SCIENTIFIC AMERICAN

JULY 1994
$3.95

How cancers defeat drug therapy.

Re-creating the origin of life.

Can science study consciousness?

***The moon was born** when a Mars-size planetoid collided with the young earth.*

SCIENTIFIC AMERICAN

July 1994 Volume 271 Number 1

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. 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). Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: Write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

DEPARTMENTS

THE COVER painting depicts the process that probably created the moon: a monumental impact. A massive projectile, perhaps as large as Mars, crashed into a young earth 4.5 billion years ago. The dust and molten debris spewed into orbit eventually accreted to form the moon. The blow is also thought to have sped up the earth's rotation to its current period. The giant impact hypothesis is just one of several ideas about the moon and the earth that emerged after the first lunar landing 25 years ago (see "The Scientific Legacy of Apollo," by G. Jeffrey Taylor, page 40).

THE ILLUSTRATIONS

Cover painting by Alfred T. Kamajian

Page	Source
31	Francesca Bray
32	Paul J. Buklarewicz
33	Joanie Popeo
34–35	Joanie Popeo (*drawings*)
36	Johnny Johnson
37	Hiroji Motowaka/ Nature Production
40–41	National Aeronautics and Space Administration
42–43	Tomo Narashima
44	NASA (*left*), G. Jeffrey Taylor (*right*)
45–46	NASA
47	Tomo Narashima (*drawings*)
48–49	Jack Harris/Visual Logic
50–53	Jared Schneidman/JSD
55	Jared Schneidman/JSD (*top*), M. C. Escher, courtesy of Cordon Art B.V. (*bottom*)
59	Dana Burns-Pizer
60	Jared Schneidman/JSD (*drawings*), Marcos Intaglietta (*photographs*)
61	Joanne R. Less and Mitchell C. Posner (*top* and *bottom*)
62	Lawrence J. Nugent
63	Jared Schneidman/JSD
64	Dana Burns-Pizer
65	Fan Yuan
67	Roberto Osti
68	From *History of the Indies,* by Fernández de Oviedo (Seville, 1535)
69	Roberto Osti
70	Ian Worpole
71	Jeff Foott (*top*), Roberto Osti (*bottom*)
72	Thomas J. O'Shea (*left*), Ian Worpole (*right*)
74	Courtesy of L'Harmas de Fabre, National Museum of Natural History, France
75	Courtesy of Georges Pasteur
76	Cathy Truc, CNRS, Marseilles (*photographs*), Patricia J. Wynne (*drawings*)
77	R. Coutin (*photograph*), René Antoine F. de Réaumur, courtesy of Michel Emerit (*drawings*)
78	Patricia J. Wynne
79	Jared Schneidman/JSD
80	E. R. Degginger/ Bruce Coleman, Inc.
82–83	Philippe Morel
84	Patricia J. Wynne (*drawing*), Heidi Knecht (*photographs*)
85–86	Heidi Knecht
87	Musée de l'Homme, Paris
88–89	James Aronovsky
90	Jason Goltz
91	Jessica Boyatt
92	Walter J. Freeman, University of California, Berkeley
93	N. Hirokawa, Wiley-Liss, © 1991
94	Jonathan R. Rehg
104–107	Patricia J. Wynne

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Michelle Press, *Managing Editor;* John Rennie, *Associate Editor;* Timothy M. Beardsley; W. Wayt Gibbs; Marguerite Holloway; John Horgan, *Senior Writer;* Kristin Leutwyler; Philip Morrison, *Book Editor;* Madhusree Mukerjee; Corey S. Powell; Ricki L. Rusting; Gary Stix; Paul Wallich; Philip M. Yam

ART: Joan Starwood, *Art Director;* Edward Bell, *Art Director, Graphics Systems;* Jessie Nathans, *Associate Art Director;* Johnny Johnson, *Assistant Art Director, Graphics Systems;* Nisa Geller, *Photography Editor;* Lisa Burnett, *Production Editor*

COPY: Maria-Christina Keller, *Copy Chief;* Nancy L. Freireich; Molly K. Frances; Daniel C. Schlenoff

PRODUCTION: Richard Sasso, *Vice President, Production;* William Sherman, *Production Manager; Managers:* Carol Albert, *Print Production;* Janet Cermak, *Makeup & Quality Control;* Tanya DeSilva, *Prepress;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Eric Marquard, *Special Projects;* Ad Traffic: Carl Cherebin; Kelly Ann Mercado

CIRCULATION: Lorraine Leib Terlecki, *Associate Publisher/Circulation Director;* Katherine Robold, *Circulation Manager;* Joanne Guralnick, *Circulation Promotion Manager;* Rosa Davis, *Fulfillment Manager*

ADVERTISING: Kate Dobson, *Associate Publisher/Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* Randy James, Elizabeth Ryan. CHICAGO: 333 N. Michigan Ave., Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager.* WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Lisa K. Carden, *Advertising Manager;* Tonia Wendt. 235 Montgomery St., Suite 724, San Francisco, CA 94104; Lianne Bloomer. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Nancy Mongelli, *Assistant Marketing Manager;* Ethel D. Little, *Advertising Coordinator*

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; Vivienne Davidson, Linda Kaufman, Intermedia Ltd., Paris; Karin Ohff, Groupe Expansion, Frankfurt; Barth David Schwartz, *Director, Special Projects,* Amsterdam. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.; SINGAPORE: Hoo Siew Sai, Major Media Singapore Pte. Ltd.

ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie M. Beaumonte, *General Manager*

SCIENTIFIC AMERICAN, INC.

415 Madison Avenue, New York, NY 10017-1111
(212) 754-0550

CHAIRMAN AND CHIEF EXECUTIVE OFFICER: John J. Hanley

CO-CHAIRMAN: Dr. Pierre Gerckens

CORPORATE OFFICERS: *President,* John J. Moeling, Jr.; *Chief Financial Officer,* R. Vincent Barger; *Vice Presidents,* Robert L. Biewen, Jonathan Piel

DIRECTOR, ELECTRONIC PUBLISHING: Martin Paul

CHAIRMAN EMERITUS: Gerard Piel

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

Psychics and Weapons

In "Bang! You're Alive" ["Science and the Citizen," SCIENTIFIC AMERICAN, April], on research into nonlethal weaponry, writer John Horgan addressed my interest in the paranormal. I am a member of the Society for Scientific Exploration and do endorse the rigorous scientific study of various anomalous phenomena. My personal and professional interests in such topics have included involvement in studies by the National Research Council and other governmental scientific bodies.

Those interests, however, have nothing to do with my development of nonlethal technologies and concepts. They do not in any way constitute part of my work at Los Alamos National Laboratory. Belief systems, whether religious, political or otherwise, should not be reported in articles on scientific topics. Similarly, they have no bearing on the validity of nonlethal weapons. The urgent need to provide new options to military and law enforcement agencies should be self-evident.

Your article has done a disservice to our nation. Innuendo and obfuscation don't belong in science.

JOHN B. ALEXANDER
Los Alamos National Laboratory
Los Alamos, N.M.

Horgan replies:
The government pays Alexander to oversee a multimillion-dollar research program. His "interest" in alien abductions and paranormal phenomena, about which most scientists are deeply skeptical, raises questions about his judgment and is therefore a legitimate part of the story.

Privileged Communications

In "Wire Pirates" [SCIENTIFIC AMERICAN, March], Paul Wallich writes: "Within the U.S., patent rights to public-key encryption are jealously guarded by RSA Data Security.... Although software employing public-key algorithms has been widely published, most people outside the government cannot use it without risking an infringement suit."

This is wrong and is a myth perpetuated by those who don't bother to check their facts. RSA Data Security provides necessary patent licenses for public-key technology at reasonable rates and actively promotes the widespread use of public-key technology. Licensees include IBM, AT&T, Motorola, Microsoft and other companies, large and small. Moreover, the technology is available royalty free for noncommercial and educational use. More than three million installed software packages utilize RSA; it is far and away the most widely used public-key cryptographic technique.

Thus, although RSA is patented, it is generally an easy matter to obtain a relevant patent license. Only those who are ignorant of the patent or disregard it run any actual risk.

JIM BIDZOS
President
RSA Data Security, Inc.
Redwood City, Calif.

Wallich replies:
As Bidzos knows, the widely published public-key software to which that passage refers is PGP, a free program available worldwide to tens of millions of computer users. PGP makes unlicensed use of algorithms for which RSA holds U.S. patents. (Viacrypt, a small company that had previously purchased a general license from RSA, distributes a commercial version of PGP.) Although RSA makes some of its software available royalty free for noncommercial use within the U.S., until recently the company blocked efforts to incorporate that software into the free version of PGP. On May 9 the Massachusetts Institute of Technology announced a U.S.-only, noncommercial version of PGP that uses RSA-licensed software.

Congress and Altruism

In the middle of Natalie S. Glance and Bernardo A. Huberman's "The Dynamics of Social Dilemmas" [SCIENTIFIC AMERICAN, March], I started thinking about term limits and the effect they would have on parliamentary compromise. If "cooperation is most likely in small groups with lengthy interactions," then term limits on Congress and other legislatures would make our already fractious politics even more vitriolic.

DAVID OLSON
Princeton, N.J.

I believe the authors' conclusions are seriously flawed, in part because they do not fully take into account the effects of irrational behavior and altruism. Many human decisions are based not on perceived good to the individual but on perceived good to others, even at the expense of the individual decision maker. Most religions actively espouse such behavior, and most individuals incorporate some degree of altruism into their decisions.

Failure to incorporate irrationality, altruism and other relevant cultural biases into these sorts of computer models of human behavior renders those models grossly inaccurate and highly misleading.

STEPHEN C. FOX
New York City

Glance and Huberman reply:
When altruism is pervasive, cooperation is easily achieved. When irrationality reigns, anything can happen. But our results will still hold when the influence of altruism is not dominant in a social group. The need all over the world to enforce taxation is an example of how dilemmas persist in all countries.

Altruism and piety confer benefits on individuals that are not quantifiable and perhaps not even acknowledged at a conscious level but are benefits nonetheless. Religious beliefs allow a person to have an infinite horizon for future interactions, because he or she expects benefits to continue eternally. Within this framework, a religious individual is behaving rationally.

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.

ERRATUM

The special issue of *Scientific American* entitled *Ancient Cities,* published in April, misstated the chronology of the pre-Columbian city of Teotihuacán, in what is now Mexico. The city was founded in the first century B.C. and declined to insignificance after A.D. 750, centuries before the period of Aztec dominance in the 14th through 16th centuries A.D.

50 AND 100 YEARS AGO

JULY 1944

“The most discussed of the new insecticides is dichloro-diphenyl-trichloroethane, shortened to DDT but also called Gesarol. This compound has remarkable power to kill insects, particularly body lice—the ‘cooties’ of World War I. Prevalence of typhus, carried by body lice, in the Mediterranean theater of this war has emphasized its value. DDT’s effectiveness in war may well be overshadowed by its value in peace. Painstaking investigations have shown it to be signally effective against many of the most destructive insects that feed upon crops.”

“In many war plants, workers may be seen tapping objects, one after another, in front of a microphone. Little or no sound can be heard by human ears, yet every now and then a light flashes and the operator tosses a piece aside as defective. This is just one of several new techniques which utilizes supersonic frequencies for inspection purposes in industry. Cracks, differences in hardness, changes in dimensions, and variations in the composition of many materials can be quickly detected by this method.”

“Success in one of the longest and most persistent searches of chemists was realized recently with the announcement of the synthesis of quinine by two American chemists, Robert B. Woodward, of Harvard University, and William E. Doering, now of Columbia University, consultants for the Polaroid Company. One of the early attempts to produce this important alkaloid led Sir William Perkin to produce the first synthetic dye in 1856 and thus laid the foundation of the modern dye industry.”

SCIENTIFIC AMERICAN

JULY 1894

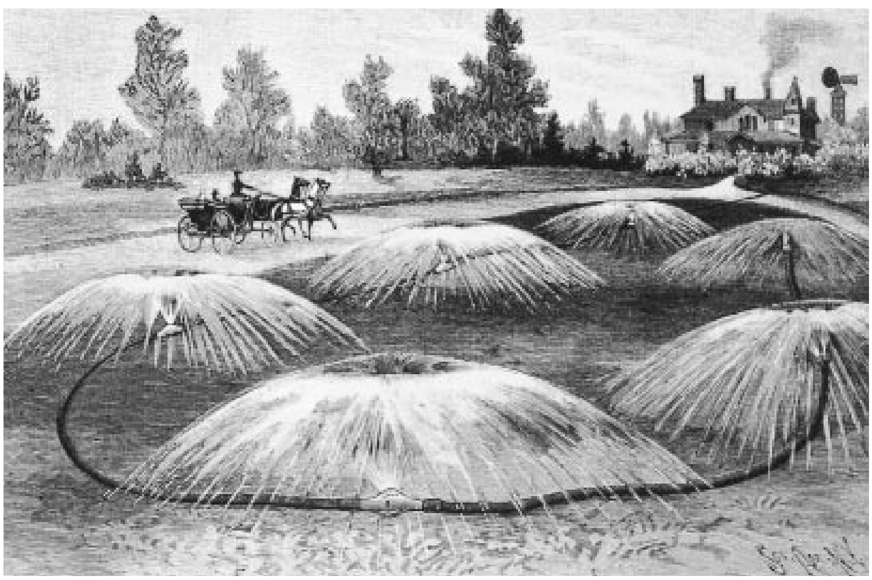
“The papers of the entire country have been full of accounts of a great railroad strike now in progress. It started in consequence of an announcement made by the Pullman Car Company that they could not continue to run their works without a reduction of wages. Pullman cars are run on roads all over the United States, and a boycott aimed at the Pullman Car Company took the form of a refusal on the part of the American Railway Union to permit its members to take a part in running any trains that were made up in whole or in part of Pullman-made cars. In this way, from a cause involving a few hundred workmen, the strike has assumed large proportions and has finally become a contest between the United States government and the American Railway Union.”

“A Frenchman, M. Bersier, has devised a plan by which the compass performs the part of the helmsman. An electric current is placed to work on the desired course, and when the vessel gets off the course for which the electrical instrument is set, the current starts a motor in either direction and moves the rudder until the vessel returns to her proper course.”

“Dr. Troitzki, writing in the Russian medical periodical *Vratch,* states he has found that new and uncut bread contains no micro-organisms. As soon, however, as bread is cut and is allowed to lie about uncovered, harmless and also pathogenic microbes find it an excellent nutrient medium. *Streptococcus pyogenes aureus* retains its vitality on the crumb of wheatmeal bread for twenty-eight to thirty-one days; the typhoid bacillus remains active twenty-five to thirty days; while the bacillus of cholera lives twenty-three to twenty-five days.”

“In his address to the *Chambre Syndicale des Produits Chimiques,* Mr. Berthelon, the illustrious chemist, suggested as a subject for the attention of the next generation of engineers the substitution of the heat of the sun as a source of energy for that derived from coal. The sinking of a shaft three or four kilometers deep is not beyond the power of modern and especially of future engineering. At such a depth, water would be found with a temperature of 160 degrees to 200 degrees Centigrade, which would develop enough power for any number of machines.”

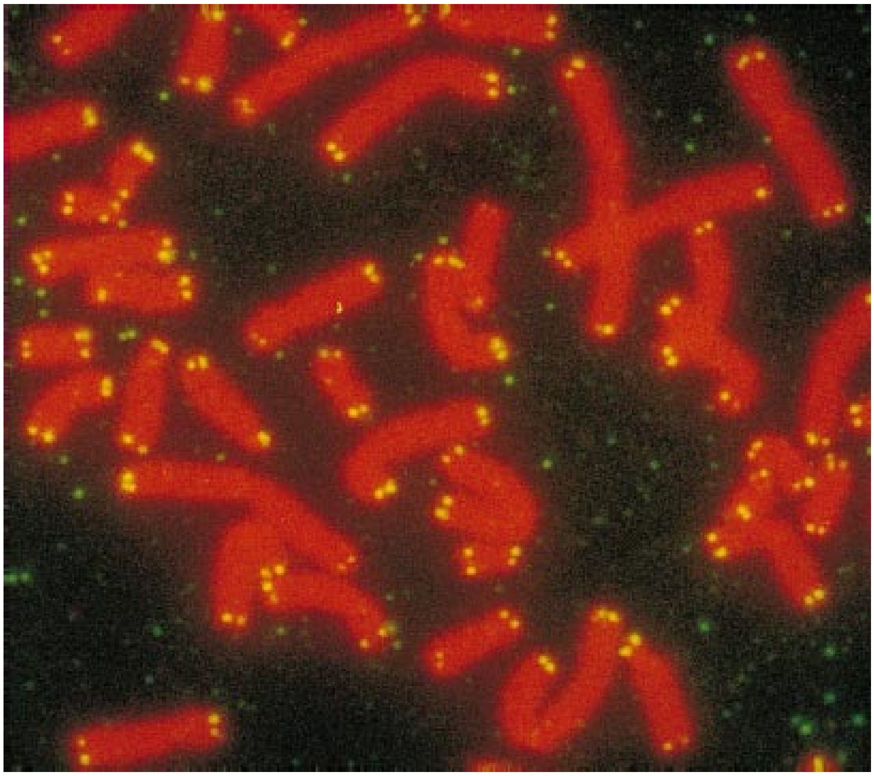
“In order to preserve a lawn in freshness during the parching days of summer the grass must be repeatedly watered. A common method is to have a hollow standard provided at its top with a rotary perforated head. This, when connected with the water supply of a hose, throws a gentle rain over a considerable space. Then the standard is moved into a new position, and so on. The object of the present invention [*see illustration below*] is to effect the instantaneous irrigation of every part of the lawn without the interposition of a special attendant, such irrigation being effected by simply turning the water faucet, which any member of a household may do.”



Fountain pipes for lawn irrigation

Immortal's Enzyme

By rebuilding their eroding DNA, cancer cells stay young


ROBERT K. MOYZIS

TELOMERES (yellow) ***at the ends of chromosomes shrink as human cells age and divide. An enzyme that maintains the telomeres may make tumor cells immortal.***

Cancer cells are like Dorian Gray: internally corrupt and destructive but miraculously blessed with eternal vigor. Researchers in California and Ontario now believe they have identified the secret of that malignant immortality. It appears to be an "immortalizing enzyme" absent from most normal tissues that allows tumor cells to divide ad infinitum. Counteracting that enzyme might be the key to developing completely novel therapies that, unlike those available today, would leave patients largely unscathed.

"I'm optimistic that this represents a unique opportunity for inhibiting cancer cells," reflects Huber Warner, a deputy associate director at the National Institute on Aging.

Immortality is the norm among tumor cells and single-cell organisms—conditions permitting, they reproduce themselves forever. Normal human cells, however, generally have a finite life expectancy. They may divide for a few dozen generations, but they eventually stop and die. Many cell biologists suspect that the erosion of structures called telomeres is to blame.

Telomeres are specialized segments of highly repetitive DNA found at the tips of chromosomes. They seem to stabilize the ends of the chromosomes and prevent them from sticking together or degenerating. (Molecular biologists are fond of comparing telomeres to the protective plastic caps on shoelaces.) Because of a quirk in the replicative machinery, when a strand of DNA is duplicated during mitosis, a few subunits at one end are always lost. With each tick of the mitotic clock, another piece of telomere is whittled away. A person's telomeres thus shrink as he or she ages. Investigators have hypothesized that cells lose their ability to divide when the telomeres fall below some critical length.

Tumor cells and single-cell organisms are immortal apparently because they can stabilize their telomeres. In the mid-1980s researchers showed that protozoans make an enzyme, telomerase, that adds new sequences to the telomeres and preserves their length. Human cells, too, carry the gene for telomerase, but most of them do not express it after birth. The one clear exception in humans is in the testis, which seems to use telomerase to rebuild the telomeres of sperm cells.

Experiments on human cells transformed in culture by tumorigenic viruses suggested that tumors also relied on telomerase. Two years ago Calvin B. Harley, now at Geron Corporation in Menlo Park, Calif., and Carol W. Greider of Cold Spring Harbor Laboratory observed that the transformed cells grew uncontrollably and displayed other hallmarks of tumor cells. Only a few of these abnormal cells, however, exhibited telomerase activity—and those were the only cells that became immortal.

Harley and Silvia Bacchetti of McMaster University and their colleagues have now verified that the cells of at least one kind of cancer, ovarian carcinoma, do express a telomerase. Extracts of abnormal cells taken from cancer patients showed telomerase activity, whereas extracts from normal cells did not. The telomeres of the cancer cells were shorter than those of normal cells, but they were stable—a fact consistent with the idea that the mutation activating the telomerase occurred sometime after the mutations that initiated the tumorigenic changes.

"All the traditional oncogenes and tumor suppressor genes are involved in aspects of growth control, but they don't by themselves make cells immortal," Harley says. "We're proposing that there's a new category of immortalizing oncogenes and that the telomerase gene is its only member." By extension, he adds, the unidentified gene that represses telomerase activity in normal cells would qualify as a new kind of tumor suppressor gene.

Harley's report, published in the *Proceedings of the National Academy of Sciences,* concerns only ovarian cancer. He says his group has looked for telomerase activity in "a large number of other tumors," and he expects to report on those findings soon as well.

If telomerase is the immortalizing enzyme in tumors and yet is missing from most normal cells, it represents a ripe target for a new kind of anticancer treatment. Harley and Greider think drugs inhibiting the activity of telomerase should rob tumor cells of their immortality. Telomerase inhibitors would not kill tumor cells, but they would arrest the cells' proliferation, which might boost the effectiveness of other anticancer agents.

Unlike conventional chemotherapy and radiation treatments, which damage all the dividing cells in a patient's body, inhibitors should have few bad side effects. That supposition, however, still needs to be examined closely, Warner and others note. The ability to produce healthy sperm would probably be impaired, although that risk is already common to conventional therapies.

Another concern is the rare but important stem cells in the intestinal lining, the bone marrow and other tissues that must frequently replace themselves. It is possible that the stem cells, which produce the replacement cells by dividing throughout a person's lifetime, may also need telomerase. The negative assays for telomerase activity in normal tissues may have missed traces in the stem cells. "If 99 percent of the cells don't have telomerase, you might not see the 1 percent that does," Greider remarks.

More precise assays will be possible when the telomerase enzyme is isolated and its gene is cloned—efforts in which both Greider and Harley are now engaged. Fortunately, Harley says, the current biochemical assays are good enough for testing possible telomerase inhibitors. Geron is now screening thousands of compounds for antitelomerase activity. With luck, Harley thinks, some of the drugs might be ready for clinical testing in two or three years.

Quite apart from their relevance to cancer, studies of telomerase might also carry a premium for people worried about old age. "If you take human cells and put telomerase back into them, can you lengthen telomeres again?" Greider asks. If so, cells in the body could conceivably be rejuvenated, which might forestall aging or some of its effects. Some gerontologists dispute that idea, arguing that age is not so easily thwarted. Ask Dorian. *—John Rennie*

A Visit to an Exotic Planet

This false-color radar image of Mount Pinatubo in the Philippines (*below*) was made by the Space Radar Laboratory, which flew on the space shuttle *Endeavor* in April. Reddish-brown areas represent ash spewed from the volcano during its potent June 1991 eruption. Darker areas indicate the location of mudflows from the volcano. Distinguishing between mud and ash is extremely difficult from the ground, notes Diane L. Evans of the Jet Propulsion Laboratory in Pasadena, Calif., the project scientist for the Space Radar Laboratory. This image will help determine which areas around the volcano can be safely resettled and which ones may still undergo potentially lethal mudslides, she says.

In the current era of fiscal austerity, American lawmakers and taxpayers alike are looking for practical results from big-budget scientific research. The Space Radar Laboratory represents the National Aeronautics and Space Administration's literally down-to-earth approach. The $366-million suite of instruments employs radar signals to reveal such key environmental markers as the density of biomass, the amount of moisture in the soil and the quantity of water contained in snow covering.

During the 11-day flight of *Endeavor,* the Space Radar Laboratory scanned about 12 percent of the earth's landmass. Other targets included the volcanic features of the Galápagos Islands (*opposite page, top*) and erosion formations around Death Valley in California (*opposite page, bottom*). The radar images of Death Valley will clarify the effects of ancient climate shifts in that region. NASA plans to launch the laboratory again in August, to give scientists a chance to examine seasonal and human-generated environmental changes.

JET PROPULSION LABORATORY/NASA

Evans hopes NASA will pony up the necessary funds (about $100 million) to transform the laboratory into a free-flying satellite. It would then take its place as a full-time companion to NASA's multibillion-dollar Earth Observing System, the centerpiece of the agency's "Mission to Planet Earth"—an ambitious scheme for using remote-sensing technology to monitor global change.

The Space Radar Laboratory has also generated considerable interest among researchers looking outward to other worlds, because its radar system is similar to the one on board the *Magellan* spacecraft that has been mapping Venus. A number of planetary scientists, including Ellen R. Stofan of the Jet Propulsion Laboratory, are poring over images from the *Endeavor* flight to learn more about the environmental and geologic disparities between the earth and Venus, its wayward sister world. *—Corey S. Powell*

Superhack

Forty quadrillion years early, a 129-digit code is broken

In August 1977 three professors from the Massachusetts Institute of Technology dared *Scientific American* readers to decode a cipher they printed in Martin Gardner's "Mathematical Games" column. The numerical teaser was one of the first published examples of their newly invented encryption system, called RSA. The trio—Ronald L. Rivest, Adi Shamir and Leonard M. Adleman—offered a $100 reward for the return of a plain-text sentence, an event they predicted might not occur for some 40 quadrillion years. This past April, Bell Communications Research scientist Arjen K. Lenstra, three computer hobbyists and over 600 volunteers from the Internet claimed the check early, after only eight months of work.

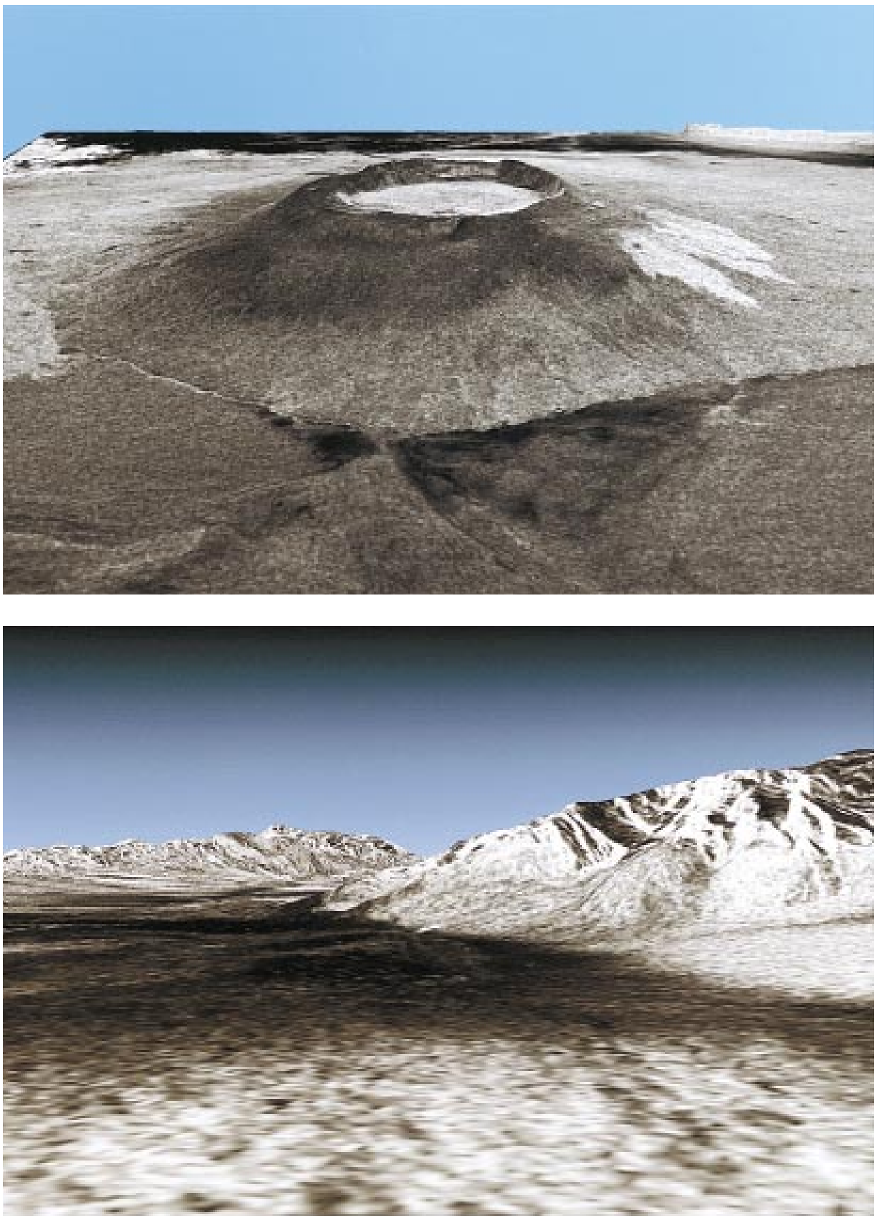
"It was inconceivable 17 years ago that this code could ever be broken," Lenstra says. Indeed, RSA is considered one of the most secure commercial encryption measures available. To encode a message using RSA, the text is converted into a number, which is then raised to a certain exponent; from that, a fixed, large number, or modulus, is subtracted repeatedly until the result is smaller than the modulus itself. The user can publish the exponent and large modulus (called the public keys) so that anyone can create a secret message with them. To determine the inverse function and recover that message, however, requires knowledge of those two prime numbers (those divisible only by 1 and themselves) that when multiplied yield the public key. Although it is trivial to generate products of large, prime numbers, it is inordinately difficult to factor these products.

In their *Scientific American* challenge, Rivest, Shamir and Adleman used what at the time was an indomitable 129-digit number as their public modulus, later nicknamed RSA-129. That Lenstra and his colleagues have cracked this key attests to the remarkable strides made over the past two decades both in mathematics and in the ability to marshal computing power from machines distributed around the globe. To break RSA-129 required some 100 quadrillion computer instructions, perhaps one of the largest and most difficult single computations ever performed.

"In order to harness enough computer power, we needed lots of computers and lots of people," says Paul Leyland, a computer systems manager at the University of Oxford, who helped to initiate the project. The mathematical attack the team used is called a multiple polynomial quadratic sieve, a tool that enabled them to split the job into many smaller tasks. Eventually the algorithm sifts out likely factors for large numbers from the millions of candidates that it generates.

Leyland, in conjunction with Derek Atkins, a graduate student at M.I.T., and Michael Graff, an undergraduate student at Iowa State University, coordinated the efforts of Internet participants on five continents, who donated time from some 1,600 computers to create 8.2 million pieces of data. Atkins verified and stored the contributions in a database at M.I.T. and then sent the entire collection to Lenstra. In two days a massively parallel supercomputer at Bellcore churned out a 64-digit factor and a 65-digit cofactor for RSA-129. What did the 1977 cipher say? THE MAGIC WORDS ARE SQUEAMISH OSSIFRAGE. Rivest explains that they chose the words at random. "I don't know that we ever expected to see them pop up again," he adds.

Fortunately, those who use RSA software (more than three million copies have been sold) need not be squeamish about the protection their system offers. As yet, no truly efficient algorithm for reckoning prime factors from massive composites has been found, although one may someday exist. In the meantime, Lenstra likens using number sieve methods to searching for millions of



JET PROPULSION LABORATORY/NASA

needles in a haystack. RSA users can still elude nosy hackers by choosing keys having more digits than those Lenstra and his colleagues have put down.

"The significance of this accomplishment is that it helps us benchmark the system and helps the user know how large the numbers need to be," Rivest says. For this purpose, RSA Data Security in Redwood City, Calif., which produces RSA encryption software packages, sponsors a series of factoring contests, ranging up to RSA-500. Lenstra notes that many organizations already hold in house the computing power needed to factor numbers 129 digits long, and so RSA-150, next on the hit list, cannot be far from falling. "The lesson to be learned is that a system believed to be secure now may not be tomorrow," Lenstra says. At least not while he's around. *—Kristin Leutwyler*

Managed Care, Circa 1300

Bernat de Berriac, M.D., could have taught Hillary Clinton a lesson or two. In the early 1300s de Berriac received five to 20 sous a year from each of several dozen men of modest means from Castelló d'Empúries, a village in Catalonia. For this paltry sum, the youthful doctor agreed to treat these men and their wives, children and servants "for every illness that requires the art of medicine."

Dental coverage was not included (that was mostly the realm of the barber). But it probably should have been, since pulling teeth was one job the 14th-century health provider could handle. The advent of the medieval precursor to the managed care plan was documented in a history that won a prize for the best recent book from the American Association for the History of Medicine for Michael R. McVaugh, a historian at the University of North Carolina at Chapel Hill.

In *Medicine Before the Plague,* published by Cambridge University Press, McVaugh shows how the discipline of medicine began to emerge as a formal profession. A major preoccupation of this era, not to mention the late 20th century, was who was covered and how a physician would get paid.

Quality of care, too, was an issue. Many of these freshly minted products of the academy had not yet achieved the status that made them desirable marital quarry. Indeed, they had to distinguish themselves from society's unlettered masses or even less educated practitioners. "In the first decades of the century," McVaugh comments, "every physician would at the outset have had to convince his patients that he knew something they did not—what was wrong with them, and how it could be cured—and that they should concede him authority and power over them in treatment."

To lure patients, early practitioners entered into a contractual arrangement—in effect, a form of managed care. In the 14th century it was easy for special-interest groups to prevail. Royalty and the church got the best pickings. Count-kings paid a lifetime retainer, a *violarium,* to three or four physicians and surgeons, in addition to barbers and apothecaries. In exchange, these newly emerging medical professionals were on call at any time, at the dispatch of a real—not an electronic—page.

Even in the 1300s, society worried about coverage for its less fortunate. Combing through old histories, contracts and wills, McVaugh found that physicians agreed to become service providers in what were inexpensive, prepaid health insurance plans. Many towns set up a post–Dark Ages version of a public health service by simply putting a doctor on the payroll.

It was still more than half a millennium until the arrival of sulfa drugs and the AMA. But surgeons of the time knew how to cut. And preventive medicine, then as now, enjoyed a vogue. Witness the promises by the physician Abraham des Castlar when he agreed to serve Castelló d'Empúries in the year 1316: "I will look at and assess all the urines brought to me by the citizens, whom I will advise as to bloodletting and diet, and generally as to their manner of life, and I will visit two or three times all the sick of the town who ask me to attend them." *—Gary Stix*

LOREN C. MACKINNEY COLLECTION *University of North Carolina at Chapel Hill*

GOTHIC LETTER from a 14th-century manuscript in the Biblioteca Apostolica Vaticana has at its center an illustration of a surgeon ministering to a patient who clings to the surgeon's assistant.

Lethal Legacy

Soviet reactor sites menace Eurasia

The cold war has ended without an exchange of nuclear attacks. For that, everyone can be (perhaps guardedly) thankful. Yet the period of tentative peace has been marred by a persistent radioactive legacy.

The explosion and fire that destroyed reactor number 4 at the Chernobyl power plant on April 26, 1986, is generally—and correctly—described as the world's worst nuclear accident. The amount of radiation that the burning reactor released into the atmosphere will never be known exactly: 50 million curies is a widely quoted figure, although one recent study concludes the amount was perhaps even five times greater.

As deadly as it was, the release at Chernobyl was puny compared with the colossal exudation of much longer lived radionuclides from reactors that the former Soviet Union used, and Russia still uses, to produce plutonium for bombs. Russian officials who are collaborating with the U.S. Department of Energy (DOE) to devise cleanup techniques have disclosed data that make even radiation-hardened nuclear engineers blanch.

A recent assessment by Clyde W. Frank of the DOE, together with Don J. Bradley of Battelle's Pacific Northwest Laboratories in Richland, Wash., confirms that the world's biggest environmental release occurred over decades at a site called Tomsk-7 in central Siberia. Tomsk-7, whose existence was clas-

IVLEVA *Magnum*

***SARCOPHAGUS AT CHERNOBYL** that houses the remains of reactor number 4 is littered with wreckage. The structure may collapse within a few years. A rising water table is adding to fears that more radioactive material may escape.*

sified until about 1990, is thought to have poured and pumped about a billion curies of high-level waste, or 20 Chernobyls' worth, into lakes in the region and into underground formations. By way of comparison, the largest releases in the U.S.—at the DOE's Hanford site in Washington State, at the Savannah River complex in South Carolina and at Oak Ridge National Laboratory in Tennessee—are believed to range from 700,000 to a million curies. The DOE estimates that the total amount of released radioactivity in Russia is about 400 times the amount in the U.S.

Frank points out that much of the radioactive material released at Tomsk has been "stored" in fractured rocks that are capped by clays and so are partly isolated, at least for the present, from the ecosphere. That form of disposal was considered safe in the U.S. in the 1940s and 1950s. But in terms of curies, Tomsk beats Oak Ridge by a factor of 1,000. Little is known about how well isolated the Tomsk burial ground really is, notes Thomas B. Cochran of the Natural Resources Defense Council. The lakes at Tomsk that have been used as dumps drain, via the Ob River, into the Arctic Ocean.

Production of plutonium at Tomsk continues, though at a low level, Frank says. The same is true of another Russian production site, Krasnoyarsk-26. Frank and Bradley report that the Krasnoyarsk facility, which is built underground, has released some "hundreds of millions" of curies from military production reactors into an underground reservoir. An article published in *Izvestia* in January describes the Krasnoyarsk site as located 100 meters above and only 750 meters from the Yenisei River, which also flows into the Arctic Ocean. Russian experts cited by the newspaper were reportedly "extremely skeptical" that the waste would remain isolated for long. Indeed, according to *Izvestia,* plutonium and radioactive isotopes of cobalt, cesium and strontium have already been detected along the Yenisei.

Then there is Chelyabinsk. The complex, now called Chelyabinsk-65, was the U.S.S.R.'s first plutonium production factory and the site of an infamous accident in 1957 in which a high-level waste tank exploded, releasing about two million curies of fission products into the atmosphere. Contamination spread over hundreds of square kilometers. Yet in terms of curies, the accident pales in significance alongside the routine dumping of wastes into surface waters in the region.

Lake Karachai, near Chelyabinsk, has been used for decades as a repository for medium-level waste from production reactors there. The lake is thought to be the most radioactive body of surface water on the earth. In 1967 a drought dried out part of the lake, and dust formed from the exposed sediment made a radioactive wind that spread dangerous levels of radiation up to 70 kilometers away. Anyone at lakeside would have acquired a lethal dose of radiation in one hour. Special machinery is now being used to fill Karachai and cap it.

Russian officials acknowledge that at least 130 million curies have been released at Chelyabinsk. That figure could be an underestimate, asserts Murray Feshbach, a professor of demography at Georgetown University. Feshbach cites official and unofficial sources that refer to more than a billion curies released at Chelyabinsk. In the early days of the site—between 1951 and 1953—waste was dumped straight into the Techa River. Studies have documented markedly elevated rates of leukemia in local inhabitants.

Collaboration with Russia on cleanup technologies is proving fruitful, Frank says. Russian workers have for some years been investigating cobalt dicarbollide, a chemical that has a strong affinity for cesium and strontium, which makes it useful for decontaminating liquids. Russian investigators recently participated in the first test of the material on actual waste at the Idaho National Engineering Laboratory. Another technology, one developed in the U.S., uses resin-coated magnetic beads to achieve a similar result.

However promising the techniques, the expense of cleaning up the cold war's legacy is likely to be prohibitive. The DOE has estimated that costs at military nuclear sites in the U.S. alone are likely to top $300 billion.

There is also the legacy of peace. The sarcophagus that frailly houses the remains of Chernobyl-4 is crumbling faster than expected and could collapse, according to an assessment carried out in March by the International Atomic Energy Agency (IAEA). If a collapse occurred, an estimated 10 tons of highly radioactive dust could be shot into the air. Moreover, *Pour La Science,* the French edition of *Scientific American,* reported in March that the structure could release radioactivity into the area's rising water table, despite the presence of a barrage.

The IAEA inspectors also found serious safety deficiencies at the two reactors still operating at Chernobyl, which are of the same design as the ill-fated number 4 reactor.

Ukraine maintains that it cannot afford to shut down the remaining Chernobyl reactors. Others might wonder if the world can afford to let Ukraine keep them running. *—Tim Beardsley*

MARK PETERSON SABA

PROTESTING VIOLENCE: ***candlelight vigil is held outside City Hall in Los Angeles in March 1993, during the second trial of police officers accused of beating motorist Rodney King. Fifty-eight people died in riots that followed acquittals at the first trial.***

Desperate Measure

Does violence need its own institute?

Violence in America batters the senses. Even those who have had the good fortune to avoid a personal encounter are subject to a constant vicarious assault through the media. In Congress a senator or representative's willingness to get tough on crime is a measure of his or her political viability. It is hard to argue against the sense of outrage: the per capita homicide rate in the U.S. in 1985 was four times higher than that in most European countries, according to World Health Organization data. Average prison time in the U.S. for each violent crime tripled between 1975 and 1989, yet reported levels of violence changed little over that period.

Given the magnitude of the problem and the failure of the courts to deter violent crime, one might guess that research on the subject should be a high priority. Or should it? Such inquiries arouse suspicions that the effort would serve to reinforce racial stereotypes and support government attempts to control behavior.

At present, such fears seem far-fetched. The amount the federal government spends on research into violent crime, compared with other causes of death, is minuscule. According to the study *Understanding and Preventing Violence,* published last year by the National Research Council, in 1989 only \$31 was spent on violence-related studies per year of potential life lost to violence in 1989. The equivalent calculations for cancer and AIDS yielded figures of \$794 and \$697, respectively.

The conviction that more research might help is what lies behind a congressionally inspired move to bring into the picture a player normally associated with semiconductors and supercomputers. Representative Louis Stokes of Ohio and Senator Barbara A. Mikulski of Maryland earlier this year asked the National Science Foundation (NSF) to look into the feasibility of establishing an interdisciplinary science and technology center focusing on research into violent behavior. In late May the foundation was planning to accept the challenge, although agency officials said the work would probably be distributed in a consortium rather than carried out at a specific center.

Jeffrey A. Roth, the principal staff officer for the National Research Council's report, states he expects the NSF's program to concentrate on nongenetic behavioral research, including work on neurochemical factors that may predispose people to violent behavior. "I think a better understanding of the neurobiology can open an avenue for benign interventions," he comments. But the line between neurochemical factors and genes is arguably nonexistent. In 1992 the National Institutes of Health canceled a meeting it had proposed to sponsor at the University of Maryland called "Genetic Factors in Crime" after the idea came under attack from African-American groups and others as racist. A replacement meeting is planned for 1995.

One of the critics was Peter R. Breggin, who heads the independent Center for the Study of Psychiatry in Bethesda, Md. Breggin observes that "bringing the National Science Foundation into fundamentally social and political issues is fraught with difficulties." He argues that rather than looking for individual differences that may predict antisocial behavior, "the real issue is whether America wants to face up to what a racist society it is."

Workers in the field of violence are, on the other hand, predictably cheered by the possibility of a new government program. The field has suffered from being fragmented, observes Colin Loftin of the University of Maryland. He thinks the National Science Foundation could make a difference. Such research could, he suggests, begin to address the problem of how to measure violence and how to assess prevention programs. "So little is now known that almost anything we learn would have an impact," he says. *—Tim Beardsley*

"You're the Top..."

Fermilab finds the top quark—sort of

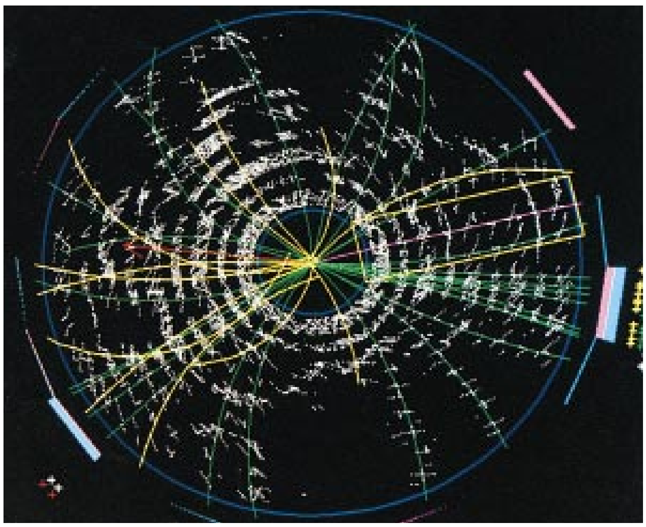
Rumors had been circulating since last summer, so it seemed to be just a matter of time before an official announcement would be made. When the call came, the expectant media circus descended on the Fermi National Accelerator Laboratory (Fermilab) in Batavia, Ill. Yet the Fermilab speakers hesitated to deliver the media goods. "Now, we're not claiming a discovery," cautioned William C. Carithers, Jr., one of the spokespersons for the hundreds of physicists who garnered the results. "What we see is the first direct hint that the top quark is there." Fermilab director John Peoples, Jr., reinforced the hedge: "I assure you, we are going to have far more evidence for it soon."

Certainly the anticipated prize was worth the attention. For two decades, the top quark has been one of the Holy Grails of high-energy physics. Out of the six kinds of quarks thought to make up all matter, it was the only one that had not been observed. The theory that characterizes particles, called the Standard Model, indicates that quarks are organized into three pairs. The first pair includes the up and down quarks, which in different combinations produce protons and neutrons. (The proton contains two up quarks and a down; the neutron grips one up and two down.) The other two pairs consist of the charm and strange quarks, and the top and bottom quarks. These latter pairs make up more exotic, short-lived particles seen in high-energy physics laboratories. Along with quarks, the family of particles known as leptons (neutrinos, electrons, muons and tau particles) composes the elementary constituents of the universe.

The quest for the top quark began after its partner, the bottom quark, was found at Fermilab in 1977. The top quark remained elusive mostly because of its heft (the heavier a particle is, the more energy is needed to create it in accelerators). Early estimates placed that value at a few tens of billions of electron volts (GeV). But when accelerators failed to turn up the top quark, theorists realized the particle must be heavier than they thought.

Scientists had the best shot at finding the top quark once they completed the Tevatron at Fermilab in 1983. The world's most powerful accelerator, it smashes protons and antiprotons together at 1.8 trillion electron volts. At this energy level, physicists believed a top quark should be made once for every few billion collisions. The search demanded the efforts of 440 investigators from 36 institutions, prompting praise for the merits of international cooperation and jokes about the number of physicists needed to install a lightbulb. By 1989 the Tevatron had set a lower limit on the top quark's mass at 91 GeV—a whopping number, considering that the next most massive quark, the bottom, weighed in at only 5 GeV.

FERMI NATIONAL ACCELERATOR LABORATORY



PARTICLE DECAY TRACKS signal top-quark production.

To get the results they announced at the press conference, the CDF team members (as they are known, for Collider Detector at Fermilab) collected data from August 1992 to June 1993. According to the Standard Model, a top quark and its antimatter twin could appear in proton-antiproton collisions. The top and antitop quarks would decay into bottom and antibottom quarks and a pair of so-called *W* bosons. The CDF workers looked for decay products, such as electrons, muons, neutrinos and mesons, of these particles.

The almost year-long experiment produced more than one trillion collisions. After sifting the data, the workers discovered 12 events that harbored signatures of the top quark. The CDF estimates the top quark to weigh 174 GeV, give or take 10 percent. That makes it the heaviest observed elementary particle yet discovered, but within the parameters set by the Standard Model.

The CDF members stop short of crying eureka because of the statistical nature of the find. Besides the rarity of top-quark events, other particles appear in the wreckage that mimic the decay signatures, thus creating a noisy background. For statistical inquiries such as this one, Carithers says, one might want four or five "standard deviations" from the background in order to proclaim a discovery. After several months of tedious number crunching, the CDF data displayed almost three deviations, corresponding to a one-in-400 possibility that the results are false. The odds for being wrong seem low. Still, "scientists tend to be a conservative lot," points out Melvyn J. Shochet, the other CDF spokesperson. "One would like to have the probability even smaller than that."

Anxiety also stems from another top-finding device at Fermilab—the D0 ("dee zero") detector, named after its location along the accelerator path. Paul D. Grannis, a D0 spokesperson, says the group's data neither support nor refute top-quark production as measured by CDF. "That's something that worries us," says CDF member Jose Benlloch.

Despite the uncertainties, most researchers at Fermilab do not consider the announcement premature. "The good thing is that for the first time, we see a positive signal," Benlloch says. The hope, too, was to squelch the persistent rumors about a discovery. "Since these reports were going to come out, it behooves us all to have the opportunity to explain what we have done," Shochet says. Fermilab has already initiated another round of collisions, and by the end of the year the CDF team should be able to verify the results.

Assuming Fermilab has indeed found the top quark, what is left for high-energy physics? Plenty, workers say. "I don't think you should view this as the completion of the Standard Model," opines R. Keith Ellis, a Fermilab theorist. "The Higgs is still missing, and who knows what else." The Higgs boson is the hypothesized mechanism that would explain why particles have the masses they do. It is thought to lie well beyond the reach of the Tevatron (the Superconducting Super Collider was meant to find the Higgs mechanism).

Because it is so heavy, the top quark must be the particle most strongly coupled to the Higgs. Fermilab theorist Christopher T. Hill thinks the top and antitop quarks may represent one component of it. "Maybe the top is a staging area to get to new particles through its decay modes," Hill speculates.

Also intriguing investigators are two minor anomalies in the data. First, one channel seemed to contain all signal and no background. Second, two top events appeared in a "control sample" that should not have produced any at all. Hill points out that if such anomalies persist (which is unlikely), it would constitute evidence for physics beyond the Standard Model.

Fermilab hopes to become a "top-quark factory" by 1998, when the main injector is slated for completion. This upgrade promises to boost the density of colliding protons in the Tevatron and thus yield many more data. With a plethora of top events, new decay modes, if they exist, should become apparent. "We'll be using top as a laboratory to understand how its elementary particles interact," Shochet states. That would also include studies of the *W* particle and bottom quark, which would also be produced in profusion. Indeed, at least one worker thinks Fermilab will be as significant as—and in some ways superior to—the "*B* factory" planned for Stanford University to explore why matter dominates antimatter in the universe.

Given a sense of progress, high-energy physicists are breathing a sigh of relief. "There was tremendous theoretical expectation that the top quark is there," says Steven Weinberg of the University of Texas at Austin. "A lot of us would have been embarrassed if it were not." Thanks to Fermilab, physics may not have to suffer the indignity of being topless anymore. —*Philip Yam*

Spell for Old Bones

At 31, she made it into the *Guinness Book of Records* as the youngest ever president of a major college. Thirteen years later, having tripled the endowment of Barnard College, made it fully residential for the first time—no mean achievement in Manhattan—and given birth to two daughters, Ellen V. Futter is the newest president of the American Museum of Natural History. On its 125th anniversary, the largest natural history museum in the world is ready to march out of mustiness.


JASON GOLTZ

SUPERACHIEVER FUTTER: "Winning is more fun."

"We felt Ellen could brighten the place up and get rid of the dust and moss," says William T. Golden, who chairs the museum's board of trustees. "It's already a much happier ship. I'm very pleased that one can get things done." Futter's strengths, he says, are dedication, decisiveness, humanity and inexhaustible energy. All these the museum gets for a salary "not dissimilar" to that at Barnard (about $250,000 before benefits) and an apartment on the East Side.

In person, Futter is smaller than one would expect, but not fragile. "I love competition," she says, talking about school sports, at which she excelled. One imagines those in the outfield taking a few steps back when Futter got up to bat. Sports taught her teamwork, she explains, her parents taught her empathy, and law school reinforced her love of justice. Even her interest in natural history has early roots. "I've always collected rocks, shells and butterflies," Futter says. "My daughters are amazed that I can still just reach out and capture a butterfly."

Futter has been spending what time she can exploring the museum—two thirds of which is off-limits to visitors. Few of those millions realize that behind the striped pots and stuffed deer lie some front-rank laboratories. When asked what her favorite scientific projects are, Futter demurs: "A mother can't have favorites. I love all of my children equally." Attached to the maternal metaphor, Futter noted at her welcoming reception that the museum's dramatic fossil mount of a 50-foot barosaurus, rearing up to defend its baby from an attacking allosaurus, reminded her of her protective role.

"This museum is part school, part university and part public forum," Futter says. Both Futter and Michael J. Novacek, the museum's dean of science, talk of how two dominant concerns of our times—the public's poor knowledge of the natural world and the vanishing diversity of life—present the museum with a particular opportunity to enhance its profile.

They may be right. An account of the museum's history, written in 1968 by John M. Kennedy of Yale University, reveals that the institution prospered when it was successfully identified with the values and concerns of the age.

The American Museum of Natural History came into being in 1869, when some wealthy New Yorkers, among them Theodore Roosevelt, Sr., and J. Pierpont Morgan, resolved to bring the uplifting lessons of nature to the city's working classes. An appeal to the public brought in hundreds of rocks, shells and pressed flowers; next the trustees purchased the bird and fossil collections of several naturalists. One trustee also acquired a life-size exhibit of a lion attacking an Arab on a camel. "It will add greatly to the popular interest of the museum," he wrote, "and aid us in getting subscriptions." (The sentiment was echoed by the museum's curators in 1991, when the newly unveiled barosaurus trio drew comments about authenticity.)

The acquisitions grew rapidly. Believing that collections made for the purposes of research were best left "to the Europeans," the trustees tended to buy objects that displayed well, such as large bones—inadvertently setting the stage for the museum's future eminence in paleontology. In 1877 the birds and bones moved into a grand new building donated by the city on the west side of Central Park.

Attendance promptly fell: there was no easy way for the city's populace to get that far north. Besides, most of the trustees being strict Presbyterians, the museum was closed on the Sabbath—the one day of the week when the working classes were not working.

