hives often contain workers sired by more than a dozen drones, whereas paper wasp workers are sired primarily by only one male. In consequence, honeybee hivemates are a mixture of full and half sisters, and paper wasp nestmates are mostly full sisters.

To distinguish between full and half siblings, a worker honeybee must have knowledge of the genes received from its father, as well as such information about the bee under examination. Thus, some mechanism of self-inspection is required—a phenomenon Richard Dawkins of the University of Oxford has dubbed the "armpit effect." Wayne M. Getz and Katherine B. Smith of the University of California at Berkeley showed that bees raised in isolation learned their own odor and then favored similarly smelling full sisters over maternal half sisters whose slightly different genetic makeup resulted in a different odor. Whether honeybees learn from themselves under crowded hive conditions is unclear.

Once recognition has taken place, the individual must decide what action to take, depending on the context of the encounter. For example, paper wasp workers are more intolerant of unrelated wasps when they invade the nest—where they might try to steal eggs—than they are when they meet the same nonkin elsewhere. According to a theoretical model developed by Reeve, for discrimination to occur, the similarity between the observed individuals' physical characteristics and the observer's template must be above some critical value. This value reflects how often organisms encounter relatives as opposed to nonrelatives as well as the costs of rejecting kin compared with those of accepting nonkin.

This model helps to explain certain errors in discrimination. For example, Anne B. Clark of SUNY at Binghamton and David F. Westneat, Jr., of the University of Kentucky have found that male red-winged blackbirds (*Agelaius phoeniceus*) feed all the chicks in their nest, even though—because females mate with more than one male—about one in four chicks is not their offspring. Presumably, it is more efficient in a reproductive sense for a male parent to feed all the chicks in its nest, which wastes only a little effort on unrelated young, than to risk allowing one of its progeny to starve.

Cannibalistic Kin

Let us now return to the question of why many organisms can distinguish their relatives. The evolutionary significance of kin recognition is dramatically illustrated by species in which some individuals have the potential to harm their relatives. Certain protozoans, rotifers, nematodes and amphibian larvae exist in two distinct forms that differ in dietary preference—they can be either cannibalistic or omnivorous. Which path an individual takes depends mainly on the environment in which it was raised, although both types can be found within one family.

Cannibalistic animals also return us to inclusive fitness theory. According to this line of thinking, cannibals should have evolved to avoid eating their own kin because of the genetic costs of such a practice: any family that exhibited such behavior would probably not survive very long.

To test this prediction, we studied patterns of kin recognition in spadefoot toad tadpoles (*Scaphiopus bombifrons*), which develop in temporary ponds in the desert. These tadpoles possess a special means of acquiring extra nourishment in order to hasten their growth so they can escape their rapidly drying ponds.

All spadefoot tadpoles begin life as omnivores, feeding primarily on detritus. Occasionally, however, one eats another tadpole or a freshwater shrimp. This event can trigger a series of changes in the tadpole's size, shape and musculature and, most important, in dietary preference. These changed tadpoles become exclusively carnivorous, feast-

LARRY DITTO *Bruce Coleman Inc.*

RED-WINGED BLACKBIRD males feed all chicks in the nest. Most of these young birds are indeed offspring, so the adults benefit in a reproductive sense by taking care of all the birds in their nests rather than risk letting kin starve.

Family Matters

After four barren years at the Philadelphia Zoo, Jessica, a rare Lowland gorilla (*right*), was moved to the San Diego Zoo. Jessica became pregnant right away and gave birth to Michael on Christmas Eve in 1991.

Kin discrimination may explain why Jessica did not mate until she was introduced to males other than those she had lived around since birth. In nature, such familiar individuals would usually be relatives, and Jessica may have viewed her companions as such. To avoid potential inbreeding, animals generally do not have much sexual interest in their close relatives.

In species that have dwindled to a single small population, identifying familiar nonrelatives as kin can be a particular problem. With an understanding of kin recognition, zookeepers can prevent animals from making such mistakes and perhaps facilitate breeding in endangered species.

RON GARRISON *Zoological Society of San Diego*

ing on other animals—including members of their own species.

Whether a tadpole will actually eat members of its own family depends on the balance between the costs and benefits of such discriminating taste. This balance changes depending on the tadpole's development and its hunger level. For example, if the tadpole remains an omnivore, it tends to congregate in schools that consist primarily of siblings. Its cannibalistic brothers and sisters, however, most often associate with and eat nonsiblings.

Carnivores nip at other tadpoles, and after this "taste test," they either eat the animals if they are not related or release them unharmed if they are siblings. Interestingly, carnivores are less likely to avoid eating brothers and sisters when they are hungry than when full. Apparently the tadpoles stop discriminating kin when their own survival is threatened—after all, a carnivorous tadpole is always more closely related to itself than to its sibling.

Arizona tiger salamanders (*Ambystoma tigrinum*) also come in two types: a small-headed omnivore that eats mostly invertebrates and a large-headed carnivore that feeds mainly on other salamanders. All larvae start off as omnivores, and they typically stay that way if they grow up among siblings. But the larvae often transform into cannibals if they grow up among nonkin. By not developing into a cannibal in the presence of siblings, the salamanders reduce their chances of harming relatives. Together with James P. Collins of Arizona State University, we found that cannibals prefer not to dine on close kin when also offered smaller larvae that are distantly related. By temporarily blocking the animals' noses, we determined that the discrimination is based on chemical cues.

New Challenges

In addition to the standard inclusive fitness theory arguments, there may be other reasons why organisms recognize kin. For example, Pfennig and his graduate student Michael Loeb, along with Collins, ascertained that tiger salamander larvae are afflicted in nature with a deadly bacterium. Furthermore, the team determined that cannibals are especially likely to be infected when they eat diseased members of their species. Perhaps natural selection favors cannibals that avoid eating kin and thereby avoid pathogens that are transmitted more easily among close relatives with similar immune systems. Such reasoning implies that kin recognition may have evolved not only to ensure relatives' survival but also simply to preserve an animal's own life.

These recent results have challenged traditional understandings of kin recognition and have demonstrated that biologists have much more to learn about the process. In the course of such work, we hope to gain more insights into the evolution of social interactions as varied as nepotism and cannibalism. Because of the fundamental connection between the immune system and the mechanism of kin recognition, we also hope further study will reveal details on how these systems operate.

Research on kin recognition also may have practical uses. Mary V. Price and Nickolas M. Waser of the University of California at Riverside have discovered that mountain delphinium (*Delphinium nelsonii*) can recognize pollen of related plants. Also, Stephen J. Tonsor of Michigan State University and Mary F. Wilson of the Forestry Sciences Laboratory in Juneau, Alaska, found that some flowering plants, such as pokeweeds (*Phytolacca americana*) and English plantains (*Plantago lanceolata*), grow faster when potted with full or half siblings than when potted with nonrelatives. If these kinship effects are widespread, they could be used to advantage in planting crops.

Scientists have been investigating kin recognition for more than half a century, and we now have a good deal of information about a variety of plants and animals. Ongoing work will allow us to formulate a broad understanding of the significance of this phenomenon.

The Authors

DAVID W. PFENNIG and PAUL W. SHERMAN have shared an interest in kin recognition for more than a decade. Pfennig received his Ph.D. from the University of Texas before joining Sherman as a National Science Foundation postdoctoral fellow at Cornell University. Currently Pfennig is assistant professor of ecology, ethology and evolution at the University of Illinois, where his research focuses on the evolution of kin recognition and developmental polymorphism. Sherman, who received his Ph.D. from the University of Michigan, is professor of animal behavior at Cornell. He studies the social behavior of various vertebrates, including ground squirrels and naked mole rats.

Further Reading

THE EVOLUTION OF CONSPECIFIC ACCEPTANCE THRESHOLDS. Hudson K. Reeve in *American Naturalist,* Vol. 133, No. 3, pages 407–435; March 1989.

KIN RECOGNITION. Edited by Peter G. Hepper. Cambridge University Press, 1991.

COMMUNAL NESTING PATTERNS IN MICE IMPLICATE MHC GENES IN KIN RECOGNITION. C. Jo Manning, Edward K. Wakeland and Wayne K. Potts in *Nature,* Vol. 360, No. 6404, pages 581–583; December 10, 1992.

KIN RECOGNITION AND CANNIBALISM IN POLYPHENIC SALAMANDERS. David W. Pfennig, Paul W. Sherman and James P. Collins in *Behavioral Ecology,* Vol. 5, No. 2, pages 225–232; Summer 1994.

