SCIENTIFIC AMERICAN

NOVEMBER 1993
$3.95

Reading the genes of extinct species.

Observing cannibal stars.

Can the environment survive free trade?

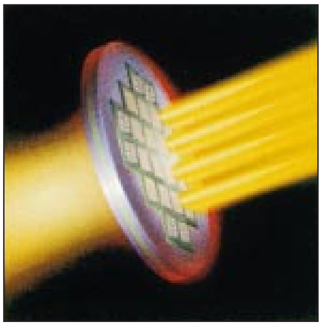
***Silicon switch** provides deft control over electrical power flow, enhancing grid efficiency and reliability.*

SCIENTIFIC AMERICAN

November 1993 Volume 269 Number 5

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. Authorized as second-class mail by the Post Office Department, Ottawa, Canada, and for payment of postage in cash. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada 800-333-1199; other 515-247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

DEPARTMENTS



THE COVER image depicts an MOS-controlled thyristor, a device for handling high-voltage electricity. Current entering and leaving the device is represented by the bright, glowing regions. Thyristors combine high-power electronics with the same kinds of silicon fabrication techniques used to make integrated circuits. By increasing the capacity of high-voltage transmission lines, utilities could defer up to $50 billion in spending over the next 30 years (see "High-Power Electronics," by Narain G. Hingorani and Karl E. Stahlkopf, page 78).

THE ILLUSTRATIONS

Cover image by Michael Goodman

Page	Source
41	Peter Sis
42–43	Steven Ferry/Matrix
44	Ian Worpole
46–48	Steven Ferry/Matrix
49	Peter Sis
50–51	Jean Guichard/Sygma
52–54	Ian Worpole
56	Steven Ferry/Matrix
57	Peter Sis
59	Roberto Osti
60	Jean-Louis Dubois/Jacana
60A	Roberto Osti
60F–62	Dimitry Schidlovsky
64	Jared Schneidman
65	Alfred T. Kamajian
66–70	Jared Schneidman
72–73	Courtesy of the University Art Collection, Syracuse University
74–75	Courtesy of Dodd, Mead & Company, © 1954 (*top center*), Syracuse University (*left* and *right*)
76	Courtesy of *Esquire*, © 1952 (*top*), Syracuse University (*bottom*)
77	*Dateline*, © 1964, courtesy of Syracuse University
78–79	Stephanie Rausser
80–81	Ian Worpole
82–83	Michael Goodman
85	Courtesy of Karl E. Stahlkopf
87	Jonathan Blair/Woodfin Camp
88–89	Ed Bridges/American Museum of Natural History (*top right*), Tomo Narashima (*bottom*)
91	Patricia J. Wynne
92	Thomas Stephan/Black Star
94–95	The MIT Museum (*left*), Jessica Boyatt (*right*)
96–97	*top:* Courtesy of Cheryl Ann Butman; Chris Smith/U.S. Department of Health and Human Services; Jim Starford/Black Star; Marguerite Holloway; Georgia Litwack; courtesy of Deborah Gordon; *bottom:* Bettmann Archive; courtesy of Öffentliche Kunstsammlung Kupferstichkabinett, Basel; Steve Murez/Black Star; painting by Hermione Dassel (1851), courtesy of Vassar College Archives; Bettmann Archive; Caroline L. Hunt, *The Life of Ellen H. Richards,* Boston (1912)
98–99	*top:* Georgia Litwack; Georgia Litwack; Georgia Litwack; Ricardo Azoury/SABA; John Reader, Science Photo Library/Photo Researchers, Inc.; Sam McNaughton, courtesy of NYZS The Wildlife Conservation Society; *bottom:* Bettmann Archive; Bettmann Archive; UPI/Bettmann; UPI/Bettmann; Bettmann Archive; courtesy of Marine Biology Laboratory, Woods Hole, Mass.
100–101	*top:* AP World Wide Photos; Schurlock Studio; Katherine Lambert; Per Breiehagen/Black Star; Georgia Litwack; courtesy of Carol Wood; *bottom:* Joseph Larner from *Nobel Prize Women in Science,* by Sharon Bertsch McGrayne, Birch Lane Press, 1993; UPI/Bettmann; UPI/Bettmann; UPI/Bettmann; UPI/Bettmann; Donald L. D. Caspar from *Nobel Prize Women in Science*
112	Patricia J. Wynne
114–115	Johnny Johnson

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Michelle Press, *Managing Editor;* John Rennie, Russell Ruthen, *Associate Editors;* Timothy M. Beardsley; W. Wayt Gibbs; Marguerite Holloway; John Horgan, *Senior Writer;* Philip Morrison, *Book Editor;* Corey S. Powell; Philip E. Ross; 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, *Quality Control;* Tanya DeSilva, *Prepress;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Eric Marquard, *Special Projects;* Leo J. Petruzzi, *Manufacturing & Makeup;* Ad Traffic: Carl Cherebin

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

ADVERTISING: Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* William Buchanan, *Manager, Corporate Advertising;* Peter Fisch, Randy James, Elizabeth Ryan. Michelle Larsen, *Director, New Business Development.* CHICAGO: 333 N. Michigan Avenue, 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; Kate Dobson, *Advertising Manager;* Tonia Wendt. Lisa K. Carden, Lianne Bloomer, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research 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.; Hoo Siew Sai, Major Media Singapore Pte. Ltd.

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

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

PRESIDENT AND CHIEF EXECUTIVE OFFICER: John J. Hanley

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

CHAIRMEN OF THE BOARD: Dr. Pierre Gerckens, John J. Hanley

CHAIRMAN EMERITUS: Gerard Piel

PRINTED IN U.S.A.

Genes and Behavior

John Horgan's article "Eugenics Revisited" [SCIENTIFIC AMERICAN, June], as its title suggests, would rather try to embarrass behavioral geneticists and impugn their motives as politically suspect than enlighten the reader about a long-standing controversy.

The two boxes and the captions of the five illustrations betray Horgan's intent: claims of genetic influence on psychological characteristics are alleged to be overblown or doubtful and to have been recently retracted or deemed unpublishable. One half-page box reminds the reader that Hitler was an enthusiastic eugenicist and thus, presumably, had much in common with the modern behavioral genetics researcher. But as those who are familiar with contemporary behavioral genetics literature will know, these baseless accusations are merely an attempt to win with scare tactics that which has not been won in the research laboratory.

Apart from direct assessments, the status of an individual's identical twin is the single best predictor of risk for schizophrenia, manic-depressive illness, alcoholism, IQ and personality. Moreover, evidence from twin studies is consistent with both adoption studies, which show that adoptees resemble psychologically their biological parents more than their adoptive parents, and family studies, which demonstrate that the psychological similarity among relatives is directly related to their degree of genetic relatedness.

Explanations for behavioral genetics findings summarized in Horgan's article are either laughable (as when he says the similarity in sexual orientation between twins owes to having been dressed alike as children), disingenuous (as when twins reared apart are said to owe their similarity to contact between the twins even though both the Minnesota and the Karolinska groups have tested and rejected that possibility), or misleading (as when Horgan features the only study of alcoholism in male twins that failed to report significantly greater concordance between identical than nonidentical twins—even though that study involved a far smaller sample than any of the five other studies).

Ironically, a case for behavioral genetics is made in the article on "Autism," by Uta Frith, in the same issue. A generation ago behavioral scientists ascribed autism to, among other things, the inadequacies of "refrigerator mothers." As Frith points out, twin studies have shown that "autism can have a genetic basis," and biobehavioral models of autism are now favored. Horgan, and the select group of critics he promotes, may long for the bygone days of radical environmentalism, but thankfully those days are past.

MATT MCGUE
Department of Psychology
University of Minnesota

Co-signers include 16 scientists from eight institutions in the U.S., Australia, Sweden and the Netherlands; list available from McGue.

Horgan replies:
I'll respond to just three points made by McGue et al. First, nowhere did I impugn their motives as politically suspect. But since they raise the issue, let me note that the major sponsor of the Minnesota twin studies is the Pioneer Fund, a private foundation that has also supported William Shockley and other proponents of racial theories of intelligence. Second, the chief critics of the Minnesota twin studies are not "radical environmentalists" but other behavioral geneticists, who believe the methods of the Minnesota group are biased toward high heritability. Finally, a growing number of investigators suspect that viral infections or physical traumas occurring during pregnancy might cause autism—possibilities that do not fall neatly into either the nature or nurture category.

Cochlear Implants

I commend John Rennie on "Who Is Normal?" ["Science and the Citizen," SCIENTIFIC AMERICAN, August]. As one of a few audiologists who respect and actively elicit the view of the American deaf community, I am thrilled to see at last such an unbiased article about this controversial topic of cochlear implants.

I would have to agree with Robert Shannon, who stated so assuredly, "I don't think . . . that deaf people are well integrated into society at large." I would have to add that this is true for African-Americans, Hispanic-Americans, female Americans, gay Americans, poor Americans, handicapped Americans and other oppressed minorities. "Society at large" in this country means white, upper middle class, Protestant, well educated and male. We, as a country, shape, coerce and even demand our inhabitants to conform to this mold or be classified as a second-class citizen. How unfortunate. How sad.

We should no more be trying to make deaf children hearing or Little People taller than we should try to make African-Americans white or women into men. If we can stop making assumptions long enough to listen to those who are deaf, listen to those who are Little People, listen to those who are African-American . . . then we can hear the *truth.*

HOLLY M. GEESLIN
Indianapolis, Ind.

Can't Get There from Here

In "Australia's Polar Dinosaurs" [SCIENTIFIC AMERICAN, July], Patricia Vickers-Rich and Thomas Hewitt Rich speculate that the tendency toward dwarfism shown by populations on islands may be a response to selective pressure to increase the number of individuals so as to ensure a diverse gene pool. Yet selective pressure can reduce the average size of a population only if a small individual achieves greater reproductive fitness than its larger cousins. The prospect of retaining a diverse gene pool many generations into the future cannot have the effect of increasing the frequency of a gene for small size.

ANDREW PAGE
Langport, England

Vickers-Rich and Rich reply:
You are correct. Space considerations forced us to abbreviate our presentation of the mechnisms causing dwarfism in island populations. Page 196 of our book *Wildlife of Gondwana* (Reed Books, Sydney, 1993) has a more thorough treatment of this topic.

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

50 AND 100 YEARS AGO

NOVEMBER 1943

"Air-conditioning of submarines is now possible through use of a non-toxic, non-explosive fluid, called 'Freon-12,' fluorine refrigerant, which is non-poisonous, has no odor, and will not support flame. It does not explode should it come into contact with the electric stoves of a sub's galley, nor does it interfere with the chemicals which purify the air. The men aboard the underseas vessels so equipped can even smoke."

"Glass with non-reflecting surfaces, developed for military uses by American Optical and RCA, can be applied, with desirable results, to post-war manufacture of many useful items. Among the new products are windshields sans dangerous reflections, less conspicuous spectacle lenses, more easily read instruments, faster camera lenses, shop windows free from reflections, more efficient microscopes and other light-transmitting instruments."

"Newspapers and magazines of today frequently predict a post-war future including a private airplane in every garage. General Aircraft Corporation has opened up a bit in regard to its plans. Here is a prophetic quotation: 'Our business man leaves his home in the morning in his 'car,' drives to the airport. While having his 'car' filled with gas, the attendants put on the wings, a five-minute job. After flying to his destination, he has the wings removed, drives his 'car' downtown, makes his necessary calls, drives back to the airport, and, donning his wings, goes on to his next destination by air.'"

"Rayon and other fibers are cutting deeply into cotton's tire-cord monopoly and are threatening other strongholds. Science, however, is starting to alter the situation. Designs for cotton goods are being developed in many forms; chemical treatments are being worked out to change the feel, the appearance, and the quality of cotton fabrics; cotton is being made water-proof, rot-proof, fire-proof, and spot-proof; agricultural experts are developing plants which will produce better grades of the fiber in larger quantities."

SCIENTIFIC AMERICAN

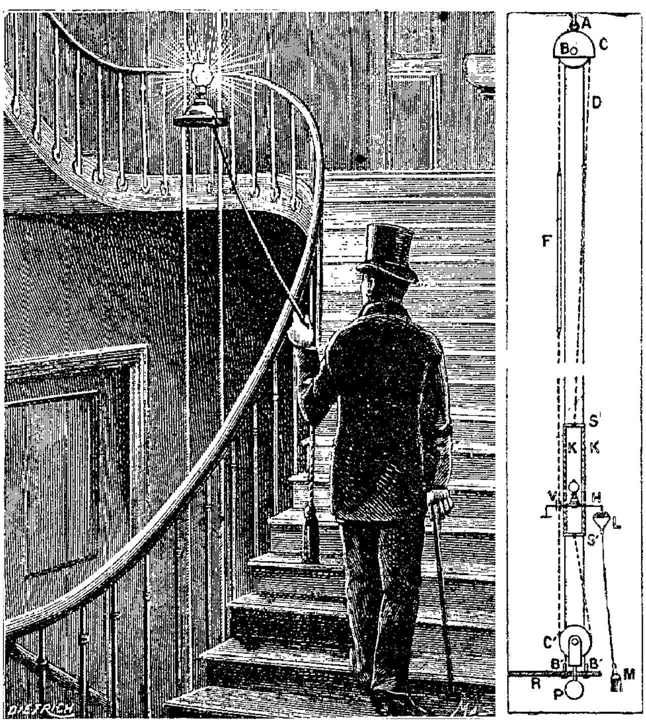
NOVEMBER 1893

"If ordinary placental mammals have evolved from pouched animals like the modern marsupials, rudiments of the pouch ought certainly to be recognizable in some of them. Dr. H. Klaatsch has just made the interesting announcement that such rudiments can actually be observed in most placentals. Something of the kind has already been found in the lemurs, and one author has supposed that rudiments of the pouch can also be detected in the sheep."

"'Once I placed an aerial motor on a pair of Fairbanks scales and set it going,' says Thomas A. Edison. 'It lightened the scales, but it didn't fly. Another time I rigged up an umbrella-like disk of shutters and connected it with a rapid piston in a perpendicular cylinder. These shutters would open and shut. If I could have got sufficient speed, say a mile a second, the inertia or resistance of the air would have been as great as steel, and the quick operation of these shutters would have driven the machine, but I couldn't get the speed. I believe that before the air ship men succeed they will have to do away with the buoyancy chamber.'"

"The American Telephone and Telegraph Company recently gave an exhibition of their long-distance telephone lines to a small party of guests who assembled at the Telephone building in Cortlandt Street. Among those assembled were Dr. Von Helmholtz and Prof. Alexander Graham Bell. A number of receivers were arranged so as to give each of the party a connection to the line. Connection was made with Boston, Chicago, and Washington in turn, and conversations were held with the officers at those points. A cornet was also played which was heard through 500 miles of wire as distinctly as though it were in an adjoining room."

"It is indispensable for the sake of economy, and especially for safety, to shut off the gas at the meter for the night in every house. The movable night lamp, which operates at an expense of but one cent a night, presents the advantage of accompanying those who go up or down stairs after the gas has been put out. It suffices to grasp at the bottom of the staircase a light counterpoise fixed to the lamp by a cord, and the lamp then ascends with the person and affords him light progressively. When the story at which one is to stop is reached, the lamp, upon the weight being released, descends of itself to the bottom of the stairway. In order to descend with a light, it suffices to raise the lamp through the chain that supports it (an operation that requires three seconds) and to grasp the counterpoise. The lamp then follows the person to the bottom of the staircase."



Movable lamp for stairway

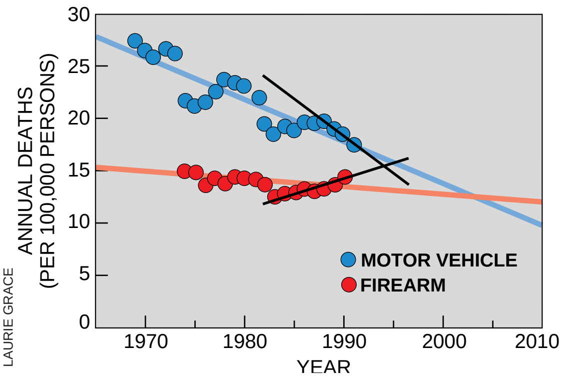
Grim Statistics

Gunfire may surpass auto accidents as a cause of death

The European tourists who were shot by highway "hunters" in Florida were driving cars that were legally required to have seat belts and may even have been equipped with airbags.

Whereas nationwide concern with automobile safety has led to improved crash-worthiness and tougher laws for drunken driving, the number of deaths caused by gunfire continues to increase. Will the declining curve of auto-related mortality intersect with the rising curve of deaths from firearm use?

The most authoritative statistics indicate that the question is not "will?" but "when?" According to Garen Wintemute of the University of California at Davis, guns may move into first place during the next decade.



LAURIE GRACE

LONG-TERM MORTALITY TRENDS for motor vehicles and firearms* (colors) *converge in 2003, short-term ones* (black) *in 1994.

Wintemute's comparison of gun and automobile mortality statistics (*left*) was published in the *Journal of the American Medical Association*. The date on which the nation achieves the crossover—some reports reveal that Louisiana and Texas have already done so—depends on the stability of current trends. Deaths from gunshot wounds have increased rapidly during the past five years (after a decade of decline), whereas automobile fatalities are falling faster than usual, as they tend to do in bad economic times. If this new pattern persists, more people will die from gunfire than in auto accidents during 1994. But if long-term historical trends reassert themselves, the crossover will wait until a few years after the turn of the century. —*Paul Wallich*

Dot's Incredible

Controlling single electrons in a quantum dot

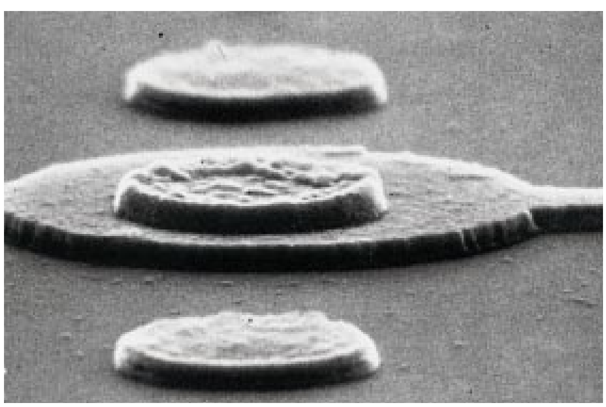
Manipulating small numbers of atomic particles seems to have become a standard part of the repertoire of physics. So devotees of the art are being dazzled by a supreme feat of nanoscale sleight of hand, which has been achieved by researchers at AT&T Bell Laboratories.

The Bell Labs workers, Raymond C. Ashoori, now at the Massachusetts Institute of Technology, and Horst L. Stormer and their colleagues, report in *Physical Review Letters* that they can control the behavior of as few as one or two electrons in a patch of semiconducting material that is only a few tens of nanometers square. This level of resolution was previously thought to be unattainable. The success should enable investigators to explore quantum phenomena that have never been observed in an experimental setting and might serve as a basis for significant technological advances.

The semiconductor specks are known as quantum dots, or artificial atoms. Although many real atoms actually constitute a quantum dot, the electronic properties of a dot make it the equivalent of an individual atom. Like a real atom, a quantum dot harbors distinct numbers of electrons. But rather than being held in place by the charge of a nucleus, the electrons in an artificial atom are confined by boundaries of a material. Trapped in such a box, the electrons occupy discrete energy levels, just as they do when bound by a real nucleus. A quantum dot is constructed from a film of semiconducting material, such as gallium arsenide, sandwiched between two insulating layers. The lithographic processes used to etch circuit patterns can form the artificial atoms [see "Diminishing Dimensions," by Elizabeth Corcoran; SCIENTIFIC AMERICAN, November 1990; and "Quantum Dots," by Mark A. Reed, January].

Detailed studies of the properties of quantum dots have been difficult. The standard method of examining their electronic characteristics—measuring the charge flowing through them—was limited in resolution. "The current is small, and you have to put 30 to 40 electrons into the artificial atom before current flows," according to Marc A. Kastner, an M.I.T. investigator who also explores artificial atoms.

But Ashoori had a dream of looking at electrons one by one as they accumulate to form an artificial atom. While working at Bell Labs, he and his colleagues decided to try measuring changes in the amount of charge (that is, the capacitance) caused by the dot rather than the amount of current flowing though it. The technique, single-electron capacitance spectroscopy, calls for placing an artificial atom between two electrically conducting plates. "We then apply a 'tickling' voltage to induce an electron from one of the plates



RAYMOND C. ASHOORI

QUANTUM DOTS are fabricated inside metal disks about one micron in diameter. A contact loop collars the middle disk and transmits the signals from the tunneling electrons inside to measurement devices.

to tunnel," Ashoori explains. The laws of quantum physics give the electron, which does not have enough energy to move from the plate to the semiconductor, a temporary boost. The particle can then tunnel through the energy barrier to make the trip. When it does so, it becomes bound to the artificial atom. The electron does not bond to a real atom, because according to the quantum mechanics of solids it is a free electron. Free electrons do not feel the presence of real atoms in the material.

Ashoori knows when an electron has tunneled to the artificial atom, because the particle's movement induces a minuscule but detectable charge to form in the other plate. By changing the voltage across the plates, the investigators can make electrons tunnel one by one to the artificial atom. Only the temperature of the sample, which must be kept near absolute zero, limits the resolution.

The physicists grant that the capacitance technique may have some practical use. It might, for instance, act as a foundation for a photodetector that counts single electrons. The device would be superior in performance to existing detectors by a factor of 10. The dots themselves might also be employed as the ultimate tiny circuit element. Ashoori and Stormer point out, however, that the true strength of the work lies in basic research. "It is a toy box, an incredibly powerful microscope," Ashoori says.

But why look at artificial atoms when there are plenty of natural ones lying around? The answer is that an artificial atom differs in promising ways from the real McCoy. Quantum dots are several hundred times larger (a hydrogen atom is about 0.1 nanometer in diameter), and the "walls" that trap electrons in a dot are not as symmetric as the nuclear charge that holds electrons. Such differences, the researchers say, open a new realm of physics.

For example, tests of quantum effects that require temperatures, field strengths and other conditions well beyond those achievable with today's equipment become possible in artificial atoms. One is the influence a magnetic field exerts on confined electrons. According to the Pauli exclusion principle, no two electrons can occupy precisely the same state. The two electrons in a helium atom, lying in their lowest energy state, distinguish themselves by orienting their "spins" in opposite directions. An external magnetic field, however, tends to force the spins to align, which would put the two electrons in the same quantum state. So, theory predicts, one electron must jump to a higher energy level.

To conduct the experiment on real atoms, workers would have to use a magnet that would generate an external field of about 400,000 teslas. Even the sun does not produce such a mighty field. The superconducting magnets used in magnetic resonance imaging typically create fields of about 0.5 to 1.5 teslas. In an artificial atom, Ashoori notes, a field of less than two teslas suffices to make an electron jump to a new energy level. Using quantum dots, physicists may also be able to probe much more rigorously such unusual phenomena as quantum chaos and the quantum Hall effect.

Customizing quantum dots is also a possibility. "The nice thing is," Stormer comments, "you can make any kind of artificial atom—long, thin atoms and big, round atoms." Then, one can string together many of these quantum dots, creating an artificial molecule. The artificial molecules can in turn be joined to make artificial solids—an intriguing prospect to many physicists. "What is driving the excitement," Kastner explains, "is the hope that there is something there we didn't expect." —*Philip Yam*

Were Four Corners Victims Biowar Casualties?

Could a mysterious disease that has taken at least 16 lives in the Four Corners region of the Southwest since this past May be related to the U.S. biological warfare program? In June, federal and state investigators blamed the outbreak on hantaviruses. Although hantavirus-related illnesses were unknown in the U.S. before this year, they have been studied by military and civilian researchers since the 1950s, when U.S. troops fighting in Korea became infected with a flulike disease that attacks the kidneys.

The virus, named after Korea's Hantaan River, is carried by rodents and is transmitted by airborne particles of the feces or urine of infected animals. The Four Corners illnesses were almost certainly caused in this way, asserts C. Mack Sewell, an epidemiologist for the state of New Mexico, who notes that the virus had previously been detected in deer mice in the area.

Rumors have nonetheless persisted among Native Americans and others in the Four Corners region that Fort Wingate, an army base near the epicenter of the epidemic, was somehow involved. In June, Senator Jeff Bingaman of New Mexico queried the Pentagon about possible biological warfare activities at the base. The Pentagon acknowledged that the fort was once used as a storage depot for chemical weapons but denied that biological weapons were ever held or tested there.

Yet Fort Wingate has served as a target site, or "impact zone," for missiles launched from other military bases, according to a former congressional investigator who requested anonymity. One possible launch site is the Dugway Proving Grounds in Utah, several hundred miles to the north. The army has conducted experiments at Dugway with both chemical and biological agents for decades. Dugway earned notoriety in 1968 when a jet aircraft from the site accidentally released nerve gas over a nearby ranch and killed thousands of sheep.

The investigator suggests that tests initiated at Dugway may have infected the Fort Wingate region with biological agents years ago. The epidemic may then have been triggered by demolition or other disturbances related to the decommissioning of Fort Wingate early this year.

There is also reason to doubt that all the Four Corners illnesses stemmed from hantavirus, the investigator notes. Fewer than half of the victims tested positive for hantavirus. Moreover, deaths were attributed not to kidney failure—the usual outcome of hantavirus infection—but from hemorrhaging of the lungs. Congress recently appropriated $6 million for a study of the Four Corners outbreak.

Whatever the conclusions of the study, the suspicion engendered by the incident shows the need for greater openness within—and perhaps demilitarization of—the biological defense program, argues Leonard A. Cole of Rutgers University, an authority on the history of biological warfare. "It would be in the army's interest to eliminate the conspiratorial attitude toward these outbreaks," he points out. This year, Congress required the Department of Health and Human Services to examine the feasibility of shifting some biological defense research from the army to the National Institutes of Health. —*John Horgan*

Insects Are Forever

Staying power, not flower power, made bugs diverse

Anyone who has ever shared an apartment with cockroaches has suspected as much, but now it's official: insects almost never go away. After surveying the fossil literature, two researchers have concluded that at the family level, insects have shrugged off catastrophes that exterminated fragile, dainty creatures—such as the dinosaurs. "Because of the low rate of extinction, you have insect lineages that are very long lived, approaching 100 million years in some cases," notes Conrad C. Labandeira, one of the new study's authors and a paleoentomologist at the National Museum of Natural History of the Smithsonian Institution. That family durability seems to explain why bugs are so numerous and varied today.

By almost any standard, insects are phenomenally successful. They were the first animals to invade the land and, later, the air. They are the most diverse group, too: by some estimates, about 876,000 insect species have been identified, and entomologists believe a full tally would be in the millions. (By comparison, taxonomists know of only about 4,000 mammal species.) According to Douglas Futuyma, an expert on insect evolution at the State University of New York at Stony Brook, insects' success has often been attributed to a presumably exceptional talent for becoming new species. Agricultural scientists know, for example, that insects can readily evolve new traits, such as resistance to pesticides. Some experiments also suggested that specific groups of insects, such as the fruit flies in the Hawaiian Islands, also diverged into separate species very quickly.

But the report recently appearing in *Science* indicates that adaptability may have been less important for insects than sheer, stubborn endurance. Since the mid-1980s, Labandeira and J. John Sepkoski, Jr., of the University of Chicago have been searching the fossil record for evolutionary patterns in insect diversity and survival. They note that many scientists have assumed that insects do not fossilize well. "There's been this received wisdom that because insects aren't durably calcified like mollusks or the bones of vertebrates, there wouldn't be much of a fossil record," Labandeira remarks. In fact, the literature from old German, Russian and Chinese sources was rich enough for Labandeira and Sepkoski to gather information about 1,263 extinct and extant insect families. Only about 825 families of four-legged animals (vertebrate tetrapods) have been documented as fossils.

Those data demonstrated that families of insects rarely disappeared, even when other animal groups were perishing en masse. The researchers found, for example, that 84 percent of the insect families living 100 million years ago, during the Cretaceous period, are also present today. In contrast, only 20 percent of the Cretaceous tetrapod families are still around. The mass extinctions at the end of the Cretaceous destroyed about one quarter of the tetrapod families (including all the dinosaurs), but the effect on insects was negligible. Indeed, the only extinction event that had a major impact on insect diversity was the huge one at the end of the Permian period, 250 million years ago. It wiped out 65 percent of the insect families then living, probably because nearly all vegetation died at the same time.

Labandeira and Sepkoski's findings do not contradict the possibility of rapid speciation in insects. Labandeira says that, if anything, long-term survivorship of families and rapid turnover of species may go hand in hand. Because greater species diversity promotes the survival of a family and surviving groups have more opportunities to diversify, the trend is self-perpetuating: nothing succeeds like success.

To the surprise of some biologists, Labandeira and Sepkoski also observed that the appearance of flowering plants, or angiosperms, 125 million years ago did not cause a burst of insect diversity. "As a matter of fact, the rate of diversification abated," Labandeira emphasizes. That finding was unexpected because insects and flowering plants often live in intimate, species-specific associations.

One explanation, the researchers posit, is that the evolutionary effects of the angiosperms might have been invisible to their study: the diversity they promoted might have been at the species rather than the family level. And Futuyma notes that the order Lepidoptera (butterflies and moths) is underrepresented in the fossil record. Because lepidopteran insects have some of the closest associations with flowering plants, he thinks their absence might disguise some diversification.

Yet Labandeira and Sepkoski also offer the theory that for insects, the angiosperms were not very novel challenges. They discovered that most types of mouthparts found in modern insects were present 100 million years before angiosperms evolved. Insects that were already dining on gymnosperms, conifers and other seed plants did not need radical adaptations to take advantage of the new flora. "We live in an angiosperm-dominated world," Labandeira reflects. "It's hard for us to picture how

ED BRIDGES/*American Museum of Natural History*

FOSSILS OF INSECTS suggest that their taxonomic families are highly resistant to extinction, which may explain why insects are so diverse today. This snake fly fossil from a limestone deposit in Brazil is 120 million years old.

well insects thrived in a world with different vegetation."

He and Sepkoski end their paper with a warning that humanity's extensive deforestation efforts might trigger a calamitous loss of insect diversity. That statement might seem paradoxical, given insects' historical resilience. Labandeira acknowledges that it was more of a cautionary speculation than an analysis and that "anything that's happening today may be mild compared with what happened during the late Permian." Still, some insect groups are highly important to ecosystems, and deforestation can eliminate them ruthlessly. If hardy insect clans are suffering, other fauna and flora may be even more debilitated. Think about that the next time you reach for a flyswatter. —*John Rennie*

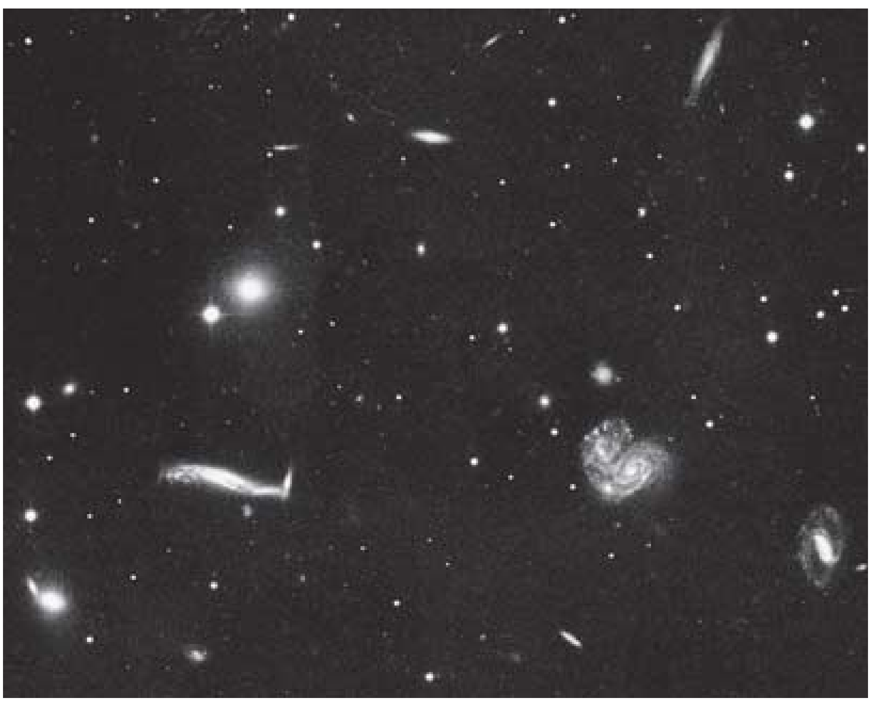
Goldilocks Cosmology

Theorists toss another ingredient into the cosmic recipe

At times, the story of modern cosmology sounds oddly like the tale of Goldilocks and the three bears. Some theorists have proposed that the mass of the universe is dominated by fast-moving invisible particles known as hot dark matter; others favor a universe dominated by sluggish cold dark matter. In either case, the unseen material helps to explain how large structures (such as galaxies and clusters of galaxies) emerged from the hot, expanding mass that existed after the big bang. But neither kind of dark matter seems entirely able to account for the observed organization of the cosmos. A number of researchers are therefore exploring a third scenario in which the universe contains a nearly even blend of hot and cold dark matter. And in good Goldilocks fashion, they argue that such a mix may work "just right."