The trustees asked Morris K. Jesup to study the problem. Becoming president in 1881, Jesup lavished on the museum the same careful attention that had benefited his Western railroads. First he had the lone curator, Alfred Bickmore, rearrange the exhibits. His labels, Jesup told Bickmore, were "too long and scientific." Bickmore also began to give lectures in natural history to local teachers. In 1884 Jesup sailed to Europe and studied the research collections of two major museums. Deciding that the American Museum as well should engage in research, he hired Joel A. Allen as the curator of a brand-new department of mammalogy and ornithology.

Like their peers in the 20th, scientists of the late 19th century were acutely aware of the speed with which species were vanishing from the earth. Allen, in particular, wished to preserve a record of American birds and mammals for posterity. The process of collecting mammals appealed to the trustees, and in 1888 the museum's first expedition—a group of sportsmen with guns—went out in search of bison. Also in that year the museum found itself forced by the city, now dominated by Irish Catholic voters, to open on Sundays.

Jesup decided as well to develop a paleontology department. Unable to entice the famed Othniel C. Marsh of Yale, he hired Henry Fairfield Osborn. Osborn came to the museum in 1891 and that summer sent an expedition to the best of Marsh's fossil quarries. Thereby he acquired some fine specimens—along with the privilege of naming them. Then, having conspired to have Marsh's funding cut off, he settled down to amass as complete as possible a record of past North American vertebrates.

Jesup also hired Franz Boas, a reputed anthropologist. Artifacts that Boas collected now represent—as he had envisaged—our only record of the cultures of several North American tribes that went the way of the bison. One trek to the northwestern coast, designed to resolve whether Indians had come from Asia, brought back a war canoe that now serves as a rendezvous point at the museum. A monograph on the Yukaghir tribe of Siberia, based on another Jesup foray, was recently acquired by the tribe's descendants. Translating it into their language and teaching it in schools, modern-day Yukaghirs hope to regain some sense of their culture, destroyed by Soviet occupation.

Like any other rambling, old Victorian house, the natural history museum has its dark secrets. Robert Peary, the Arctic explorer, brought back six Inuit Indians, four of whom died soon after arriving in Manhattan. (Only last year were their skeletons shipped back to Greenland.) One survivor was able to return home; the other, a little boy, Minik, was brought up by a museum employee and died, a bitter young man, at age 27. Another individual displayed at the museum and at the Bronx Zoo, the pygmy Ota Benga, took his own life.

And, like a Victorian family, the museum survived its tragedies. By the time Jesup died in 1908, he had greatly enhanced the institution's scientific status. The trustees had given generously, to projects that caught their attention—and that brought in the public. The study of nature, these self-made men hoped, would acquaint the city's immigrants with the "real America."

Osborn, who followed Jesup as president, turned the museum into a national institution. Expeditions to the Gobi Desert led by Roy Chapman Andrews turned up fossilized dinosaur eggs and the earliest known skulls of mammals, reinforcing Osborn's scientific eminence. Enjoying the sponsorship of the trustees—J. P. Morgan was his maternal uncle—Osborn managed as well to catch the imagination of middle-class Americans.

He successfully identified the museum with such dominant values of the age as boldness, adventurousness and hard work. The expeditions, skillfully dramatized by the museum's press officer, made heroes of men such as Andrews. (The "Indiana Jones" movies may have been inspired by him.) In the 1920s hundreds of boys wrote each year to the museum, asking how they could get a job there. Often they offered to work without pay and to "do anything, just sweeping the floors."

Until the stock market crash of 1929, the museum grew headily. New exhibits—such as the African mammal hall, the work of the great hunter, naturalist and taxidermist Carl Akeley—opened frequently. The administrative offices also expanded, in response to the need for perpetual fund-raising. Research did not always do so well. Osborn thought of naturalists with doctorates as "too theoretical" and anthropology as "gossip of the natives"; Boas eventually left the museum for Columbia University. In later years the museum recaptured his legacy through the presence of Margaret Mead, his student.

JASON GOLTZ

SUPERMOTHER BAROSAURUS protecting her young.

During the Great Depression, the museum lost many of its original trustees. Tired of worrying about money, Osborn retired in 1933. From then on, the museum was run not so much by its president as by its director. Over the years the directors consolidated the administrative and scientific departments. New sources of funding, such as the National Science Foundation, opened up after World War II. Curators began to compete with academics at universities for research grants.

In 1988 the trustees decided that, once more, the president should be the chief executive officer. During the tenure of George D. Langdon, Jr., the first paid president, the museum embarked on a program of modernization—which Futter has inherited.

Today the museum's labyrinthine storage areas contain more than 30 million specimens and artifacts. Among these, reportedly, are busts once used in the service of eugenics, now lining a ghostly attic—and a colony of beetles almost a century old that still serves to clean the flesh off delicate bones.

A staff of 200 scientists, among them 42 curators, conduct research in laboratories tucked behind exhibition halls. They also go on numerous expeditions (although the original rule of "finders, keepers" is much altered these days). Thousands of screaming schoolchildren gallop through the halls every day, falling over in awe at the sight of the barosaurus. ("Visiting Mommy's office is now a lot more fun," says Futter of her daughters, aged eight and 12.) On the fourth floor, windows that have been boarded up for decades are being opened, letting light into new and future exhibits of fossil vertebrates.

Curators continue to be deeply involved in planning the museum's exhibits. The Hall of Human Biology, which opened in 1992, took six years of intense work by the anthropologist Ian Tattersall. Museum scientists are reorganizing the fossil halls according to the cladistic paradigm of evolutionary theory, for which the museum "is Mecca." Says Ward C. Wheeler, who helped to start up the new molecular systematics laboratory: "The expertise on evolution at this museum is incredible. You need to know something, it's here. Or you need something, it's here, too."

To researchers, the museum's collection presents an outstanding resource: the world's most extensive record of life-forms, past and present. Valuable as well is the curators' experience in identifying species. Over recent decades, as biology has become ever more specialized, fewer scientists can recognize biodiversity when they see it.

Such experience is useful in guiding conservation efforts. Curator Melanie Stiassny, for example, points out that lemurs and some cichlid fish, found only in pristine pockets in Madagascar, are of ancient and unique lineage. They therefore deserve our utmost attention.

"We are sitting on two of the most pressing issues of our time," Futter notes. Both of these issues—the state of the natural environment and science education—the American Museum is determined to make its own. If history is any guide, the strategy bodes well. "This museum," says Futter with characteristic conviction, "is poised for enormous success." *—Madhusree Mukerjee*

SCIENTIFIC AMERICAN

Agriculture for Developing Nations

The capital-intensive, highly mechanized Western model may not suit every developing region. Systems of intensive polyculture, exemplified by rice cultivation, may be better

by Francesca Bray

People in the rich industrial countries have fixed ideas about the development of agriculture. Children at school learn about the technical progress from digging stick to hoe and from the cattle-drawn wooden ard to the tractor-driven, steel-shared plow. Economists and sociologists describe the shift from small family farms to large, efficient commercial enterprises. Human labor and skills yield to increasingly complicated machines. Although at times we feel pangs of nostalgia for the old ways, we know that the Western model traces the inevitable path of human progress.

Or does it? Mounting frustration over attempts to plan agricultural development around the world has made it clear that the way farming developed in Europe and North America may not, after all, be the best model in the poor countries of Africa, Asia and Latin America—nor, indeed, for the survival of the biosphere. The Earth Summit held in 1992 in Rio de Janeiro marked the official endorsement of a new, critical approach to the world's problems with resources. Its key words are not "growth" and "development" but "conservation" and "sustainability." The basic philosophy of classical agricultural and economic development—more is better, for everybody—is now seriously in question.

Yet the world still faces urgent problems of poverty, hunger and disease. Rural populations are especially deprived and vulnerable. The great question is whether agricultural policies based on conservation and sustainability can solve these acute problems. Or is conventional growth-driven development, for all its drawbacks, the only way to improve rural living standards? I shall argue here that the Western model may not be the ideal for every developing region.

FRANCESCA BRAY is professor of anthropology at the University of California, Santa Barbara. Taking her bachelor's degree in Chinese studies and her Ph.D. in social anthropology at the University of Cambridge, she worked at the Needham Research Institute in Cambridge from 1973 to 1981. Her field there was the history of Chinese agriculture, and she wrote *Agriculture,* a volume in the series *Science and Civilisation in China,* edited by Joseph Needham. From 1981 to 1983 she held a Leverhulme Research Fellowship to study the rice economies of Asia, and for the next four years she worked at the CNRS in Paris. She came to the U.S. in 1987, serving as professor of anthropology at the University of California, Los Angeles, until she transferred to Santa Barbara last year. Her books include *The Rice Economies: Technology and Development in Asian Societies,* published in 1986 and reissued this year, and the forthcoming *Fabrics of Power,* a study of the technologies that defined women's lives in imperial China. Her next research project will be on innovation in Chinese medicine.

As critics of classical development policies have pointed out, the world now produces more than enough food for everyone, but development has often worsened the inequities of distribution. In fact, this trend is hardly surprising if one examines the criteria that define development in agriculture. The "modernization" of agriculture, as generally understood, entails the application of science, technology and capital to increase the output of just a few crops that have world markets—among them wheat and rice for human consumption, corn and soybeans for animal feed, and cotton for industry.

This approach gives rise to issues of equity and conservation. In terms of equity, the system favors rich farmers and puts poor ones at a disadvantage. In addition, specialization and economies of scale reduce economic diversity and employment opportunities in rural areas. The system entails three problems in conservation. Monoculture reduces biodiversity. The intensive use of fossil fuels and chemical inputs creates pollution; often the inputs of energy equal or even exceed the output of crops. Large-scale mechanized operations hasten soil erosion and other environmental degradation.

For these reasons, the trend of Western agricultural development toward industrial farming has come under increasing challenge from conservationists and also from social groups that feel threatened by it—among them In-

TRADITIONAL RICE FARMING entailed large amounts of hand labor. This scene of rice cultivation in the beautiful but poor Malaysian state of Kelantan was photographed some 20 years ago by the author. The woman is transplanting seedlings.

dian rebels in the Mexican state of Chiapas and owners of small farms in France. Are there alternatives to the Western model, or do we need to invent new models? Environmentalists have identified several apparently sustainable local farming traditions—all of them forms of polyculture. All farming systems were originally polycultures providing a range of basic requirements for subsistence. In some Mediterranean areas even today, one can find farmers planting wheat and barley around their olive trees. Much of the North American wheat belt used to support mixed grain and dairy farming. A form of polyculture that has recently attracted attention from agronomists because of its inherent sustainability is the system whereby corn, beans and squash are planted in the same hole. They complement rather than compete with one another because their root systems draw nutrition and moisture from different levels of the soil. In fact, the roots of the bean actually fix nitrates and so furnish natural fertilizer for the corn. This highly intensive form of land use was first developed well over 2,000 years ago. It sustained great civilizations such as the Maya and continues to support dense pockets of population all over Central America.

Are there other types of agricultural systems that might support sustainable but intensive development on a scale large enough to address the dire problems of rural poverty faced by many developing nations? The answer requires first a definition of "sustainable." To my mind, a sustainable agricultural system cannot be judged simply by the ecological soundness of its farming methods. It must also provide a living for all its population, farmers and nonfarmers alike. The majority of the world's poor live in the countryside; rural populations are still growing, and urban services and industries now absorb less labor than they once did. A sustainable agricultural system must therefore be able to create employment as well as to produce food. It should be flexible and diversified, able to yield not only subsistence but also marketable surpluses, and it should sustain an internal rural exchange of goods and services instead of depending heavily on the external world for both inputs and markets.

I want to propose that it is easier to plan development toward sustainable rural economies if we take as our model not the farming systems of the West, which inherently tend toward systems of monoculture and economies of scale, but systems of polyculture that use land intensively and offer a basis for economic diversification. Some food staples lend themselves more easily than do others to intensive polyculture. In this article, I use wet-rice farming in East Asia as my case, because its historical record is sufficiently rich to demonstrate a coherent pattern of technical and economic evolution. I do not suggest, however, that the world's problems will be solved if every region switches to wet rice. Almost any combination of food staples that uses land intensively will do.

SUPPLEMENTARY CROP of *kabocha* (Japanese pumpkin squash) is grown on a bund, or small dike, surrounding a rice field in Japan. Such a concentrated use of land is characteristic of farming in East Asian polyculture.

The view of "proper" agricultural progress that the West has inflicted on the rest of the world has its historical roots in the development of farming in northwestern Europe and the grain belts of the New World—the regions that supplied food for the urban centers of the industrial revolution. But this dynamic, in which labor is a scarce resource and output is increased by substituting technical innovations for manpower and animals, is not inevitable; it is predicated on the conditions of production specific to those regions.

Northern Europe, where this dry-grain farming system evolved, has a short growing season. The staple cereals—wheat, barley and rye—bear seed heads, or panicles, with relatively few grains, at best a few dozen compared with 100 or more grains on a panicle of rice or millet. Each plant usually has no more than three or four stems, or tillers. In principle, one seed could produce some 200 offspring, but the biblical parable reminds us that many seeds die where they fall. Farmers in medieval Europe had to keep as much as a third of their crop for the next year's seed; another large portion went to feeding draft animals over the winter. Because the only fertilizer available was manure, land had to be left fallow often and could be planted with cereal only once every two or three years. In short, this farming system used land *extensively* and could not support high population densities. The typical 11th-century English holding, as recorded in the Domesday Book, was 30 acres (12 hectares).

Draft animals played a crucial role in this farming system. Yields were so low that it was impossible to till enough land for subsistence by manpower alone. Some plow teams consisted only of a pair or two of oxen, but in the heavy clay soils typical of northern Europe, where a plowshare had to cut deeply to turn the soil over, as many as a dozen oxen might form a team. Where draft animals and heavy implements figure prominently in agricultural production, it is clear that large farms, which can afford more animals and equipment and can organize their use more efficiently, will have a significant advantage over smaller holdings. The larger the farm in medieval Europe, the more likely it was to produce a surplus.

Urban markets for food grew in the 12th and 13th centuries, and the old feudal systems, under which serfs worked both their own strips of land and the lord's domain, began to break down. Manorial lords started to consolidate and enclose large holdings and farm them with wage labor. The laborers were often peasants who had lost traditional rights to land as its ownership became privatized. If landowners let their land to tenants, it was not to subsistence smallholders but to better-off farmers—small capitalists like the English yeomen, who could bear the risks of investment in animals and equipment. Capitalist relations in agriculture had formed in many parts of northwestern Europe before the 15th century. Markets in land and labor were well developed. The social relations necessary for the foundation of a modern mechanized agriculture were thus in place, but the necessary technical expertise was lacking.

Development of this agricultural system was driven by the superior performance of large, centrally managed units of production. The 18th century recorded improvements that included new crop varieties and breeds of animal, better plows and drainage systems, and crop rotations that combined cereals with fodder crops such as clover and turnips. All the experts agreed that only large farms were suitable for these "high farming" methods. Economies of scale dictated who could afford such improvements.

Before mechanization, many high-farming innovations required increased labor as well as capital. In northwestern Europe, farmers had to compete with the new and expanding industries for workers; in the sparsely populated New World, labor was simply very scarce. Inventors had been tinkering with farm machinery as early as the 16th century, but without much success. By the early 19th century the need for such machines was felt acutely.

That was the time when engineers could at last draw on materials and expertise from the industrial sphere—steel, steam power and chemicals—to develop labor substitutes for agriculture. The first successful mechanical threshers came on the British market in the 1830s (provoking riots by agricultural laborers as they saw their precarious livelihoods threatened). Horse-drawn reapers, harvesters and mechanical drills followed, and eventually in the 20th century the tractor replaced the horse. Chemical fertilizer eliminated the necessity for crop rotations and facilitated monoculture. Herbicides and pesticides further reduced the need for labor. The amount of agricultural land per agricultural worker in the U.S. today is 137 hectares, and a medium-size farm of the type usually run by a single family ranges between 20 and 100 hectares.

This is the historical experience from which our image of "normal" agricultural progress derives. Just as Western patterns of industrialism spread from nation to nation, defining our notions of a modern economy, so, too, after World War II the characteristics of the Western agricultural revolution defined the worldwide agenda of agricultural modernization. Such progress seemed normal and inevitable to the postwar agricultural economists and scientists, mostly from or trained in the U.S., who worked out a package of technical and economic aid to modernize agriculture in the poorer nations.

The new technology they developed gave such impressive initial results that it quickly came to be called the green revolution. The technology centers on the use of high-yielding varieties of wheat, corn and rice. These varieties are hybrids that farmers cannot breed themselves and that need chemical fertilizers and herbicides to thrive. In experimental stations the hybrids produced such high yields that they were soon called miracle seeds. As Indian economist Vandana Shiva points out, however, comparisons between old and new varieties measure only the output of that one crop, not of the whole mixed cropping system that it often displaces, so the overall gains may be much less than claimed.

Because of the emphasis on monoculture, the agricultural agencies that supply technical information, seed and credit to farmers usually advocate large-scale cultivation and the consolidation of holdings to make mechanization feasible. Under these conditions, salable surpluses and profit margins (but not necessarily yields) are generally proportional to the size of the farm, and small farms lose their viability.

The primary aim of the green revolution policies of the 1960s and 1970s was the eradication of world hunger: the modernization of underproductive farming systems would increase the world output of staple grains. In this respect, the green revolution has been a great success. The world's production of the main staple grains (wheat, corn and rice) would today be more than adequate to feed the world's population if it were not for problems of maldistribution.

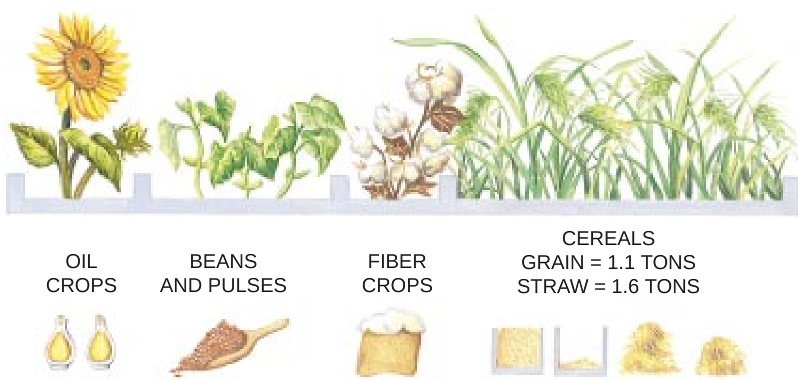
But as farmers have been encouraged to concentrate on monoculture, they have become more vulnerable to crop pests and price fluctuations. The variety of local diets has been drastically reduced, as have employment opportunities. The new technology uses enormous amounts of chemicals and fossil fuels. In energy terms, it is less efficient than many traditional farming systems. Monoculture, mechanical plowing, the extension of crops into woodlands and

Agricultural Productivity

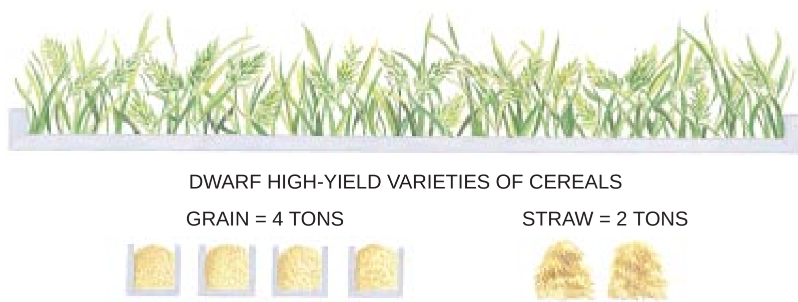
Economic calculations of agricultural productivity usually take into account only the yield of a particular crop per unit of land and overlook other uses to which the land may be put. The drawings show the result of such a calculation comparing a polyculture (growing several different crops on a hectare of land) with a monoculture in which only rice—a dwarf, high-yielding variety common in green revolution agriculture—is grown.

In the polyculture (*top*) one hectare of land is used for several crops in a year, producing as the main crop 1.1 tons of a cereal grain (rice) and 1.6 tons of straw used for fodder and fuel, but also producing as secondary crops quantities of oil, beans and fiber. The monoculture (*bottom*) produces four tons of rice and two tons of straw. Because the typical calculation of productivity applies only to yields of a single crop, the comparison puts the polyculture in an unfavorable light—1.1 tons of grain per hectare as against four tons for the monoculture crop. The other yields of the polyculture are ignored.

TRADITIONAL POLYCULTURE



GREEN REVOLUTION MONOCULTURE



pastures, and the use of chemical products all contribute to environmental degradation.

The second aim of green revolution policies was to generate rural prosperity through the production of marketable surpluses. It seemed clear that the application of science and capital would yield more efficient and productive farming practices. Theories in vogue at the time recognized that the capital requirements of this kind of modernization would initially favor wealthier farmers but assumed that soon the benefits would trickle down to the entire population.

In fact, many regions have experienced a severe economic polarization. Rich farmers add to their holdings while poor ones are edged out of farming into a dependent wage-labor force. The people who can afford to farm rely increasingly on the urban economy for goods, services and markets. Opportunities for work in the countryside diminish, but urban industry cannot generate enough jobs, and the unemployed congregate in city slums.

The parallels between the green revolution and the 18th-century modernization of Western farming are clear. If advocates of the green revolution neglected to consider the negative social and ecological consequences of their plans, it was largely because this style of development, with its reliance on capital and machinery, seems to represent the inevitable path to modernization.

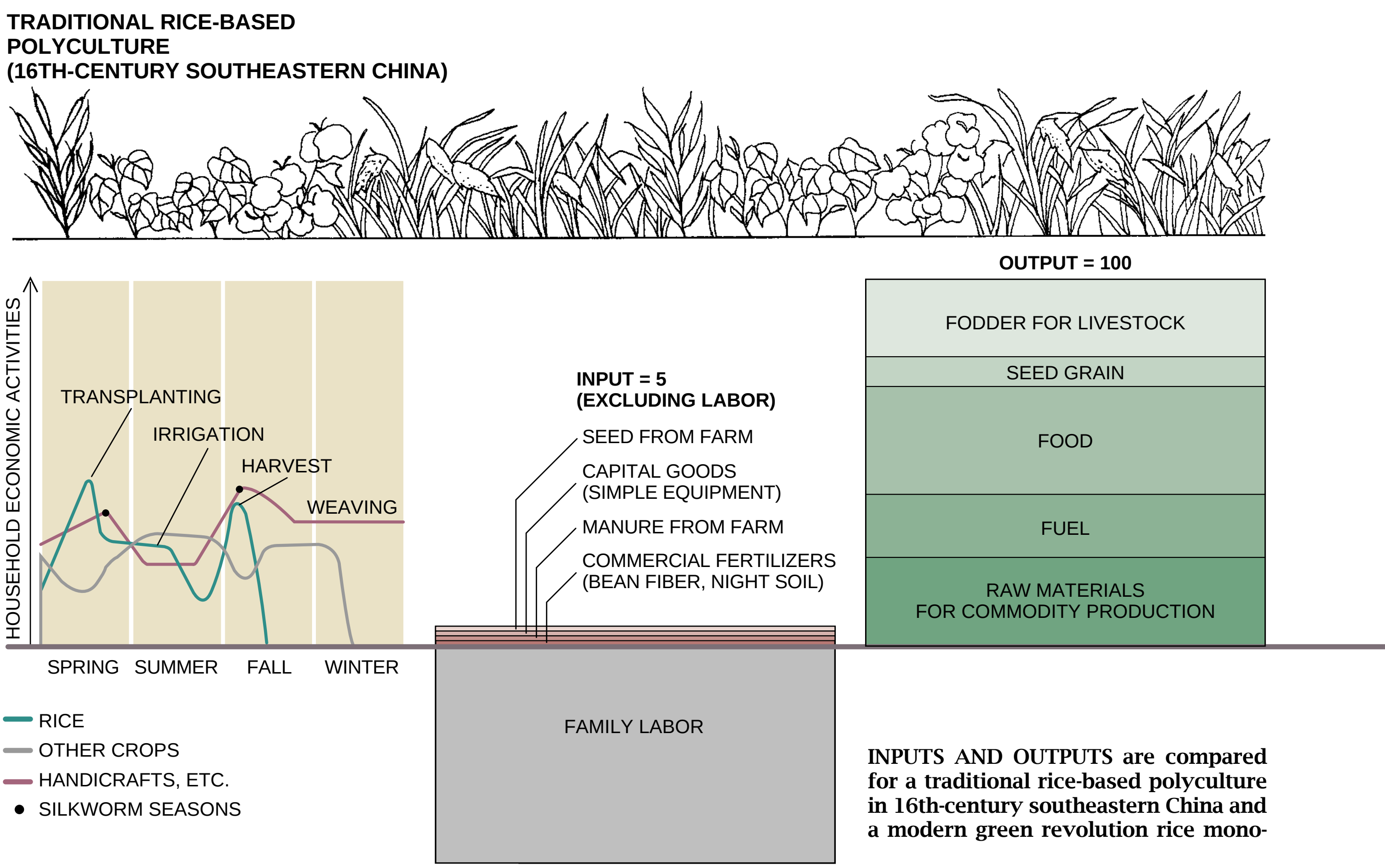
It was in 1976, during a year spent studying farmers' reactions to the green revolution in the beautiful but poor Malaysian state of Kelantan, that I began to think about alternative models of agricultural development. Before I went to Kelantan, I had spent several years researching the history of rice cultivation in China. As I read more about agricultural development, I realized that many Japanese experts had reached conclusions similar to mine based on their historical experience. They, too, saw a logic in the historical intensification of Asian rice cultivation that was quite different from what had happened in the West. They also felt that the introduction of green revolution technology often represented a disastrous break with the past, and they suggested that there would be many advantages to adopting the "Japanese model."

Looking at the conditions of production and the consequences of development, one finds that the Japanese (or, better, East Asian) model, which centers on the production of wet rice, differs radically from the dry-wheat model of northern Europe. In China, Japan, Vietnam and Korea, the use of land was intensified over the centuries because of the increasing availability of skilled labor. There were few economies of scale, smallholdings predominated and intensive cropping patterns sustained a mixed farming system and a highly diversified rural economy that could provide a living for large populations.

Water is a crucial factor in shaping the development of rice cultivation. Rice is a monsoon crop; it can be grown in dry fields, but water is its natural habitat. The earliest find of domesticated rice so far is in a Neolithic Chinese village near Shanghai, situated at the edge of a shallow marsh and dated to approximately 5000 B.C. Other early sites dotted around southeastern continental Asia are also close to marshes or other natural water supplies.

A good rice field or paddy is one in which the water supply can be accurately regulated and drained. As a result, paddies are usually quite small by Western standards: a field 20 yards square would be considered large in China. Young rice seedlings need damp soil but rot in standing water; once they are about a foot tall, they like to have several inches of standing water through the period of flowering and ripening, after which the field should be drained for several days before harvesting.

Rainwater can easily be impounded in a field surrounded by bunds (small



INPUTS AND OUTPUTS are compared for a traditional rice-based polyculture in 16th-century southeastern China and a modern green revolution rice mono-

dikes), but it may evaporate before the rice is fully grown. Rice farmers in some regions therefore adopted rain-fed tank irrigation systems very early. Other forms of irrigation include the channeling of small streams into hillside terraces and the construction of diversion channels from larger rivers—in which case the water usually has to be pumped up into the fields. All these forms were common in China and Japan by medieval times and allowed rice farming to spread from small river valleys up mountainsides and down into the deltaic floodplains. Constructing bunds, irrigation networks, tanks or terraced fields requires large initial investments of labor, but thereafter maintenance is relatively cheap and easy. So it is not surprising that rice farmers have often preferred intensifying production in their existing fields to extending the cultivated area.

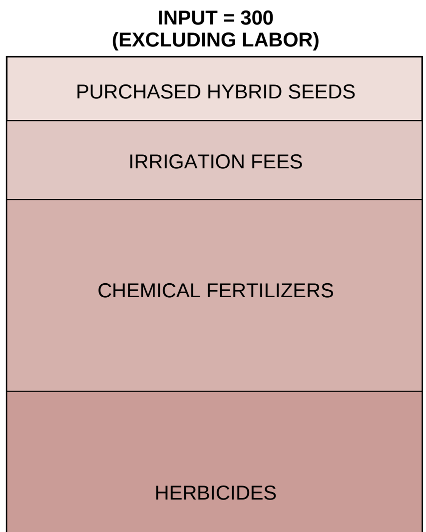
Water enhances the sustainability of rice systems. Unlike dry fields, rice paddies gain rather than lose fertility over the years. Whatever the original structure and fertility of the soil, over several years of continuous wet-rice cultivation the top few inches of soil turn to a fine, gray, low-acidity mud with a layer of hardpan below that retains the water. Nitrogen-fixing organisms that occur naturally in the water serve as a manure. Traditional rice varieties usually respond well to organic fertilizers; lime and soybean waste were widely used in both China and Japan by the 17th century, giving annual yields of up to six tons per hectare in some double-cropping areas.

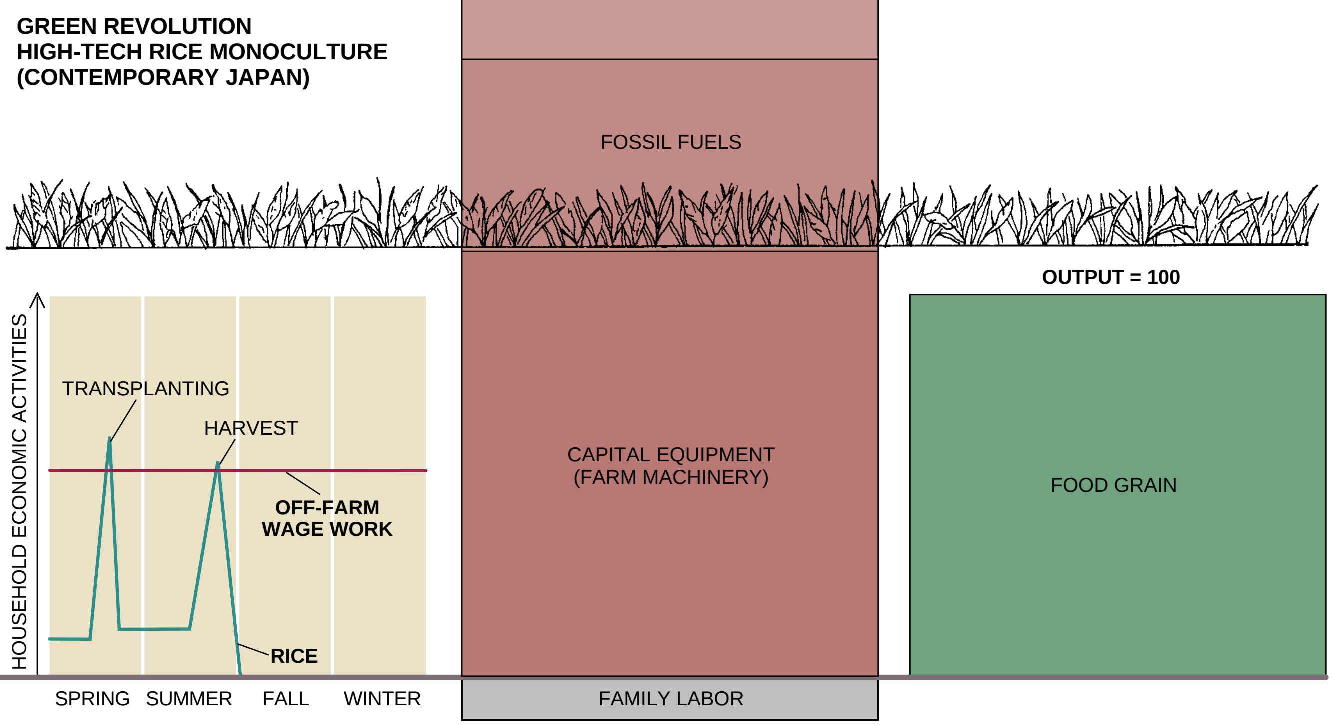
Rice plants have several seed-bearing stems, and each seed head contains on average about 100 grains. The technique of transplanting rice seedlings augments these traits. A small patch of fertile land is meticulously tilled, manured and sowed with carefully selected pregerminated seed. Meanwhile the main field is soaked, plowed and harrowed to create a fine silky mud. After a month or so, the seedlings are pulled up, the sickly ones are discarded and the tops of the leaves of the healthy ones are chopped off. Then the seedlings are replanted in shallow water in the main field.

This procedure is labor intensive, but it permits the careful selection of healthy plants and the efficient use of small amounts of manure. Moreover, the plant responds to the transplanting process by growing more tillers. By the time the seedlings are transplanted, they need only a few weeks in the main field. Hence, the land can be used for other crops in the off-season.

Wet-rice cultivation has enormous potential for expanding the uses of land. Tanks, channels and bunds may occupy as much as a fifth of the land, but no space need be wasted. Fish nibble the weeds in the tanks or eat snails in the paddy, and ducks feed on the fish. Narrow bunds serve for growing vegetables, and broad bunds may be planted with mulberries to feed silkworms, whose droppings are used as manure. After the rice is harvested, the field can be drained to grow barley, vegetables, sugarcane or tobacco.

The alternation of winter rice with summer wheat became common in the lower Yangtze region of China 1,000





culture in Japan. The height of the labeled bars reflects the relative amount of that input or output. The curves at the bottom left of each diagram indicate how the people of the farm household apportion their productive time. In the polyculture economy the women do little work in the fields but are heavily involved in handicrafts such as silk production. In the monoculture economy, women do more of the fieldwork because many of the men have off-site jobs.

years ago. A judicious choice of fast-maturing varieties and the abundance of water afforded 17th-century farmers in the Canton region two or even three crops per year plus a few side crops of vegetables; yearly yields totaled as much as seven tons per hectare. Because fields were small, farm implements were small, light and cheap. A single water buffalo served the needs of a typical farm; if production was really intensive, the farmer might give up plowing altogether in favor of hoeing.

In general, rice farming did not require much capital outlay compared with dry-wheat farming, and there were few economies of scale to be practiced. Although a landlord in south China might own as much land as his English counterpart, his home farm would be of modest size, and the rest would be let out in small parcels to many tenants, chosen not for their capital assets but for their skills and experience. The system did not polarize rural society and drive poor people out. The relative advantage of smallholdings guaranteed access to land for large numbers of peasants, even if it was through the exploitative relation of tenancy.

The labor requirements of wet-rice farming are high but intermittent. Peasants in medieval China and Japan could therefore use rice farming as the basis for the commercial production of vegetables, sugar, silk or tea or for the household manufacture of textiles, liquor, bean curd or handicrafts. Rice served as the foundation of a rural economy that both required and absorbed the labor of a dense population.

Economic historians have often equated this system with “agricultural involution,” by which individuals work harder and harder for ever decreasing returns. The assertion might be true if calculations were based only on rice yields, as if one were dealing with a monoculture. But when all the other goods produced in such an economy are taken into account, the system appears in a much more favorable light. Although its capacities for expansion are not infinite, they are considerable. During several centuries of population growth, China’s rice regions established the foundation for a rural economy in which many of the people made salable goods at home. Only after 1800 did rural living standards begin sharply declining—a trend that was exacerbated by the effects of multiple wars.

A similar process of rural development took place in Japan, creating the basis for the building of the modern state. This achievement is one reason Japanese agronomists see their system as an exportable model. Yet in Japan as in the West, industrialization was achieved through ruthless patterns of exploitation. Between 1600 and 1800 the rural economy expanded in conjunction with the growth of trade and cities. Techniques for growing rice were improved, and land became so productive that the Meiji government of 1868–1912 was able to fund the construction of a modern industrial state mostly through raising agricultural taxes. But this level of extraction left tenant farmers in a state of near destitution that the state did not feel obliged to address until the introduction of universal suffrage in 1945.

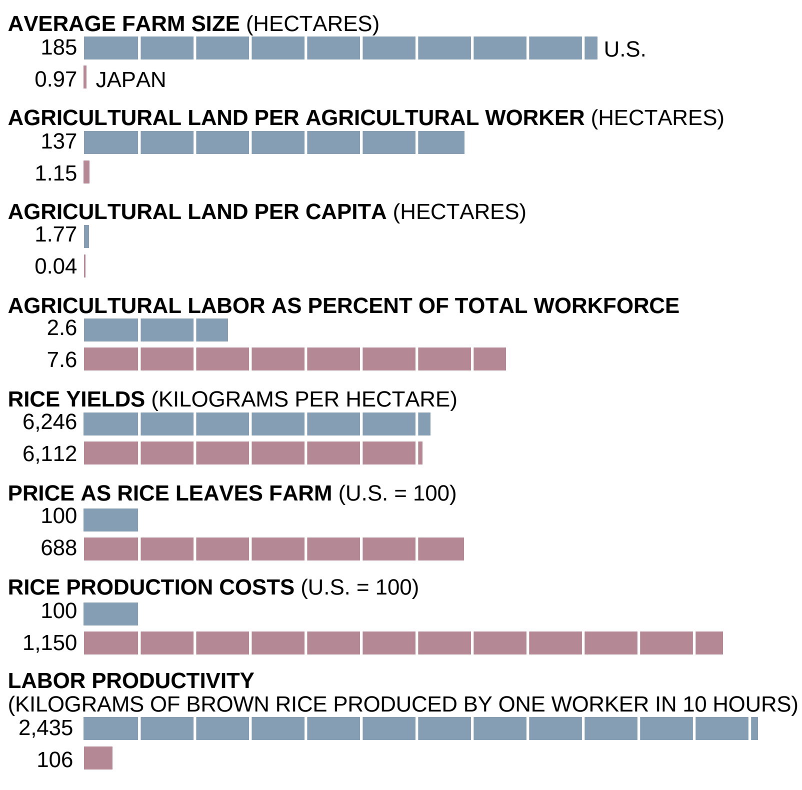
The new regime set out to guarantee rice self-sufficiency and to eliminate rural poverty. Land reforms were enacted to do away with tenancy and set stringent limits on the purchase of land. This policy institutionalized the tiny but independent family rice farm, supplying a framework for the successful long-term balancing and integration of rural and urban development.

Yet Japanese agriculture today is in a state of crisis. Except for the aberrant poor harvest of 1993, caused by bad weather, rice is overproduced and wastefully produced, in large part because of heavy subsidies and price support paid by the government since the 1950s. The strategy of increasing rural incomes by raising rice prices has backfired. Until the 1960s, Japanese farmers used moderate inputs and simple machinery. Since the 1960s, mechanization has taken over in rice production with small-scale tractors, transplanters and harvesters. Almost all farmers own a full range of expensive machinery, and the average use of fertilizer per hectare is 1,110 kilograms (compared with 160 in the U.S. and 48 in Thailand). As long ago as 1977, the Japanese economist Taketoshi Udagawa calculated that energy inputs amounted to three times the food energy of the rice. It costs 15 times as much to produce a kilogram of rice in Japan as in Thailand and 11 times as much as in the U.S.

No one in Japan today would call this policy economically sustainable. Nor is it any longer conservationally sound. Although the irrigated fields and channels still protect Japan’s narrow river valleys from floods, the channels and soil are saturated with chemicals. No fish or frogs swim in the paddies now.

Japan’s current crisis makes it clear that the East Asian model of agriculture, too, can go awry. Yet it would be tragic if the Japanese gave in tamely to

Comparison between the U.S. and Japan



the advice they are hearing to adopt the Western style rather than seeking creative endogenous solutions that might be ecologically and socially more rewarding.

Such solutions may be already at work. Through recent economic reforms in Japan, Taiwan and China, the patterns of land use and economic diversification based on rice cultivation have brought about a modernization characterized by an unusual degree of balance between rural and urban development. The rising ratio of farm-household income to the household income of industrial workers in Japan shows the trend: 69 percent in 1960, 92 in 1970, 115 in 1980 and 113 in 1988.

What are the implications for sustainable rural development elsewhere? This is a problem faced not only by nations with large and impoverished rural populations, such as Mexico and India, but also by wealthy nations, such as France, that want to avoid further rural depopulation. Monoculture is not an irreversible trend, but in today's global economy, rural diversification does require structured support and fair prices for agricultural products. In Japan, where consumers will pay high prices for fruits and vegetables, large numbers of rice farmers have been persuaded to switch part of their land to orchards and truck farms. In China, the state abandoned the Maoist policy of "putting grain first" in the late 1970s. It allowed farmers to combine a basic level of grain farming with all kinds of other crops and livestock. At the same time, farm prices were increased to a realistic level. Agricultural production shot up overnight. Farmers produced not just food but also the raw materials for the development of rural industry. Moreover, they became wealthy enough to consume a wide range of industrial goods. China's current spectacular growth rates can be understood only against this background of rural revitalization.

The examples of premodern China and Japan show that intensive polyculture, precisely because it does not depend on expensive inputs, can yield a livelihood for poorer farmers, offer widespread access to land and generate other employment opportunities. Ideally, polyculture should not only support rural diversification but also lessen dependence on industrial inputs. Mayan peasants can grow corn without buying chemicals because beans naturally manufacture nitrates. But a farmer does not have to operate at the scale of peasant subsistence to do without chemicals. In California, organic vegetable growers and wine producers are developing interplanting techniques (another form of polyculture) to substitute for chemical pesticides. They grow more kinds of plants, hire more workers and buy fewer chemicals—and they are doing a big business.

MODERN RICE FARMING is increasingly done with machines designed for the small scale of rice fields. Here a farmer in Japan operates a machine that transplants rice seedlings in a wet paddy after they have grown to a length of about a foot.

The examples I have cited should be a stimulus to look closely at other non-Western agricultural systems. If we are to find long-term solutions to the truly modern problem of feeding the world without destroying it, we have much to learn from such systems.

FURTHER READING

JAPANESE AGRICULTURE: PATTERNS OF RURAL DEVELOPMENT. Richard H. Moore. Westview Press, 1990.

THE VIOLENCE OF THE GREEN REVOLUTION: THIRD WORLD AGRICULTURE, ECOLOGY AND POLITICS. Vandana Shiva. Zed Books and Third World Network, 1991.

JAPANESE AND AMERICAN AGRICULTURE: TRADITION AND PROGRESS IN CONFLICT. Luther Tweeten, Cynthia L. Dishon, Wen S. Chern, Naraomi Imamura and Masaru Morishima. Westview Press, 1993.

THE RICE ECONOMIES: TECHNOLOGY AND DEVELOPMENT IN ASIAN SOCIETIES. Francesca Bray. University of California Press, 1994.

The Scientific Legacy of Apollo

The retrieved lunar rocks have helped settle questions about the moon's origin, its composition and even the early conditions that affected life on the earth

by G. Jeffrey Taylor

When Neil A. Armstrong and Edwin "Buzz" Aldrin, Jr., dug into the moon's surface 25 years ago, they were doing more than collecting dry, dark dirt. They were time traveling. Their journey in *Apollo 11* across 380,000 kilometers of space sent them back billions of years. Armstrong, Aldrin and the 10 astronauts who followed returned with samples that contain a fascinating history of the moon and the earth. The rocks have indicated the moon's violent and surprising origin, its composition and its age. Instruments placed on the surface enabled geophysicists to reconstruct the satellite's internal structure and activity. Without the Apollo program, none of these discoveries could have been made.

By traveling to the moon, we also learned about the earth. Volcanism, folding, faulting, mountain building, weathering and glaciation have erased or modified most of the earth's ancient history. Fortunately, the moon was not so energetic a geologic engine. It was active enough in its first billion years to produce an intriguing and complex array of products, but not so vigorous that it completely eradicated the chronicle of what had happened. By comparing the moon's craters, lava flows and volcanic debris with corresponding formations on the earth, workers can test theoretical models of the mechanisms that created such features here.

The Apollo missions, of course, did not instantly modify the thinking about our nearest celestial neighbor. It took several years to analyze the samples and to form reasonable theories based on those empirical findings. The landings recovered 382 kilograms of moon material from six sites. The rocks quickly oxidize when exposed to air, so they are preserved in a dry, nitrogen-filled chamber at the National Aeronautics and Space Administration Lyndon B. Johnson Space Center in Houston.

Among the first questions the samples resolved was the moon's age. Isotopic dating showed that the moon formed at the same time as did the

G. JEFFREY TAYLOR, who received his Ph.D. in geology from Rice University in 1970, is a professor at the Hawaii Institute of Geophysics and Planetology, School of Ocean and Earth Sciences and Technology, University of Hawaii at Manoa in Honolulu. He chairs the lunar exploration science working group, a committee that advises the National Aeronautics and Space Administration on future lunar missions. He has recently become active in studies of the dynamics of lava flows on the earth, the moon, Mars and Venus. His belief that education is a prime justification for a vigorous space program has led him to develop instructional materials for use in grades 4 through 12.

EARTHRISE over the Mare Smythii region, located on the eastern limb of the moon, was taken 25 years ago by *Apollo 11.* It epitomizes the idea that we can learn about the earth by studying the moon.

earth, 4.5 billion years ago. The rocks also indicated that the moon was geologically active until about two billion years ago. Other major questions took longer to answer.

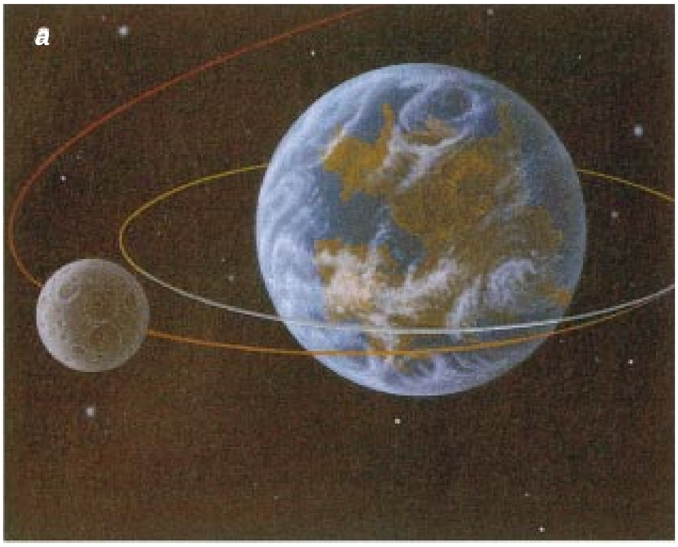
In fact, investigators did not achieve a consensus on a theory of the moon's origin until 1984, 12 years after the last Apollo mission flew. The agreement emerged from a conference I organized with William K. Hartmann of the Planetary Sciences Institute in Tucson and Roger J. Phillips, now at Washington University. The meeting was held in Kona, on the big island of Hawaii. Given the tenacity with which scientists cling to their views, none of us suspected that one of the hypotheses of lunar origin would spring forth as a leading candidate above the others. Certainly none of us thought the postconference favorite would not be one of the three classic hypotheses. Each of these hypotheses had what some considered to be fatal flaws. Each also had ardent supporters. It is a testament to human persistence and imagination that so many scientists tried so hard to adapt their preferred idea to a growing list of facts. Many houses of cards came tumbling down in Kona.

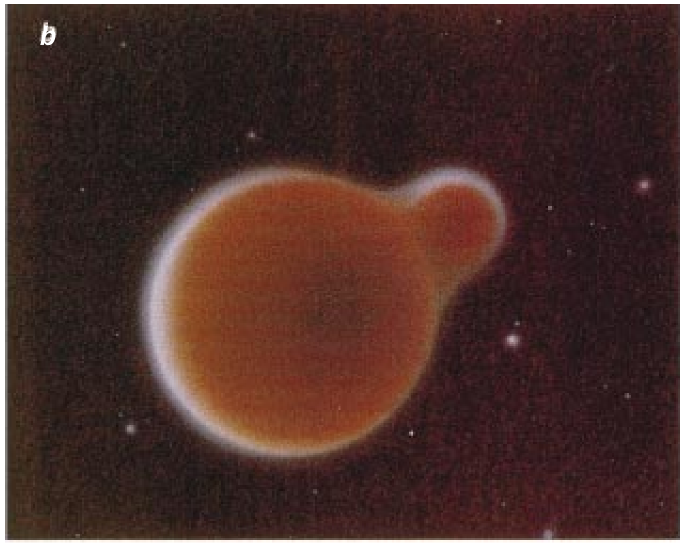
The least favorite classic idea going into the conference was the capture hypothesis. In its original form the capture hypothesis held that the earth seized a fully formed moon that came whizzing in from elsewhere in the solar system. In principle, such capture is possible but unlikely. A body passing near the earth would probably collide with it or get a gravitational boost that would alter its orbit so much that it could never meet up with the earth again. The chances of the orbits of the moon and the earth being exquisitely right for a capture is so minuscule that all but a few scientists had rejected the idea.

The Apollo mission helped to put that theory to rest. Lunar samples showed that the moon and the earth have similar quantities of oxygen isotopes, suggesting a close kinship. If the moon had formed elsewhere in the solar system, it would probably have had a different isotopic oxygen composition from that of the earth.

The second classic lunar genesis idea presented was the fission hypothesis. This theory has a long and honorable history. George Darwin, the second son of Charles's 10 children, first proposed it. He postulated that the earth, during a period after it formed a core, was at one time spinning extremely fast. It bulged so much at the equator that eventually a small blob spun off, becoming the moon. The scenario would account nicely for a crucial feature of the moon deduced by astronomers more than 100 years ago. Based on the satellite's orbital characteristics and size, the investigators calculated that the moon must be less dense than the earth. The low density implies that the moon must have only a small metallic core, if it harbors one at all. The fission idea would explain this fact: a fissioned moon is composed mostly of the earth's mantle (the layers between the crust and the core).

Subsequent calculations showed that the earth would have to have been rotating once every 2.5 hours in order to have spun off the material that became the moon. This short day is among the chief problems with the hypothesis: no one can figure out how the earth would have been spinning so fast in the first place. The models that described planetary formation as an accumulation of dust grains indicated that the earth would end up spinning rather slowly, if at all. Incorporating events that add angular momentum—most notably, impacts of planetesimals up to a few hundred kilometers across—did not help. Computer simulations showed that for every object that struck the earth to add clockwise spin, another impact would cause the planet to spin counterclockwise. Even if there were a mechanism for imparting enough angular momentum into the earth, advocates of the fission hypothesis had to find a way to eliminate much of the rotational energy. The earth-moon system of today does not have nearly the amount of momentum needed to initiate separation of the two bodies from one another. Nevertheless, the calculations left enough room for intellectual maneuvering to keep the fission hypothesis





SCENARIOS OF LUNAR FORMATION varied widely before the Apollo missions. The capture hypothesis (*a*) depicts the moon as a body caught by the earth's gravity. In the fission idea (*b*), rapid rotation caused a piece of the early earth to

from being discounted on dynamical grounds alone.

The Apollo program provided a new test. If the moon split from the earth in this manner, it ought to have exactly the same composition as the earth's rocky material near the surface (specifically, the crust and mantle). The moon and the earth do have identical amounts of oxygen isotopes, which indicates that the two bodies are related in some way. But the compositional similarity ends there. Crucial data came from lunar samples, a network of seismometers left behind and spectroscopic studies by the Apollo 15 and 16 missions. They enabled researchers to conclude that the moon and the earth have different chemical compositions.

For example, the moon has much less volatile material—substances that boil away easily—than does the earth's mantle. The satellite completely lacks any water-bearing minerals: it is bone-dry. It also lacks other kinds of volatile elements, from common ones such as potassium and sodium to more exotic chemicals such as bismuth and thallium. Scientists also discovered that the moon is enriched in nonvolatile substances relative to the earth. Called refractories, these elements are the opposite of volatiles: they boil at high temperatures. It appears that refractories such as aluminum, calcium, thorium and the rare-earth elements are present in the moon at concentrations that are about 50 percent higher than those in the earth. Another bit of damaging evidence against the case for fission comes from the ratio of iron oxide to magnesium oxide. The ratio of these common compounds seems to be about 10 percent higher in the moon than it is in the crust and mantle of the earth.

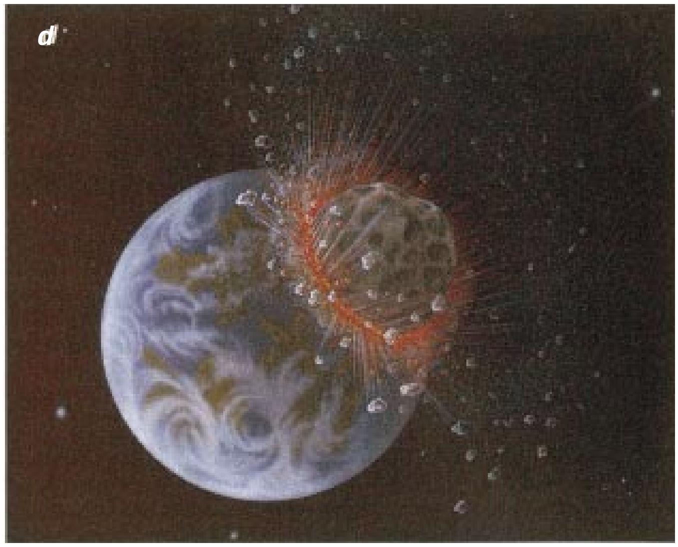
Despite the evidence, fission proponents did not yield. They developed schemes to drive off volatiles and enrich refractories; they widened the error bars in the iron oxide to magnesium oxide ratios sufficiently to claim that the two bodies are indistinguishable. But in the long run the Apollo program findings have convinced most investigators that the fission model fails the compositional test.

The third classic idea is the double planet hypothesis, by which the moon and the earth formed concurrently from a cloud of gas and dust. Thus, the raw materials for the moon came from a ring of material in orbit around the earth. As the earth grew, so did the ring and the embryonic moon within. This hypothesis always had trouble explaining why the moon has such a small metallic core compared with that of the earth. Richard J. Greenberg and Stuart J. Weidenschilling and some of their colleagues at the Planetary Sciences Institute and at the University of Arizona tackled the question during the year before the Kona conference. They suggested that the orbiting ring of material acts as a compositional filter. The rocky parts of incoming bodies break up easily and are incorporated into the ring; metallic parts pass through to become part of the earth.