From Complexity to Perplexity

by John Horgan, *senior writer*

Can science achieve a unified theory of complex systems? Even at the Santa Fe Institute, some researchers have their doubts

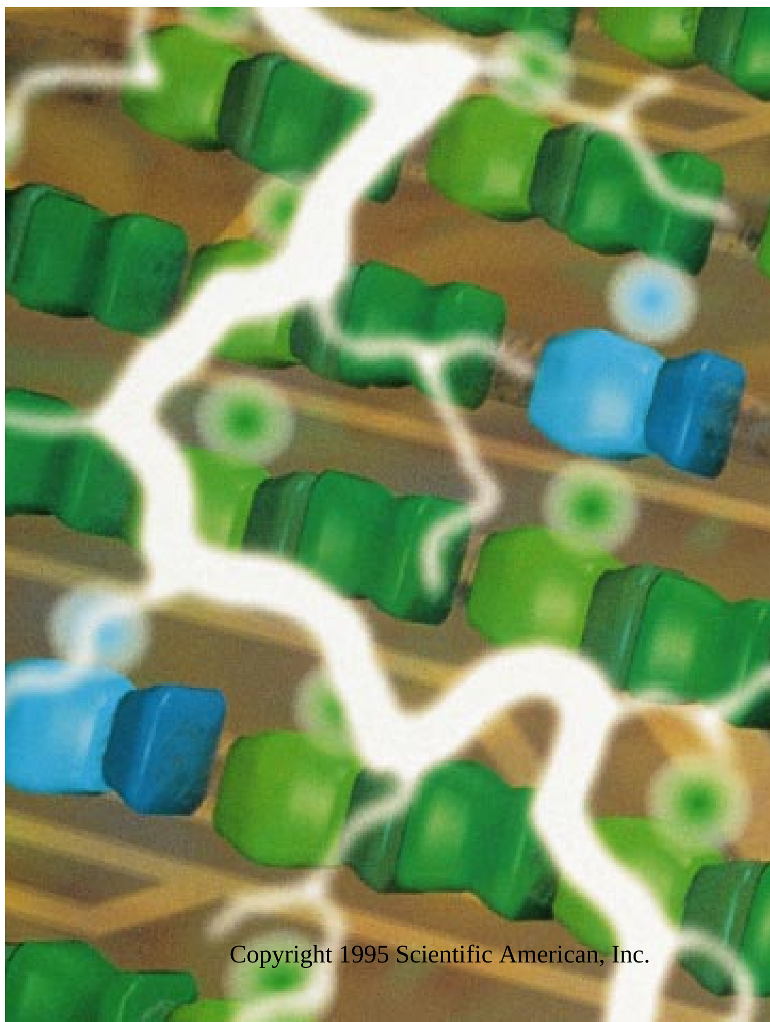
Champagne and big ideas are bubbling at the Museum of Indian Arts and Culture in Santa Fe, N.M. The museum is hosting a dinner for the Santa Fe Institute, where complex people ponder complex things. Some of the institute's brightest luminaries are there, including Murray Gell-Mann, Nobel laureate and co-discoverer of quarks, with his permanently skeptical squint; artificial-life proselytizer Christopher G. Langton, clad in his uniform of jeans, clodhoppers, leather vest and silver bracelet; the ruddy-faced nonlinear economist W. Brian Arthur, who has recently been taking calls from the White House; and world-class intellectual riffer Stuart A. Kauffman, whose demeanor is at once cherubic and darkly brooding. Mingling with these scientific pioneers are various "friends of the institute," ranging from mega-philanthropist George Soros to the best-selling novelist Cormac McCarthy.

Before everyone tucks into the filet mignon, David Liddle, a computer entrepreneur who chairs the board of trustees, reviews the institute's accomplishments. "There is a lot to be proud of," he says. There certainly is, at least from a public-relations standpoint. The institute is not large: it supports only six full-time researchers in Santa Fe; 50 "external faculty" members work elsewhere. Nevertheless, in the decade since its founding, the institute has enjoyed much favorable attention from the press, including *Scientific American,* and has been celebrated in several popular books. It has become renowned as a leading center of complexity studies, a place where scientists impatient with the stodgy, reductionist science of the past are creating a "new, unified way of thinking about nature, human social behavior, life and the universe itself" (as one book jacket put it).

What Liddle does not say is that even some scientists associated with the institute are beginning to fret over the gap between such rhetoric and reality. Take Jack D. Cowan, a mathematical biologist from the University of Chicago who helped to found the institute and remains on its board. Cowan is no scientific prude; he has explored the neurochemical processes underlying the baroque visual patterns evoked by LSD. But some Santa Fe theorists exhibit too high a "mouth-to-brain ratio" for his taste. "There has been tremendous hype," he grumbles.

Cowan finds some work at Santa Fe interesting and important, but he deplores the tendency of research there "to degenerate into computer hacking." Too many simulators also suffer from what Cowan calls the reminiscence syndrome. "They say, 'Look, isn't this reminiscent of a biological or physical phenomenon!' They jump in right away as if it's a decent model for the phenomenon, and usually of course it's just got some accidental features that make it look like something." The major discovery to emerge from the institute thus far, Cowan suggests, is that "it's very hard to do science on complex systems."

Some residents blame the media for the exaggerated claims associated with the institute. "Ninety percent of it came from journalists," Arthur asserts. Yet the economist cannot help



but play the evangelist. “If Darwin had had a computer on his desk,” he exclaims, “who knows what he could have discovered!” What indeed: Charles Darwin might have discovered a great deal about computers and very little about nature.

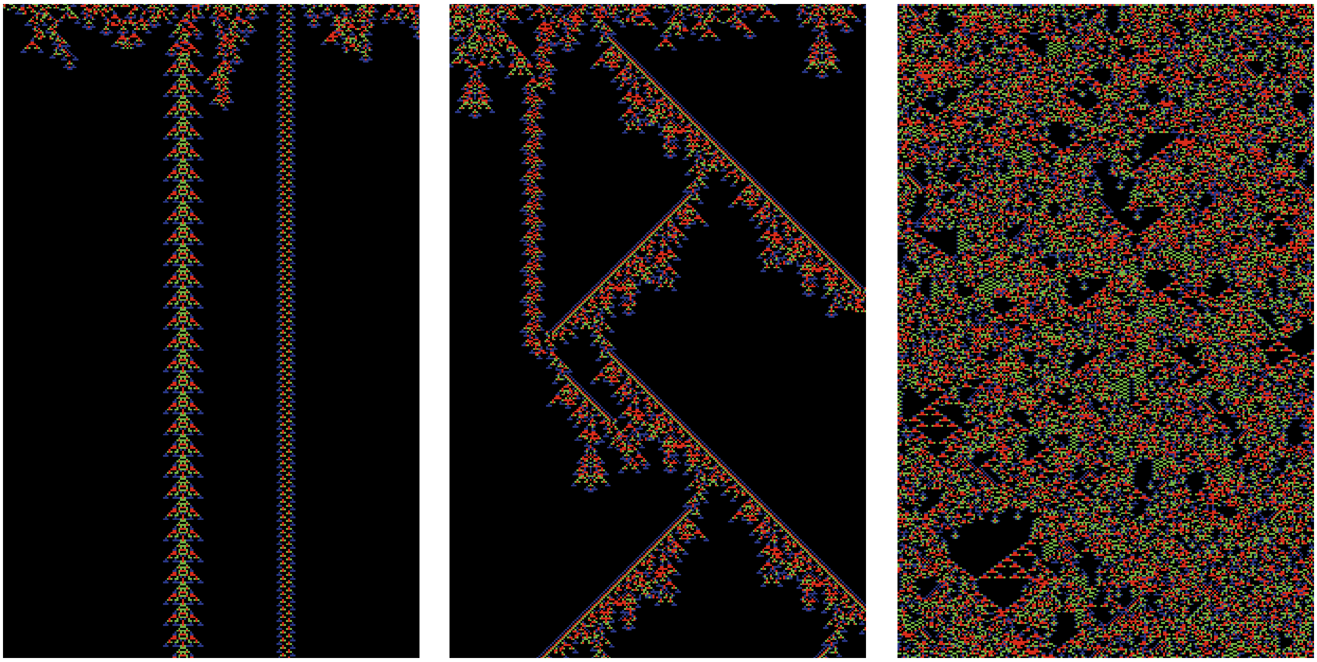
The grandest claim of Santa Fe’ers is that they may be able to construct a “unified theory” of complex systems. John H. Holland, a computer scientist with joint appointments at the University of Michigan and the Santa Fe Institute, spelled out this breathtakingly ambitious vision in a lecture two years ago: “Many of our most troubling long-range problems—trade balances, sustainability, AIDS, genetic defects, mental health, computer viruses—center on certain systems of extraordinary complexity. The systems that host these problems—economies, ecologies, immune systems, embryos, nervous systems, computer networks—appear to be as diverse as the problems. Despite appearances, however, the systems do share significant characteristics, so much so that we group them under a single classification at the Santa Fe Institute, calling them complex adaptive systems [CAS]. This is more than terminology. It signals our intuition that there are general principles that govern all CAS behavior, principles that point to ways of solving the attendant problems.” Holland, it should be said, is considered to be one of the more modest complexologists.

Some workers now disavow the goal of a unified theory. “I don’t even know what that would mean,” says Melanie Mitchell, a former student of Holland’s who is now at the SFI. “At some level you can say all complex systems are aspects of the same underlying principles, but I don’t think that will be very useful.” Stripped of this vision of unification, however, the Santa Fe Institute loses much of its luster. It becomes just another place where researchers are using computers and other tools to address problems in their respective fields. Aren’t all scientists doing that?

Scientists familiar with the history of other would-be unified theories [*see box on pages 108 and 109*] are not sanguine about the prospects for their brethren at Santa Fe. One doubter is Herbert A. Simon of Carnegie Mellon University, a Nobel laureate in economics who has also contributed to artificial intelligence and sociobiology. “Most of the people who talk about these great theories have been infected with mathematics,” he says. “I think you’ll see a bust on the notion of unification.” Rolf Landauer of IBM, who has spent his career exploring the links between physics, computation and information, agrees. He accuses complexologists of seeking a “magic criterion” that will help them unravel all the messy intricacies of nature. “It doesn’t exist,” Landauer says.

The problems of complexity begin with the term itself. Complexologists have struggled to distinguish their field from a closely related pop-science movement, chaos. When all the fuss was over, chaos turned out to refer to a restrict-

HYPED TO DEATH? Artificial life, a major subfield of complexity studies, is “fact-free science,” according to one critic. But it excels at generating computer graphics. This image from Thomas R. Ray of the University of Delaware and the Santa Fe Institute illustrates how life evolves in cyberspace. Green and blue objects are self-replicating programs; lightning represents random “mutations”; and the skull is the Reaper, a routine that exerts selection pressure by “killing” less fit programs.