Cosmologists have tended to shy away from mixed dark matter models, in part because "the subject is often guided by aesthetic simplicity. Most people thought mixed dark matter was very ugly," reflects Nick Kaiser of the Institute for Theoretical Astrophysics at the University of Toronto. Kaiser and his coworkers Robert A. Malaney and Glenn D. Starkman think they have addressed such reservations by finding an attractive way to create two kinds of dark matter through a single mechanism.

In a recent paper in *Physical Review Letters,* the researchers envision a universe that initially contained a population of massive neutrinos, neutral particles that barely interact with normal matter. Physicists commonly assume that neutrinos have no mass, but mounting evidence suggests otherwise. Kaiser and his collaborators propose that the massive neutrinos could have decayed in such a way as to stimulate the formation of slow-moving (and hence "cold" in cosmological parlance) dark matter particles. The workers call this mechanism "neutrino lasing," by analogy to the stimulated creation of photons of light in a conventional laser. The heavy neutrinos themselves decay into lighter, high-speed particles (possibly another form of neutrino) that constitute a component of hot dark matter. In this way, a single, fairly elegant set of events can account for the existence of two separate components of dark matter.



MOUNT WILSON AND PALOMAR OBSERVATORIES

CLUSTERS OF GALAXIES, such as this one in the constellation Hercules, may have assembled under the gravitational coercion of vast clumps of unseen dark matter. But the simplest dark matter models do not match the observed cosmic structure.

Neutrino lasing occurs at such high energies that "it could be very very difficult indeed" to devise a laboratory test to prove the existence of the phenomenon, Kaiser admits. "What we are presenting here is a new piece of physics," he explains; now it is up to the particle physicists to find a place for it in the broader context of their theories.

Even if the idea does not pan out, neutrino lasing is far from the only way to create mixed dark matter. "There are lots of more mundane ways to do it," says Robert K. Schaefer of the Bartol Research Institute at the University of Delaware. Indeed, from a particle physics point of view, "it's sort of natural" to have both hot and cold dark matter, he says. Schaefer sees great promise in two-component dark matter cosmological models. New observations have competing models "scrambling after data points," he claims, whereas the latest findings are "settling more and more toward mixed dark matter."

Some cosmologists still object to the notion of mixed dark matter on aesthetic grounds. "I've seen people get up after talks and say, 'This is the ugliest model I've ever seen'—there's no scientific rationalization," Schaefer reports. Jeremiah P. Ostriker of Princeton University agrees that the lack of simplicity is a poor argument against mixed dark matter models. "Who's to say that nature will be simple? Biological systems are a mess," he laughs.

Ostriker objects to the simplest mixed dark matter models for a very different reason: in his opinion, "they don't work." Astronomical observations reveal that galaxies and quasars existed within a couple of billion years after the big bang and large clusters of galaxies not long thereafter. Mixed dark matter cosmologies cannot readily explain how such objects could have formed so soon after the big bang.

Kaiser readily concedes that difficulty but thinks the various bits of evidence indicating at what era large galaxy clusters began to form remain equivocal. "You pay your money, and you take your choice," as he puts it. Ostriker, in contrast, feels the inability of mixed dark matter to account for the appearance of the early universe means “it is probably not correct.”

Mixed dark matter represents only one of a number of theoretical tweaks that cosmologists are using to fine-tune their models to fit the observations. Some workers posit that cosmic strings—hypothetical, extremely dense defects in the structure of space left over from the first moments of creation—could have acted as seeds around which galaxies formed. Other theories invoke alternative but equally hypothetical mechanisms for creating dense structures very early in the history of the universe.

Each time new data come in, Ostriker notes, the easiest thing to do is “just add another epicycle” to existing cosmological theory. Mixed dark matter adds one layer of complexity to the previous models, most of which incorporated cold dark matter alone. “But nature could be nasty; there could be cold dark matter, hot dark matter *plus* strings,” Ostriker muses. Or the universe could be far simpler than most astronomers imagine. Despite many claims to the contrary, Ostriker maintains that there is still no solid evidence for exotic dark matter. If such dark matter does not exist, then one could build a model “based on hydrogen, tables and chairs—stuff we know about,” he comments.

The joy of speculating about the early history of the universe—as well as the frustration—is that the possibilities are nearly endless. Goldilocks had but three bowls of porridge to sample. Only the human imagination limits the menu of cosmology. —*Corey S. Powell*

Sausage Factory

How Congress passes the pork to Back-Home U.

Back in 16th-century England, when livestock grazed on a common, farmers would identify their swine by special marks on the animals’ ears. In 20th-century America, earmarks of a different kind are increasingly being used to distribute federal pork to colleges and universities.

An investigation conducted by Congressman George E. Brown, Jr., of California, chairman of the House Committee on Science, Space and Technology, shows that during the 1980s the practice of cajoling Congress into supporting academic projects that had not been requested by the executive branch, subjected to competitive review or scrutinized by any congressional authorizing committee grew to majestic proportions—at least by university standards.

A select but expanding group of colleges now routinely taps federal funds by lobbying influential members of Congress to insert special provisions, earmarked to fund specific projects, into the appropriations bills and reports for federal agencies. More than $700 million so earmarked was appropriated in 1992. In 1980 the total was $11 million. Brown points out that individual earmarked appropriations ranged in value from a few hundred thousand dollars to more than $40 million this past year.

In general, authorizing committees in Congress approve the overall direction of agencies’ spending, and then appropriations committees vote the funds to be used. But by sliding in an earmarked provision at a late stage in the appropriation process—often in the conference committee, which reconciles House and Senate versions of a bill—a member can secure funds for a project that might not survive a measured consideration.

Brown complains that the practice “destroys rational efforts to set priorities tied to national needs” and “fails to protect the taxpayers’ investment.” Many unreviewed allocations, he notes, were forced onto unwilling federal agencies that consequently had little choice but to spend the money or risk a confrontation. In this way, the Department of Energy was pushed into building hospitals, for example, and the Federal Aviation Administration was directed to spend $30 million this past year on an “airway science” program that it would like to terminate.

Although less than 5 percent of higher education institutions receive earmarked funds, Brown’s “just say no” campaign faces an uphill battle. The number of universities retaining lobbyists in Washington—at fees of up to $60,000 a month, according to Brown’s staff—is escalating. Brown found that 21 out of 50 academic institutions that received allocated funds in fiscal 1993 had employed registered lobbyists the previous year. Moreover, the same schools keep showing up time and again in the chow line. Thus, eight of the top 10 recipients in 1992 were among the top 20 between 1980 and 1992. Iowa State University, the University of Alaska, Oregon Health Sciences University and Louisiana State University head the list of all-time winners.

Martin C. Jischke, president of Iowa State, says the earmarked projects at his university, which include a center for advanced technology development and an experimental food irradiation facility, are “important and defensible.” Moreover, he asserts, “there was no competitive federal program to which we could apply.”

Other recipients of unreviewed targeted funds responded to Brown’s survey by saying they as lesser lights in the scholastic firmament would be unable to compete with better-established schools. But that plea does not stand up in most cases. Many of the institutions that receive the largest of such allocations fall in the top 25 percent of recipients of peer-reviewed research grants from the National Science Foundation, according to Brown’s staff.

Recipients insist on their right to lobby Congress and point out that because some funds are “leveraged”—that is, the institution itself provides funds to match the federal dollars—they represent cost-efficient federal spending. But many such funds are not leveraged. And the contention that they send federal dollars to poor states is contradicted by another of Brown’s findings. What many recipients have in common, Brown’s study shows, is a senator or congressman in an influential position on an appropriations subcommittee.

Some political tides may be flowing in Brown’s favor. Congressman William H. Natcher of Kentucky, the new chairman of the House Appropriations Committee, has set his cap firmly against earmarking funds for academic groups. And as budgets get tighter, members of authorizing committees in the House are becoming increasingly sensitive to the threat that such appropriations pose, observes Peter Smith of the Association of American Universities.

But universities themselves seem to find it hard to speak with one voice on the subject. In 1987 members of the association, which represents major research institutions, voted 43 to 10 to observe a moratorium on seeking earmarked funds. Since then, several members who voted in favor of the moratorium have “slipped” and now accept earmarked money, Smith says. —*Tim Beardsley*

“Pollution, Pollution...”

Federal air standards permit dangerous particulate levels

It’s enough to make Tom Lehrer sit down at the piano again. Findings from a recent study indicate that loopholes in government standards have let one of the most harmful forms of air pollution become a dangerous fact of everyday urban life.

The study, presented at the annual meeting of the American Lung Association by C. Arden Pope, a visiting scientist at the Harvard School of Public Health, measured the effects of air pollution on residents in six U.S. cities. Pope found a 26 percent higher risk of premature death in the city most polluted with ambient particles as compared with the least polluted city surveyed. The workers also noted "robust associations" between chronic exposure to airborne particulate matter and increased mortality.

As disturbing as the results is the fact that the air in all six cities carried particles whose density per volume of atmosphere was well below legal thresholds. The current Environmental Protection Agency regulations mandate that density of particles less than 10 microns in diameter shall not exceed 150 micrograms per cubic meter of air during a 24-hour period. The Harvard data correlate morbidity and mortality statistics with the presence of particles one quarter the size specified in the EPA regulations. The density of such particles did not on average exceed one third that specified by the EPA benchmark.

The Harvard team culled its statistics from an analysis of 8,111 residents in the six cities, whom it followed for 14 to 16 years. Pope, who came from Brigham Young University to participate in the Six Cities Study, says the conclusions point toward fine particles and particles from the combustion of fossil fuels as the most pernicious air pollutants.

Such pollutants include carbon, hydrocarbons, dust, acid aerosols and sulfates. Common respiratory problems that can develop from exposure to these pollutants are chronic obstructive pulmonary disease, cardiovascular disease and asthma.

"If you ask the average layperson, these results would probably come as no surprise, but it has taken a while for science to catch up with common sense," says Alfred Munzer, president of the American Lung Association (ALA). "This is the first time that we have hard data to show not just the morbidity caused by particle pollution but the increase in mortality as well."

A 1992 report by Joel Schwartz, an epidemiologist at the EPA, and Douglas W. Dockery of Harvard, who also contributed to the Six Cities Study, estimated that respirable particles may cause some 60,000 premature deaths in the U.S. every year. Previous accounts compiled by Pope, Schwartz and Dockery have linked death rates and particulate pollution levels around the country for decades. Critics argued that these studies did not adequately control for individual risk factors, such as tobacco smoking, and so did not warrant remedial legislation or regulatory action.

ROBERT PROCHNOW

PARTICLE MONITOR, which neighbors a midtown Manhattan bus stop, was moved back from the curb in 1990 to meet site criteria of the Environmental Protection Agency. In its previous position, the equipment consistently collected fine particle levels from diesel exhaust far exceeding federal standards.

But the Six Cities Study directly addressed such factors. The findings have forced doubters out of that defensive ditch. Indeed, the statistics have prompted the ALA to announce that it intends to file suit against the EPA to demand a timely review of policy regarding particle pollution. "It is imperative that we revise our standards, particularly when it comes to particles," Munzer says.

The EPA has not reviewed its standards for particulate matter since 1987, even though the Clean Air Act of 1970 requires it to do so every five years. Schwartz, who was one of the first epidemiologists to document the dangers of particles, is frustrated with the delay. He notes that EPA officials have targeted 1999 as the earliest possible date for a policy change regarding particulate matter. "I think the difference between reviewing particle standards at Thanksgiving 1999 and Christmas 1999 is more important than all the other regulations the EPA plans to put out between now and then," he emphasizes.

EPA officials say they are moving as fast as they can. "We have planned to set up an expedited schedule to review particle standards," says Robert D. Brenner, chief of policy for the EPA's Office of Air Pollution. "Now we know that there is a serious particulate problem, but that doesn't necessarily tell you how to set the standards," he cautions. Schwartz points out that to speed the review process, funds and workers would have to be reallocated from other projects.

One problem for scientists, both in academia and at the EPA, is that these particles, no more than 10 microns in diameter (less than half the width of an average human hair), are extremely difficult to examine. They come from a variety of sources: construction work, cars driving over dirt and paved roads, wind erosion, tobacco smoke, fireplaces and even backyard barbecues. Moreover, chemical reactions catalyzed by sunlight create harmful particle concoctions that are difficult to isolate from other forms of air pollution.

Or so argue the investigators. The skeptics want more proof. Some critics contend that a biological explanation of how particles cause disease or death must be demonstrated before the policy can be changed. One unanswered question is whether the particles themselves cause damage or whether blame belongs to the chemicals they carry deep into the lungs.

Many researchers feel policy decisions cannot be left up in the air until the biological grounds for damage associated with particles are pinned down. Munzer hopes the threat of an ALA lawsuit will put pressure on the EPA to make a change soon. "These things usually work, but it shouldn't be necessary," he laments. "The process for including scientific progress in public policy needs to be a much more rapid one." —*Kristin Leutwyler*

Unraveling Alzheimer's

A major cause of the disease yields to researchers

Workers at the Duke University Medical Center have identified what seems to be a critical factor in the development of Alzheimer's disease, the degenerative brain disorder that afflicts four million people in the U.S. The factor may be associated with about 80 percent of the cases of the illness. Identification of the factor, a form of a gene responsible for the manufacture of a lipoprotein, has been confirmed by 10 other laboratories.

The Duke researchers—Margaret A. Pericak-Vance, Ann M. Saunders and Allen D. Roses, among others—have found a strikingly clear association between the onset of Alzheimer's and a particular variant of a gene that codes for a known blood protein, apolipoprotein E. The suspect gene, *APOE*-ε4, can be detected with a test that is already widely used for diagnosing a serious cholesterol-transport disorder.

Investigators had previously discovered in a few cases of the relatively rare early-onset form of Alzheimer's a mutation in the gene responsible for the production of amyloid beta-protein, which is deposited in the brains of patients. Some other cases appeared to be linked to a different unknown gene. But these findings had not seemed relevant to the majority of patients.

The first clue that a variant of the apolipoprotein E gene might be involved fell to a group in Roses's laboratory led by Warren J. Strittmatter. They found last year that apolipoprotein E binds to the beta-amyloid deposited in the brain of patients suffering from Alzheimer's. The gene for apolipoprotein E is on chromosome 19. Pericak-Vance had just identified some cases of Alzheimer's in families in which the disease was inherited along with the section of chromosome 19 that includes the apolipoprotein E gene.

Roses's group immediately started studying apolipoprotein E. The payoff was not long in coming. Saunders discovered that patients in families afflicted with Alzheimer's are more likely than are other people to have the particular form of the apolipoprotein E gene known as *APOE*-ε4.

In August, Saunders published a confirming report in *Neurology* that the association holds in so-called sporadic cases, in which there are no other affected family members. Other workers have found evidence of the association in autopsied patients, in living patients and in pilot epidemiological surveys. "Until this, there hadn't been a lot of progress since Alois Alzheimer found plaques and tangles in the brains of his patients in 1907," Roses comments.

"This is a major finding. It's not only right, it's important, too," remarks John A. Hardy of the University of South Florida. "All the papers suggest that having one *APOE*-ε4 gene increases your risk of Alzheimer's threefold to fourfold and that people with two *APOE*-ε4 genes are very likely to develop Alzheimer's." About 15 percent of the U.S. population has one *APOE*-ε4 gene and are thus at elevated risk. Between 1 and 1.5 percent of the elderly has two such genes, and Roses has found that they tend to acquire Alzheimer's earlier than do people with one *APOE*-ε4 gene.

Although nobody is sure exactly why *APOE*-ε4 makes the development of Alzheimer's more likely, Roses says he has definite ideas, whose therapeutic potential he plans to pursue. Dennis J. Selkoe, an Alzheimer's researcher at Harvard University and founding scientist of Athena Neurosciences, a South San Francisco biotechnology company, suggests an explanation. He speculates that the protein made by *APOE*-ε4 interferes with the removal of amyloid beta-protein from the brain. Or possibly, Selkoe says, the apolipoprotein encourages deposition of the material.

A drug that inhibited the interaction of the two proteins might slow down or eliminate whatever goes wrong in Alzheimer's, according to Ivan Lieberburg of Athena. Lieberburg has discussed a collaboration on a therapeutic approach with Roses's group. The need for therapies is urgent. The only existing drug for Alzheimer's, Warner-Lambert's Cognex, which the Food and Drug Administration approved in September, benefits just a small proportion of patients, points out Mark J. Alberts, a Duke researcher. Cognex inhibits an enzyme that, in turn, destroys acetylcholine, an important neurotransmitter that is in short supply in the brains of Alzheimer's patients. But Cognex, whose chemical name is tacrine hydrochloride, does not slow the cell death that causes the shortage. Other approaches to therapy now under investigation share this limitation.

Until the hope for a satisfactory therapy is realized, the findings raise an ethical problem that has previously emerged only in the context of rarer conditions, such as Huntington's disease. To wit, should researchers tell those who ask whether they have the predisposing gene? The millions of people in the U.S. with two *APOE*-ε4 genes appear to have a risk for Alzheimer's of more than 80 percent, according to Roses's and others' data. Learning that such a fate was probably in store could cause people great anguish, the Duke researchers acknowledge. Yet already the Duke team has had "hundreds" of requests from individuals who want to be tested, mainly relatives of Alzheimer's patients, Saunders says. The group has obliged some of them.

"We're not encouraging people to get tested," Saunders cautions. The investigators fear that insurance companies may want to use the test to identify individuals who have the *APOE*-ε4 gene. Some insurers have already approached the Duke team, Saunders states.

Nobody is claiming that *APOE*-ε4 is the one true cause of Alzheimer's. Still, the finding could catalyze other breakthroughs. A worker at the National Institute of Neurological Disease and Stroke, who is pursuing an entirely different line of research on the disease, thinks his findings could be linked to the Duke discovery. Daniel L. Alkon, who studies mechanisms of memory, has, together with René Etcheberrigaray and others, proposed in a paper in the *Proceedings of the National Academy of Sciences* that skin-tissue cells from Alzheimer's sufferers have defective molecular channels of a particular type. The channels in question move potassium ions across cell walls, and Alkon says he was able to identify accurately as victims of Alzheimer's 70 individuals by examining their fibroblast potassium channels.

Alkon already has ideas about how aberrations causing the defective channels might be linked to unusual processing of amyloid beta-protein and, thus, indirectly to its binding partner apolipoprotein E. The mystery of Alzheimer's may be slowly coming unraveled. —*Tim Beardsley*

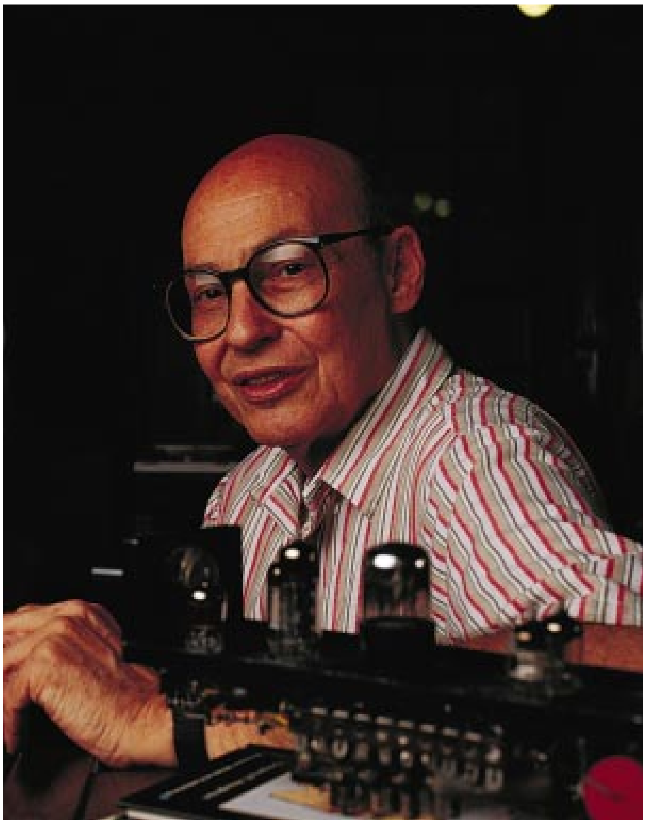
The Mastermind of Artificial Intelligence

Marvin Minsky's ideas about the mind may—or may not—offer lasting insights. But they certainly reveal much about the mind of Minsky. According to Minsky, the mind is not a unified entity but a "society" of elements that both complement and compete with one another. Minsky's emphasis on multiplicity seems to transcend science; he views single-mindedness with a kind of horror. "If there's something you like very much, then you should regard this not as you feeling good but as a kind of brain cancer," he explains, "because it means that some small part of your mind has figured out how to turn off all the other things."

Minsky even recoils at the tendency of ordinary mortals, once they have invested the time in learning to do something, to keep doing it. To counter this trait, which he calls the investment principle, Minsky has trained himself to "enjoy the feeling of awkwardness" aroused by confronting an entirely new problem. "It's so thrilling not to be able to do something," he remarks.

This credo has served Minsky well in his role as a founding father of artificial intelligence. Called AI by insiders, the field is dedicated to the proposition that brains are nothing more than machines, albeit extremely complicated ones, whose abilities will someday be duplicated by computers. In pursuit of the goals of AI, Minsky, who turned 66 in August, has drawn on computer science, robotics, mathematics, neuroscience, psychology, philosophy and even science fiction. His ideas have in turn influenced all these fields as well as AI itself. Colleagues were scheduled to honor him on October 18 at the Massachusetts Institute of Technology, from which he has ruled the AI roost for more than 30 years.

But the same traits that made Minsky a successful pioneer of AI have led him to become increasingly alienated from the field as it matures. Before my meeting with Minsky, in fact, other AI workers warn me that he might be somewhat cranky; if I do not want the interview cut short, I should not ask him too directly about the current slump in AI or what some workers characterize as his own waning influence in the field. One prominent theorist pleads with me not to take advantage of Minsky's penchant for hyperbole. "Ask him if he means it, and if he doesn't say it three times, you shouldn't use it," the theorist urges.



JESSICA BOYATT

MINSKY poses with a component from a neural-network "learning machine" that he and a colleague built in 1951.

Minsky is rather edgy when I meet him in his office at the Artificial Intelligence Laboratory. He fidgets ceaselessly, blinking, waggling his foot, pushing things about his desk. Unlike most scientific celebrities, he gives the impression of conceiving ideas and tropes from scratch rather than retrieving them from memory. He is often but not always incisive. "I'm rambling here," he says glumly after a riff on the nature of verification in AI collapses in a heap of sentence fragments. Even his physical appearance has an improvisational air. His large, round head seems entirely bald but is actually fringed by hairs as transparent as optical fibers. He wears a crocheted belt that supports, in addition to his pants, a belly pack and a holster containing pliers with retractable jaws. With his paunch and vaguely Asian features, he resembles a high-tech Buddha.

Minsky is unable, or unwilling, to inhabit any emotion for long. Early on, as predicted, he plays the curmudgeon. His only rival in grasping the mind's complexity is dead: "Freud has the best theories so far, next to mine, of what it takes to make a mind," Minsky declares. If AI has not progressed as fast as it should have, that is because modern researchers have succumbed to "physics envy"—the desire to reduce the intricacies of the brain to simple formulae—and to the dreaded investment principle. "They are defining smaller and smaller subspecialties that they examine in more detail, but they're not open to doing things in a different way." Even M.I.T.'s own AI lab, which he founded, is guilty. "I don't consider this to be a serious research institution at the moment," he sneers.

But a metamorphosis occurs when, touring the AI lab, we stop to chat with some researchers in a lounge. Minsky engages in amiable shoptalk about chess-playing computers. He then recounts how the science fictionist Isaac Asimov, who just died, always refused Minsky's invitations to see the robots being built at M.I.T. out of fear that his imagination "would be weighed down by this boring realism."

One lounger, noticing that he and Minsky have the same pliers, yanks his instrument from its holster and with a flick of his wrist snaps the retractable jaws into place. "En garde," he says. Minsky, grinning, draws his weapon, and he and his challenger whip their pliers repeatedly at each other, like punks practicing their switchblade technique. Minsky expounds on both the versatility and—an important point for him—the drawbacks of the pliers; his pair pinches him during certain maneuvers. "Can you take it apart with itself?" someone asks. Minsky and his colleagues share a laugh at this reference to a fundamental problem in robotics.

Returning to Minsky's office, we encounter a young, extremely pregnant graduate student. She is scheduled for an oral doctoral exam the next day. "Are you nervous?" Minsky inquires. "A little," she replies. "You shouldn't be," he says and gently touches his forehead to hers. I realize, watching this scene, that there are many Minskys.

Too many, according to Minsky. As a child, the son of a New York City surgeon, he was a prodigy in both mathematics and music. Minsky still occasionally finds himself composing "Bach-like things"—an electric organ crowds his office—but he tries to resist the impulse by convincing himself that music suppresses thought. "I had to kill the musician at some point," he says. "It comes back every now and then, and I hit it."

Minsky started to think about thinking—or, more specifically, about learning—in high school. Although he received undergraduate and graduate degrees in mathematics (from Harvard and Princeton universities, respectively), he scavenged in other disciplines for ideas he felt could illuminate the mind. In 1951 he and a colleague built a machine, made of vacuum tubes, motors and servomechanisms, that could "learn" how to navigate a maze. It was the first neural network ever built. Minsky followed this engineering project with a doctoral thesis on automated learning.

In 1959 he and John McCarthy—who is credited with having coined the term "artificial intelligence"—founded what became the M.I.T. Artificial Intelligence Laboratory. McCarthy left four years later to found his own laboratory at Stanford University, and since then, he and Minsky have had an intellectual parting of the ways. McCarthy has championed AI models based on logic, whereas Minsky contends that logic requires precise definitions that the real world fails to respect. The definition of a bird as a feathered animal that flies, he points out, does not apply if the bird is dead or caged or has had its feet encased in concrete "or has been meditating and decided flying is egotistical."

He has been even harder on neural networks, the technology he helped to nurture. In 1969 he and Seymour Papert of M.I.T. presented a detailed critique of a then popular neural network in a book entitled *Perceptrons*. The book is often said to have dealt neural networks a nearly mortal blow; funding fell off rapidly, and the field languished for more than a decade before it slowly began regenerating. Minsky's intention was not to destroy the field, as some observers have claimed—"That's crazy," he snaps—but to outline the limits of the technology.

Although Minsky applauds the recent resurrection of neural networks, he charges that some "semicommercial" researchers are not as forthcoming as they should be. "They write a paper saying, 'Look, it did this,' and they don't consider it equally wonderful to say, 'Look, it can't do that.' Most of them are not doing good science, because they're hiding the deficiencies." Minsky insists that no single approach can reproduce the intricacies of the mind, because the mind itself employs many fallible methods that back up and check one another. The mind, he muses, is a tractor-trailer, rolling on many wheels, but AI workers "keep designing unicycles."

Some aspects of the mind have proved harder to understand or reproduce than Minsky expected. He confirms the often-told anecdote that in the early 1960s he assigned artificial vision, now recognized as a profoundly difficult problem, to a student as a summer project. But he expects all the major questions in AI to be solved as imaging and electrode techniques reveal the brain in ever finer detail. "Everything we've done up to now I regard as like chemistry before Lavoisier," he remarks.

Minsky poured his thoughts about thinking into *The Society of Mind,* published in 1985. The book consists of 270 essays, most of them only one page long, which range from rather technical discussions of neural wiring to philosophical excursions into the nature of human identity. In the book's prologue, Minsky contended that the work's atomized structure reflects its major theme, that "you can build a mind from many little parts, each mindless by itself." "As far as I know, nobody read the book," Minsky grumbles.

Minsky has nothing but contempt for those who believe that computers, while they may be able to mimic certain aspects of human intelligence, can never become truly conscious. "They're idiots," he fumes. (Minsky is kinder to me when I make the mistake of suggesting that there might always be a qualitative difference between humans and artificial machines; he calls me a "racist.")

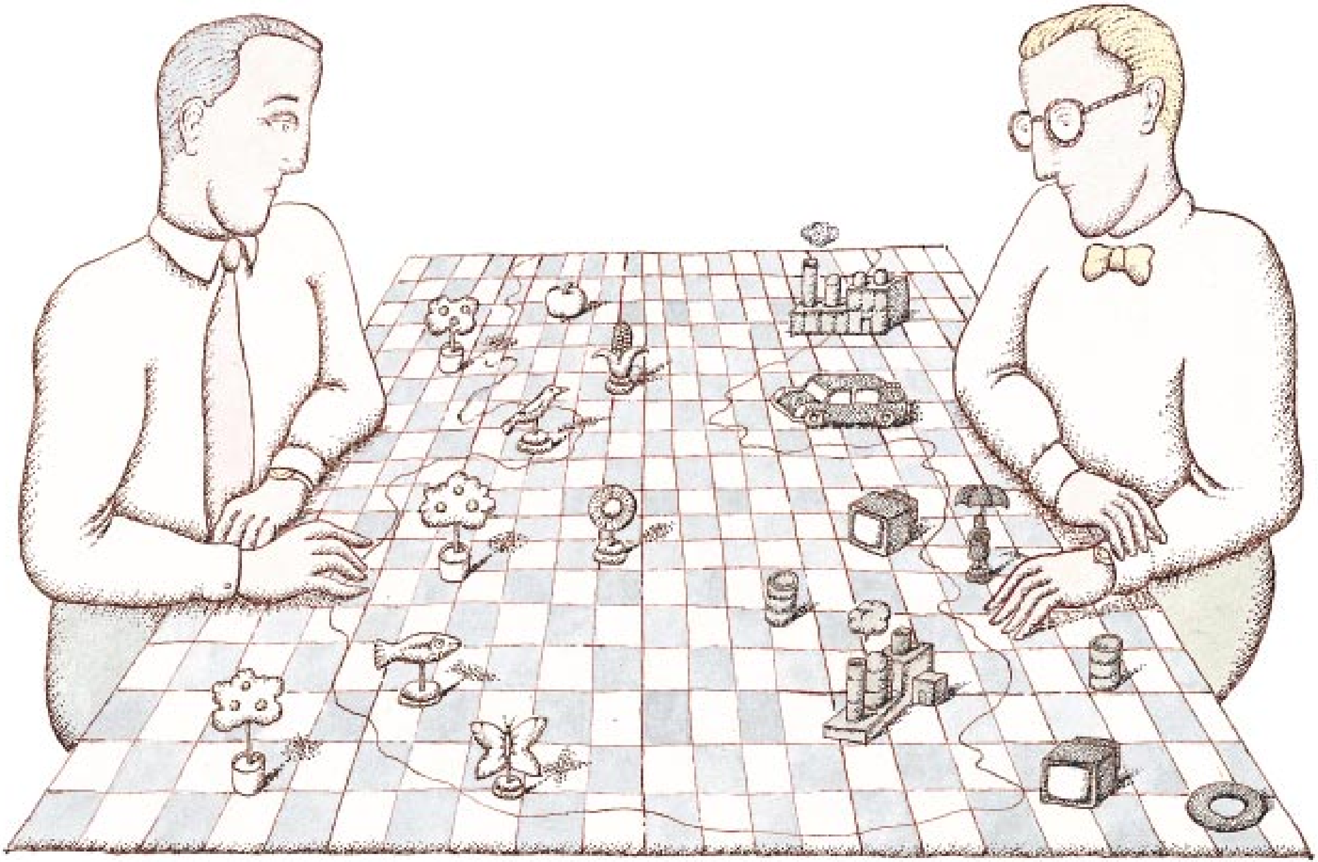
The mystery of consciousness is "trivial," Minsky declares. "I've solved it, and I don't understand why people don't listen." Consciousness, Minsky explains, involves one part of the mind monitoring the behavior of other parts. This function requires little more than short-term memory, or a "low-grade system for keeping records." In fact, computer programs such as LISP, which have memory features that allow their processing steps to be retraced, are "extremely conscious," Minsky asserts—more so than humans, who have pitifully shallow short-term memories.