Much debate centered on the efficacy of the process, and many researchers raised doubts about whether the incoming bodies would have been separated into cores and mantles. Although the binary planet hypothesis explains the similarity of the composition of the earth and the moon with respect to oxygen isotopes, it does not account for the differences in volatiles and refractories. Most important, it runs into the angular momentum problem. That is, it does not explain how the earth's rotation came to be 24 hours, which is faster than predicted by simple accretion models, and how the ring could have acquired enough circular motion to stay in orbit.

As we were organizing the Kona meeting, I wondered what new twists the proponents for these fatally flawed ideas would devise. Although modifications were indeed presented, we were all surprised by the enthusiastic reception given to a long-ignored idea: the giant impact theory. Among the most surprised was conference co-organizer Hartmann, and he was one of the inventors of the notion. By the end of the deliberations, a clear consensus formed in support of the idea that the impact of a large projectile with the growing earth dislodged the material that would form the moon. Of course, skeptical diehards and others desperately clinging to one of the old ideas remained, but the giant impact theory of lunar origin enchanted most participants.

The idea was not really brand-new. Like an actor who achieves "overnight success" after years of small roles, the giant impact theory was a bit player for a long time. Hartmann and his colleague Donald R. Davis proposed the impact theory in 1975. They had been investigating the accumulation of planets from smaller objects and noticed that numerous large bodies would have wan-



break off and become the moon. In the double planet hypothesis (*c*), dust grains accumulated to form the earth and moon.

The now dominant giant impact theory (*d*) states that a huge collision flung into orbit debris that became the moon.

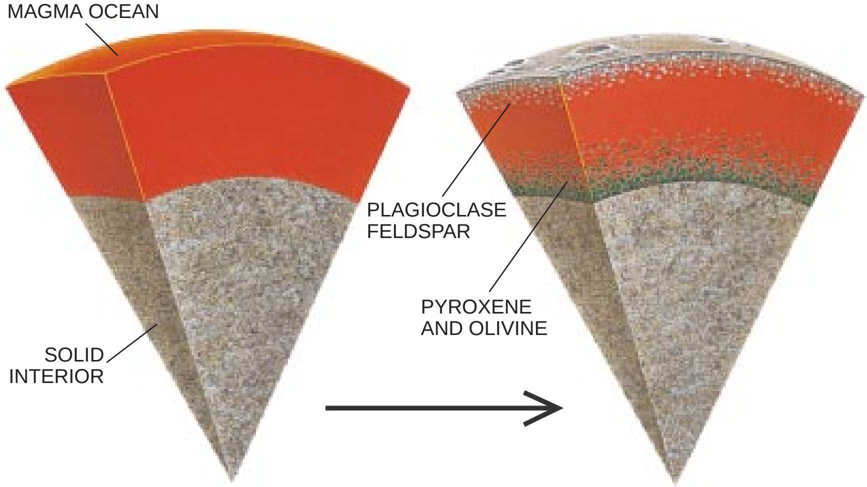
dered near the earth. A few could have been as large as Mars. Hartmann and Davis hypothesized that the earth collided with such an object. As a result, some of the debris was launched into orbit, providing raw material for the moon. Alastair G. W. Cameron of the Harvard-Smithsonian Center for Astrophysics and William R. Ward of the Jet Propulsion Laboratory in Pasadena, Calif., independently suggested the same idea one year later as they tried to resolve the angular momentum problem. They also addressed the details of the mechanism by which the material could achieve orbit and not fall back to the earth.

The work of Hartmann and Davis had been anticipated almost 30 years before by the late geologist Reginald A. Daly of Harvard University. Two distinguished pioneers of lunar science, Ralph B. Baldwin of the Oliver Machinery Company in Grand Rapids, Mich., and Don E. Wilhelms of the U.S. Geological Survey in Menlo Park, Calif., found that Daly had suggested in 1946 that the moon formed from the earth by the glancing impact of a planet-size object. Although Daly's analysis contains errors, the giant impact idea is clearly stated in this insightful but completely ignored paper. Even if Daly's work had been widely read, it might have been discounted: the paper was published before scientists realized that impacts were an important planetary process [see "Prematurity and Uniqueness in Scientific Discovery," by Gunther S. Stent; SCIENTIFIC AMERICAN, December 1972].

Although the giant impact hypothesis lay in obscurity until the announcement at Kona, nothing since that 1984 conference has shaken its firm position as the leading theory. It simply explains too many observations. The moon lacks metallic iron at its center because the core of the impactor stuck to the earth, so the moon formed from the silicate parts of both objects. The difference in the ratio of iron oxide to magnesium oxide between the earth and the moon exists because the moon formed mostly from the impactor. (One assumes that the projectile harbored less iron oxide than does the earth.)

The moon is dry because of the incalculable amount of heating that took place during the collision: the high temperatures evaporated all water and other volatiles. The refractories enriched the moon's composition because they recondensed quickly after heating and so were incorporated. The identical oxygen isotopic composition of the earth and the moon arises because the impactor and the earth formed in the same region of the evolving solar system. Finally, the hypothesis explains the most difficult problem: the angular momentum of the earth-moon system. The projectile must have struck the earth off-center, away from the central axis. This type of blow would have sped up the earth's rotation to its current value.

The most enticing aspect of the giant impact theory is that such a collision is a natural consequence of planet formation. No unusual or ad hoc circumstances need to be invoked. Such catastro-



OCEAN OF MAGMA is thought to have once encircled the moon over its solid interior. Over 100 million years, the magma slowly crystallized. Lighter material, mostly plagioclase feldspar, rose to the surface and eventually formed the lunar crust. Heavier compounds, chiefly olivine and pyroxene, sank and became the moon's mantle.

phes, while enormous, are not unlikely. Indeed, planetary scientists now appeal to giant impacts to explain the composition of Mercury and the large tilt of Uranus. Without this colossal event early in the history of the solar system, there would be no moon in the sky. The earth would not be rotating as fast as it does, nor would it have such strong tides. Days might even last a year, as they do on Venus. But then, we probably would not be here to notice.

In addition to settling the question of the moon's origin, the Apollo samples enabled researchers to deduce the satellite's structure and evolution. The moon's features appear to have been reworked quite a bit by internal processes, though to a much lesser extent than those of the earth. A gigantic body of magma hundreds of kilometers deep, which apparently surrounded the moon and helped to form the lunar crust and mantle, precipitated the events. Indeed, this magma ocean theory has reigned as a central tenet of lunar science since *Apollo 11* returned the first samples.

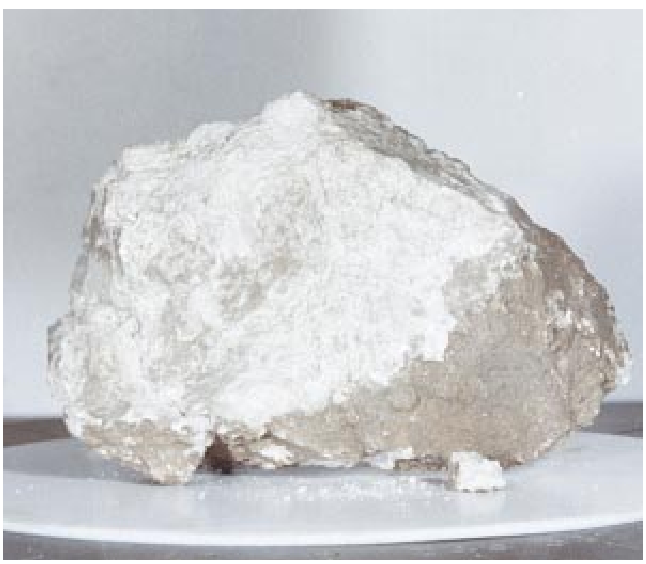
The *Eagle* landed on the surface of a mare in the Sea of Tranquillity, one of the dark gray areas that paint the features of the man-in-the-moon. They are the remnants of vast lava flows that leaked out onto the surface billions of years ago. Armstrong and Aldrin returned mostly samples from underfoot—basaltic bedrock rich in titanium. They also returned samples of the lunar regolith. The regolith is the loose debris that falls back to the moon after meteoroid impacts have kicked them up. It is the moon's version of soil, covering most of the surface to depths as great as 20 meters.

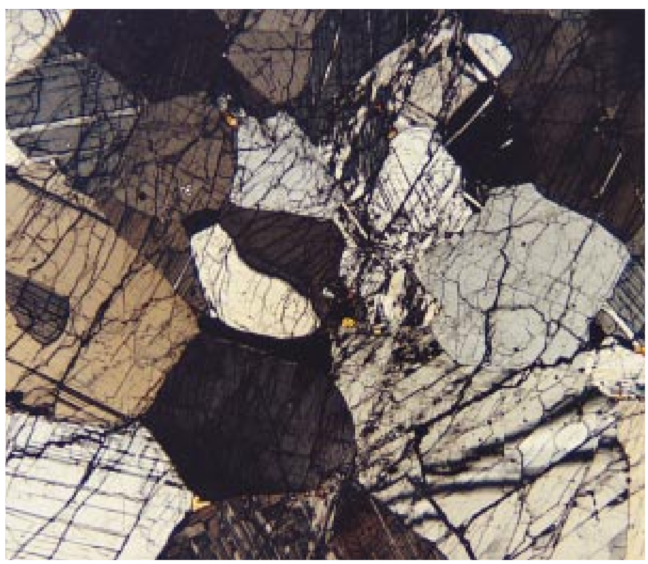
The samples of regolith contain a small percentage of white stones and pebbles composed chiefly of calcium and aluminum silicates known as plagioclase feldspar. Some rocks, called anorthosites, were composed of nothing but such feldspar. John A. Wood of the Harvard-Smithsonian Center for Astrophysics and Joseph V. Smith of the University of Chicago independently suggested that these anomalous particles decorating the *Apollo 11* regolith were tossed there by impacts on distant highland terrain (the light-colored areas of the moon). Thus, they argued, the highlands must be dominated by feldspar-rich rocks. This bold extrapolation was confirmed by *Apollo 16* and other craft that had touched down on the highlands. Remote chemical measurements conducted by the orbiting *Apollo 15* and *Apollo 16* command modules, along with telescopic observations of the highlands by my colleague B. Raymond Hawke of the University of Hawaii and his co-workers, also provided critical evidence.

But that extrapolation was not enough for Wood and Smith. They wondered why the highlands were so rich in plagioclase feldspar. The material might have accumulated on top of magma, like ice cubes floating in a glass of water. Such events happen on the earth in large magma bodies called layered intrusions. The structures form when dense minerals sink and lighter ones rise. Wood and Smith proposed that feldspar floated in a magma "sea," eventually creating the lunar crust. The heavy minerals composed of iron and magnesium silicates—olivine and pyroxene—sank to create the mantle. Moreover, the workers argued, if all the highlands are rich in feldspar, then the magma must have been everywhere, encircling the moon. Thus was born the magma ocean, an idea not even conceived of before the Apollo missions.

Further support for the magma ocean concept comes from a seemingly unconnected group of rocks, the mare basalts. These rocks are rich in olivine and pyroxene—the heavy material that sank in the magma ocean. They erupted onto the surface in the form of lava three billion years ago. Most important, these basalts lack a trace element called europium. The plagioclase feldspar from the highlands, however, is enriched in this element. In fact, the enrichment in the highlands is about equal to the depletion in the mare basalts. The findings bolster the presumption that both the maria and the highlands emerged from the magma ocean. During their formation, the feldspar-rich highlands simply grabbed more of the europium than did the mare basalts.

The presence of a magma ocean prompts a question. How was it created in the first place? Specifically, from where did the energy necessary to liquefy planetary material come? The process of core formation may have supplied some energy: the sinking of metallic iron releases heat. The immense impact that led to the formation of the moon contributed a further boost of energy. Geophysicists have examined the problem in detail and concluded



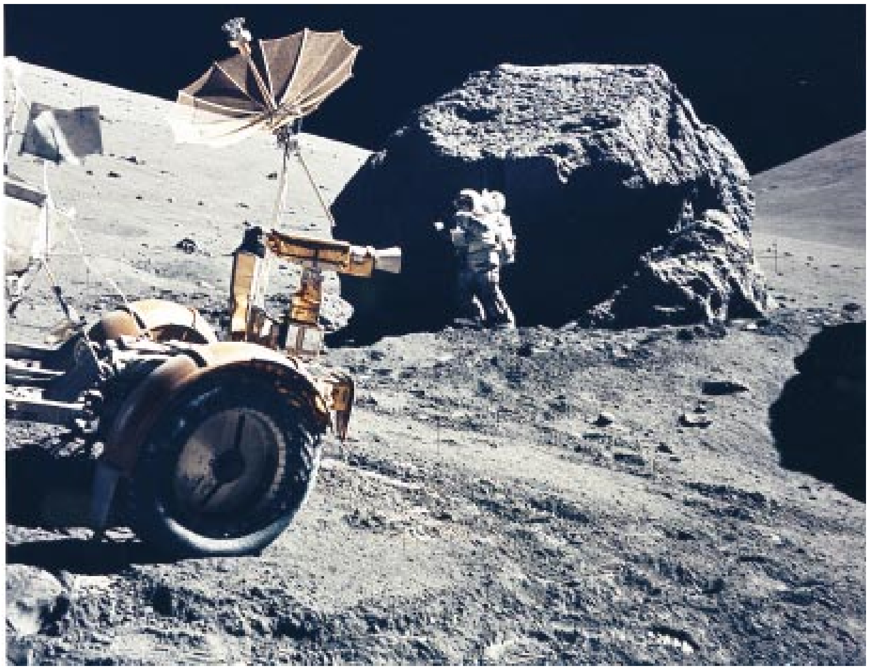


ANORTHOSITE is a type of rock from the lunar highlands (*left*). A slice of such a rock, photographed through polarized light, reveals its composition (*right*). Different compounds polarize light in various ways and, with filters, can be identified by color. All grayish-to-white minerals are feldspar. Tiny orange grains are iron-magnesium minerals.

that the giant impact led to the formation of a large amount of molten material. So great was the melting that up to 65 percent of the projectile and the earth became magma.

The concept of the magma ocean is now being applied to other planets as well. It is changing the way scientists look at the evolution and early history of the solar system. In the laboratory, experimenters try to determine how minerals form in magma and how trace elements are partitioned between the magma and crystallized materials. Another community of researchers investigates processes that could have operated in a magma ocean on the earth 4.5 billion years ago (the active geology of the planet since then has removed all evidence of a magma ocean). I have been marshaling evidence that some asteroids, especially those that formed iron cores, also bore magma oceans early in their histories. All this research was spawned because creative and bold scientists attached special importance to a few dozen little chunks of white rock in a charcoal-gray pile of dirt from the moon.

Although the evidence is compelling, some investigators are still skeptical about the magma ocean theory: they cite the existence of lunar highlands that lack feldspar. The final proof requires a global survey from a lunar orbit of the highlands' crust. The *Clementine* probe, a Department of Defense mission to test advanced sensors, has recently completed its mission to map the moon spectroscopically. (NASA has no planned missions to explore the moon.) Information from *Clementine*'s sensors may provide the crucial data to establish the amount of plagioclase feldspar in the crust.



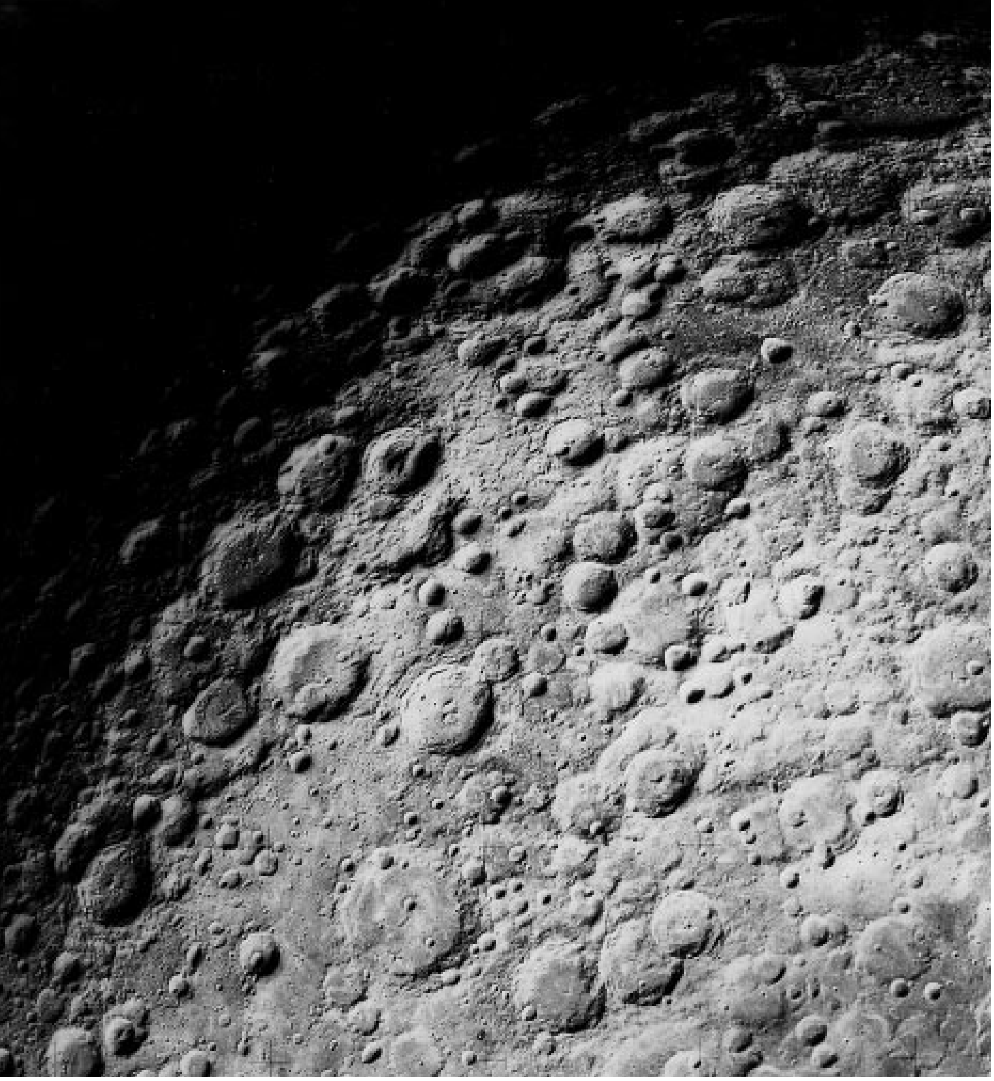
APOLLO 17 **ASTRONAUT Harrison H. Schmitt stands next to a rock that contains materials that had been melted after a huge projectile struck the moon 3.86 billion years ago. The impact created the Serenitatis basin, a 920-kilometer-wide depression. The selection of specific types of samples has helped detail the importance of impact cratering in planetary formation.**

After the giant impact and the structural formation induced by the magma ocean, the moon experienced another step in its evolution: impact cratering. This major geologic process in fact still affects the planets. Its importance was not always appreciated. Before the space age, many scientists claimed that volcanism formed lunar craters and cavities. But as the Apollo missions approached, understanding of impact processes and products increased tremendously. Geologists proved that many circular structures on the earth were formed by collision with an extraterrestrial object. They studied them in order to establish the key characteristics of such features. Others made craters in laboratories, using high-speed guns that propelled projectiles at velocities of several kilometers a second, which smashed into targets.

The first person to assemble strong evidence for an impact origin for lunar craters was the renowned geologist Grove K. Gilbert, whose contributions ranged from basic mapping to hydrogeology. In 1893 he published a classic paper called *The Face of the Moon,* the first geologic study of the earth's satellite. Gilbert correctly identifies the maria as vast lava plains. He also describes the craters and explains why they could not be volcanic. Like Daly's paper about the origin of the moon, this one was also forgotten for decades. In fact, the impact-cratering idea did not return until the early 1940s, when Baldwin began studying the moon. Ironically, Baldwin became aware of Gilbert's important work in a letter from Daly in 1948.

The highlands are especially battered, and the rocks show it. Most samples have been melted, mixed, crushed and compressed by shock waves. These rocks, called breccias, are as complicated as an M. C. Escher print. The ages of highland breccias are surprising. In 1974 Fouad Tera, Dimitri A. Papanastassiou and Gerald J. Wasserburg of the California Institute of Technology pointed out that there are two sharp delineations of ages of highland rocks. The first is at around 4.4 billion years, which Tera and his colleagues took to be the end of primary lunar differentiation (basically, when the magma ocean stopped crystallizing). The other takes place at about 3.9 billion years. This second age, they reasoned, represents a time of intense bombardment that completely wiped out any evidence of previous bombardments; the impacts "reset" the ages of the surface rocks. They dubbed this period of fierce impact bombardment the "lunar cataclysm." The idea states that most of the basins and large craters on the moon formed in a narrow time interval, roughly 3.85 to four billion years ago. Indeed, of the samples dated, the ages of virtually all the rocks from the Apollo program flights and the Russian automated Luna 20 mission are in the 3.85- to 3.95-billion-year range.

Some people did not like the cataclysm idea. Baldwin argued that the apparent clustering of ages was an illusion. The data in effect were contaminated by the widespread distribution of ejected debris. Specifically, the debris originated from the immense event responsible for the formation of the Imbrium basin, a 1,300-kilometer-wide dent that corresponds to the man-in-the-moon's right eye. Baldwin also argued that the elevated parts of large basins have gradually sunk down, indicating they formed before 3.95 billion years ago, perhaps as much as 4.3 billion years ago. Hartmann believes the cluster of ages around 3.9 billion years represented the last of a declining flux of projectiles left over from planetary accretion. The paucity of samples from



CRATERED HIGHLANDS on the far side east of Mare Smythii may have been formed during a geologically brief episode—the lunar cataclysm. This hypothesis supposes that the moon was exposed to an intense bombardment between 3.85 billion and four billion years ago.

before that period was the result of what he called a "stone wall." As older rocks were reheated by impacts, their ages were continuously reset to 3.9 billion years. Therefore, only the last impact events are recorded. Hartmann's and Baldwin's arguments persuaded most investigators. So the possibility that a cataclysm—the dramatic pickup in the impact rate between 3.85 and four billion years ago—was cast aside or at least ignored.

The cataclysm's exile lasted more than a decade, until Graham Ryder of the Lunar and Planetary Institute in Houston vigorously revived the idea in 1990. Ryder makes three points. One is that ages of rocks are not so easily reset. Recent work on the effects of impacts on ages demonstrates that the only materials whose ages are affected are those that melt during the impact and, perhaps, a small percentage of other rocks in the target. Most rocks are crushed up and tossed around but not heated substantially.

Ryder also argues against Hartmann's stone-wall idea. He draws attention to samples of lava flows found among the *Apollo 14* specimens. These rocks range in age from 3.9 to 4.3 billion years. They indicate that the ages of samples can be preserved even though they would have been the most prone to demolition because of their position on the lunar surface.

Ryder's third main contention challenges the notion that all the samples reflect the age of the huge Imbrium basin. Indeed, most workers now feel the idea is probably too simplistic. The highlands harbor many chemically distinct groups of rocks melted by impact, a fact that suggests several collisions. The ages of the rocks cluster between 3.85 and 3.95 billion years.

Still, the question of the exact number of properly dated events remains. So does the issue of what number of impacts constitutes a cataclysm. The differing viewpoints expressed in the cataclysm hypothesis and the stone-wall idea stem in part from two ways of looking at the moon. Ryder advocates cataclysm because he has some confidence in what lunar samples are telling us. Hartmann is more concerned with how planets accrete and thus prefers the stone-wall theory. Reconciling the two perspectives requires more samples from the moon. Especially useful would be specimens from deposits of impact melt inside large basins. Dating them would yield direct, unambiguous ages of each basin from which the samples were collected.

In all likelihood, the bombardment that cratered the moon was not unique. It appears to have occurred throughout the inner solar system. Ancient cratered terrains exist on Mars and Mercury—indeed, Mercury looks much like the moon. (Venus is so active that its earliest features could not have survived.) Numerous projectiles probably hit the early earth as well. The lunar craters can provide some estimate of the size of some of the objects that crashed into the earth. The moon has 35 basins larger than 300 kilometers wide. Even if only half of these basins formed between 3.85 and four billion years ago, the conclusion is inescapable that during the same period the earth would have experienced more than 300 comparable impacts. (The earth is a bigger target, both in terms of cross-sectional area and mass, so it tends to be hit by about 20 times as many projectiles.) Of these, between 15 and 20 would have been monumental, forming basins larger than 2,500 kilometers in diameter. This size, equal to that of the largest lunar basin, is about half the distance across the continental U.S.

Such impacts would have had dramatic consequences. One affected characteristic would have been the earth's geology. Large impacts would have altered any convection patterns in the mantle that may have driven early plate tectonics. They would also have rapidly excavated hot material from the mantle. Brought to the surface, the hot rocks would have melted instantly, producing vast amounts of magma. The residual craters may have collected sediments. Such material may have eroded from the higher areas on the rims of the craters or from the central peaks. The sedimentation could have led to the formation of the first continents.

Life would have had a difficult time thriving during this time. The most dramatic events would have been capable of vaporizing all liquid water on the planet. It seems unlikely that life would have survived anywhere. It would thus have been forced to start over. In fact, Christopher F. Chyba of the NASA Ames Research Center at Moffett Field, Calif., suggests that several sterilization events occurred before the impact rate settled down about 3.8 billion years ago. Only since then would life have been able to take a permanent hold on the earth. Indeed, strong evidence exists that organisms arose by 3.6 billion years ago, only about 200 million years after the bombardment subsided. Recent work indicates that self-replicating molecules can develop quickly, so the 200-million-year period is a reasonable time frame for organisms to emerge.

Large impacts have also been invoked to explain major extinctions on the earth. The impact hypothesis is especially well documented at the Cretaceous-Tertiary boundary, which resulted in the end of half of the living species, including the dinosaurs, 65 million years ago. The main lines of evidence stem from a global enrichment of iridium at the boundary and the presence of shocked forms of quartz and feldspar. Teams spearheaded by Alan R. Hildebrand of the University of Arizona and Virgil L. "Buck" Sharpton of the Lunar and Planetary Institute have identified the site of the probable impact. The crater, called Chicxulub, is completely hidden by sediments forming the Yucatán Peninsula. It was originally discovered in 1981 through gravity surveys and through drilling activities by Pemex, Mexico's national oil company. The structure is 300 kilometers in diameter and 65 million years old.

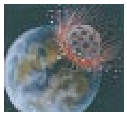
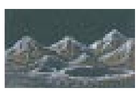
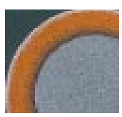
Some scientists have proposed that such mass extinctions are not happenstance events but in fact recur periodically. Proving this hypothesis by looking at features on the earth is not possible. Besides uncertainties in the fossil record, too few terrestrial craters have been dated accurately. Without a correct historical record, a search for periodicity is futile.

The moon's surface may hold the evidence: it is teeming with craters formed during the past 600 million years. Friedrich Hörz of the NASA Johnson Space Center estimates that 5,000 of them are larger than five kilometers in diameter. Even in a local region, say, within a 100-kilometer radius, there are 500 craters wider than one kilometer. But an accurate determination of age requires samples.

Indeed, to fill in missing data about the earth's origin and early history, we must return to the moon. The origin problem seems to be solved, but the details remain sketchy. The existence of the magma ocean is not proved to everyone's satisfaction. We need to determine the overall composition of the moon, which can be done with spectroscopic surveys from orbit and seismographic studies on the surface. More samples from key localities in the lunar highlands would allow us to unravel the processes that operated inside a complicated magma ocean body. The bombardment history of the moon will never be worked out without samples from identifiable impact deposits inside craters. Moreover, new missions do not have to be as expensive or complex as was Apollo. Automated probes can do the job for us.

Of course, the nation and the world may decide they cannot afford to send a fleet of orbital and surface missions to the moon. In that case, we will never know the details of the formation, early melting and bombardment history of the moon and the earth. Only by continuing the legacy of Apollo can we hope to complete our understanding of our place in the solar system.

Apollo's Influence on Lunar Science

TOPIC	PRE-APOLLO VIEW	CURRENT VIEW
Origin 	Captured, derived from the earth or formed with the earth as a dual planet	Giant impact on the earth, followed by formation of the moon from debris
Craters 	Most impact, some volcanic	Almost all impact; dynamics of ejected debris determined
Presence of volatiles (such as water) 	Unknown, although some scientists thought water had flowed on the moon's surface	Mostly dry, but water brought in by impacting comets may be trapped in very cold places at the poles
Rock ages 	Uncertain, but probably ancient (more than a few billion years)	Highlands: most rocks older than 4.1 billion years, with anorthosites 4.4 billion years. Maria: some as young as about two billion years, others as old as 4.3 billion years
Magma ocean 	Not conceived of	Anorthosites formed from magma ocean; other highland rocks formed after that
Composition of maria 	Unknown	Wide variety of basalt types
Composition of highlands 	Unknown	Wide variety of rock types, but all containing more aluminum than do mare basalts
Composition of mantle 	Unknown	Varying amounts of mostly olivine and pyroxene

FURTHER READING

LUNAR SOURCE BOOK: A USER'S GUIDE TO THE MOON. Edited by Grant Heiken, David Vaniman and Bevan M. French. Cambridge University Press, 1991.

TO A ROCKY MOON: A GEOLOGIST'S HISTORY OF LUNAR EXPLORATION. Don E. Wilhelms. University of Arizona Press, 1993.

Synthetic Self-Replicating Molecules

Molecules crafted in the laboratory can make copies of themselves, "mutate," compete for resources and assemble, giving a paradigm for life

by Julius Rebek, Jr.

Imagine a molecule that likes its own shape: finding a copy of itself, it will fit neatly with its twin, forming for a while a complete entity. If the original molecule is presented with the component parts of itself, it will assemble these into additional replicas. The process will continue as long as the supply of components lasts.

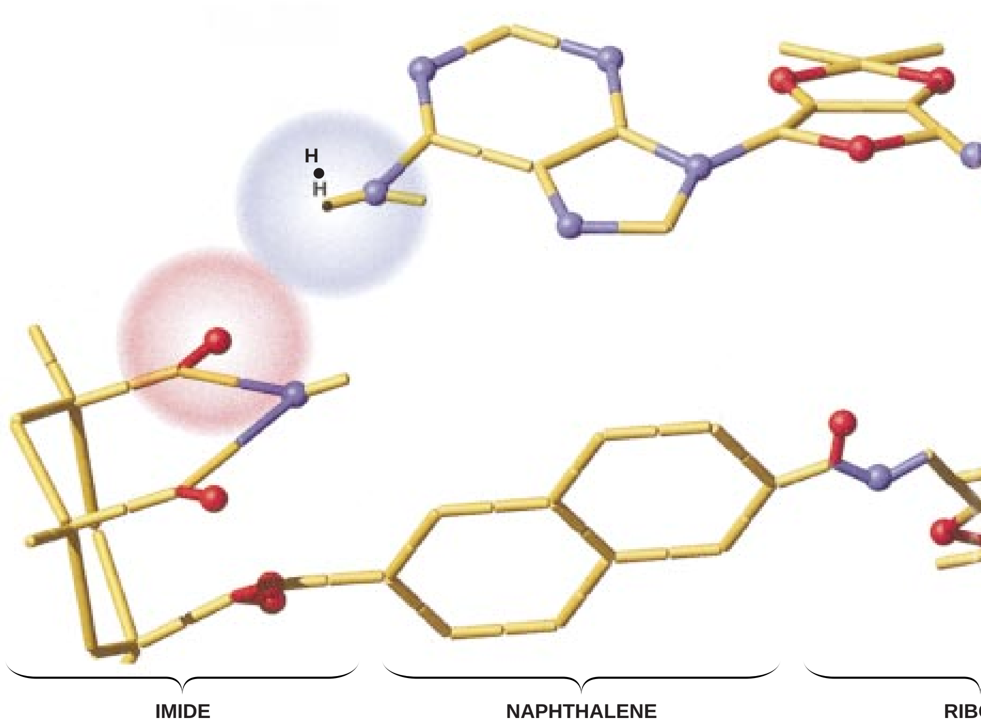
My colleagues and I at the Massachusetts Institute of Technology have designed such self-assembling molecules and crafted them in the laboratory. Our efforts are intended to illuminate the ways in which life might have arisen. Probably it began when molecules came into existence that were capable of reproducing themselves. Our organic molecules, although they operate outside of living systems, help to elucidate some of the essential principles of self-replication.

Attempts to imitate life are still very young compared to the beginnings of life on the earth perhaps three and a half billion years ago. No one can say for sure what the atmospheric or terrestrial conditions were at the time or which molecule crossed the critical frontier between organic chemistry and biology.

In 1953 Stanley L. Miller, then at the University of Chicago, made one of the first—if not the first—attempts to recreate this transition. He formed amino acids in a mixture of water, methane, ammonia and hydrogen—substances thought to have been present on the primitive earth—by subjecting the chemicals to an electric discharge. James D. Watson and Francis Crick's unraveling of the structure of DNA—also in 1953—further fueled this quest.

For some decades now, the most widely accepted recipe for the origin of life has specified DNA or RNA in lukewarm water, neither acidic nor alkaline, plus only those reagents that were presumably part of the primitive earth's atmosphere. Practitioners of prebiotic chemistry such as Miller and Leslie E. Orgel of the Salk Institute for Biological Studies in San Diego have provided deep insights into molecular replication under these constraints. Indeed, countless metric tons of DNA replicate every day in living creatures in precisely this manner, under the surveillance of a host of enzymes.

But recent findings indicate that the primitive earth was perhaps not quite as hospitable as the "warm pond" paradigm would have [see "In the Beginning...," by John Horgan; SCIENTIFIC AMERICAN, February 1991]. Such suspicions, coupled with discoveries of organisms living at (literally) blood-curdling temperatures or near sulfurous volcanic vents at the bottom of the ocean, have led scientists to suggest that life possibly arose by some very different route. Perhaps the only qualities essential to the molecules that become live are detailed in two remark-



ARNI, a self-complementary molecule (*bottom*), gathers the components it needs—an adenine ribose (*top left*) and a naphthalene imide (*top right*)—to replicate. Blue balls indicate nitrogen atoms and red balls, oxygen. Halos surround atoms involved in hydrogen bonding.

able books by Richard Dawkins, *The Blind Watchmaker* and *The Selfish Gene.* Although written 15 years ago, Dawkins's sketches astonishingly prefigure the results of the past four years of my work on self-replication.

Molecules, natural or synthetic, are able to replicate when their shapes and chemistry have a feature called complementarity. By virtue of the way a molecule occupies space and the way its attracting atoms or groups of atoms are distributed along its arms, one molecule may fit snugly into the chemical nooks and crannies of another. The "goodness of fit" between two such complementary molecules thus depends not only on their spatial structure but also on the different kinds of chemical bonds that hold them together in groups. Such groups, or "complexes," form and dissipate rapidly in microseconds or nanoseconds—times that are very short, yet long enough for chemical reactions to take place.

The forces holding complexes together are many times weaker than the covalent bonds binding atoms into molecules. One kind of force—important in complexes—is called a hydrogen bond. This bond comes about when a hydrogen atom possessing a partial positive charge is attracted to, for instance, an oxygen atom that has a partial negative charge. More general attractions of this class go by the name of polar interactions.

Another kind of force, the van der Waals force, is more subtle: if correctly positioned, electrons of one molecule can jostle away those of another, creating a charge imbalance that results in attraction. Yet a third kind of attraction is "aromatic stacking"—an arrangement that flat organic molecules (often having a pleasant odor; hence the name) sometimes assume when they do not like the solvent they find themselves in. By sidling up to one another, flat surface to flat surface, they can squeeze out all the solvent molecules between them and achieve a more stable, stacked configuration.

JULIUS REBEK, JR., was born in Hungary in 1944. After spending the years between 1945 and 1949 in Austria, his family settled in Kansas. Rebek received his bachelor's degree from the University of Kansas and a Ph.D. from the Massachusetts Institute of Technology, where he worked with Daniel S. Kemp on peptide chemistry. As an assistant professor at the University of California, Los Angeles, he developed the three-phase test for reactive intermediates. In 1976 he moved to the University of Pittsburgh and in 1989 back to M.I.T., where he is the Camille Dreyfus Professor of Chemistry. Rebek is a Fellow of the American Academy of Arts and Sciences and was recently elected a member of the National Academy of Sciences.

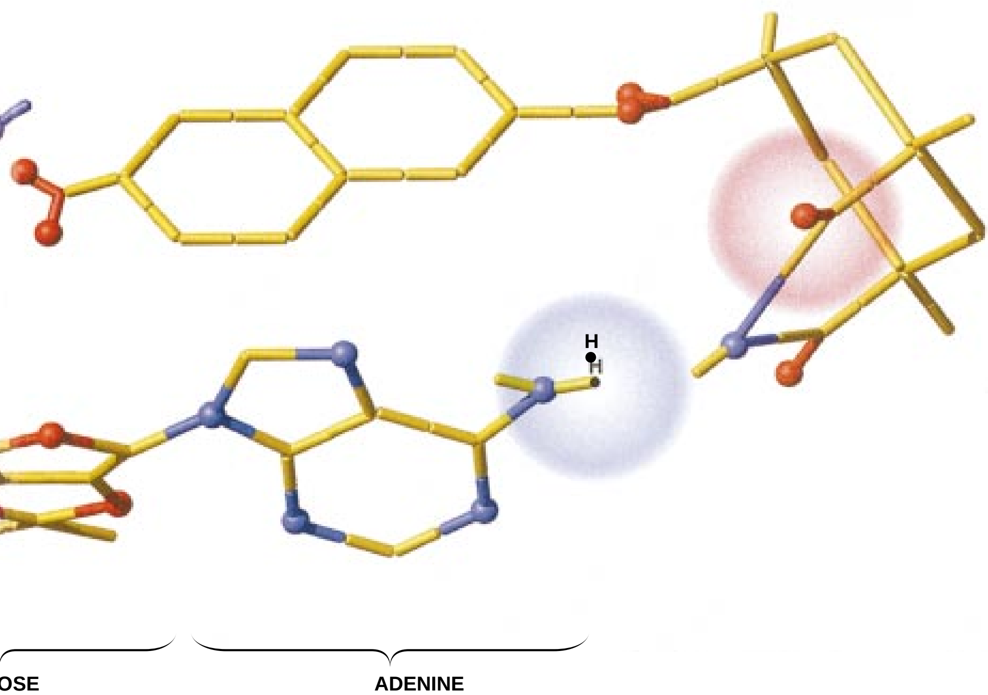
Once a complex forms, the molecular surfaces that match up with one another are relatively protected. Destructive solvents, dissolved acids, bases or oxidants cannot get to them. Strong covalent bonds then have time to join the complementary parts. Sometimes two of three molecules in a complex link together; the third merely serves to ease the process.

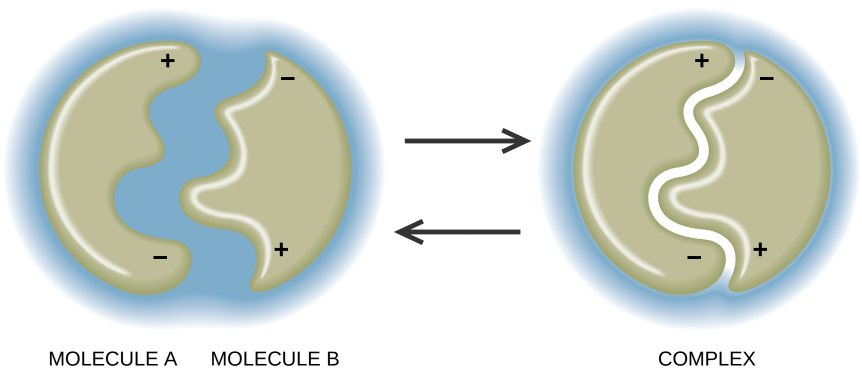
Such a coupling gives rise to a rather popular scheme for replication—the one preferred by DNA. A simple depiction of this scheme uses concave and convex shapes. A concave molecular surface—lined with appropriately enticing atoms—can recognize and surround its convex complement. Further, it can act as a mold for assembling the convex molecule from its component parts. In turn, the convex molecule serves as a template for gathering and fusing the component parts of the concave one. These two replication events—each molecule forming the other—establish what is called a bi-cycle [*see bottom illustration on next page*]. Our recent experiments indicate that a bi-cycle can be extremely efficient.

There is an alternative paradigm of replication: two complementary molecules in a complex can join at some site that is not on the recognition surface. They form a single molecule, one end of which is complementary to the other—and the whole is complementary to itself [*see bottom illustration on page 51*]. The recognition surfaces at the ends of this new, self-complementary molecule are still accessible to other molecules. The ends can each gather a fragment identical to that at the other end.

Once gathered, the two new components cannot move freely and travel through space in tandem; the chances of their becoming linked to each other are greatly enhanced. Thus, the self-complementary entity makes a copy—and in similar manner, many copies—of itself. No enzymes are needed: the molecule catalyzes its own formation.

This is the method we have used in the laboratory to make molecules capable of reacting with one another in ways reminiscent of life. Among them are molecules that bear a passing resemblance to genetic materials—specifically, to nucleic acid components known as adenines. Adenines are flat; besides, they have hydrogen and nitrogen atoms that can form hydrogen bonds with the oxygen and hydrogen atoms of their complementary molecules, called imides. Our imides are con-





MOLECULAR RECOGNITION occurs when two fragments whose geometric and chemical properties complement one another form a complex. The + and – signs indicate electrostatic attractions. Moreover, the solvent is squeezed out between the molecules, helping to stabilize the short-lived complex.

structed from a humpbacked molecule, Kemp's triacid, the skeleton of which folds over in such a way that large, concave structures can easily be fashioned from it. So the imide features a hydrogen bonding site crookedly attached to an aromatic stacking surface; these fit perfectly with the hydrogen bonding site and the flat stacking surface of adenine.

When associated together in a complex, the adenine and the imide become covalently attached, forming a self-complementary molecule. Our early attempts to get this molecule to self-replicate were thwarted by its unforeseen floppiness. Although some flexibility is helpful for molecular recognition—a leather boot is easier to slip on than a wooden one—a lot of flexibility can make fitting very difficult—try slipping on a sock without using your hands.

Molecules become floppy if they have single bonds, involving only two electrons each. Such a bond allows the parts it joins to rotate with respect to each other, giving rise to many different shapes. When Tjama Tjivikua, my graduate student from Namibia, linked the adenine to the imide by a covalent bond, he had to work with a chain of carbon atoms. The chain was so long and flexible that the resulting self-complementary structure doubled over on itself, rather like a jackknife folding shut. So snugly did the adenine fit into the imide that the self-satisfied molecule no longer associated with other molecules or replicated.

Happily, this situation was curable. The remedy called for inserting a larger and more rigid molecule in place of the single chain to prevent folding. Our choice was a larger stacking surface, a naphthalene, bolstered by a less flexible link between the two components, a cyclic ribose group.

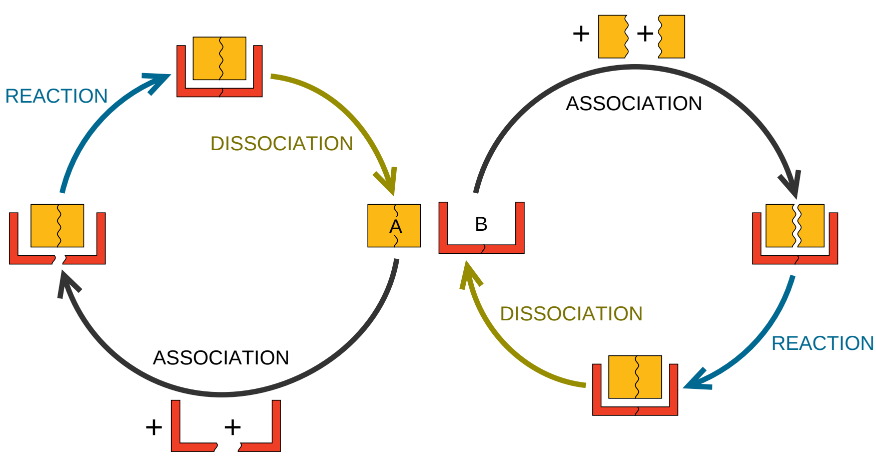
This new J-shaped molecule, adenine ribose naphthalene imide (ARNI for short), provided us with our first instance of replication. Using high-performance liquid chromatography to detect minute changes in chemical concentrations, Tjivikua and Pablo Ballester, a postdoctoral visitor from Majorca, achieved the result. They compared the rate of formation of ARNI in a solution that contained only its components with the rate of formation when some ARNI was added. The presence of ARNI increased the rate of formation, clear evidence of the presence of a self-replicating system.

If one plots the progress of a reaction through time, one generally derives a curve that assumes the shape of a reclining parabola. The product forms fastest at the beginning, when the reactants are at their highest concentrations; the rate of formation slows down as the reactants are consumed. For an autocatalytic reaction—one in which the product, like our ARNI, catalyzes its own formation—the growth curve should be S-shaped, or "sigmoidal" [*see top illustration on opposite page*]. The reaction begins slowly. As the product appears and begins to act as a catalyst, the reaction accelerates. An upward curve results. Finally, as the materials are consumed, the reaction grinds to a halt.

The degree of sigmoidal curvature depends on several factors, the most important of which is the efficiency of the autocatalytic step. If the background reaction—in which the components combine by themselves, without getting help from the self-replicating molecule—is too strong, it can swamp the signal from the self-replicating process. In 1990 Günther von Kiedrowski and his co-workers in Göttingen showed that a self-replicating nucleic acid could exhibit such sigmoidal growth—proving that the autocatalyzed synthesis is in this case more efficient than the random one.

Although ARNI did not show sigmoidal growth, our next attempt, ARBI, did. We slowed the background reaction rates by giving ARBI a slightly longer stacking element, a biphenyl instead of naphthalene. We now had proof of a bona fide synthetic self-replicating molecule.

Is it alive? Not by most current defi-



REPLICATION BI-CYCLE involves two molecules of complementary shapes, represented by block A and sleeve B, into which it fits. In the left cycle, the block (*middle*) collects the two parts of the sleeve (*bottom*) around it to form a complex (*left*); the parts then react to form a whole sleeve (*top*). The block and sleeve quickly dissociate. In the right cycle, it is the sleeve that assembles the fragments of the block. Thus, the two complementary molecules catalyze each other's formation.

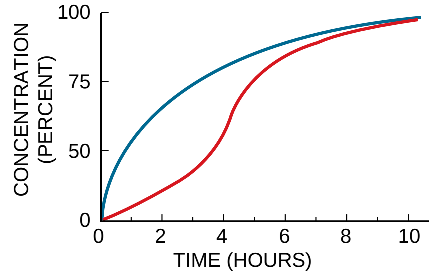
nitions. Our (or its) critics were quick to point out that as a life-form, ARBI had severe limitations: the molecule would make copies only of itself. To allow evolution, a self-replicating molecule has to be capable of "making mistakes": occasionally synthesizing other molecules that can perhaps be better replicators. Unlike art and music critics, those in science at least indicate in which direction improvements may lie. We responded by devising molecules that were capable of making—indeed, that were incapable of not making—mistakes.

In organic chemistry, a "mistake" is made when there is a lack of selectivity between reaction partners. We needed a molecule that would catalyze not only its own formation but also that of a molecule of similar shape. Besides, at least one of these two molecules had to be able to change into a more efficient replicator.

A molecule can readily be manipulated into making replicas of a competitor. Instilling the capacity to "evolve" takes more planning. In the course of our search for a solution, we tapped into some earlier findings on the hydrogen bonding sites of adenine. There are two ways in which an imide can attach to an adenine. It can find a site along the Watson-Crick edge, which is involved in the replication of DNA, or it can dock along the Hoogsteen edge, a region normally exposed in DNA (though sometimes joined in such exotic forms as triple helices).

We had already shown that simple adenines can attach to our imides along either edge. For example, roughly equal amounts of Watson-Crick and Hoogsteen complexes are formed with an imide attached to a naphthalene surface. But if one of the hydrogens of the amino (NH_2) group of adenine—involved in hydrogen bonding—is replaced by a larger group of molecules, the situation changes. This new group positions itself in such a way as to block access to the Watson-Crick edge

S CURVE (*red*) is the signature of an autocatalytic reaction when the concentration of the reaction product is plotted against time. At first, the molecule forms slowly. As it catalyzes its own synthesis, the reaction suddenly accelerates, only to slow down as the reactants are consumed. Conventional reactions (*blue*) simply have a parabolic shape.



while leaving the Hoogsteen edge largely open. For example, when a small methyl group joins the adenine, more than 85 percent of our synthetic imide receptors are found to bind along the Hoogsteen edge.

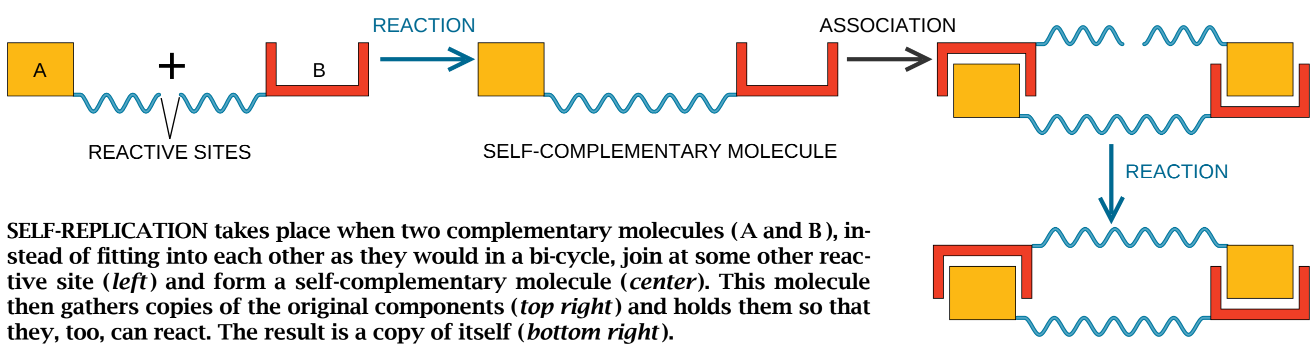
We decided to exploit the change in the rate of replication that comes from blocking the Watson-Crick edge. Accordingly, we prepared two different adenines, one bearing a benzyloxycarbonyl, or "Z," group (a popular blocking group in protein synthesis) and another bearing a Z group with an additional nitrogen group: Z-NO_2. The plan was this: an altered adenine and the imide would assemble on the product template as before. But the blocking groups would be dangling off the molecule in sites quite far from where the covalent bond forms. The Z group at one end would not know if the blocking group at the other end was a Z or Z-NO_2; synthesis should take place regardless of the groups' identities.

Moreover, blockade of the Watson-Crick edge would limit hydrogen bonding to the Hoogsteen edge. Thus, the modified replicators ZARBI and ZNARBI—made from adenines containing Z and Z-NO_2, respectively—would have to replicate slowly. At this point, the choice of blocking group becomes critical. Although many attached groups can lead to the formation of replicators that make mistakes, the nitro group (NO_2) is somewhat special. Investigators have known for 30 years that they can remove the group easily by irradiating it with particular wavelengths of ultraviolet light. Once the Z-NO_2 is excised, the Watson-Crick edge is freed. The new, lighter molecule can now match up along this edge as well as the Hoogsteen edge, making it, we hoped, twice as efficient as the molecules from which it is derived.

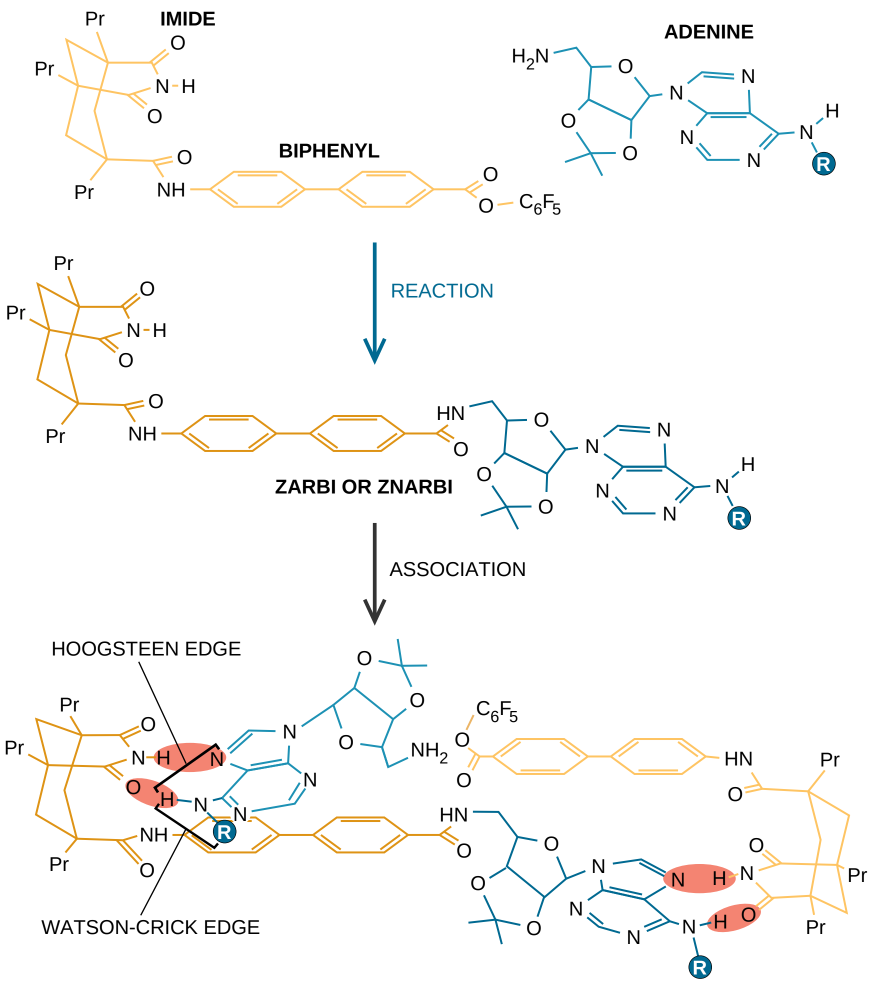
We subjected our adenines bearing Z or Z-NO_2 groups to our now standard tests for self-replication. When the adenines joined with the biphenyl receptors to produce ZARBI and ZNARBI, the new molecules did indeed behave as replicators—though admittedly not very efficient ones. The replicators compensated for their clumsiness with their ability to make mistakes. ZARBI would catalyze its own formation and also act as a template for assembling its rival, ZNARBI. The latter reciprocated, catalyzing its own formation and also that of ZARBI.