EDGE OF CHAOS concept is illustrated by this triptych of cellular-automaton images generated by Christopher G. Langton of the Santa Fe Institute. Langton and others have suggested that the complexity of a system may be equivalent to its capacity for computation; this capacity peaks in a regime (*center*) between highly ordered states (*left*) and chaotic ones (*right*). Others have challenged Langton's interpretation of his experiments and the basic premise of the edge of chaos.

ed set of phenomena that evolve in predictably unpredictable ways. Various attempts have been made to provide an equally precise definition of complexity. The most widely touted definition involves "the edge of chaos." The basic idea is that nothing novel can emerge from systems with high degrees of order and stability, such as crystals. On the other hand, completely chaotic systems, such as turbulent fluids or heated gases, are *too* formless. Truly complex things—amoebae, bond traders and the like—appear at the border between rigid order and randomness.

Most popular accounts credit the idea to Christopher Langton and his co-worker Norman H. Packard (who coined the phrase). In experiments with cellular automata, they concluded that a system's computational capacity—that is, its ability to store and process information—peaks in a narrow regime between highly periodic and chaotic behavior. But cellular-automaton investigations by two other SFI researchers, James P. Crutchfield and Mitchell, did not support the conclusions of Packard and Langton. Crutchfield and Mitchell also question whether "anything like a drive toward universal-computational capabilities is an important force in the evolution of biological organisms." Mitchell complains that in response to these criticisms, proponents of the edge of chaos keep changing their definition. "It's a moving target," she says.

Other definitions of complexity have been proposed—at least 31, according to a list compiled several years ago by Seth Lloyd of the Massachusetts Institute of Technology, a physicist and Santa Fe adjunct. Most involve concepts such as entropy, randomness and information—which themselves have proved to be notoriously slippery terms. All definitions have drawbacks. For example, algorithmic informational complexity, proposed by IBM mathematician Gregory J. Chaitin, holds that the complexity of a system can be represented by the shortest computer program describing it. But according to this criterion, a text created by a team of typing monkeys is more complex—because it is more random—than *Finnegans Wake.*

The Poetry of Artificial Life

Such problems highlight the awkward fact that complexity exists, in some murky sense, in the eye of the beholder. At various times, researchers have debated whether complexity has become so meaningless that it should be abandoned, but they invariably conclude that the term has too much public-relations value. Complexologists often employ "interesting" as a synonym for "complex." But what government agency would supply funds for research on a "unified theory of interesting things"? (The Santa Fe Institute, incidentally, will receive about half its $5-million 1995 budget from the federal government and the rest from private benefactors.)

Complexologists may disagree on what they are studying, but most concur on how they should study it: with computers. This faith in computers is epitomized by artificial life, a subfield of complexity that has attracted much attention in its own right. Artificial life is the philosophical heir of artificial intelligence, which preceded it by several decades. Whereas artificial-intelligence researchers seek to understand the mind by mimicking it on a computer, proponents of artificial life hope to gain insights into a broad range of biological phenomena. And just as artificial intelligence has generated more portentous rhetoric than tangible results, so has artificial life.

As Langton proclaimed in the inaugural issue of the journal *Artificial Life* last year, "Artificial life will teach us much about biology—much that we could not have learned by studying the natural products of biology alone—but artificial life will ultimately reach beyond biology, into a realm we do not yet have a name for, but which must include culture and our technology in an extended view of nature."

Langton has promulgated a view known as "strong a-life." If a programmer creates a world of "molecules" that—by following rules such as those of chemistry—spontaneously organize themselves into entities that eat, reproduce and evolve, Langton would consider those entities to be alive "even if it's in a computer." Inevitably, artificial life has begotten artificial societies. Joshua M. Epstein, a political scientist who shuttles between Santa Fe and the Brookings Institution in Washington, D.C., declares that computer simulations of warfare, trade and other social phenomena will "fundamentally change the way social science is done."

Artificial life—and the entire field of complexity—seems to be based on a seductive syllogism: There are simple sets of mathematical rules that when followed by a computer give rise to extremely complicated patterns. The world also contains many extremely complicated patterns. Conclusion: Simple rules underlie many extremely complicated phenomena in the world. With the help of powerful computers, scientists can root those rules out.

This syllogism was refuted in a brilliant paper published in *Science* last year. The authors, led by philosopher Naomi Oreskes of Dartmouth College, warn that "verification and validation of numerical models of natural systems is impossible." The only propositions that can be verified—that is, proved true—are those concerning "closed" systems, based on pure mathematics and logic. Natural systems are open: our knowledge of them is always partial, approximate, at best.

"Like a novel, a model may be convincing—it may ring true if it is consistent with our experience of the natural world," Oreskes and her colleagues state. "But just as we may wonder how much the characters in a novel are drawn from real life and how much is artifice, we might ask the same of a model: How much is based on observation and measurement of accessible phenomena, how much is based on informed judgment, and how much is convenience?"

Numerical models work particularly well in astronomy and physics because objects and forces conform to their mathematical definitions so precisely. Mathematical theories are less compelling when applied to more complex phenomena, notably anything in the biological realm. As the evolutionary biologist Ernst Mayr of Harvard University has pointed out, each organism is unique; each also changes from moment to moment. That is why biology has resisted mathematicization.

Langton, surprisingly, seems to accept the possibility that artificial life might not achieve the rigor of more old-fashioned research. Science, he suggests, may become less "linear" and more "poetic" in the future. "Poetry is a very nonlinear use of language, where the meaning is more than just the sum of the parts," Langton explains. "I just have the feeling that culturally there's going to be more of something like poetry in the future of science."

A Critique of Criticality

A-life may already have achieved this goal, according to the evolutionary biologist John Maynard Smith of the University of Sussex. Smith, who pioneered the use of mathematics in biology, took an early interest in work at the Santa Fe Institute and has twice spent a week visiting there. But he has concluded that artificial life is "basically a fact-free science." During his last visit, he recalls, "the only time a fact was mentioned was when I mentioned it, and that was considered to be in rather bad taste."

Not all complexologists accept that their field is doomed to become soft. Certainly not Per Bak, a physicist at Brookhaven National Laboratory who is on the Santa Fe faculty. The owlish, pugnacious Bak bristles with opinions. He asserts, for example, that particle physics and condensed-matter physics have passed their peaks. Chaos, too, had pretty much run its course by 1985, two years before James Gleick's blockbuster *Chaos* was published. "That's how things go!" Bak exclaims. "Once something reaches the masses, it's already over!" (Complexity, of course, is the exception to Bak's rule.)

Bak and others have developed what some consider to be the leading candidate for a unified theory of complexity: self-organized criticality. Bak's paradigmatic system is a sandpile. As one adds sand to the top of the pile, it "organizes" itself by means of avalanches into what Bak calls a critical state. If one plots the size and frequency of the avalanches, the results conform to a power law: the probability of avalanches decreases as their size increases.

Bak notes that many phenomena—including earthquakes, stock-market fluctuations, the extinction of species and even human brain waves—display this pattern. He concludes that "there *must* be a theory here." Such a theory could explain why small earthquakes are common and large ones uncommon, why species persist for millions of years and then vanish, why stock markets crash and why the human mind can respond so rapidly to incoming data.

"We can't explain everything about everything, but something about everything," Bak says. Work on complex systems, he adds, will bring about a "revolution" in such traditionally soft sciences as economics, psychology and evolutionary biology. "These things will be made into hard sciences in the next

Discord over Definitions

Can researchers create a unified theory of complex systems if they cannot agree on what complexity is? Seth Lloyd has compiled a list of 31 different ways to define complexity. Among them are:

Entropy. Complexity equals the entropy, or disorder, of a system, as measured by thermodynamics.

Information. Complexity equals the capacity of a system to "surprise," or inform, an observer.

Fractal dimension. The "fuzziness" of a system, the degree of detail it displays at smaller and smaller scales.

Effective complexity. The degree of "regularity" (rather than randomness) displayed by a system.

Hierarchical complexity. The diversity displayed by the different levels of a hierarchically structured system.

Grammatical complexity. The degree of universality of the language required to describe a system.

Thermodynamic depth. The amount of thermodynamic resources required to put a system together from scratch.

Time computational complexity. The time required for a computer to describe a system (or solve a problem).

Spatial computational complexity. The amount of computer memory required to describe a system.

Mutual information. The degree to which one part of a system contains information on, or resembles, other parts.

years in the same way that particle physics and solid-state physics were made hard sciences."

In his best-seller *Earth in the Balance,* Vice President Al Gore said Bak's theory had helped him to understand not only the fragility of the environment but also "change in my own life." But Sidney R. Nagel of the University of Chicago asserts that Bak's model does not even provide a very good description of a sandpile. He and other workers at Chicago found that their sandpile tended to oscillate between immobility and large-scale avalanches rather than displaying power-law behavior.

Bak retorts that other sandpile experiments confirm his model. Nevertheless, the model may be so general and so statistical in nature that it cannot really illuminate even those systems it describes. After all, many phenomena can be described by a Gaussian or bell curve. But few scientists would claim that human intelligence scores and the apparent luminosity of galaxies must derive from common causes. "If a theory applies to everything, it may really apply to nothing," remarks the Santa Fe researcher Crutchfield. "You need not only statistics but also mechanisms" in a useful theory, he adds.