Like many AI practitioners, Minsky predicts that computers will someday evolve far beyond humans, who are nothing but "dressed-up chimpanzees." Humans may be able to "download" their personalities into computers and thereby become smarter and more reliable. This trick may yield infinite life, among other perks. Minsky envisions making copies of himself that could undergo experiences he would otherwise shun. "I regard religious experience as very risky, because it can destroy your brain. But if I had a backup copy—"

Meanwhile the *Ur*-Minsky remains restless. Hollywood may provide one outlet for his energies. That becomes clear when Laurel, an administrator at the AI lab, sticks her head in the office to ask what's new. Minsky replies that the Disney corporation has hired him to design a "magic carpet ride," based on its hit movie *Aladdin,* for one of its theme parks. Minsky has been working on a virtual-reality scheme at a laboratory Disney has set up for special effects. "I love it," Minsky says of the laboratory. "It's just like the AI lab used to be."

Noting that Stephen W. Hawking, the English cosmologist, recently appeared on "Star Trek," Laurel suggests that Minsky is well suited for playing "an alien genius" on the television show. Evil or benign? I ask. "Oh, either," Laurel replies. Minsky seems intrigued, but he worries that he may be unable to rehearse scenes properly. "I can't say the same thing twice," he confesses.

Minsky is also working on a new book, *The Emotion Machine.* "That's a person," Minsky says of the title. One goal of the book, he notes, is to help people think constructively about thinking. "I'm interested in people who are trying to do some work but keep watching television or going to baseball games." The book will advise such people to make "a little block diagram" of their minds, with different, competing agents labeled. I try to imagine Minsky as a self-help guru, propounding his AI-oriented program on "Donahue." Then, recalling the way he comforted the pregnant graduate student, I think, Why not? —*John Horgan*



Debate: Does Free Trade Harm the Environment?

Reconciling economic growth with environmental protection is one of the greatest challenges now facing policymakers. Unfortunately, these twin goals are widely seen as antithetical: the prescriptions for promoting one often seem to discourage the other. In the following pages, two economists debate whether unrestricted international trade, as embodied in proposals for the General Agreement on Trade and Tariffs (GATT) and the North American Free Trade Agreement (NAFTA), will harm or help the environment. Jagdish Bhagwati of Columbia University argues that freeing trade from inefficient restrictions may be the best way to achieve environmental protection while also safeguarding prosperity and liberty. To the contrary, insists Herman E. Daly of the World Bank: free trade left to itself may harm both the environment and human welfare. The authors offer starkly different predictions about the possible consequences of the new trade agreements.

The Case for Free Trade

Environmentalists are wrong to fear the effects of free trade. Both causes can be advanced by imaginative solutions

by Jagdish Bhagwati

Economists are reconciled to the conflict of absolutes: that is why they invented the concept of tradeoffs. It should not surprise them, therefore, that the objective of environmental protection should at times run afoul of the goal of seeking maximum gains from trade. In fact, economists would be suspicious of any claims, such as those made by soothsaying politicians, that both causes would be only mutually beneficial. They are rightly disconcerted, however, by the passion and the ferocity, and hence often the lack of logic or facts, with which environmental groups have recently assailed both free trade and the General Agreement on Tariffs and Trade (GATT), the institution that oversees the world trading system.

The environmentalists' antipathy to trade is perhaps inevitable. Trade has been central to economic thinking since Adam Smith discovered the virtues of specialization and of the markets that naturally sustain it. Because markets do not normally exist for the pursuit of environmental protection, they must be specially created. Trade therefore suggests abstention from governmental intervention, whereas environmentalism suggests its necessity. Then again, trade is exploited and its virtues extolled by corporate and multinational interests, whereas environmental objectives are embraced typically by nonprofit organizations, which are generally wary of these interests. Trade is an ancient occupation, and its nurture is the objective of institutions crafted over many years of experience and reflection. Protection of the environment, on the other hand, is a recent preoccupation of national and international institutions that are nascent and still evolving.

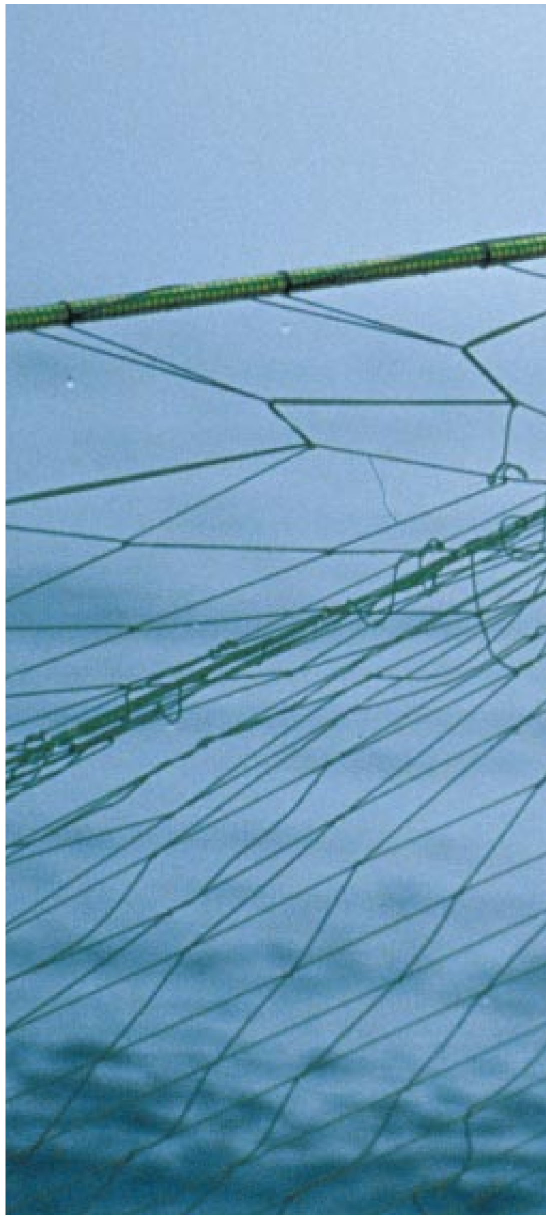
JAGDISH BHAGWATI is Arthur Lehman Professor of Economics and professor of political science at Columbia University and was Ford International Professor of Economics at the Massachusetts Institute of Technology. He has served as the economic policy adviser to the director-general of the General Agreement on Tariffs and Trade. Five volumes of his collected essays have been published by MIT Press. His most recent books are *Protectionism* (MIT Press, 1988) and *The World Trading System at Risk* (Princeton University Press, 1991). He also writes frequently for the *New York Times*, the *Wall Street Journal* and the *New Republic.*

Last year the environmentalists' hostility to trade exploded in outrage when an impartial GATT Dispute Settlement Panel ruled in favor of Mexico and free trade and against the U.S. and the welfare of the dolphin. The U.S. had placed an embargo on the import of Mexican tuna on the grounds that the fish had been caught in purse-seine nets, which kill dolphins cruelly and in greater numbers than U.S. law permits. The GATT panel ruled, in effect, that the U.S. could not suspend Mexico's trading rights by proscribing unilaterally the methods by which that country harvested tuna.

This decision spurred the conservationists' subsequent campaigns against free trade and GATT. GATT has no shortage of detractors, of course. In fact, some of its recent critics have feared its impotence and declared it "dead," referring to it as the General Agreement to Talk and Talk. But the environmentalist attacks, which presume instead GATT's omnipotence, are something else again.

An advertisement by a coalition of environmental groups in the *New York Times* on April 20, 1992, set a new standard for alarmist, even scurrilous, writing, calculated to appeal to one's instincts rather than one's intellect. It talks of "faceless GATT bureaucrats" mounting a "sneak attack on democracy." This veiled reference to Pearl Harbor provides an example of a common tactic in trade controversy: Japan-bashing. The innuendos have continued unabated and are manifest in the endless battles in Congress over the supplemental environmental accords for the North American Free Trade Agreement (NAFTA). The hostility is also intruding on the conclusion of the Uruguay Round of GATT talks, now in their seventh year, with the environmentalists opposing the establishment of the new Multilateral Trade Organization, which is meant to provide effective discipline and a necessary institutional structure for GATT.

It is surely tragic that the proponents of two of the great causes of the 1990s, trade and the environment, should be



DOLPHIN VERSUS FREE TRADE: the U.S. outlaws fishing methods that result in the death of dolphins such as this one,

locked in combat. The conflict is largely gratuitous. There are at times philosophical differences between the two that cannot be reconciled, as when some environmentalists assert nature's autonomy, whereas most economists see nature as a handmaiden to humankind. For the most part, however, the differences derive from misconceptions. It is necessary to dissect and dismiss the more egregious of these fallacies before addressing the genuine problems.

The fear is widespread among environmentalists that free trade increases economic growth and that growth harms the environment. That fear is misplaced. Growth enables governments to tax and to raise resources for a variety of objectives, including the abatement of pollution and the general protection of the environment. Without such revenues, little can be achieved, no matter how pure one's motives may be.

How do societies actually spend these additional revenues? It depends on how getting rich affects the desire for a better environment. Rich countries today have more groups worrying about environmental causes than do poor countries. Efficient policies, such as freer trade, should generally help environmentalism, not harm it.

If one wants to predict what growth will do to the environment, however, one must also consider how it will affect the production of pollution. Growth affects not only the demand for a good environment but also the supply of the pollution associated with growth. The net effect on the environment will therefore depend on the kind of economic growth. Gene M. Grossman and Alan B. Krueger of Princeton University found that in cities around the world sulfur dioxide pollution fell as per capita income rose. The only exception was in countries whose per capita incomes fell below $5,000. In short, environmentalists are in error when they fear that trade, through growth, will necessarily increase pollution.

Economic effects besides those attributable to rising incomes also help to protect the environment. For example, freer trade enables pollution-fighting technologies available elsewhere to be imported. Thus, trade in low-sulfur-content coal will enable the users of local high-sulfur-content coal to shift from the latter to the former.

Free trade can also lead to better environmental outcomes from a shift in the composition of production. An excellent example is provided by Robert C. Feenstra of the University of California at Davis. He has shown how the imposition of restraints on Japanese automobile exports to the U.S. during the 1980s shifted the composition of those exports from small to large cars, as the Japanese attempted to increase their revenues without increasing the number of units they sold. Yet the large cars were fuel inefficient. Thus, protective efforts by the U.S. effectively increased the average amount of pollution produced by imported cars, making it more likely that pollution from cars would increase rather than diminish in the U.S.

Although these erroneous objections to free trade are readily dismissed (but not so easily eliminated from public discourse), there are genuine conflicts between trade and the environment. To understand and solve them, economists draw a distinction between two kinds of environmental problems: those that are intrinsically domestic and those that are intrinsically transnational.

Should Brazil pollute a lake lying wholly within its borders, the problem would be intrinsically domestic. Should it pollute a river that flows into Argentina, the matter would take on an intrinsically transnational character. Perhaps the most important examples of transnational pollution are acid rain, created when sulfur dioxide emissions in one country precipitate into rain in another, and greenhouse gases, such as carbon dioxide, which contribute to global warming wherever they are emitted.

ensnared off the U.S. Atlantic coast. But when the U.S. attempted to apply its standard to Mexico by imposing an embargo on tuna imported from that country, an international tribunal rejected the policy last year as an illegal restriction of trade.

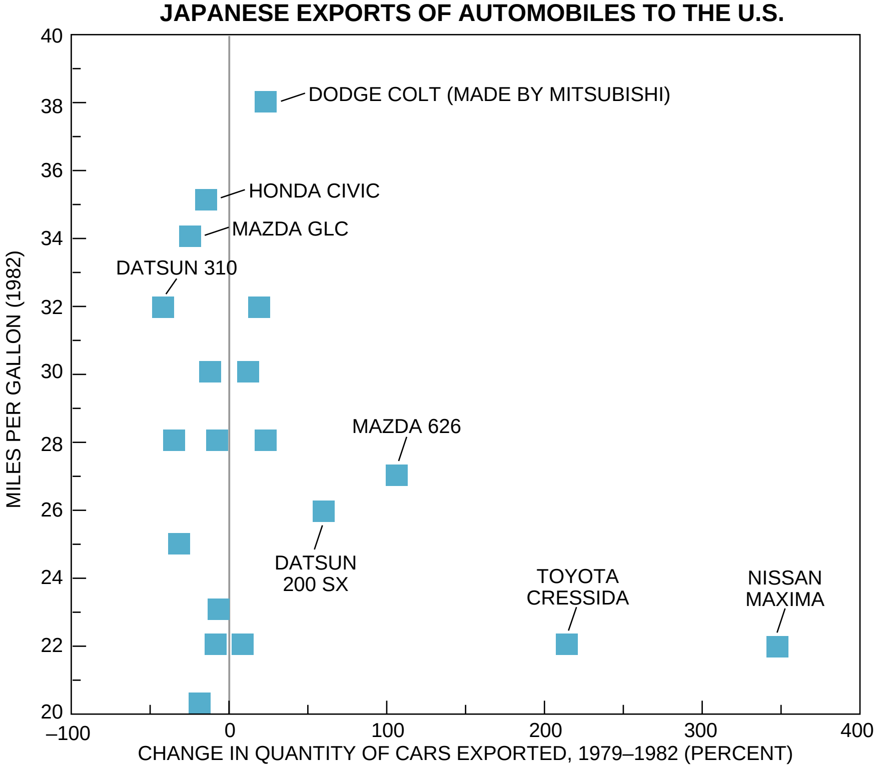
Why do intrinsically domestic environmental questions create international concern? The main reason is the belief that diversity in environmental standards may affect competitiveness. Businesses and labor unions worry that their rivals in other countries may gain an edge if their governments impose lower standards of environmental protection. They decry such differences as unfair. To level the playing field, these lobbies insist that foreign countries raise their standards up to domestic ones. In turn, environmental groups worry that if such "harmonization up" is not undertaken prior to freeing trade, pressures from uncompetitive businesses at home will force down domestic standards, reversing their hard-won victories. Finally, there is the fear, dramatized by H. Ross Perot in his criticisms of NAFTA, that factories will relocate to the countries whose environmental standards are lowest.

But if the competitiveness issue makes the environmentalists, the businesses and the unions into allies, the environmentalists are on their own in other ways. Two problem areas can be distinguished. First, some environmentalists are keen to impose their own ethical preferences on others, using trade sanctions to induce or coerce acceptance of such preferences. For instance, tuna fishing with purse-seine nets that kill dolphins is opposed by U.S. environmental groups, which consequently favor restraints on the importation of such tuna from Mexico and elsewhere. Second, other environmentalists fear that the rules of free trade, as embodied in GATT and strengthened in the Uruguay Round, will constrain their freedom to pursue even purely domestic environmental objectives, with GATT tribunals outlawing disputed regulation.

Environmentalists have cause for concern. Not all concerns are legitimate, however, and not all the solutions to legitimate concerns are sensible. Worry over competitiveness has thus led to the illegitimate demand that environmental standards abroad be treated as "social dumping." Offending countries are regarded as unfairly subsidizing their exporters through lax environmental requirements. Such implicit subsidies, the reasoning continues, ought to be offset by import duties.

Yet international differences in environmental standards are perfectly natural. Even if two countries share the same environmental objectives, the *specific* pollutions they would attack, and hence the industries they would hinder, will generally not be identical. Mexico has a greater social incentive than does the U.S. to spend an extra dollar preventing dysentery rather than reducing lead in gasoline.

JAPANESE EXPORTS OF AUTOMOBILES TO THE U.S.



SOURCE: Robert C. Feenstra, University of California, Davis

PERVERSE CONSEQUENCES for the environment may result from trade restrictions. This graph shows Japanese car exports to the U.S. before and after Japan's acquiescence in voluntary export restraints. Sales of small, fuel-efficient models declined, whereas those of the larger "gas guzzlers" soared.

Equally, a certain environmental good might be valued more highly by a poor country than by a rich one. Contrast, for instance, the value assigned to a lake with the cost of cleaning up effluents discharged into it by a pharmaceutical company. In India such a lake's water might be drunk by a malnourished population whose mortality would increase sharply with the rise in pollution. In the U.S. the water might be consumed by few people, all of whom have the means to protect themselves with privately purchased water filters. In this example, India would be the more likely to prefer clean water to the pharmaceutical company's profits.

The consequences of differing standards are clear: each country will have less of the industry whose pollution it fears relatively more than other countries do. Indeed, even if there were no international trade, we would be shrinking industries whose pollution we deter. This result follows from the policy of forcing polluters of all stripes to pay for the harm they cause. To object, then, to the effects our negative valuation of pollution have on a given industry is to be in contradiction: we would be refusing to face the consequences of our environmental preferences.

Nevertheless, there is sentiment for enacting legislation against social dumping. Senator Davil L. Boren of Oklahoma, the proponent of the International Pollution Deterrence Act of 1991, demanded import duties on the grounds that "some U.S. manufacturers, such as the U.S. carbon and steel alloy industry, spend as much as 250 percent more on environmental controls as a percentage of gross domestic product than do other countries.... I see the unfair advantage enjoyed by other nations exploiting the environment and public health for economic gain when I look at many industries important to my own state." Similarly, Vice President Al Gore wrote in *Earth in the Balance: Ecology and the Human Spirit* that "just as government subsidies of a particular industry are sometimes considered unfair under the trade laws, weak and ineffectual enforcement of pollution control measures should also be included in the definition of unfair trading practices."

These demands betray lack of economic logic, and they ignore political reality as well. Remember that the so-called subsidy to foreign producers through lower standards is not given but only implied. According to Senator

EMPLOYMENT IN MEXICAN TUNA FISHERY may offset the saving of dolphins that would result were the industry to forgo purse-seine nets. Countries should not be faulted for placing human welfare ahead of our culture-specific concerns.

Boren, the subsidy would be calculated as "the cost that would have to be incurred by the manufacturer or producer of the foreign articles of merchandise to comply with environmental standards imposed on U.S. producers of the same class of merchandise." Anyone familiar with the way dumping calculations are made knows that the Environmental Protection Agency could come up with virtually any estimates it cared to produce. Cynical politics would inevitably dictate the calculations.

Still, there may be political good sense in assuaging environmentalists' concerns about the relocation of factories to countries with lower standards. The governments of higher-standards countries could do so without encumbering free trade by insisting that their businesses accede to the higher standards when they go abroad. Such a policy lies entirely within the jurisdictional powers of a higher-standards country. Moreover, the governments of lower-standards countries would be most unlikely to object to such an act of good citizenship by the foreign investors.

Environmentalists oppose free trade for yet another reason: they wish to use trade policy to impose their values on other communities and countries. Many environmentalists want to suspend the trading rights of countries that sanction the use of purse-seine nets in tuna fishing and of leg-hold traps in trapping. Such punishments seem an inappropriate use of state power, however. The values in question are not widely accepted, such as human rights, but idiosyncratic. One wonders when the opponents of purse-seine nets put the interests of the dolphin ahead of those of Mexico's people, who could prosper through more productive fishing. To borrow the campaign manifesto of President Bill Clinton: Should we not put people first?

Moreover, once such values intrude on free trade, the way is opened for an endless succession of demands. Environmentalists favor dolphins; Indians have their sacred cows. Animal-rights activists, who do not prefer one species over another, will object to our slaughterhouses.

The moral militancy of environmentalists in the industrialized world has begun to disillusion their closest counterparts in the undeveloped countries. These local environmentalists accuse the rich countries of "eco-imperialism," and they deny that the Western nations have a monopoly on virtue. The most radical of today's proenvironment magazines in India, *Down to Earth,* editorialized recently: "In the current world reality trade is used as an instrument entirely by Northern countries to discipline environmentally errant nations. Surely, if India or Kenya were to threaten to stop trade with the U.S., it would hardly affect the latter. But the fact of the matter is that it is the Northern countries that have the greatest [adverse] impact on the world's environment."

If many countries were to play this game, then repeated suspensions of trading rights would begin to undermine the openness of the trading system and the predictability and stability of international markets. Some environmentalists assert that each country should be free to insist on the production methods of its trading partners. Yet these environmentalists ignore the certain consequence of their policy: a Pandora's box of protectionism would open up. Rarely are production methods in an industry identical in different countries.

There are certainly better ways to indulge the environmentalists' propensity to export their ethical preferences. The U.S. environmental organizations can lobby in Mexico to persuade its government to adopt their views. Private boycotts can also be undertaken. In fact, boycotts can carry much clout in rich countries with big markets, on which the targeted poor countries often depend. The frequent and enormously expensive advertisements by environmental groups against GATT show also that their resources far exceed those of the cash-strapped countries whose policies they oppose.

Cost-benefit analysis leads one to conclude that unilateral governmental suspension of others' trading rights is not an appropriate way to promote one's lesser ethical preferences. Such sanctions can, on the other hand, appropriately be invoked multilaterally to defend universal moral values. In such cases—as in the censure of apartheid, as practiced until recently in South Africa—it is possible to secure widespread agreement for sanctions. With a large majority converted to the cause, GATT's waiver procedure can be used to suspend the offending country's trading rights.

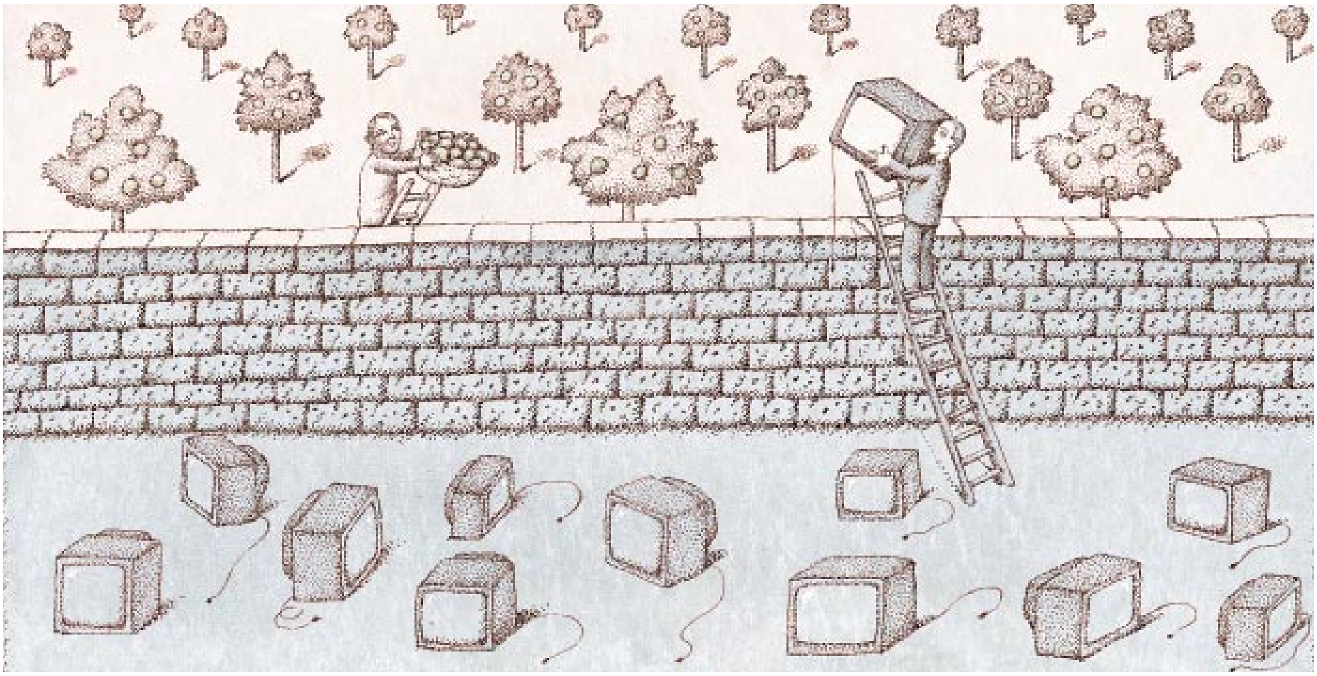
Environmentalists are also worried about the obstacles that the current and prospective GATT rules pose for environmental regulations aimed entirely at domestic production and consumption. In principle, GATT lets a country enforce any regulation that does not discriminate against or among foreign suppliers. One can, for example, require airbags in cars, provided that the rule applies to all automobile makers. GATT even permits rules that discriminate against trade for the purpose of safety and health.

GATT, however, recognizes three ways in which regulations may be set in gratuitous restraint of trade; in following procedures aimed at avoiding such outcomes, GATT upsets the environmentalists. First, the true intention—and effect—of a regulation may be to protect not the environment but local business. Second, a country may impose more restrictions than necessary to achieve its stated environmental objective. Third, it may set standards that have no scientific basis.

The issue of intentions is illustrated by the recently settled "beer war" between Ontario and the U.S. Five years ago the Canadian province imposed a 10-cents-a-can tax on beer, ostensibly to discourage littering. The U.S. argued that the law in fact intended to discriminate against its beer suppliers, who used aluminum cans, whereas local beer companies used bottles. Ontario had omitted to tax the use of cans for juices and soups, a step that would have affected Ontario producers.

The second problem is generally

PURE DRINKING WATER is essential for Mexican villagers, who wait in line to collect it rather than risk contracting cholera from local sources. The relative value of environmental benefits varies in different countries: Mexico can better improve public health by concentrating its resources on the purification of water than by reducing the lead in gasoline.



BENEFITS OF TRADE flow from the economies achieved when countries specialize in enterprises in which they enjoy comparative advantage. Such specialization will proceed better when all sides trust in the stability of the trading regime.

tougher because it is impossible to find alternative restrictions that accomplish exactly the same environmental results as the original policy at lower cost. An adjudicating panel is then forced to evaluate, implicitly or explicitly, the trade-offs between the cost in trade disruption and the cost in lesser fulfillment of the environmental objective. It is therefore likely that environmentalists and trade experts will differ on which weights the panel should assign to these divergent interests.

Environmentalists tend to be fearful about the use of scientific tests to determine whether trade in a product can be proscribed. The need to prove one's case is always an unwelcome burden to those who have the political power to take unilateral action. Yet the trade experts have the better of the argument. Imagine that U.S. growers sprayed apples with the pesticide Alar, whereas European growers did not, and that European consumers began to agitate against Alar as harmful. Should the European Community be allowed to end the importation of the U.S. apples without meeting *some* scientific test of its health concerns? Admittedly, even hard science is often not hard enough—different studies may reach different conclusions. But without the restraining hand of science, the itch to indulge one's fears—and to play on the fears of others—would be irresistible.

In all cases, the moderate environmentalists would like to see GATT adopt more transparent procedures for adjudicating disputes. They also desire greater legal standing to file briefs when environmental regulations are at issue. These goals seem both reasonable and feasible.

Not all environmental problems are local; some are truly global, such as the greenhouse effect and the depletion of the stratospheric ozone. They raise more issues that require cooperative, multilateral solutions. Such solutions must be both efficient and equitable. Still, it is easy to see that rich countries might use their economic power to reach protocols that maximize efficiency at the expense of poorer countries.

For instance, imagine that the drafters of a protocol were to ask Brazil to refrain from cutting down its rain forests while allowing industrialized countries to continue emitting carbon dioxide. They might justify this request on the grounds that it costs Brazil less to keep a tree alive, absorbing a unit of carbon dioxide every year, than it would cost the U.S. or Germany to save a unit by burning less oil. Such a trade-off would indeed be economically efficient. Yet if Brazil, a poorer country, were then left with the bill, the solution would assuredly be inequitable.

Before any group of countries imposes trade sanctions on a country that has not joined a multilateral protocol, it would be important to judge whether the protocol is indeed fair. Nonmembers targeted for trade sanctions should have the right to get an impartial hearing of their objections, requiring the strong to defend their actions even when they appear to be entirely virtuous.

The simultaneous pursuit of the two causes of free trade and a protected environment often raises problems, to be sure. But none of these conflicts is beyond resolution with goodwill and by imaginative institutional innovation. The aversion to free trade and GATT that many environmentalists display is unfounded, and it is time for them to shed it. Their admirable moral passion and certain intellectual vigor are better devoted to building bridges between the causes of trade and the environment.

FURTHER READING

AMERICAN RULES, MEXICAN JOBS. Jagdish Bhagwati in *New York Times,* Section A, page 21, col. 1; March 24, 1993.

"CIRCUMVENTING" DEMOCRACY: THE POLITICAL MORALITY OF TRADE NEGOTIATIONS. Robert E. Hudec in *New York University Journal of International Law and Politics,* Vol. 25, No. 2, pages 401–412; September/October 1993.

ENVIRONMENTAL IMPACTS OF A NORTH AMERICAN FREE TRADE AGREEMENT. Gene M. Grossman and Alan B. Krueger in *The Mexico-U.S. Free Trade Agreement.* Edited by Peter M. Garber. MIT Press, 1993.

TRADE AND ENVIRONMENT: DOES ENVIRONMENTAL DIVERSITY DETRACT FROM THE CASE FOR FREE TRADE? Jagdish Bhagwati and T. N. Srinivasan. Mimeograph. Yale University, 1993.

The Perils of Free Trade

Economists routinely ignore its hidden costs to the environment and the community

by Herman E. Daly

No policy prescription commands greater consensus among economists than that of free trade based on international specialization according to comparative advantage. Free trade has long been presumed good unless proved otherwise. That presumption is the cornerstone of the existing General Agreement on Tariffs and Trade (GATT) and the proposed North American Free Trade Agreement (NAFTA). The proposals in the Uruguay Round of negotiations strengthen GATT's basic commitment to free trade and economic globalization.

Yet that presumption should be reversed. The default position should favor domestic production for domestic markets. When convenient, balanced international trade should be used, but it should not be allowed to govern a country's affairs at the risk of environmental and social disaster. The domestic economy should be the dog and international trade its tail. GATT seeks to tie all the dogs' tails together so tightly that the international knot would wag the separate national dogs.