Now came the challenge of demonstrating a chemical version of mutation: a permanent, heritable structural change affecting the survival capacity of an organism—or in our case its analogue, a self-replicating molecule. Changes in a molecular replicator's structure might be caused by shifts in temperature, acidity, salinity or many other factors. We chose irradiation with light.

We first placed ZARBI and ZNARBI in direct competition for reproductive resources. Qing Feng, then my graduate student, and Jong-In Hong, a postdoctoral fellow, allowed the adenine derivatives bearing the Z and Z-NO_2 groups to compete for a limited amount of the complementary biphenyl receptor.



SELF-REPLICATION takes place when two complementary molecules (A and B), instead of fitting into each other as they would in a bi-cycle, join at some other reactive site (*left*) and form a self-complementary molecule (*center*). This molecule then gathers copies of the original components (*top right*) and holds them so that they, too, can react. The result is a copy of itself (*bottom right*).



"MUTANT MOLECULES" are produced when an adenine with an extra group R joins with a biphenyl imide to form a self-complementary molecule. The R can be a Z (benzyloxycarbonyl) or a Z-NO_2 group, giving, respectively, a ZARBI or a ZNARBI molecule. The latter can collect an adenine along the Hoogsteen edge (the R blocks the Watson-Crick edge) and fuse it with a biphenyl imide, catalyzing its own formation and also that of its competitor. The red ovals indicate hydrogen bonding.

ZNARBI proved to be a slightly better replicator. When all the receptor was used up, we irradiated the reaction vessel with ultraviolet light having a wavelength of 350 nanometers. After a few hours of irradiation, the Z-NO_2 blocking groups had all been removed from both the ZNARBI replicators as well as from the adenine progenitors. That is, the ZNARBI molecules had all been converted to ARBIs and the Z-NO_2-bearing adenines to plain adenines. A "mutation" had occurred, prompted by a change in the environment. ZARBI and the Z-adenine remained unaltered.

Next we added more biphenyl receptor. ZARBI found the radiation product, ARBI, to be its competitor. The sleeker ARBI, having the advantage of replicating in either the Hoogsteen or the Watson-Crick mode, rapidly took over the resources of the system.

A simple evolutionary interpretation can be sketched for this experiment. Think of ZARBI as the original in this sequence; its replication requires the presence of Z-adenine and the biphenyl receptor. If nitric acid is added, then some of the Z-adenines are used up to form Z-NO_2 adenines; the latter give rise to ZNARBIs. ZNARBI is a better replicator compared with its ancestor ZARBI. When the system is irradiated, a second change takes place. ZNARBI converts to the simpler and more efficient ARBI. This last molecule proves to be the best replicator of the three.

Although mutation is considered to be the driving force for most evolutionary alterations, another significant paradigm for change is recombination. Two chromosomes can split, exchange strings of DNA and rejoin, thus combining their characteristics. Also, certain computer programs attempt to "teach" strings of information to solve a problem. If the strings are allowed to split and recombine at random, they soon give rise to much better problem solvers. Mutation allows for single, small changes; recombination, on the other hand, allows the creation of hybrids that are very different from the progenitors.

Our interest in demonstrating recombination at the molecular level led us to develop an entirely new set of self-replicating molecules. The principle was the same: two complementary molecules were joined by a covalent bond to give a single, self-complementary whole that could aid its own synthesis. Feng and another student, Tae Kwo Park, devised a replicating system based on a different component of nucleic acids, thymine. Some time earlier Park had developed a synthetic receptor that would recognize thymine's imide nucleus and also lie on thymine's flat aromatic surface. This receptor featured a U-shaped molecular skeleton. The bottom of the U was a large, rigid aromatic spacer known as xanthene; one arm of the U featured an amine and the other arm a diaminotriazine, the receptor for thymine. When the latter two became joined by a covalent bond, a self-complementary unit was generated, called diaminotriazine xanthene thymine, or DIXT. We were able to show that DIXT was also self-replicating.

The stage was now set for a recombination experiment. Could the adenine-based replicators and the thymine-based replicators, when placed in the same vessel, shuffle their components into new combinations? They did indeed. Even so, we were surprised by the results. One of the new recombinants, ART (adenine ribose thymine), was the most prolific replicator we had yet encountered, whereas the other one, DIXBI (diaminotriazine xanthene biphenyl imide), was unable to replicate at all—it was "sterile."

How did this difference in the ability to replicate come about? The efficiency of the ART replicator is easily rationalized. ART looks a good deal like a piece of DNA, possibly the best replicator in existence. Furthermore, its ribose piece makes the recognition surfaces parallel to one another, a very helpful configuration. This and the high affinity of adenine for its complement thymine make for an easily assembled complex—the intermediate stage in replication.

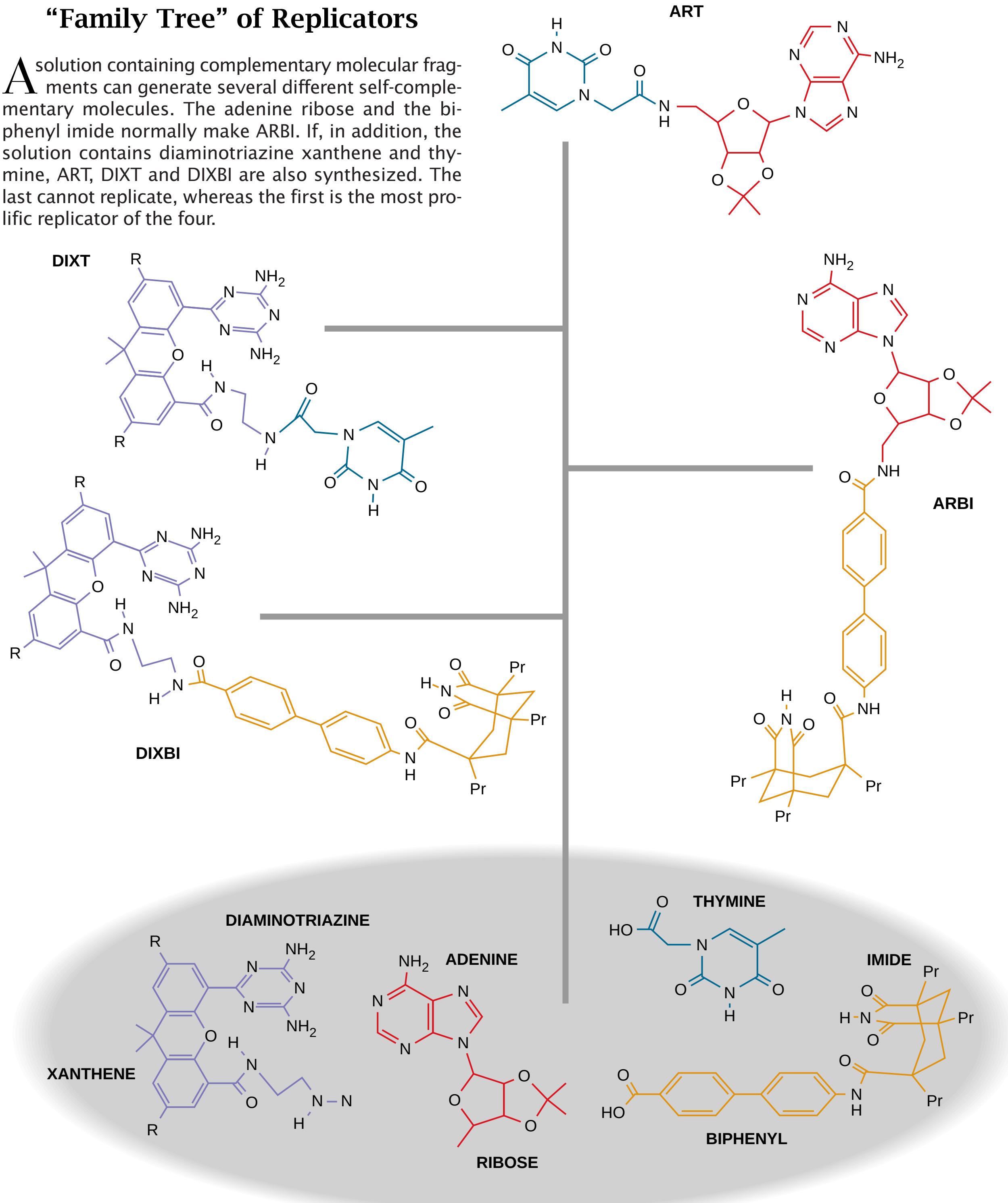
The inefficiency of DIXBI can also be traced to its overall molecular shape.

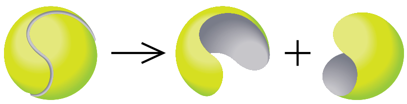
DIXBI is composed of two U-shaped molecules connected by a rigid biphenyl spacer; its overall structure can adopt a C or an S shape. The recognition surfaces are exposed inside the C shape, where there is not enough room for a replicating complex to form. In the S shape the recognition surfaces are far apart, so that when a complex forms, the reactive pieces are too far away from one another to bond covalently. Thus, even though DIXBI is self-complementary, it cannot achieve replication.

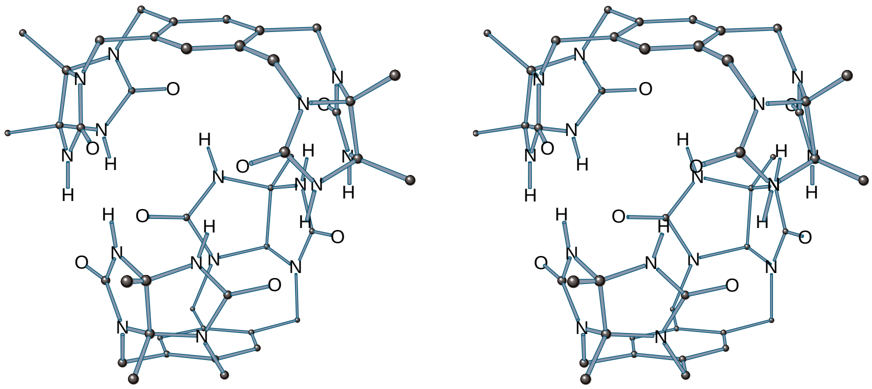
With this experiment we were able to show that a relatively small pool of components can give rise to a "family tree" of replicators. Three of these are effective at self-replication, but one branch of the tree dies out. To push this analogy further, it would be appropriate for the sterile molecule to be

"Family Tree" of Replicators

A solution containing complementary molecular fragments can generate several different self-complementary molecules. The adenine ribose and the biphenyl imide normally make ARBI. If, in addition, the solution contains diaminotriazine xanthene and thymine, ART, DIXT and DIXBI are also synthesized. The last cannot replicate, whereas the first is the most prolific replicator of the four.







TENNIS BALL cut along its seam yields two self-complementary shapes that inspire a design for a cell wall. To the right is a molecule that can assemble with its twin into a hollow sphere. The drawings are stereoscopic; if you can cross your eyes enough to overlap the images, you can see a three-dimensional molecule.

chopped up and converted into pieces that the effective replicators could use for themselves. We have made some progress in this direction. It requires equipping our molecules with acids and bases that can manipulate other molecules more actively than simple recognition will allow.

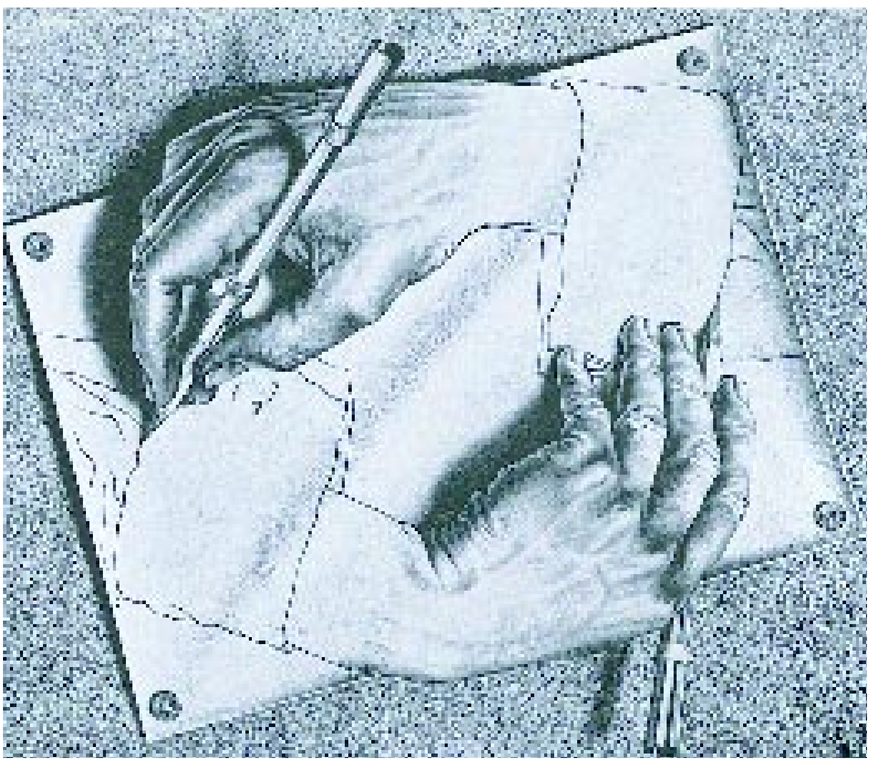
Although it has been enjoyable to pursue replication and even evolution with synthetic molecules, we have been looking to the next step in expressing life as a series of molecular reactions. We feel, as do other workers, that a key attribute of life is a boundary: a container or a cell wall that separates inside from outside and prevents desirable molecules from floating away—while keeping undesirable ones at bay.

Inspired by a naturally occurring membrane, we have made some small, initial steps toward this goal. Viruses use a protein shell as a container; the shell is made up of many identical copies of a single protein unit. The units are also self-complementary—but the recognition surfaces are oriented so that they automatically assemble into a closed shell. Indeed, Crick had predicted that many identical copies of proteins would compose the viral coat, since there is not enough information in the viral genome for many different molecules to be involved.

When we used self-complementarity as our guide, a minimalist design struck us, based on the structure of a tennis ball. Cut along its seam, a tennis ball gives two identical pieces, the convex ends of which are complementary in shape to the concave middles. René Wyler, a Swiss postdoctoral fellow, has now synthesized a structure that mimics the shape of the tennis-ball pieces, while adding chemical complementarity. The units fit together with hydrogen bonds along the seam.

There are good indications that a smaller molecule, such as a solvent chloroform molecule, can fit within our molecular tennis ball. But it is too small to accommodate even our most minimal replicators. We are now working with Javier de Mendoza of the Free University of Madrid to develop a larger molecule—a softball—that may have an interior roomy enough to hold some of our replicators.

Once we have made it past the problem of containment, the biggest obstacle to the molecular life agenda will be these questions: How can our fledgling organism harness energy? From sunlight or from other molecules? How can the component pieces of the replicators and their containers be replenished? These are the challenges of the next decade. Whether they are met or not, the efforts of chemists to answer them will surely provide insight into the organic chemistry of life—how it came about and how it continues to flourish.



***DRAWING HANDS,* a 1948 lithograph by M. C. Escher, illustrates the principles of self-complementarity and self-replication.**

FURTHER READING

CHEMICAL EVOLUTION: ORIGIN OF THE ELEMENTS, MOLECULES, AND LIVING SYSTEMS. Stephen F. Mason. Clarendon Press, 1991.

MOLECULAR REPLICATION. Leslie Orgel in *Nature,* Vol. 358, No. 6383, pages 203–209; July 16, 1992.

A TEMPLATE FOR LIFE. Julius Rebek in *Chemistry in Britain,* Vol. 30, No. 4, pages 286–290; April 1994.

Barriers to Drug Delivery in Solid Tumors

Many tumors resist full penetration by anticancer agents. Such resistance may help explain why drugs that eradicate tumor cells in laboratory dishes often fail to eliminate malignancies in the body

by Rakesh K. Jain

An agent that destroys cancer cells in a culture dish should, in theory, be able to kill such cells in the body. Certain drugs that display potent anticancer activity in the laboratory have indeed saved the lives of many patients, particularly those suffering from various pediatric cancers or malignancies of the blood. Sadly, however, the existing pharmacopoeia has not markedly reduced the number of deaths caused by the most common solid tumors in adults, among them cancers of the lung, breast, colon, rectum, prostate and brain.

No single process is likely to explain the disappointing results. Nevertheless, recent research indicates that a much overlooked property of tumors—namely, resistance to penetration by drugs—can play a significant role in undermining therapy. After cancer-fighting drugs are delivered orally or by injection, they travel via the bloodstream to their targets. To eradicate tumors, the agents must then disperse throughout the growths in concentrations high enough to eliminate every deadly cell. Yet studies my colleagues and I have conducted during the past 20 years show that solid cancers frequently impose formidable barriers to such dispersion. Investigation of those barriers is now generating exciting ideas for overcoming them.

Success in combating the obstacles should enhance many different approaches to therapy, all of which have as a common denominator that their effectiveness depends on optimal accumulation of therapeutic agents in tumors. Whenever possible, physicians treat solid cancers by surgical removal or by irradiation. If, however, parts of an original (primary) mass cannot be excised or if the tumor is thought to have metastasized, doctors may rely on systemic drug therapy to eliminate any remaining cancerous tissue.

Systemic treatment typically consists of chemotherapeutic drugs that are toxic to dividing cells (including, unfortunately, healthy ones). More recently, genetic engineering and other technologies have provided a second class of drugs—proteins and other biological products. This class encompasses several molecules of the immune system, such as tumor necrosis factor, interleukins, interferons and monoclonal antibodies. It also includes white blood cells known as lymphokine-activated killer (LAK) cells and tumor-infiltrating lymphocytes, as well as various agents designed to carry out gene therapy.

Radiation therapy may similarly incorporate delivery of blood-borne drugs. Radiation works in part by converting oxygen molecules into highly destructive forms called free radicals. Yet tumors are frequently oxygen deficient. In an attempt to increase the vulnerability of malignancies to the treatment, investigators are testing the value of administering sensitizing agents that mimic oxygen or somehow elevate oxygen levels in a tumor. Malignancies can also be killed by heating, and so hyperthermic therapies sometimes involve drugs that increase a tumor's response to heat. In photodynamic therapy, a compound that is relatively harmless until it is exposed to laser light is injected and given time to collect in a tumor. Then light is focused on the mass.

I was a graduate student in chemical engineering at the University of Delaware when I first became intrigued by the possibility that tumors might resist invasion by drugs. In 1974 James Wei, my Ph.D. adviser, arranged for me to assist Pietro M. Gullino of the National Cancer Institute in measuring the uptake of drugs by malignancies in animals. As I learned more and more about cancer, it became obvious to me that the failure of blood-borne agents to distribute throughout tumors could readily undermine drug therapy. I began to wonder whether solid tumors possessed features that blocked such distribution. I also began to think that my engineering background, including my understanding of fluid and molecular transport, could aid in exploring that possibility.

The likelihood that tumors could impede drug penetration was suggested partly by their structure. Contrary to popular perception, malignant growths are not merely clusters of proliferating

RAKESH K. JAIN, who was born in India, is Andrew Werk Cook Professor of Tumor Biology in the department of radiation oncology at Harvard Medical School. He is also professor in the Harvard University–Massachusetts Institute of Technology division of health sciences and technology and director of the Edwin L. Steele Laboratory at Massachusetts General Hospital. Before moving to Boston, he was a professor of chemical engineering at Carnegie Mellon University for more than a decade. Jain's research uniquely combines engineering science with tumor biology. He has received many awards for his contributions to the understanding of tumor pathophysiology, recently including a National Cancer Institute Outstanding Investigator Grant.

IDEALIZED SOLID TUMOR has been partly cut away to reveal some of its blood vessels. Before a blood-borne drug can begin to attack malignant cells in a tumor, it must accomplish three critical tasks (*detail*). It has to make its way into a microscopic blood vessel lying near malignant cells in the tumor (*1*), exit from the vessel into the surrounding matrix (the interstitium) (*2*) and, finally, migrate through the matrix to the cells (*3*). Unfortunately, tumors often develop in ways that hinder each of these steps.

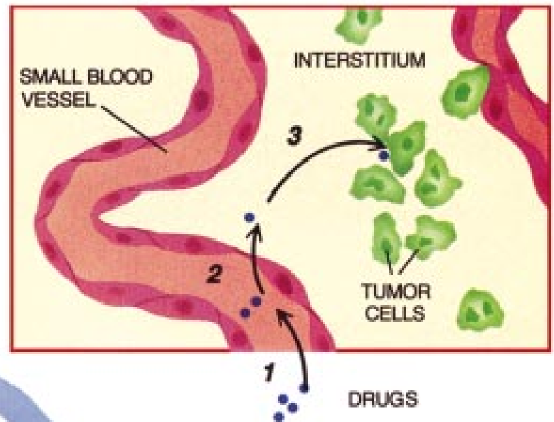
cells. Cancer cells often occupy less than half the volume of a tumor. One to 10 percent of the volume is contributed by blood vessels weaving through the tumor mass. The remaining space is filled primarily by an abundant collagen-rich matrix—the interstitium—that surrounds cancer cells and can separate them from the vasculature. (Healthy tissue contains an extracellular matrix as well, but the interstitium in tumors is usually more extensive.)

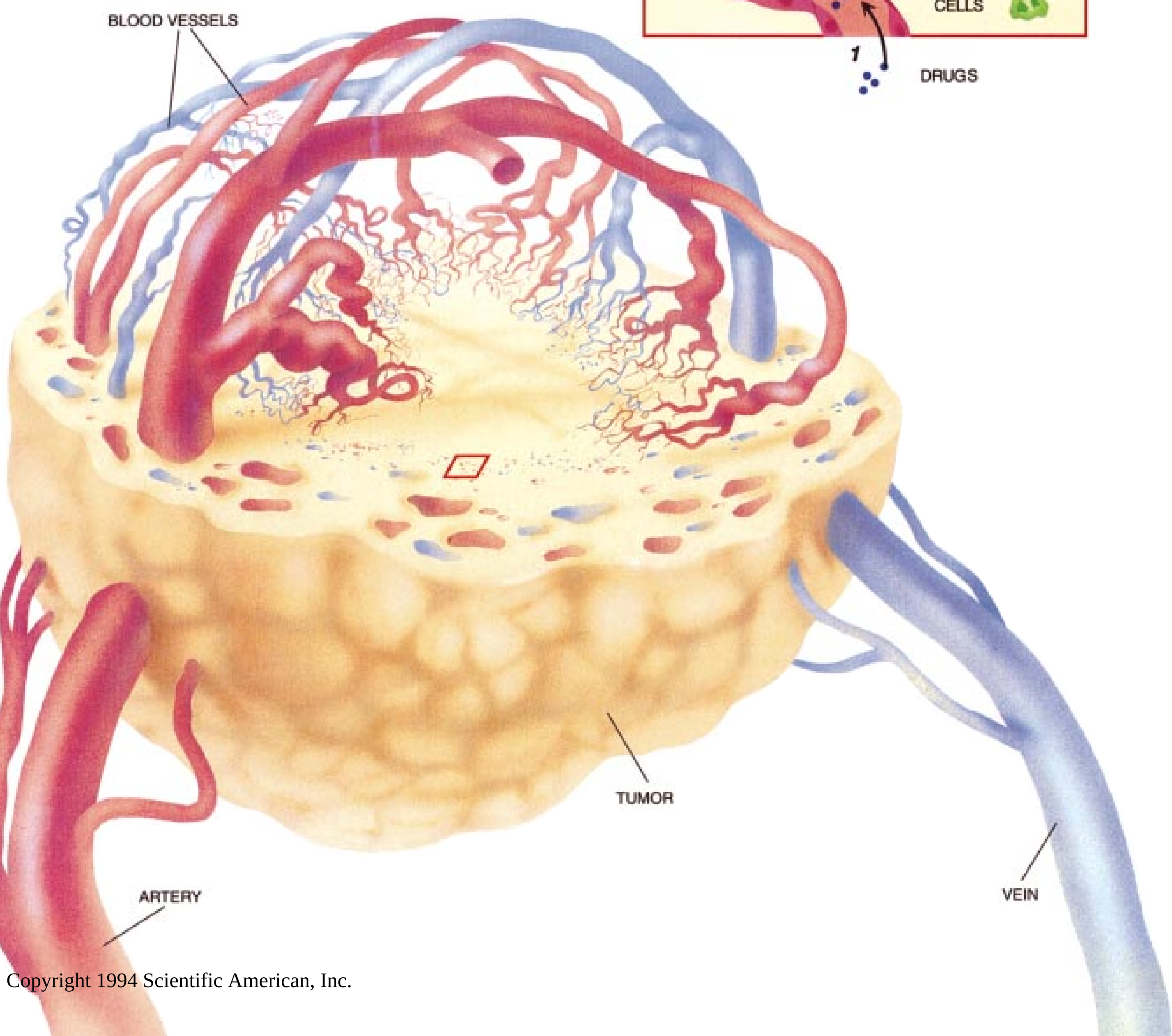
To reach cancer cells in a tumor, then, a therapeutic agent must make its way into the blood vessels of the tumor and across the vessel wall into the interstitium. Finally, it must travel, often great distances, through that matrix to the cells. Accomplishing any of those steps, I reasoned, could be problematic.

I was unable to pursue this idea when I obtained my first faculty position, but my opportunity came in 1978, the year I moved to Carnegie Mellon University. My colleagues and I—a multidisciplinary group—have been studying barriers to drug penetration ever since, most recently at Harvard University and at Massachusetts General Hospital, which I joined in 1991.

We apply several methods in our work. For example, at Carnegie Mellon, we immediately adopted a method introduced by Gullino in 1961. In this approach we grow tumors in rodents so that each cancerous mass is connected to the circulatory system by a single artery and a single vein. That arrangement enables us to measure how much drug flows into and out of a tumor. From such information, the amount of drug that is retained can be calculated. We also developed a related procedure that permits the study of colon cancers arising spontaneously in humans. When surgeons remove tumors from patients, they can occasionally provide us with individual nodules fed by one artery and drained by one vein. We then maintain the circulation artificially.

Although these approaches yield valuable information, the inner workings of the tumors remain something of a black box. To obtain detailed insight





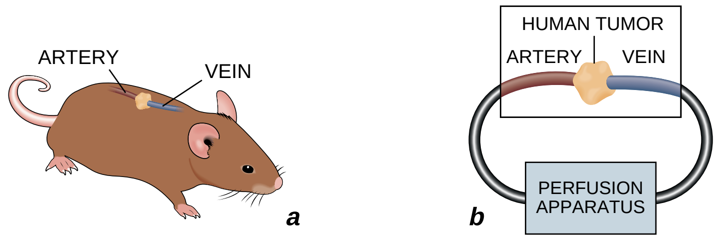
How Drug Delivery Is Studied Experimentally

The author's group applies two main experimental strategies for determining the fate of drugs in tumors. In "isolated tumor" techniques the workers implant into a rodent a tumor that is fed by a single artery and drained by a single vein (*a*), or they occasionally obtain such a tumor from a human patient and keep blood flowing artificially (*b*). Then they measure the amount of drug entering and leaving the tumor and calculate the amount absorbed.

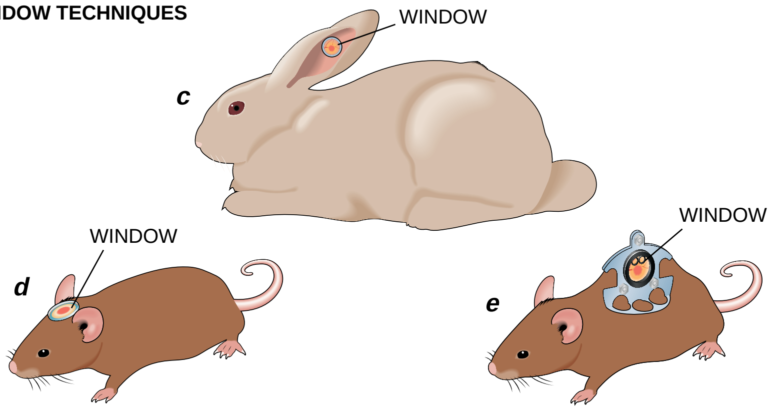
In "window" techniques the investigators grow a tumor in the ear of a rabbit (*c*) or on the brain (*d*) or dorsal skin (*e*) of a rodent and put a glass cover slip on the tumor. By focusing a microscope on the visible tissue, they can directly observe such phenomena as the development of new blood vessels and the spread of a drug through a tumor.

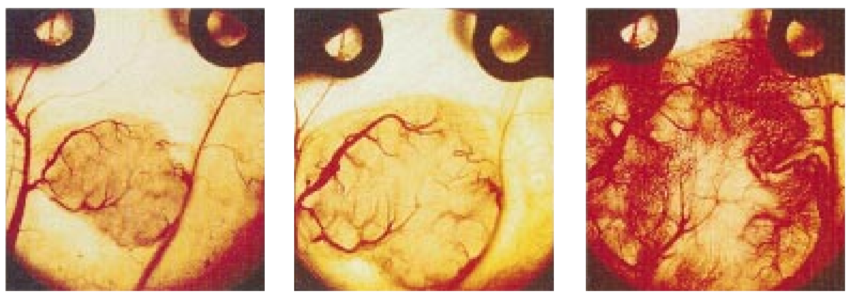
The photographs track changes that occurred between the fifth day and 20th day (*left to right*) after cancer cells were implanted in the dorsal skin of a rodent. By day 20, the periphery of the resulting tumor had gained a tangle of blood vessels, but the center had lost much of its blood supply (*white area*). The tumor thus lost vessels needed for bringing blood-borne drugs directly to the central region.

ISOLATED-TUMOR TECHNIQUES



WINDOW TECHNIQUES





into the flow of blood and the distribution of drugs and other substances in a tumor, we apply modified versions of "window" techniques that were introduced for use in rabbits by J. Calvin Sandison of the University of Pennsylvania in the 1920s and for use in mice by Glenn H. Algire of the National Cancer Institute in the 1940s.

We implant tumor cells in the ear of a rabbit or on the brain or back skin of a rodent. Then we cover them with a transparent glass cover slip or sandwich them between two cover slips. As the resulting tumor grows against the glass, we can observe it under a microscope. By attaching fluorescent labels to injected agents, we can even trace the passage of various substances through the tumor. Since 1991 our access to immunodeficient mice has enabled us to view human tumors, not merely those of animal origin. (The lack of an immune system prevents the mice from rejecting the human grafts.) Mathematical modeling, a classic engineering strategy, aids our investigations as well. It allows us to combine theory with experiment, to formulate and test predictions, and to minimize the number of animals in our experiments.

Early on, our work and that of others revealed that the vascular system of tumors can be highly disorganized, both in its structure and its operation. This disorganization, in turn, can form one important barrier to drug delivery. In normal organs, blood vessels are arrayed predictably and provide blood to all areas of the constituent tissue. Arteries delivering oxygenated blood from the heart divide into smaller arterioles and then into microscopic capillaries. From the capillaries, fluid, nutrients and oxygen pass into the surrounding matrix and cells. The capillaries feed into venules, which take up wastes and excess fluid from the tissue and deliver them to veins for removal.

Although tumors initially obtain blood from the existing vasculature in the region, they eventually produce new small blood vessels, which branch excessively, twist into tortuous shapes and grow in unpredictable directions that can change from day to day. Consequently, some areas of the tumor may be well vascularized, whereas others have little or no blood supply.

These findings thus indicated that one of the first problems a blood-borne drug encounters en route to cancer cells is an uneven distribution of blood vessels. Indeed, through our transparent windows we can see that regions lacking blood vessels receive no drug directly from the circulation. (Tumor cells in those blood-starved areas may seem, on superficial inspection, to be dead, but they frequently revive if nourishment returns.)

What is more, the aberrant branching and twisting of the vasculature often contribute to an observable slowing of blood flow—a phenomenon that is exacerbated by the unusual viscosity of the blood in tumors. The slowed flow hinders delivery of drugs to poorly perfused regions of the tumor. It also participates in causing other drug-delivery problems, as will be seen.

The nonuniform blood supply is by no means the only obstacle to the spread of a drug in a tumor. A second impediment takes the form of abnormally high pressure in the interstitial matrix (as measured by the force the matrix exerts on a probe inserted into it). The pressure can retard the passage of large molecules across vessel walls into the interstitial matrix. It can thus contribute to the low concentration of drug molecules frequently seen (by fluorescence microscopy or other imaging techniques) in the interstitial matrix of animal and human tumors growing in mice.

We began to suspect that interstitial pressure might pose problems for drug delivery when we took a close look at the forces that control the movement of molecules from the blood into the matrix. Such passage occurs across or between endothelial cells, which line vessel walls in a single layer. We knew that molecules leave blood vessels (extravasate) primarily by two mechanisms: diffusion and convection. (Cells, which I shall discuss separately, use a third option as well.) Diffusion is the movement of molecules from an area of high concentration to an area of lower concentration. Convection is the transport of molecules by a stream of flowing fluid. Unlike diffusion, which is unaffected by pressure gradients, convection is governed by them: fluid flows from areas of high pressure to areas of low pressure, carrying molecules with it.

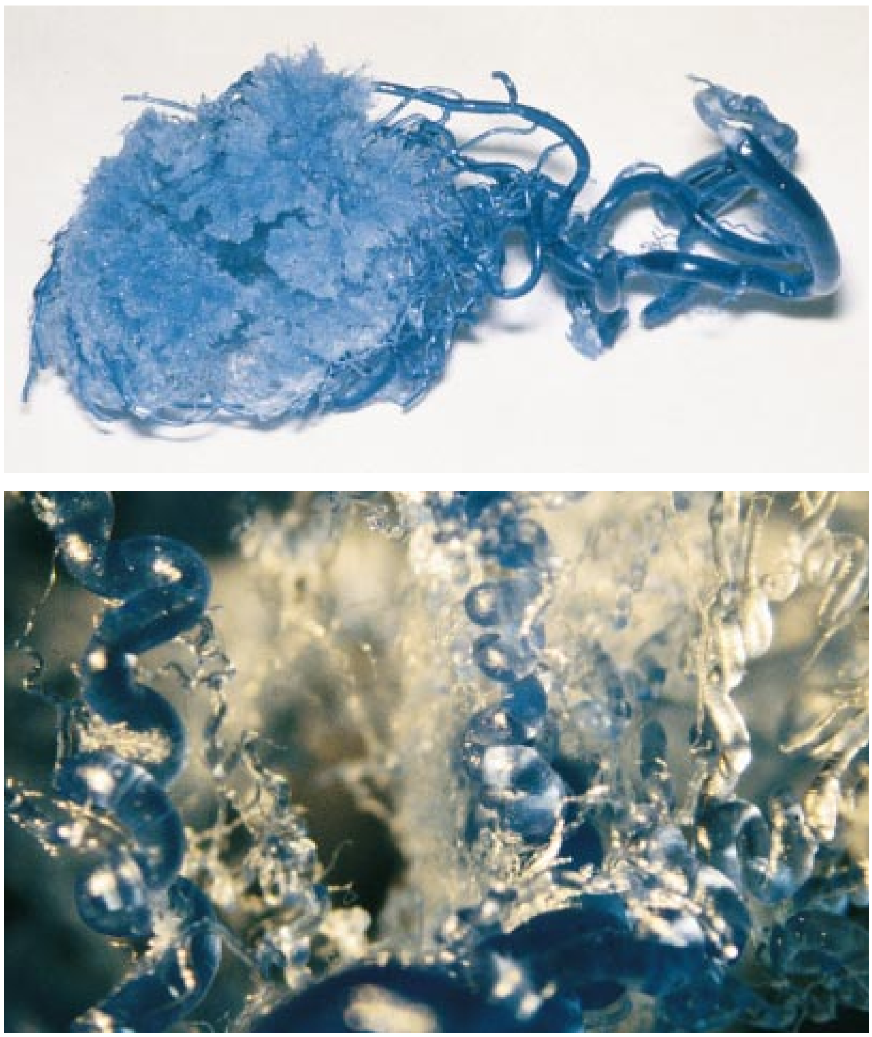
The difference between the two processes can be illustrated by a simple example. When a blob of ink is dropped into a glass of still water, the ink molecules spread gradually outward—by diffusion—until a uniform concentration is achieved. If, however, water in a glass is stirred after a drop of ink is added, the swirling water rapidly distributes the ink wherever the fluid travels.

We knew, too, that small molecules—such as oxygen and conventional chemotherapeutic drugs (which have a molecular weight lower than 2,000 daltons)—leave blood vessels and migrate through normal tissue mainly by diffusion. But large molecules—including genetically engineered drugs (which have a molecular weight greater than 5,000 daltons)—move mainly by convection.

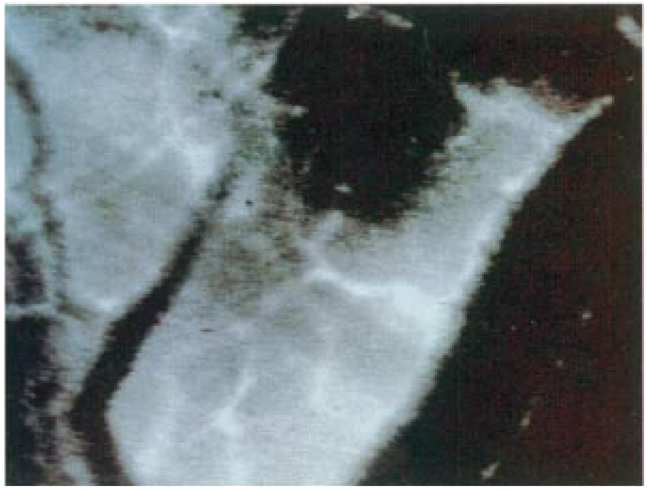
In healthy tissue, convective movement of large molecules from the blood into the interstitium occurs because the pressure in the capillary network is higher than the pressure in the interstitial tissue (which is approximately zero). But what happens in tumors? In 1987 I proposed that if the interstitial pressure in solid tumors was abnormally high, the convective passage of large drug molecules into the interstitium would be impeded. Some big molecules would still enter the matrix by diffusion but not rapidly, because the rate of passage by diffusion becomes slower as size increases.

In order to begin testing this hypothesis, my colleagues and I set out to determine whether interstitial pressures in human tumors are indeed elevated. Two early discoveries lent support to that possibility. First, a review of the scientific literature revealed that in 1950 J. S. Young and his co-workers at the University of Aberdeen had measured interstitial pressure at the center of rabbit tumors that were transplanted into other rabbits; the pressure in the tumors was higher than that in normal tissue. Other groups had subsequently published similar findings. The relevance to spontaneous human tumors was unclear, though, and so these discoveries were largely ignored.

Second, when one of my graduate students and I developed a mathematical model of pressure distribution in solid tumors in 1988, the model suggested that interstitial pressure would be high. In fact, it produced the unexpected prediction that the pressure would be equally high throughout the bulk of a



CAST OF BLOOD VESSELS in a half-pound human tumor is shown in two views. It was made by injecting a blue polymer into the vessels of a surgically excised colon cancer and then eliminating all tissue. The region resembling a crystal in the full cast (*left part of top image*) was formed by a chaotic cluster of microscopic vessels; the hole at the center of the region arose because the area lacked a blood supply. The close-up view (*bottom*) highlights one of many structural abnormalities of tumors: a number of the vessels are tortuous and twist into corkscrewlike coils that can contribute to a marked slowing of blood flow.



FLUORESCENTLY LABELED MOLECULES (*white*) were photographed soon after they entered small blood vessels in the outer rim of a tumor (*left*) and, later, after they began seeping into the interstitial matrix surrounding the vessels (*right*). The molecules were able to cross some vessel walls into the matrix but not others, such as the right wall in the rightmost vessel. They also made their way through parts of the matrix but not through others, such as the black region at the top. The differences result to a great extent from variability in the porosity of both the vessels and the matrix in the periphery.

tumor. Then the pressure would drop steeply to near zero in the periphery, where it would approach the low pressures in the surrounding normal tissue.

The finding of a uniformly elevated pressure profile startled us because every other parameter that had been measured in tumors (or has since been measured) was nonuniform. Not only is the distribution of blood vessels uneven and changeable, but the rate of blood flow can also change with time, even in a single vessel. Moreover, some vessels are extremely porous, or leaky, whereas others are not. And a single vessel can be abnormally leaky in one region but relatively nonporous in another.

No experimental data in the scientific literature could confirm or refute the model's predictions, and so a postdoctoral fellow in my group undertook to measure interstitial pressures throughout animal tumors. As predicted, he found that the pressure in large solid tumors (those greater than about half a centimeter in diameter) was uniformly high everywhere except in the outer rim. Other laboratories have now confirmed these results in animals. Since 1990 we have collaborated with physicians at the University of Pittsburgh, at the University of Munich and at Massachusetts General Hospital to measure pressures in solid tumors of patients undergoing treatment. The pressures are as high as those in animals, sometimes higher.

Part of the 1987 hypothesis has thus been borne out: interstitial pressure in large tumors is abnormally high. But is the elevation sufficient to cause serious impairment of the convective flow of large drug molecules? That is, is the pressure in interstitial tissue equal to or higher than the pressure in the microvascular network? Our mathematical model suggested they would be about equal. Simultaneous measurements of microvascular and interstitial pressures in animal tumors have recently validated this prediction. The measurements also indicate that the pressure in tumor blood vessels is higher than it is in normal capillaries. We think this elevation stems mainly from the direct and indirect compression of the vessels by proliferating cells. The unusual architecture of the vasculature and the high viscosity of blood in tumors apparently contribute as well.

Together the experimental and theoretical findings fit well into the following scenario for the development of pressure-related barriers to drug accumulation in the interstitium. A tumor initially grows in the midst of normal tissue and makes use of the existing vasculature. At this stage, it displays low interstitial and vascular pressure and relies on the existing lymphatic system to drain excess fluid from the interstitial matrix. As the tumor grows, it produces new, often leaky blood vessels but is unable to form its own lymphatic system. Meanwhile the abnormal geometry of blood vessels slows the rate of blood flow. This slowing, combined with compression by tumor cells and other factors, elevates the pressure in the vessels. Fluid seeps copiously out of them into the matrix and, in the absence of a functional lymphatic system, is not removed efficiently.

As the fluid builds in the interstitium, so does the pressure. Eventually the pressure in the vessels and the matrix equalizes. At that point, small molecules flowing in the blood continue to escape readily by diffusion (albeit only in areas that still have a blood supply). But larger molecules often remained confined in blood vessels, except in the periphery, where interstitial pressure is close to normal. Some large molecules do cross into the interstitium by diffusion, but they do so slowly. Hence, the body can eliminate most of them from the circulation before they can accumulate to optimal concentration in tumors.

The challenges faced by therapeutic agents persist even after they find their way into the interstitium. To be fully effective, they must spread throughout the matrix to cells not directly fed by blood vessels. Small molecules make the trip fairly easily (by diffusion) if they are not degraded, reabsorbed by microvessels or stopped by other processes. But large, convection-dependent molecules have a much harder time; the uniform pressure in most of a tumor's interior keeps convection from operating there.

Ironically, convection does function in the periphery—regrettably, in the wrong direction. Recall that pressure plummets at the tumor margin, where it approaches that of the surrounding normal tissue. In consequence, fluid flows from the outermost boundary of the high-pressure region into the periphery and away from the cancer. Gullino measured the magnitude of this movement in 1974. He found that approximately 10 percent of the blood fluid leaving a solid tumor oozes out from its periphery rather than draining via a vein. We and others have since confirmed his finding. The liquid escaping from the tumor surface carries drug molecules out and away from the tumor in the process.

For molecules that remain in the interstitium, the often slow process of diffusion is the only available means of dispersion to poorly perfused areas. In mathematical terms, the time required to move some distance by diffusion is proportional to the square of the distance. That is, if it takes a molecule one minute to travel one micron, the molecule will require four minutes (2^2) to move two microns and 16 minutes (4^2) to move four microns. In contrast, the time required to move by convection, if it were operating, would be directly proportional to the distance alone. In other words, a molecule that spent one minute traveling across one micron of tissue by convection would require only two minutes to travel two microns and four minutes to travel four microns.

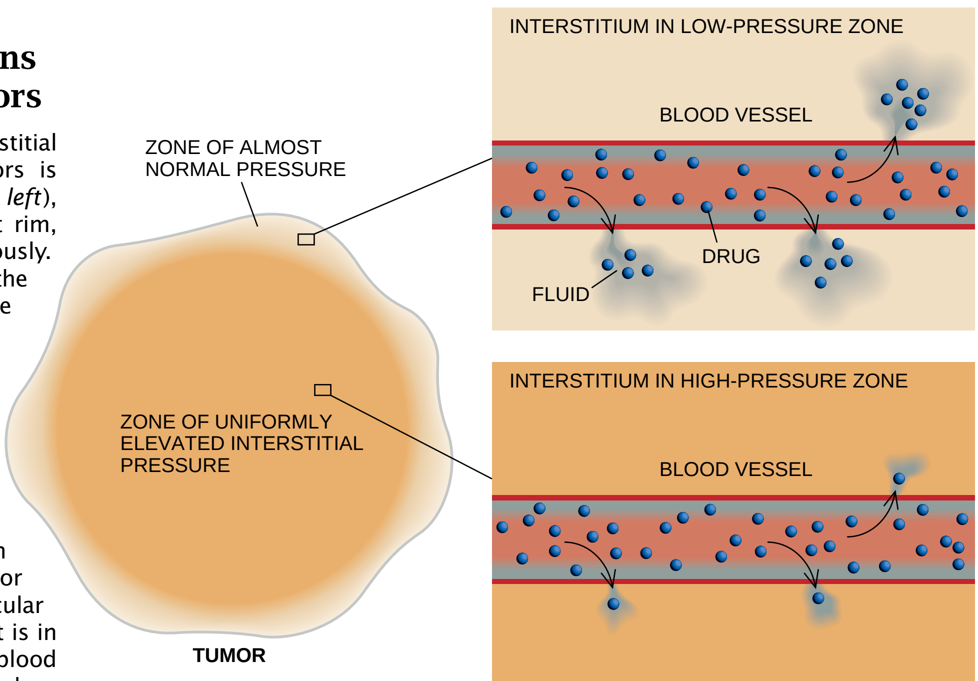
How long might it take for macromolecules at the periphery of a tumor to spread to the center by diffusion? To gain a sense of the answer, we injected molecules having different sizes, shapes and charge distributions into animals and measured the time it took for them to pass through the tissue visible in our transparent windows. We then fed this information into a mathematical model that allowed us to calculate the time it would take for the molecules to reach uniform concentrations in tumors of different sizes and physiological characteristics. We found, for instance, that a continuously supplied monoclonal antibody having a molecular weight of 150,000 daltons could take several months to reach a uniform concentration in a tumor that measured one centimeter in radius and had no blood supply in its center.

Such slow spread could easily hinder the ability of macromolecules, including the new generation of genetically engineered drugs, to eradicate tumors. Agents delivered to the bloodstream do not survive there indefinitely. Repeated doses would therefore have to be delivered to maintain elevated concentrations in the blood for as long as would be needed to achieve complete penetration of a tumor. Ongoing delivery would not only be expensive, it could also be harmful to some normal tissues that took up the drugs. (For any drug to be acceptable as a therapy, it must work without causing serious irreversible damage to healthy tissue.) Repeated dosing could also lead the immune system to manufacture antibodies and other agents that would degrade the drug before it could achieve its maximal effect.

Other impediments to passage in the interstitium are also under investigation by us and by others. These studies indicate that many genetically engineered drugs are sticky; this property

Pressure Patterns in Human Tumors

Pressure in the interstitial tissue of solid tumors is uniformly elevated (*top left*), except in the outermost rim, where it drops precipitously. The elevated pressure in the inner zone can impede movement of large drug molecules into the matrix from the bloodstream—for a simple reason. Large molecules travel mainly by convection, flowing in fluid from a high- to a low-pressure region. In the outer zone of a tumor (*top detail*), the vascular pressure is higher than it is in the interstitium, and so blood fluid (*gray*) laden with drug molecules (*blue*) seeps (*arrows*) into the interstitium. In the inner zone (*bottom detail*), the interstitial pressure is about equal to that in the blood vessels; hence, convection virtually ceases. High interior pressure, which was predicted by a mathematical model, has now been found in humans (*table*) as well as in animals. The pressure readings listed are measured in millimeters of mercury.



INTERSTITIAL PRESSURE IN HUMAN TUMORS

TYPE OF TISSUE	NUMBER OF PATIENTS	MEAN PRESSURE
NORMAL BREAST	8	0.0
NORMAL SKIN	5	0.4
RENAL CELL CARCINOMAS	1	38.0
CERVICAL CARCINOMAS	26	22.8
COLORECTAL LIVER METASTASES	8	21.0
HEAD AND NECK CARCINOMAS	27	19.0
BREAST CARCINOMAS	8	15.0
METASTATIC MELANOMAS	12	14.3
LUNG CARCINOMAS	26	10.0

slows the rate of diffusion, just as sticky shoe soles would handicap someone running a race. Moreover, enzymes in the interstitial compartment might degrade some drugs before they have the opportunity to act on target cells. Tumors also display metabolic disorders that can reduce the effectiveness of some drugs and radiation. For instance, the relative lack of oxygen in tumors may lead cancer cells to secrete high levels of lactic acid. A number of drugs will break down or fail to work in an acidic environment.

Oddly enough, the very factors that act as impediments to penetration can occasionally be beneficial. For example, stickiness can help retain a drug in a tumor. Some drugs work better in an acidic or hypoxic environment. Further, if a compound can overcome the multifarious roadblocks and eventually accumulate in a poorly vascularized area, the lack of a blood supply becomes a bonus. In what we refer to as the reservoir phenomenon, the accumulated drug, which has few avenues of escape, can serve as a reservoir that releases the drug gradually into neighboring regions of a tumor.

Certain barriers that confront drug molecules also hinder the delivery of white blood cells administered as anticancer weapons. The heterogeneous blood supply of tumors is clearly a major barrier, just as it is for molecules. We are still trying to determine the extent to which cancers resist extravasation and interstitial migration by white cells. The cells' ability to leave blood vessels may depend less on pressure gradients than on their ability to attach to endothelial cells in vessel walls and to deform themselves in ways that increase their contact with the wall. These behaviors help the cells to squeeze between endothelial cells and thus propel themselves out of the vascular system.

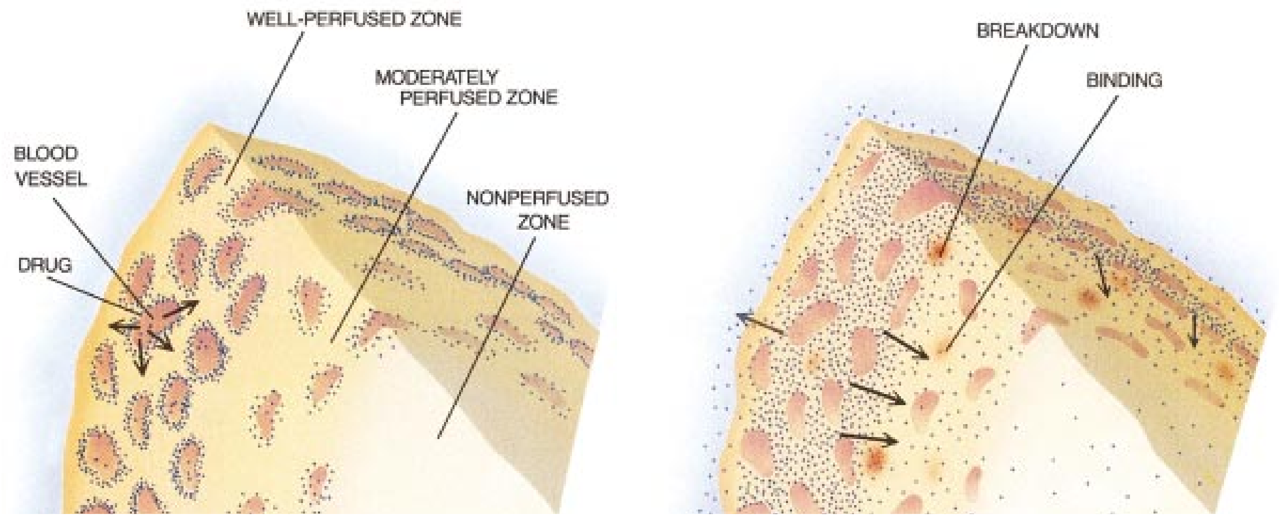
Once cells are in the interstitial matrix, they migrate by attaching to the matrix and crawling through it. The efficiency of this movement is again influenced by the cells' adhesive properties and deformability as well as by various characteristics of the tumor tissue. For instance, certain molecules in the tumor milieu can facilitate or hamper cell motility; others control the direction in which the cells migrate. More study of these processes is needed.

I should note that when therapeutic cells or molecules eventually reach cancer cells, their problems do not end. They may encounter added resistance from the cells themselves. Other laboratories are studying that phenomenon [see "Multidrug Resistance in Cancer," by Norbert Kartner and Victor Ling; SCIENTIFIC AMERICAN, March 1989].

If, as our data indicate, the obstacles to drug dispersion in tumors can be formidable, what might be done to circumvent or eliminate the barriers? No perfect solutions have emerged, but intriguing approaches have been proposed. As ever, early detection and treatment can be beneficial. Compared with established tumors, small ones tend to have a more uniform circulatory system and lower interstitial pressure. They should therefore be easier for conventional and novel drugs to penetrate.