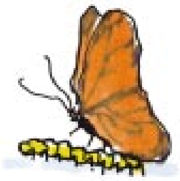
Another skeptic is Philip W. Anderson, a condensed-matter physicist and Nobel laureate at Princeton University who is on the SFI's board. In "More Is Different," an essay published in *Science* in 1972, Anderson contended that particle physics and indeed all reductionist approaches have only a limited ability to explain the world. Reality has a hierarchical structure, Anderson argued, with each level independent, to some degree, of the levels above and below. "At each stage, entirely new laws, concepts and generalizations are necessary, requiring inspiration and creativity to just as great a degree as in the previous one," Anderson noted. "Psychology is not applied biology, nor is biology applied chemistry."

"More is different" became a rallying cry for chaos and complexity. Ironically, Anderson's principle suggests that

Cybernetics and Other Catastrophes

Complexologists are not the first scientists in this century to think they could create a mathematical theory of, well, almost everything. Some notable predecessors:

Cybernetics. In his 1948 opus, *Cybernetics, or Control and Communication in the Animal and the Machine,* the mathematician Norbert Wiener sought to show how a theory based on feedback and other engineering concepts could explain the operation not only of machines but also of biological and social phenomena. Wiener's dream remains unfulfilled, but his neologism—based on the Greek term *kybernetes,* or steersman—remains enshrined in cyberculture.



Catastrophe theory. The French mathematician René Thom developed catastrophe theory as a purely mathematical formalism in the 1960s. But then Thom and others began claiming that it could provide insights into a broad range of phenomena displaying abrupt discontinuities, from the metamorphosis of a caterpillar into a butterfly to the collapse of civilizations. By the

The Other Complexologists

The Santa Fe Institute does not have a monopoly on big, complexity-related ideas. In fact, two other prominent complexologists claim recently to have achieved breakthroughs that transcend any work being done at Santa Fe.

One is Ilya Prigogine, a Belgian chemist. In 1977 Prigogine won a Nobel Prize for studies of so-called dissipative structures, notably "pumped" chemical cells that never achieve equilibrium but oscillate between multiple states. On these experiments, Prigogine, who oscillates between the International Solvay Institutes in Belgium and the University of Texas at Austin, has constructed a tower of ideas about self-organization, emergence, the links between order and disorder—in short, complexity.

Prigogine's great obsession is time. He contends that physics has not paid sufficient heed to the obvious fact that time proceeds in only one direction. Prigogine, now 78, has recently formulated a new theory that, he claims, finally does justice to the irreversible nature of reality. The probabilistic theory eliminates the philosophical paradoxes that have plagued quantum mechanics and reconciles it with classical mechanics, nonlinear dynamics and thermodynamics. As a bonus, Prigogine says, the theory will help bridge the chasm between the sciences and the humanities and bring about the "reenchantment" of nature.

Futurist Alvin Toffler (renowned mentor of Speaker of the House Newt Gingrich) has likened Prigogine to Isaac Newton and prophesies that the science of the Third Wave future will be Prigoginian. But many scientists assert that while Prigogine excels at waxing philosophical, he has made little or no concrete contribution to science. "I don't know of a single phenomenon he has explained," says Pierre C. Hohenberg of Yale University, a specialist in pattern formation.

Lurking in the wings is another rather large presence, Stephen Wolfram. The British-born physicist and wunderkind helped to spawn the field of complexity in the early 1980s before veering off to develop and market his successful software program, Mathematica.

Wolfram is disappointed with what has transpired at the Santa Fe Institute over the past decade. "I have to smile when I look at all the hypey books," he says. Wolfram reveals that he is writing a book that will resolve many of the field's fundamental problems, "but I can't discuss it yet."

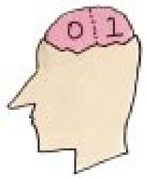
He then discusses it, just a little. Ever since the time of Newton, he says, science has been dominated by a belief in the power of mathematics—and differential equations in particular. In his book Wolfram will argue that simple sets of logical rules—those that give rise to cellular automata, for example—might provide a much more powerful language for describing reality.

Wolfram wants to avoid the exaggerated claims that have plagued the Santa Fe Institute. "I don't think what I produce will be a theory of everything," he remarks; he will be "very happy" if his achievement turns out to be merely as important as, say, Newton's discovery of calculus.

late 1970s, after a frenzy of interest, catastrophe theory had itself collapsed; one critic concluded that Thom's work "provides no new information about anything."

Chaos. Some chaos specialists insist that their field—which addresses systems that display bifurcation, sensitivity to initial conditions and other mathematically defined behavior—remains vital. But the French mathematician David Ruelle, a pioneer of the field, noted four years ago that "in spite of frequent triumphant announcements of 'novel' breakthroughs, [chaos] has had a declining output of interesting discoveries."

Information theory. Created by Claude E. Shannon in 1948, the theory provided a way to quantify the information content in a message. The hypothesis still serves as the theoretical foundation for information coding, compression, encryption and other aspects of information processing. Efforts to apply information theory to other fields, ranging from physics and biology to psychology and even the arts, have generally failed—in large part because the theory cannot address the issue of meaning.



MICHAEL CRAWFORD

these antireductionist efforts may never culminate in a unified theory of complex systems, one that illuminates everything from immune systems to economies. Anderson acknowledges as much. "I don't think there is a theory of everything," he comments. "I think there are basic principles that have very wide generality," such as quantum mechanics, statistical mechanics, thermodynamics and symmetry breaking. "But you mustn't give in to the temptation that when you have a good general principle at one level it's going to work at all levels."

Anderson favors the view of nature described by the evolutionary biologist Stephen Jay Gould of Harvard, who emphasizes that life is shaped less by deterministic laws than by contingent, unpredictable circumstances. "I guess the prejudice I'm trying to express is a prejudice in favor of natural history," Anderson says.

Anderson's views flatly contradict those of Stuart Kauffman, one of the most ambitious of all the artificial lifers. Kauffman has spent decades trying to show—through elaborate computer simulations—that Darwinian theory alone cannot account for the origin or subsequent evolution of life.

Kauffman says he shares the concern of his former teacher John Maynard Smith about the scientific content of some artificial-life research. "At some point," he explains, "artificial life drifts off into someplace where I cannot tell where the boundary is between talking about *the world*—I mean, everything out there—and really neat computer games and art forms and toys." When *he* does computer simulations, Kauffman adds, he is "always trying to figure out how something in the world works, or almost always."

Kauffman's simulations have led him to several conclusions. One is that when a system of simple chemicals reaches a certain level of complexity or interconnectedness (which Kauffman has linked both to the edge of chaos concept and to Bak's self-organized criticality), it undergoes a dramatic transition, or phase change. The molecules begin spontaneously combining to create larger molecules of increasing complexity and catalytic capability. Kauffman has argued that this process of "autocatalysis"—rather than the fortuitous formation of a molecule with the ability to replicate and evolve—led to life.

"Obscurantism and Mystification"

Kauffman has also proposed that arrays of interacting genes do not evolve randomly but converge toward a relatively small number of patterns, or "attractors," to use a term favored by chaos theorists. This ordering principle, which Kauffman calls "antichaos," may have played a larger role than did natural selection in guiding the evolution of life. More generally, Kauffman thinks his simulations may lead to the discovery of a "new fundamental force" that counteracts the universal drift toward disorder required by the second law of thermodynamics.

In a book to be published later this year, *At Home in the Universe,* Kauffman asserts that both the origin of life on the earth and its subsequent evolution were not "vastly improbable" but in some fundamental sense inevitable; life, perhaps similar to ours, almost certainly exists elsewhere in the universe. Of course, scientists have engaged in interminable debates over this question. Many have taken Kauffman's point of view. Others, like the great French biologist Jacques Monod, have insisted that life is indeed "vastly improbable." Given our lack of knowledge of life elsewhere, the issue is entirely a matter of opinion; all the computer simulations in the world cannot make it less so.

Kauffman's colleague Murray Gell-Mann, moreover, denies that science needs a new force to account for the emergence of order and complexity. In his 1994 book, *The Quark and the Jaguar,* Gell-Mann sketches a rather conventional—and reductionist—view of nature. The probabilistic nature of quantum mechanics allows the universe to unfold in an infinite number of ways, some of which generate conditions conducive to the appearance of complex phenomena. As for the second law of thermodynamics, it permits the temporary growth of order in relatively isolated, energy-driven systems, such as the earth.

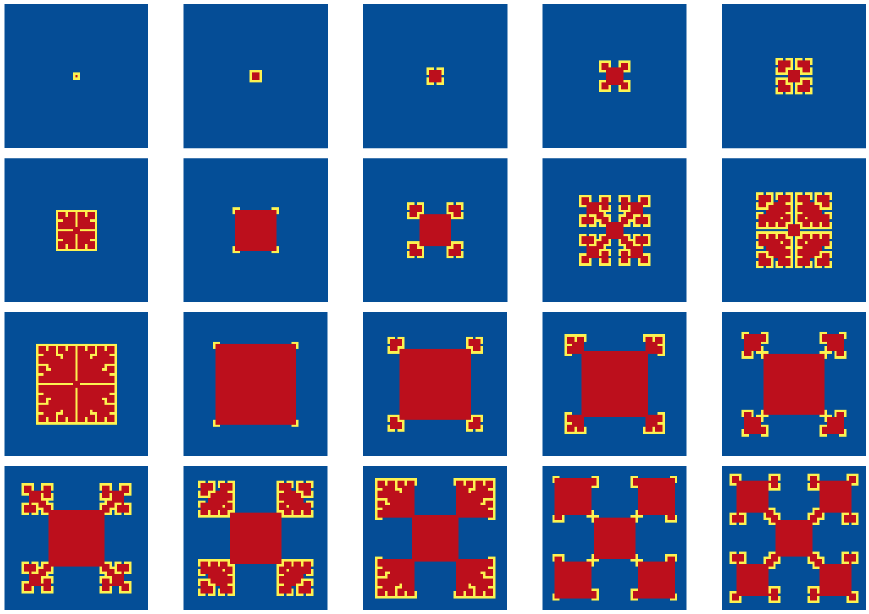
"When you look at the world that way, it just falls into place!" Gell-Mann cries. "You're not tortured by these strange questions anymore!" He emphasizes that researchers have much to learn about complex systems; that is why he helped to found the Santa Fe Institute. "What I'm trying to oppose," he says, "is a certain tendency toward obscurantism and mystification."