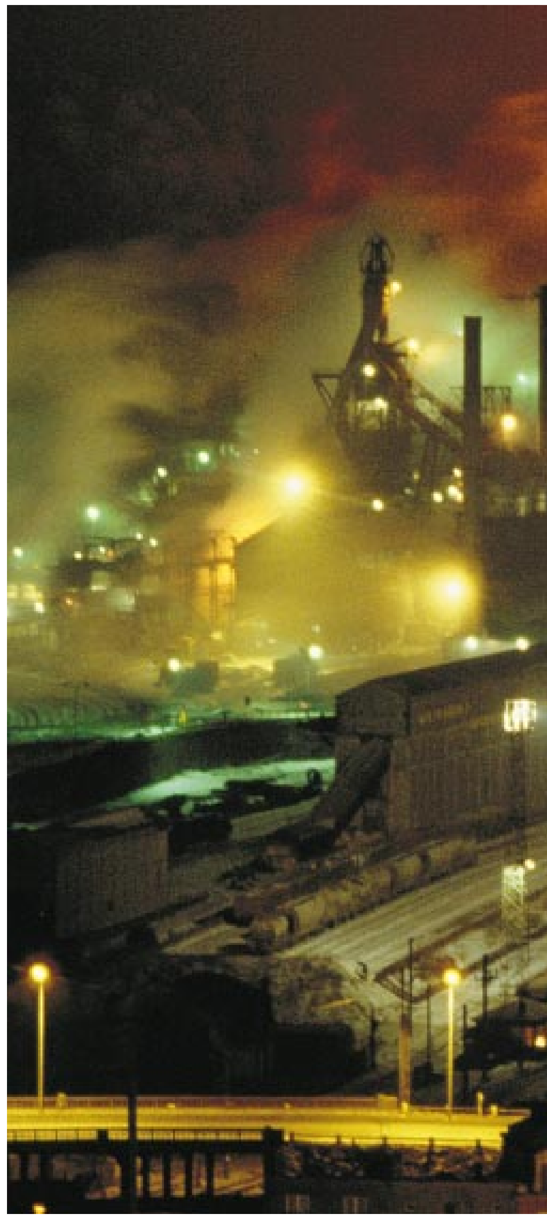
The wiser course was well expressed in the overlooked words of John Maynard Keynes: "I sympathize, therefore, with those who would minimize, rather than those who would maximize, economic entanglement between nations. Ideas, knowledge, art, hospitality, travel—these are the things which should of their nature be international. But let goods be homespun whenever it is reasonably and conveniently possible; and, above all, let finance be primarily national." Contrary to Keynes, the defenders of the proposed Uruguay Round of changes to GATT not only want to downplay "homespun goods," they also want finance and all other services to become primarily international.

Economists and environmentalists are sometimes represented as being, respectively, for and against free trade, but that polarization does the argument a disservice. Rather the real debate is over what kinds of regulations are to be instituted and what goals are legitimate. The free traders seek to maximize profits and production without regard for considerations that represent hidden social and environmental costs. They argue that when growth has made people wealthy enough, they will have the funds to clean up the damage done by growth. Conversely, environmentalists and some economists, myself among them, suspect that growth is increasing environmental costs faster than benefits from production—thereby making us poorer, not richer.

A more accurate name than the persuasive label "free trade"—because who can be opposed to freedom?—is "deregulated international commerce." Deregulation is not always a good policy: recall the recent experience of the U.S. with the deregulation of the savings and loan institutions. As one who formerly taught the doctrine of free trade to college students, I have some sympathy for the free traders' view. Nevertheless, my major concern about my profession today is that our disciplinary preference for logically beautiful results over factually grounded policies has reached such fanatical proportions that we economists have become dangerous to the earth and its inhabitants.

The free trade position is grounded in the logic of comparative advantage, first explicitly formulated by the early 19th-century British economist David Ricardo. He observed that countries with different technologies, customs and resources will incur different costs when they make the same products. One country may find it comparatively less costly to mine coal than to grow wheat, but in another country the opposite may be true. If nations specialize in the

HERMAN E. DALY is senior economist in the environment department of the World Bank in Washington, D.C. Before joining the bank in 1988, he was alumni professor of economics at Louisiana State University. He holds a B.A. from Rice University and a Ph.D. from Vanderbilt University. Daly has taught in Brazil as a Ford Foundation Visiting Professor and as a Senior Fulbright Scholar. He has also served as a research associate at Yale University and as a visiting fellow at the Australian National University. Co-founder and associate editor of *Ecological Economics,* Daly has written several books, including *Steady-State Economics.* The views expressed here by Daly should not be attributed to the World Bank.



POLLUTING is one way in which industries can "externalize" some of the costs associated with production. Industries have profit incentives to produce goods

products for which they have a comparative advantage and trade freely to obtain others, everyone benefits.

The problem is not the logic of this argument. It is the relevance of Ricardo's critical but often forgotten assumption that factors of production (especially capital) are internationally immobile. In today's world, where billions of dollars can be transferred between nations at the speed of light, that essential condition is not met. Moreover, free traders encourage such foreign investment as a development strategy. In short, the free traders are using an argument that hinges on the impermability of national boundaries to capital to support a policy aimed at making those same boundaries increasingly permeable to both capital and goods!

That fact alone invalidates the assumption that international trade will inevitably benefit all its partners. Furthermore, for trade to be mutually beneficial, the gains must not be offset by higher liabilities. After specialization, nations are no longer free *not* to trade, and that loss of independence can be a liability. Also, the cost of transporting goods internationally must not cancel out the profits. Transport costs are energy intensive. Today, however, the cost of energy is frequently subsidized by governments through investment tax credits, federally subsidized research and military expenditures that ensure access to petroleum. The environmental costs of fossil-fuel burning also do not factor into the price of gasoline. To the extent that energy is subsidized, then, so too is trade. The full cost of energy, stripped of these obscuring subsidies, would therefore reduce the initial gains from long-distance trade, whether international or interregional.

in countries with permissive pollution, health and labor standards and then to sell the goods elsewhere. Yet that competitive pressure can drive down higher standards. Tariffs that eliminate these unfair advantages are therefore essential for protecting the global efficiency of resource use.

Free trade can also introduce new inefficiencies. Contrary to the implications of comparative advantage, more than half of all international trade involves the simultaneous import and export of essentially the same goods. For example, Americans import Danish sugar cookies, and Danes import American sugar cookies. Exchanging recipes would surely be more efficient. It would also be more in accord with Keynes's dictum that knowledge should be international and goods homespun (or in this case, homebaked).

Another important but seldom mentioned corollary of specialization is a reduction in the range of occupational choices. Uruguay has a clear comparative advantage in raising cattle and sheep. If it adhered strictly to the rule of specialization and trade, it would afford its citizens only the choice of being either cowboys or shepherds. Yet Uruguayans feel a need for their own legal, financial, medical, insurance and educational services, in addition to basic agriculture and industry. That diversity entails some loss of efficiency, but it is necessary for community and nationhood.

Uruguay is enriched by having a symphony orchestra of its own, even though it would be cost-effective to import better symphony concerts in exchange for wool, mutton, beef and leather. Individuals, too, must count the broader range of choices as a welfare gain: even those who are cowboys and shepherds are surely enriched by contact with countrymen who are not *vaqueros* or *pastores.* My point is that the community dimension of welfare is completely overlooked in the simplistic argument that if specialization and trade increase the per capita availability of commodities, they must be good.

Let us assume that even after those liabilities are subtracted from the gross returns on trade, positive net gains still exist. They must still offset deeper, more fundamental problems. The arguments for free trade run afoul of the three basic goals of all economic policies: the efficient *allocation* of resources, the fair *distribution* of resources and the maintenance of a sustainable *scale* of resource use. The first two are traditional goals of neoclassical economics. The third has only recently been recognized and is associated with the viewpoint of ecological, or steady-state, economics. It means that the input of raw materials and energy to an economy and the output of waste materials and heat must be within the regenerative and absorptive capacities of the ecosystem.

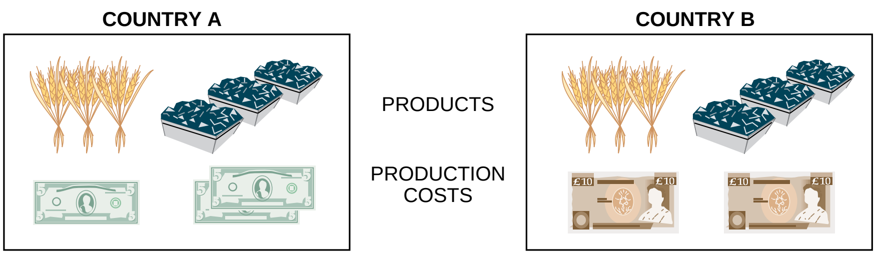
In neoclassical economics the efficient allocation of resources depends on the counting and internalization of all costs. Costs are internalized if they are directly paid by those entities responsible for them—as when, for example, a manufacturer pays for the disposal of its factory wastes and raises its prices to cover that expense. Costs are externalized if they are paid by someone else—as when the public suffers extra disease, stench and nuisance from uncollected wastes. Counting all costs is the very basis of efficiency.

Economists rightly urge nations to follow a domestic program of internalizing costs into prices. They also wrongly urge nations to trade freely with other countries that do not internalize their costs (and consequently have lower prices). If a nation tries to follow both those policies, the conflict is clear: free competition between different cost-internalizing regimes is utterly unfair.

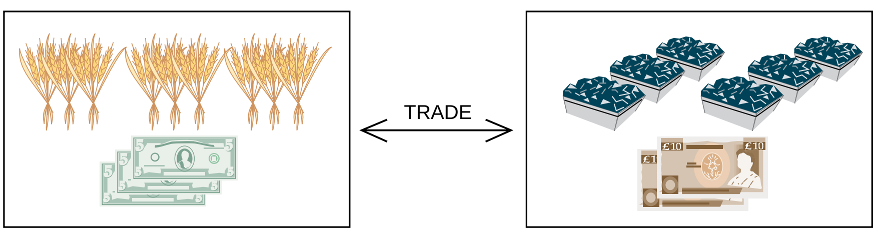
International trade increases competition, and competition reduces costs. But competition can reduce costs in two ways: by increasing efficiency or by lowering standards. A firm can save money by lowering its standards for pollution control, worker safety, wages, health care and so on—all choices that externalize some of its costs. Profit-maximizing firms in competition always have an incentive to externalize their costs to the degree that they can get away with it.

For precisely that reason, nations maintain large legal, administrative and auditing structures that bar reductions in the social and environmental standards of domestic industries. There are no analogous international bodies of law and administration; there are only national laws, which differ widely. Consequently, free international trade encourages industries to shift their production activities to the countries that have the lowest standards of cost internalization—hardly a move toward global efficiency.

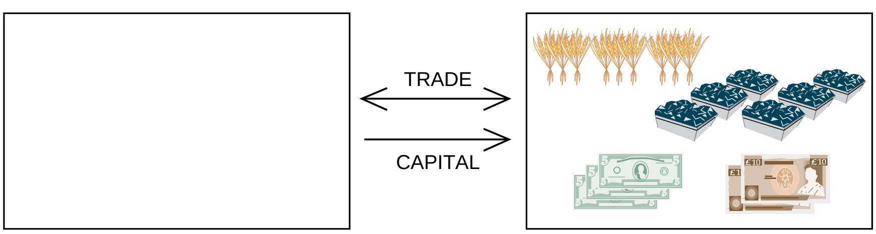
How Comparative Advantage Works



When there is no international trade, each country's production is limited entirely by its own capital and resources. Some products are comparatively less expensive to produce than others on a per unit basis.



When there is free trade, countries can specialize based on comparative advantage. All of a country's capital can be invested in making one product. Absolute cost differences between the countries do not matter. The hidden assumption is that capital cannot cross borders.



If capital is also mobile, capital can follow absolute advantage rather than comparative advantage. As in this example, one country may end up producing everything if it has lower absolute costs.

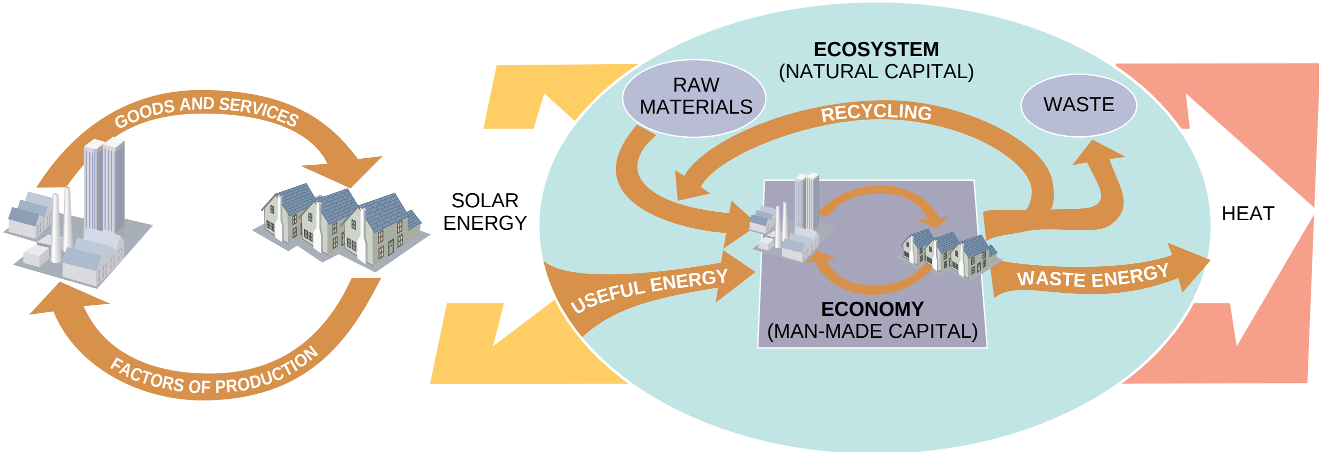
Attaining cheapness by ignoring real costs is a sin against efficiency. Even GATT recognizes that requiring citizens of one country to compete against foreign prison labor would be carrying standards-lowering competition too far. GATT therefore allows the imposition of restrictions on such trade. Yet it makes no similar exception for child labor, for uninsured risky labor or for subsistence-wage labor.

The most practical solution is to permit nations that internalize costs to levy compensating tariffs on trade with nations that do not. "Protectionism"—shielding an inefficient industry against more efficient foreign competitors—is a dirty word among economists. That is very different, however, from protecting an efficient national policy of full-cost pricing from standards-lowering international competition.

Such tariffs are also not without precedent. Free traders generally praise the fairness of "antidumping" tariffs that discourage countries from trading in goods at prices below their production costs. The only real difference is the decision to include the costs of environmental damage and community welfare in that reckoning.

This tariff policy does not imply the imposition of one country's environmental preferences or moral judgments on another country. Each country should set the rules of cost internalization in its own market. Whoever sells in a nation's market should play by that nation's rules or pay a tariff sufficient to remove the competitive advantage of lower standards. For instance, under the Marine Mammal Protection Act, all tuna sold in the U.S. (whether by U.S. or Mexican fishermen) must count the cost of limiting the kill of dolphin associated with catching tuna. Tuna sold in the Mexican market (whether by U.S. or Mexican fishermen) need not include that cost. No standards are being imposed through "environmental imperialism"; paying the costs of a nation's environmental standards is merely the price of admission to its market.

Indeed, free trade could be accused of reverse environmental imperialism. When firms produce under the most permissive standards and sell their products elsewhere without penalty, they press on countries with higher stan-



DIFFERENT VIEWS OF ECONOMIES distinguish neoclassical and steady-state economics. Neoclassical economics pictures the economy as an isolated system (*left*) in which exchange value circulates between industries and households. Neither matter nor energy enters or leaves the system, so the economy can be of any size. In the steady-state view (*right*) the economy is only one component of a larger ecosystem in which materials are transformed and energy is converted to heat. As the economy grows larger, its behavior must conform more closely to that of the total ecosystem.

dards to lower them. In effect, unrestricted trade imposes lower standards.

Unrestricted international trade also raises problems of resource distribution. In the world of comparative advantage described by Ricardo, a nation's capital stays at home, and only goods are traded. If firms are free to relocate their capital internationally to wherever their production costs would be lowest, then the favored countries have not merely a comparative advantage but an absolute advantage. Capital will drain out of one country and into another, perhaps making what H. Ross Perot called "a giant sucking sound" as jobs and wealth move with it. This specialization will increase world production, but without any assurance that all the participating countries will benefit.

When capital flows abroad, the opportunity for new domestic employment diminishes, which drives down the price for domestic labor. Even if free trade and capital mobility raise wages in low-wage countries (and that tendency is thwarted by overpopulation and rapid population growth), they do so at the expense of labor in the high-wage countries. They thereby increase income inequality there. Most citizens are wage earners. In the U.S., 80 percent of the labor force is classified as "nonsupervisory employees." Their real wages have fallen 17 percent between 1973 and 1990, in significant part because of trade liberalization.

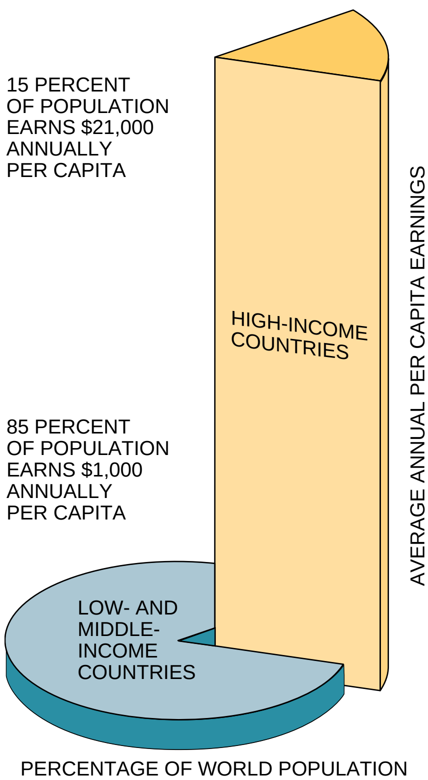
Nor does labor in low-wage countries necessarily gain from free trade. It is likely that NAFTA will ruin Mexican peasants when "inexpensive" U.S. corn (subsidized by depleting topsoil, aquifers, oil wells and the federal treasury) can be freely imported. Displaced peasants will bid down wages. Their land will be bought cheaply by agribusinesses to produce fancy vegetables and cut flowers for the U.S. market. Ironically, Mexico helps to keep U.S. corn "inexpensive" by exporting its own vanishing reserves of oil and genetic crop variants, which the U.S. needs to sustain its corn monoculture.

Neoclassical economists admit that overpopulation can spill over from one country to another in the form of cheap labor. They acknowledge that fact as an argument against free immigration. Yet capital can migrate toward abundant labor even more easily than labor can move toward capital. The legitimate case for restrictions on labor immigration is therefore easily extended to restrictions on capital emigration.

When confronted with such problems, neoclassical economists often answer that growth will solve them. The allocation problem of standards-lowering competition, they say, will be dealt with by universally "harmonizing" all standards upward. The distribution problem of falling wages in high-wage countries would only be temporary; the economists believe that growth will eventually raise wages worldwide to the former high-wage level and beyond.

Yet the goal of a sustainable scale of total resource use forces us to ask: What will happen if the entire population of the earth consumes resources at the rate of high-wage countries? Neoclassical economists generally ignore this question or give the facile response that there are no limits.

The steady-state economic paradigm suggests a different answer. The regenerative and assimilative capacities of the biosphere cannot support even the current levels of resource consumption, much less the manyfold increase required to generalize the higher stan-



RAISING THE INCOMES in the more populous, less wealthy nations will be difficult. Over the next 40 years, the population will double. To reach the higher level of per capita income, the low- and middle-income countries would have to increase their use of resources by a factor of almost 36 ($21 \times 2 \times 0.85$). To avoid augmenting the damage to the environment, they would need to boost resource-use efficiency by the same factor.

dards worldwide. Still less can the ecosystem afford an ever growing population that is striving to consume more per capita. As a species, we already preempt about 40 percent of the land-based primary product of photosynthesis for human purposes. What happens to biodiversity if we double the human population, as we are projected to do over the next 30 to 50 years?

These limits put a brake on the ability of growth to wash away the problems of misallocation and maldistribution. In fact, free trade becomes a recipe for hastening the speed with which competition lowers standards for efficiency, distributive equity and ecological sustainability.

Notwithstanding those enormous problems, the appeal of bigger free trade blocs for corporations is obvious. The broader the free trade area, the less answerable a large and footloose corporation will be to any local or even national community. Spatial separation of the places that suffer the costs and enjoy the benefits becomes more feasible. The corporation will be able to buy labor in the low-wage markets and sell its products in the remaining high-wage, high-income markets. The larger the market, the longer a corporation will be able to avoid the logic of Henry Ford, who realized that he had to pay his workers enough for them to buy his cars. That is why transnational corporations like free trade and why workers and environmentalists do not.

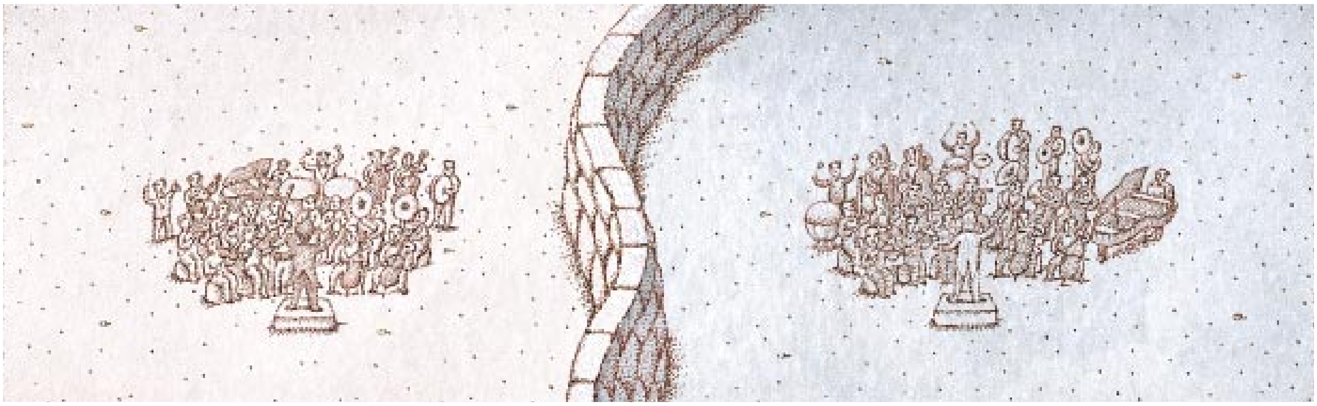
In the view of steady-state economics, the economy is one open subsystem in a finite, nongrowing and materially closed ecosystem. An open system takes matter and energy from the environment as raw materials and returns them as waste. A closed system is one in which matter constantly circulates internally while only energy flows through. Whatever enters a system as input and exits as output is called throughput. Just as an organism survives by consuming nutrients and excreting wastes, so too an economy must to some degree both deplete and pollute the environment. A steady-state economy is one whose throughput remains constant at a level that neither depletes the environment beyond its regenerative capacity nor pollutes it beyond its absorptive capacity.

Most neoclassical economic analyses today rest on the assumption that the economy is the total system and nature is the subsystem. The economy is an isolated system involving only a circular flow of exchange value between firms and households. Neither matter nor energy enters or exits this system. The economy's growth is therefore unconstrained. Nature may be finite, but it is seen as just one sector of the economy, for which other sectors can substitute without limiting overall growth.

Although this vision of circular flow is useful for analyzing exchanges between producers and consumers, it is actively misleading for studying scale—the size of the economy relative to the environment. It is as if a biologist's vision of an animal contained a circulatory system but not a digestive tract or lungs. Such a beast would be independent of its environment, and its size would not matter. If it could move, it would be a perpetual motion machine.

Long ago the world was relatively empty of human beings and their belongings (man-made capital) and relatively full of other species and their habitats (natural capital). Years of economic growth have changed that basic pattern. As a result, the limiting factor on future economic growth has changed. If man-made and natural capital were good substitutes for one another, then natural capital could be totally replaced. The two are complementary, however, which means that the short supply of one imposes limits. What good are fishing boats without populations of fish? Or sawmills without forests? Once the number of fish that could be sold at market was primarily limited by the number of boats that could be built and

***MAQUILADORAS,* or factories near the border between the U.S. and Mexico, have become a troublesome source of pollution for that area. Some U.S. manufacturers have built such factories in Mexico to take advantage of that country's lower labor costs and pollution-control standards. If commerce becomes less regulated, such problems may become more common. Mexican environmentalists closed this plant after showing that it was contaminating its vicinity with lead.**



NATIONAL SELF-SUFFICIENCY is a good commonly overlooked by free traders. Just as nations are better off having their own symphony orchestras and other cultural offerings, they should also keep their vital industries local.

manned; now it is limited by the number of fish in the sea.

As long as the scale of the human economy was very small relative to the ecosystem, no apparent sacrifice was involved in increasing it. The scale of the economy is now such that painless growth is no longer reasonable. If we see the economy as a subsystem of a finite, nongrowing ecosystem, then there must be a maximal scale for its throughput of matter and energy. More important, there must also be an optimal scale. Economic growth beyond that optimum would increase the environmental costs faster than it would the production benefits, thereby ushering in an antieconomic phase that impoverished rather than enriched.

One can find disturbing evidence that we have already passed that point and, like Alice in *Through the Looking Glass,* the faster we run the farther behind we fall. Thus, the correlation between gross national product (GNP) and the index of sustainable economic welfare (which is based on personal consumption and adjusted for depletion of natural capital and other factors) has taken a negative turn in the U.S.

Like our planet, the economy may continue forever to develop qualitatively, but it cannot grow indefinitely and must eventually settle into a steady state in its physical dimensions. That condition need not be miserable, however. We economists need to make the elementary distinction between growth (a quantitative increase in size resulting from the accretion or assimilation of materials) and development (the qualitative evolution to a fuller, better or different state). Quantitative and qualitative changes follow different laws. Conflating the two, as we currently do in the GNP, has led to much confusion.

Development without growth is sustainable development. An economy that is steady in scale may still continue to develop a greater capacity to satisfy human wants by increasing the efficiency of its resource use, by improving social institutions and by clarifying its ethical priorities—but not by increasing the resource throughput.

In the light of the growth versus development distinction, let us return to the issue of international trade and consider two questions: What is the likely effect of free trade on growth? What is the likely effect of free trade on development?

Free trade is likely to stimulate the growth of throughput. It allows a country in effect to exceed its domestic regenerative and absorptive limits by "importing" those capacities from other countries. True, a country "exporting" some of its carrying capacity in return for imported products might have increased its throughput even more if it had made those products domestically. Overall, nevertheless, trade does postpone the day when countries must face up to living within their natural regenerative and absorptive capacities. That some countries still have excess carrying capacity is more indicative of a shortfall in their desired domestic growth than of any conscious decision to reserve that capacity for export.

By spatially separating the costs and benefits of environmental exploitation, international trade makes them harder to compare. It thereby increases the tendency for economies to overshoot their optimal scale. Furthermore, it forces countries to face tightening environmental constraints more simultaneously and less sequentially than would otherwise be the case. They have less opportunity to learn from one another's experiences with controlling throughput and less control over their local environment.

The standard arguments for free trade based on comparative advantage also depend on static promotions of efficiency. In other words, free trade in toxic wastes promotes static efficiency by allowing the disposal of wastes wherever it costs less according to today's prices and technologies. A more dynamic efficiency would be served by outlawing the export of toxins. That step would internalize the disposal costs of toxins to their place of origin—to both the firm that generated them and the nation under whose laws the firm operated. This policy creates an incentive to find technically superior ways of dealing with the toxins or of redesigning processes to avoid their production in the first place.

All these allocative, distributional and scale problems stemming from free trade ought to reverse the traditional default position favoring it. Measures to integrate national economies further should now be treated as a bad idea unless proved otherwise in specific cases. As Ronald Findley of Columbia University characterized it, comparative advantage may well be the "deepest and most beautiful result in all of economics." Nevertheless, in a full world of internationally mobile capital, our adherence to it for policy direction is a recipe for national disintegration.

FURTHER READING

FOR THE COMMON GOOD. H. E. Daly and J. B. Cobb, Jr. Beacon Press, 1989.

INTERNATIONAL TRADE AND THE ENVIRONMENT. Edited by Patrick Low. World Bank, 1992.

POPULATION, TECHNOLOGY AND LIFESTYLE: THE TRANSITION TO SUSTAINABILITY. Edited by Robert Goodland et al. Island Press, Washington, D.C., 1992.

MYTHS AND MISCONCEPTIONS OF FREE TRADE. Ravi Batra. Scribner's, 1993.

INTERNATIONAL TRADE AND ENVIRONMENT. Edited by Carl Folke et al. Special issue of *Ecological Economics,* Vol. 9, No. 1; February 1994 (in press).

Chemical Signaling in the Brain

Studies of acetylcholine receptors in the electric organs of fish have generated critical insights into how neurons in the human brain communicate with one another

by Jean-Pierre Changeux

As long ago as 1904, the British scientist T. R. Elliot proposed correctly that neurons (nerve cells) often communicate with one another and with other cell types not electrically but chemically. He suggested that an action potential, or electrical impulse, propagating along an excited neuron triggers the release of chemicals (now called neurotransmitters) from the excited cell. In turn, the liberated chemicals may cause another cell to take in or extrude selected ions. By thus altering the flow of charge across the membrane of this second cell, the neurotransmitters can give rise to a new impulse.

Since then, investigators have identified perhaps 50 neurotransmitters and have learned that a single neuron may secrete several of them. Workers have also struggled to explain just how neurotransmitters, particularly those in the brain, manage to regulate ionic transport, and hence impulse production, in the cells they influence.

The solution to this last problem has emerged slowly, but research in my laboratory at the Pasteur Institute in Paris and in other laboratories during the past 25 years has made great headway. We have ascertained that receptors for neurotransmitters—which protrude from cell membranes—play a pivotal role in mediating the conversion of chemical signals into electrical activity. And we have begun to clarify how certain major receptors carry out this challenging task. Those molecules have been found to constitute a remarkable superfamily of neurotransmitter receptors that are known as neurotransmitter-gated ion channels.

Much of this insight derives from intensive study in the 1970s and 1980s of a receptor initially isolated from the electricity-generating organ of an electric fish. Yet the story of how investigators solved the mysteries of impulse transmission by chemicals more properly begins decades earlier, with the pioneering contributions of John Newport Langley of the University of Cambridge. In 1906 he proposed that bodily tissues bear receptors for drugs. In so doing, he provided one of the first significant clues to the means by which neurotransmitters exert their effects.

Langley based his proposal on studies he conducted into how the poison curare kills its victims. When he undertook these investigations, he was aware that curare causes asphyxiation: it blocks motor nerves from inducing contraction of respiratory muscles. He wondered, though, whether it acted on the nerves or on the muscles. To find out, he placed large doses of nicotine, a substance that normally induces contraction, directly onto strips of skeletal muscle from chicken (at a site where motor nerves are normally connected). A contraction followed. Then he applied curare. The drug blocked nicotine's action. Langley concluded that curare interacts directly with muscle tissue, which displays on its surface an "especially excitable component," or "receptive substance," capable of combining with either nicotine or curare.

Neurobiologists now understand that nicotine and curare couple with the part of the receptor molecule designed to bind to the neurotransmitter acetylcholine. Bound nicotine serves as an agonist: it mimics the stimulatory effect of a naturally produced substance, acetylcholine. Bound curare, in contrast, is a competitive antagonist: it issues no stimulatory signal and, at the same time, prevents acetylcholine and nicotine from doing so.

Despite the brilliance of the receptor concept, its value eluded the scientific community for decades. Skepticism arose in part because scientists lacked the tools for isolating receptors. Moreover, they had trouble imagining how binding of a chemical to a receptor molecule at the cell surface could influence the flow of ions through channels in the cell membrane.

I helped to ease these objections in the mid-1960s, when, as a graduate student working on my doctoral dissertation, I suggested a theoretical solution to this conceptual difficulty. A few years earlier, structural studies of hemoglobin and various enzymes had indicated that these molecules included several separate sites capable of associating with other substances. I—together with my teachers Jacques Monod and François Jacob and their colleague Jeffries Wyman—postulated that certain enzymes may be activated by allosteric, or indirect, means, by which binding at one site influences behavior of another site without any assistance from an additional source of energy. We assumed

ACETYLCHOLINE RECEPTOR, which consists of five subunits (*left*), was the first neurotransmitter receptor to be isolated. Later work showed it to include not only neurotransmitter binding sites but also an ion-transporting channel (*right*). (The beta and delta subunits and part of one alpha subunit have been cut away for clarity.) The channel is closed when the receptor is at rest, but it opens rapidly when the two alpha subunits both combine with acetylcholine.