For large tumors, various two- and three-step approaches to treatment are under study. In one of them, an antibody that binds selectively to some molecular constituent of the tumor is linked to an enzyme to form what is called an abzyme. The enzyme chosen is one that has no apparent effect on the body but is capable of converting an inactive form of a selected, small drug molecule (a prodrug) into a tumor-killing agent. A very large dose of the abzyme is injected into the bloodstream, so that it can accumulate in a tumor in spite of slow extravasation and slow interstitial diffusion. High doses can be used because neither the antibody nor the enzyme component significantly harms normal tissue. The prodrug is injected once the abzyme accumulates in the tumor and is cleared from normal tissues and the systemic circulation. Being small, it can readily diffuse out of the tumor circulation; in the interstitial matrix, it should encounter the abzyme, become activated and wend through the tumor, eradicating malignant cells.

Another strategy for evading the barriers to drug dispersion would be to inject patients with liposomes (fatty vesicles) that have been filled with a drug of low molecular weight. The newest generation of liposomes persists for a long while in the blood. Hence, the vesicles should have time to exit from leaky areas of vessels and to reach reasonably high levels in the surrounding interstitium. There the liposomes would grad-

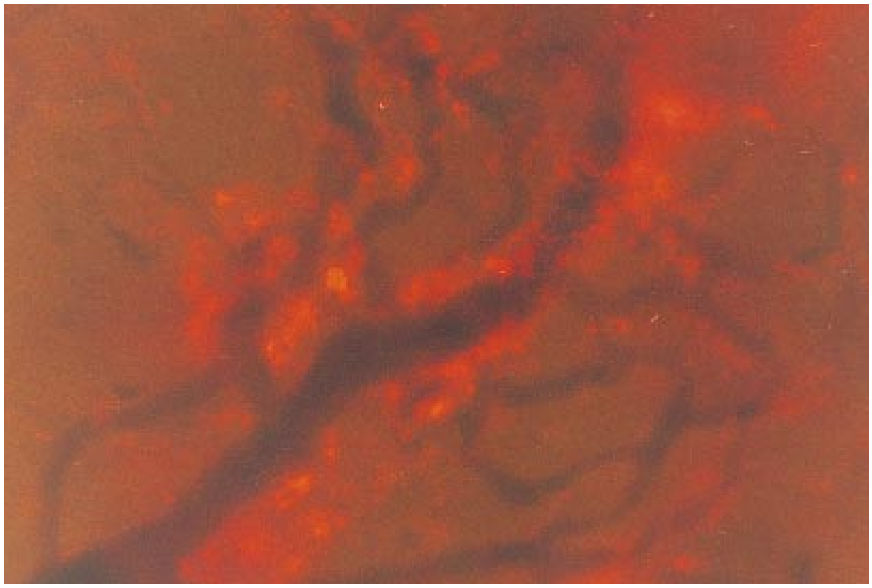


SEGMENT OF A TUMOR includes well-perfused, moderately perfused and nonperfused zones (*left*); a full tumor can have many similarly heterogeneous areas. Drug molecules (*blue*) that reach the interstitium from blood vessels (*red*) immediately attempt to diffuse from well-perfused zones (where the concentration of the drug is the highest) to poorly perfused, low-concentration zones (*black arrows in right image*). But many of the molecules never complete the journey. Some of them are washed away in fluid oozing from the periphery (*gray arrow*). Others may undergo processes—such as binding to the interstitial matrix or degradation—that significantly delay or halt their progress.

ually release the drug, which would proceed to disperse throughout the tumor.

The abzyme and liposome strategies, like others that exploit small molecules, have an Achilles' heel, however. As noted earlier, small molecules can degrade quickly. They can also seep back into tumor vessels and be cleared away from tumors as easily as they diffuse out of blood vessels. They may thus disappear before they have fulfilled their cell-killing responsibilities. Furthermore, high rates of extravasation occur only at sites where vessel walls are leaky.

As research improves understanding of the factors that govern blood flow and the movement of molecules and cells within tumors, workers should be able to invent tools that do not merely evade barriers to drug delivery but actually eliminate them. To be more precise, investigators should be able to increase perfusion in poorly vascularized areas, increase permeability of tumor vessels, reduce interstitial pressure and increase the rate of interstitial transport. Unfortunately, very few laboratories are studying the causes of nonuniform blood flow and high pressure in tumors. Work would proceed more quickly if more investigators became involved.



BRIGHT SPOTS clustered around blood vessels (*dark tubes*) are fluorescently labeled liposomes that have seeped from leaky areas of the vessels into the interstitial matrix at the periphery of a tumor. Liposomes may one day serve as vehicles for delivering drugs to the interstitium. There the liposomes would presumably release drugs over the prolonged period needed to enhance penetration in the matrix. The extent of liposome transport to the interstitium would be improved, however, if the permeability of nonleaky tumor vessels could somehow be increased.

Nevertheless, some progress has been made, particularly in lowering pressure. For example, we recently found that the drugs pentoxifylline and nicotinamide can reduce interstitial pressure in human tumors grown in animals. These agents are also known to increase the oxygen supply in various tumors, which is a benefit for radiation therapy. We also find that irradiating cervical cancers in women sometimes lowers interstitial pressure. Whether such treatment will improve the uptake of drugs in patients remains to be seen. Still, several studies in the scientific literature show that after human tumors grown in animals are irradiated, they accrue increased amounts of injected antibodies.

Another pressure-related strategy may combat tumors confined to one region. The approach calls for mixing a drug with a large amount of fluid and injecting the mixture directly into the center of a tumor. This act would increase the pressure at the core of the tumor relative to the surrounding tumor tissue. As a result, the drug would spread along the induced pressure gradient by convection, from the core through the surrounding region to the periphery. This approach is now being carried out in brain and other tumors.

Alternatively, if a tumor's vascular system could be destroyed completely, no drug would have to extravasate or cope with the interstitium. The persistent, total lack of nourishment would be expected to starve and eventually kill tumor cells. A variety of drugs—among them tumor necrosis factor and monoclonal antibodies that recognize endothelial cells or the subendothelial matrix—have the potential to shut down the blood supply completely. Under some conditions, heating and photodynamic therapies can also impair the vasculature. A number of antiangiogenesis approaches, pioneered by Judah Folkman of Harvard, are under active investigation in many laboratories and clinics.

The white blood cells that are currently being tested against various cancers in human patients may prove useful as antivascular therapies as well. We have recently found that LAK cells adhere to the tumor vasculature and impair the flow of blood in the vessels. This discovery is consistent with the suggestion that in cases where these cells prove helpful to patients, they do so in part by disrupting the tumor's blood supply. Thus, toxicity to tumor cells may not be the only effect of LAK cells in the body. The finding further implies that combining antivascular therapies with therapies that are designed to attack cancer cells could well improve the effectiveness of both types of treatment.

As the age of molecular medicine and gene therapy dawns, scientists need to put expanded effort into uncovering the reasons why therapeutic agents that show encouraging promise in the laboratory often turn out to be ineffective in the treatment of common solid tumors. I hope ongoing research into the barriers to drug delivery will ultimately ensure that existing and future anticancer drugs live up to their tantalizing potential.

FURTHER READING

TRANSPORT OF MOLECULES ACROSS TUMOR VASCULATURE. R. K. Jain in *Cancer and Metastasis Reviews,* Vol. 6, No. 4, pages 559–593; 1987.

TRANSPORT OF MOLECULES IN THE TUMOR INTERSTITIUM: A REVIEW. R. K. Jain in *Cancer Research,* Vol. 47, No. 12, pages 3039–3051; June 15, 1987.

DETERMINANTS OF TUMOR BLOOD FLOW: A REVIEW. R. K. Jain in *Cancer Research,* Vol. 48, No. 10, pages 2641–2658; May 15, 1988.

PHYSIOLOGICAL RESISTANCE TO THE TREATMENT OF SOLID TUMORS. R. K. Jain in *Drug Resistance in Oncology.* Edited by Beverly A. Teicher. Marcel Dekker, 1993.

TUMOR ANGIOGENESIS. Judah Folkman in *Cancer Medicine.* Edited by James F. Holland et al. Lea and Febiger, 1993.

TRENDS IN CANCER EPIDEMIOLOGY: A WAR NOT WON. Tim Beardsley in *Scientific American,* Vol. 270, No. 1, pages 130–138; January 1994.

Manatees

These giant aquatic grazers outchewed their rivals in the New World. Now humans, their sole enemy, hold the key to their survival

by Thomas J. O'Shea

Once upon a time, a young maiden was bathing by the banks of a river. Startled by the sight of approaching men, she jumped in, covering her bottom with a fan. Shyness then doomed her to a life in water: the maiden became a manatee, her fan metamorphosing into its distinctive spatulate tail.

So runs a legend from Mali in West Africa, echoing curiously the origins of manatees as mammals who left the land for a life in sea and river. Whereas myriad tales in native cultures from Africa to the Americas tell of manatees, only recently are the beasts yielding their secrets to scientific inquiry. Among their unique adaptations to life as marine herbivores are an unending supply of teeth—constantly replenishing worn ones—and an anomalously low metabolic rate that allows them to fast for up to seven months. Ponderous and slow, manatees have humans as their only enemy.

Manatees belong to the mammalian order Sirenia—so named because (at least to some eyes) they looked like sirens. The aquatic sirenians probably arose in the Old World; their antecedents were terrestrial mammals that also gave rise to elephants, hyraxes and perhaps aardvarks. We do not know what forces of natural selection drove an ancient mammal to exploit the niche of a large marine herbivore. Yet the beginnings of such an ecological strategy can still be seen today. Domestic sheep on islands off Scotland forage for marine algae in the intertidal zone, even swimming between patches; pigs of the Tokelau Islands in the South Pacific habitually forage along coral reefs, wading with heads submerged.

Fossil evidence suggests that Old World sirenians reached an isolated South America in the Eocene or Oligocene era, more than 35 million years ago. The earliest known true manatee (who lacked, however, the dental technology of modern manatees) lived in the middle Miocene era, 13 to 16 million years ago. The sirenian order currently includes three species of manatees in the family Trichechidae, plus their older relative, the dugong, in the family Dugongidae. A second dugongid, the Steller's sea cow, was hunted to extinction within 25 years of its discovery in 1741.

Up to the late Miocene, dugongids exploiting sea-grass meadows colonized the marine waters of the western Atlantic and Caribbean. Manatees, though feeding on both freshwater and saltwater plants, were restricted to the rivers and estuaries of South America. Dugongs are now found only in the warm, shallow parts of the Indian Ocean and the western Pacific. The animals were evidently displaced about a million years ago by upstart New World manatees, which, as paleontologist Daryl P. Domning of Howard University convincingly argues, outchewed their older relatives.

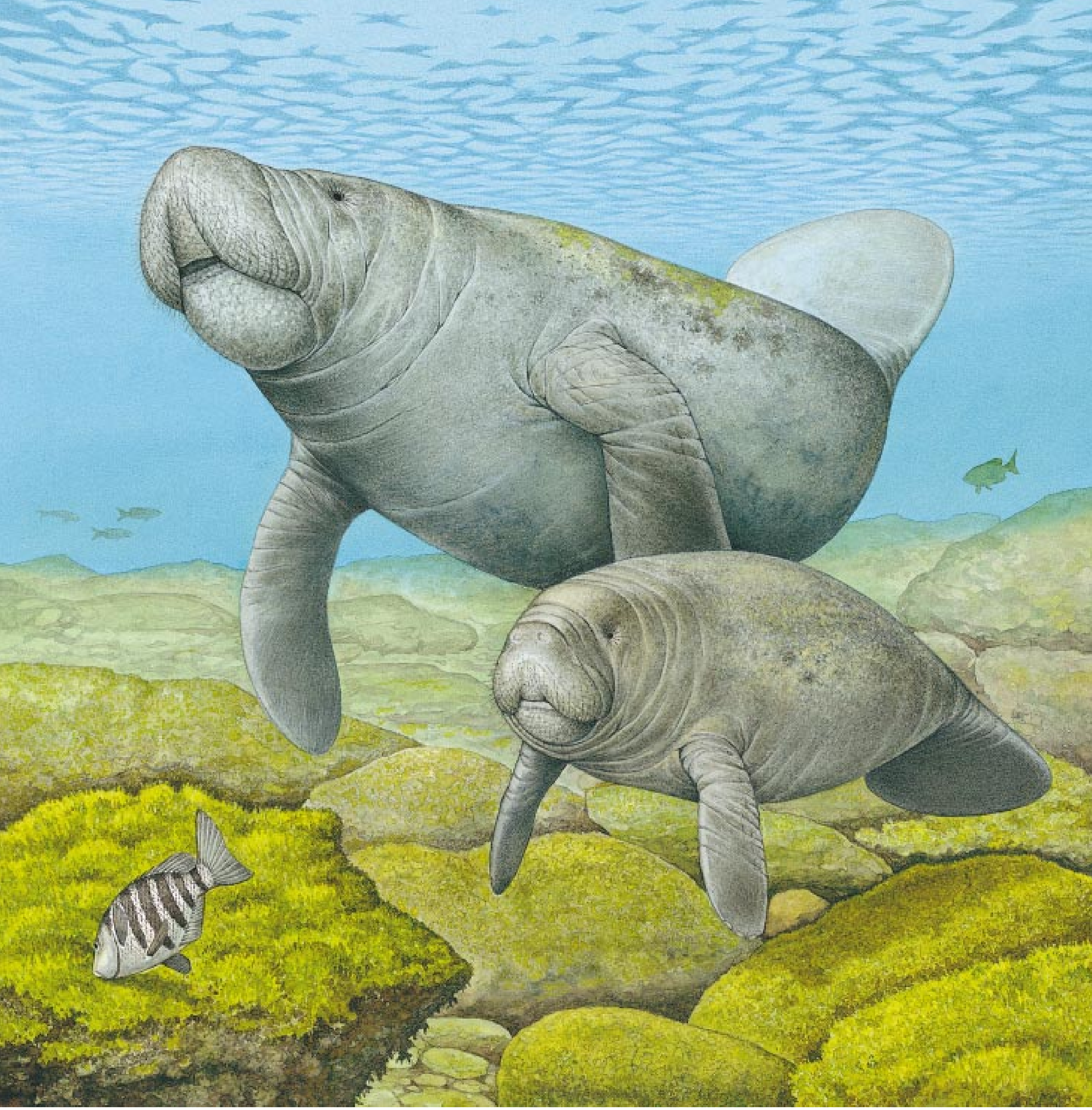
The chewing prowess of manatees derives from the fact that they never run out of teeth. Manatees possess only premolars and molars (one row on either side of the jaw), but these are continuously replaced by new teeth sprouting at the rear of the row—rather like wisdom teeth—and moving forward. The worn-down front teeth drop out, and the bony tissue separating the tooth sockets continuously breaks down and re-forms to allow the new teeth to move forward at roughly one or two millimeters per month. This process takes place throughout life: even the oldest specimens show new cheek teeth erupting at the rear of each tooth row.

Such an adaptation points to an increased abrasiveness of the manatee diet at some earlier time. Horses, for example, evolved high-crowned cheek teeth in response to the emergence of the true grasses, which have an elevated content of sandpapery silica. Currently true grasses are an important part of the diet of all three species of manatees; in contrast, the staple of the exclusively marine dugong is marine angiosperms (sea grasses). The ancient Caribbean dugongids probably also lived on this diet.

Domning notes that the newly abundant true grasses seem to have invaded South American ecosystems in the Miocene. Continental glaciation in the Pliocene and Pleistocene periods lowered the sea level and increased erosion and runoff of sand and soil. Sand deposition likely increased the amount of abrasive material ingested with the sea grasses. The more sand-tolerant trichechids invaded these habitats as well. By about a million years ago they had broadened their feeding niche to include sea grasses and replaced dugongids in the Atlantic and Caribbean.

In the Amazon region, mountain building in the late Miocene era created a transient closed basin; trichechids isolated here became the Amazon manatee (*Trichechus inunguis*). The West Indian manatee (*T. manatus*) is apparently a little-changed descendant of coastal South American trichechids of the Pliocene-Pleistocene time; at present it can be distinguished into subspecies from Florida and the Antilles. A similar form also reached West Africa by way

THOMAS J. O'SHEA is a member of the National Biological Survey of the U.S. Department of the Interior and the assistant director of the National Ecology Research Center in Fort Collins, Colo. Conservation biology and mammalogy are his primary interests. O'Shea has studied manatees since 1979, in Florida and in several tropical countries. He led the Sirenia Project of the U.S. Fish and Wildlife Service for more than seven years and served as deputy chair of the Sirenia Specialist Group of the IUCN–World Conservation Union. An adjunct associate professor in the Department of Wildlife and Range Sciences at the University of Florida and an adjunct curator at the Florida Museum of Natural History, he serves as a scientific adviser to the U.S. Marine Mammal Commission.

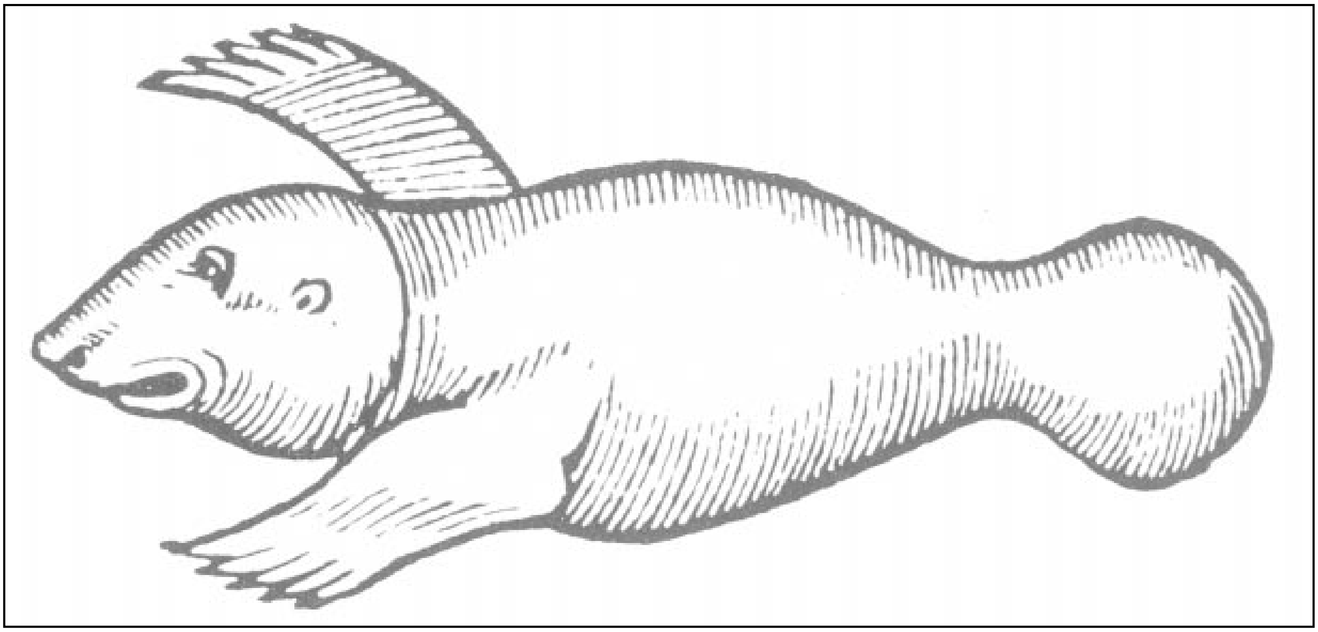


FLORIDA MANATEE and her calf laze in placid waters. Calves stay with their mothers for up to two years, communicating by faint squeaks and learning seasonal migration patterns. The two may recognize each other long after weaning.

of transoceanic currents comparatively recently—perhaps since the late Pliocene—to give rise to the West African manatee, *T. senegalensis.*

Manatees have had a long and intimate relationship with humans, mostly as food. These large animals—my colleagues and I once weighed a Florida manatee at 3,650 pounds—have been hunted and relished by members of coastal and riverine cultures throughout the tropical Atlantic. In West Africa, manatees are lured into box traps with cassava, speared from platforms on stilts, entangled in nets, shot by harpoons on baited triggers and trapped by fences on outgoing tides. South American Indians place logs across streams to trap manatees in receding waters.

The creatures also feature in numerous superstitions and legends in native cultures. In Mali, manatees in the river Niger are considered evil spirits; only a few tribesmen know the proper incantations to hunt and kill a manatee without dying or going mad. When a hunt is successful, cuts of meat are distributed



EARLIEST KNOWN REPRESENTATION of a manatee, from Fernández de Oviedo's 1535 *History of the Indies,* is based on Christopher Columbus's description of the creatures. His crew, the first Europeans to see manatees, took them to be mermaids.

according to social status. Some other protocols must be followed, too. If a pregnant woman eats certain cuts, the unborn child is believed to be in danger of becoming an adult of low moral character. The oils and skin of the manatee are made into remedies for various ailments, and potions are made from its ribs.

Across the Atlantic, in the headwaters of the Amazon in Ecuador, a Siona Indian shaman told my fellow scientists and me a legend about the origin of the Amazon manatee. An ancient god was deceived and trapped by a tapir, who cruelly subjected the god to attack by piranhas. The god escaped and in revenge banished one of the tapir's daughters to live forever in the water—as a manatee.

Inhabitants of the coasts and rivers of Central America, the Caribbean and northeastern South America prize manatees as food and as a source of medicine. Great stealth and machismo are needed to harpoon manatees from dugout canoes on the Orinoco River. Accordingly, the tiny middle-ear bones of manatees are worn as magical charms against evil and disease. The Piraoa Indians of Venezuela, however, have a taboo on hunting manatees and river dolphins. The Piraoa think they will die if they eat manatee meat; they believe the creatures to be bewitched humans who dwell in underwater cities at the bottom of the Orinoco.

In 1493 Christopher Columbus and his crew became the first Europeans to see manatees in the New World. They reported the creatures as mermaids. In the decades that followed, Europeans acquired a closer acquaintance with manatees. Indians used manatee hides as shields against Spanish explorers armed with crossbows. In the late 1600s William Dampier, the English buccaneer, fed his crews with boatloads of manatee meat from Panama, supplied by Miskito Indians.

Later, manatees were exploited commercially, both as bush meat for laborers at South American frontier posts and in processed form for resale in distant markets. Shiploads were exported to the West Indies from Guyana, Suriname and Brazil in the 17th and 18th centuries. Exports ended with the close of the 19th century, when manatee populations had become greatly reduced. Legal trade in Amazon manatees continued in Brazil until 1973. Perhaps as many as 7,000 were killed in peak years of the late 1950s.

The end of commercial exploitation coincided with the rise of other threats to the survival of manatees. By the 1970s it became clear that increasing numbers were dying in accidents and in encounters with man-made artifacts. Inexpensive synthetic gill nets that became widespread in tropical fisheries incidentally entangled and suffocated manatees. Rivers and estuaries became polluted and turbid because of deforestation and erosion, blocking the light needed by aquatic plants. Manatee habitats became endangered, especially in areas where humans are plentiful. An estimated 800 to 1,000 people move to Florida every day; many newcomers settle in coastal areas, where wetlands are drained and replaced by housing—complete with canals and boat docks. Water quality has dropped; in Tampa Bay, 80 percent of the sea-grass beds have vanished over recent decades. In Florida, boat propellers inflicted gruesome wounds. Deaths of manatees accidentally caused by humans doubled during the 1980s, as the number of registered boats plying the waters also grew dramatically.

Clearly, the manatee needed protection, or else it would rapidly become legend alone. So little was known about its habits and habitat, however, that it was difficult to formulate plans to save the manatee. Then, in the 1970s, the Marine Mammal Protection Act and the Endangered Species Act initiated a spurt of research, concentrated in Florida.

Biologists began, for one thing, to dissect systematically carcasses cast on the beach. The dissections augmented earlier knowledge about the anatomy of manatees. In addition to replacing their teeth constantly, manatees have evolved other features helpful to their aquatic lifestyle.

A manatee's lungs have single lobes and lie above the abdominal cavity—along the back—enabling the creature to remain horizontal under water. Its gastrointestinal tract is long, and the animal digests food in the hindgut. There is also a "cardiac gland" in a pouch off the stomach. The gland contains specialized secretory cells, and the pouch protects them from the abrasive diet.

The animal's heavy, dense bones lack marrow; they help to keep the manatee submerged. Only the backbone contains some marrow for producing red blood cells. The animal can adjust its depth by changing the volume of its lungs. Its lips are large, studded with bristles, and prehensile—forming a kind of abbreviated trunk. The long, paddlelike flippers are used to manipulate food and to walk stealthily along the bottom. Counts of growth layers in ear bones suggest that the manatee's normal lifespan is about 60 years.

By examining the stomach contents of manatees, Catherine A. Beck and her colleagues at the National Biological Survey in Gainesville, Fla., have found that the diet has much variety. For instance, manatees eat most local aquatic plants, as well as acorns from overhanging oak trees. Unfortunately, their stomachs also include refuse, such as plastics, condoms, fishing line and steel hooks—some of which have led to their death.

Also in the 1970s biologists began to study live, captive manatees, thereby uncovering some intriguing facts about the rate at which the animals consume energy. The metabolic rate is related to the amount of oxygen a mammal consumes per unit time divided by its body weight. Most species consume energy at rates that depend on their size. For example, small mammals have a large surface area for their weight. Because heat is lost through the surface and mammals have to expend energy to maintain a constant body temperature, smaller mammals use more energy relative to their size and have higher metabolic rates.

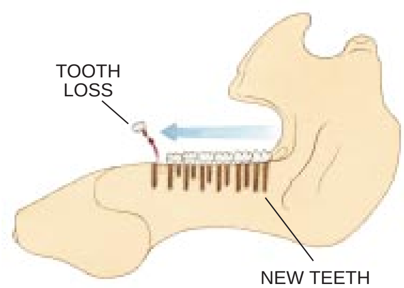
On a mouse-to-elephant curve of metabolic rates for mammals of different weights, most kinds of marine mammals fall where they should—going by their respective sizes. But Blair Irvine of the U.S. Fish and Wildlife Service and C. James Gallivan and the late Robin C. Best of the National Institute for Amazon Research in Brazil found that manatees deviate sharply from this pattern. Amazon manatees metabolize energy at only 36 percent of the rate expected for mammals of their size; in Florida, manatees come in at a mere 15 to 22 percent of the predicted values.

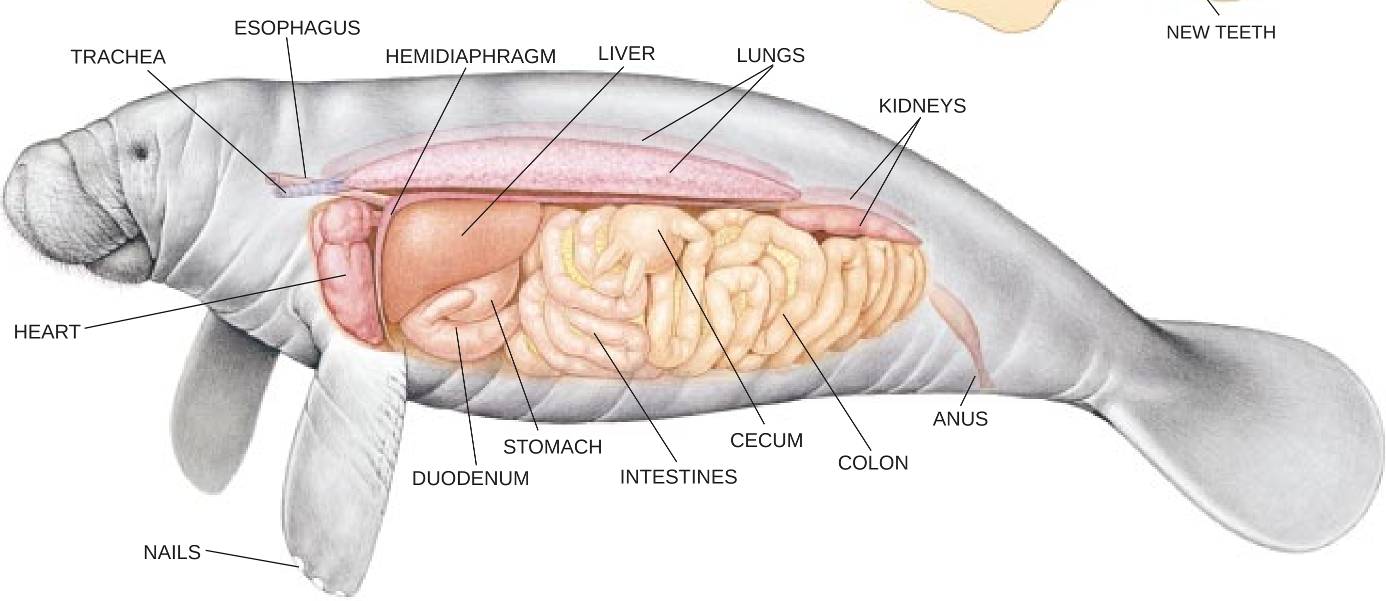
What are the ramifications of such low metabolic rates? One of the most astounding is the capacity that the Amazon manatee has for prolonged fasting. As in earlier days of sirenian evolution, seasonal flooding in the upper Amazon gives rise to large floating meadows of grasses and other vegetation. Amazon manatees swim through the treetops of the flooded forests in times of plenty, but during the dry season they can become isolated for months in lakes and pools devoid of vegetation [see "Flooded Forests of the Amazon," by Michael Goulding; SCIENTIFIC AMERICAN, March 1993]. Like mammals of colder zones preparing for winter dormancy, the creatures put on large quantities of fat during the flood season, which allows them to survive the lean times of the dry season. Best calculated that Amazon manatees can go without feeding for almost seven months by subsisting on stored fat and by expending energy at their customarily low metabolic rates.

We do not know if there is a similar link between metabolism, fattening and seasonal fasting for other species of manatees. But the influence of low metabolic rates on the distribution of Florida manatees is well known. The animals do not live north of Florida or Georgia on a year-round basis. Unable to metabolize energy fast enough to counter heat lost to cool surrounding waters, manatees probably cannot expand beyond subtropical climes. Nearly all sirenians throughout geologic time have lived in warm regions.

Almost every year, however, manatees are sighted north of their typical range. These summer wanderers return south in autumn or die and wash ashore in Virginia or the Carolinas later in winter. The carcasses are marked by an absence of fat deposits and other signs of long-term exposure to cold. Wanderers have been verified as far north as the

ANATOMICAL DRAWING of a manatee reveals several adaptations to life as an aquatic grazer. Lungs lying along the back help to keep the animal horizontal under water. Flippers and a prehensile upper lip are used to manipulate food. New teeth (*right*) are constantly generated at the rear of each tooth row and move forward. Eventually, teeth worn out by chewing abrasive grasses drop out. The food is digested as it passes through the long hindgut (cecum to anus) for several days.

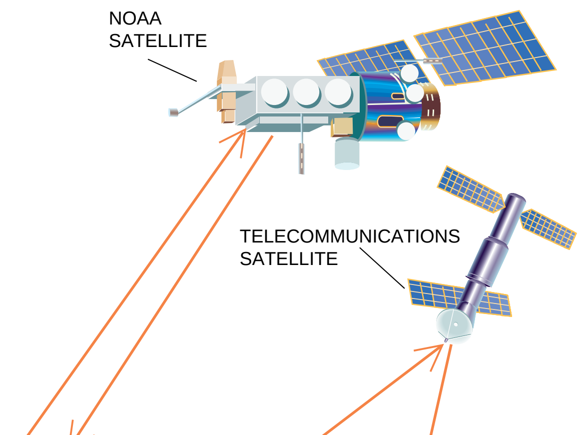


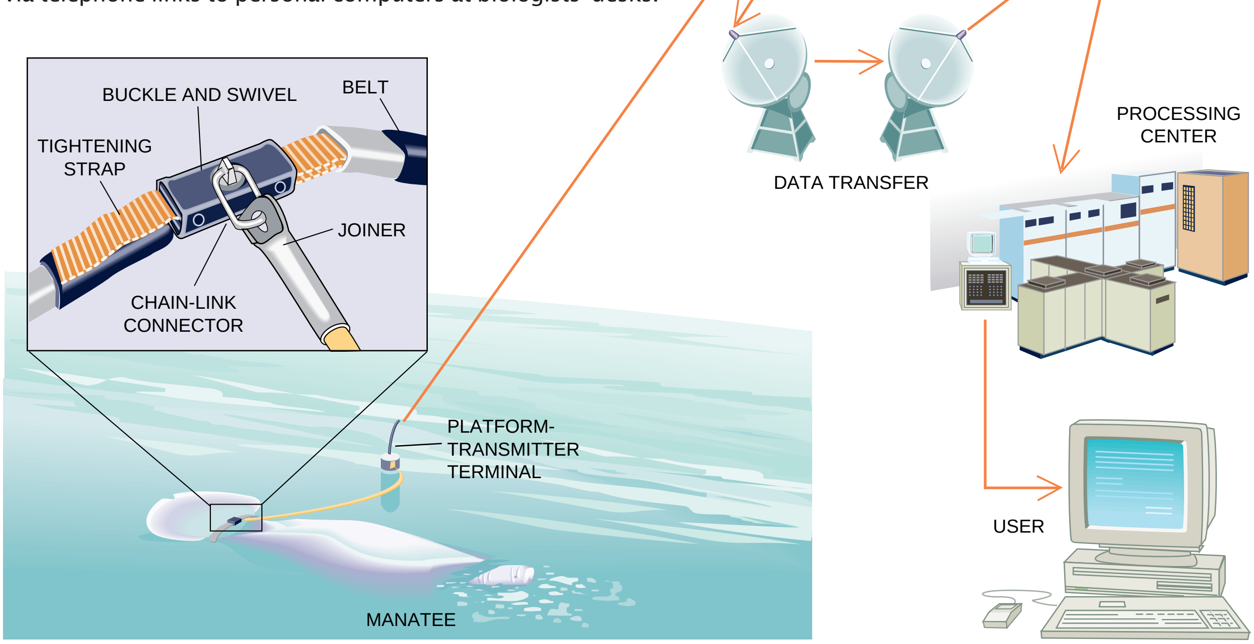


Satellite Tracking

The travels of Florida manatees are monitored by satellite telemetry. A floating platform-transmitter terminal is attached to a belt on a manatee. The belt falls off after a few months. The inset shows the belt buckle; the swivel minimizes drag, while the joiner frees the manatee if the tether snags.

The terminal emits signals at an ultrahigh frequency. The transmissions—encoding the identity of a manatee, the water temperature and the angle at which the platform is tipped—are received by a passing satellite of the U.S. National Oceanic and Atmospheric Administration. Equipment on board the satellite calculates the transmitter's position and relays all the information to stations on the earth. Data next pass through telecommunications satellites to processing centers and then via telephone links to personal computers at biologists' desks.





lower Potomac; I strongly suspect that the "Chessie" sea monsters of Chesapeake Bay in the 1970s and 1980s were errant Florida manatees.

Even in Florida, manatees find winter temperatures stressful. They respond by migrating to the southern third of the peninsula or to local sources of warmer water. These include artesian springs, such as Crystal River on the Gulf Coast, and effluents of pulp mills and electric power plants. On the coldest days, 300 or so animals aggregate at several of these sites. For many years, winter aggregations gave biologists the only window into the lives of manatees in the wild.

In the early 1950s Joseph C. Moore of Everglades National Park noted that individual manatees can be distinguished by boat-propeller scars; he was thus able to make some basic behavioral observations. His initial study was followed in the 1960s by Daniel Hartman, then a graduate student at Cornell University. Hartman tracked individual females and their offspring at Crystal River. Currently the National Biological Survey maintains a computer-based scar-pattern catalogue identifying hundreds of Florida manatees from sites throughout their range.

Dozens of wild females have now been observed for more than a decade—some for 20 to 25 years—giving us insight into their reproductive cycle. A female attains sexual maturity as young as three years of age and continues to reproduce for more than 20 years. One calf is born every two to three years; there are occasionally even twins.

Social interactions seem to revolve around reproduction. Manatees are basically solitary animals, but when a female is in estrus, she is pursued by a herd of six to 20 jostling and wrestling males. About a year after mating, apparently with several males, the female selects a secluded area for birthing. Mother and calf stay together as a nursing pair for at least a year, maintaining contact by faint, squeaky vocalizations.

A few females and their calves have been seen together for up to a year after weaning and perhaps continue to recognize one another for much longer. Sometimes nursing females will "adopt" and suckle calves that are not their own. Manatees will occasionally socialize in transient groups, which individuals join and leave seemingly at random.

In 1978 speed zones for boats were established at winter aggregation sites. Still, manatees and boats continued to collide at other times and places. If they were to protect the creatures in their diverse habitats, policymakers needed to know more about the distribution and migration patterns of manatees. In the 1980s advances in radiotelemetry provided a means of observing manatees in their travels over long distances.

Manatees in the waters of Brazil and Florida were first to be tracked by radio in the 1970s. Investigators attached a transmitter to a belt—designed to corrode and fall off after the study—around the constriction between a manatee's body and its spatulate tail. The technique was not, however, useful in coastal habitats, where saltwater impedes the passage of radio waves.

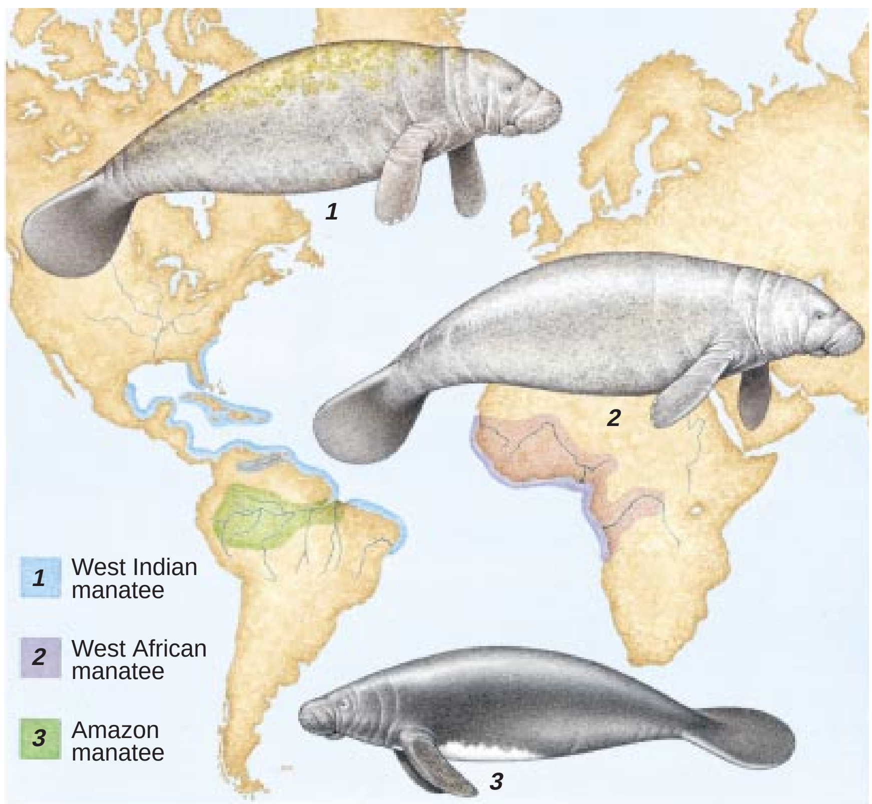
So my colleagues Galen B. Rathbun, James P. Reid and James A. Powell designed a floating transmitter attached to the belt by a two-meter-long flexible nylon tether. The tether was equipped with swivels to minimize drag and with breakaway links to prevent the manatee from getting trapped should the tether snag. The device put the antenna in the air during most manatee activities in shallow waters. We could locate animals as far as 20 to 30 miles away from light aircraft and five to 10 miles away from boats or shore.

Soon we were able to take the floating-transmitter concept a step further. Bruce Mate of Oregon State University had been trying to track great whales in the open ocean. He solved this tremendous logistical problem by attaching to the whales "platform-transmitter terminals" monitored by satellites. These transmitters, used to track weather balloons and vessels at sea, emit ultrahigh-frequency signals. Satellites receive the signals and pass on the encoded information to processing centers on the earth. From the centers it travels via telephone links to personal computers. Within hours or less of the last pass of a satellite over a transmitter, a scientist can know a whale's location.

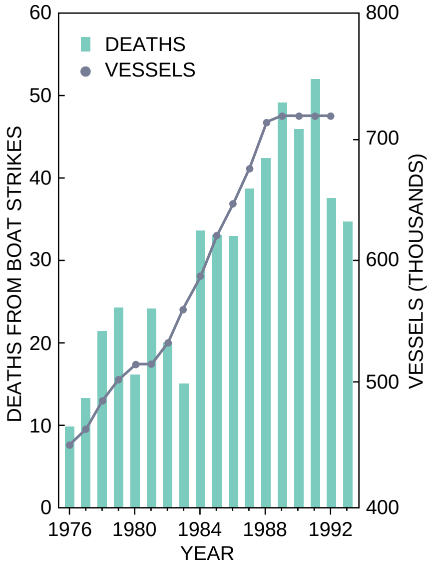
In 1985 Rathbun, Mate and Reid released a manatee with a floating satellite transmitter off Florida's Gulf Coast. Its signals were received by satellites of the National Oceanic and Atmospheric Administration. Since then, we have tracked more than 100 manatees with tethered transmitters, most by satellite. The technique has been adopted by the Florida Department of Environmental Protection, is being used in Puerto Rico and has been applied to dugongs in Australia.

The satellite link reveals a manatee's location (within 100 meters), the water's temperature and the number of times the transmitter is tipped—giving clues about the animal's activity. In Florida, this information is correlated with maps of sea-grass beds, warm waters and other manatee resources at the Marine Research Institute of the Florida Department of Environmental Protection in St. Petersburg. Earlier, it was believed Florida manatees moved slowly

WATER HYACINTH, a floating plant, is held down with a flipper and stripped of its broad leaves by a Florida manatee in the Crystal River National Wildlife Refuge.



THREE SPECIES OF MANATEES now live in tropical and subtropical waters. The West Indian manatee, shown with algae on its back, most closely resembles the earliest trichechid. The African manatee is descended from dispersed West Indian manatees. The Amazon species feeds on grasses and has the most complex dentition.



SOURCE: Florida Department of Environmental Protection

BOAT-KILLED MANATEE displays slashes inflicted by a propeller. The graph details the increase in the number of boats registered in Florida and in the number of manatees killed by boats since the mid-1970s. The correlation, while not proving a cause-and-effect relationship, is highly suggestive. The rise in the number of boats also indicates that their use for recreation is increasing. Consequently, so is their destructive effect on the manatee habitat.

and were essentially nomadic. Now we know they can travel fairly fast—sometimes 50 kilometers a day—and their seasonal movements can be quite directed. Some females, for instance, will often graze in roughly the same areas every summer and head for the same hot spots every winter. Offspring appear to learn the mother's migration patterns.

Generally, though, manatees are flexible in the kinds of habitats they occupy. We have tracked individuals from southern Georgia and northeastern Florida—where the primary forage is saltmarsh grass available only on banks at high tide—traveling in less than five days to the Merritt Island National Wildlife Refuge, where they feed almost exclusively on submerged, rooted marine angiosperms. After lingering in this region, rich in classic Florida wildlife such as wading birds, sea turtles, bottlenosed dolphins and alligators—which sometimes seize and detach the trailing transmitters—the same manatees may continue southward, arriving in the urban environs of Fort Lauderdale and Miami for the coldest weeks of the year.

Increased knowledge about the adaptability of manatees to diverse habitats and about their rate of reproduction has made us guardedly optimistic about the survival of manatees in Florida—once they are adequately protected. Administrative efforts by the state of Florida, the U.S. Fish and Wildlife Service and local governments to reduce accidents have a high potential for paying off. We have combined the resighting histories of manatees in the scar-pattern catalogue with recent statistical theories to estimate the year-to-year survival rate of adult manatees. Chances of survival are good in areas with solid histories of protection, like Crystal River. The population there has grown from about 60 animals some 20 years ago to nearly 300 now. In Florida as a whole, 1,856 have been counted by air in winter.

We do not know how many manatees were missed in these counts. Although the general pattern from the 1970s through the 1980s pointed to an increase in manatees in several regions of Florida, trends in most recent years leave doubt about whether the population is growing at all. Along with lower estimates of survival in less protected regions (from our mathematical models) and more manatees being found dead from human activities, uncertainties in the recent population data call for continued efforts for conservation focused on some key areas of Florida. Should such efforts maintain their momentum—and barring unforeseen catastrophes—the Florida manatee could become a rare success story for endangered species.

Ultimately what will save these creatures is a sympathetic public. In this regard, there has been some genuine progress. Manatees have become extremely popular. For example, walking into my daughter's elementary school classroom, I was pleasantly surprised to see—in the permanent alphabet displayed above the blackboard—along with "A" for apple, "M" for manatee.

Internationally, the situation is less encouraging. Although most of the 40 to 45 nations in which manatees live offer them legal protection, the laws are not well enforced. Also, few attempts have been made (outside of Florida) to protect their habitat. Guatemala created the world's first manatee reserve some 30 years ago. The secretive animals are rarely glimpsed there, but the reserve is still maintained. Along the coast of Panama, where Dampier provisioned his crews centuries ago, meager numbers of manatees persist in rivers of the Bocas del Toro region, and Panamanian conservation groups are working on their behalf.

It will be an uphill battle. But similar attempts are starting elsewhere. A new generation of conservation biologists from the tropical Atlantic is increasingly enthusiastic and concerned about manatees. The energy of these scientists was evident at the First International Conference on Manatees and Dugongs held at the University of Florida this past spring. Efforts to learn more about manatees in the tropics and to apply this information to conservation seem to be growing, providing seeds of hope for the future.

The Warauno Indians of the Orinoco Delta Territory of Venezuela refer to the Milky Way Galaxy as "the road of the manatee." I remain hopeful that the underwater roads of the manatee will continue to be traveled along our tropical Atlantic rivers and coastlines here on the earth, to the marvel and delight of future generations.

FURTHER READING

ECOLOGY AND BEHAVIOR OF THE MANATEE (*TRICHECHUS MANATUS*) IN FLORIDA. Daniel S. Hartman. American Society of Mammalogists, Special Publication No. 5, June 27, 1979.

DISTRIBUTION, STATUS, AND BIOGEOGRAPHY OF THE WEST INDIAN MANATEE. L. W. Lefebvre, T. J. O'Shea, G. B. Rathbun and R. C. Best in *The Biogeography of the West Indies: Past, Present, and Future.* Edited by C. A. Woods. Sandhill Crane Press, 1989.

MANATEES AND DUGONGS. J. E. Reynolds III and D. K. Odell. Facts on File, 1991.

WATERBORNE RECREATION AND THE FLORIDA MANATEE. T. J. O'Shea in *Wildlife and Recreationists: Coexistence through Management and Research.* Edited by Richard L. Knight and Kevin J. Gutzwiller. Island Press, Covelo, Calif. (in press).

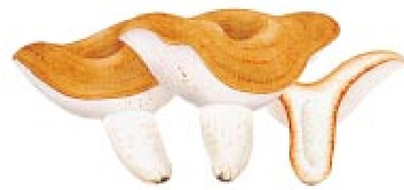
Jean Henri Fabre

This reclusive entomologist became one of the most popular educational authors of his day. A look at his greatest work reveals both the underappreciated achievements and the failings of his science

by Georges Pasteur

Boletus appendiculatus

Order Aphyllophorales



Lactarius sanguifluus

Much of the general public has forgotten the entomologist and author Jean Henri Fabre, yet during the 19th and early 20th centuries, he was one of the most celebrated educators on the subject of nature. Victor Hugo once called him the "Homer of insects." Edmond Rostand rebaptized him the "Virgil of insects." Toward the end of his life, Fabre was nominated for the Nobel Prize in Literature; the overenthusiasm of his supporters, who pestered the Swedish Academy, may have ruined whatever chance he had.

Today Fabre is most widely remembered by the people of Japan. Exhibitions about him are commonplace there, and schools frequently use his works as references. Since 1923 the Japanese publishing market has featured no fewer than 47 complete or partial translations of Fabre's 10-volume compendium entitled *Souvenirs entomologiques* (*Entomological Memories*), as well as editions of his other books.

GEORGES PASTEUR became involved in Fabran studies at the suggestion of the American entomologist George P. Georghiou. A field vertebrate student in Africa, Pasteur was trained in insect behavior by studying the consequences of selection for geotaxis and phototaxis on fruit flies in the laboratory of Theodosius Dobzhansky at the Rockefeller University in 1968 and 1969. Pasteur has been the head of the laboratory of ecological genetics at the École Pratique des Hautes Études in Montpellier since 1978 and is associated with the National Museum of Natural History in Paris. In addition to several books in French, he is co-author with Nicole Pasteur of *Practical Isozyme Genetics.*

Generations of scientists around the world have also revered Fabre's memory. They fondly recall when, as teenagers, they read extracts of *Souvenirs* and were roused toward their vocations. Fabre's inspirational effect is not without irony: some of his scientific principles were unorthodox in his own time and are still more so today. In addition, although Charles Darwin greatly admired Fabre and referred to him as the "inimitable observer" in *On the Origin of Species,* Fabre refused to accept the theory of evolution.

Yet the charm of *Souvenirs entomologiques* overwhelms whatever quirks Fabre may have had. As Fabre explained in an 1882 letter to Jean Baptiste Dumas, the former mentor of Louis Pasteur, he had adopted a "light form" in *Souvenirs* so it would "find readers in addition to naturalists." That it did: *Souvenirs* was hugely popular from the moment the first volume appeared in 1879.

The more than 4,000 pages of *Souvenirs* combine original science with prose that is uncannily rich and vivid, at times funny, at times poetic and always witty. They alternate recollections and digressions on all manner of topics with suspenseful tales of investigations into insect behavior. They expose Fabre's emotions and opinions about everything. And they blend unflagging optimism and inexhaustible admiration for nature in a relaxed atmosphere, like that of warm summer days spent in fields of joyously buzzing insects. *Souvenirs* thus offers an extraordinary glimpse into both the character and the science of its author.

Born in 1823 to a poor rural family in the south of France, Fabre—"Henri," at home—had to fight throughout his childhood for scholarships and later for menial jobs to pay for his education. Because he could scarcely afford school at all in some years, he developed an outstanding aptitude for self-teaching. By himself, he learned both ancient Greek and English. Fabre became a bachelor of letters in 1844 and of mathematics in 1846, then a licensee in mathematics in 1847 and in the physical sciences in 1848. (Since the Middle Ages, French universities have conferred a license, intermediate between a baccalaureate and a doctorate, akin to the American master's degree.) He later became a licensee in the other natural sciences. His educational progress was especially remarkable given that he rarely if ever attended courses in these subjects. In 1855 Fabre became a *docteur ès sciences* from the Sorbonne, toiling in his kitchen to complete one dissertation on animals and two on plants.

Such beginnings paved the reclusive path that Fabre was to follow throughout his career. While working as a lycée (high school) teacher of physics, Fabre

Copyright 1994 Scientific American, Inc.

became the quintessential hermit. His solitary habits ruffled many of his colleagues' feathers, which is one reason why he never attained a university position. A more important reason was Fabre's lack of money. The University of Poitiers did offer him a position in the early 1860s, but Fabre had to refuse it because he was not financially independent: at that time, faculty members were paid only token salaries, and research grants were unknown.

Fabre staved off poverty by two means. First, putting his knowledge of chemistry to work, he tackled the problem of extracting the red dye alizarin, which was widely used in textiles, from madder, a common European herb. Fabre succeeded where others had failed, and by 1866 he held a patent for obtaining pure alizarin powder from madder. Unfortunately, his luck was short-lived. By 1867 German chemists had learned how to synthesize alizarin from other compounds. As a result, madder culture was abandoned altogether.

A second moneymaking scheme was more enduringly fruitful. Fabre became a 19th-century Isaac Asimov: he wrote books that made science and technology familiar to adults and children alike, substituting rationality for the old wives' tales that dominated most people's thinking. Between 1862 and 1891 he wrote 95 such books, and all proved smashing successes. Indeed, during Fabre's lifetime, many of them outsold *Souvenirs.* Several stayed in print through the 1920s—*Le ciel* (*The Sky*), first published in 1865, saw 11 printings. A beautifully illustrated Japanese edition of *Histoire de la bûche* (*The Wonder Book of Plant Life*), which is probably the best known of the books, appeared as recently as 1986.

"INIMITABLE OBSERVER" Jean Henri Fabre, as Charles Darwin called him, conducted extraordinarily detailed entomological studies on the grounds of his home in southern France. His books on insects and many other topics have inspired a love of science in generations of readers. Among his other accomplishments, Fabre was also a brilliant painter of mushrooms (*opposite page*).

In 1868, because of those books' fame, the education minister personally asked Fabre—a substitute teacher in a provincial high school—to tutor the son of Napoleon III. True to his solitary character, Fabre declined that honor. (Indeed, in *Souvenirs,* he describes how anxious he was to leave all the while he was at court.) That decision was a smart one, in view of what befell the emperor and his court soon afterward!