Maybe complexologists, even if they cannot create a science for the next millennium, can limn the borders of the knowable. The Santa Fe Institute seemed to raise that possibility last year when it hosted a symposium on "the limits of scientific knowledge." For three days, a score of scientists, mathematicians and philosophers debated whether it might be possible for science to know what it cannot know. After all, many of the most profound achievements of 20th-century science—the theory of relativity, quantum mechanics, Gödel's theorem, chaos theory—prescribe the limits of knowledge.

Some participants, particularly those associated with the institute, expressed the hope that as computers grow in power, so will science's ability to predict, control and understand nature. Others demurred. Roger N. Shepard, a psychologist at Stanford University, worried that even if we can capture nature's intricacies on computers, those models might themselves be so intricate that they elude human understanding. Francisco Antonio Doria, a Brazilian mathematician, smiled ruefully and murmured, "We go from complexity to perplexity." Everybody nodded.

THE AMATEUR SCIENTIST conducted by Alun L. Lloyd

Computing Bouts of the Prisoner's Dilemma



ALUN L. LLOYD

DEFECTORS* (red *and* yellow) *gradually invade a world of cooperators* (blue). *The game began with one defector at the center, with the score for cheating,* b, *set to 1.85.

Cooperating with others and exploiting them for personal gain are the two main ways members of a society can interact. To gain further insight into the social dynamics of competing individuals, researchers have formalized the choices within a mathematical framework known as game theory. They have devised various strategies about when to cooperate and have pitted those schemes against one another to determine the most successful. Such analyses indicate that cooperation often emerges naturally in simple societies; moreover, members can often fend off cutthroat exploiters from the outside [see "The Arithmetics of Mutual Help," by Martin A. Nowak, Robert M. May and Karl Sigmund, page 76].

Many of these notions derive from the classic game called the Prisoner's Dilemma. A player can cooperate with opponents or try to cheat them (called defecting). The opponent, of course, faces the same choice. With this game and some programming, a reader can explore the concept of mutual help, without going to prison.

I describe a spatial setup, where players inhabit the squares of an oversize chessboard and spend their time repeatedly playing rounds of the Prisoner's Dilemma against their neighbors. To simplify the programming, I ignored corners and edges of the chessboard and instead considered the squares to wrap back around on themselves. All the games in each round are played at the same time.

Within each round, every player takes on, one at a time, its eight nearest neighbors and itself (this self-interaction is included to make the computer program simpler). The players earn points depending on the strategy they and their opponents play. Each one gets a point if both cooperate; none if both defect. The player receives nothing if it cooperates and the opponent defects. The highest score, which I have labeled *b,* is for cheating (a player defects as the opponent cooperates). The value of *b* will ultimately control the outcome of the game. Just pick a value greater than 1; I used 1.85. A table known as a payoff matrix summarizes the scoring; it lists the rewards for the four different possible combinations of strategies [*see illustration on page 112*].

The nine payoffs resulting from the play are added up to give each player's score in that round. Each player then looks to see if any of its eight neighbors earned a score higher than it did. If so, the player will adopt the more successful strategy for the next round. For instance, if the opponent with the highest score in the player's neighborhood cooperated, the player will cooperate in the next round.

If you play the game once, against just one other player, your best choice is to defect. Betrayal maximizes your score regardless of what your opponent does. In the spatial game, however, the outcome is harder to predict because the strategy each player adopts in the next round depends on the scores of its eight neighbors as well as its own. In turn, each neighbor's score depends on its nearest neighbors, which means that the configuration of the nearest 24 players affects the outcome at each square of the chessboard.

I wrote the program in a dialect of the BASIC language called QBASIC, which comes with recent versions of MS-DOS for PC-compatible computers. It will work with Microsoft's QuickBASIC on the PC or the Macintosh and can easily be converted to run with other computer languages. The code is listed on page 114, and the explanation of its function appears in the box on page 112.

Once the program is working, you can adjust several parameters to influence the outcomes. As I mentioned, the game turns largely on the value of *b,* the advantage for hoodwinking. As *b* increases, defectors do better when they come across cooperators. Surrounding defectors always exploit lone cooperators, which always do worse. But groups of cooperators can help one another to flourish. One example is a square of four cooperators in a sea of defectors. Each cooperator will score four points (one from playing with each of its three

ALUN L. LLOYD is a graduate student in the department of zoology at the University of Oxford. His interests include mathematical biology and nonlinear dynamics (that is, chaos). His research is supported by the Wellcome Trust.

neighbors and one from playing against itself), but the defectors neighboring the cooperators will score at most $2b$ because they are next to, at most, two cooperators. If b is not too large, cooperators will score higher than defectors. Lone defectors will do well because they are surrounded by cooperators.

The outcome of the competition between such different circumstances is that clusters of defectors and cooperators are seen to grow and shrink, to collide and tear one another apart. Often a dynamic equilibrium results. From a distance, an overall pattern is discernible, but individual squares are constantly changing.

Because each square's payoff is some multiple of b plus some multiple of 1, there is only a certain set of b values for which the behavior changes. The first experiment you could carry out is to run the program for different b values between 1 and 3. Try, for instance, 1.15, 1.35, 1.55, 1.77, 1.9 and 2.01. For the smaller b values, the defectors tend to be isolated, but for larger ones they form connected structures.

Once you have seen some of the different behaviors, you may want to describe them in more quantitative terms. Count the numbers of cooperators and defectors (identified by a total of four colors, depending on the strategies played over two rounds) in each round. How do these quantities vary as the contest continues? For some b values, these frequencies get progressively closer to some fixed value; for others, they vary periodically (for instance, taking one of two values, depending on whether the round is even or odd) or even in an un-

BASIC Ideas of Cooperating and Defecting

Getting the most out of the Prisoner's Dilemma games means understanding the construction of the program. The program keeps track of the players and their scores by organizing them into arrays.

Each square on the chessboard is labeled by its row and column number, written as a pair (i,j). I chose the number of squares, N, on the chessboard to be 60. The strategies are labeled by numbers: 1 for cooperation and 2 for defection. The schemes adopted by each player can then be recorded as an array of numbers, $s(i,j)$. The payoffs are recorded in another array, $pm(x,y)$—which stands for payoff matrix—where x is the strategy a player adopts and y the strategy the opponent plays. Hence, a player adopting a strategy of defecting (2) against an opponent cooperating (1) will receive $pm(2,1)$, which equals b, the score for cheating.

Before play begins, the program decides which strategy each player will use in the first round. For every square, the computer picks a number at random between 0 and 1. If it is less than a certain value—say, 0.1—then the program places a defector on that square; otherwise, a cooperator goes there. The cutoff number roughly indicates the proportion of players who will defect in the first round.

The program goes through every square (i,j) on the board, calculating each player's payoff and recording it as $payoff(i,j)$. The payoff is worked out by adding the scores from the games with the nine players, whose positions relative to the player on (i,j) are given by $(i+k, j+l)$, where k and l take the values −1, 0 or 1.

Notice in the program there is a slight complication because the chessboard wraps around on itself. If j equals 1, we are looking at the first column; its neighbors to the "left" actually lie in the last column. Similar troubles plague the last column and the first and last rows. To solve the problem, I included an array, $bc(m)$—for boundary conditions—which redirects the computer in these cases.

Now that we know the position of a player (i,j) and each neighbor $(bc(i+k),bc(j+l))$, the array $s(i,j)$ indicates the strategies each will play. The payoff from this single game can be calculated by looking in the array containing the payoff matrix: $pm(s(i,j), s(bc(i+k),bc(j+l)))$. Nine such payoffs are summed before moving to the next square on the board.

Once the payoff for every square has been calculated, the program finds the most successful strategy in each neighborhood. Then it updates the array of strategies accordingly, storing these new strategies in the array $sn(i,j)$. The program goes through every square, recording its payoff in the variable hp, for highest payoff, and its strategy in $sn(i,j)$. The program then looks at the neighboring squares in turn. If one of the neighbor's payoff is greater than hp, hp is set equal to that payoff, and the neighbor's strategy is recorded in $sn(i,j)$. Once all the neighbors have been examined, the variable hp contains the highest value of the payoff in the neighborhood, and $sn(i,j)$ has the strategy used by that player.

After all the new strategies, $sn(i,j)$, have been decided, they are copied into the array $s(i,j)$, and the next round begins. The progress of the game is followed by coloring a square grid of points on the screen. The coloring depends on the strategy each player adopts in the current and previous rounds. So there are four colors: blue (is cooperating, did cooperate), red (is defecting, did defect), green (is cooperating, did defect) and yellow (is defecting, did cooperate). Using this scheme, we can see not only which players are cooperators (blue and green) or defectors (red and yellow) but also which players' strategies are changing (green and yellow) and which are not changing (red and blue). In the program, the array $c(x,y)$ tells the computer which colors represent the previous and current strategies (x and y, respectively).

If the program runs too slowly, reduce the size of the board. Doubling the number of squares means that the program will take approximately four times as long to play each round. If you do wish to increase the board size, you may have to increase the sizes of the arrays s, sn and bc accordingly.