JEAN-PIERRE CHANGEUX has been director of the molecular neurobiology unit of the Pasteur Institute in Paris since 1972. He is also a professor at the institute and at the College of France. Changeux has won many awards for his contributions to neuroscience.

that attachment of some substance to a docking site on an enzyme could propagate a conformational change throughout the enzyme, thereby rendering a distant site able to act on a substrate (the substance transformed by an enzyme).

In my dissertation I noted briefly that receptors for neurotransmitters might function similarly. They might contain both a neurotransmitter binding site and a separate region that forms an ion channel. Attachment of the neurotransmitter to the binding site could elicit a conformational change in the molecule that would culminate in the opening of its channel component. To evaluate the merit of this idea, my co-workers and I had to analyze the composition of some kind of receptor in detail. For this, we needed a good supply. Unfortunately, no receptor had yet been isolated, and so that task became our mission.

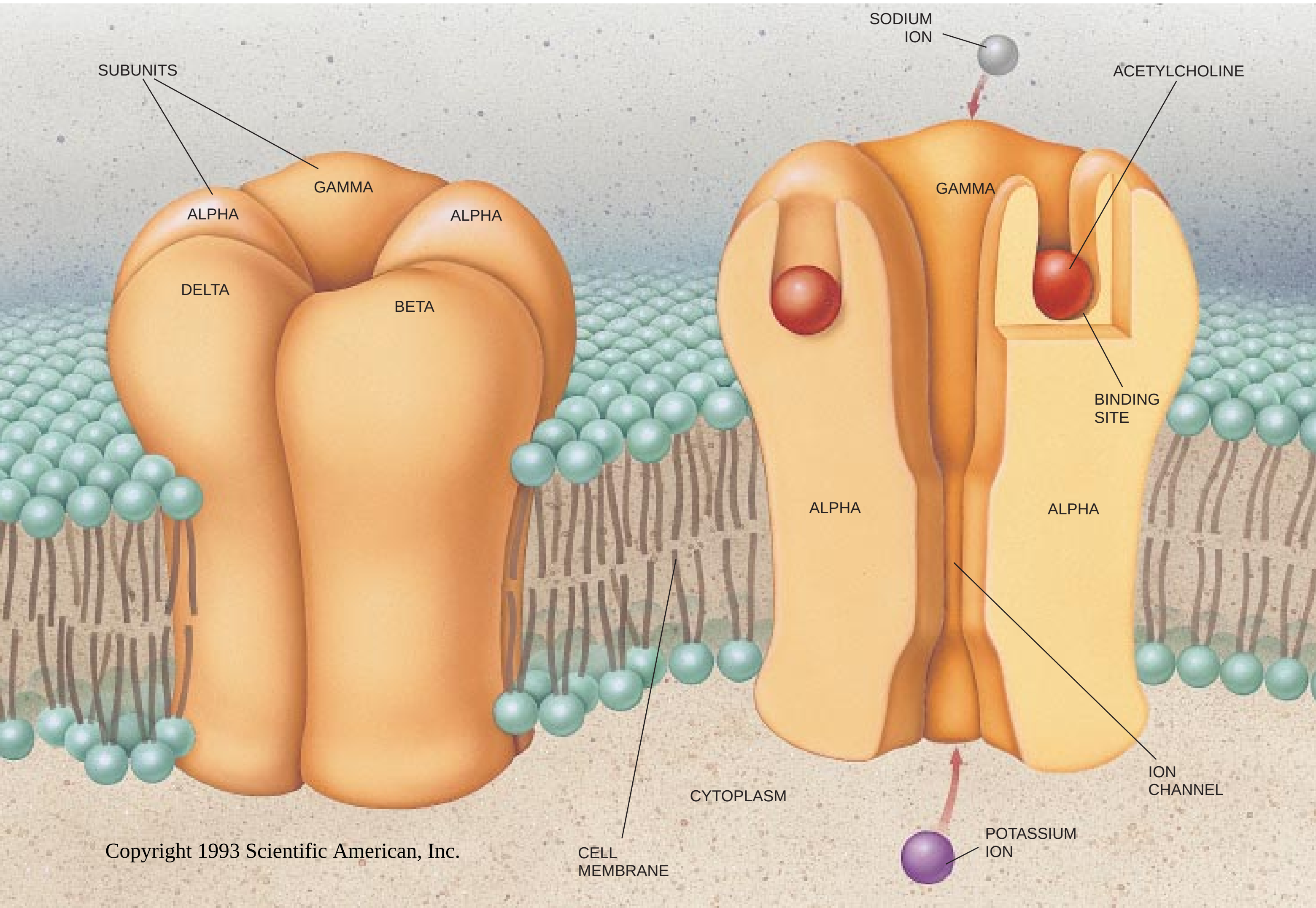
Our choice of receptor was inspired by discoveries made by David Nachmansohn after he fled Nazi Germany. In the late 1930s, while at the University of Paris–Sorbonne, Nachmansohn and his colleagues showed that acetylcholine not only induces muscle to contract, it also causes electricity-generating organs of electric fish to produce current. Furthermore, the organs offer two particular advantages for researchers. The constituent cells, called electrocytes, are huge and thus relatively easy to handle. Additionally, they number in the billions, which means electric organs harbor an abundance of acetylcholine receptor molecules.

With these advantages in mind, we decided to isolate the acetylcholine receptor in the electric organ of the electric eel (*Electrophorus electricus*). First, we had to break up the electrocytes in the organ to create preparations that could be analyzed chemically. Michiki Kasai, now at Osaka University, and I therefore ground up electric tissue. Then we separated out micron-sized fragments of membranes from the innervated regions of electrocytes. Langley's studies of muscle tissue suggested we would find a high concentration of the receptor in the innervated areas. These membrane fragments have a wonderfully useful property: they close up into microsacs, or tiny vesicles, that can be filled with radioactively labeled sodium (Na^+) and potassium (K^+) ions.

As would be expected if functional copies of the receptor were present in these microsacs, addition of acetylcholine to a suspension of the vesicles dramatically altered the flow of ions into and out of the vesicles—just as occurs in intact electrocytes when they respond to acetylcholine. Moreover, in agreement with early suggestions of an allosteric mechanism, no additional energy supply was required for the reaction to take place. Hence, we felt reasonably sure that the receptor was present and functional in the microsac membranes.

Still, we needed some way to distinguish the receptor from the rest of the material in the membranes. At that time, the only way to pinpoint a molecular species on a membrane was to radioactively label a substance that homed to it and bound tightly. We were having difficulty finding a suitable homing material when Chen-Yuan Lee of the National Taiwan University came to our rescue.

Lee happened to visit my laboratory in the spring of 1970 to present his research on the structure and action of snake venom. The bites of several snakes, such as the banded krait (*Bungarus multicinctus*) and the cobra, are fatal because their venom contains tox-



ic molecules that, like curare, block signal transmission by motor neurons. Among these molecules are alpha toxins. Lee reported that even at low concentrations, alpha-bungarotoxin from the banded krait almost irreversibly blocks the effects of acetylcholine on the muscles of evolutionarily advanced vertebrates. I realized then that alpha-bungarotoxin might provide the specificity we needed for identifying the acetylcholine receptor on the microsacs we derived from the electric eel. To our delight, toxin supplied by Lee did our bidding perfectly.

We could finally take up the task of purifying the receptor, which we soon found was a protein. By 1974 our group and several others had succeeded. My team reached this goal by applying a technique called affinity chromatography. We created insoluble beads to which arms that ended in a structural analogue of curare had been attached. Then we doused the beads with microsac membranes that had been dissolved in a detergent solution to separate the constituent molecules. The free receptor molecules bound to the analogue, and the rest of the solution floated away. Next we poured copies of the curare analogue over the beads. Now the receptor molecules bound preferentially to the added analogue and came off the beads. By passing the resulting complexes of receptor and analogue through a membrane permeable only to the curare substitute, we eliminated the analogue and acquired a pure supply of the receptor.

Eager to learn something about the structure of our prize, we presented Jean Cartaud of the Jacques Monod Institute in Paris with samples to view in an electron microscope. He found that when observed from above, the receptor resembled a rosette with a depression in the center. Other analyses, by Arthur Karlin of the Columbia University College of Physicians and Surgeons and by Michael Raftery of the California Institute of Technology, revealed that the overall molecule is composed of five protein chains, or subunits: two alpha chains, which have an identical molecular weight, and three chains called beta, gamma and delta, which vary in molecular weight. Moreover, Karlin demonstrated that the alpha subunits bear the primary responsibility for recognizing acetylcholine. (Today it is clear that the binding site on each of the two alpha subunits must be occupied in order for the ion channel to open.)

These results intrigued us. They suggested that each subunit might form one "petal" of the rosette and that the central depression might reflect the extracellular entryway to a membrane-spanning ion channel. We needed more data in order to test that idea, but in the interim we had to cope with another nagging problem. Could we be sure that receptor molecules contained an ion channel, not solely the site that bound acetylcholine?

In 1974 my co-worker Gerald L. Hazelbauer and I tackled this question by incorporating proteins from purified extracts of microsacs into lipid membranes that enclosed radioactively labeled sodium or potassium ions. As would be predicted if the channel were present, binding by acetylcholine triggered the flow of ions, and binding by alpha-bungarotoxin and curare blocked changes in ionic flux. Later we confirmed the results with purified receptors themselves. Hence, by 1980 we and other groups had clearly shown that the pure protein does indeed contain all the structural elements needed for the chemical transmission of an electrical signal—namely, an acetylcholine binding site, an ion channel and a mechanism for coupling their activity.

To attack the deeper problem of how the acetylcholine receptor worked, we had to decipher the sequence of its constituent amino acid. Such information provides clues to the shape adopted when a protein, which is little more than a string of amino acids, folds in on itself. And knowledge of the folded structure offers clues to the functions of the various domains in that structure.

Introduction of automated sequencing and genetic engineering techniques in the late 1970s facilitated this effort. In 1979 my colleague Anne Devillers-Thiéry, Donny Strosberg of the Jacques Monod Institute and I elucidated the sequence of the first 20 amino acids at one end (called the amino terminal) of the alpha subunit in the acetylcholine receptor of the European, or marbled, electric ray (*Torpedo marmorata*). Subsequently, Raftery and Leroy E. Hood and their colleagues at Caltech identified essentially the same sequence in the Californian electric ray (*T. californica*) and went further. When they characterized the 54 amino acids abutting the amino terminal of the alpha, beta, gamma and delta subunits, they unexpectedly found striking similarity: 35 to 50 percent of the sequence was identical, or homologous, in all four subunits.

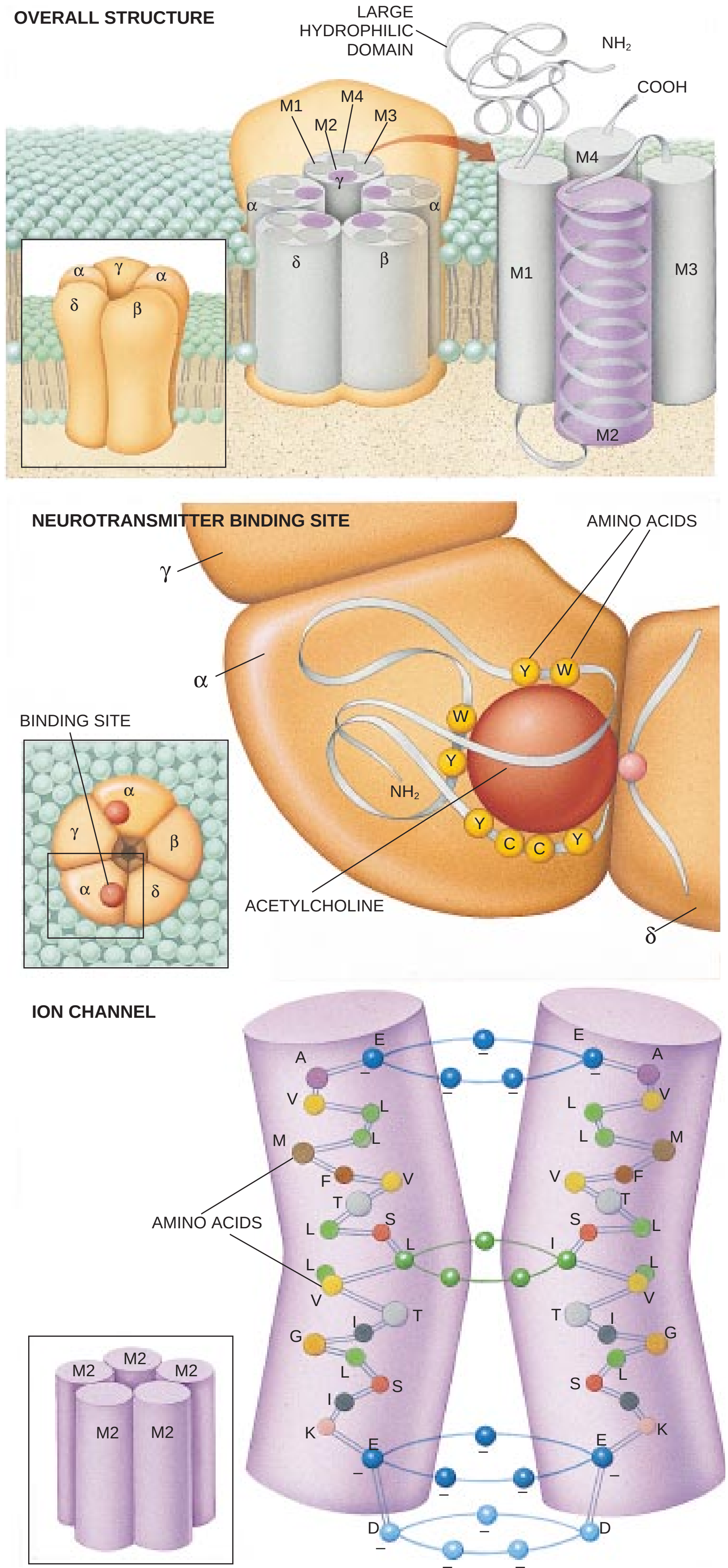
Molecular biologists interpret such identity to mean that the genes specifying the amino acid sequences of the subunits are descendants of some ancestral gene that duplicated twice (and underwent subsequent alteration) in the course of evolution. The homology further implied that the complete subunits were similar to one another and therefore probably did arrange themselves quasisymmetrically around a central axis, forming the petals on the rosette seen in the electron microscope.

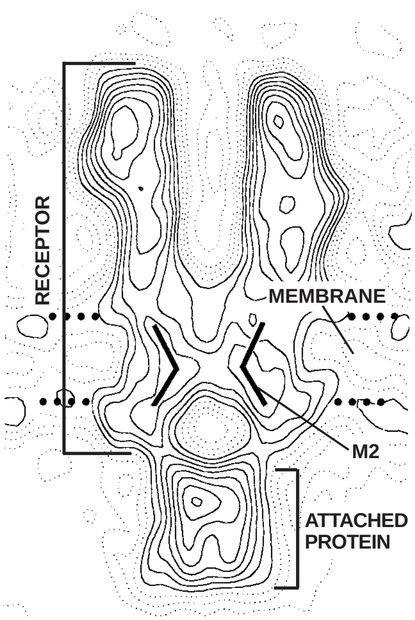
By 1983 more complete information had emerged. Shosaku Numa and his team at Kyoto University had solved the full sequences of the alpha and then

ELECTRIC RAY from the *Torpedo* genus harbors an electricity-generating organ that contains billions of copies of an acetylcholine receptor. Identification of the receptor's amino acid sequence led to the discovery that acetylcholine receptors in the muscle and brain of humans are structurally similar to the *Torpedo* receptor.

Anatomy of the Acetylcholine Receptor

Major structural features of the acetylcholine receptor have begun to yield to scrutiny. Its five subunits (*inset in top panel*), which can be depicted schematically as cylinders (*top, center*), are each formed from a protein that has folded in on itself (*detail at top right*). Every subunit includes a large hydrophilic (water-loving) region abutting the amino (NH_2) terminal, as well as four hydrophobic (water-hating) membrane-spanning segments: M1, M2, M3 and M4. The neurotransmitter binding sites, viewed from above in the middle panel (*inset*), consist primarily of amino acids (*yellow spheres in detail*) residing in the large hydrophilic region of the alpha subunits. (Letters in the spheres represent specific amino acids.) Neighboring subunits contribute as well (*pink sphere*). The ion channel is composed of five M2 segments (*inset in bottom panel*) and contains several rings of amino acids that affect the functioning of the receptor. Among these, three negatively charged rings (*blue in detail showing two M2 segments in a specific acetylcholine receptor*) draw positively charged ions (*not shown*) through the channel. An uncharged, leucine ring (*green*) at the center, where M2 segments probably bend, participates in closing the ion channel when the receptor becomes desensitized to acetylcholine. The electron density map (*bottom left*), showing a slice through the receptor, indicates the probable orientation of two M2 segments (*dark bars*). Nigel Unwin of the Medical Research Council in Cambridge, England, supplied that image.





the beta, gamma and delta subunits of the *T. californica* receptor. Three other laboratories, including mine, had also identified the sequence of the *T. californica* gamma subunit and supplied that of the *T. marmorata* alpha subunit. Then Numa published the sequences of all the subunits in the acetylcholine receptor on human muscle. The muscle receptor turned out to differ little from those of electrocytes.

These studies put workers in an excellent position to ascertain something about the architecture of the folded subunits and how they might fit together. In trying to predict the structure of a folded protein, scientists often scan the linear amino acid sequence for stretches that are rich in either hydrophilic or hydrophobic amino acids. Hydrophilic substances are attracted to water, such as that in cytoplasm or in the fluids that bathe cells; hydrophobic, or water-hating, substances prefer to associate with other hydrophobic entities, such as the lipids that form cell membranes.

Every subunit that had been sequenced began at the amino-terminal side with a large hydrophilic region and housed four separate hydrophobic segments of about 20 amino acids. The hydrophobic areas are referred to, beginning with the one closest to the lengthy hydrophilic domain, as M1, M2, M3 and M4. This arrangement suggested that every subunit chain snaked through the thickness of the cell membrane four times, so that all four of the hydrophobic regions spanned the membrane [*see box on preceding page*].

In one model, at least, the large hydrophilic domain protruded into the extracellular space. So situated, that domain on the alpha subunits (the subunits primarily responsible for grasping acetylcholine) would be well positioned to serve as a neurotransmitter binding site. It further seemed reasonable to guess that one membrane-spanning segment from each of the five subunits associated with its counterparts in the other subunits to form the ion channel. The like segments could form such a channel if together they encircled the central axis of the rosette.

Later research confirmed this scenario and added important details. More immediately, though, the sequencing of the subunit proteins enabled researchers to dispel confusion over the operation of acetylcholine receptors in the brain. The trouble derived from the response of certain of these receptors to snake venoms.

By the early 1980s investigators knew that nicotine-sensitive, or nicotinic, acetylcholine receptors were present in the brain of higher vertebrates. Neurobiologists were puzzled, however, when alpha-bungarotoxin seemed to block the functioning of certain of the receptors but not others. What is more, a *Bungarus* toxin called neuronal bungarotoxin apparently attached to some cerebral receptors but, again, not to others. (The picture is actually more complicated. The brain additionally includes a class of acetylcholine receptors named muscarinic receptors that I shall not discuss here. Those receptors are very different from the nicotinic types. They are formed from a single protein chain and do not include an ion channel. They exert their effects through such intracellular mediators as G proteins.)

The smoke cleared when James W. Patrick and Stephen F. Heinemann and their colleagues at the Salk Institute for Biological Studies in San Diego and Marc Ballivet of the University of Geneva used their knowledge of the structure of electrocyte and muscle subunits to decipher the amino acid sequences of cerebral subunits. The investigators guessed that the amino acid sequences of the cerebral subunits probably resembled those of electrocytes and muscle even though certain cerebral receptors behaved somewhat differently from their counterparts in electrocytes and muscle. If the proteins were similar, then the genes specifying their amino acid sequences would be similar as well. That being the case, a process called DNA hybridization could be expected to help isolate the cerebral subunit genes and thereby uncover the amino acid sequences of the corresponding subunits.

DNA hybridization techniques capitalize on a prominent characteristic of genes. Genes consist of two strands of nucleotides (the building blocks of DNA). One strand from a gene will readily combine, or hybridize, with the other strand from the same or a closely related gene. Aware of this propensity, the investigators hoped they could retrieve the cerebral subunit genes from a larger pool of brain-derived DNA by "fishing" for them with "hooks" made of nucleotide sequences that actually direct the synthesis of electrocyte or muscle subunits. The procedure worked beautifully. Seven alpha-subunit types (each of which is numbered) were found to be produced in the brains of vertebrates, including humans. Three nonalpha types, often classified as beta subunits, were discovered as well. This diversity suggested that the variable responses of cerebral receptors to snake venoms derive from slight differences in the amino acid sequences of one or more subunits.

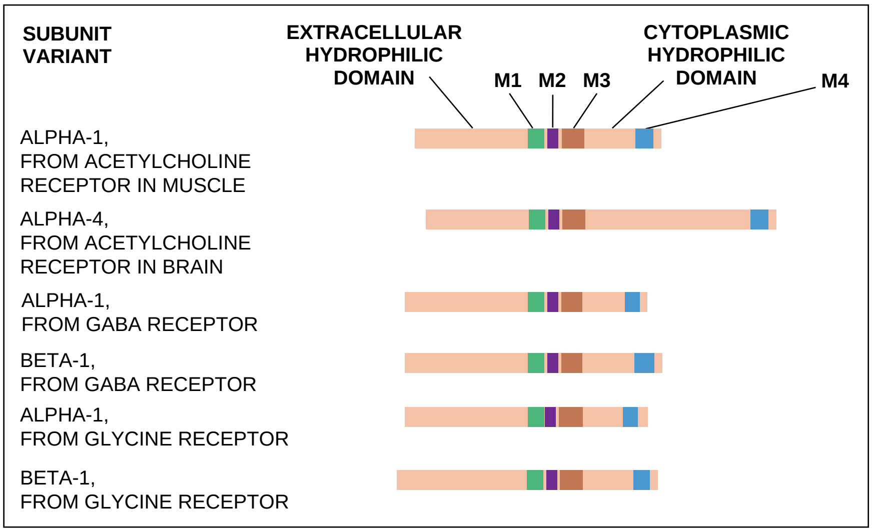
Subsequent studies have demonstrated that the protein products of most of the subunit genes identified to date can yield functional receptors in living cells if the cells make at least one alpha and one nonalpha variant. The experiments yielding this conclusion often involved injecting the genes into the nucleus of oocytes, or immature eggs, from the frog *Xenopus.* In response, the protein-making machinery of the oocytes transcribed the genes into messenger RNA and, after transporting the RNA to the cytoplasm, translated it into the specified proteins. Then the proteins associated with one another in groups of five to produce receptors.

Evaluations of many receptors produced by *Xenopus* oocytes also demonstrated that substitution of one subunit variant for another in a receptor can indeed change some properties of the receptor. As an example, neuronal bungarotoxin blocks the response to acetylcholine receptors composed of beta-2 and either alpha-3 or alpha-4 subunits, but the toxin does not interfere with the activity of molecules composed of beta-2 and alpha-2 subunits.

At about the same time as the heterogeneity of nicotinic acetylcholine receptors was emerging, researchers were busy attempting to decipher the structure and operation of receptors for other neurotransmitters. When that work began, few would have guessed that the receptors for gamma-aminobutyric acid (GABA) and glycine would have much in common with acetylcholine receptors. After all, nicotinic acetylcholine receptors excite cells by opening a channel permeable to cations (positively charged ions). Receptors for GABA and glycine, in contrast, facilitate transport of chloride anions (Cl^-); the flow of chloride anions into cells inhibits generation of electrical impulses and can thereby counteract the effects of excitatory receptors.

Nevertheless, studies conducted in the 1980s revealed that glycine and GABA receptors consist of multiple subunits. That in itself was not remarkable. More strikingly, however, Heinrich Betz of the University of Heidelberg and Eric A. Barnard of the Medical Research Council in Cambridge, England, respectively determined the complete sequences of the glycine and GABA receptors and found that the distribution of hydrophilic and hydrophobic domains strongly resembled that of nicotinic acetylcholine receptors.

In other words, it began to seem likely that the subunits in the GABA and



UNFOLDED PROTEIN CHAINS (*multicolored bars*) that constitute subunits in receptors for acetylcholine, GABA and glycine have much in common. All harbor a large, extracellular hydrophilic domain, a smaller, cytoplasmic hydrophilic domain and four hydrophobic segments (M1, M2, M3 and M4) believed to span the cell membrane. These similarities suggest that the molecules all belong to one superfamily of structurally related neurotransmitter receptors.

glycine receptors, in common with those of nicotinic acetylcholine receptors, weave through the cell membrane four times. Evidence also indicated that the complete receptors for GABA and glycine carry both a neurotransmitter binding site and an ion channel. More recent work suggests that some serotonin receptors have a similar architecture as well. These receptors, like acetylcholine receptors, control the cross-membrane transport of cations and are thus excitatory.

The architectural similarities among the receptors explain why neurobiologists now consider acetylcholine, GABA, glycine and serotonin receptors to constitute the superfamily of genetically and structurally related neurotransmitter-gated ion channels. (Receptors for the prevalent neurotransmitter glutamate may be distantly related. They include a neurotransmitter binding site and an ion channel but differ in structure.) There is also evidence that neurotransmitter-gated ion channels are allosteric proteins. As would be expected for allosteric molecules, the estimated distance between the neurotransmitter binding site and the ion channel is large—about 30 angstroms.

As was found for acetylcholine receptors, the subunits of other members of the superfamily come in multiple varieties. Hence, a GABA receptor in one part of the brain might well have somewhat different properties than does a variant elsewhere in that organ. For instance, benzodiazepines, which are so abundantly consumed as tranquilizers by industrialized populations, potentiate the inhibitory action of only certain GABA receptor species. They do so by binding to a site that is distinct from the GABA binding site.

As the precise influence on behavior of every subspecies of every subunit in neurotransmitter-gated ion channels is deciphered, pharmacologists should be able to design drugs that will selectively impede or enhance those effects. Such agents, in turn, might help ameliorate any number of debilitating conditions, including mood disorders, tissue damage associated with stroke and, perhaps, Alzheimer's disease.

Of course, to devise such drugs, researchers require a rather full understanding of receptor structure. They need to know the specific amino acids responsible for binding neurotransmitters, for directing the flow of ions in and out of cells and for otherwise modulating receptor function. One useful way to gather such information is known as affinity labeling. Some traceable version of a molecule that interacts with a receptor is allowed to bind irreversibly to that target; the bound substance thus highlights the amino acids that constitute the binding site.

Between 1988 and 1990, my colleagues Michael Dennis, Jérôme Giraudat, Jean-Luc Galzi and I uncovered much of the acetylcholine binding site by identifying amino acids in a *Torpedo* receptor that were labeled by the compound *p*–(*N,N*–dimethyl) aminobenzenediazonium fluoroborate, also called DDF. We learned that several aromatic amino acids (those carrying ring-shaped side chains) are critical to DDF binding, and we confirmed binding by a pair of cysteine amino acids identified in a *Torpedo* receptor by Karlin. The labeled amino acids are distributed within three distinct regions of the large hydrophilic domain of the amino-terminal region. It became evident that they collectively form a kind of negatively charged cup in which the positively charged part of acetylcholine could lodge.

What is even more exciting, we went on to show that these amino acids actually do play a critical role in receptor function. With Daniel Bertrand of the University of Geneva Medical Center, we proved this point in a receptor that consists entirely of alpha-7 subunits from the chicken brain. (This receptor is one exception to the rule requiring the presence of both alpha and beta subunits for receptor formation.) Specific mutation, by what is called site-directed mutagenesis, of the amino acids that DDF labeled in the *Torpedo* receptor strikingly impeded the alpha-7 receptor's response to acetylcholine.

Taken together, the affinity-labeling and mutagenesis studies confirm that the large hydrophilic region of the alpha subunit is exposed to the extracellular environment. There it sits, ready to receive acetylcholine released from nerve endings and to trigger the opening of the ion-transporting channel.

Affinity labeling also delineated the structure of the ion channel in a *Torpedo* receptor. Difficult analyses convinced us by the end of 1985 that the drug chlorpromazine attaches to amino acids on the membrane-crossing, M2 hydrophobic segment of at least one subunit—the delta chain. This work, and a similar report by Ferdinand Hucho and his co-workers at Berlin University, suggested that the channel's inner wall is formed by five M2 segments, one contributed by every subunit.

Numa and Bert Sakmann, then at the Max Planck Institute for Biophysical Chemistry in Göttingen, confirmed this possibility. By site-directed mutagenesis, they determined that at least three rings of negatively charged amino acids (especially glutamate) participate in transporting ions through the channel. Each ring lies in a plane parallel to the surface of the cell and consists of five amino acids, one supplied by the M2 segment of every subunit. A single ring resides at the extracellular surface of the membrane (at the top of the channel). A second, termed the intermediate ring, lies at the bottom of the channel, and the third ring lies directly below the second, in the cytoplasm proper.

Given that the distribution of hydrophobic subunits in the GABA, glycine and serotonin receptors matches that of acetylcholine receptors, we wondered if their M2 segments formed the channel in those receptors as well. They do, even though those receptors transport negatively, rather than positively, charged ions. The difference in charge preference apparently stems from variance in just a few amino acids. When Bertrand's team and mine transferred into the alpha-7 receptor three M2 amino acids from a GABA receptor (including the one giving rise to the intermediate ring), those few changes converted the alpha-7 receptor channel from a cationic to an anionic transporter.

Whether a receptor carries an anion- or cation-transporting channel, its main function is to open that channel in response to signals from a neurotransmitter. Yet neurotransmitter receptors have another fascinating skill as well. By altering their conformation, they can apparently increase or decrease their readiness to respond to neurotransmitters. In that way, they can regulate the pool of receptors available to respond to external signals and can thus influence the efficiency of signal transmission.

My associates and I realized that receptors could possess this regulatory power when we began to consider a phenomenon noted by several investigators over the years. Receptors react differently to discrete pulses of highly concentrated acetylcholine (such as those usually delivered by neurons) than they do to the continuous availability of lower concentrations (such as is provided in many experiments).

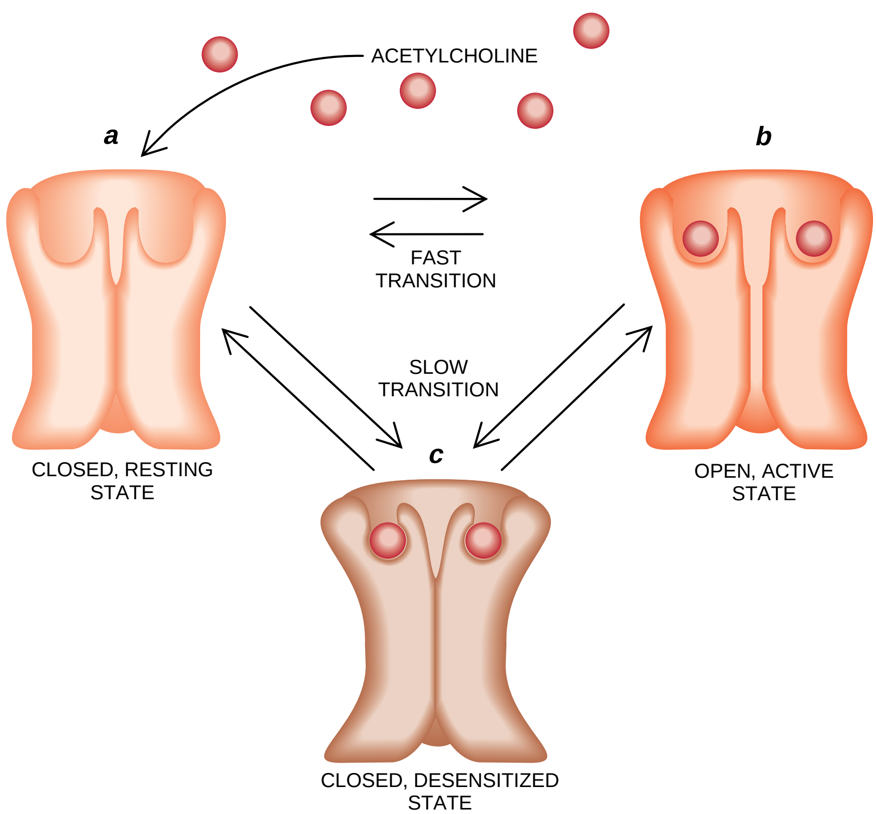
Excited neurons secrete large concentrations of acetylcholine in discrete bursts at synapses, the specialized junctions now known to connect neurons. The molecules freed during a single pulse pour into the synaptic cleft (the space separating communicating cells). Many of them make their way from the excited, presynaptic cell to receptors on the surface of a postsynaptic cell. Under normal circumstances, the affinity of most receptor molecules for the neurotransmitter is low. Consequently, immediately after acetylcholine binds to receptors and causes the channel to open, the receptors release their hold on the neurotransmitter, which is promptly degraded. Within milliseconds of being bound, the receptors revert to their closed, unbound state and are ready to react once again.