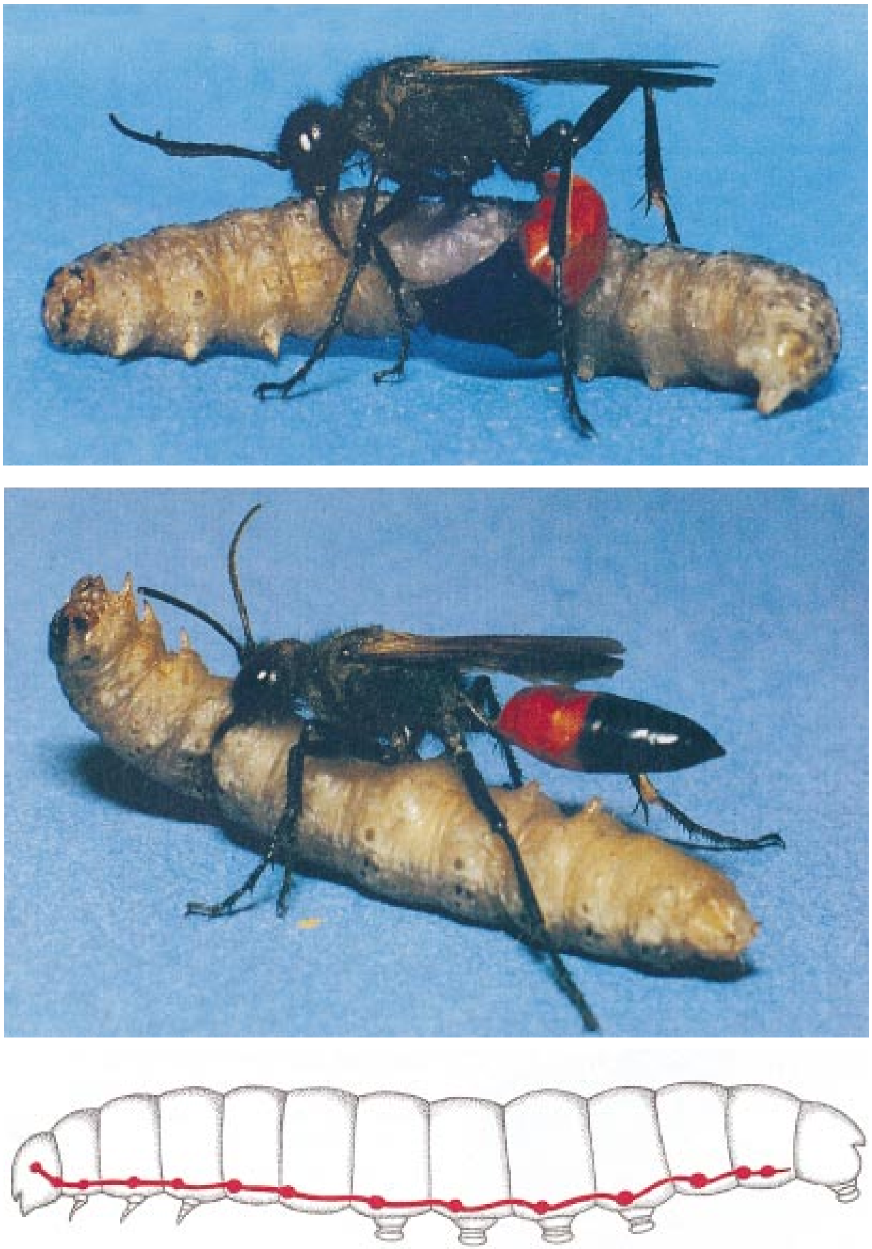
Between that episode and the 1870 Franco-Prussian War, Fabre's firm belief in bettering the condition of women got him into trouble. He had started an immensely popular series of evening lectures for girls. The education of girls, however, had always been the business of the Catholic Church, which did not entirely approve of Fabre's curriculum. He was soon blamed nationwide for such crimes as teaching girls about the sexual functions of flowers. The war ended both the lecture series and the uproar, but in 1871 Fabre and his family were nonetheless turned out of their home by their landladies, two devout Catholic spinsters.

That eviction proved pivotal for Fabre. He had not been happy with his life through that moment: he thoroughly regretted never having fulfilled the dreams of his youth, such as becoming a university professor. The sudden misfortune of being without a home allowed him to change all that.

Thanks to a loan from his wealthy friend John Stuart Mill, Fabre, then just 48 years old, was able to retire. Thereafter he devoted himself to writing, observing nature and experimenting with insects. In 1879, the year that the first volume of *Souvenirs* appeared, the Fabre family moved into L'Harmas, a two-acre fallow lot about six miles west of Orange, which Fabre selected as "a living entomology laboratory." (Today L'Harmas serves as the Fabre Annex of the National Museum of Natural History.)

For more than four decades, until his death in 1915, Fabre continued his work. In addition to *Souvenirs,* he wrote poems in two languages and composed melodies based on them. He also painted 700 graceful, sublimely colored images of mushrooms that are among the most exquisite ever produced.

In retrospect, one can see that if Fabre had realized his dream of becom-



PARASITOID WASPS, such as *Podalonia,* immobilize insects (*top*) by stinging them, then drag them away (*middle*) and lay eggs in their bodies. As Fabre observed, the wasps instinctively “know” their prey’s anatomy (*bottom*): they sting a caterpillar near nerve centers in each segment (*red*) to paralyze without killing it.

ing a university professor, he would have retired no earlier than 1893, which would probably have been too late to launch a research program like the one he started at L’Harmas. At best he would have left only a short equivalent of *Souvenirs*—much to the detriment of humanity’s cultural heritage.

The insect studies that Fabre explored in *Souvenirs entomologiques* represent his major intellectual interest, and for that reason it deserves the fullest attention. I offer a summary of its scientific contents from my perspective as a late 20th-century biologist.

The most noteworthy aspect of Fabre’s work was his emphasis on experimentation. Until Fabre, the few experiments that had ever been conducted on animal behavior were essentially limited to birds. That oversight is not so strange as it may sound. The protocols of biological experimentation, which depend on controlled studies and probabilistic analyses, are not obvious. True experimentation did not appear in the history of science until Galileo. Few biologists have simultaneously possessed a faculty for selecting the right organisms to use and a gift for conceiving tests that would yield results of indisputable meaning.

To be fair, one should note that Fabre was not alone in introducing experimentation to animal behavior studies. The British scientist John Lubbock (1834–1913) pioneered the use of mazes and problem solving in the study of insect behavior, and those research methods have become the foundation for decades of standard laboratory practices involving all kinds of animals. But Lubbock’s greatest influence was on laboratory research (especially in America), whereas Fabre was more influential on studies in natural settings (chiefly in continental Europe).

Entomologists remember Fabre in part for his detailed study of insect metamorphoses. All insects and true bugs begin life as larvae from eggs and end as adults, but in between they pass through several different types of intermediate stages. Some insects, such as silverfish, assume their adult form very early and grow to maturity by enlarging and molting. Grasshoppers and true bugs mature through a process of incomplete metamorphosis, in which the larvae become nymphs that look much like adults but lack wings; only when fully mature do the nymphs become winged adults. Beetles, flies and other insects experience full metamorphosis, in which the larvae change into immobile pupae and later emerge as fully formed winged adults. Fabre examined all these life cycles in detail and also discovered a fourth one. Young blister beetles pass through two or more intermediate larval and pupal stages before completing their metamorphosis into adults. Fabre called this pattern hypermetamorphosis—the only scientific term he ever coined.

Notwithstanding those contributions, Fabre has on the whole not received the full credit from scientific circles that he deserves. Part of the problem is that *Souvenirs* is a literary work with few tables and no diagrams and therefore stood little chance of receiving serious attention from the experts. Researchers who did read it were generally struck by the same virtues that appealed to the public imagination and not always by what was scientifically important.

Scientists have overlooked that Fabre was the first to demonstrate taxis in animals. A taxis is a reflex in which an organism orients its body toward or away from a source of stimulation. That concept is generally attributed to the German-American scientist Jacques Loeb, who developed it in the 1880s as an animal counterpart to tropism in plants—the tendency for plants to grow toward light and against gravity. Yet Fabre recognized some taxes in insects as early as the mid-1870s, and he embarked on a six-year experimenting spree with wild bees to test the idea. He showed, for example, that bees emerging from their pupae have an innate tendency to move upward (against gravity) and toward fresh air, the dominant tendency depending on the species. If 10 *Osmia* bee pupae, for example, were placed inside vertical tubes sealed at the top and open at the bottom, at most three bees near the low opening escaped—the others stubbornly moved to the top, became trapped and eventually died.

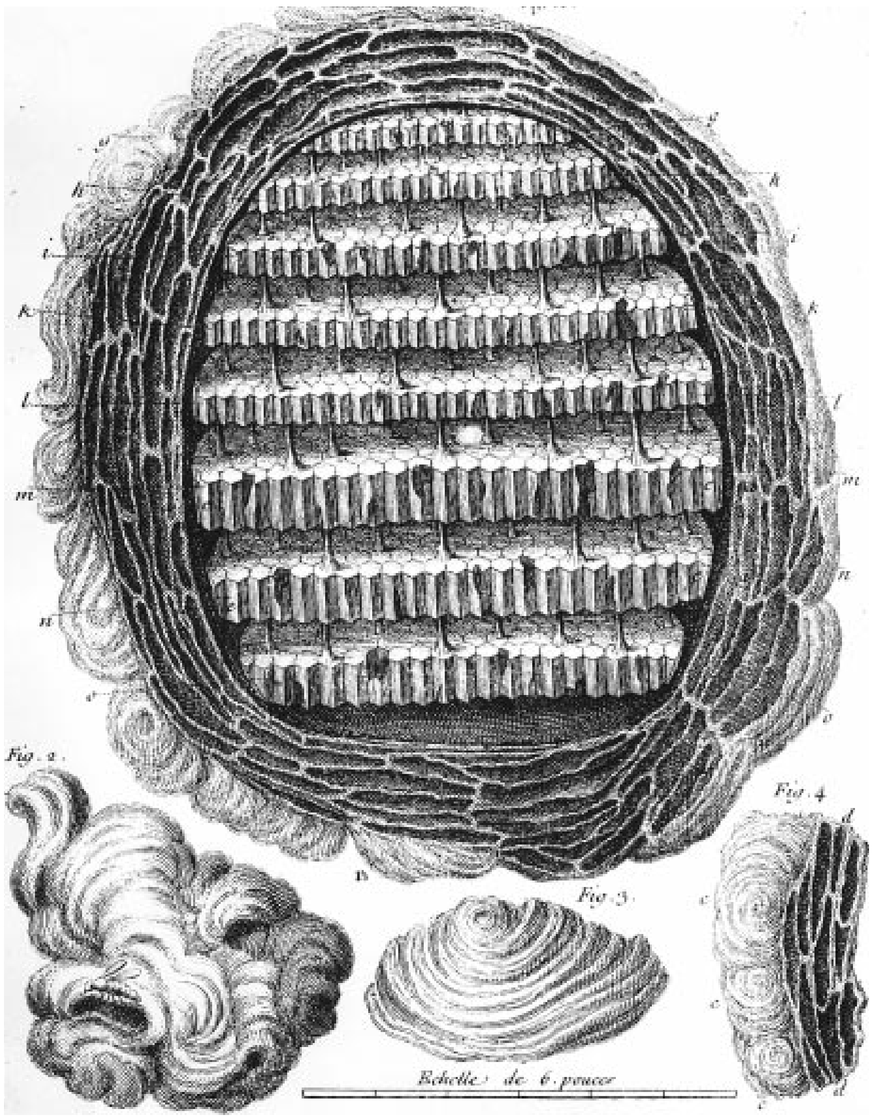
I also know of one instance in which Fabre found the truth about a phenomenon that many modern biologists still misunderstand. The phenomenon is mimicry among *Volucella* flies, which closely resemble bees and wasps and lay their eggs in those insects' nests. The oldest mimicry hypothesis ever published, dating back to 1817, stated that the similarity between some *Volucella* flies and bumblebees helped the flies to function as brood parasites. That is, the bee-mimicking flies could enter the bee nests and lay their eggs without fear of attack. The fly larvae would later eat the nest's contents.

A crucial aspect of this hypothesis is that the presence of the flies in the nest harms the bees. Otherwise, the host bees have no reason to attack their guests. If the bees and flies are not antagonistic, then the best explanation for the flies' mimicry is that it protects them from birds. (Flies and beetles often don the warning colors of wasps to discourage birds.)

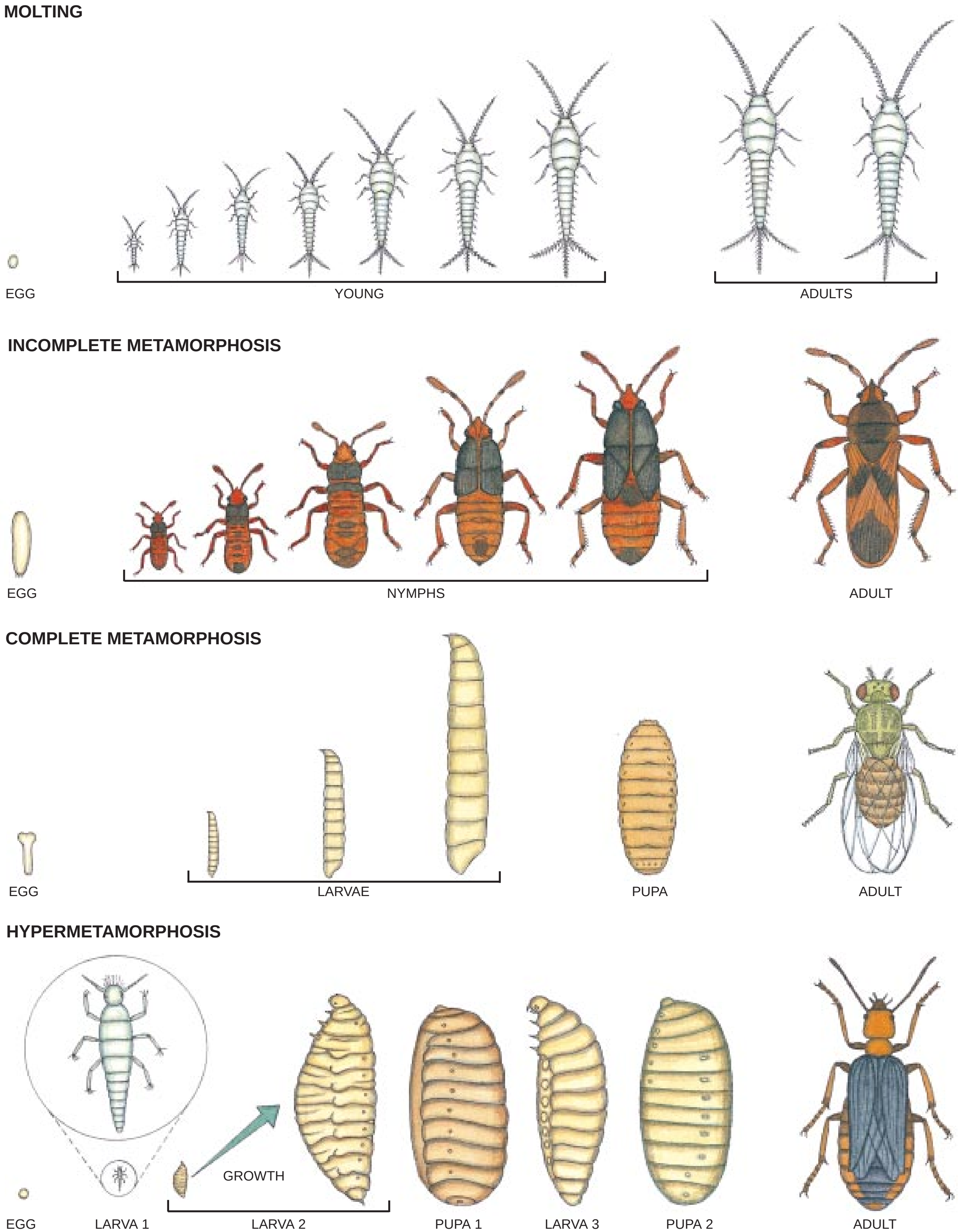
In the 1860s, when many scientists cited mimicry as evidence supporting Darwin's ideas about evolution, that hypothesis was enthusiastically extended to other cases of insect mimicry. Among those was *Volucella zonaria,* whose larvae develop in the nests of European common wasps. But few scientists in either the 19th or 20th century have bothered to check whether the natural history of *V. zonaria* supports the brood parasitism rationale for mimicry.

Fabre was more careful. In *Souvenirs* he describes how the common wasp nest, a completely dark underground structure, is meticulously cleaned by worker wasps, which hurl any debris, dying insects and cadavers into a subjacent charnel pit. Ordinary flies that do not look at all like wasps constantly fly in and out of the nest without being attacked unless they land on the broodcomb, where wasp larvae develop. Any flies unwise enough to go there—including ones that are better mimics than *V. zonaria*—are instantly killed. All these observations and others suggest that visual mimicry is of no benefit to the *Volucella* flies inside the dark nest or at its entrance.

As Fabre also saw, *Volucella* flies safely lay their eggs in the outer layers of the nest, away from the comb. Most of the fly larvae fall into the charnel and there subsist on the bodies of dead wasps. The fly larvae that do enter the nest are ignored by the worker wasps and eventually enter the comb. Fabre found that the fly larvae do not compete for food with the wasp larvae. Instead the flies crawl into the cells housing the wasp larvae, squeeze past them and drink the fluid excretions accumulating at the bottom. The cohabiting fly and wasp larvae never harm one another unless the wasp larva is injured; the fly larva will then start feeding at the injured spot. Generally speaking, fly larvae can survive on any living or rotting animal matter.



***VOLUCELLA* FLIES (*above*) mimic wasps and lay their eggs inside wasp nests (*right*). Many entomologists have mistakenly described the flies as parasites of the wasps. Fabre was the first to realize that the flies and wasps are mutually supportive. The drawing of the wasp nest is by René Antoine F. de Réaumur (1683–1757), whom Fabre admired.**



INSECTS MATURE through a variety of developmental patterns, all of which Fabre observed in the course of his work. Silverfish are essentially adult in shape as larvae; they grow to maturity by molting. True bugs have immature nymph forms before becoming winged adults. Flies experience complete metamorphosis, in which larval grubs become pupae and emerge as adults. Fabre observed all these developmental cycles and also discovered a fourth: hypermetamorphosis among the meloid beetles, which pass through several larval and pupal forms before becoming adults.

Fabre concluded that the relation between the flies and the wasps is not parasitism but mutualism: the *V. zonaria* larvae scavenge for food in the wasp nest and in return help to clean it. Yet more than a century later textbooks continue to assert that *Volucella* flies are probably brood parasites and that the adult mimicry protects the flies from wasp attacks, as though those questions had not been answered otherwise. Fabre's demonstration does not preclude the possibility that other species of flies are wasp parasites nor that *Volucella* larvae clothe themselves in chemical or tactile signals that help them to mimic wasp larvae. Still, modern scientists should be more cognizant of the experiments Fabre has already conducted on their behalf.

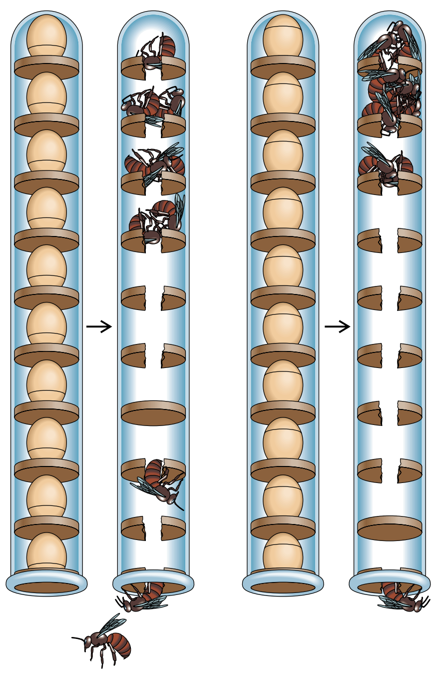
If Fabre himself had been asked what his most important contribution was, he would have pointed to his work distinguishing animal intelligence from that of humans. In his time, a vast body of anecdotes purported to show that animals could reason. Fabre proved, however, that although insects display what look like incredibly sophisticated mental powers in their natural settings, they manifest absolute stupidity when faced with unusual situations. No matter how useless or counterproductive an insect's normal response became in Fabre's experimental settings, the insects responded repeatedly in the same futile way. Those demonstrations, and similar experiments with pets and wild birds, convinced Fabre that the mental essence of humans is fundamentally different from that of animals.

That difference made the behavior of animals no less fascinating to him. Nothing excited Fabre more than the manifestations of instinct among solitary wasps, which gripped his imagination for 50 years. He showed that many solitary wasps paralyze rather than kill their prey and that they lay their eggs in the bodies of the immobilized animal—activities in which the wasps seemed to show an uncannily precise knowledge of their prey's nervous system.

Fabre realized why the parasitic wasps acted as they did. Paralysis keeps the larva's food fresh and safely immobile. The wasp deposits its egg on a soft superficial part of the prey so that the larva can begin to eat far from brain ganglia essential to the prey's survival. A marvel of instinct, the larva eats its way through the body, reaching the brain only at the end of its growth.

Stinging and egg laying have now been observed in hundreds of wasp species, specifically by Andre L. Steiner of the University of Alberta [see "How Parasitic Wasps Find Their Hosts," by James H. Tumlinson, W. Joe Lewis and Louise E. M. Vet; SCIENTIFIC AMERICAN, March 1993]. The bewildering diversity of recorded behaviors defies any all-encompassing explanation. Nevertheless, Fabre was essentially correct about the species he studied. He did notice individual variations in the behavior of wasps, such as the location and number of stings they delivered. Impressed as he was by the permanence of these complex behaviors, however, Fabre saw individual variations as exceptions to fixed patterns in species.

The Darwinian revolution promoted a different view of behavioral variation. To Darwinists, a typical behavior was not a fixed archetype. Rather it was merely the most common behavior found in a population and therefore bound to evolve through natural selection. Yet until the nature of heredity and population genetics was understood, belief in directional selection—that is, consistent selection for certain traits—could not be more than an act of faith. That is why many scientists in the 19th and early 20th centuries who were fully convinced of the fact of evolution could not accept natural selection as its mechanism. Fabre, however, did not separate the two. Because he could not admit natural selection, he rejected evolution, which he saw as a theory for armchair naturalists. Furthermore, the notion of a continuum of intelligence stretching through animals to humans seemed adverse to cherished conclusions of his work.



INSTINCTIVE ORIENTATION by animals, or taxis, was first studied by Fabre. He found, for example, that *Osmia* bees emerging from their pupae primarily move upward and secondarily toward fresh air. If a tube simulating 10 pupal cells was inverted, most of the bees died at the closed top. No more than three, sensing external air, escaped by moving down (*left*). If Fabre rotated the pupae back to their natural orientation, at best one bee escaped (*right*).

Still, *Souvenirs* clearly shows that the subject haunted Fabre—in part because, as he confessed, the mere rejection of evolution "explains nothing." His awareness of paleontology gave him tormenting glimpses of his own inconsistency. Occasionally he could not help expressing himself as an evolutionist, as in this reflection from 1901, which was astonishingly ahead of its time: "I take interest in…a bunch of crow-silks;…I catch a fleeting view in my imagination of ancient times where those conferva [cyanobacteria, or blue-green algae], first born of the plants, roughed out a respirable atmosphere for life to come. What I have under my eyes, between the tiled walls of my trough, tells me the history of this planet wrapping itself in pure air."

Did Fabre ever read *On the Origin of Species?* There is no way to settle that nagging question. By the time L'Harmas became French national property in 1922, Fabre's last son had sold his father's library. In a letter still preserved at L'Harmas, Darwin thanks Fabre for sending him *Souvenirs,* adding, "I do not believe that anyone in Europe has more truly admired your investigations than I have done." Perhaps Darwin returned the gesture by sending Fabre a copy of *Origin.* Since Fabre later mocked Darwin's statement that "a little dose of judgment or reason…often comes into play, even with animals low in the scale of nature," I surmise that Fabre did read at least that far in the chapter on instinct—that is, to the bottom of its first page.

Do not infer from their disagreements that Fabre underestimated Darwin. He probably received more professional letters from Darwin than from anyone else, and he brushed up his English for those occasions. Through correspondence, Darwin participated intensely in Fabre's experiments on homing instincts in mason bees. When Darwin died, Fabre reportedly wept.

Yet Fabre's rejection of evolution as a general concept is a sobering example of how even the brightest of scientists can be blinded by a preconceived idea—a lesson that is especially ironic in Fabre's case because he was obsessed

Excerpts from the Writings of Jean Henri Fabre

Who does not know the magnificent [Great Peacock] Moth, the largest in Europe, clad in maroon velvet with a necktie of white fur? The wings, with their sprinkling of grey and brown, crossed by a faint zig-zag and edged with smoky white, have in the centre a round patch, a great eye with a black pupil and a variegated iris containing successive black, white, chestnut and purple arcs....

"Well, on the morning of the 6th of May, a female emerges from her cocoon in my presence, on the table of my insect laboratory. I forthwith cloister her, still damp with the humours of the hatching, under a wire-gauze bell-jar. For the rest, I cherish no particular plans. I incarcerate her from mere habit, the habit of the observer always on the lookout for what may happen....

"It was a lucky thought. At nine o'clock in the evening, just as the household is going to bed, there is a great stir in the room next to mine. Little Paul, half-undressed, is rushing about, jumping and stamping, knocking the chairs over like a mad thing. I hear him call me:

"'Come quick!' he screams. 'Come and see these Moths, big as birds! The room is full of them!'... In the kitchen I find the servant, who is also bewildered by what is happening and stands flicking her apron at great Moths whom she took at first for Bats....

"We enter the room, candle in hand. What we see is unforgettable. With a soft flick-flack the great Moths fly around the bell-jar, alight, set off again, come back, fly up to the ceiling and down. They rush at the candle, putting it out with a stroke of their wings; they descend on our shoulders, clinging to our clothes, grazing our faces.... Coming from every direction and apprised I know not how, here are forty lovers eager to pay their respects to the marriageable bride born that morning amid the mysteries of my study."

—from *The Life of the Caterpillar,* translated by Alexander Teixeira de Mattos (Dodd, Mead and Company, 1918)

Great Peacock moth

"The more maternity asserts itself, the higher does instinct ascend. In this respect, no creatures are more deserving of our attention than the Hymenoptera [bees, wasps and ants], upon whom the cares of maternity devolve in their fullness. All these favourites of instinct prepare board and lodging for their offspring. They become master-craftsmen in a host of trades for the sake of a family which their faceted eyes will never behold, but which is nevertheless no stranger to the mother's powers of foresight. One turns cotton-spinner and produces cotton-wool bottles; another sets up as a basketmaker and weaves hampers out of bits of leaves; a third becomes a mason and builds rooms of cement and domes of road-metal.... Next comes the provisions for the expected nurselings: piles of honey, loaves of pollen, stores of game, preserved by a cunning paralysing-process....

"So far as the rest of the insect race is concerned, the mother's cares are generally most summary. In the majority of cases, all that is done is to lay the eggs in a favourable spot, where the larva, at its own risk and peril, can find bed and breakfast."

—from the preface to volume 5 of *Souvenirs entomologiques,* quoted in *The Sacred Beetle and Others,* translated by Alexander Teixeira de Mattos (Dodd, Mead and Company, 1918)

with avoiding preconceptions in his work. Indeed, Fabre's animosity toward prejudice extended to an inordinate disdain for theories in general. One cannot practice biology without theories: they arise from the falsifiable hypotheses on which all experiments are based. Theories become bad only when they turn into dogmas.

Unfortunately, dogmatic is what Fabre tended to be. For instance, his rejection of evolution and evolutionary explanations for mimicry completely distorted his studies of the crab spider, *Thomisus onustus.* Put *T. onustus* on any flower, and within a day or two it will obligingly take on the flower's color: white on white flowers, yellow on yellow flowers, pink with purple veins on pink flowers veined with purple. Fabre writes that when prey appear, "the thomise, a bandit in wait under the cover of flowers, comes out of its lurking-place." But the spider never hides: it is always in full view, albeit in an inconspicuous position, and on the flower clusters called spurge inflorescences, it looks like just one more tiny bloom. The "inimitable observer" was misled by his own biases.

In the same way, Fabre rejected the evolutionary explanation for why some insects simulate death. By pretending to be dead, those insects can avoid predation by lizards and insectivorous birds. Fabre decided, however, that insects could not simulate death, because they have no concept of it. He preferred to think that the immobility of the insects was a form of hypnosis "from dread." The chapters in *Souvenirs* on this topic are charming, but Fabre's evidence is subjective, and his experiments are irrelevant.

In a 4,000-page work from another time, weaknesses are sure to appear. We can forgive them in someone who opened a new world of knowledge to his contemporaries and who provided lay readers with an opportunity to criticize the methods and scrutinize the opinions of a successful scientist. Judging from the accomplishments of scientists in Japan, the U.S. and elsewhere who devotedly read Fabre, I would venture that his opposition to the Darwinian theory of evolution has not appreciably hurt his enlightening effect on young minds.

FURTHER READING

THE OXFORD COMPANION TO ANIMAL BEHAVIOUR. Edited by David McFarland. Oxford University Press, 1981.

ACTES DU CONGRÈS JEAN-HENRI FABRE: ANNIVERSAIRE DU JUBILE (1910–1985). Edited by Yves Delange. Leopard d'Or, Paris, 1986.

STINGING BEHAVIOUR OF SOLITARY WASPS. A. L. Steiner in *Venoms of the Hymenoptera: Biochemical, Pharmacological, and Behavioural Aspects.* Edited by T. Piek. Academic Press, 1986.

THE STUDY OF INSTINCT. Niko Tinbergen. Oxford University Press, 1989.

LES CHAMPIGNONS DE JEAN-HENRI FABRE. Edited by Claude Caussanel. Citadelle, Paris, 1991.

Late Ice Age Hunting Technology

Cro-Magnon artisans designed many kinds of spearpoints. By re-creating these weapons, we can better appreciate the sophisticated skill ancient hunters possessed

by Heidi Knecht

Archaeologists try to piece together a complete picture of past events by examining objects left at the scene. Although artifacts buried under layers of sediment for long periods are often decayed, scattered and damaged, they offer a tremendous amount of information about the people who made them and about how and why they were produced. In recent years, archaeologists have increasingly adopted an experimental approach to gain fresh perspective on the clues they unearth. By trying to replicate certain objects and use them as our ancestors may have, researchers can test their theories and gain insight into the course of daily life millennia ago.

Using this method, I investigated the hunting tools and practices of the Cro-Magnons, who lived during the Late Ice Age (40,000 to 12,000 years ago). From studies of Cro-Magnon food debris and artifacts, workers have long known that those ancient people were highly successful at killing large game. All the same, until recently no one had considered how Late Ice Age humans had designed and used their sophisticated hunting weapons. Cro-Magnon arsenals may well have included traps and snares made from wood and plant or animal fibers. These hunters also used spears, consisting of a sharp point (made from stone, antler, bone or ivory) that was hafted, or fastened, to a wooden shaft. For the most part, only the spearpoints themselves have survived. Luckily, archaeologists have recovered such points from many different sites throughout Europe.

I analyzed in detail four distinct types of Late Ice Age spearpoints dating from between 40,000 and 22,000 years ago, hoping to rediscover some of the procedures our ancestors used to make them. Applying those same techniques, I then fashioned spearpoints in the laboratory identical to those found at archaeological sites and tested them in the field. Through these experiments, I have come to realize the specialized knowledge Cro-Magnon hunters must have had not only of their prey but also of the mechanical properties of their weaponry. They dramatically changed the design of their spearpoints during different periods of the Late Ice Age. By gaining a hands-on appreciation of each of these four designs, we can see more clearly the goals that inspired them so very long ago.

Archaeologists assume that Cro-Magnon hunters must have studied the habits of their prey extensively. They needed to know where herds of red deer, horses, ibexes and other large mammals roamed from season to season. They also had to be aware that the value of animals varied at different times of the year. In the early summer, for instance, reindeer may have been infested with insects that would have rendered their hides unsuitable for making clothing and tents. In addition, to take down animals as powerful as bison and mammoths, these hunters needed the skill to craft and use an array of lethal weapons.

During the beginning of the Late Ice Age, artisans most frequently created their deadly spearpoints from bone and antler. Several factors probably governed the selection of these two raw materi-

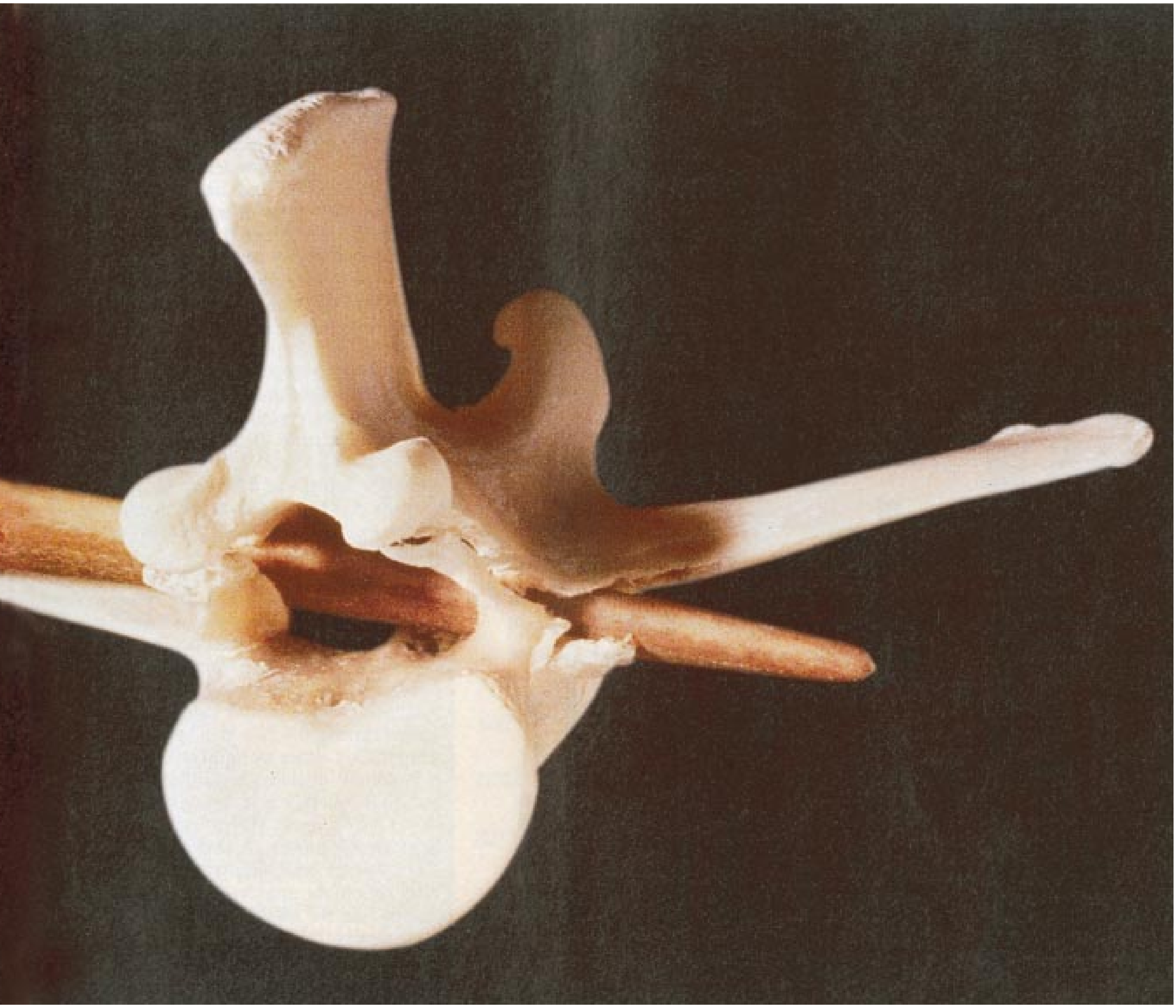
HEIDI KNECHT has done her most recent fieldwork in the southwest of France, excavating Late Ice Age sites and analyzing projectile points. She plans to continue her experimental work under a 1994 research grant from the Higher Council of Archaeological Research in France. She received a Ph.D. in anthropology in 1991 from New York University and currently teaches anthropology at the University of Miami.

PIERCED VERTEBRA harbors a split-based spearpoint that the author made,

als and the decision to use one instead of the other. The first factor was their relative availability. Bones, of course, came part and parcel with the rest of an animal after any successful kill. In contrast, only certain species, such as reindeer and red deer, carry antlers and only during certain seasons. Moreover, antlers are hardest and largest and therefore could supply more durable points in greater numbers during the fall rutting season or after they had been shed. And, except for reindeer, only the males of a species sport antlers.

The decision to use bone or antler for fashioning spearpoints also depended on whether either material was needed for some other purpose. For example, bone was often used to fuel fires kindled for warmth and cooking. Marrow, an excellent source of fat, which is needed to metabolize meat protein, resides in the center of a long bone and could be extracted only by breaking the bones apart. By shattering long bones and then boiling them, the Cro-Magnons could make bone grease—a kind of soup. All these activities destroyed long bones to such an extent that they could not be used for making tools or weapons.

Nevertheless, the choice between antler and bone does not seem to have been based solely on which materials were most available during any one season. Cro-Magnon hunters manipulated these two materials using different methods, in ways that imply that they understood the comparative advantages of each. Antler is a type of bone, but its pattern of growth and development differs dramatically from that of skeletal bone. As a result, antler and bone have distinct structural and mechanical properties. John D. Currey, a biologist at the University of York, has measured the bending strength, elasticity and fracture properties of the two materials and found significant differences between them. When force is applied along the longitudinal axis of bone and antler samples, they show nearly the same bending and compressive strengths. Bone, however, is stiffer and more mineralized, and so it is more



using procedures similar to those practiced during the Late Ice Age. The remarkable durability of this replicated weapon demonstrates that Cro-Magnon artisans had considerable knowledge about the materials they used in their designs.

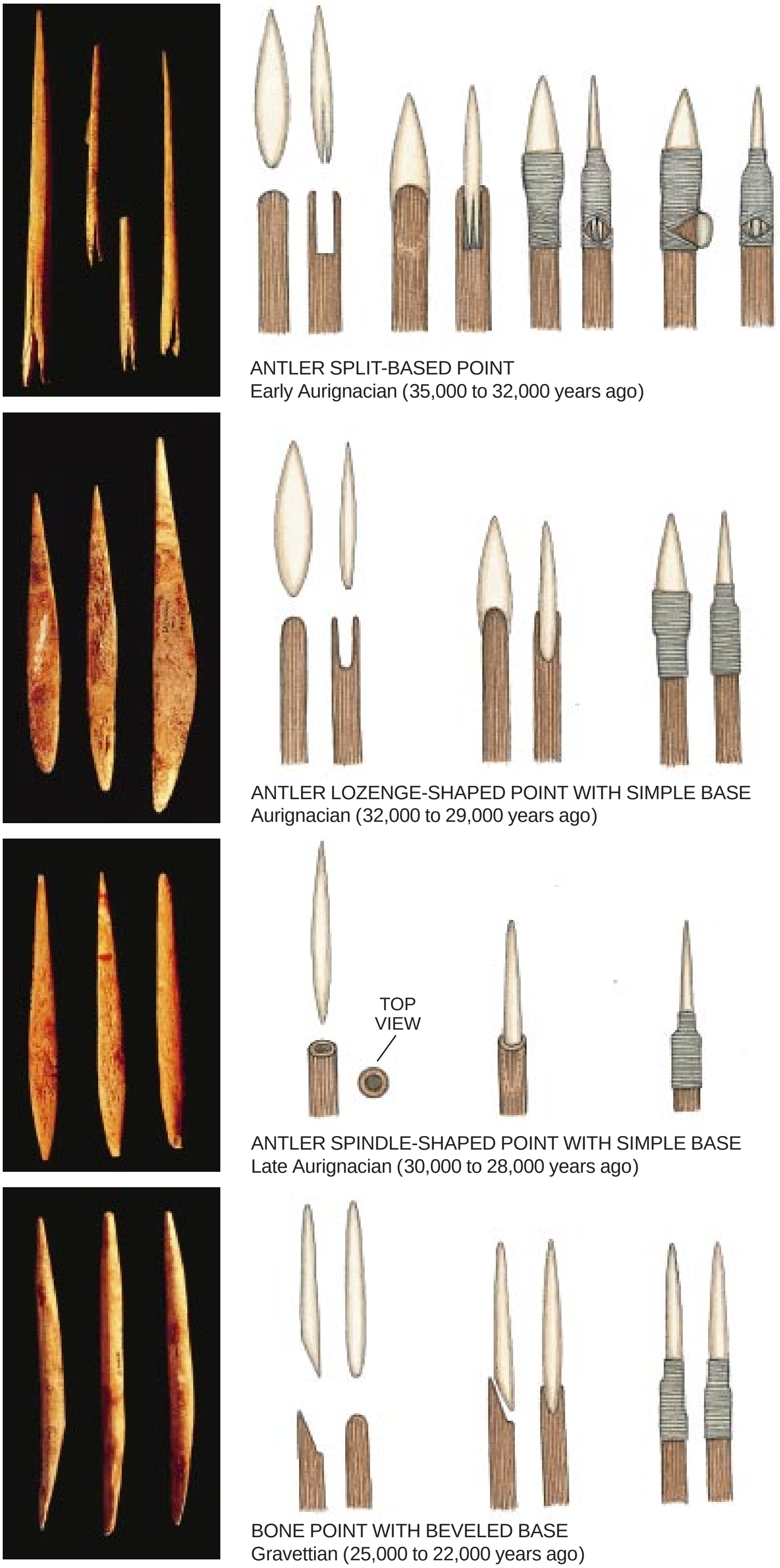
likely to fracture on impact. The fact that Late Ice Age artisans developed specific ways for handling antler and bone does not suggest that they knew why the two products performed as they did. But, through experience, they no doubt would have accumulated quite a bit of knowledge about which of the two was more suitable for a given task.

During the Aurignacian period, 40,000 to 28,000 years ago (the earliest period during which anatomically modern humans were living in western Europe), Cro-Magnon artisans preferred antler for making weapons. They made antler spearpoints in three successive designs. The earliest ones were lozenge-shaped points with split bases; then, beginning around 30,000 years ago, thicker lozenge-shaped points having rounded bases were introduced; and 2,000 years after that, spindle-shaped points with gently tapered bases came into use. During the Gravettian period, which followed the Aurignacian, Cro-Magnon artisans introduced a fourth type of spearpoint, one that was made from bone and had a beveled base [*see illustration at left*].

What criteria guided those who fashioned these highly regular points? To try to answer that question, I first considered what characteristics these four basic variations share. Of necessity, the points are all shaped so that they can pierce the hide of an animal, penetrate muscles and internal organs, and inflict a life-threatening wound. The shapes of the bases, despite their dissimilarities, were also functional. Cro-Magnon artisans developed the different forms so that the points could be firmly fastened to the wooden shaft and thus stay attached on impact.

Of course, the way a point was hafted onto its shaft affected the aerodynamics of the spear as a whole. To avoid making the finished product too heavy or unwieldy, Cro-Magnon hunters needed to take into account how forcefully the average hunter could throw a spear. During the Late Ice Age, spears were hurled either by hand or by atlatl, a handheld stick with a hook at one end that was inserted into a socket at the end of a spear. Prehistoric hunters began using atlatls at least as early as 22,000 years ago. Although archaeologists debate the precise reasons ancient people had for using atlatls, these devices clearly increased the force and accuracy with which a hunter could throw a spear.

All four points, whether thrown by hand or by atlatl, would have served equally well as the lethal end of a spear. Why did their design change over the



SPEARPOINTS changed dramatically during the Late Ice Age in form (*photographs*) and in the way they were hafted onto a shaft (*drawings*). The later designs were more easily repaired. Spindle-shaped points could be reversed if the tip wore out. Because the bevels of the bone points were uniform, they could be interchanged.

course of the Late Ice Age? In search of an explanation, I conducted two series of experiments. First, I used Late Ice Age procedures to manufacture bone and antler spearpoints in the laboratory. I was able to reconstruct step by step the production process used by Cro-Magnons because archaeologists have been fortunate enough to recover spearpoints in different stages of completion, as well as waste and by-products from the process. To check that my method was accurate, at every opportunity I compared the scratches and other marks left on the surface of my objects with those found on archaeological specimens. If the marks did not match, I tried alternative techniques or tools until they did agree. I limited myself to using tools made only from stone, bone and antler. In the end, I was able to fabricate points nearly identical to the genuine artifacts from each period.

In the next stage of my experiments, working with several colleagues on a small farm near Les Eyzies in the southwest region of France, I measured the aerodynamics and durability of the replicated points, as well as how readily they penetrated an animal's hide and bones. We hafted our spearpoints to wooden shafts using strips of animal hide and natural resin and then shot them into goat carcasses suspended in lifelike positions. For the purpose of our exercises, we launched the spears using a calibrated crossbow, which allowed us to control the speed and force with which the spear hit the target. We could therefore be certain that the kinetic energy of the spear would be close to that of a spear thrown either by hand or by atlatl.

We tested each of our spearpoints several times in the hopes of breaking them. One goal of these trials was to determine whether bone and antler points break in telltale ways. Many Ice Age spearpoints excavated from archaeological sites are damaged, and it is often impossible to tell how, or even when, they broke. We learned that if the points were damaged when they hit bones inside the target, bone points usually broke along their midsection. Antler spearpoints, however, showed damage in the form of transverse snaps only near the tip. From such observations, we will be better able to judge whether certain points broke during manufacture or during use or whether instead they were crushed in the ground long after they had been deposited.

We also recorded how much damage the spearpoints did to the bones of the animal. Points made from bone and antler were far more rugged than we had expected. When they entered the animal at a force equivalent to that of a spear released from an atlatl, the points often passed unscathed through vertebrae, ribs and even femurs.

REPLICATED SPEARS were formed by hafting points (these are lozenge-shaped) onto shafts using animal hide and natural resin. Cro-Magnon hunters most likely applied resin to the spearpoint and shaft to increase the adhesion between them.

Producing bone and antler spearpoints similar to the Late Ice Age artifacts proved to be no simple task. In our efforts, it became ever more apparent that even during the earliest phase of the Aurignacian, Late Ice Age workers had mastered a highly coherent parcel of techniques. Over time, their methods changed to meet the needs of Late Ice Age hunting in innovative ways. Indeed, it seems from our exercises that the way in which the spearpoints changed shape over time offered more advantages to hunters, who were often on the move when in search of game. The later points were not only lethal, they were also more easily repaired, and during the Gravettian they were readily interchangeable as well.

At the beginning of the Aurignacian, Cro-Magnon hunters manufactured split-based points by splitting and wedging sections of antler. To extract a morsel of antler, the artisans halved a segment of antler along its length. To make the split in the base of the point, they gently struck a wedge held against a platform constructed at the base of the spearpoint. They controlled the direction and length of the split by shaping the point so that it was wider and thicker in that region where the split should end.

Given that splitting and wedging were common techniques for forming spearpoints, it is not surprising that the hunters apparently used another wedging tactic to haft the split-based points. They inserted the base of a spearpoint into a U-shaped slot cut into the end of a wooden shaft and then wrapped the assembly using a ligature made from plant or animal fiber. Next they forced a small wedge made from antler or wood into the split in the base of the point. Archaeologists have recovered antler wedges of this kind from one archaeological site, the Abri Castanet in the Vallon de Castelmerle, Sergeac, France. The wedge then splayed the wings of the split base open inside the shaft [*see top illustrations on opposite page*]. A combination of the forces acting between the shaft, the point, the wedge and the ligature secured the attachment. Curiously, despite the care it took to execute this design, the makers of split-based points do not seem to have been primarily concerned with repairing their products. Although most split-based points have standardized, symmetrical shapes, reworked points found at archaeological sites appear unbalanced, lopsided or misshapen.

Thirty-two thousand years ago, split-based points suddenly disappeared, and lozenge-shaped points with simple bases took their place. In contrast to their forerunners, the lozenge-shaped points seem to have been designed so that they might be resharpened. To make lozenge-shaped points, the hunters cut a fragment of antler in half along its length. The artisans shaped the surfaces and sides of the lozenge-shaped point separately, as they did when making split-based points. They trimmed the sides of lozenge-shaped points by scraping them lengthwise. Then they scraped the top and bottom surfaces of the points until they were smooth. Because the lozenge-shaped points are much wider in the middle than at either end, the spear shaft probably needed to be slotted so that the girth of the point could extend past the diameter of the shaft. A spear shaft having a diameter as wide as a lozenge-shaped point would have been cumbersome to grasp or throw.

I found that lozenge-shaped points were less difficult to make because there was no need to split the base of the point first. Moreover, these points could easily be resharpened, without even removing them from the shaft. Every time we sharpened one of our points that had been damaged in the field, its form fell within the range of variation found in the archaeological sample. Therefore, the design of lozenge-shaped points may have been a response to a growing desire for easily maintained and reusable spearpoints.

Lozenge-shaped points gradually faded out, and spindle-shaped points emerged. Both the Aurignacian lozenge- and spindle-shaped points were usually made from antler, although occasionally some were fashioned from bone or ivory instead. Unlike the split-based points, they both have simple bases that taper gently to a blunt, rounded end. The artisans, however, could manufacture more spindle-shaped points from the same amount of antler than they could lozenge-shaped points. To form spindle-shaped points, Cro-Magnons scraped a segment of antler along its length. Rather than smoothing the sides of the point separately from shaping the upper and lower surfaces, they removed material from the entire circumference of the spindle-shaped point simultaneously. In other words, spindle-shaped points were formed “in the round.” The spear maker probably inserted the tapered ends of the spindle-shaped points into a socket hollowed in the end of the shaft. In this way, the spindle-shaped points could be hafted very simply, again making no distinction between their surfaces and sides. The symmetrical design looked exactly the same all the way around the perimeter of the point and the shaft.

We resharpened spindle-shaped points attached to the shaft with the same ease that we experienced reworking damaged lozenge-shaped points. Moreover, because the two ends of spindle-shaped points are identical, a reworked spindle-shaped point could also be recycled by turning it around. The original base would then serve as a fresh point. Whereas only one side of a lozenge-shaped point could have been reused, both sides of a spindle-shaped point could have been resharpened repeatedly.

Spindle-shaped points prevailed for the remainder of the Aurignacian. Then, during the Gravettian, there appear single-beveled points made from long bones of large mammals rather than from antlers. The artisans broke the bones into large splinters and shaped them to a gently tapered point by scraping them along their length. Next they whittled or scraped the end of the piece to form the sloped base. The Gravettian designers probably beveled one end of a wooden shaft at the same angle as the bevel on the base of the point. The two faces, pointing toward each other, would then fit together perfectly. To increase adhesion, the spear maker applied resin to each of the beveled surfaces.

In addition, on virtually all single-beveled points, the surface of the bevel is textured and marked either by the pores of the remaining spongy bone or by purposefully scored grooves. This irregular surface further increased the hold between the base and the end of the shaft—smooth surfaces would have slid against each other, and the point could have come loose in the haft. The artisans sometimes flattened and scored the sides of single-beveled points as well, a feature that probably helped to secure the lashing or adhesive even more.

The shift from antler points to bone points is particularly curious. Given that antler is a tougher and more flexible material than bone, spearpoints made from antler are less likely to break. Moreover, according to the different manner in which antler and bone points fracture when they strike an animal or the ground, a bone point loses more of its length with each successive break. After several breaks, a bone point becomes too short to be resharpened. Therefore, it cannot be resharpened as many times as can an

CALIBRATED CROSSBOW enabled experimenters to control the speed and force of a spear in target trials. They set the crossbow so that the kinetic energy of their spears equaled that of a spear thrown by hand or by atlatl. They could then be certain that the replicated weapons would not outperform those used by prehistoric hunters.

ATLATL made from reindeer antler, dates to the Magdalenian (18,000 to 11,000 years ago) and was recovered from the cave of Enlène in France. The hook against which the end of a spear would rest is in the form of a bird. Another bird is carved in bas-relief along the handle. Perhaps a piece of ivory or stone was at one time inserted into the handle to form the bird's eye.

antler point of the same original length.

In designing single-beveled points, however, the artisans did in fact use bone in the best way possible to avoid breakage. They used segments of bone near the epiphyses, the ends of a long bone. Epiphyseal bone exhibits more uniform mechanical properties than does the material found along the middle of a long bone, regardless of the direction in which it experiences stress. As a result, this bone would have been less likely to fracture during impact, penetration and any subsequent movement of the wounded animal. Moreover, single-beveled points attached to a wooden shaft are more streamlined than are their antler predecessors, offering yet another advantage. Indeed, the cross-sectional shape of such a spear remains the same along its entire length, a characteristic that maximizes its elastic resilience.

In our experiments, we had no difficulty retooling broken single-beveled points while they were still attached to the shaft, as was the case for lozenge- and spindle-shaped points. As evidence that Gravettian hunters valued this characteristic, archaeologists have found broken single-beveled points from several Ice Age sites that look as though someone had been in the process of resharpening them. Also, if the single-beveled base of a spearpoint was damaged, we found that another bevel could readily be made on the remaining part of the shaft to accommodate the new haft. Significantly, the angle of the bevel was highly standardized. This uniformity meant that the same spear shaft could host a number of resharpened or new points. A hunter could carry spare points and replace a broken point while away from home on a hunting expedition. Of course, the few stone tools needed to repair a broken point could also be taken along.

These four types of spearpoints are but a small fraction of the myriad kinds that hunters manufactured during the Late Ice Age. In addition to bone and antler points, several types of points were made from stone as well. At sites dating to the Gravettian in southern Germany, archaeologists have found collections of points made from mammoth ribs. In addition, it appears that hunters eventually developed more sophisticated hafting mechanisms that were designed so that the point would detach from the shaft when it hit a target. Left inside the animal, the point could have caused even more internal damage.

Indeed, during different periods of the Late Ice Age, artisans explored a number of strategies to form lethal spearpoints. In their various designs, we can see distinct solutions to the same problem—that of manufacturing highly effective weapons. By continuing to reproduce these weapons and to test their performance capabilities, we can reach a fuller understanding of the manner in which our ancestors tackled the challenges that hunting presented.

FURTHER READING

OSSEOUS PROJECTILE POINTS: BIOLOGICAL CONSIDERATIONS AFFECTING RAW MATERIAL SELECTION AND DESIGN AMONG PALEOLITHIC AND PALEOINDIAN PEOPLES. R. Dale Guthrie in *Animals and Archaeology,* Vol. 1: *Hunters and Their Prey.* Edited by Juliet Clutton-Brock and Caroline Grigson. BAR International Series 163, pages 273–294; 1983.