PLAYER'S STRATEGY \ OPPONENT'S STRATEGY	COOPERATE	DEFECT
COOPERATE	1	0
DEFECT	b	0

The payoff matrix

SCIENTIFIC
AMERICAN

CORRESPONDENCE

Reprints are available; to order, write Reprint Department, Scientific American, 415 Madison Avenue, New York, NY 10017-1111, or fax inquiries to (212) 355-0408.

Back issues: $8.95 each ($9.95 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, NY 10017-1111, fax (212) 980-8175 or call (800) 777-0444.

***Scientific American*-branded products** available. For free catalogue, write Scientific American Selections, P.O. Box 11314, Des Moines, IA 50340-1314, or call (800) 777-0444.

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., 222 Rosewood Drive, Danvers, MA 01923. Specify CCC Reference Number ISSN 0036-8733/95. $3.00 + 0.50.

Editorial correspondence should be addressed to The Editors, Scientific American, 415 Madison Avenue, New York, NY 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, NY 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. E-mail: SCAinquiry@aol.com.

The Program for the Prisoner's Dilemma

```
DEFINT C, I-N, S                         define variables, array sizes
DEFSNG B, H, P
DIM s(120, 120), sn(120, 120)
DIM bc(121), c(2, 2)
DIM payoff(120, 120)

LET b = 1.85                             advantage for cheating
LET N = 60                               size of board
LET p = 0.1                              proportion of defectors
LET pm(1, 1) = 1                         set up payoff matrix
LET pm(1, 2) = 0
LET pm(2, 1) = b
LET pm(2, 2) = 0
                                         set up colors
LET c(1, 1) = 1                          = 409 on Mac
LET c(2, 2) = 4                          = 205 on Mac
LET c(1, 2) = 2                          = 341 on Mac
LET c(2, 1) = 14                         = 69 on Mac

RANDOMIZE TIMER                          initialize board
FOR i = 1 TO N
 FOR j = 1 TO N
  LET s(i, j) = 1
  IF (RND < p) THEN LET s(i, j) = 2
NEXT j, i

FOR i = 1 TO N                           set up boundary conditions
 LET bc(i) = i                           no problem if i between 1 and N
NEXT i
LET bc(0) = N                            redirect neighbors of edges
LET bc(N + 1) = 1

SCREEN 12                                not needed on Mac

FOR M = 1 TO 1000                        begin playing game

 FOR i = 1 TO N                          calculate payoffs for each player
  FOR j = 1 TO N
   LET pa = 0
   FOR k = -1 TO 1
    FOR l = -1 TO 1
     LET pa = pa + pm(s(i, j), s(bc(i + k), bc(j + l)))
   NEXT l, k
   LET payoff(i, j) = pa
 NEXT j, i

 FOR i = 1 TO N                          find largest payoff in each
  FOR j = 1 TO N                         neighborhood and calculate
   LET hp = payoff(i, j)                 new strategies
   LET sn(i, j) = s(i, j)
    FOR k = -1 TO 1
     FOR l = -1 TO 1
       IF payoff(bc(i + k), bc(j + l)) > hp THEN
         LET hp = payoff(bc(i + k), bc(j + l))
         LET sn(i, j) = s(bc(i + k), bc(j + l))
       END IF
    NEXT l, k
 NEXT j, i

 FOR i = 1 TO N                          display strategies
  FOR j = 1 TO N
   COLOR (c(sn(i, j), s(i, j)))          ForeColor on Mac
   PSET (i, j)
   LET s(i, j) = sn(i, j)
  NEXT j, i
NEXT M

END
```

predictable fashion. Calculate the fraction of players that change their strategies in each round. If it equals 0, you have reached a static equilibrium. Graphically, the situation shows up as all squares being two colors. (Such a pattern occurs when *b* is set to 2.01.)

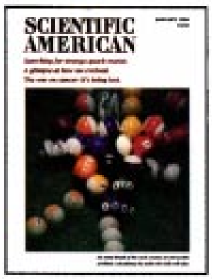
Is the initial proportion of defectors to cooperators important for these results? Try running the program several times with different ratios. The clustering of defectors and cooperators is also crucial. If the number of cooperators is small, on some occasions they may be clustered and can flourish. On others, they may be isolated and are doomed.

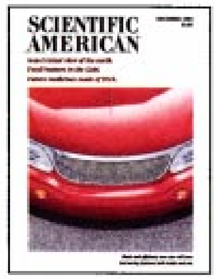
This program can generate some pretty patterns if the initial configuration is symmetrical. It is easy to modify the code so that every square at first is a cooperator. Remove the line in the program that decides whether a square should start off as a defector. Now set just one square, near the center, to be a defector. (Insert the line $s(30,30) = 2$, for instance, just after the loop that now sets all squares to be cooperators.) Choose a b value between 1.8 and 2 and start the program. You should see that the single defector can invade the world of cooperators [*see illustration on page 110*].

Another easy modification is to play the game only against the eight neighbors. To do this, do not add on the payoff when the square plays against itself (when $k = 0$ and $l = 0$ in the program). Similar behaviors will be seen but with slightly different b values.

Or try different boundary conditions—for instance, unwrap the chessboard and let the edge players compete against their five neighbors and the corner players their three. This scenario can be messier to program, as edge and corner squares must be treated differently. Another alternative would be to insist that all corner and edge squares always cooperate or always defect. An interesting change would be to play the game against only the four nearest neighbors (up, down, left and right). For those of you who like a challenge, try setting up the game on a honeycomb-pattern board, so that each square has six neighbors.

Further explorations could include altering more than one entry in the payoff matrix. Small changes, such as setting the entry at the bottom right in the payoff matrix (both player and opponent defecting) to a tiny positive value such as 0.01, do not alter overall behavior very much. More radical changes will have a noticeable effect. The number of possible manipulations of the program is almost limitless, as will be the patterns produced.





REVIEWS

Next of Kin

Review by John Mitani

KANZI: THE APE AT THE BRINK OF THE HUMAN MIND, by Sue Savage-Rumbaugh and Roger Lewin. John Wiley & Sons, 1994 ($24.95).

The more that scientists have studied primates, the harder they have found it to identify the traits that distinguish us from our evolutionary cousins. Language seems to represent one of the last bastions of human uniqueness. But in *Kanzi,* primatologist Sue Savage-Rumbaugh of Georgia State University and writer Roger Lewin launch an assault on the proposition that even human speech is special.

Kanzi's opening chapter clearly states the central question: Does a qualitative difference exist between humans and other animals? Opinions are highly polarized. Many social scientists adhere to a strict mind-body dichotomy. While acknowledging that evolutionary continuity links humans and animals in their physical features, this group suggests that mental experiences are not comparably shared. That viewpoint is especially prevalent among linguists, in part because of the influential work of Noam Chomsky, who has argued that conscious thought is made possible by the specifically human attribute of language.

Savage-Rumbaugh subscribes to an alternative, evolutionary perspective, which assumes that other animals experience many of the same mental states we do. She has tackled the mind-body dichotomy head-on by teaching captive apes a languagelike form of referential communication based on lexigrams, a system of computer-controlled symbols.

In her new book, Savage-Rumbaugh and Lewin summarize the findings of her 20-year research program, highlighting recent studies involving a male bonobo, or pygmy chimpanzee, named Kanzi. The authors relate Savage-Rumbaugh's initial skepticism of the strong claims made regarding some of the early, highly publicized attempts to teach apes sign language. Those studies received withering attacks from linguists and psychologists, who noted that subjects appeared to be mimicking researchers rather than producing spontaneous utterances. Savage-Rumbaugh was one of the few to take up the challenge of redesigning ape language studies, beginning first by questioning some of the premises underlying previous research.

The Chomskyan notion that humans stand apart from all other animals derives from the proposition that we possess a unique, innate mechanism to decode the syntactical structure of language. An emphasis on syntax as the key to human linguistic ability led several investigators to search for similar production abilities in apes. In contrast, Savage-Rumbaugh began by clearly distinguishing language production from comprehension. She has consistently emphasized that the crucial precursor of language competence is the comprehension of nonspoken referential symbols. As a result, she has focused on what a word means to apes, rather than on how these animals produce words.

Kanzi chronicles Savage-Rumbaugh's initial efforts with two male chimpanzees, Sherman and Austin. In their early research Savage-Rumbaugh and her colleagues demonstrated that these chimps could do far more than imitate their teachers. The chimps were able to inform one another and to take turns in communicating during food-sharing tasks—a remarkable behavioral breakthrough, given that such sharing is exceedingly rare in the wild. This research underscored the difference between language production and comprehension and suggested that a continuing emphasis on the latter might facilitate further linguistic development in apes.

The authors next turn to Savage-Rumbaugh's more recent studies conducted with bonobos. Bonobos display some unusual social and sexual habits: they appear to be more socially attuned and far calmer than chimpanzees, and unlike chimps, bonobos use sex in a variety of contexts, including food sharing. Savage-Rumbaugh ran into difficulties during the language training sessions with Matata, a bonobo female. Living up to her Kiswahili name (which translates as "problem"), Matata failed to master the system of lexigrams.

These sessions produced a highly fortuitous result, however; they exposed Matata's young, adopted son, Kanzi, to the artificial language. Kanzi proved able to comprehend and spontaneously produce a limited number of lexigrams without training. This startling turn of events led to a shift in effort—a move away from rigorous training sessions and toward a more laissez-faire approach in which Kanzi was treated as a developing human infant, acquiring what he

KANZI seems capable of comprehending some basic rules of syntax. But are his mental experiences really similar to ours?

wanted to learn, when he wanted to learn it.