In contrast, when acetylcholine is supplied continuously to receptors, the receptor molecules begin to lose their responsiveness. After initially opening the ion channel, they slowly take on a "desensitized" conformation over the course of seconds or minutes. That is, they bind avidly to acetylcholine but maintain a closed channel and do not transport ions. Even small concentrations of acetylcholine will be held by these closed-channel receptors for relatively long periods, during which the receptors cannot react to new signals.

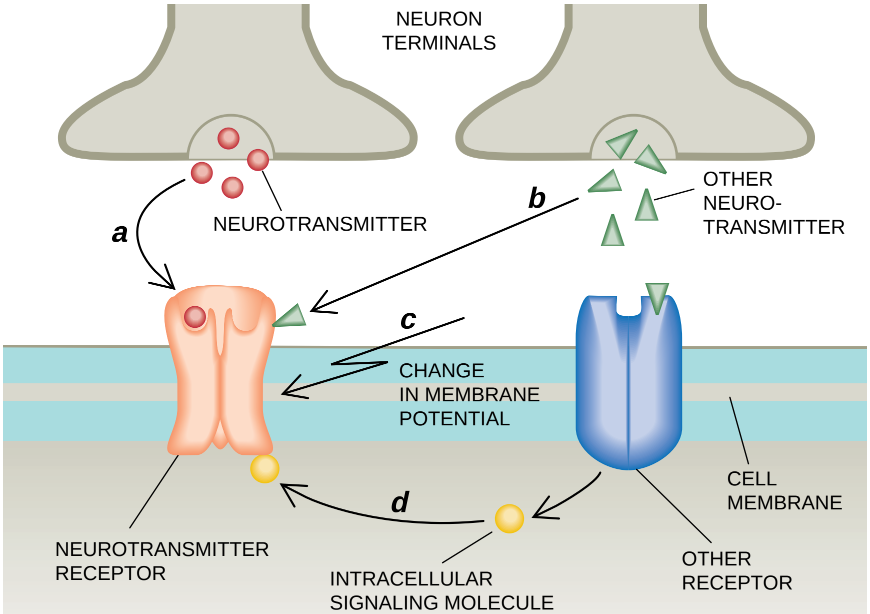
So it seems that acetylcholine receptors can adopt at least three interconvertible states that can differ in their affinity for the neurotransmitter and in the efficiency of signal transmission. In addition to the high-affinity, desensitized state, in which the channel remains closed, there is a low-affinity, resting (but activable) state in which the channel is closed but easily opened if both alpha subunits are bound suddenly by acetylcholine. The low-affinity, open-channel condition is the third state. All three states switch back and forth spontaneously but at different rates than occur when acetylcholine is present.

Site-directed mutagenesis has helped clarify the process by which desensitization occurs; it appears that leucine amino acids are involved. In affinity-labeling studies carried out by my group, chlorpromazine labeled a ring of uncharged leucine amino acids near the center of the ion channel. When Bertrand's team and mine replaced the leucines in this ring with a smaller uncharged amino acid, we created a receptor that resembled a normal desensitized receptor in that it bound tightly to acetylcholine. Yet its channel was fixed in an open state. This result implies that the leucine ring locks the ion channel closed when the receptor is in the desensitized conformation.

My associates and I have long pondered the benefits that might accrue to an organism from bearing receptors able to adopt multiple states. Of course, a desensitization mechanism would protect receptor-bearing cells from becoming overexcited in response to dangerously high levels of acetylcholine. But I believe that



DYNAMIC NATURE of the acetylcholine receptor is evident in its ability to adopt multiple conformations. In the resting state (*a*), the receptor has low affinity for acetylcholine, and its ion channel is closed. If it is exposed briefly to a high concentration of the neurotransmitter, it assumes the active, open-channel conformation for milliseconds (*b*) before releasing the acetylcholine and reverting to the resting state. If acetylcholine is supplied continuously, resting and activated receptors can slowly assume a desensitized state (*c*). In this condition, the receptor holds acetylcholine with high affinity for seconds or minutes, maintains a closed channel and will not respond to new pulses of acetylcholine.



MANY FORCES can affect the conformation of a neurotransmitter receptor and hence the efficiency with which it responds to signals received from a neuron acting on it. Aside from the concentration and delivery rate of the neurotransmitter (*a*), some influences include binding by additional neurotransmitters or other extracellular chemicals (*b*), changes in the electrical potential across the cell membrane (*c*) and binding by intracellular signaling molecules (*d*), such as ions.

there is another explanation as well.

In 1982 my colleague Thierry Heidmann and I further proposed that the ability of acetylcholine receptors to alter their conformation slowly could more routinely serve to increase or decrease the efficiency of signal transmission at a synapse. In so doing, such receptors could participate in learning. Many theorists, following the lead of Donald O. Hebb, postulate that learning depends on changes in the efficiency of signal transmission across the synapses linking two neurons that are activated simultaneously.

Our hypothesis is far from proved, but it is plausible. If the ability to adopt many states were important to regulating signal transmission, this flexibility should appear in other neurotransmitter receptors as well. Research shows that GABA, glycine and serotonin receptors also are able to assume desensitized conformations.

The demonstration of state changes in other receptors is not the only support for the possibility that receptors regulate synaptic efficiency. These molecules sit in a particularly good position to control the degree of responsiveness needed at any given moment. Crossing the cell membrane as they do, they are exposed to chemical and electrical signals issued both from outside and from within the cell. If each of these signals pushed the receptor toward one conformation or another, the final arrangement would reflect the summed, or integrated, influence of various, possibly contradictory, forces.

Among the signals that impinge on receptors are the intracellular concentration of calcium ions and changes in the electrical potential across the cell membrane. Hyperpolarization of the membrane and elevation of the calcium concentration in muscle accelerate desensitization of the acetylcholine receptor. Richard L. Huganir and Paul Greengard of the Rockefeller University further established in 1986 that phosphorylation of the receptor promotes desensitization.

If neurotransmitter receptors did in fact control the efficacy of intercellular signaling, we would anticipate that they would be capable of increasing, not merely decreasing, their sensitivity to neurotransmitters. Such potentiation has been observed. Extracellular calcium enhances the stimulatory effect of nicotinic acetylcholine receptors in the brain, and glycine enhances the effect of glutamate receptors.

In general, then, we suspect that if a receptor is tottering between two conformations, one of which is sensitive and the other of which is refractory to stimulation, the balance can be shifted toward one of the two states by chemical or electrical signals. The resulting conformation, in turn, enhances or depresses the ability of the receptor to convey signals promptly.

Clearly, our understanding of chemical signaling in the brain has advanced dramatically in the past quarter century. The decision to isolate the acetylcholine receptor from a fish organ was risky: if the effort failed or if that receptor was unrelated to any others, we would have wasted time and energy. Fortunately, the gamble paid off more than we could have hoped. The acetylcholine receptor in electric fish was identified, and its sequence was deciphered. Then biotechnology, especially recombinant DNA methodology, enabled workers to characterize related receptors in human muscle and in the brain and to learn that nicotinic acetylcholine receptors are structurally related to, and form a superfamily with, those responsive to GABA, glycine and serotonin.

Meanwhile it became clear that the subunits composing each type of receptor are themselves variable: a receptor in one part of the brain may well possess properties that differ from those of their immediate kin elsewhere in cerebral tissue. This likelihood raises the tantalizing possibility that drugs targeted to specific receptors on defined categories of neurons can be developed for the highly selective treatment of signaling disorders in the brain.

FURTHER READING

NEURONAL MAN: THE BIOLOGY OF MIND. J. P. Changeux. Translated by Laurence Garey. Pantheon Books, 1985.

FUNCTIONAL ARCHITECTURE AND DYNAMICS OF THE NICOTINIC ACETYLCHOLINE RECEPTOR: AN ALLOSTERIC LIGAND-GATED ION CHANNEL. J. P. Changeux in *FIDIA Research Foundation Neuroscience Award Lectures,* Vol. 4. Raven Press, 1990.

EXPLORATIONS OF THE NICOTINIC ACETYLCHOLINE RECEPTOR. A. Karlin in *Harvey Lectures: 1989–1990,* Vol. 85, pages 71–107; 1991.

THE FUNCTIONAL ARCHITECTURE OF THE ACETYLCHOLINE NICOTINIC RECEPTOR EXPLORED BY AFFINITY LABELLING AND SITE-DIRECTED MUTAGENESIS. J. P. Changeux, J. L. Galzi, A. Devillers-Thiéry and D. Bertrand in *Quarterly Reviews of Biophysics,* Vol. 25, No. 4, pages 395–432; November 1992.

NICOTINIC ACETYLCHOLINE RECEPTOR AT 9 Å RESOLUTION. N. Unwin in *Journal of Molecular Biology,* Vol. 229, No. 4, pages 1101–1124; February 20, 1993.

X-ray Binaries

In these systems, ultradense neutron stars feed on their more sedate companions. Such stellar cannibalism produces brilliant outpourings of x-rays and drastically alters the evolution of both stars

by Edward P. J. van den Heuvel and Jan van Paradijs

All the shimmering stars that pierce the night sky shine because of the same fundamental process: nuclear fusion. When two or more atomic nuclei collide and fuse into one, they release virtually unimaginable amounts of energy. The fusion of one gram of hydrogen, for example, liberates as much energy as the combustion of 20,000 liters of gasoline. In stars such as the sun, fusion reactions burn brilliantly for billions of years. They are not the only source of stellar energy, however. In 1971 astronomers recognized a class of bizarre, x-ray-emitting stars, known as x-ray binaries, whose intense emissions require an energy source far more efficient than even fusion.

Theorists have deduced that these objects consist of a normal star orbiting a collapsed stellar corpse, usually a neutron star. Neutron stars are so dense that the entire mass of the star is squeezed into what is essentially a single atomic nucleus 20 kilometers across. The stars in these binaries lie so close together that gas can flow from the normal star to the neutron star. That captured material forms a rapidly swirling disk whose inner edge, just above the neutron star's surface, races around at nearly the speed of light. Friction within the disk eventually causes the gas to fall inward, or accrete, onto the neutron star. In the process, violent collisions between particles heat the gas to temperatures of 10 million to 100 million kelvins. Under such incredibly hot conditions, the gas emits torrents of energetic x-rays. Pound for pound, accretion unleashes 15 to 60 times as much energy as does hydrogen fusion.

Astronomers now recognize that accretion powers a rich diversity of astrophysical objects. These range from infant stars to quasars, objects about the size of the solar system that outshine entire galaxies, most likely as a result of gas spiraling into a supermassive black hole. X-ray binaries serve as ideal showcases for learning in detail how the accretion process works. They are bright and relatively nearby, residing well within our galaxy.

The study of x-ray binaries also provides a glimpse into the life cycle of some of the most exotic and dynamic stellar systems in the sky. In these stellar duos, one or both members spends some time feeding off its partner. That transfer of material stunningly alters both stars' development. One star may pay for its gluttony by prematurely ending its life in a spectacular supernova explosion. On the other hand, placid, elderly neutron stars may receive an infusion of rotational energy that causes them to become a prominent source of rapidly pulsed radio waves.

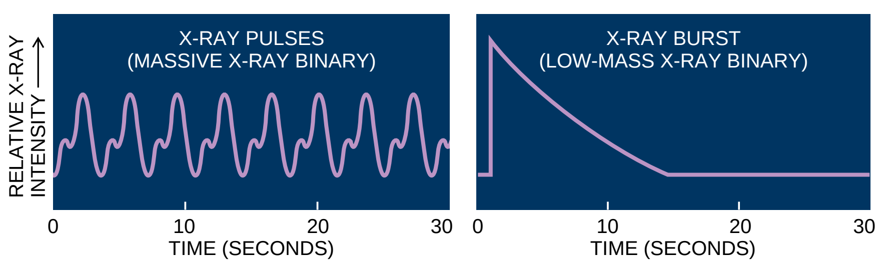
Despite their prominence in the x-ray sky, x-ray binaries escaped the notice of researchers until the dawn of the space age in the 1960s. Celestial x-rays are absorbed high in the upper atmosphere, precluding their detection from the ground. The advent of space technology opened up an entirely new field of investigation by making it possible to loft telescopes above the obscuring layers of the earth's atmosphere.

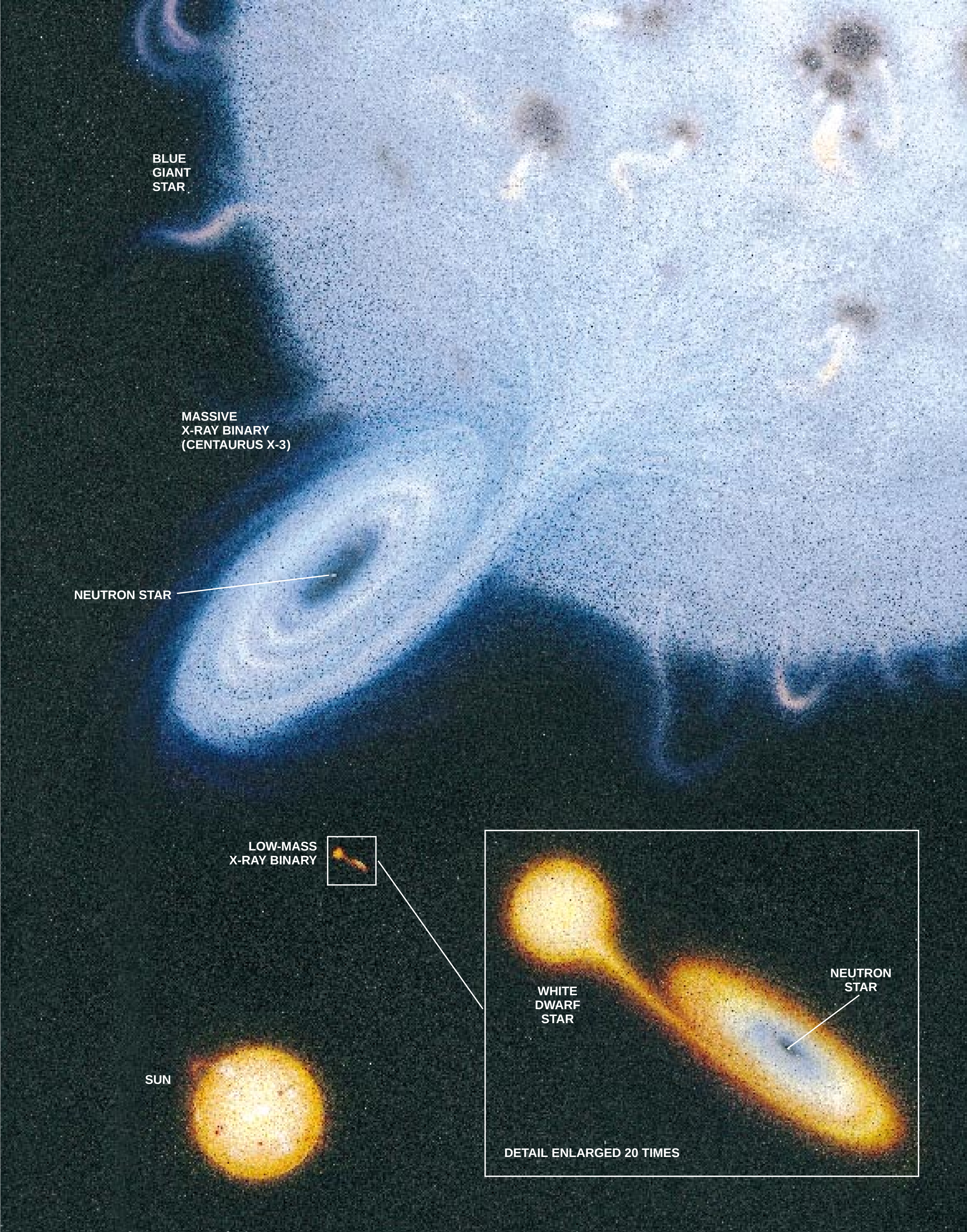
In 1962 Riccardo Giacconi, now the director of the European Southern Observatory, and his associates at American Science and Engineering in Cambridge, Mass., placed an x-ray detector on board a rocket and discovered the first known celestial x-ray source, Scorpius X-1. The name indicates that it is the brightest x-ray-emitting object in the constellation Scorpius. Scorpius X-1 shines about 1,000 times brighter in x-rays than in visible light. The identity of the object emitting this radiation was a total mystery.

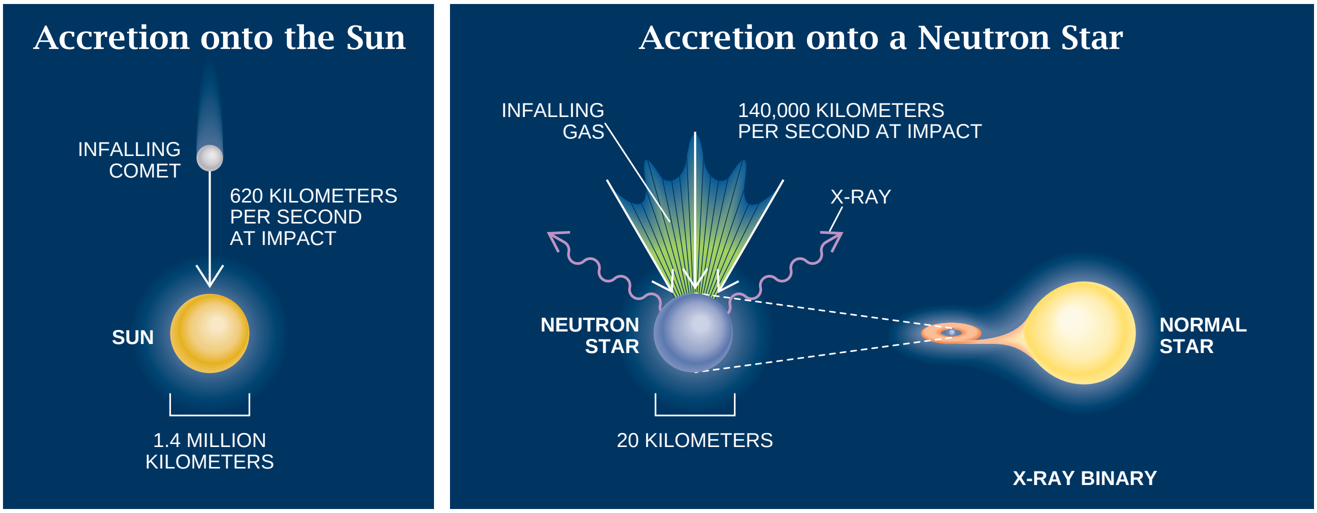
In the following years, x-ray detec-

EDWARD P. J. VAN DEN HEUVEL and JAN VAN PARADIJS have collaborated on the study of celestial x-ray sources since the late 1970s. Van den Heuvel received his Ph.D. in mathematical and physical science from the University of Utrecht in 1968. In 1974 he joined the faculty of the University of Amsterdam, where he is now chairman of the astronomy department. He is also a co-founder and the director of the Center for High-Energy Astrophysics, operated jointly by the University of Amsterdam and the University of Utrecht. Van Paradijs earned his Ph.D. in astronomy from the University of Amsterdam in 1975. He has been a professor of astronomy at the university since 1988.

X-RAY BINARIES comprise two very different classes of double-star systems. In both cases, a neutron star lies at the heart of the x-ray source. Most young x-ray binaries, such as Centaurus X-3 (*top right*), contain a bright blue star having 10 to 40 times the mass of the sun. The youthful neutron star emits pulses of x-rays as it rotates (*below, left*). Low-mass x-ray binaries usually contain far older, sunlike stars. In the tiny low-mass system 4U 1820–30 (*bottom right*), both stars must be compact objects, presumably a neutron star and a larger but less massive white dwarf. Erratic bursts of x-rays occur when gas collects on the surface of an old neutron star and undergoes a thermonuclear detonation (*below, right*).




BLUE
GIANT
STAR
MASSIVE
X-RAY BINARY
(CENTAURUS X-3)
NEUTRON STAR
LOW-MASS
X-RAY BINARY
WHITE
DWARF
STAR
NEUTRON
STAR
DETAIL ENLARGED 20 TIMES
SUN



INFALL OF MATTER, or accretion, can be nature's most efficient mechanism for generating energy. The amount of energy liberated by accretion depends on the gravity at an object's surface. Matter falling onto the sun (*left*) attains only a tiny fraction of the velocity of material accreting onto an ultradense neutron star (*right*). Friction converts kinetic energy into thermal energy; infalling gas in an x-ray binary reaches temperatures of 100 million kelvins, causing it to emit energetic x-rays.

tors placed on rockets and very high altitude balloons revealed a few dozen similar "x-ray stars." Astronomers truly began to understand these objects only after 1970, when the National Aeronautics and Space Administration launched *Uhuru,* the first x-ray satellite, which was designed and built by a team led by Giacconi. Suddenly, astronomers could study the x-ray sky around the clock. Within its first few months of service, *Uhuru* revealed two intriguing x-ray sources, Centaurus X-3 and Hercules X-1. Both objects vary in brightness in a rapid, extremely regular manner: once every 4.84 seconds for Centaurus X-3, once every 1.20 seconds for Hercules X-1. These sources turned out to be the first of a whole class of pulsed x-ray stars.

The pulses provided a critical clue to the nature of these objects. In 1967 Antony Hewish and S. Jocelyn Bell of the University of Cambridge, along with several co-workers, discovered pulsars, a class of stars that emit regular blips of radio emission. After some initial puzzlement, theorists realized that radio pulsars are swiftly spinning neutron stars whose powerful magnetic fields generate a lighthouse beam of radio waves that flashes by the observer once each rotation. The similarly short and constant variations of the newfound x-ray stars hinted that they too were associated with neutron stars.

Another noteworthy trait of Centaurus X-3 and Hercules X-1 is that they experience regular eclipses, in which they dip to a small fraction of their normal brightness. These eclipses proved that the objects must be binary stars, presumably a neutron star orbiting a larger but much more sedate stellar companion that occasionally blocks the neutron star from view. Centaurus X-3 has an orbital period of 2.087 days; for Hercules X-1, the period is 1.70 days.

The pieces of the puzzle began to fall into place. The short orbital periods of the pulsating x-ray stars demonstrated that the two stars sit very close to each other. In such proximate quarters the neutron star can steal gas from its companion; the gas settles into a so-called accretion disk around the neutron star. The inner parts of the disk greatly surpass the white-hot temperatures on the surface of the sun (about 6,000 kelvins). As a result, the accretion disk shines mostly in the form of x-rays, radiation thousands of times as energetic as is visible light. Accretion is so efficient that some x-ray binaries emit more than 10,000 times as much energy in x-rays than the sun radiates at all wavelengths.

The x-ray pulsations occur because the neutron star has a strong magnetic field whose axis is inclined with respect to its axis of rotation. Close to the neutron star, the magnetic field directs the infalling, electrically charged gas toward the star's magnetic poles. There the gas crashes onto the surface, giving rise to two columns of hot (100 million kelvins), x-ray-emitting material. As the star rotates, these columns move in and out of view as seen from the earth, explaining the variation in the star's apparent x-ray flux. Several researchers independently arrived at this explanation of pulsating and eclipsing binary x-ray sources; indeed, by 1972, it had already become accepted as the standard model for such objects.

Careful timing of the pulsations of x-ray binaries showed that they are not perfectly regular. Instead the period of pulsation smoothly increases and decreases over an interval equal to the orbital period. This phenomenon results from the motion of the x-ray source around the center of gravity of the binary-star system. While the source is moving toward the earth, each pulse travels a shorter distance than the one before and so arrives a minuscule fraction of a second early; while the source is moving away from the earth, each pulse arrives a similar amount late.

The amplitude of this effect reveals the velocity at which the source moves along the line of sight to the earth. Centaurus X-3 swings back and forth at 415 kilometers a second. That velocity implies that the companion star has at least 15 times the mass of the sun, typical of a brilliant, short-lived blue star. Since the early 1970s, astronomers have uncovered about 30 pulsating x-ray binaries. In nearly all cases, the companion stars are luminous blue stars having masses between 10 and 40 times that of the sun.

The bright stars in x-ray binaries show periodic changes in the frequency of dark lines, or absorption lines, in their spectra. These changes, known as Doppler shifts, result from the orbital motion of the visible star around the x-ray source. Radiation from an approaching object appears compressed, or bluer; likewise, radiation from a receding object looks stretched, or redder. The degree of the Doppler shift indicates the star's rate of motion. Because the corresponding velocity of the x-ray source can be deduced from the variations of the pulse period, one can use Newton's law of gravity to derive the mass of the embedded neutron star.

The measured neutron star masses fall primarily between 1.2 and 1.6 times the mass of the sun, in good agreement with theoretical expectations. Researchers have found, much to their excitement, that several nonpulsating x-ray binaries seem to contain stars having more than about three solar masses. Current theory holds that neutron stars exceeding that mass limit will produce a gravitational field so intense that it collapses without limit. The result is one of nature's most intriguing objects: a black hole, an object whose gravity has cut it off from the rest of the universe.

As astronomers have found more x-ray binaries, they have come to recognize the existence of two distinct populations: those containing large and luminous blue stars and those containing much older, less massive stars more akin to the sun. The x-ray binaries that include massive blue stars must be very youthful. A star more than 15 times as massive as the sun squanders its supply of hydrogen fuel in less than 10 million years, a blink of the eye compared with the 15-billion-year age of the Milky Way. Hence, the double-star systems from which these x-ray binaries evolved must have been born only a few million years ago in interstellar gas clouds. Like these clouds and other young, hot stars, pulsating massive x-ray binaries tend to concentrate in the plane of the Milky Way, but not toward the galactic center.

About half of the strong x-ray sources in our galaxy, including Scorpius X-1, belong to a very different stellar population. These x-ray binaries concentrate predominantly in the central lens-shaped bulge of the galaxy and in globular clusters, dense spherical swarms of stars. Such regions harbor mostly older stars, those having ages between about five and 15 billion years.

In general, these elderly x-ray binaries do not undergo regular pulsations. They differ from the massive, pulsing x-ray binaries in other ways as well. The visible-light spectra of the aged x-ray binaries appear utterly unlike those of normal stars. Instead they grow steadily brighter toward the blue end of the spectrum; some of their radiation emerges at distinct wavelengths, or colors. Theoretical models indicate that such a spectrum would be produced by an inflowing disk of gas heated by intense x-rays streaming from inner parts of the disk, just above the neutron star's surface.

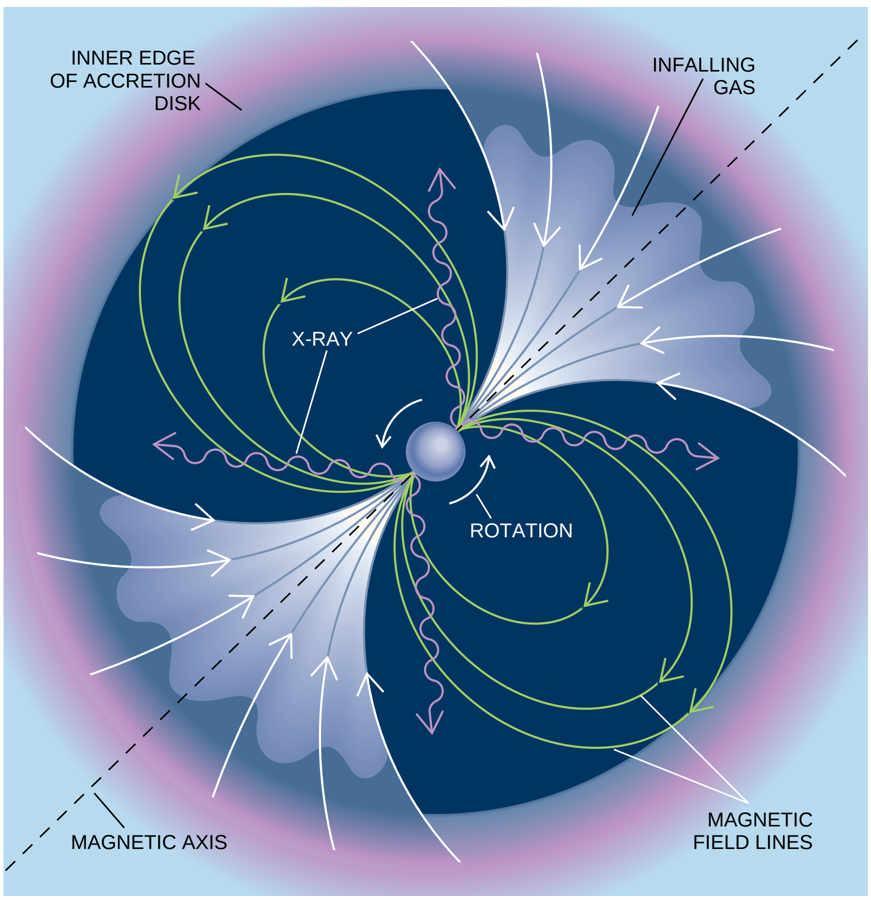
Emission from the disk almost completely drowns out the light from the companion star. That disparity implies that the companion must be fairly faint, which in turn indicates that its mass is no greater than that of the sun. These double-star systems are therefore known as low-mass x-ray binaries. Solar-mass stars remain stable for at least 10 billion years, consistent with the age of the stellar population in which low-mass x-ray binaries reside.

Low-mass x-ray sources undergo occasional extreme flare-ups, or x-ray bursts, which have yielded a great deal of information about these systems. Within a few seconds of the beginning of a burst, the object's x-ray brightness increases by a factor of 10 or more, peaks for a few seconds to a few minutes and then decays to the original level in about a minute. X-ray bursts recur irregularly every few hours or so.

Researchers have deduced that the x-ray bursts result from runaway nuclear fusion reactions in the gas accreted onto the surface of a neutron star. Between bursts, new matter flowing from the companion star replenishes the nuclear fuel. That steady accretion gives rise to the persistent emission of x-rays seen between the bursts. Despite the spectacular nature of the bursts, low-mass x-ray binaries emit more than 90 percent of their total energy during times of quiescence—a testimony to the great efficiency of accretion compared with fusion.