THE OPTIMAL DESIGN OF HUNTING WEAPONS: MAINTAINABILITY OR RELIABILITY. Peter Bleed in *American Antiquity,* Vol. 51, No. 4, pages 737–747; October 1986.

VISUAL THINKING IN THE ICE AGE. Randall White in *Scientific American,* Vol. 261, No. 1, pages 92–99; July 1989.

HUNTERS OF THE RECENT PAST. Edited by Leslie B. Davis and Brian O. K. Reeves. Unwin Hyman, 1990.

BEFORE LASCAUX: THE COMPLEX RECORD OF THE EARLY UPPER PALEOLITHIC. Edited by Heidi Knecht, Anne Pike-Tay and Randall White. CRC Press, 1993.

HUNTING AND ANIMAL EXPLOITATION IN THE LATER PALEOLITHIC AND MESOLITHIC OF EURASIA. Edited by G. L. Peterkin, H. M. Bricker and P. Mellars. *Archeological Papers of the American Anthropological Association,* No. 4; 1993.

SPEARS, DARTS, AND ARROWS: LATE WOODLAND HUNTING TECHNIQUES IN THE UPPER OHIO VALLEY. Michael Shott in *American Antiquity,* Vol. 58, No. 3, pages 425–443; July 1993.

Can Science Explain Consciousness?

by John Horgan, *senior writer*

What was once the greatest mystery of biology, the human brain, is gradually yielding its secrets. Investigators are probing its deepest recesses with increasingly powerful tools, ranging from microelectrodes, which can discern the squeaks of individual neurons, to magnetic resonance imaging and positron emission tomography, which can amplify the cortical symphony evoked by, say, viewing Georges Seurat's *La Grande Jatte* or sniffing barbecued spareribs. With these and other techniques, researchers have begun elucidating the physiological processes underlying such facets of the mind as memory, perception, learning and language.

Emboldened by these achievements, a growing number of scientists have dared to address what is simultaneously the most elusive and inescapable of all phenomena: consciousness, our immediate, subjective awareness of the world and of ourselves. Francis Crick should receive much of the credit—or blame—for the trend. Crick, who shared a Nobel Prize for the discovery of DNA's structure in 1953, turned to neuroscience shortly after he moved from England to the Salk Institute for Biological Studies in San Diego almost 20 years ago. Just as only the late Richard M. Nixon, famous for his red-baiting, could reestablish relations with communist China, so only Crick, who possesses a notoriously hard nose, could make consciousness a legitimate subject for science.

In 1990 Crick and Christof Koch, a young neuroscientist at the California Institute of Technology who collaborates closely with Crick, proclaimed in *Seminars in the Neurosciences* that the time was ripe for an assault on consciousness. They rejected the belief of many of their colleagues that consciousness cannot be defined, let alone studied. Consciousness, they argued, is really synonymous with awareness, and all forms of awareness—whether involving objects in the external world or highly abstract, internal concepts—seem to involve the same underlying mechanism, one that combines attention with short-term memory.

Contrary to the assumptions of cognitive scientists, philosophers and others, Crick and Koch asserted, one cannot hope to achieve true understanding of consciousness or any other mental phenomenon by treating the brain as a black box—that is, an object whose internal structure is unknown and even irrelevant. Only by examining neurons and the interactions between them could scientists accumulate the kind of empirical, unambiguous knowledge that is required to create truly scientific models of consciousness, models analogous to those that explain transmission of genetic information by means of DNA.

Crick and Koch urged that investigators focus on visual awareness, since the visual system has already been well mapped in both animals and humans. If researchers could find the neural mechanisms underlying this function, they might unravel more complex and subtle phenomena, such as self-awareness, that may be unique to humans (and therefore much more difficult to study at the neural level). We may even comprehend why we have the paradoxical sensation of free will, an ineradicable sense that our minds exist independently of and exert control over our bodies. Crick elaborates on these ideas in *The Astonishing Hypothesis,* a book published this year and dedicated to Koch.

Crick's exhortations have helped incite an intellectual stampede in which mainstream researchers jostle with philosophers, computer scientists, psychiatrists and other distinctly exotic species thirsting for insights into the mind. Meetings are proliferating. In April more than 300 investigators gathered at the University of Arizona Health Sciences Center in Tucson for a meeting entitled "Toward a Scientific

What is consciousness? Can neurobiology explain it, or—as some philosophers argue—does this most elusive and inescapable of all phenomena lie beyond experiment's reach?

FRANCIS CRICK of the Salk Institute for Biological Studies has spurred interest in consciousness by declaring it to be a legitimate subject for science.

Basis for Consciousness." The annual conference of the Society for Neuroscience, the field's largest and most prestigious meeting, will host its first symposium on consciousness when it convenes in Miami in November. New publications have sprung up to feed the burgeoning interest, including *Psyche,* an E-mail journal based in Australia, and *Journal of Consciousness Studies,* a British quarterly scheduled for launching this summer.

Of course, neuroscientists are still far from agreeing on how consciousness should be studied or even defined. One prominent worker claims to have already "solved" consciousness: Gerald M. Edelman of the Scripps Research Institute, who shared a Nobel Prize in 1972 for research on antibodies. Edelman contends that our sense of awareness stems from a process he calls neural Darwinism, in which groups of neurons compete with one another to create an effective representation of the world. Edelman has promulgated this theory in a series of books—most recently, *Bright Air, Brilliant Fire,* published in 1992.

Crick has accused Edelman of dressing up ideas that are not terribly original in an idiosyncratic and obscure jargon. Most neuroscientists agree with this assessment (and find ludicrous the suggestion in a recent *New Yorker* magazine profile that Edelman might garner a second Nobel Prize for this work). But even those who admire Crick's efforts suspect his outlook might be too narrow. Gerald D. Fischbach of Harvard University, a former president of the Society for Neuroscience, says that it is not clear whether the kind of "electrophysiological" theory called for by Crick would suffice to explain consciousness, in the sense that discovering the structure of DNA accounted for heredity. "I don't think the field is mature enough to answer that question yet," he says.

HARD-CORE MYSTERIAN Colin McGinn, a philosopher at Rutgers University, believes consciousness and its attendant conundrums, such as free will, are mysteries too profound for humans to plumb, scientifically or otherwise.

The New Mysterians

Tomaso Poggio of the Massachusetts Institute of Technology, an authority on perception who was Koch's thesis adviser, thinks Crick may place too much emphasis on mechanisms that might coordinate, or bind together, the firings of neurons responding to a visual scene. Conversely, Crick may unduly neglect the role that the brain's plasticity, or ability to change its circuitry, might play in creating consciousness and other aspects of the mind, according to Poggio. Antonio R. Damasio of the University of Iowa, who maps our mental faculties by studying brain-damaged patients, holds that because a theory of consciousness must show how each of us acquires a sense of self, it must take into account not just the brain but the entire body. Damasio also believes that because consciousness is shaped by an individual's interactions with the environment and with other people, a neural model of consciousness will probably have to be supplemented by cognitive and social theories.

As neuroscientists debate these issues among themselves, others have challenged whether conventional neuroscience—despite its success in illuminating other attributes of the mind—can ever account for consciousness. Members of this eclectic group hail primarily from traditions outside mainstream neuroscience, such as physics and philosophy. Such individuals often seem less interested in clarifying consciousness than in mystifying it. For that reason, Owen Flanagan, a philosopher at Duke University, has dubbed some doubters "the new mysterians" (after the 1960s rock group Question Mark and the Mysterians, who performed the hit song "96 Tears").

One contingent of mysterians, whose most prominent member is Roger Penrose, a physicist at the University of Oxford, proposes that the mysteries of the mind must be related to the mysteries of quantum mechanics, which generates nondeterministic effects that classical theories of physics (and neuroscience) cannot. Although at first ignored and then derided by conventional neuroscientists, this alternative has steadily won popular attention through Penrose's efforts.

Another group of mysterians, which consists for the most part of philosophers, doubts whether any theory based on strictly materialistic effects—quantum or classical—can truly explain how and why we humans have a subjective experience of the world. "The question is, how can any physical system have a conscious state?" says Jerry A. Fodor, a philosopher at Rutgers University. Scientists who think science alone can answer that question "don't really understand it," Fodor declares.

None of these philosophers advocates dualism, a philosophy that holds that the mind exists independently of and can influence matter. But they reject the hard-core materialism of Crick, who claims in his new book that "your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules." Such a framework, they say, is inadequate for comprehending mental phenomena. Some other theory is required to make the relation between matter and mind "transparent," as the philosopher Thomas Nagel of New York University has put it.

Both these philosophical views and the quantum-consciousness theories are rejected by Terrence J. Sejnowski, a neural-network researcher at the Salk Institute. "I call these arguments from ignorance," says Sejnowski, who regularly shares tea with Crick and applauds his efforts to make consciousness a scientific subject. Sejnowski thinks that by following Crick's strict program neuroscientists "might actually get somewhere" in addressing biology's most profound riddle. Pointing out that life once seemed impossibly complex—before the discovery of DNA's structure revealed how information is passed from one generation to another—Sejnowski argues that much of the mystery veiling the mind will evaporate once scientists learn more about how the brain works.

"Just Like Woodstock"

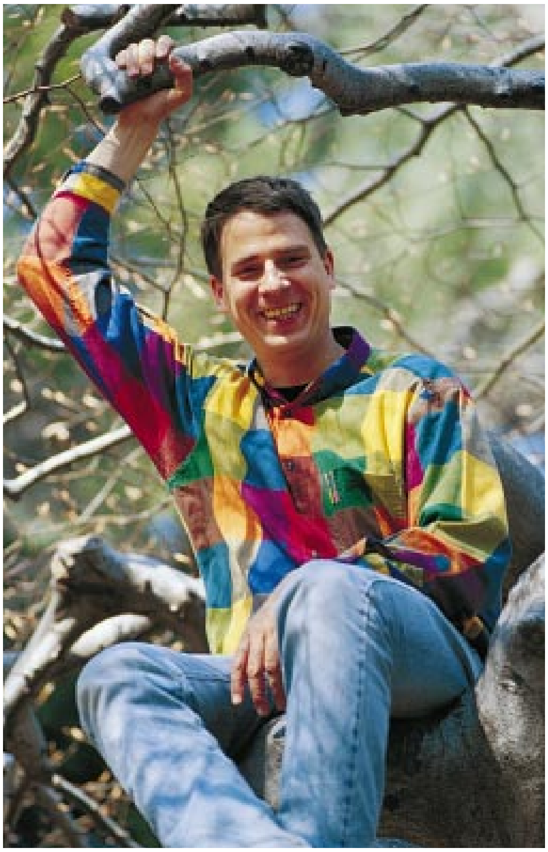
Still, Sejnowski expresses the hope that the challenges from the philosophical skeptics and the quantum-consciousness claque, while misguided, may spur neuroscientists to be more creative as well as rigorous in their own research on consciousness. "When you get these wild disagreements," he says, "you also have an opportunity."

By that criterion, there should have been opportunities galore at the recent meeting on consciousness at the University of Arizona, billed as "the first interdisciplinary scientific conference on consciousness." "Wow, it's just like

Woodstock," one speaker exclaims as he scans an auditorium crammed with neuroscientists, philosophers, psychiatrists, cognitive scientists and others who defy categorization. Culture clashes abound. Indeed, the meeting offers a snapshot of a field—call it consciousness studies—in the throes of creation.

Crick is not here, but Koch is. So is Steen Rasmussen, a biochemist and computer scientist from the Santa Fe Institute, headquarters of the trendy fields of chaos and complexity. He suggests that the mind may be an "emergent"—that is, unpredictable and irreducible—property of the brain's complex behavior, just as James Joyce's *Ulysses* is a surprising outcome of applying the rules of spelling and grammar to the alphabet.

Other investigators offer more radical takes on consciousness. Brian D. Josephson of the University of Cambridge, who won a Nobel Prize in 1973 for discovering a subtle quantum effect that now bears his name, calls for a unified field theory that can account for mystical and even psychic experiences. Andrew T. Weil, a physician at the University of Arizona who is an authority on psychedelia, asserts that a complete theory of mind must address the reported ability of South American Indians who have ingested psychedelic drugs to experience identical hallucinations.



CHRISTOF KOCH of the California Institute of Technology, Crick's collaborator, suggests that when it comes to consciousness, philosophers should heed the advice of Ludwig Wittgenstein: "Whereof one cannot speak, thereof one must be silent."

Just when one thinks that one has encountered every possible "paradigm" (a term much abused at this conference), another rears its head. After a speaker describes human thoughts as "quantum fluctuations of the vacuum energy of the universe"—which "is really God," she assures her listeners—an audience member stands to proclaim that physicists are discovering profound links between information, thermodynamics and black holes, and these findings may help unlock the mysteries of consciousness. "There's not a black hole in our brains," he adds, "but—" "I think there *is* a black hole in my brain!" an overloaded listener interrupts.

Those who find lectures in the auditorium too staid can forage for more nonlinear fare in the hallway outside. "That's where all the really interesting stuff is," confides Spiros Antonopoulos, a journalist who sports a nose ring and chin braid and is covering the meeting for the cyberpunk magazine *Fringeware Review.* There one can join arguments over whether only humans are conscious or whether computers, bats or even paramecia share this trait.

The conference does not lack empirical findings. Among the most intriguing are those involving patients whose sense of awareness has been damaged by some trauma or disease. Several investigators have studied patients who display the strange condition known as blindsight: they respond to visual stimuli—even catching a ball tossed to them, for example—while insisting that they cannot see anything.

Victor W. Mark, a neurologist from the University of North Dakota, shows a videotape of a young woman suffering from epilepsy so debilitating that to relieve it surgeons severed her corpus callosum, the bundle of neurons linking the two hemispheres of her brain. Although the operation alleviated her seizures, she was left with two centers of consciousness vying for dominance. When asked if her left hand feels numb, she shouts, "Yes! Wait! No! Yes! No, no! Wait, yes!" her face contorted as each of her two minds, only one of which can feel the hand, tries to answer. The researcher then hands her a sheet of paper with the words "yes" and "no" written on it and tells her to point to the correct answer. The woman stares at the sheet for a moment. Then her left forefinger stabs at "yes" and her right forefinger at "no."

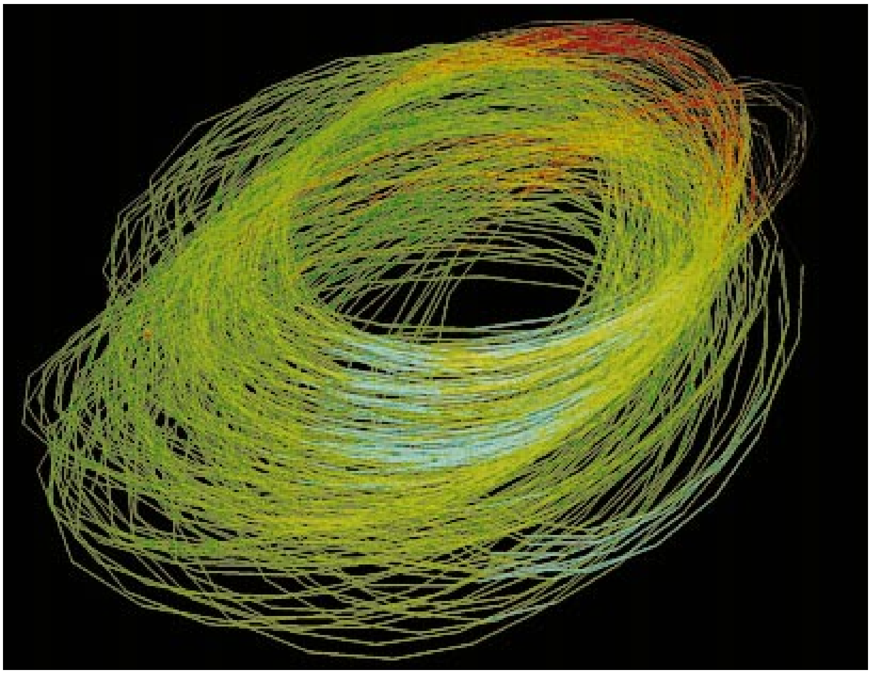
To materialists such as Crick, the significance of such phenomena is obvious: when the brain is injured in certain ways, consciousness (and not necessarily perception) is impaired. Clearly, consciousness has no existence independent of what has been called the "meat machine" but is lodged firmly within it. Some audience members, however, extract different messages from Mark's video presentation, as if it were a Rorschach blot. One listener suggests that even healthy individuals experience some fracturing of the self—albeit in a much less dramatic form. A psychiatrist then wonders aloud whether the woman's two selves could be trained to get along better with each other through conflict resolution.

The Binding Problem

Koch, during his presentation, strives to bring the discussion down to earth. Striding restlessly around the stage in jeans and scarlet lizard-skin cowboy boots, he delivers a high-speed, German-accented summary of the message he and Crick have promulgated for the past four years: scientists should concentrate on questions that can be experimentally resolved and leave metaphysical speculations to "late-night conversations over beer." To emphasize this point, Koch reiterates a line of the "philosopher" Clint Eastwood: "A man has gotta know his limitations."

He then details his and Crick's argument that consciousness stems from a process that combines attention with short-term memory. (Koch notes that the turn-of-the-century philosopher William James first had this insight.) The phenomenon of attention involves more than simple information processing, Koch observes. To demonstrate this point, he shows a slide of a pattern that can be viewed either as a vase or as a pair of human profiles facing each other. Although the visual input to the brain remains constant, the pattern that one is aware of, or attends to, keeps changing. What neural activity corresponds to the change in attention?

The answer to this question is complicated by the fact that "there is no



CHAOTIC FIRING of large groups of neurons may allow us to respond rapidly to—and become aware of—new perceptions, according to Walter J. Freeman of the University of California at Berkeley. This image is an abstract "phase portrait" of the olfactory cortex's electroencephalographic response to a smell.

single place where everything comes together" in forming a perception; even a single scene is processed by different neurons in different parts of the brain. One must therefore determine what mechanism transforms the firing of neurons scattered throughout the visual cortex into a unified perception. "This is known as the binding problem," Koch explains, noting that it is considered by many neuroscientists to be the central issue of their field.

One possible answer to the binding problem has been suggested by experiments on animals showing that neurons in different regions of the brain occasionally oscillate at the same frequency, roughly 40 times per second. Koch asks the audience to imagine the brain as a Christmas tree with billions of lights flickering apparently at random. These flickerings represent the response of the visual cortex to, say, a roomful of people. Suddenly, a subset of lights flickers at the same frequency, 40 times per second, as the mind focuses on the face of an old flame.

Koch concedes that the evidence implicating 40-hertz oscillations in awareness is tenuous; it has shown up most clearly in anesthetized cats. Another form of binding could be simple synchrony: neurons merely fire at the same time and not necessarily the same frequency. Evidence for synchrony is also trickling in from animal experiments.

After Koch's lecture, Valerie Gray Hardcastle, a philosopher at Virginia Polytechnic Institute and State University, takes the podium to proclaim that "simple solutions to the binding problem must fail." Pointing out that monitoring individual neurons may yield spurious associations, she suggests that scientists would do better to examine the behavior of populations of neurons or even of the entire brain.

Walter J. Freeman, a lean, white-bearded professor at the University of California at Berkeley, raises similar objections to Koch's remarks. Freeman's criticism is significant, because he was one of the first prominent scientists to investigate 40-hertz oscillations. He asserts that neither oscillations nor synchrony will play more than a minor role in the solution to the binding problem and that "the current wave of enthusiasm is misplaced."

Freeman advocates a more complex approach to the binding problem. Large groups of neurons, he explains, display chaotic behavior; that is, their firing seems random but actually contains a hidden order. Like all chaotic systems, these neural patterns are extremely sensitive to minute influences. The sight of a familiar face, therefore, can trigger an abrupt shift in the firing pattern corresponding to a shift in one's awareness. Freeman concedes, however, that even his theory is only—at best—one piece of the puzzle.

Freeman thinks another piece of the puzzle may be supplied by Benjamin Libet of the University of California at San Francisco, one of the few researchers who has studied consciousness by conducting extensive experiments on humans. Libet is a sharp-featured man with the narrow-eyed, somewhat defensive mien of someone who has had to fight hard for acceptance of his work. (Crick, in his new book, recalls that Libet once confided that he dared to study consciousness in human subjects only after receiving tenure.)

Timing Is Everything

One set of experiments done by Libet involved patients in whose cortices electrodes had been implanted for medical reasons. Libet delivered a mild electrical pulse both to these electrodes and to the skin of the subjects. In either case, the subjects became aware of pulses only if they lasted for more than 0.5 second. When Libet stimulated the cortical neurons as much as 0.5 second before stimulating the skin, the subjects reported, paradoxically, having felt the skin pulses first.

Libet theorizes that once the volunteers became aware of the stimulation of their skin, they experienced the sensation as if they had been aware of it from the beginning and not after 0.5 second. We subjectively compensate for the time lag in our awareness of tactile sensations, Libet contends, through a process he calls "backward referral in time." He likens this ability to the one that allows an observer moving past a picket fence to maintain a constant image of the house behind it.

Libet then describes an equally surprising set of experiments on healthy subjects whose brain waves were monitored by an electroencephalograph (EEG) machine. They were instructed to flex a finger at a moment of their choosing while noting the instant of their decision as indicated on a clock. The volunteers took some 0.2 second to flex after they had decided to do so. But, according to the EEG, the subjects' brains displayed neural activity 0.3 second *before* they decided to act. In a sense, the brain made the decision to move before the mind became conscious of it. "The actual initiation of volition may have begun even earlier in a part of the brain we weren't monitoring," Libet comments.

After Libet delivers his carefully crafted speech, one audience member asks whether his findings bear on the question of free will. "I've always been able to avoid that question," Libet replies with a grimace. He proposes that we may exert free will not by initiating intentions but by vetoing, acceding or otherwise responding to them after they arise.

Other observers fault Libet for overgeneralizing from his data. Flanagan, the Duke philosopher who coined the term "mysterians," points out that, strictly speaking, Libet's subjects were not acting of their own free will, since they had been instructed in advance to flex their fingers. Flanagan adds that Libet's time-lag findings may hold true only for tactile responses and not for other sensory modes.

In fact, Flanagan thinks there may be many modes of consciousness; our awareness of an odor, for example, not only stems from a different set of neurons but also is in some sense qualitatively different from our visual awareness. Flanagan contends that neuroscientists must therefore resist the temptation to look for any single mechanism—such as 40-hertz oscillations or Freeman's chaotic neural behavior or Libet's time-delay factor—that accounts for consciousness.

In his 1992 book *Consciousness Reconsidered,* Flanagan argues on behalf of a philosophy called constructive naturalism, which holds consciousness to be a common biological phenomenon occurring not only in humans but in many other animals—and certainly all the higher primates. Other adherents to this position include Daniel C. Dennett of Tufts University (author of *Consciousness Explained,* also published in 1992) and Patricia S. Churchland of the University of California at San Diego. "We say you can acquire knowledge of consciousness by triangulation," Flanagan remarks, that is, by combining neural and psychological data from experiments on humans and animals with subjective reports from humans.

Quantum Microtubules

All these philosophers—and most neuroscientists—are united in their skepticism that consciousness depends in some important way on quantum effects. Since at least the 1930s some physicists have speculated that quantum mechanics and consciousness might be linked. They based their speculation on the principle that the act of measurement—which ultimately involves a conscious observer—has an effect on the outcome of quantum events. Such notions have generally involved little more than hand waving, but they have become more prominent lately because of Penrose.

Penrose has earned respect as an authority on general relativity and for inventing the geometric forms known as Penrose tiles, which fit together to form quasiperiodic patterns. In his 1989 best-seller, *The Emperor's New Mind,* he vigorously attacked the claim of artificial-intelligence proponents that computers can replicate all the attributes of humans, including consciousness. A sequel, *Shadows of the Mind,* will be published this fall.

An elfin man with a shock of black hair who manages to seem simultaneously distracted and acutely alert, Penrose summarizes the new book's themes at the Tucson conference. He first tells a story about how Deep Thought, a computer that has beaten some of the world's best chess players, was stumped by an endgame situation that even a clever amateur player would have known how to handle. Penrose concludes that "what computers can't do is understand."

The key to Penrose's argument is Gödel's theorem, a 60-year-old mathematical demonstration that any moderately complex system of axioms yields statements that are self-evidently true but cannot be proved with those axioms. The implication of the theorem, according to Penrose, is that no deterministic, ruled-based system—that is, neither classical physics, computer science nor neuroscience—can account for the mind's creative powers and ability to ascertain truth.

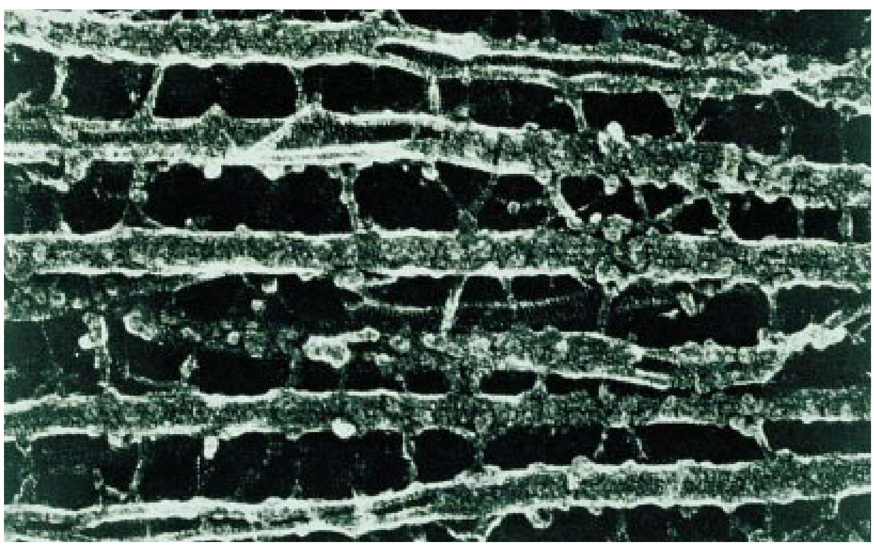
In fact, Penrose thinks the mind must exploit nondeterministic effects that can be described only by quantum mechanics or "a new physical theory that will bridge quantum and classical mechanics and will go beyond computation." He even suggests that nonlocality, the ability of one part of a quantum system to affect other parts instantaneously (Einstein dubbed it "spooky action at a distance"), might be the solution to the binding problem.

Although Penrose was once rather vague about where quantum effects work their magic, he now hazards a guess: microtubules, minute tunnels of protein that serve as a kind of skeleton for cells, including neurons. This endorsement delights Stuart R. Hameroff, an anesthesiologist at the University of Arizona, who organized the Tucson meeting and is the leading proponent of the microtubule hypothesis.

Hameroff, an aging hipster with a goatee and ponytail, manages to squeeze a remarkable number of scientific buzzwords into his talk on microtubules: emergent, fractal, self-organizing, dynamical. He claims to have found evidence that anesthesia eliminates consciousness by inhibiting the movement of electrons in microtubules.

Erecting a mighty theoretical edifice on this frail claim, he proposes that microtubules perform nondeterministic, quantum-based computations that "somehow" give rise to consciousness. Each neuron is thus not a simple switch but "a network within a network," Hameroff elaborates. He acknowledges that microtubules occur in most cells, not just neurons, but the implications of this fact do not faze him. "I'm not going to contend that a paramecium is conscious, but it does show pretty intelligent behavior," he says.

Other quantum-consciousness devotees make Penrose and Hameroff look like paragons of rigor. For example, Ian



MICROTUBULES, minute tunnels of protein serving as a kind of skeleton in almost all cells (not just neurons), may generate quantum effects crucial to consciousness, according to Roger Penrose, a physicist at the University of Oxford, and Stuart R. Hameroff, an anesthesiologist at the University of Arizona.

N. Marshall, a British psychiatrist, presents what he believes is evidence that thought stems from quantum effects. He and several colleagues claim to have found that the ability of subjects to perform simple tests while hooked up to an EEG machine varies depending on whether that machine is plugged in or not. Their conclusion? When the EEG machine is turned on, it "observes" the brain and therefore alters its thoughts—just as observing an electron alters its course. In other words, Heisenberg's uncertainty principle applies to the entire brain.

One listener appalled by such assertions is John G. Taylor, a physicist and neural-network specialist at King's College London. He insists that all the quantum-consciousness enthusiasts, and even Penrose, ignore the most basic facts about quantum mechanics. For example, nonlocality and other quantum effects they have seized on as vital to consciousness are observed only at temperatures near absolute zero—or at any rate far below the ambient temperatures of the brain. Like most neuroscientists, Taylor also objects to the quantum approach on pragmatic grounds. Before turning to the extreme reductionist approach advocated by Penrose and others, researchers should explore possibilities that are more plausible and experimentally accessible—and that have already proved successful for explaining certain aspects of perception and memory. "If that fails, then maybe we should look elsewhere," he adds.

DAVID J. CHALMERS of Washington University holds that philosophy must bridge the "explanatory gap" between a physical theory of consciousness and our subjective experience.

The Explanatory Gap

One of the few points of view missing at the Arizona meeting is that of Colin McGinn, a philosopher at Rutgers University and perhaps the most unflinching of all mysterians. In his 1991 book *The Problem of Consciousness,* McGinn argued that because our brains are products of evolution, they have cognitive limitations. Just as rats or monkeys cannot even conceive of quantum mechanics, so we humans may be prohibited from understanding certain aspects of existence, such as the relation between mind and matter. Consciousness, in other words, will remain forever beyond human understanding, according to McGinn.

At least one philosopher at the Arizona meeting veers dangerously close to this glum conclusion. David J. Chalmers, an Australian at Washington University who bears an uncanny resemblance to the subject of Thomas Gainsborough's painting *Blue Boy,* agrees with McGinn that no strictly physical theory—whether based on quantum mechanisms or neural ones—can explain consciousness.

All physical theories, Chalmers claims, can describe only specific mental *functions*—such as memory, attention, intention, introspection—correlating to specific physical processes in the brain. According to Chalmers, none of these theories addresses the really "hard" question posed by the existence of the mind: Why is the performance of these functions accompanied by subjective experience? After all, one can certainly imagine a world of androids that resemble humans in every respect—except that they do not have a conscious experience of the world.

Science alone cannot supply an answer to this question, Chalmers declares. Unlike McGinn, however, Chalmers holds that philosophers can and must construct a higher-level theory to bridge that "explanatory gap" between the physical and subjective realms. In fact, Chalmers has such a theory. He asserts that just as physics assumes the existence of properties of nature such as space, time, energy, charge and mass, so must a theory of consciousness posit the existence of a new fundamental property: information. The concept of information, Chalmers explains, has aspects that are both physical and "phenomenal" (a philosopher's term that is roughly equivalent to "experiential" or to "subjective").

Koch finds such arguments irksome. He notes that if everyone shared the belief of McGinn and Chalmers that science cannot solve consciousness, the prophecy would be self-fulfilling. Science may not be able to resolve all the mysteries of the mind, Koch concedes, but philosophy has a much slimmer chance of providing lasting insights about the mind-body problem or the question of free will. He adds that when considering such ancient conundrums, philosophers should heed the advice offered by their illustrious forebear Ludwig Wittgenstein: "Whereof one cannot speak, thereof one must be silent."

Yet of all the outcomes of the surging interest in consciousness, the least likely is silence. The evening after Chalmers gives his speech in Tucson (which is extremely well received), Koch confronts him to complain that his "double-aspect theory of information" is untestable and therefore useless. "Why don't you just say that when you have a brain the Holy Ghost comes down and makes you conscious!" Koch exclaims. Such a theory is unnecessarily complicated, Chalmers responds drily, and it would not accord with his own subjective experience. "But how do I know your subjective experience is the same as mine?" Koch sputters. "How do I even know you're conscious?"

Later, Koch and Chalmers head off to the hotel bar to continue their discussion over beer. They come to a reconciliation of sorts. Koch expresses interest in Chalmers's views on computation and cognition—a paper on which Chalmers just happens to have in his knapsack. Chalmers concedes that perhaps neuroscience may still provide direction and inspiration for philosophy.

The field of consciousness studies is struggling with its own binding problem. Yet from encounters such as this, progress may emerge—just as the interactions between neurons rather than their individual properties give rise to the miracle of the mind.

The Wall

Chip makers' quest for small may be hitting it

Gordon Moore, chairman of Intel, the world's largest semiconductor corporation, recalls that the company's first chip-manufacturing plant in the late 1960s cost $3 million. "Today that is the cost of one piece of equipment in one of our plants," Moore laments. The price tag for the newest factories that Intel and other suppliers build can exceed $1 billion. In the year 2000 that figure could be higher by a factor of nearly 10.

The trend may terminate the inexorable decline in the unit price of computer logic and memory that manufacturers have ridden so profitably for the past 25 years. Unit prices have declined because semiconductor makers have been able to double the density of transistors on a chip, on average, every 18 months.

Consequently, during recent decades, the cost per function—a function being a bit of memory or a logic switch on a chip—has declined every year by 25 to 30 percent. That fact has let computer buyers purchase a more powerful machine each year for the same price.

These changes could affect the growth in the semiconductor industry. "There's a fear that if the 25 to 30 percent decrease in cost per function changes, the industry's rate of growth will slow," says Glen Cheney, president of Semi/Sematech, a consortium of equipment, materials and service suppliers. James D. Meindl, a professor at the Georgia Institute of Technology, puts it more succinctly: "It's a potential showstopper."

The problem is that life gets harder for chip makers as dimensions shrink below a micron in size. At some time, perhaps well past the turn of the century, conventional integrated-circuit technology will peter out. At about 0.05 micron, the dimensions of the individual devices will be so small that quantum effects will disrupt their behavior. Then a new technology based on probabilistic laws will be needed.

The quest for small has already begun to grow difficult as industry tries to eke improvements from the classical manufacturing process. In it, thin layers of silicon, oxides or metal are grown or deposited on the surface of a silicon disk, called a wafer. Selected surface areas for each chip are then patterned and removed using a photolithographic and etching process. Tiny ridges and valleys formed through etching are implanted, or "doped," with materials such as boron or arsenic to control the conductivity of the component parts of transistors and other circuit elements. When complete, the wafer is "diced" up into individual chips that are then packaged with a ceramic or plastic covering.

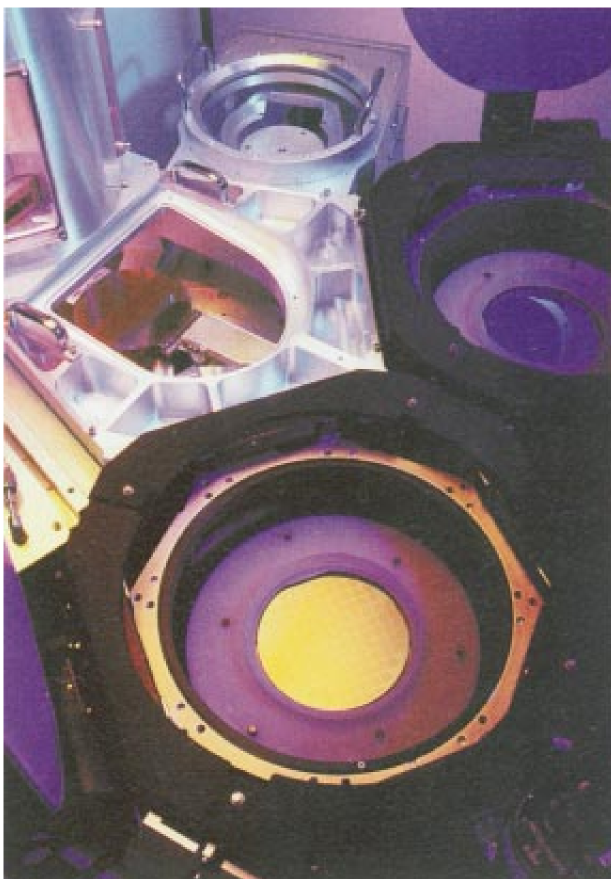
The production of microchips requires the various steps, such as deposition and lithography, to be repeated many times before the wafer is cut into chips. Today's state-of-the-art chips, which house 16 million bits of memory, require 200 processing steps. That number could triple if industry is to ship billion-bit dynamic random-access memory chips by the year 2001. For such devices, engineering tolerances can be on the order of 15 to 50 nanometers, less than the width of a coiled DNA molecule.

Methods used to enhance productivity have begun to fade. Cleaner processing has reduced the number of defects on a silicon wafer. The measure of a defect-free wafer, called yield, now stands at 90 percent. "There's no place left to go," says Thomas Seidel, chief technologist at Sematech, a consortium of semiconductor manufacturers.

Another productivity booster—expanding the diameter of a wafer so that more chips can be processed by an individual machine—has also lost some of its allure. In past years, a single chip maker would absorb the research and development cost of introducing a larger wafer. By so doing, it would get a manufacturing advantage before its competitors. Ultimately, the leader provided its rivals with a nice subsidy. Having filled the front-runner's order, equipment manufacturers would then sell the machinery needed to work with the larger wafer size to chip makers playing follow the leader.

IBM, for example, was the first to design new processes and then order a completely new set of machines for making eight-inch wafers, which currently reflects the state of the art. Yet even IBM may balk at the $2 billion or more needed for the next wafer size, which could measure anywhere from 10 to 16 inches in diameter.

Sematech and the Semicon-



APPLIED MATERIALS

PRECISION-ETCHING equipment is needed to create microchip structures whose smallest dimensions are less than one micron. Capital costs soar, however, as the size of transistors on the chip surface gets smaller.

ductor Research Corporation, which coordinates university research efforts, have established programs to find ways to pick up the pace of production. "There seems to be a rule of thumb that has operated in the past that no piece of equipment will exceed 30 wafers per hour throughput," complains Craig Barrett, Intel's chief operating officer. "I don't know if that's a rule of nature or a rule of equipment vendors."

Equipment suppliers, such as Applied Materials in Santa Clara, Calif., the industry's largest manufacturer, are responding. They have created technology that allows several production steps to be executed without removing the silicon workpiece from the vacuum chamber, which saves processing time.

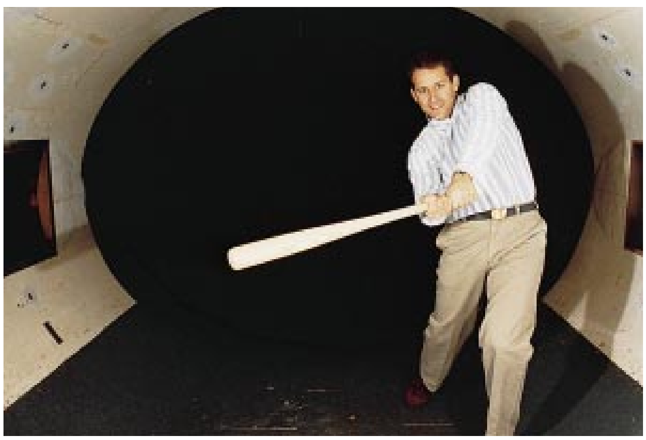
As capital budgets mount, risk sharing is becoming an increasingly popular way to relieve the pain. Witness the collaboration between IBM, Siemens and Toshiba on developing a 256-megabit memory chip. Another anodyne is capacity reduction. The industry, which suffered from a glut of capacity in the 1980s, has become a lot leaner. It could become even more so. "The number of players is likely to be quite limited and not to grow in the future because of the rising capital costs," Meindl says. According to industry market analyst VLSI Research, the number of plants has already dropped from about 2,000 a decade ago to 876 today.

There is more evidence of an industry entering early middle age. VLSI Research undertook a provocative financial analysis of three major industry producers. The firm used chaos theory to analyze a factor it calls "technology drag," the burden placed on companies by having to introduce ever more costly technologies. The study showed that Intel, the industry leader, had generated $3 in revenue for each dollar invested during the early 1970s. By 1998, unless something changes, it will make only a dollar for each dollar invested. Intel will wait and see. "Anybody's ability to analyze what's happening a few years down the road is extremely limited," Moore replies.

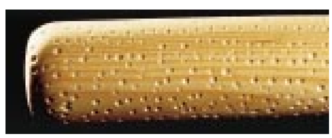
To cope with mounting capital costs, VLSI Research recommends that manufacturers moderate the blistering pace of technological one-upmanship that has led to four generations of advanced chips coexisting in the same market. It might work, but the chances that the patient will voluntarily take its medicine are slim. *—Gary Stix*

To Fenway, with Love

Golf balls fly faster, and therefore farther, because they have dimples that relieve aerodynamic drag. This past February an aeronautics instructor at the Massachusetts Institute of Technology won a patent for extending the same idea to baseball bats. Knocking shallow pea-size depressions into a bat, he discovered, can add speed to a swing and distance to a hit.



DONNA COVENEY *MIT Photo*



DIMPLED BASEBALL BAT patented by Jeffrey C. DiTullio of M.I.T. produces less drag, yielding longer hits.

The eureka moment came to Jeffrey C. DiTullio after using a wind tunnel to study drag on cylinders. "While I was sitting in traffic on the way home, it occurred to me that if we could reduce the drag on the baseball bat, then hitters could swing it faster." A quicker swing would both add distance to each hit and give the batter more time to judge the pitch.

DiTullio knew that the main source of drag on a moving cylinder is not friction but pressure. As a bat carves its swath, it separates the air in front of it into two boundary layers that recombine some distance behind it. In the wake of the bat, between the separated boundary layers, the lack of air creates a partial vacuum that pulls on the bat and sucks energy out of the swing.

A smooth bat will slice the air cleanly, leaving a large wake of low pressure. But a bat roughened by shallow dimples will trip up the boundary layers, sending them tumbling in turbulent eddies into the space just behind the speeding bat. That relieves the vacuum and cuts the drag.

With bats donated by the Boston Red Sox, DiTullio hammered out three prototypes. In the M.I.T. wind tunnel, DiTullio says, the dimpled bats showed 60 percent less drag against a 70-mile-per-hour wind than did smooth bats that were otherwise identical. In theory, that ought to allow a professional batter to swing a dimpled bat 5 percent faster—enough to add about 15 feet to a hit. But in field trials with players from the Red Sox farm team, DiTullio measured a less dramatic improvement of just 3 percent.

Still, in many sports, equipment manufacturers would jump at the chance to offer a 3 percent boost in performance at little or no extra cost. M.I.T. has high hopes that it can sell the idea to bat manufacturers such as Hillerich & Bradsby, maker of the famed Louisville Slugger. But while "some companies are mildly interested," reports David McFeeters-Krone of M.I.T.'s technology licensing office, "we are still actively seeking licensees."

M.I.T. will have to make a good pitch indeed to get a dimpled bat into the majors. The league has rejected many previous innovations in bat design, from J. A. Hillerich's 1904 "fish scale" bat (covered in wedged points to reduce foul hits) to the metal and composite bats that have consumed 75 percent of wooden bat sales since the mid-1970s.

Dimples would work as well on aluminum or graphite bats as on wood, and DiTullio's design may find easier acceptance among amateur ballplayers than among purist professionals. If so, M.I.T., like the inventor of the dimpled golf ball, may strike it rich. Otherwise, like the inventor of the fish-scale bat, it may just strike out. *—W. Wayt Gibbs*

Private Screening

At-home HIV tests stir up controversy

EPITOPE, INC.

ORAL SAMPLE COLLECTOR from Epitope, which can be used for gathering specimens for HIV screening, is placed between the gum and cheek. The sample of saliva and blood components is then sent by the health provider to a laboratory for testing.

AIDS is a late 20th-century scourge. It is also, in a market economy, an opportunity for drug and medical device manufacturers, albeit a controversial one. For example, research by the University of California at San Francisco's Center for AIDS Prevention Studies predicts that over 40 million people might get tested for the human immunodeficiency virus (HIV) if they could be assured of their privacy. The testing might add 100,000 individuals to the roster of those known to carry HIV. At least one company perceives the study as pointing to the need for a home HIV test, a market that could reach into the hundreds of millions of dollars.

Direct Access Diagnostics, a subsidiary of Johnson & Johnson, has already applied to the Food and Drug Administration for approval of a test that would let someone prick a finger with a lancet at home. The individual then blots the blood on filter paper and mails the specimen to the manufacturer. The manufacturer assays the sample, using conventional antibody-testing methods.

The person tested would get the results of the anonymous test by calling the service—the patient is identified only by code number. If the test shows that the person is positive for HIV, a trained counselor would respond. Negative test results would be conveyed by a recorded message. A caller receiving good news would still have an opportunity to talk to a counselor and listen to recorded tapes on AIDS-related issues.

Direct Access's product has already begun to draw fire. An organization that represents AIDS counselors and testing personnel says a telephone call is not enough. The National Alliance of Lesbian and Gay Health Clinics (NALGHC), which speaks for 11 private health centers nationwide, has criticized the test in a letter to the FDA and held an April press conference to voice its objections.

Christopher J. Portelli, the NALGHC's coordinator, emphasized that patients need intense face-to-face counseling on health care options and practices that can prevent infection of others. "It's very easy to drop the receiver when a patient hears bad news," he says. Carriers of the virus who do not get adequate counseling are susceptible to suicide and domestic violence, Portelli warns. A test packet purchased at a local pharmacy, he asserts, could also facilitate coercive testing by law enforcement officials or school administrators.

The American Medical Association worries about collecting blood samples at home. "I think it [a test] should be done via a health provider with an explanation and a physical," explains Raymond Scalettar, a member of the AMA's board of trustees. "This is not like any home test. To some, a positive AIDS test is a death sentence."

Elliott J. Millenson, president of Direct Access, says he has labored for seven years to address the kind of problems cited by Portelli and Scalettar. Before becoming president of the Johnson & Johnson concern, Millenson headed another company, University Hospital Laboratories, which developed an identical test that was rejected by the FDA in 1990 because of the issue of adequate telephone counseling. After being turned down, Millenson convinced Johnson & Johnson to back him in the venture that has become Direct Access Diagnostics.

Since Millenson's first rejection, he has worked to set up a thorough training program for telephone counselors. His firm also established an extensive computer database, which uses information from the Centers for Disease Control, to provide referrals for health care, legal help and financial counseling.

Millenson focuses on the need for collecting home samples: he cites a 1991 survey from the Centers for Disease Control that showed that 8 percent of U.S. adults planned to be tested for HIV during the following year; that figure, he says, could have tripled if a home test were available.

In fact, not everyone concerned with AIDS health care problems agrees with the NALGHC's stance. Other activists, researchers, minority organizations and state officials have written the FDA in support of the product and service.

As it awaits a final agency decision, Direct Access has shown more than a touch of sensitivity to criticism. When it learned of the NALGHC's opposition, the company stopped payment on a $5,000 contribution to the Whitman-Walker Clinic in Washington, D.C., where the NALGHC's press conference was held. Whitman-Walker is the city's largest private facility for AIDS testing and a member of the NALGHC. The company also dismissed a physician at Whitman-Walker from an expert panel it has set up to study the health impact of the product if it does gain approval.

The FDA seems uncertain about the politically charged issue of home testing. The agency has thus contemplated consulting higher-level officials in the Department of Health and Human Services, of which the FDA is one agency. Agency higher-ups might then make a policy decision about this type of test.

Debate over new forms of HIV testing may become still more acute. Several companies are working on making HIV testing as easy as spitting in a cup. Saliva Diagnostic Systems (SDS) in Vancouver, Wash., is developing a test for professionals and consumers that its officials compare with a home pregnancy test. A strip coated with an antigen protein is exposed to saliva. It changes color within minutes if HIV antibodies are present.

The FDA is unlikely to endorse so readily a test that, if not closely regulated, could be subject to abuse: an employer could test job applicants, or people might head for a singles bar with a condom and a test kit. A measure of the agency's cautiousness on new HIV testing could be witnessed in 1991 when it pulled from the market OraSure, an oral HIV sample collector developed by Epitope in Beaverton, Ore. A health provider places the swab between the gum and cheek to gather saliva and components of blood that filter into epithelial tissue. The sample is then tested for antibodies in a laboratory. Epitope had sold its product to insurance companies for evaluating applications for new policies, a market that was less regulated by the FDA. The company, which has garnered support for the device from some AIDS groups, still awaits an FDA decision on OraSure as a device for use by health professionals.

Because of FDA conservatism, SDS—and two other companies developing saliva tests—has targeted markets outside the U.S. such as developing nations in Africa and Southeast Asia. Ronald L. Lealos, president of SDS, points out that an advantage of marketing in places like Thailand is that the company does not have to cope with the charged political atmosphere that surrounds the disease in the U.S. "It doesn't have any of these 'gay disease' type of connotations over there," Lealos notes. "That's not the way it's perceived, so there isn't a political aura surrounding it."

If effective, saliva may become a way to detect HIV infection where laboratory facilities are scarce. There are fewer antibodies in saliva than in blood. If the tests are made sufficiently sensitive, they should serve as accurate diagnostic tools, says Thomas C. Quinn, a senior investigator at the National Institute of Allergy and Infectious Diseases. "Testing saliva can be a convenient noninvasive technique," Quinn comments. Quinn has, in fact, used rapid blood tests in Africa. Such an assay is approved for use in U.S. health care facilities.

No test for saliva or other oral fluids has yet to make its way through the FDA's rigorous clinical trials. Last December in a letter to Epitope, P. Michael Dubinsky, an official at the FDA's Center for Biologics Evaluation and Research, ordered the company to refrain from making any statement suggesting that the FDA was not worried about the "safety and efficacy" of OraSure. Dubinsky wrote that the "FDA has had some very real concerns regarding reduced sensitivity when testing OraSure samples as compared to blood samples using the same test method."

A company vice president, John H. Fitchen, claims that OraSure yields results comparable to those of blood tests in clinical use. The stock market, however, is cynical about the company's ability to get the agency's approval. In early May, Fitchen says, one fifth of the firm's 10 million shares traded on the American Stock Exchange were being sold short by investors betting that the stock will plummet if the FDA does not approve Epitope's application.

They may want to hedge by investing in Trinity Biotech in Dublin, Ireland. Trinity is supplying its rapid saliva test to the government of China. Judy Phelan, the product manager, says the Chinese are interested in evaluating the test for mass HIV screening of people who enter the country. —*Gary Stix*

Blue Video

The quest for the blue laser sparks intense competition

When it comes to high-density optical-disk recording, there is no widget so earnestly sought by so many workers as the low-cost, blue-violet, room-temperature semiconductor laser. The virtue of low cost is obvious; the desirability of a blue-violet beam and room-temperature operation may need elaboration.

During the optical recording process, a laser burns microscopic pits into the surface of the master disk. The pits are what carry the information that is retrieved when another laser reads the pits. "To achieve more than two hours of high-quality, full-motion video on a CD-size disk," says Pioneer spokesman Kinro Shimizu, "the blue laser is essential."

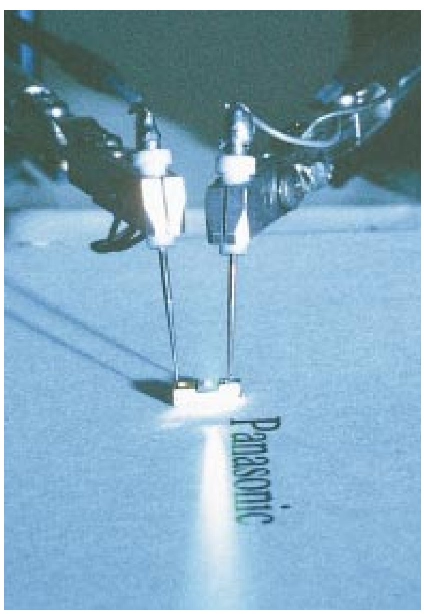
According to Shimizu, the size of a pit that can be inscribed or read by a blue laser is half the length and width of a pit that a red laser can inscribe. That means a blue laser can burn four times as much data into a given surface area. Reducing the size of the pits is no mean feat itself: if a conventional optical disk were blown up to the size of a baseball field, each pit would be roughly the size of a grain of sand. By narrowing the track pitch and the length between pits, Shimizu adds, recording density can be increased an additional 2.5 to three times in total.

To be useful in a consumer product, the laser must meet another test. It has to operate at room temperature. A cryogenic support system does not fit well with everyday domestic living. During the past year, corporate and university workers have posted significant progress with respect to wavelength and room-temperature operation.

This past March, Pioneer workers directed a 100-milliwatt beam from an infrared semiconductor laser through a nonlinear crystal of potassium titanyl phosphate made by Du Pont. The emerging beam, a second harmonic at 425 nanometers, was sufficiently stable to read data from a high-density optical disk under laboratory conditions.

Less than a month later, in early April, Matsushita researchers coaxed blue light from a semiconductor laser, using the same second-harmonic generation method as Pioneer did. Yet another Matsushita team, led by Tsuneo Mitsuyu and his colleagues in the Device Research Group, developed a laser diode that emits pulses of light with a wavelength of 509 nanometers. Mitsuyu is now developing a solid-state laser that generates a continuous blue signal. To do that, he says, requires a new semiconductor material with a wide band gap, such as zinc selenide, a II–IV compound in terms of the periodic table.

Hitachi's Central Research Laboratory in Tokyo has also been working with zinc selenide compounds for several years. Masahito Migita, a senior re-



MATSUSHITA ELECTRIC, INC., CO.

BLUE LIGHT at a wavelength of 509 nanometers is emitted from this laser diode, recently demonstrated by a team led by Tsuneo Mitsuyu of Matsushita Electric, the maker of Panasonic products.

searcher at Hitachi, is looking at two other possible compounds from which to make blue-violet laser diodes: zinc cadmium magnesium sulfur selenium telluride—which has been induced to lase, if only briefly—and gallium aluminum indium nitride.

Like their counterparts at Pioneer and Matsushita, Hitachi researchers are experimenting with the second-harmonic generation method. Other major Japanese electronics firms, such as NEC and Toshiba, are also involved in the blue-laser quest.

In the summer of 1993 a Sony research team operated a continuous-wave blue laser successfully at room temperature. Sony's experimental laser diode emits light at a wavelength of 489.9 nanometers—closer to the green part of the spectrum than to the violet. A group from Brown University and Purdue University independently achieved similar results at about the same time.