This "training" regime proved strikingly successful. After four months Kanzi had acquired 20 symbols; in 17 months he used 50 symbols. Kanzi also could produce combinations of words spontaneously. Building on this latter finding, Savage-Rumbaugh and Patricia Greenfield of the University of California at Los Angeles investigated whether Kanzi had achieved the linguist's holy grail, syntax. Detailed examination of Kanzi's utterances has brought Savage-Rumbaugh to the exciting conclusion that the bonobo does possess a rudimentary syntactical ability; Kanzi appears to be sensitive to word order and other syntactical cues.

The methods employed in ape language studies are finding practical application in teaching communication skills to children who have severe mental disabilities. The authors view this work as a logical extension of their research with Kanzi. The implied similarity of language ability leads naturally into a final reiteration of Savage-Rumbaugh's assertion that mental continuity exists between apes and humans.

The results reported about Kanzi will doubtless stir controversy. Some researchers have already rejected Savage-Rumbaugh's finding that Kanzi spontaneously acquires and uses language. The question of true animal "language" arouses such highly antagonistic reactions that one wonders whether advances in understanding will ever be possible. Many linguists and other social scientists will continue to hold fast to a philosophical stance that does not permit them to accept Savage-Rumbaugh's claims. In contrast, a growing number of cognitive psychologists and like-minded ethologists will applaud these reports suggesting that animals possess advanced cognitive skills.

Is there hope for a resolution? If one begins with a priori assumptions about what animals can or cannot do, scientific inquiry effectively comes to a halt. In the absence of further data, it seems prudent to put the questions to the animals themselves, in the manner that Savage-Rumbaugh has pioneered. It will also be important to refrain from overinterpreting the results and to begin thorough investigations into the limits of animal cognition. Such research may guide a new generation of studies—one that promises to illuminate the differences as well as the similarities between apes and humans—that will clarify our place in nature.

JOHN MITANI is professor of anthropology at the University of Michigan.

Mind out of Matter

Review by David J. Chalmers

SHADOWS OF THE MIND: A SEARCH FOR THE MISSING SCIENCE OF CONSCIOUSNESS, by Roger Penrose. Oxford University Press, 1994 ($25).

What are the most important unsolved problems in late 20th-century science? Different people would make different choices, but a consensus list would surely include the following: Can computers be as intelligent as people? How can we make sense of the reality underlying quantum mechanics? And what is the physical basis of consciousness? In his latest book the British mathematical physicist Roger Penrose suggests that these three questions are intimately related and has the audacity to take on all three at once.

Shadows of the Mind offers more detailed and concrete versions of arguments Penrose raised in his 1989 book, *The Emperor's New Mind.* His central claim is that human thinkers can do things that no computer could ever do. He rests his argument on Gödel's theorem, which states that for every consistent formal system that has the power to do arithmetic, there will always be a true statement—the "Gödel sentence"—that the system cannot prove. (A formal system is a set of logical or computational rules; it is termed consistent if it never produces contradictory statements.) Yet we humans can see that the statement is true, which Penrose, like many others before him, takes as a sign that our mind can go beyond the powers of any formal system.

The natural reply is that humans cannot always recognize the truth of arbitrary Gödel sentences, either. In logical terms, we cannot see that a system's Gödel sentence is true unless we can first determine that the system is consistent. There is no reason to believe we can always do this, so humans may be limited in the same way that formal systems such as computers are. Perhaps each one of us is a complex formal system that cannot determine that system's consistency; given the complexity of the brain, that inability would not seem surprising. In that case, Penrose will never perceive the truth of his own Gödel sentence.

But Penrose has a further move ready. In a fantasy dialogue with a robot mathematician, he advances a hypothetical argument about what he could prove if he knew he was identical to a given formal system, F. Using some intriguing self-referential reasoning, he asserts that he could then know the truth of a Gödel statement that F itself could never prove—even if F knew its own identity. If he is right, he cannot be identical to F after all. Generalizing this logic, Penrose concludes that he cannot be identical to *any* formal system.

Esoteric though it may seem, this argument is stronger than the one that went before. Its greatest weakness may lie in Penrose's assumption that he can know unassailably that he himself is consistent. This assumption may lead to a paradox in its own right, in which case it must be withdrawn. Without the assumption, the argument fails.

If Penrose is right, however, there must be some noncomputable element in the physical processes of the brain. Yet nothing in existing physical theories accounts for such noncomputable processes. That tension drives the book's stimulating second half, in which Penrose speculates on the future of physics and its relation to the science of the mind. These chapters follow Penrose's search for the two elements that his theory needs: a noncomputable element in the laws of physics and a mechanism in the brain that can exploit it.

In physics, Penrose pins his hopes on future elaborations of the theory of quantum mechanics. Like many other scientists and philosophers, he believes no current theory gives an acceptable picture of quantum-mechanical reality. In particular, none of these theories offers a compelling account of why a quantum wave function (the mathemat-

MECHANICAL BRAIN could not produce conscious thought, argues Roger Penrose.

ical function describing the position and momentum of a particle) sometimes collapses into a discrete state, as quantum mechanics says it must. Penrose rejects the idea that this collapse happens only on observation by a conscious observer, and he also dismisses exotic interpretations that avoid collapse entirely. Instead he proposes that new fundamental laws are required to explain the process by which collapse occurs. It is here he thinks a noncomputable element might be found.

Penrose suspects that the new laws will derive from efforts to reconcile quantum mechanics with general relativity. Combining the two theories into a new description of "quantum gravity" carries one to the conclusion that space-time can sometimes double up on top of itself. The collapse of a wave function may occur when this superposition of space-time becomes too intense, Penrose speculates.

Nothing about his proposal implies that the process of collapse is noncomputable, however. The only real impetus for thinking it is comes from Penrose's original argument about human abilities and Gödel's theorem. It is not often that theories of physics are driven by theories of the mind! Even if noncomputability turns out to be a red herring, however, Penrose's speculations on physical theories remain of considerable interest in their own right.

In his final step Penrose attempts to close the circle and show how these noncomputable physical laws could be relevant to the science of the mind. Drawing on the ideas of the anesthesiologist Stuart Hameroff, Penrose suggests that the effects of quantum gravity are mediated by microtubules, protein structures found in the skeletons of neurons. Microtubules are small enough that quantum effects might conceivably influence their functioning but large enough that they could in turn affect the way entire neurons behave. If so, the noncomputable process of quantum collapse within a microtubule might be amplified into macroscopic, noncomputable processes in the workings of the brain.

The direct support for this hypothesis seems surprisingly thin. Penrose says little about just *how* quantum action in a microtubule might mediate neural functioning, and he offers no neurobiological data that his hypothesis might explain. For now, these ideas must be taken merely as interesting musings.

And what of human consciousness, Penrose's ultimate target? At the start of *Shadows of the Mind,* he notes that the biggest mystery of all is how electrical activity in the brain gives rise to the experience of consciousness. It is hard to understand why an inner life should arise from the mere enaction of a computation, no matter how complex. But Penrose's alternative—quantum processes in microtubules—does not seem any more satisfying. Why should consciousness emerge from collapsing wave functions in microtubules? At best, Penrose's thoughts about quantum gravity might help explain certain aspects of human behavior, such as our ability to prove mathematical theorems. That this behavior is accompanied by conscious experience remains as perplexing as ever. For all Penrose's work, the science of consciousness is still missing its most important component.

Many readers will remain unconvinced that Penrose has refuted artificial intelligence and unlocked the secrets of the mind. But this is the kind of book that is successful whether or not it is convincing. Its function is to stimulate and to challenge, and here it succeeds completely. It is impossible to read *Shadows of the Mind* without experiencing a sense of wonder at the puzzles of modern science and the powers of the human mind.

DAVID J. CHALMERS is McDonnell Fellow in Philosophy, Neuroscience and Psychology at Washington University.

The Stars on a Platter

Review by Timothy Ferris

REDSHIFT: MULTIMEDIA ASTRONOMY. Maris Multimedia, 1993–1994 ($64.95, Macintosh or Windows). **ASTRONOMER.** Expert Software, 1994 ($14.95, Windows). **HYPERSKY.** Willmann-Bell, 1993 ($49.95, DOS or Windows; CD-ROM data set an additional $79.95). **BEYOND PLANET EARTH.** Discovery Channel, 1993 ($49.95, Windows). **JOURNEY TO THE PLANETS.** Multicom Publishing, 1994 ($34.95, Macintosh or Windows).

Multimedia technology has routinely been touted as a formidable educational tool, capable of combining the intellectual depth of a university library with the emotional appeal of a Hollywood movie. Yet to date, many of the multimedia programs that aim to inform have proved in actuality to be dull, shallow and technically unstable, lending credence to the cynical gibe that CD-ROM is the technology of the future—and always will be. These five astronomy programs betray many of multimedia's current limitations. But a couple of them also hint at the estimable power this emerging medium can put in the hands of nonspecialists who want to use their computers to explore the night sky.

The standout here is *RedShift,* a versatile "planetarium program"—meaning that it can display the night sky in a fashion useful to visual observers and to those equipped with small telescopes. It also sports such multimedia features as the ability to record and play back astronomy-related movies and to display more than 700 high-resolution color images of planets, nebulae and galaxies. The disk's vast database contains roughly a quarter of a million stars, 40,000 deep-space objects, 5,000 asteroids and 100 periodic comets.

RedShift's display looks almost as serious as a 747 cockpit, but it is easy to use. It can be set up to provide a wealth of information, including the current solar and sidereal time, horizon position and celestial coordinates. A zoom function permits full-screen depictions of planetary disks that accurately portray such specifics as the locations of Jupiter's satellites or which hemisphere of Mars currently faces the earth.