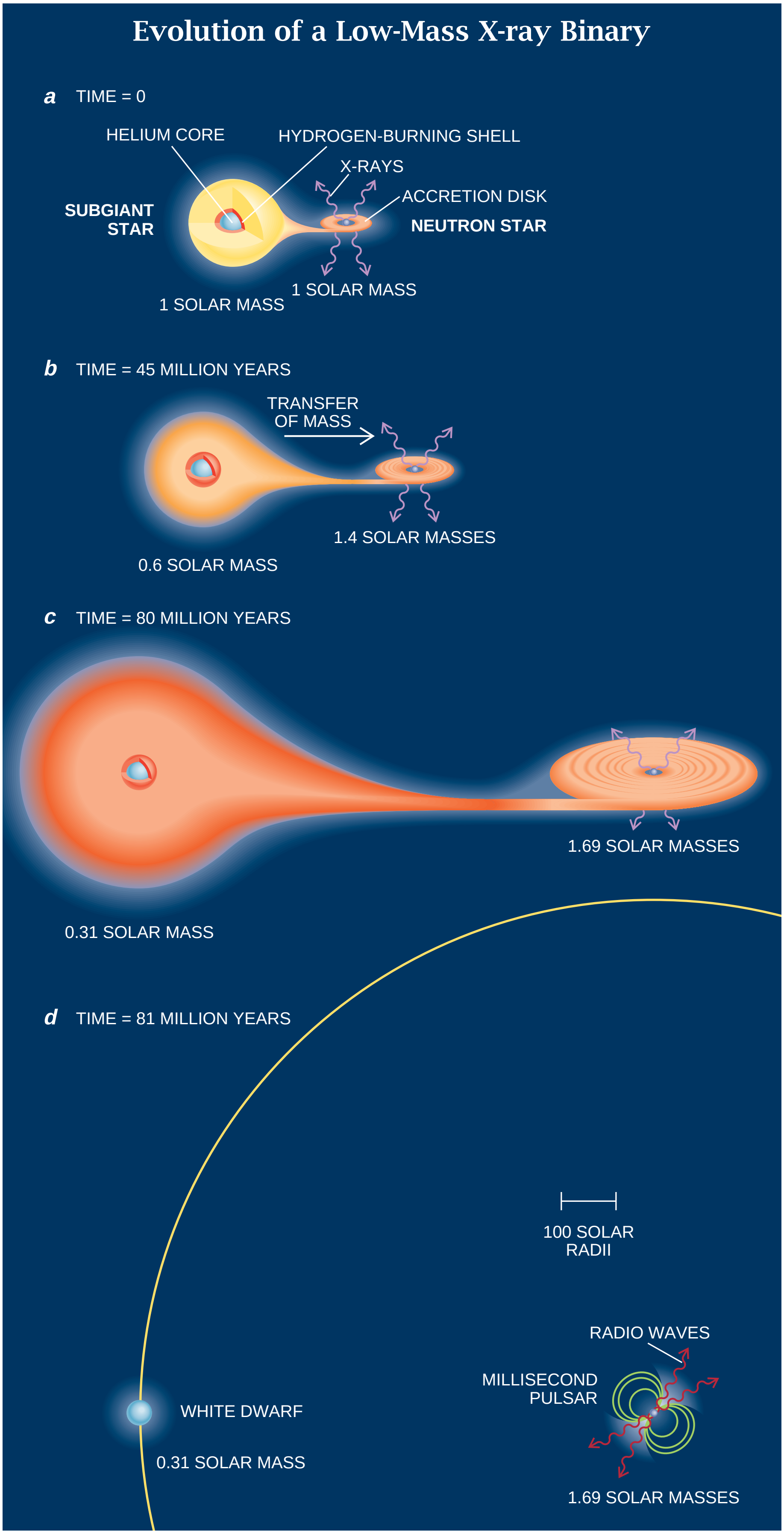
X-ray bursts occur only in low-mass binary-star systems and x-ray pulses almost solely in high-mass ones; not a single system displays both forms of behavior. The critical factor responsible for this disparity is probably the strength of the neutron star's magnetic field. High-mass x-ray binaries must contain neutron stars having powerful magnetic fields, capable of generating easily detectable pulsations. Neutron stars in low-mass x-ray binaries seem to possess far weaker fields. This explanation is bolstered by theoretical models indicating that a powerful magnetic field would inhibit the nuclear instabilities that produce x-ray bursts.

The disparate characteristics of low-mass and high-mass x-ray binaries un-



MAGNETIC FIELD of a young neutron star prevents infalling gas from reaching the star's surface, except at the two magnetic poles. Two hot, x-ray-emitting columns of gas, each about a kilometer across, collect at the neutron star's poles. The star's rotation axis is inclined with respect to its magnetic axis, so an observer perceives regular pulses of x-rays as the magnetic poles rotate in and out of view.

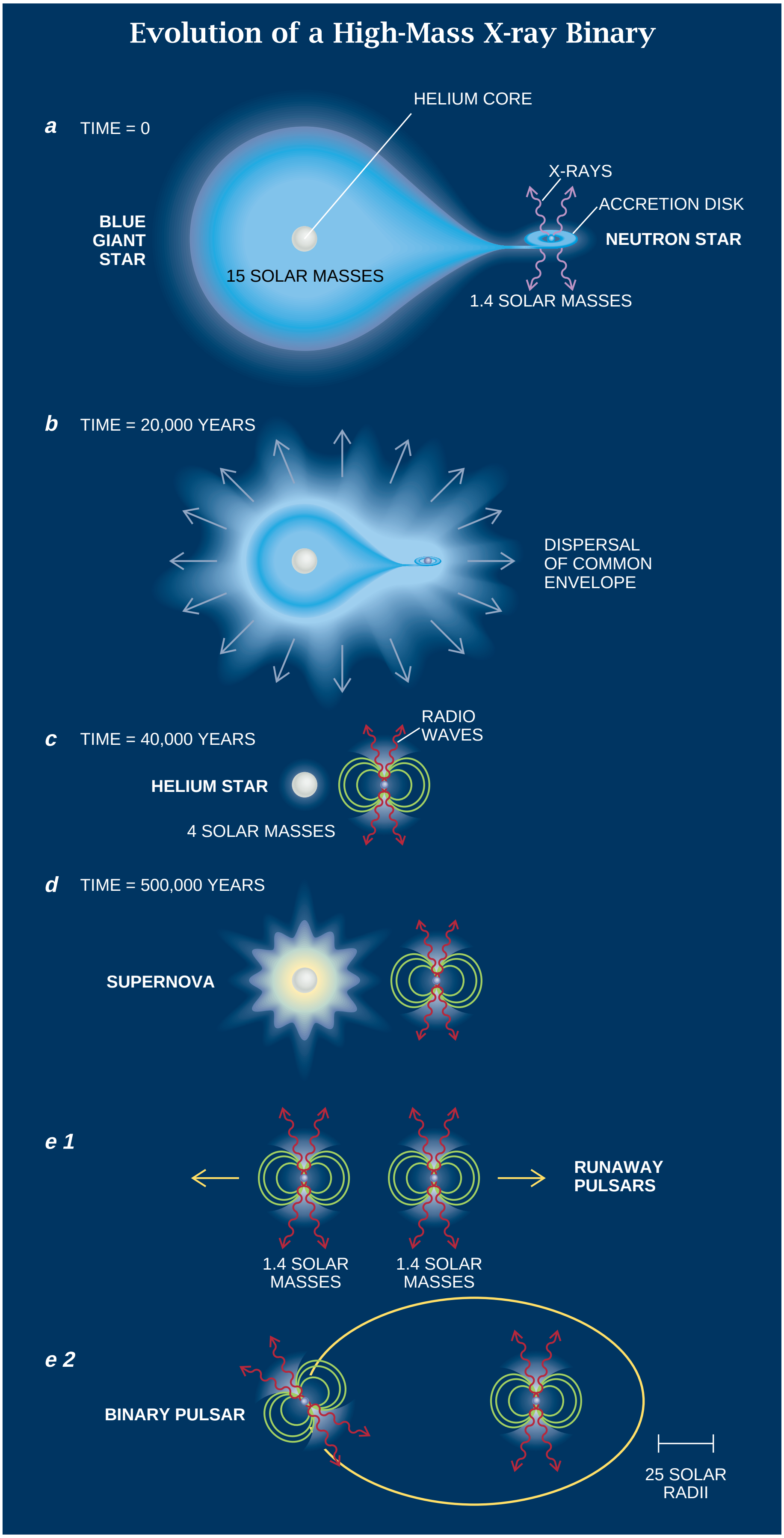
LOW-MASS X-RAY BINARY initially consists of a neutron star pulling material from its companion (*a*). The low-mass star is an elderly subgiant having a dense, inert helium core. The transfer of mass causes the stars' orbit to widen. At the same time, the low-mass star steadily expands and cools as it evolves (*b*). The neutron star gradually consumes the subgiant's outer envelope (*c*). The exposed helium core (now considered a white dwarf) remains in a circular orbit around the neutron star (*d*). The rotating neutron star is now a millisecond pulsar that emits pulses of radio waves but no x-rays. (This scenario is based on calculations by Paul C. Joss and Saul A. Rappaport of M.I.T.)



derscore the very different ways in which these systems must have formed and evolved. Almost immediately after the discovery of high-mass x-ray binaries in 1971, workers recognized that such objects represent a normal stage in the evolution of close double-star systems in which both objects have more than a few times the mass of the sun. The more massive star quickly consumes its fuel and expands into a bloated red giant, whose outer layers spill over onto the companion star, exposing the red giant's helium-rich center. A few hundred thousand years later, this helium star explodes as a supernova, shedding much of its outer mass; its remnant core collapses into a neutron star. The neutron star attracts gas from its companion and becomes a source of x-rays.

The formation of a low-mass x-ray binary involves a more specialized set of circumstances. Some of these binaries could have started out as a massive star and a stellar lightweight orbiting each other. The small companion star would have too little gravity to capture material from the primary star, so its mass would not appreciably increase. When the primary annihilates itself as a supernova, much of the system's mass would escape into interstellar space. In most cases, that loss of mass would disrupt the binary and send the two stars sailing off on separate courses. In the rare instance in which the stars remain bound to each other, they could evolve into a low-mass x-ray binary.

There is also a gentler way by which a neutron star may form next to a low-mass star. If the primary star initially has less than eight times the mass of the sun, it will not blow up. Instead it will produce a white dwarf, a stellar cinder far denser than a normal star but much less so than a neutron star. In a white



MASSIVE X-RAY BINARY contains a bright blue star and an accreting neutron star (*a*). The blue star expands until its outer envelope engulfs both its helium-rich core and the neutron star (*b*). The orbital motion of the two stars inside the envelope heats the envelope and blows it away, leaving behind a helium star and a radio pulsar (*c*). If the helium star has more than 2.5 solar masses, it explodes as a supernova (*d*) and forms a second pulsing neutron star. The explosion may disrupt the system (*e1*); otherwise, the result is two neutron stars locked in a rapid, highly eccentric orbit (*e2*). Less massive helium cores do not explode; they end up as white dwarfs in circular orbits about the neutron star.

dwarf the star's gravity has crushed its constituent atoms into a soup of electrons and nuclei; a white dwarf having the mass of the sun would be about the size of the earth.

As the low-mass star evolves, it will gradually expand; if the two stars are in a close orbit, gas from the low-mass star will accrete onto the surface of the white dwarf. The mass of the white dwarf may eventually exceed a critical value, about 1.4 solar masses. At that point, the dwarf star collapses into a neutron star. This kind of quiet collapse ejects very little material, so the system can remain tightly bound. Later, the stars spiral in closer toward each other, accretion begins and the system becomes a low-mass x-ray binary.

In such binaries, the neutron star's gravity exerts a strong pull on its much larger but less massive companion. The combination of gravitational and centrifugal forces gives rise to a pear-shaped region of stability, called a Roche lobe, surrounding the low-mass star. Any material lying outside the Roche lobe will flow toward the neutron star. The transfer of material causes the distance between the two stars to increase if the mass-losing star is the less massive of the two, as is the case in low-mass x-ray binaries. When the size of the orbit increases, so does the size of the Roche lobe. Once the lobe grows bigger than the companion star, the flow of matter ceases and the neutron star stops emitting x-rays. Evidently, some mechanism acts to keep feeding gas to the neutron star.

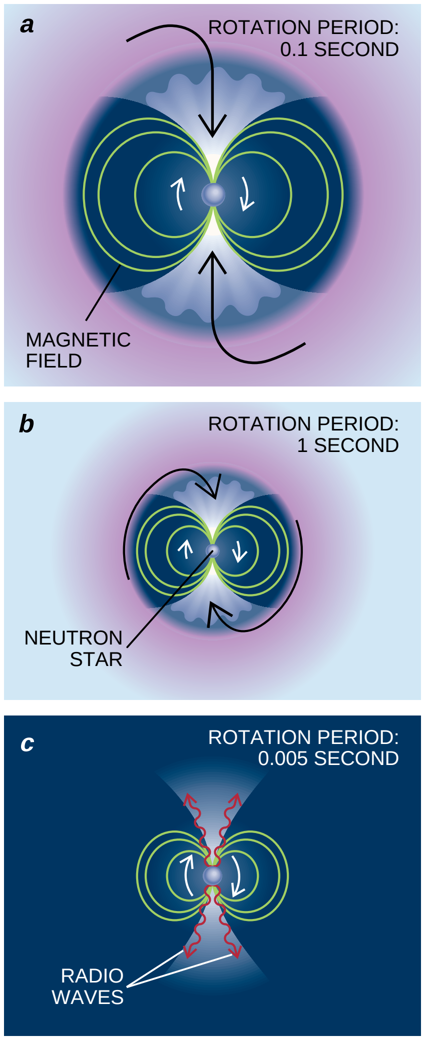
In one class of low-mass x-ray binaries—tightly bound systems whose periods are less than about 10 hours—the flow of gas is maintained by a steady shrinking of the stars' mutual orbit. As the stars orbit, they shed gravitational waves that carry off angular momentum, which causes the stars to draw closer together. That effect negates the tendency of mass transfer to move the stars apart. The stars ultimately settle into a slowly shrinking orbit in which a steady trickle of gas migrates from the companion to the neutron star. In this way, the neutron star accretes about one thousandth of an earth mass each year, sufficient to account for the observed luminosity of many low-mass x-ray binaries (about 3×10^{30} watts).

The brightest x-ray sources in the central regions of the galaxy emit about 10 times that much energy. These bright objects constitute a second class of low-mass x-ray binaries that have relatively long orbital periods of about one to 10 days. Such leisurely orbits imply that the separation between the two stars, as well as the diameter of the normal companion, must be quite large. Here the flow of matter must result from the physical swelling of the companion star as a consequence of physical changes in its interior.

Such changes occur in the later evolutionary stages of a sunlike star. Hydrogen fusion produces helium, which accumulates as a dense core; hydrogen fusion takes place in a shell around this core. As the star ages, the hydrogen-burning shell migrates outward, causing the star's outer envelope to expand and cool. That expansion more than compensates for the increasing distance between the stars caused by the transfer of angular momentum. X-ray binaries having a period of about five to 10 days reach equilibrium if they experience a mass transfer rate of about five thousandths of an earth mass per year, about the rate required to power the bright sources around the galactic center.

In 1982 a group of researchers—Ronald F. Webbink of the University of Illinois, Saul A. Rappaport of the Massachusetts Institute of Technology, G. J. Savonije of the University of Amsterdam and Ronald E. Taam of Northwestern University—investigated the fate of these low-mass x-ray binaries. Their calculations predict that, regardless of their initial traits, these systems always reach the same evolutionary end point. The giant star soon loses its entire hydrogen-rich envelope; its naked helium core remains as a white dwarf containing between 0.25 and 0.45 solar mass. The stars' final orbit is extremely circular because of the tens of millions of years of tidal interaction between the neutron star and its low-mass partner.

After the supply of accreting material dries up, binary-star systems no longer emit detectable amounts of x-rays. The last evolutionary stages of x-ray binaries nonetheless offer a fascinating glimpse at what happens to very old neutron stars. During these later phases, the neutron star's most distinctive emission is in the form of radio waves, not x-rays.



ROTATION OF NEUTRON STAR is strongly influenced by accretion in an x-ray binary. The neutron star's magnetic field defines the inner edge of the surrounding disk of matter, where gas falls onto the star. When the star is young, its field is strong, so the inner edge of the disk is distant and comparatively slow-moving (*a*). As the magnetic field decays, the inner edge of the accretion disk moves inward (*b*). The star now accretes rapidly moving material that causes its rate of rotation to increase. By the time the accretion ceases, the neutron star may be rotating hundreds of times per second (*c*).

In 1983, while working on the 300-meter radio telescope in Arecibo, Puerto Rico, Valentin Boriakoff, then at Cornell University, Rossolino Buccheri of the Italian National Research Council in Palermo and Franco Fauci of the University of Palmero discovered the binary radio pulsar PSR 1953+29. Its properties closely resemble those of the extinct x-ray binaries modeled by Webbink and his colleagues. The pulsar's radio signals displayed no signs of the eclipses or absorption produced by normal stars. The researchers recognized that the pulsar's companion must itself be a compact object. Because of its low mass, it is probably a white dwarf.

One of the most surprising aspects of PSR 1953+29 is its period of radio pulsation: a remarkably swift 6.1 milliseconds, or 160 rotations a second. At about the same time as the discovery of PSR 1953+29, Donald C. Backer of the University of California at Berkeley and his co-workers found another pulsar, PSR 1937+21, which has a period of only 1.6 milliseconds. Astronomers now recognize these objects as the prototypes of a category of rapidly spinning neutron stars known as millisecond pulsars.

The inferred history of x-ray binaries made it clear why these pulsars spin so quickly. In low-mass x-ray binaries (and in many massive x-ray binaries as well), orbital motion prevents matter from falling directly onto the neutron star. Instead it goes into orbit about the star, forming an accretion disk. Material from the disk's inner edge falls onto the neutron star. During the later stages of accretion, that infalling material greatly speeds up the star's rotation.

Nearly all binary radio pulsars possess companions that have evolved into white dwarfs or neutron stars. At some stage, those companions were giants that overflowed their Roche lobes and dumped matter onto the neutron stars, increasing the stars' rate of rotation. During that time, the double stars would have appeared as x-ray binaries. After the companion star lost its outer layers and the accretion process ceased, a naked millisecond pulsar remained.

The power of a pulsar's radio emission varies in proportion to the fourth power of the rate of rotation. Millisecond neutron stars can be detected only because they were "spun up" by their companions during the x-ray binary phase. Radio pulsars that acquired their rapid rotation in this way are now called recycled pulsars, a term suggested by V. Radhakrishnan of the Raman Research Institute in Bangalore, India.

Starting in 1987, a number of observers—in particular, groups led by Andrew G. Lyne of the University of Manchester, Shrinivas R. Kulkarni of the California Institute of Technology and Alexander Wolszczan of Pennsylvania State University—have found that globular clusters are incredibly rich hunting grounds for binary and millisecond pulsars. Studies of globular clusters have already revealed 32 radio pulsars; 70 percent of these pulsars rotate in less than 10 milliseconds, indicating that they are recycled. This celestial bounty results from the extremely dense nature of globular clusters. In their central regions, these clusters may contain more than 10,000 stars per cubic light-year, a million times the density of stars in the sun's corner of the galaxy. Under such crowded conditions, neutron stars face good odds of passing close to and capturing a stellar companion. Globular clusters harbor 200 to 1,000 times as many x-ray binaries per million stars as does the galaxy as a whole.

In addition to the binary pulsars discussed thus far, astronomers have identified another rarer class of such objects that have substantially different characteristics. Their orbits often are extremely eccentric and their companions contain 0.8 to 1.4 times the mass of the sun. These objects probably arose from high-mass x-ray binaries in the following way.

In massive x-ray binaries, accretion causes the two stars to spiral ever closer together (the opposite of the situation for low-mass x-ray binaries). That process, combined with the swelling of the companion star as it evolves, causes the companion to overflow its Roche lobe completely, engulfing the neutron star. Frictional drag quickly sends the neutron star spiraling in toward its companion. At a certain point, the friction generates so much heat that it drives off the gaseous hydrogen envelope. What remains is the neutron star in a close orbit around the stripped core of the companion, which consists of helium and heavier elements.

If the heavy-element core is sufficiently massive, it will later detonate into a supernova and produce a second neutron star. The force of the explosion and the precipitous loss of mass cause the stars' orbit to become highly elliptical; in many cases, the stars break free entirely to become runaway radio pulsars. If the orbit survives, the neutron stars follow their eccentric courses almost forever; over the ages, their orbits will slowly narrow because of the emission of gravitational waves. One of the most thoroughly studied binary pulsars, PSR 1913+16, consists of two neutron stars that race through a highly elliptical orbit once every seven hours and 45 minutes. This system's extreme properties allow it to serve as a sensitive test-bed for many aspects of Einstein's theory of relativity, as Joseph Taylor of Princeton University has beautifully demonstrated.

Recent studies of binary pulsars have overturned a long-standing idea about how neutron stars change over the eons. Based on statistical analyses of pulsars, most astronomers concluded that a neutron star's magnetic field decays without any outside assistance and in due time vanishes completely. The existence of recycled pulsars proves, however, that some magnetic field persists even in extremely old systems. Moreover, the companion stars in binary pulsars offer a way to determine just how old those stars are.

Three of the millisecond pulsars have observable white dwarf companions, which serve as natural chronometers. A white dwarf steadily radiates away the heat left behind from its days as the core of a red giant star. Over the eons, white dwarfs grow progressively cooler and redder; the color of a white dwarf therefore betrays its age.

In 1986 Kulkarni measured the color of the white dwarf companion to PSR 0655+64 and concluded it must be at least 500 million years old. Using similar reasoning, three sets of researchers—J. F. Bell of the Mount Stromlo and Siding Spring Observatories and his colleagues; John Danziger of the European Southern Observatory and his co-workers; and Charles D. Bailyn of Yale University—determined that the white dwarf in the binary pulsar system PSR J0437-4715 is about two billion years old. The pulsars in these systems must be considerably older because they would have formed long before their companions evolved into white dwarfs, yet they retain substantial magnetic fields. Otherwise, they could not be detected.

Recent work by Frank Verbunt, Ralph A.M.J. Wijers and Hugo Burm of the Center for High-Energy Astrophysics in the Netherlands further demonstrates the persistence of neutron-star magnetic fields. The researchers studied three anomalous, low-mass x-ray binaries that also are x-ray pulsars, indicating that they each contain a strongly magnetized neutron star. No matter how a neutron star originates, it always loses at least a few tenths of a solar mass in the form of neutrinos. When this happens, the binary system widens, shutting off the flow of gas. Accretion cannot occur until the binary system has shrunk through the emission of gravitational radiation or until the companion star has begun to evolve into a giant.

Both these mechanisms take considerable time to take effect. This knowledge allowed Verbunt and his co-workers to set lower limits to the ages of the accreting neutron stars in low-mass x-ray binary pulsars. In the case of Hercules X-1, they found that the strongly magnetized neutron star is at least 500 million years old. Neutron stars' magnetic fields evidently do not spontaneously decay, at least not on such time scales.

And yet the magnetic fields of almost all binary radio pulsars are 100 to 10,000 times weaker than those of normal, youthful radio pulsars, regardless of whether the binary pulsar descended from a high-mass or a low-mass x-ray binary. The weakness of their fields seems to be attributable to some factor that all binary pulsars have in common. The most obvious common factor is accretion. In 1986 Taam and one of us (van den Heuvel) proposed, on observational grounds, a link between field decay and the accretion process. Theorists have recently advanced several models to explain the details of this relation.

One model holds that newly accreted layers on the surface of a neutron star form an electrically conductive layer that allows only a small fraction of the star's magnetic field to reach the outside. Another possibility, recently proposed by G. Shrinivasa of the Raman Research Institute, is that it is the gradual slowing of a neutron star's rotation that causes its magnetic field to dissipate. Such deceleration occurs before and during the early stages of accretion. Once the magnetic field has weakened below a critical threshhold, the action of accretion reverses the spin-down trend, but that infusion of rotational energy cannot restore the magnetic field to its original strength.

In any case, there is every indication that millisecond radio pulsars will retain their fields and continue to pulse untold billions of years into the future. Thus it happens that long after their era of flamboyant x-ray emission, x-ray binaries settle down to become some of the most steady, unchanging entities in the cosmos.

FURTHER READING

THE X-RAY UNIVERSE. Edited by Wallace Tucker and Riccardo Giacconi. Harvard University Press, 1985.

A NEW CLASS OF PULSARS. Donald C. Backer and Shrinivas R. Kulkarni in *Physics Today*, Vol. 43, No. 3, pages 26–35; March 1990.

PULSARS TODAY. Francis Graham Smith in *Sky and Telescope*, Vol. 80, No. 3, pages 240–244; September 1990.

X-RAY BINARIES AND RECYCLED PULSARS. Edited by E.P.J. van den Heuvel and S. A. Rappaport. Kluwer Academic Publishers, 1992.

The Art of Boris Artzybasheff

A compelling mid-20th century vision of the machines of war and peace

by Domenic J. Iacono

A generation of magazine readers during and immediately following World War II fell under the spell of the anthropomorphic machines created by Boris Artzybasheff. Artzybasheff liked machines. He admired their power to take on the labor men and women had toiled under for centuries. "I would rather watch," he explained, "a thousand-ton dredge dig a canal than see it done by a thousand spent slaves lashed into submission." But he was troubled by the other side of humanity's affair with machines—the destruction it worked with its inventions.

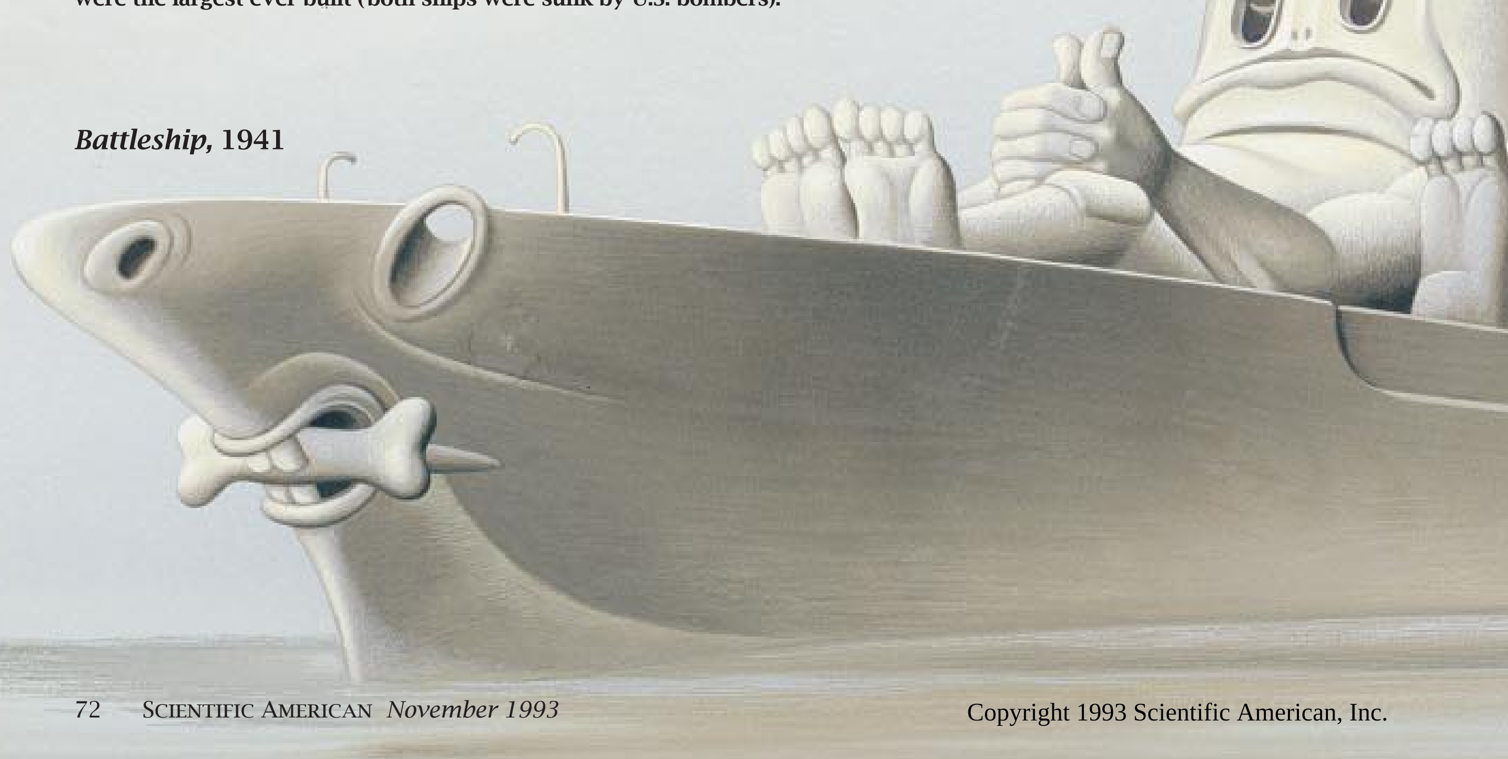
Born in 1899 in Kharkov, Ukraine, Artzybasheff came of age during the Russian Revolution. In 1919, after five months of fighting the Communists with the Ukrainian army, he escaped to the Black Sea, where he boarded a ship headed for

Continued on page 76

JAPANESE BATTLESHIP plows ahead as an admiral stands amidships, representing the vessel's bridge. A seated figure in a top hat, easily identified by the American public as Emperor Hirohito, holds an antenna. Published shortly before the attack on Pearl Harbor (in the November 3, 1941, issue of *Life*), Artzybasheff's *Battleship* presents the Japanese as a less than formidable enemy. The sharklike bow of the ship with a bone clenched in its teeth appears more like a canine retrieval stunt than a serious threat. The front gun turret with its twiddling thumbs, the fishing-pole crane and the birdlike plane all work to portray a humorous rather than sinister interpretation. Considered by some to be a racist portrayal, this image neatly fit the propagandistic treatment of Axis forces in the American media.

Before the first naval encounters of World War II, military establishments supported the concept of heavily armed and armored warships. The Japanese battleships *Yamato* and *Musashi*, with 18-inch guns and nearly 70,000-ton displacements, were the largest ever built (both ships were sunk by U.S. bombers).

***Battleship,* 1941**



Jet Propulsion Buzz Bomb, 1944

BUZZ BOMB, or V-1, was a small, pilotless airplane that used a jet engine to propel its payload of explosives. The Germans launched more than 18,000 of these drones between June 1944 and March 1945. The weapon, intended as a reprisal for the Allied air attacks against German cities, was aimed at England and at Allied positions on the continent. The American public, never the target of a missile attack, could only imagine the fear and consternation engendered by the approach of these strange whistling bombs.

The buzz bomb as conceived by Artzybasheff possessed two distinguishing features: a shrill whine and long contrails that marked its flight path. The motor rides atop the bomb like a humanized jet engine hell-bent on destruction. Its signature high-pitched whine is clearly suggested by the open-mouthed air inlet, while the narrow stream of exhaust is enclosed within the airy veils of contrails that are stirred by high-altitude winds. The painting was published with a collection of Artzybasheff's other anthropomorphic war machines in his book *As I See* (Dodd, Mead & Company, 1954).

Tank, 1941

Barrage Balloons, 1941

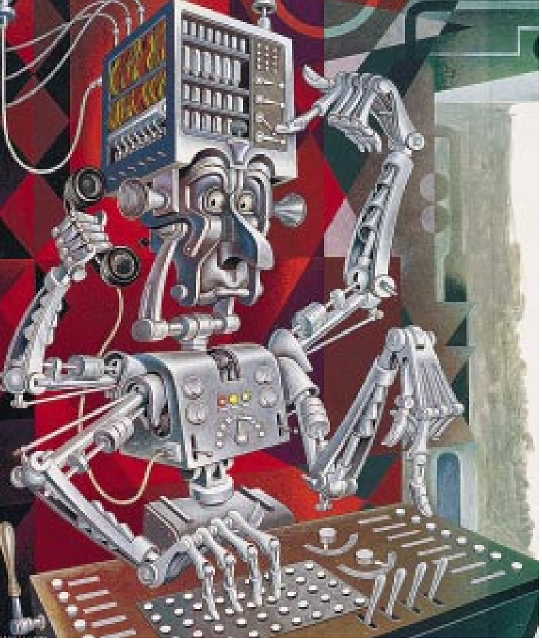
BARRAGE BALLOONS were developed as part of defensive systems intended to protect key military and government installations and densely populated areas. Captive and nonpowered, the balloons were tethered to the ground by thin, multiple wires contrived to deter low-flying airplanes. Artzybasheff depicted these lighter-than-air aircraft as floating figures whose arms and legs serve as fins and stabilizers; the outlet valves of the balloons appear as mouths gasping for air. In addition to the humanizing forms given these defensive weapons, Artzybasheff has also created an anthropomorphic machine with a psychological aspect. Barrage balloons were sacrificial safeguards: when successful in their mission, they were often destroyed. The facial features, pudgy bodies and almost melancholy postures evoke an air of resignation to their assigned task.