In the U.S., workers at IBM have successfully used potassium niobate crystals as a nonlinear medium in which to double the frequency of gallium arsenide laser diodes. AT&T recently signed a technology agreement with APA Optics in Blaine, Minn. The agreement will enable AT&T to conduct blue-laser research exploiting gallium aluminum nitride epitaxial systems that APA has developed.

But don't confuse experimental success with commercial practicality. The blue laser that will read and record a two-hour movie from a CD-size disk is still years away. Until then, consumers will have to be content with a VCR or with the somewhat cumbersome three-disk systems based on red-wavelength technology. *—Robert Patton, Tokyo*

Roach Wars

Diabolical additions to the blatticidal arsenal

As many as 50,000 cockroaches may call your house their home. If you are tempted to take extreme measures to rid yourself of these roommates, you may be justified in feeling that it's you or them. A roach (*Blatella germanica*, or German roach, is one familiar example) can spread harmful microorganisms such as salmonella. Secretions from a roach can produce dermatological or respiratory allergies. At least one laboratory researcher has even had to give up investigations of roaches because of allergic reactions. (Roaches 1, researchers 0.) Studies in emergency rooms in Wilmington, Del., and Atlanta, have found that sensitization to cockroaches is an important risk factor for asthma attacks.

But isn't there something just a wee bit cruel in biological weapons that pesticide manufacturers and biotechnology companies want to put in your hands? To be sure, the corporations cannot be blamed for trying. U.S. consumers bought $240 million worth of anticockroach toxins last year.

ROACH CITY devised by the Department of Agriculture's Agricultural Research Service lets investigators breed thousands of roaches efficiently in confined spaces.

Consider the innovation that EcoScience, a Worcester, Mass., company, has marketed. It consists of a type of fungus that literally devours the German cockroach. *Metarhizium anisopliae,* as the fungus is known, could play a protagonist in a grade B movie. When the fungus's spores contact the exoskeleton, they release enzymes that eat away the hardened cuticle, allowing the penetration of rootlike hyphae. The roach's innards are then slowly digested by the fungus, killing the insect within a week. During that time, spores also attach themselves to nestmates, who are then similarly devoured.

EcoScience says its product is an environmentally benign alternative to the chemical sprays and baits that attack a roach's nervous system. Also, the company's studies have shown that risk of fungal allergies is low, says Kevin J. Devine, the company's vice president of sales and marketing.

This fungus supposedly was deployed against agricultural pests in Czarist Russia. Now EcoScience has figured out a way to package these microorganisms for the supermarket shelf. A permeable plastic pouch encloses the round plastic chambers housing the fungi spores. The pouch controls humidity as well as levels of carbon dioxide and oxygen. This arrangement gives *Metarhizium* shelf life.

Another biological assault against the cockroach may be afoot. Biosys, a company in Palo Alto, Calif., has enlisted nematodes. The tiny worms enter a roach's anus or its spiracle (the aperture through which the insect breathes). They then deposit a bacterium that kills the roach, turning it into fodder for the nematode.

But the firm needs to find a way to slow nematode metabolism so that the worms do not starve during the year or so they may have to spend on the shelf in a hardware store or supermarket. "The product couldn't yet exist unrefrigerated in a Kmart, a Payless or a Home Depot," says Stephen A. Manweiler, a company manager who has done research on the nematodes for use against roaches and other insects.

A number of companies are developing preparations that inhibit the synthesis of chitin, the main constituent of the insect's hardened exoskeleton. Academic research teams, including one at North Carolina State University, have launched a quest to discover the chemical, known as an aggregation pheromone, that the German cockroach releases to summon a klatsch of males and females. (A sexual pheromone, by contrast, would draw only the opposite sex.) The aggregation pheromone might be used to lure the insects into the kind of sticky trap sometimes marketed as a roach motel.

Industrial firms and government and academic research laboratories take blatticide seriously. S. C. Johnson Corporation, which makes Johnson Wax and Raid, the antiroach product, breeds 80,000 roaches a week and plays host to up to a million roaches at any one time. "These animals are formidable opponents," says Keith Kennedy, a research leader at S. C. Johnson. "People working with cockroaches don't have to worry about job opportunities."

One prized strain of roach in S. C. Johnson's stable is known simply as HRDC. The acronym does not represent a long-winded Latin name. It stands for the House of Representatives, District of Columbia, from whose corridors of power the strain emerged in the early 1980s. "We use HRDC as a reality check because it is resistant against so many different chemicals," Kennedy says. "It seems to be a supertough cockroach, and it's only fitting that it came from Washington." Indeed, HRDC may be around long after our nation's capital is gone. Any organism that has outlived *Tyrannosaurus rex* has time on its side. *—Gary Stix*

Green Economics

The U.S. makes a foray into sustainable accounting

At the United Nations Conference on Environment and Development held in Rio de Janeiro in 1992, governments agreed to develop a system of accounting that would better reflect exploitation of natural resources. With traditional methods, activities such as logging, fishing and mining can only add to the gross national product. Such accounting fails to reflect the fact that exploitation reduces the value of natural assets—forests, fish and minerals—and affects their ability to provide income in future years. Exploitation that exceeds replacement is unsustainable.

The U.N. published recommendations on environmental accounting last year, and in May of this year the U.S. made its first attempt to comply. The end product, published by the Department of Commerce's Bureau of Economic Analysis, is known as the Integrated Economic and Environmental Satellite Accounts. The result, pleasantly surprising to the uninitiated but predictable to economists, seems to suggest that as far as oil, gas, coal and some minerals are concerned, the U.S. measures up well.

The new accounts try to put a value on the U.S.'s most important below-ground natural resources during the period between 1958 and 1991. The bureau used various methods to measure the value of "proved" reserves in this sector, reserves that are economically exploitable. The outcome was that the aggregated stock of proved reserves increased in value from between $103 billion and $182 billion in 1958 to between $471 billion and $916 billion in 1991. In constant dollars the real stock slipped, but by an insignificant amount, from between $544 billion and $1,077 billion in 1958 to between $530 billion and $1,030 billion in 1991.

Anyone who has looked into the Bingham copper pit near Salt Lake City will wonder what kind of legerdemain can explain these numbers. The invisible hand at work is that of technology, not that of Adam Smith. The figures show little change because the value of reserves used has almost been compensated for by the increase in reserves that can profitably be exploited. Advances in mining and extraction capabilities have turned reserves that were not worth using 30 years ago into valuable resources. Prospecting has also led to some new finds. So incorporating the depletion of mineral resources into U.S. economic statistics would have little effect.

Is the constancy of mineral resources, then, an argument for ignoring sustainability and continuing business as usual? "Absolutely not," replies Robert Repetto, a prominent researcher in environmental economics at the World Resources Institute in Washington, D.C. Repetto points out that although improvements in technology have allowed the mine tailings, or waste material, of 30 years ago to become a profitable source of ores today, the change has not taken place without a toll on the environment. Wastes are now produced in larger amounts, for example. Moreover, consumption has increased, so the ratio of reserves to consumption has gone down. The constant value of mineral reserves in the U.S. does, however, confirm what experts already knew, Repetto says: that whatever other environmental problems it may have, the U.S. is not in imminent danger of running out of minerals. If accounts were prepared that valued fish stocks, on the other hand, they might well lead to better management of that resource.

CHIP VINAI Gamma Liaison

CLEAR-CUT FORESTS are unappealing for recreation. Environmental accounting might put a value on that loss. This stand in Washington was clear-cut this year.

Environmental accounts may have a sharper message in developing countries, which generate more of their gross national product by selling raw materials and exploiting other natural resources, notes Kirk Hamilton, an economist at the World Bank. David W. Pearce, an economist at University College London, has estimated "green" economic statistics that incorporate natural resources for a selection of countries. Although Costa Rica, Germany, Japan, the U.S. and several other countries are likely to be sustainable economies, by Pearce's reckoning, many developing countries, including Papua New Guinea, Ethiopia and Mali, have a decreasing level of total capital and so are unsustainable. For unsustainable economies, developing green accounts may lead to wiser policies, Hamilton says. Even countries such as Mexico and the Philippines are only marginally sustainable, according to Pearce.

Putting a value on tangible natural resources that are sold and so have a market value can be a complicated exercise, but it is at least conceptually straightforward. Yet it is only the first step to truly green economics, which would include all significant interactions with the environment. When green accounting is extended to include factors such as the value of damage caused by pollution, however, things become more complex. Economic theorists have not come to a consensus on the value to industry of the atmosphere as a repository for unwanted gases, for example.

The counting gets more complicated still if entirely intangible values are to be included—which, arguably, they should be for some purposes. Economists have been unable to agree even in principle on how exactly to measure the value of assets such as the beauty of a wilderness or the splendor of a blue whale. "I don't see us trying to put aesthetic values into economic accounts," comments Carol S. Carson, director of the Bureau of Economic Analysis. When it comes down to hard numbers, sustainability remains an imprecise idea.

Economists may be getting greener, but they will not anytime soon be able to lead the way to a sustainable future. Difficulties arise above and beyond the measurement challenges, Hamilton argues. Many governments would probably not look after the long-term interests of their people even if green accounts showed them how to do so. Environmental accounts are unlikely to have much influence on leaders who do not want to hear. —*Tim Beardsley*

In the Long Run, We All Retire

A disquieting change may soon disrupt the retirement ritual in corporate America. The gold watch and the handshake may linger, but the pension could be in deep trouble. As an increasing number of large companies run into financial difficulty and an aging population swells the ranks of retirees, concern grows about the safety of the retirement benefits secured by privately funded pension plans.

By last year the gap between employers' pension liabilities and the corresponding assets exceeded $53 billion, nearly four times the figure of just a decade ago. The government-run Pension Benefit Guaranty Corporation (PBGC), which stands behind retirement plans whose sponsors default, has a mere $8 billion in hand to cover the potential shortfall.

Does this mean taxpayers are going to have to cough up another 40-odd billion dollars next week to bail out the PBGC? Probably not. First, much of the money that pension plans owe does not actually have to be paid out until well past the turn of the century. Second, depending on how the economy performs in the interim, there could be no pension deficit at all.

Or the gap could be much greater. No one knows. Jeremy I. Bulow of Stanford University explains that calculations of pension costs are inordinately sensitive to assumptions about interest rates and the performance of the economy in coming decades. A dollar invested today at a 7 percent annual return will produce only about half the retirement benefits in 20 years that it could at 10 percent. The sharp drop in interest rates during the past five years has sent corporations scrambling to replenish their pension funds.

Of course, back when interest rates were high, companies took advantage of the logic of compounding to recapture "excess" pension assets. During the mid-1980s, corporate treasuries took back about $20 billion from ostensibly overfunded pension plans. Although Congress subsequently changed the law to make such windfalls more difficult to reap, rising stock prices can have much the same effect by inflating the value of a plan's assets. The bull market of 1993 excused many companies from making any pension contributions whatsoever that year, says Douglas L. Kruse of Rutgers University.

The rules that govern pension contributions make it relatively easy for a company to underfund, Bulow says. For example, if management promises increased retirement benefits to sweeten stagnant wages, it has as long as 15 years to make good the resulting shortfall in pension assets. A company that undertakes such a maneuver frequently can build up a significant backlog. (Curiously, he notes, pension plans for unionized, hourly employees are generally underfunded, whereas those for salaried workers are generally fully funded.)

Such chronic underfunding essentially pushes the cost of making good on today's promises onto tomorrow's managers. Many observers, however, worry that future generations of executives may come up short or may attempt to duck their responsibility altogether. In what could become an alarming trend, according to the PBGC, some otherwise profitable companies with severely underfunded plans have entered bankruptcy proceedings, dropped their pension responsibilities into the lap of the government and emerged otherwise unchanged. Larger corporations have also moved to divest themselves of troubled divisions whose retirement gaps are particularly onerous. If the spin-off companies fail, the government, not the former owner, must make up the difference.

Pension changes shift the risk from managers and shareholders to workers.

Secretary of Labor Robert B. Reich is pushing legislation that would limit such practices. Companies would be forced to replenish seriously delinquent funds within only five years, and the PBGC would gain a voice in corporate transactions (such as sales, mergers and bankruptcy proceedings) that threaten to leave taxpayers holding the bag. Companies that try to get out from under pension obligations could be required to continue making payments for at least a few years, until the benefits they promised are on a sounder footing.

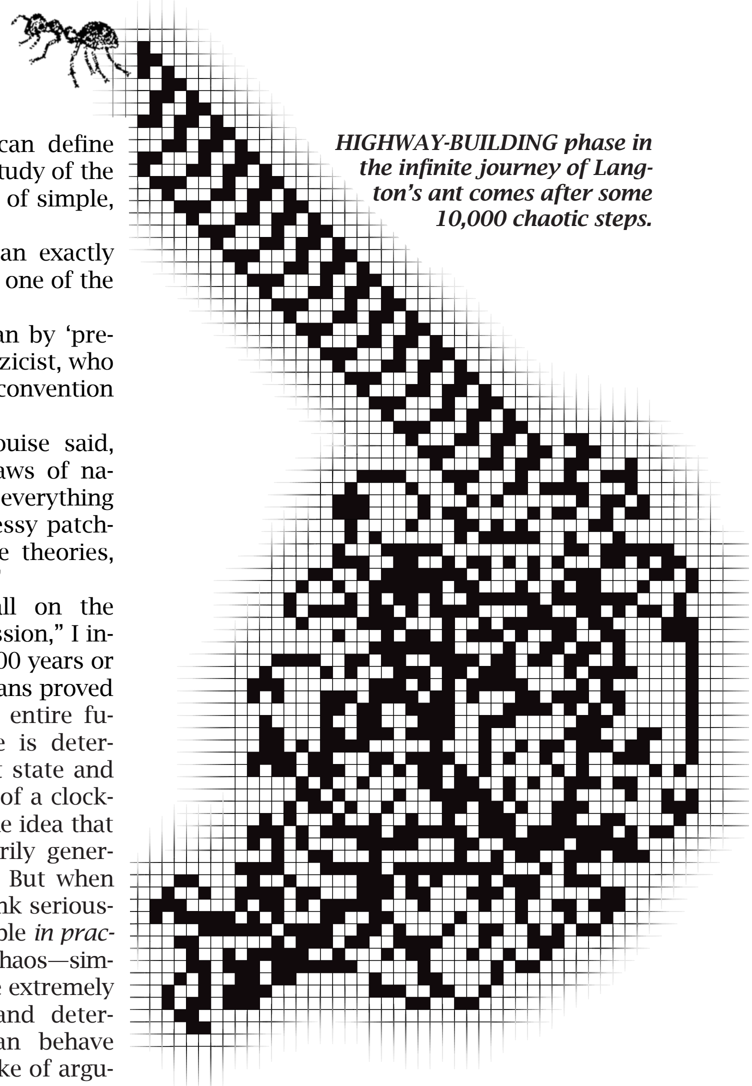
In addition, the proposed law would remove the cap on PBGC insurance premiums. Annual payments currently run $19 per person covered, plus $9 for every $1,000 in unfunded benefits up to a maximum of $53 per year. The most seriously underfunded plans would pay as much as $50 a person more under the new law, increasing their incentive to reform.

Meanwhile another trend threatens to undermine the PBGC's entire enterprise. In the 20 years since ERISA (the Employee Retirement Income Security Act, which established pension oversight) was first passed, employers have increasingly shifted away from defined-benefit plans, which promise a certain level of payments, to defined-contribution plans, which simply invest an agreed-on amount each year and pay retirees whatever it yields.

Defined-contribution plans are not subject to the same complex regulations as defined-benefit ones. Furthermore, Kruse points out, they shift the risk of changing investment conditions from managers and shareholders to workers. (In some cases, defined-contribution plans can even increase investment risks substantially because most of their assets may be in the form of the parent company's stock.) Although classical economic theory predicts that employers should have to pay some kind of premium to workers for accepting this risk, such compensation is not visible in studies. Indeed, the remaining defined-benefit plans appear to be associated with higher wages, Kruse says.

It is not clear how the uncertainty now spreading through the pension funding system will eventually affect the benefits workers receive. For all anyone knows, increased savings rates could more than make up the difference if wage earners decide to take retirement nest eggs into their own hands. Perhaps the only unimpeachable prediction, in line with John Maynard Keynes's famous dictum, is that most of the managers and policymakers charting the pension system's course today will not see the ultimate results of their actions. *—Paul Wallich*

The Ultimate in Anty-Particles



HIGHWAY-BUILDING phase in the infinite journey of Langton's ant comes after some 10,000 chaotic steps.

The hotel lobby was in a state of pandemonium, with suitcases and backpacks piled all over the place. I wondered for the hundredth time whether I'd been in my right mind even to consider attending the Quinquennial SPAM World Convention. I took one look at the check-in line and headed for the hotel bar. About half the participants at this meeting of the Society for Philosophizing about Mathematics had had the same idea.

Immediately to my left, a smartly dressed woman in her mid-sixties was engaged in a heated discussion with what appeared to be the epitome of teenage grunge. A thin woman with spiky hair wore a T-shirt labeled "Watch This Space." There was an overexcited fuzzicist who was busily explaining a flexible extension of conventional logic to three skeptical constructivists who wanted to make it more rigid. And a nerdish type in one corner was madly tapping the keys of a laptop computer.

I grinned and began to relax. The trip was going to be as intellectually rewarding as I had hoped, after all. I introduced myself to spiky hair and T-shirt, whose name turned out to be Louise. "I'm watching," I said.

"Uh?"

" 'This Space.' I'm not seeing anything unexpected."

She gave me a funny look, as if she was trying to work out whether I was being sexist. "It's for the Equation," she said. "When they find it."

"The Equation?"

"Mind you, I'm not very hopeful of a quick breakthrough now that those idiots in Congress have canceled the Superconducting Super Collider."

"Oh, that equation," I said. "The Theory of Everything." She was a fundamentalist; I should have picked that up.

"You may scoff," she said. I started to shake my head to indicate that scoffing had not been on my mind. "I merely believe that everything in the universe is governed by one fundamental law and that the central aim of science must be to find out what it is."

"Yes, but even if it exists, what makes you think it's a mathematical law?" teenage grunge queried.

"The very word 'law' indicates a precision that can be found only in mathematics. Indeed, we can define mathematics as the study of the logical consequences of simple, precise laws."

"What do you mean exactly by 'logical'?" queried one of the constructivists.

"What do you mean by 'precise'?" asked the fuzzicist, who was called Inez. The convention was under way.

"The point is," Louise said, "once we find the laws of nature, we can deduce everything else. Instead of a messy patchwork of approximate theories, we'll know the truth."

"I think you're all on the wrong level of discussion," I interjected. "It's only 100 years or so since mathematicians proved that *in principle* the entire future of the universe is determined by its present state and that led to a picture of a clockwork universe and the idea that simple laws necessarily generate simple behavior. But when people started to think seriously about what's possible *in practice,* they discovered chaos—simple laws can generate extremely complex behavior, and deterministic systems can behave randomly. For the sake of argument, suppose you're right. Suppose that at base the universe really does obey a simple set of fundamental laws, and we find them. Will that really help us understand the world in which we live?"

"Of course it will," Louise answered. "First, it will provide a basic philosophical underpinning. Second, the behavior of our human-sized world is necessarily implicit in the fundamental laws, so the laws will explain everything."

"In principle, maybe. But not in practice. For instance, in the human-scale universe, cats like to chase mice. I wouldn't agree that your 'fundamental' laws explained that unless you could show me a convincing deduction, starting from your equations and ending with the fact that cats like to chase mice. How do you plan to do that?

"Even if you could carry the computations out, they would be impossibly huge and totally incomprehensible. No, the problem with Theories of Everything is that they start with the wrong concept of 'explanation.' An explanation is an explicit argument that leads from hypothesis to conclusion, not just a vague statement that the conclusion is implicit in the assumptions. And certainly not a stack of computer printouts 1,000 miles high that purports to render the implicit explicit."

The nerd in the corner woke up. "Let me show you all something." He plunked his laptop on the table. "Look at the screen." A fine grid of squares appeared. "That's an ant, okay?"

"Where?"

"In the middle, only it's invisible. But now I'll show it to you." He clicked again. Something rushed madly to and fro across the grid, leaving behind a random trail of black and white squares.

It continued for a minute or so and then built a curiously patterned diagonal stripe and disappeared off the edge of the screen.

"Fascinating," Louise said. "Now, as I was saying about the Equation—"

"Louise," the nerd said, "if you'll let me explain what you just saw, you'll see that it is highly relevant to the topic under discussion."

"You always say that, Nathan, even when it's a game of Dougal the Dugong."

"Suspend judgment for a few minutes, and I'll justify my claim. What I've just shown you is a mathematical system known as Langton's ant. It's an amazingly simple cellular automaton invented by Chris Langton of the Santa Fe Institute."

"The complexity mob?"

"Precisely. They look for large-scale regularities in complex systems. Langton's ant is a simple example. The ant starts out on the central square, heading in some selected direction—say, east. It moves one square in that direction and looks at the color of the square it lands on—black or white. If it lands on a black square, it paints it white and turns 90 degrees to the left. If it lands on a white square, it paints it black and turns 90 degrees to the right. It keeps on following those same simple rules forever. You saw what happened when I started it on an all-white grid [*see illustration at top right*]."

"Surprisingly complex behavior for such a simple set of rules," I observed.

"You see, Louise, Langton's rules are the Theory of Everything for the universe that his ant inhabits. An anty-matter universe," he ended apologetically.

"Look at the strange sequence of shapes that the ant creates. For the first 500 or so steps, it keeps returning to the central square, leaving behind it a series of rather symmetrical patterns. Then, for the next 10,000 steps or so, the picture becomes very chaotic. Suddenly—almost as if the ant has finally made up its mind what to do—it builds what James Propp of the Massachusetts Institute of Technology, who first made the discovery, calls a highway. It repeatedly follows a sequence of precisely 104 steps that moves it two cells northwest and continues this indefinitely, forming a diagonal band [*see illustration on opposite page*]."

"Amazing," I remarked. Louise's face said, So what?

"The really amazing thing," Nathan went on, "is that it always seems to end up building a highway, even if you scatter black squares around the grid before it starts."

"I think," one of the constructivists said, "I can start it off next to an infinite line of black squares, and it will just follow them off to infinity. If I get the spacing just right."

"Sorry. Any *finite* arrangement of black squares. But nobody has ever been able to prove that the ant always builds a highway."

"Can anything general be proved about what Langton's ant does when it starts with any finite arrangement of black squares?"

"Yes," Nathan replied. "E.G.D. Cohen and X. P. Kong of the Rockefeller University proved that the ant's trajectory is necessarily unbounded. It escapes from any finite region [*see box below*]."

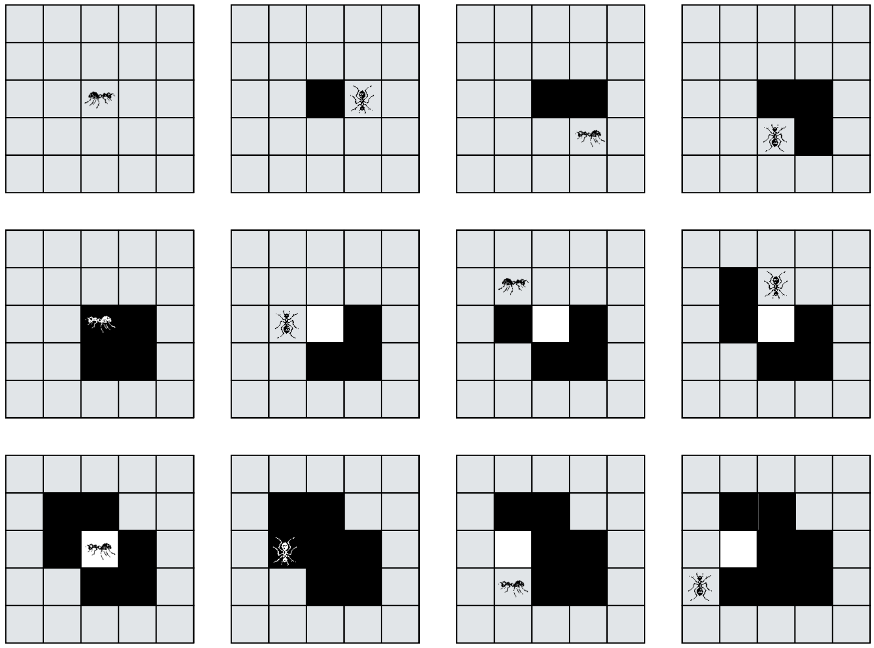
"But what does it have to do with the Equation?" Louise asked.

"We know the Theory of Everything for Langton's ant," Nathan offered. "The rules. We set them up. But despite that, nobody can answer one simple question: Starting from an arbitrary 'environment' of finitely many black cells, does the ant always build a highway?"

"So here the Theory of Everything lacks explanatory power?" I wondered.

"Precisely. It predicts everything but explains nothing. In contrast, the Cohen-Kong theorem *explains* why trajectories are unbounded."

"I can see several flaws in your argu-



FIRST 12 STEPS followed by Langton's ant are indicated. (For clarity, squares not yet visited are shown as gray: these should be treated as white when applying the rules.)

Cohen-Kong Theorem

It is easy to check that the Theory of Everything for Langton's ant is time-reversible—that is, the current pattern and heading uniquely determine the past as well as the future. Any bounded trajectory must eventually repeat the same pattern, position and heading, and by reversibility such a trajectory must be periodic, repeating the same motions indefinitely. Thus, every cell that is visited must be visited infinitely often. The ant's motion is alternately horizontal and vertical, because its direction changes by 90 degrees at each step. Call a cell an H cell if it is entered horizontally and a V cell if it is entered vertically. The H and V cells tile the grid like the black and white squares of a checkerboard.

Select a square M that is visited by the ant and is as far up and to the right as possible, in the sense that the cells immediately above and to the right of it have never been visited. Suppose this is an H cell. Then M must have been entered from the left and exited downward and hence must have been white. But M now turns black, so that on the next visit the ant exits upward, thereby visiting a square that has never been visited. A similar problem arises if M is a V cell. This contradiction proves that no bounded trajectory exists.

ment," Louise noted. "First, the Cohen-Kong theorem is a *consequence* of the Theory of Everything, which therefore has at least some explanatory power. Next, your argument is founded on ignorance. Maybe tomorrow somebody will come up with a proof that ants always build highways."

"I agree. But it is making the Cohen-Kong consequence *explicit* that explains the unbounded trajectories. You can appeal to the uniqueness of the consequences of the Theory of Everything until you're blue in the face, but that alone won't tell you whether a bounded trajectory exists. Similarly, even though we know the Theory of Everything, we will have no idea whether highway building is the universal pattern until somebody makes it an explicit consequence. Or disproves it."

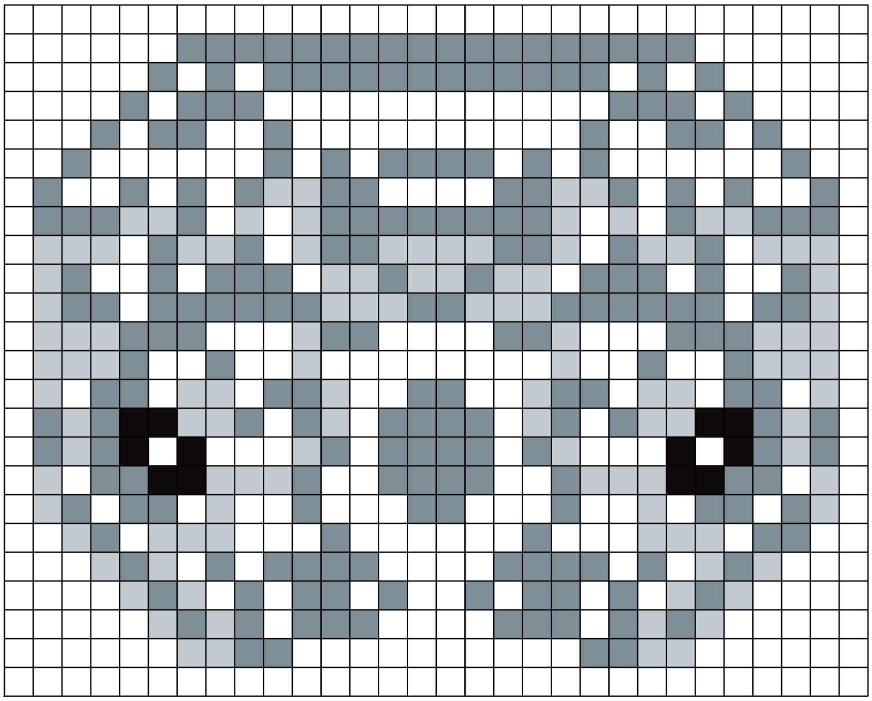
"Seems to me," I interrupted, "that you're asserting an awful lot based on just one exceptional example."

"Not really," Nathan countered. "Langton's ant is entirely typical of rule-based systems. There are lots of generalizations, and they exhibit surprising behaviors and, more surprising, common patterns. You can have fun putting one or more ants into a chosen environment and seeing what they do. You can change the rules and set up different environments—a hexagonal lattice, for instance, instead of a square grid. It's best done on a computer, where simple programs can implement the rules. I should add that there's also a practical side to these ideas: they relate to questions in statistical mechanics about arrays of particles—'ants'—that at any given moment can exist in only one of several states—'colored squares.'"

"Particles and anty-particles?"

"Thank you, Inez. Now, recently Greg Turk of Stanford University and, independently, Leonid A. Bunimovich of the Georgia Institute of Technology and S. E. Troubetzkoy of the University of Bielefeld investigated generalized ants defined by a rule-string. Suppose that instead of just black and white, the squares have n colors, labeled 0, 1, $2 \ldots n-1$. The rule-string is a sequence of n symbols 0 or 1. When the ant leaves a cell of color k, it changes it to color $k+1$ (wrapping $n = (n-1)+1$ around to 0). It turns right if the kth symbol is 1 and left if it is 0. It moves one square on and repeats.

"Langton's original rules are summed up in rule-string 10. Some rule-strings give trivial ant dynamics—for example, an ant with rule-string 1 (or even $111 \ldots 1$) travels forever around a 2×2 square. But any rule-string that contains both 0 and 1 must lead to unbounded trajectories, by the Cohen-Kong idea.



SYMMETRICAL PATTERNS are produced at each step of some rule-strings. The pattern shown here is step 16,464 of ant 1100.

"Suppose for simplicity you start with a 'clean' grid—all cells in color 0. Ant 100 creates patterns that start out looking rather like those of Langton's ant—at first symmetrical, then chaotic. After 150 million steps, however, it is still behaving chaotically. Does it ever build a highway? Nobody knows. Ant 110 does build a highway, and it takes only 150 steps to do so. Moreover, it needs a cycle of just 18 steps to create the highway, instead of the 104 used by Langton's ant. Ant 1000 is relentlessly chaotic. Ant 1101 begins chaotically but goes into highway construction after 250,000 steps, using a cycle of length 388. Ant 1100 keeps building ever more complex patterns that, infinitely often, are bilaterally symmetrical [*see illustration above*]. So it can't build any kind of highway in the usual sense.

"I defy anyone to give a simple classification of the behaviors of all these generalized ants or to predict from their rule-string just what their long-term behavior will be," Nathan declared. "Even if they all start on a clean grid."

Louise looked unhappy. "Yeah, but you haven't proved nobody can do that."

"That's true," I pointed out. "But only slightly more complex transition rules lead to examples such as John Horton Conway's game of Life. Conway proved that in Life there are configurations that form universal Turing machines—programmable computers. Alan M. Turing proved that the long-term behavior of a Turing machine is undecidable—for example, it is impossible to work out in advance whether or not the program will terminate. Translated into Life terms, that implies that the question 'Does this configuration grow unboundedly?' is formally undecidable. So here's a case where we know the Theory of Everything, and we know a simple question that it is provably impossible to answer on the basis of that theory."

"Exactly," Nathan concurred. "So why do you think a real Theory of Everything, for our universe, can in any meaningful sense be an Ultimate Answer?"

"I dunno," sighed Louise, her faith temporarily shaken. She shook her head, then brightened. "It is a bit of an anty-climax."

FURTHER READING

WINNING WAYS, VOL. 2: FOR YOUR MATHEMATICAL PLAYS: GAMES IN PARTICULAR. Elwyn R. Berlekamp, John H. Conway and Richard K. Guy. Academic Press, 1982.

COMPUTER RECREATIONS. A. K. Dewdney in *Scientific American,* Vol. 261, No. 3, pages 180–183; September 1989 and Vol. 262, No. 3, pages 118–121; March 1990.

MATHEMATICAL ENTERTAINMENTS. David Gale in *Mathematical Intelligencer,* Vol. 15, No. 2, pages 54–55; Spring 1993.

FURTHER ANT-ICS: TRAJECTORY OF GENERALIZED ANTS. Jim Propp in *Mathematical Intelligencer,* Vol. 16, No. 1, pages 37–42; Winter 1994.

BOOK REVIEWS by Philip Morrison

A Scientific Summer

THE SCIENTIFIC TRAVELER: A GUIDE TO THE PEOPLE, PLACES, AND INSTITUTIONS OF EUROPE, by Charles Tanford and Jacqueline Reynolds. John Wiley & Sons, 1992 (paperbound, $16.95). **TRAVELERS' HEALTH: HOW TO STAY HEALTHY ALL OVER THE WORLD,** devised and edited by Richard Dawood. Revised third edition. Random House, 1994 (paperbound, $18).

These two unusual travel guides offer specific, detailed matter of practical merit. Both embed well-organized advice within an attractive conceptual framework, enjoyable reading even for stay-at-homes. The two titles span the limits from travel's broad pleasures to its varied risks.

Tanford and Reynolds are distinguished researchers in chemistry and biology from Duke University, now resident in England. Their sensibilities are cultivated and joyfully informed by science; they write with engaging candor and goodwill. Their book contains images and firsthand descriptions within 17 European countries, from Spain to St. Petersburg, from Scotland to Samos. You can learn a lot, for prefaced to each destination is a lively, compact account of the history of the persons and places they celebrate. Their deep aim is wider than scholarship itself, touching on the scientist's pleasure with actuality, the lure of the very place, the very tools, of memorable change. The sheer pleasure in "being there" includes (and even goes beyond) the presence of rich context at its true scale, a sensory grasp that no words on paper can convey.

Sometimes the report itself fascinates. Join their quest for Albert Einstein in Bern, still as provincial a place "as one can find in Europe." There the young man and his new wife first settled, there the papers were written that in 1905 blazed forth the good news of Comet Einstein. Most of the city is still of Einstein's time, and Bern does not forget its famous young family and their bohemian circle. The cramped apartment two floors up where three Einsteins lived is now a museum, light of heart, that "succeeds in evoking both...the man and the scientist" for a throng of visitors. "We need more places like this in the world." But Ulm, where Einstein was born, is no such place. That city seems grudging about its famous son.

"TEMPLE OF VACCINIA," where Edward Jenner gave free vaccinations to the poor, stands in a corner of the garden at the Jenner Museum in Berkeley, England.

"The only tribute that strikes a note of genuine respect," the authors report, is a modest carved stone slab, "a gift to the city of Ulm from the people of India." Even after his fame is resplendent, Einstein is still more welcome as a Swiss Jew than as a scientist of German Ulm.

In a suburb of Den Haag you can visit the beautiful summer home of the Huygens family. Our guides contrast the place with the contemporary home of Isaac Newton in Woolsthorpe: the Newtons were country squires, and not poor, but their substantial house was "a farmhouse...for people who work all day with muddy boots," whereas park and house of the aristocratic Huygens were meant for "garden parties and musical soirées...well-dressed gentlemen and ladies with...delicate slippers."

Listed here are some 200 sites, as scenic as James Clerk Maxwell's family home (now a gutted relic of fire) near Dumfries or those marvelous painted caves of Spain and the Dordogne or the grassy old volcanoes of the Auvergne. You can drive with an altimeter up Puy de Dôme just as Blaise Pascal first proposed (although he sent his brother-in-law to make the climb). Near the cliffs at Lyme Regis in Dorset, you can now learn about Mary Anning, a carpenter's young daughter, who found there the first whole ichthyosaur skeleton. In Paris there is the Curie Institute, but physicist Lise Meitner is not at all well recorded at her lab in Berlin-Dahlem.

Dr. Johnson supplies an epigraph, via Boswell: "A man who has not been in Italy is always conscious of an inferiority." Nowadays he would include women and much extend the list of destinations overseas. Time for travel?

The second book bears the bad news (actually, not so bad), well presented by the editor-originator of this bargain 600-page compendium. Dawood is a far-roving London physician who draws on help from 70 specialist friends. Chapter by chapter, they lay out the medical background a serious reader wants. Here medicine is viewed widely, based on science, yet extending to social psychology, today's cities and marketplaces and the vivid uncommon sense of broad experience.

Accidents are the most common cause of death in travelers, counting 25 times the fatalities from infectious disease. All the same, a third of the text treats infections, in particular those prevalent within the tropics and the poorer lands. Another third reviews the hazards and discomforts of the journey itself, accidents, security and a medical guide to sex overseas. The remainder offers help to travelers with special needs (even mountaineers and divers) and many lists and tables, among them a country-by-country appraisal, not of wines but of drinking water. The authors rest their persuasive opinions on reasoned grounds, always good reading, and they recognize that in the end it is the details that help the most.

Forty percent of all travelers from low- to high-risk regions acquire diarrheal illness. Most of it is bacterial in origin; diet and water are key. Think out why french fries and citrus fruit are safe and why iced drinks and fresh salads bring high risk. The drugs that work—and one or two that don't—are named and assessed; some are antimicrobials, and some relieve by inhibiting the debilitating symptoms. Even the most dangerous of these diseases, cholera itself, has a simple, almost sure cure, oral rehydration. Viral hepatitis A is common and severe but preventable by prior injection of immune (gamma) globulin. Sunburn is costly but familiar; rabies is very rare but requires prompt action. Endemic malaria demands a degree of personal attention and the right pills; new resistant strains are increasing.

A remarkable closing essay treats the planned eradication of specific diseases on the triumphant model of now vanished smallpox. No single smallpox case is known anywhere today, where once were many millions of sufferers. The illness cost the world each year more than a third of a billion dollars and much tragedy. The World Health Organization campaign ended smallpox for good in about a decade, at a total world cost below that of one single year of coping.

A list of 21 candidates for eradication is given here. The criteria are clear, both for the epidemiological vulnerability of a disease and for the public will to end it. In 1986 the tropical Guinea worm was chosen for eradication by 1995. India and Pakistan will, it seems, end it by then, and in Africa, too, the case rates have fallen by two thirds. The method is to make sure that no drinking water carries the larvae of the worm. Three million people a year were infected among 100 million at natural risk. The worm does not kill; however, it slowly and painfully emerges out of your arm, say, as a threadlike creature up to three feet long, keeping kids from school and farmers from their fields for a month or two. Tube wells are the best protection, although they are expensive and slow to come. Filtration through fine-meshed cloth, prevention of infection and a chemical additive can all break the chain cheaply, ending the tiny water fleas that are host to the larvae.

The next target for global eradication is poliomyelitis itself. Not long back still a quarter of a million overt cases were recorded every year worldwide. WHO has passed its capital sentence on polio for the year 2000. The two Americas are already almost free; 10 years of combat cut new cases there from many thousands to a scant few in 1992. The method: wide coverage of the children with live oral vaccine (especially once it is made more heat stable by deuteration) and public surveillance to find and isolate new cases of flaccid paralysis within an urgent cordon of vaccinations.

Eradication is not always within our reach. An ill-advised promise was made to eliminate neonatal tetanus by 1995. That affliction is now held as ineradicable because the natural environmental reservoir of the agent is inexhaustible. Sterile vigilance will always be needed at human birth, as far as we now know. We have not succeeded fully against the malarial protozoan nor against its vector, the mosquito. Measles is potentially eradicable; the modest goal so far is a 90 percent reduction by 2000, for the vaccine does not work well on infants, and the public does not regard the ailment as serious.

Public health and very private health indeed are equally well served by this practical book, chock-full of ideas, hints and cogent connections, laced with both the reassuring and the scary. (Don't ignore those studies that show increased heart disorders among those who remain content at home.)

Contingency's Footprints

THE BROKEN DICE AND OTHER MATHEMATICAL TALES OF CHANCE, by Ivar Ekeland. Translated from the French by Carol Volk. University of Chicago Press, 1993 ($19.95).

Saint Olaf's Saga tells of an island settlement that had from time to time changed its fealty, now to the kingdom of Norway, again to the rulers of Sweden. It fell in Olaf's day that the two kings agreed to throw dice for a final decision. "He was to have it who threw the highest." The Swedish monarch threw two sixes and suggested that Norway retire. Olaf of Norway remarked that the two sixes were still there in the dice, and God could trivially turn them up again. Olaf then indeed rolled two sixes. By custom the king of Norway cast again. On one of the dice the six spots showed. The other die split in two, and its two halves displayed a one and a six. Olaf took possession of the island.

Olaf Haraldsson was no petty fraud; the dice were not loaded or cleverly cracked. We have this on the authority of shadowy Brother Edvin. His unique manuscript of about 1240 was reported by one peerless librarian, Jorge Luis Borges of Buenos Aires, and it is analyzed in the opening chapters of this delightful literary examination of the grand issues of chance and fate. Leave other chapters to the roll of the dice. (The gifted author, bearing both the weighty credentials of a powerful Parisian mathematician and his grandfather's rich legacy of the Edda, suggests that chance should rule the reader's choice; order here is less important.)

The learned Edvin dealt with the moral issues of casting lots and found ample precedent even in Scripture. He further expanded on a new note: Was not any material realization of pure chance bound to be as imperfect as it had proved for the Swedish king? More, he argued that even purely arithmetical means to achieve randomness were flawed. Our Continuator picks up that same theme, applying it to our time of digital computers.

Some arithmetical generator of "randomness" is today a common accessory of most software. Yet all such mathematical attempts at chance are as flawed as bone cubes. A rudimentary model makes this surprisingly clear. Write the digits you reach by simply adding 3 from a start at 0: the list begins 0, 3, 6, 9. Now throw away any excess over 10: then you hold 2, 5, 8, 1, 4, 7. The cycle closes, as you return again to 0. It seems to be one random choice.

Alas for our hopes: plot dots on a numbered plane grid, locating each dot by the horizontal and vertical entries given by the number pairs you form by taking two digits at once out of your random list. Thus, begin with a horizontal 0 and a vertical 3, then move to H6, V9, then to H9, V2 and so on. Your grid will not at all display a strew of disorderly points; instead it will show two parallel lines, one with three dots, one with seven: order.

Any truly random collection of numbers must satisfy every test of hidden order that can be devised, an impossibility for a finite collection of the digits, even if the commonplace home computer "random number" generators now cycle their list through 10 billion numbers instead of only 10. They still are not independent choices but retain some outrageously hidden order. Does chance even exist?

The monkey typing ignorantly forever must sooner or later produce the entire prose Edda, without any author at all. Here Ekeland's knowing treatment avoids the misleading accounts of so many less witty authors: the monkey must be a demon for patience and prepare to turn out as well every other book that has ever been written, a great many more still unwritten and astonishingly more false jumble and gibberish. Yet if he will shorten the random text by noting major repetitions (for instance, here follows copy 1472 of *Hamlet*), the shortened series of letters, coded into binary, will have a perfectly calculable property that shows the text to be less than random.

Any series truncated piece by piece by certain simple decisions, really by a rule that there must be no rules, becomes by any test indistinguishable from randomness. That holds even if it is made by a mechanical rule endlessly iterated, if only the run is edited by fully prescribed though unending decisions. A "purely random" series can thus be formed wholly by necessity.

"We can't get away from determinism." Not so fast. The mathematical axioms themselves are contingent. The famous theorem of Kurt Gödel affirms that an infinite number of distinct mathematics exist, "all born of the same necessity." One of them somehow excels, Plato held (and Gödel himself concurred). Yet "anyone who studies the history of mathematics cannot help being overwhelmed by doubt." The hoofed footprints of contingency are stamped everywhere.

So then is chance to be king? A physicist has an easier time than a mathematician; for him, knowledge is forever partial, approximate. He is skeptical of infinities both in the large and in the small. Perhaps logical certainty itself is the mother of doubts? If you feel like the king of Sweden, his irrefutable arithmetic thwarted when the cube split, the other chapters will wonderfully raise and courageously assuage, although not really answer, the thorny questions that you and I and Dr. Ekeland still share.

Behind the Veils of Venus

VENUS: THE GEOLOGICAL STORY, by Peter Cattermole. Johns Hopkins University Press, 1994 ($49.95).

Brightest of the wanderers, our lady planet of perpetual cloud has dazzled watching eyes since Sumer, exactly because of her white veil, past which no optical tube ever can peer. The veil is an opaque cloud layer 10 times as reflective as the sintered surface of the moon. Microwave radar first showed us the rock surface in the 1960s and fixed the planet's true diameter and unexpected spin. Venus is a few percent smaller than the earth and has a day longer than her year: 243 earth days to turn once on her axis, only 224 to orbit the sun.

Earth-based radars lift the veil only whenever Venus draws nearest the earth. But since 1962, 21 probes have looked in more closely. Both Russian and American craft have visited Venus: flybys, orbiters, dropped instrument probes and landers. The latest entrant, *Magellan,* the Jet Propulsion Laboratory's wonderful orbiter of 1989, has embarked on its final year of advanced radar mapping, taking topography, texture, thermal and gravitational data at a resolution around 100 meters, rivaling the best earth orbiters. *Magellan* reorients four times each orbit to record and send its data; it points merely by spinning an internal reaction wheel, with no need for firing little jets by the tens of thousands. Its solar panels under that hot nearby sun furnish more than a kilowatt of electric power. Its end may come late in 1994, its best atmospheric data coming with the fiery glory of deliberate terminal reentry.

This volume, marked by its clarity and wide interests, full of up-to-date images, amounts to a bare-bones account of a whole new geology as it is unfolding. The visually based narrative by a geologist at Sheffield University not only fully surveys novel Venus, its air and rock (no liquids there save lava), it also illumines by explicit comparison the history of our own planet underfoot.

The barometer on sun-drenched, cloud-darkened Venus reads 100 times higher pressure than ours; nearly all that gas is carbon dioxide. The rock surface is hot enough to melt lead and zinc; the circulation of the dense atmosphere evens out day and night, and seasons are absent as well. The history of this climate is understood from hydrogen isotopes and traces of noble gases. Solar ultraviolet photons dissociated any H_2O molecules that came to high levels, and the lightest of atoms, free hydrogen, was "remorselessly" lost to space. Early on, hot oceans fed the process, and hot rock weathered in the new oxygen supply. There was not much gas left except carbon dioxide from the carbonate minerals, unstable at those furnace temperatures.

The maps allow an instructive simple comparison: a plot of the amount of solid surface area at each altitude. For the earth, the graph is plainly bimodal: one wide peak appears at the level of the seafloor and another peak around five kilometers higher, the continents. Venus is different; it has one narrow peak at the median. No seas, no raised continents. It is not the extremes that are missing; they occupy small areas in any case, and Venus does have a few deep rifts and high mountains. Venus is smoother overall; 90 percent of the surface lies within a three-kilometer interval. The earth has many flat places, leveled by water-produced clays and silts; the broken rock slabs on Venus yield much less area at low slope.

Winds on Venus are fierce, dense and fast-moving. Impact craters are often accompanied by what appear to be wind streaks and even dunes. Yet small impact craters are missing; only larger bodies can make it down through that dense air to the ground. Signs appear of impactors that blew up in flight, leaving a halo of debris around a center swept smooth because of turbulence from an explosion high above, as the Siberian forest was felled under the shock of the Tunguska midair explosion.

It is heat that makes the biggest difference on Venus. The very surface rocks are as fiery as those buried on the earth at a depth of several tens of kilometers, a major part of the crustal thickness. Warm rocks are more ductile, so that on Venus the support of mountains is less enduring, and they spread rather quickly. High peaks have to be kept up by active motions below. The Everest of Venus, Maxwell Montes, is no nearly balanced pile, but a dynamic result of ongoing motions.

The landforms of Venus, its volcanoes, rifts and distortions, are not the result of plate collisions and subduction. More likely, they stand above or beside rising hot plumes of magma, like those below Hawaii. Venus has a weaker, blockier crust and less marked relief, and little difference is seen—though exceptions exist—among the mineral chemistry of its regions. The strange forms called coronas, like enormous flattish lava domes, "weird" spiderlike lava twists called arachnoids (no clear photo here) and other new forms record somehow the complex, interactive, deep-lying plumes common on Venus.

This is work in progress, both the flow itself and our understanding of it. One picture, looking out from ground level over lava plains in Idaho, allows a nonspecialist some feeling for the complexity of Venus. More such comparisons, and more diagrammatic guides to *Magellan* images, would make this fine account even more helpful. One stunning computer-derived radar perspective of Maat Mons, high above its Idaho-on-Venus, lifts the veil dramatically.

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

Index of articles since 1948 available in electronic format. Write SciDex®, Scientific American, 415 Madison Avenue, New York, 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 0036-8733/94. $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.

ESSAY by Robert McC. Adams

Ancient Sumer, Modern Iraq

The Gulf War and its aftermath have left a curiously contradictory picture of Iraq. On the one hand, a repressive, internally fragmented upstart among nations. On the other, a resilient society that has managed, despite a crippling import embargo and the loss of most of its oil-export income, to rebuild a significant part of its infrastructure. As a whole, the picture may seem completely at odds with Iraq's earlier history of cultural achievement. The territory now known as Iraq was, after all, the scene of the world's first literate, urban civilization in the late fourth millennium B.C. Nor was that a singular accident. Admittedly with lengthy interruptions, Sumerian society and its successors continued to be justly celebrated for civilization of a high order through the first millennium A.D.

Without dismissing the importance of these historic achievements, on which most study (including much of my own) has traditionally concentrated, sometimes a greater effort at abstraction can clear the head. In this case, it suggests that the contrast between past and present may be both less severe and more illuminating than it first appears.

A distinguishing characteristic of societies in this region has been their substitution of organizational complexity and hierarchy for more dependable and generous endowments of natural resources. Yet whereas hierarchy is a blandly neutral term, its application was not. All urban societies in this area have been erected on exactions from their heavily exploited agricultural sector.

The productivity of lowland Mesopotamia has always been proverbial, although it is also largely illusory. Subsistence has been drawn from many complementary sources, overcoming the deficiencies of each through complex redistributive or marketing arrangements. Irrigation agriculture, the most important component, illustrates the difficulties. Problems of salinization are pervasive, frequently forcing the abandonment of entire regions for years. They are exacerbated by the gross undependability of the water supply. Over-irrigating leads with certainty to future problems of salinization, but it has long provided the only insurance available to the individual cultivator that at least this year's winter crop will survive to be harvested. Uncertainty is endemic, in other words, and risks are intensified by the short-term choices to which farmers are driven.

Imagine this as an ecosystem without a self-organized, intentional human component that seeks to enhance its own well-being. The complex system of organisms substituting for people might best remain at a low-density, fluidly adaptive level. Some biological means might well evolve for storing subsistence reserves against short-term fluctuations and for regulating access to resources. As nothing more than a new kind of ant colony, this system would still map fairly well onto the basic economic functions that in Mesopotamia have been filled since the fourth millennium by towns and cities.

Now insert real people. Big reserves are better than little ones, but they attract predators. Build walls, organize armies, send out raiding parties. With human propensities for symbolizing, invoke deities as reasons for doing all this and embellish their abodes and the perquisites of those new specialists chosen to attend them.

Given some leisure and self-importance, those attendants' domains of discretion grow. They exploit a servile population of their own, while cultivating rising numbers of specialists in learning and the arts. The latter in turn reify the entire edifice as a divinely sanctioned world. It becomes endowed with cultural monuments embodying its own rich symbolism. Many millennia later those affirmations of eternal life and importance still find treasured places in libraries and museums.

Now, remember the constraints of the subsistence system. Highly productive under optimal conditions, it is subject to uncontrollable risks and to salinization. On a least-effort basis for the cultivator, it is preferably low intensity, shares risks and labor, and internalizes the capacity to draw flexibly among subsistence alternatives. Thus favored are social arrangements permitting mobility (toward more secure water and more arable land) over making improvements that will remain vulnerable by being tied to locations fixed by a particular administration.

Rural, seminomadic groups are intimately adapted to these constraints. Their interests are antithetical to state or urban elites that want to maximize resource intake for their own purposes. Such groups are better served by "tribal" structures of solidarity based on shared descent and reciprocal obligation. Reviled but also sometimes feared by elites for these reasons, they have always borne the main burden of taxes and involuntary labor. Their inferior status and continuing exploitation feeds the repressive character of the system.

Of course, repressive regimes are always at risk. Perturbations of any kind, political or environmental, can shrivel the resources on which they depend for supremacy. The succession of rulership is a recurrent crisis. In this light, the distance between modern Iraq and its past grows smaller.

Nothing is deterministic or linear in the dynamics of this system, which is characterized by stress, uncertainty and conflict but never by equilibrium. Like many social and natural systems, it wanders erratically through history along a ridge between two adjoining basins of attraction (as complexity theorists sometimes call them): chaos and order. The longer intervals of political fragmentation, urban decline and tribal resurgence leave few records and monuments and have received very little historical attention. It is the shorter intervals of self-asserted splendor at the hidden cost of harshly imposed centralization that too often are taken to represent the whole.

Local circumstances that may have shaped history no longer excuse tyranny as their modern outgrowth. Glacially, perhaps, as an avalanche of intercourse and information erodes boundaries of all kinds, the world is edging toward a common stand against the more extreme forms of internal repression as well as external aggression. Yet all of us will continue to live with ever growing hierarchies and complexities of organization—convenient buffers against a deteriorating environment, widening social and economic gulfs and increasingly grave uncertainties of many kinds that are mostly of our own making. It may be well to keep in mind this example of the long-term risk.

ROBERT McC. ADAMS was professor of anthropology at the University of Chicago before becoming secretary of the Smithsonian Institution in 1984.