Part of the fascination of a competent planetarium program resides in its ability to re-create the night sky at times and places beyond the horizons of living experience. Here, too, *RedShift* is facile yet reliable. In one test, the program readily confirmed the story told by Ferdinand Columbus that he and his father, Christopher, observed a lunar eclipse that began "with the rising of the moon" on the evening of February 29, 1504. The elder Columbus exploited that event to terrify the locals on Jamaica, threatening that God would not restore the moon unless they provisioned his empty ships. The publisher credibly states that *RedShift* can predict the locations of the inner planets to within 30 arc-seconds on any date between 4712 B.C. and A.D. 10,000.

Like many multimedia programs, *RedShift* can be a bit cranky to install, compelling users to study the arcana of video drivers and configuration settings. But once up and running, it rewards the viewer with such treats as an eye-popping movie of Jupiter as seen from a point near the orbit of Io and compellingly realistic views of the earth as seen from any desired perspective. The mapping software that makes these graphics possible, written in Kaliningrad, Russia, by a team of computer programmers who had worked at Russian Space Mission Control, ranks with some of the best computer graphics to have come out of the Jet Propulsion Laboratory in Pasadena, Calif.

The still photographs in *RedShift,* though appealing, leave one wishing they were more plentiful and of more consistent quality; the accompanying text could be lengthier and more carefully composed. (A particularly amusing gaffe attributes a double-quasar image to "gravitational lending," a trick one hopes does not catch on with Congress.) But even in its current, somewhat flawed incarnation, *RedShift* bears the earmarks of a valuable resource for students, amateur astronomers and anyone with more than a superficial interest in the night sky.

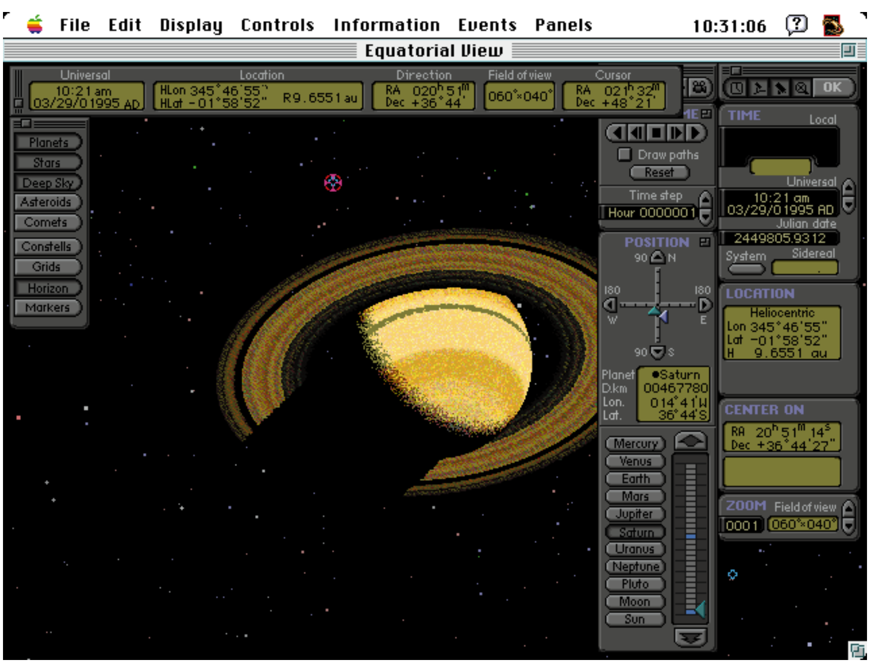
Expert's *Astronomer,* a less ambitious CD-ROM, makes no pretense of matching *RedShift*'s accuracy—and indeed, *Astronomer* errs by several hours in computing the timing of the occultation observed by Copernicus. Its night-sky display is also relatively rudimentary. And the disk's multimedia side, which is actually a separate program, amounts to little more than a set of digitized slides accompanied by brief text. But *Astronomer* is fairly cheap and simple and may be suitable for those who are not yet comfortable with mastering the likes of *RedShift.*

HyperSky's strength is vast quantities of raw data: this CD-ROM contains all 15 million stars listed in the *Hubble Space Telescope Guide Star Catalog,* plus more than three million galaxies, nebulae, star clusters and quasars. That is almost two orders of magnitude more data than *RedShift* provides—enough to have filled a small observatory library in predigital days. Unfortunately, *HyperSky*'s interface is cumbersome, and the injudicious application of a "display" command can leave even users of fast workstations twiddling their thumbs while the program laboriously paints the screen with star images packed as dense as beach sand. But the program offers a wealth of information for the serious amateur or professional astronomer sophisticated enough to know exactly what he or she is after.

Beyond Planet Earth and *Journey to the Planets* suffer from the drawbacks that have led people to associate "multimedia" with a bland corporate emphasis on style over substance. Both disks emphasize a few glitzy video sequences accompanied by synthesizer music and pretentious narrations that generate more heat than light. Beyond that, one finds mostly a wasteland of animation, still pictures, music, text and spoken words that, for all their bluster, conveys less information than even a mediocre textbook.

To maneuver at the requisite waterbug depth over this halfhearted potpourri is to be reminded that television and movies, which have been doing this kind of thing for decades, can be mightily emotive but have generally proved ineffective as tools for imparting in-depth knowledge. If multimedia technology is to realize its oft-stated promise, the programmers still have a long row to hoe.

TIMOTHY FERRIS is a faculty member at the University of California, Berkeley, and author of The Mind's Sky.



REDSHIFT ***packs impressive graphics and serious science onto a CD-ROM.***

ESSAY by Ralph E. Gomory

The Known, the Unknown and the Unknowable

We are all taught what is known, but we rarely learn about what is not known, and we almost never learn about the unknowable. That bias can lead to misconceptions about the world around us.

The known is pressed on us from the first. In school we start each course at the beginning of a long book full of things that are known but that *we* do not yet know. We understand that beyond that book lies another book and that beyond that course lies another course. The frontier of knowledge, where it finally borders on the unknown, seems far away and irrelevant, separated from us by an apparently endless expanse of the known. We do not see that we may be proceeding down a narrow path of knowledge and that if we look slightly left or right we will be staring directly at the unknown.

Even when we are right on the edge of the unknown, we may not be aware of it. Those of us who learned the history of the Persian Wars in school did not know that the events so vividly described are all based on the writings of the one source who survived—Herodotus. If you want to know almost anything that happened in the Greece of that time and it was not recorded by Herodotus, it is unknown and in all probability can never be known. But we did not think of his accounts as fragments of knowledge on the edge of the unknown; it was just more stuff from the huge pile of facts we had to learn about the history of Greece.

Because of such lessons, we grow up thinking more is known than actually is. If we had a better description of the limits of present knowledge, that description could be a part of what we are taught. Such insight would give us a better perspective on what is known and what is currently unknown.

In time, many things now unknown will become known. We will learn more about what lies below the surface of the earth, and we may learn how neurons interact to let us perceive and think. The accumulating pile of data can be misleading, however. Beyond the currently unknown are the things that are inherently unknowable.

Few unknowables are consciously recognized as such. The outcome of a spinning roulette wheel and the local weather three months from now belong to that small class. Every day, however, we bump into phenomena that may well be unknowable but that we do not recognize as such. Some of these unknowables form the bases of respected professions. Brokers make a living anticipating the fluctuations of stock prices. Presidents run for office based on claims of what they will do for a vast and poorly understood economy composed of many unpredictably interacting sectors. We do not even know if we are dealing here with the partly known, the mainly unknown or the unknowable.

16TH-CENTURY MAP reveals its unknowns only in hindsight.

NATIONAL MARITIME MUSEUM, GREENWICH

Nevertheless, we unconsciously recognize that the unknowable surrounds us. Nobody thinks about or pretends to know who will run for president 20 years from now. Nor do people try to predict the automobile accidents they will be involved in. To know that we will be struck by a car next year, we would have to know, with impossible accuracy, the particulars of the life of the driver, his habits, his timing, his way of pressing the accelerator and so forth—all the facts that are needed to bring him with perfect precision to that unpleasant encounter. It is clear that all these details are unknown, and we do not try very hard to learn about them because we instinctively realize they are also unknowable.

In distinguishing the known or the unknown from the unknowable, the level of detail can be decisive. The level of detail is what separates the delusion of the gambler from the wealth of the casino owner. The gambler attempts to predict the individual and unpredictable spins of the roulette wheel; the owner concerns himself with the quite predictable average outcome.

The prediction process is aided by the fact that the artificial is generally simpler than the natural. The roll of a bowling ball down an alley, for instance, is easier to predict than the motion of an irregular stone tumbling down a rough hillside. It is likely that the artificial will increasingly save us from the unpredictable. It may be easier to move gradually toward a completely enclosed earth whose climate could be artificially controlled than to learn to predict the natural weather.

It is in creating the artificial and controllable that science excels. Science and engineering have made it possible to construct the partially artificial surroundings we live in today, replete with huge bridges, trucks, airplanes, antibiotics and genetically altered species. We are likely to build an increasingly artificial, and hence increasingly knowable, world.

Two limitations may constrain the march of predictability. First, as the artifacts of science and engineering grow ever larger and more complex, they may themselves become unpredictable. Large pieces of software, as they are expanded and amended, can develop a degree of complexity reminiscent of natural objects, and they can and do behave in disturbing and unpredictable ways. And second, embedded within our increasingly artificial world will be large numbers of complex and thoroughly idiosyncratic humans.

RALPH E. GOMORY is president of the Alfred P. Sloan Foundation.