THE TANK entered the military's inventory during World War I, but it was in World War II that it became a dreaded weapon. In September 1939 the Germans defeated Poland with their blitzkrieg of tanks and dive-bombers, and in 1940 Hitler's heavily armored panzer divisions overran Belgium, the Netherlands and France. The speed, agility and destructive capability of the tank became dramatically clear when the Germans quickly defeated the numerically superior French and British forces. When it first appeared in *Life* magazine, Artzybasheff's *Tank* was not specifically identified as an enemy weapon, but the advance of German tank columns in Europe and Africa dominated the news. And the high-stepping legs (treads) reminded readers of the goose-step marching of German troops, while the cowering antitank guns were interpreted as the ineffective French forces stationed beyond the Maginot Line.

DOMENIC J. IACONO is assistant director and curator of collections of the University Art Collection at Syracuse University. A specialist in art created in America during the Great Depression, he holds a master of arts degree from Florida State University. In 1992 he organized an exhibition of Boris Artzybasheff's work, "War News and Strange Brews."

PROPAGANDA was used as a tactical device by both Axis and Allied countries during World War II. The Nazi Ministry of Propaganda targeted at Allied troops broadcasts by English-speaking announcers playing American music. In the U.S., the Office of War Information concentrated on developing leaflets that were distributed in German-occupied lands to counter the Nazi propaganda. Artzybasheff's portrayal of propaganda as a two-headed monster with loudspeakers blaring false reports and slander from fingertip microphones was targeted at European enemies. The figure's arms are raised in the Nazi, Fascist and Communist salutes; its legs, protected by insulators, draw power from the electric lines on which it stands.

Radio Propaganda, 1941



Continued from page 72

New York. He spent his 20th birthday on Ellis Island. The young artist soon found employment in New York City as an engraver, fashioning labels for beer and medicine bottles, and began to establish a reputation for creative design. Some of his early commissions included stage sets for the Ziegfeld Theater and Michel Fokine's Russian Ballet and a mural for a 57th Street speakeasy. Eventually he turned his attention to illustrating books, for which he won many awards, including the John Newbery Award from the American Library Association.

In 1940 the editors of *Fortune* asked Artzybasheff to develop a cover for the magazine. It was this painting of a Japanese soldier standing before a large sculpted head of the Buddha that attracted the attention of the editors at *Time* magazine. They were assembling a cadre of illustrators to originate covers portraying a different headliner each week. During the next 25 years, until his death in 1965, Artzybasheff conjured up more than 200 covers for *Time;* they included portraits of Stalin, King George VI, Hitler, MacArthur, Truman, Mao Tse Tung and Ho Chi Minh.

Perhaps the most compelling forms of Artzybasheff's art, however, are his paintings and drawings of humanized machines and mechanized humans. These pictures, which often border on the surreal, display a keen sense of how

Executive of the Future, 1952

ROBOTIC EXECUTIVE was one of the artist's first published forays into the white-collar domain, where the computer was emerging in the early 1950s as the office machine of the future. The executive, which appeared in a 1952 issue of *Esquire,* had four arms to handle the common work load of a typewriter keyboard, telephone, push-button and toggle-switch operations, and changes in mental processes. Pressure gauges and warning lights control its cold calculations.

Telephone Executive, 1952

CANTANKEROUS BUSINESSMAN, accustomed to command, has been fashioned from the receiver of a rotary telephone. Dressed in a three-piece suit, he lounges back against the dial and barks instructions to a subordinate, emphasizing his point with an index finger. Ashes from a cigar float unnoticed onto the mouthpiece. The implication is of a rude, complacent and inattentive (he is depressing the switch hook with his right arm) executive who is talking merely to hear himself talk. The drawing appeared in Artzybasheff's book *As I See.*

the machine works and what human task it was meant to replace. The images of animated weapons of war and tyranny that he made for *Life* magazine showed how men can create monsters that personify the human capacity for destruction. When asked about his ideas on war and weaponry, Artzybasheff replied, "I try to shake this thought off: It may be that a healthy planet should have no more life upon it than a well-kept dog has fleas; but what possesses the flea to concoct its own flea-powder?"

Artzybasheff's paintings display an attention to detail and process that was similar to an engineer's approach to devising a machined component. He kept extensive notes on his ideas, techniques and formulas for mixing the opaque watercolors, called gouache, that he used. He even notated his preparatory sketches, constantly making revisions and adaptations. Often Artzybasheff would build his painting from a base of detailed drawings (he called them skeletons) that were then layered with skins of color to develop the final execution of the painting. Components were frequently made up separately and then brought to the composition in their finished form. The result was a picture that embodied the imaginative vision of an artist who was justifiably labeled the master of the machine age.

***FOREIGN CORRESPONDENT,* originally painted as a cover illustration for *Dateline* magazine in 1964, reflected the importance of the media in the cold war period of the 1950s and 1960s. Composed of a camera attached to the wings of an aircraft flying over the major capital cities of the world, Artzybasheff's image of a correspondent depicts the media's ability to convey information from far-reaching sources in a timely and objective manner. The pencil and notepad suspended on the camera's strap alluded to the personal or subjective nature of the correspondent's approach to news events and issues. The subtle rendering of martini and highball glasses covering the globe was Artzybasheff's humorous suggestion of the way in which reporters often get their leads, stories and the all-important exclusive.**

***Foreign Correspondent,* 1964**

High-Power Electronics

A new generation of silicon switches enables power grids to meet the needs of utility customers with high efficiency and reliability

by Narain G. Hingorani and Karl E. Stahlkopf

Consumers of electricity are demanding customers. The silicon chips that now pervade daily life, bringing sophisticated behavior to everything from toasters to machine tools, are highly vulnerable to irregularities in their electrical diet. A loss of power for a single cycle of alternating current, one sixtieth of a second, can make computer screens go blank or interrupt other sensitive electronic equipment. At the same time that users of electric power demand quality, they also want more power. As a result, transmission networks are being pressed closer to their operating limits. Yet a range of problems hobbles expansion, and power transfers from one part of the country to another challenge the network's adequacy. All these factors increase the risk of instability and even blackout.

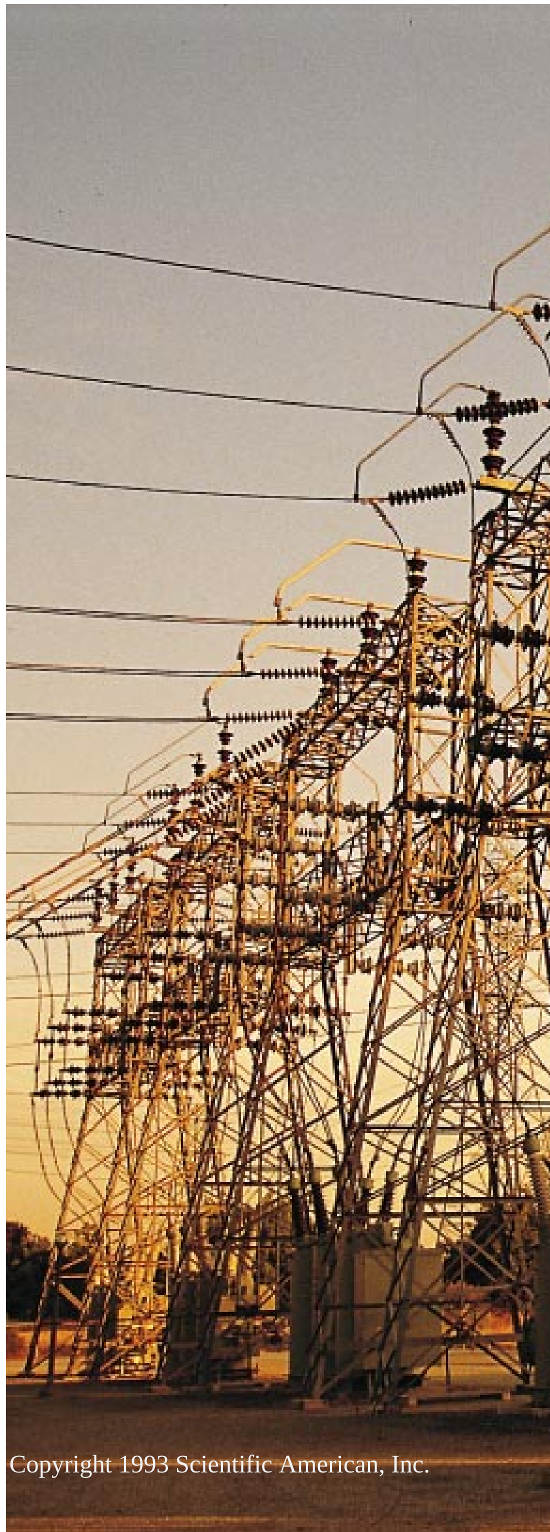
To avoid such problems, the engineers who manage the transmission of electric power must act with extreme caution. They operate the power grid well below its theoretical maximum capacity. The strategy reduces the possibility that a sudden, unforeseen increase in demand or loss of capacity might cause overloads that would ripple outward until they engulfed a significant portion of the nation's transmission network. On the infamous night of November 9, 1965, for example, a blackout struck most of the northeastern U.S. and parts of Canada. Utilities must also maintain large reserves of generating capacity as a safety margin against such contingencies.

The traditional equipment does not make the job any easier. When engineers switch in additional generating equipment or otherwise cope with failures or changing demand, they rely on massive electromechanical switches—essentially enormous cousins of the household circuit breaker—that take several AC cycles to turn on or off. These switches introduce their own electrical noise and potential instabilities into the system. They cannot be used to make the continuous, fine adjustments that full utilization of power transmission capacity would require.

The development of high-voltage silicon switches may provide utilities with a technology that enables them to cope effectively with economic constraints as they meet the needs of their customers. These switches are the basis of control systems that can guide the flow of megawatts as rapidly and efficiently as integrated circuits handle microwatts. They can fend off cascading power interruptions and significantly increase the usable capacity of many transmission lines. Indeed, they open up avenues for controlling the distribution of power that could not be exploited by their electromechanical predecessors. Utilities using them will be able to deliver more power of better quality while reducing transmission losses and thus the amount of energy they must generate.

NARAIN G. HINGORANI and KARL E. STAHLKOPF conduct research on transmission and distribution of electricity at the Electric Power Research Institute in Palo Alto, Calif. Hingorani, EPRI's vice president for electrical systems, has been studying power transmission systems at the institute for nearly 20 years. He received a bachelor's degree in electrical engineering from Baroda University in India and a doctorate from the University of Manchester in England and was elected as a fellow of the National Academy of Engineering in 1988. Stahlkopf is director of EPRI's electrical systems division. He has worked at the institute since 1973. He received bachelor's degrees in electrical engineering and in naval science from the University of Wisconsin and a doctorate in nuclear engineering from the University of California, Berkeley.

The troubles that power transmission systems experience today go back to the roots of the industry and the conflict that raged a century ago between Thomas Edison and George Westinghouse: direct versus alternating current. In DC systems, electrical charge

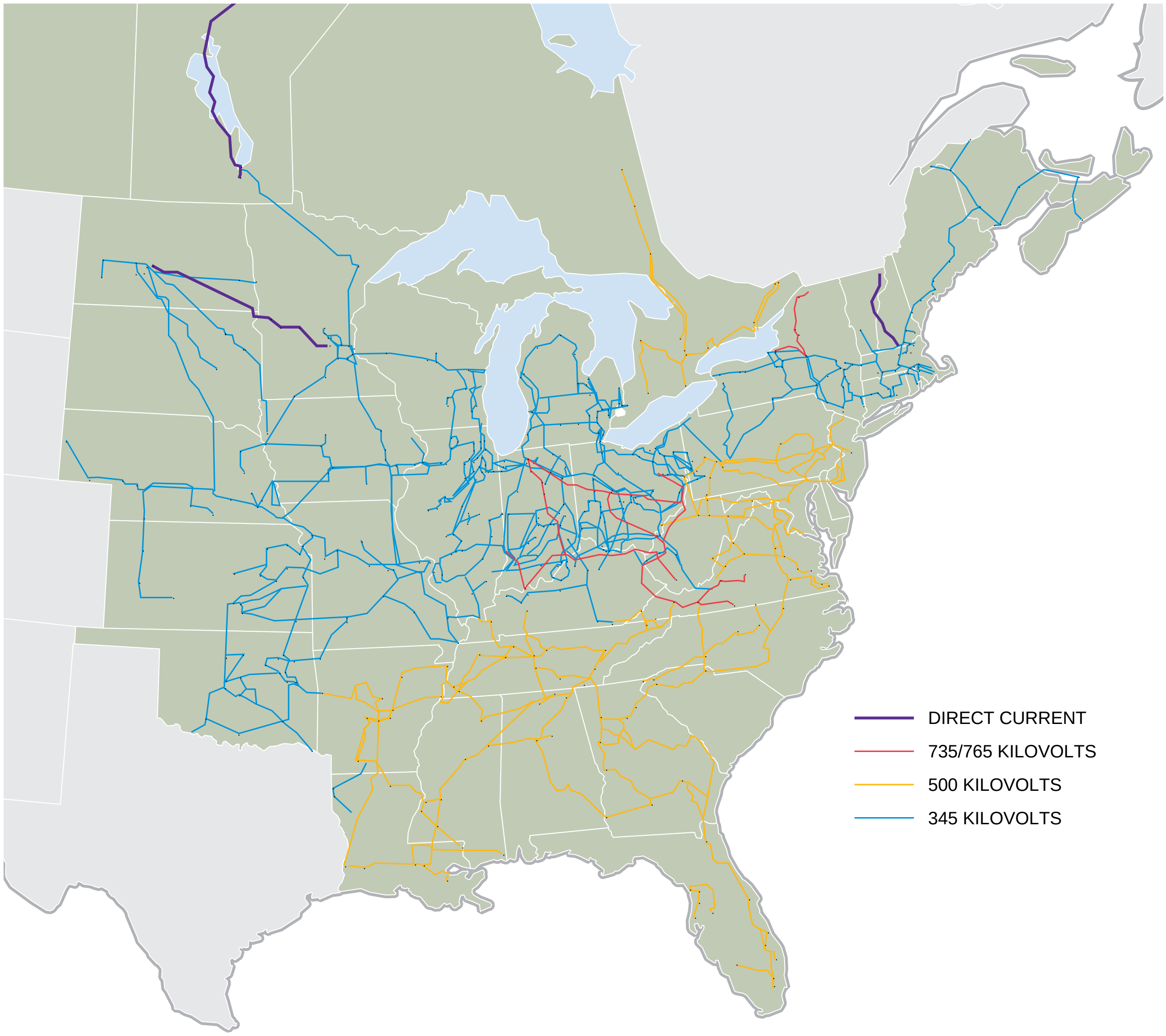


flows in only one direction: from producer to consumer. In AC systems, the flow of charge reverses itself many times each second; a terminal that had a positive potential a fraction of a second ago may now have a negative potential (and will have a positive one in another fraction of a second). A simple generator, consisting of a coil rotating in a magnetic field, will produce alternating current because the relative orientation of the coil and the magnetic field reverses with every half rotation.

Edison chose direct current for his first power plant, built to light lower Manhattan in 1882. DC at low voltage was safe, reliable and easy to control. It was also very inefficient: low voltage meant high current, and resistive losses increase as the square of the current transmitted. As a result, power plants could not serve customers who lived or worked more than a few miles away.

These limitations prompted Westinghouse to push for alternating current, which could be converted by transformers to high voltages for efficient transmission and then back to lower levels for safe use. (A transformer consists of two interwound coils; an alternating

ELECTRICAL SUBSTATION in San Jose, Calif., transforms power from a 500-kilovolt long-distance transmission line to 230 and 115 kilovolts for more local distribution. Utility managers operate such facilities and their associated transmission lines at a fraction of their potential capacity because conventional switching equipment cannot react quickly to unanticipated disturbances. Solid-state power devices, which are able to respond much faster to outages or to sudden changes in voltage, could increase the capacity of many transmission lines and distribution systems by as much as 50 percent.



POWER GRID ties together most of the U.S. and parts of Canada east of the Mississippi River (*left*). Other networks link the rest of the U.S. and Canada. By adjusting the voltage and phase of alternating current flowing through high-voltage transmission lines, utility companies are able to move power from one region to another.

current of one voltage passing through the primary coil generates an alternating current of another voltage in the secondary coil; the ratio between the two voltages depends on the number of turns in each coil.)

Edison warned against the dangers of alternating current, which can be more hazardous than equivalent levels of direct current. The advantages of alternating current became clear in 1896, however, when a high-voltage transmission line began to carry current to Buffalo, N.Y., from the first hydroelectric power plant at Niagara Falls. Other high-voltage links between widely separated load centers and power plants followed. Today most of the U.S. and Canada has been tied to one of several highly interconnected AC power systems.

Opponents of alternating current did have a point. AC power is more difficult to control than is DC. Line voltages tend to rise or fall in complicated ways as loads change. Voltage instabilities are more likely to get out of hand. These problems arise because AC generates changing electric and magnetic fields around the lines that carry it; these fields in turn affect the flow of current. Ideally, the power flowing through an AC line consists of a series of waves of rising and falling voltage and current that reverse polarity 120 times a second, in perfect sinusoidal rhythm.

When the sine waves representing voltage and current at any given point in the transmission line rise and fall in synchrony, they are said to be "in phase"—that is, the phase angle between them is zero. When a load is imposed at one end of a line, current flows through the load (doing work and creating heat in the process); this flow of electrical charge causes a slight drop in the line voltage, and so power flows from the higher-voltage end of the line to the lower-voltage one.

The relation between current and voltage in a DC line depends only on the resistance of the cables. In a line carrying AC, however, the relation is governed by a complex quantity known as impedance. Impedance is akin to resistance but varies with frequency. It can alter not only the strength of an AC flow but also its phase. Coils, for example, have a form of impedance called inductive impedance that passes direct current untouched but acts as a barrier to rapidly changing current flows. If a sudden voltage is applied to a simple inductor, such as a coil of wire, initially no current will pass; only after a characteristic time will the current flow build up to its full value. Consequently, voltage and current can get out of phase: voltage reaches its proper value on schedule, but the increase in current is delayed. When the sinusoids of commercial AC pass through an inductor, the rise and fall of current lags the peaks and troughs of voltage. In addition, overall voltage drops.

Capacitors, devices that store electrical charge, display a form of impedance that operates in opposing fashion. They present little hindrance to the passage of rapidly changing voltages but block the passage of direct current completely. Because capacitors act as reservoirs of electrical charge, they can allow large amounts of current to flow for a short time without an appreciable change in voltage. Hence, in AC systems, capacitance makes the sinusoids of current tend to run ahead of those of voltage.

The amount of power transmitted through an AC line is simply the product of current and voltage. As a result, either kind of phase mismatch between voltage and current impairs capacity; it does not matter whether the sinusoids of voltage run ahead of the sinusoids of current, or vice versa.

In a real power transmission system, these factors grow even more complicated. The impedance along any given path depends on the interaction between the cables and other equipment and the changing electric and magnetic fields engendered by the power flow. Transformers and motors create inductive impedance, whereas capacitors and long, lightly loaded transmission lines create capacitive impedance. As power flows change in response to these distortions, however, the magnitude of capacitive and inductive effects can change with them, thus altering the milliseconds-earlier balance and causing power flows to change yet again.

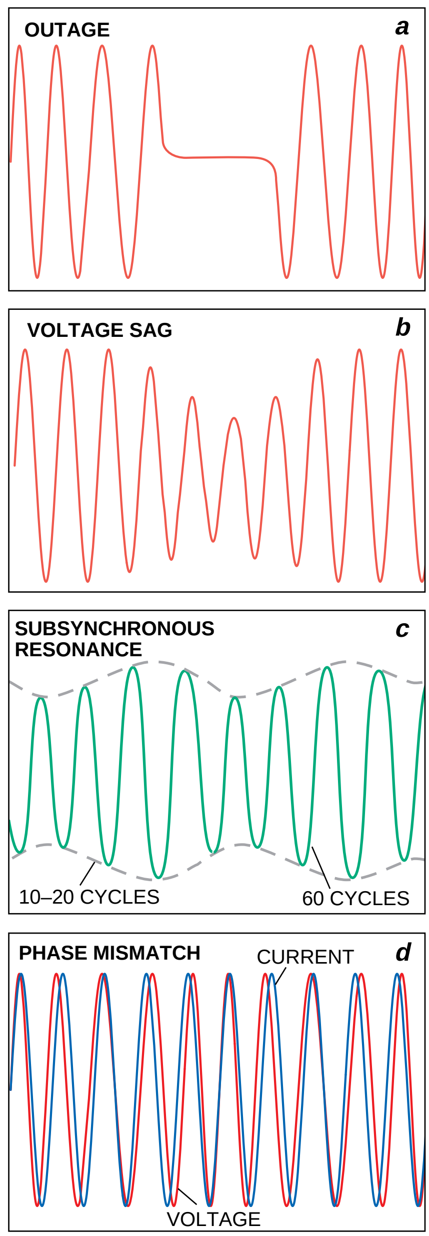
Controlling impedance is crucial to the proper distribution of power, but it is also very difficult. Once a transmission line, transformer or motor has been built, its impedance is fixed and cannot easily be changed. The impedance of transmission lines can be changed in large steps by inserting either series capacitors or inductors in the line, but this process is time-consuming and does not lend itself to the rapid control of electrical flows.

Interaction between impedance and current subjects AC networks to several different kinds of degradation. Among them are loop flow (the flow of current along unintended paths), large-scale instability (violent, uncontrollable fluctuations in power flow), poor control of line voltage and small-scale instability in the face of rapid load changes.

Loop flow occurs when many lines connect a power source and a load—hydroelectric plants in Ontario and air conditioners in New York City, say. Rather than taking the shortest physical path, power will flow along the lines that offer the least impedance. Ontario power destined for Manhattan may flow through transmission lines as far west as Ohio and Kentucky. Even though this distance may be physically longer than the "direct" path, impedance makes it appear much shorter in an electrical sense. The detour can be costly for the utilities whose lines carry the unwanted power, because the current passing through on its way elsewhere consumes transmission capacity that they could use to serve their own customers.

Large-scale instability is a result of the way that power flows from point to point in high-voltage transmission networks. Power always flows from areas with surplus capacity to areas with a deficit, but the number of megawatts transferred through specific lines depends on the impedance of the intervening lines and on the difference in phase between the sinusoids at the sending and receiving ends. The larger the phase difference, the more power is transferred. If the phase angle becomes too large, however, small phase shifts can lead to large changes in the magnitude of the flow. The system becomes unstable, producing a disturbance that can result in a widespread blackout. Utilities limit the amount of power transferred over their high-voltage lines so that instabilities occur only rarely. If there were some way of damping sudden fluctuations, however, power flows could be increased.

In addition to their problems in regulating phase angle, utilities can have difficulties in maintaining a constant line voltage. A factory with many large motors, for example, places a substantial inductive load on the line. Inductive impedance tends to pull voltage down, which can create problems for other customers connected to the same circuit. Long lines that are only lightly loaded, in contrast, may see voltage rise above acceptable levels. Finally, such events as lightning strikes, short circuits on power lines or sudden shifts in loads cause voltage instabilities. Uncontrolled, some of these disturbances can damage equipment or even cause full-scale blackouts. Usually, impedances inherent in transmission systems quickly



ALTERNATING-CURRENT transmission and distribution systems are subject to a variety of ills. Outages (*a*), which can be either momentary or lengthy, can disrupt complex manufacturing processes. Voltage sags (*b*) can also cause damage to devices that consume electricity. Subsynchronous resonance (*c*) can set up vibrations that destroy transmission and generation equipment. Phase mismatches between voltage and current (*d*) are generally not destructive, but they reduce the amount of power that can be transmitted from one location to another. Under extreme conditions, such reductions could cause blackouts.

bring the voltage back to normal, but sometimes the combination of capacitance and inductance can amplify surges or ebbs. Indeed, if subjected to certain kinds of periodic disturbance, a power system may exhibit subcycle resonance, in which it sustains alternating-current waves at frequencies in addition to the standard 60 cycles per second. Interactions between the normal power frequency and these spurious frequencies can set generators vibrating with enough energy to tear apart steel shafts three feet in diameter.

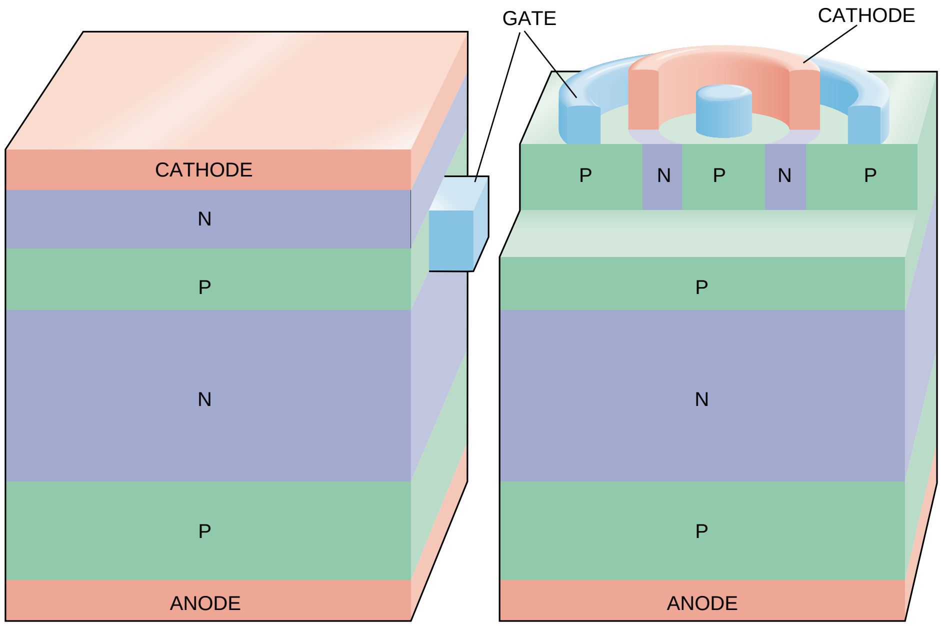
Disturbances in transmission systems are complex in cause and symptomology, but the effect on utility customers is invariable: lost work, disrupted manufacturing processes and damage to valuable equipment. Even the shortest of glitches can be exceedingly costly. In a recent survey of several industries, workers at Westinghouse found that an outage of just five cycles—83 milliseconds—cost a glass plant $200,000; an outage of little more than five minutes cost a semiconductor manufacturer $500,000. They estimate that such losses total between $3 billion and $5 billion annually in the U.S.

To avoid such outages, many companies are investing in their own power-conditioning equipment. Such equipment ranges from uninterruptible power supplies (UPS), which can provide enough energy to ride through short outages, to full-scale generation facilities that offer complete backup power for a plant. Unfortunately, current UPS technology is inefficient: between 10 and 20 percent of the energy that flows through such a device is lost. An estimated three billion kilowatt-hours go to waste every year, at a direct cost of more than $100 million (not to mention the indirect costs of burning additional fossil fuels and building new power plants and transmission lines).

In search of a way out of the choice between costly outages and wasted energy, engineers turned to semiconductor technology. For the past two decades, they have been developing high-voltage, high-power electronic devices known as thyristors. These components, disks of silicon a few inches in diameter and a fraction of an inch thick, can carry many hundreds and even thousands of amperes of current at potentials of thousands of volts. Furthermore, by acting in a small fraction of an AC cycle, they can prevent instabilities from damaging equipment or causing outages. A thyristor switch, for example, can cut off a failing power line and bring in a backup source without interrupting the alternating-current waveform.

How Thyristors Work

Solid-state power control starts with the silicon-controlled rectifier, the simplest kind of thyristor (*left*). A positive voltage applied to the gate of an SCR at the beginning of an AC cycle will cause it to conduct as long as the voltage at the anode is higher than the voltage at the cathode. (In general, current



SILICON-CONTROLLED RECTIFIER

GATE TURN-OFF THYRISTOR

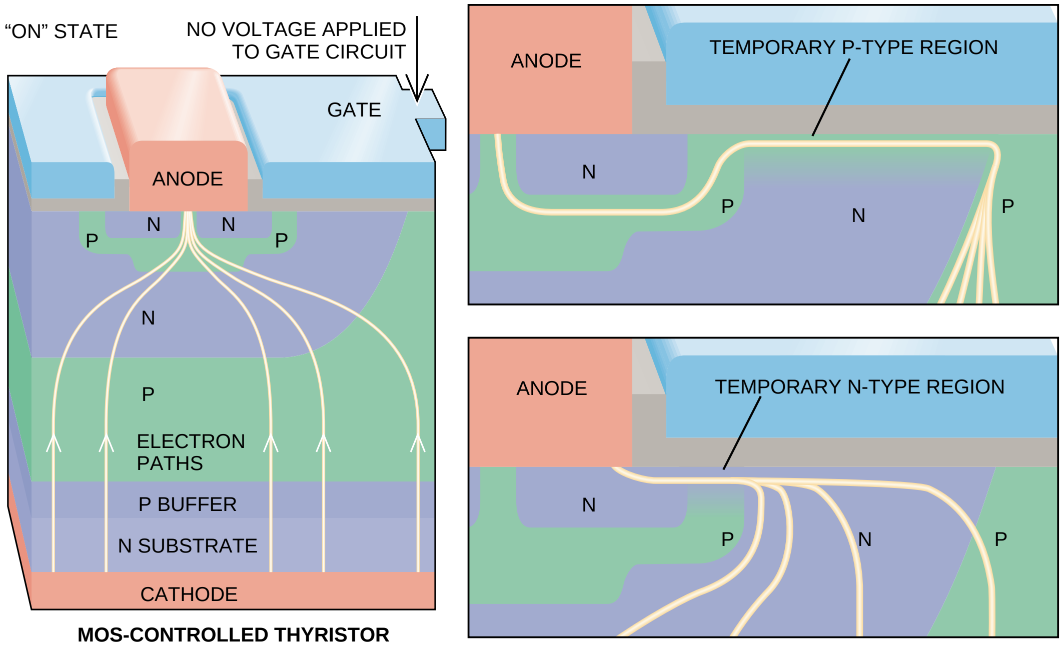
Thyristor-based systems can control impedance, voltage, current and phase angle in ways that would be impossible with mechanical switches.

These devices can improve the flexibility of transmission networks by enabling utilities to increase the loading of lines that are now limited by concerns over loop flow, stability or other problems. Indeed, in some cases, such controllers may be able to double the capacity of crucial transmission corridors; utilities can thus defer the construction of new lines and power plants. The Electric Power Research Institute, a collaborative research and development arm of U.S. utilities and several international affiliates, has coined the term "FACTS" (flexible AC transmission system) to cover the wide range of applications for thyristor-based circuits.

As do conventional transistors and integrated circuits, the new high-power devices depend on the conducting properties of crystalline silicon doped with impurities. A typical transistor might consist of two layers of silicon doped with phosphorus or other elements that donate free electrons (negative, or *n* regions) separated by a layer doped with boron or other elements that accept free electrons (positive, or *p* regions). A small voltage applied to the *p* layer can change its conductivity and thus switch on or off the flow of current from one *n* layer to the other.

In general, electricity flows across a junction between the *n*- and *p*-type material when a positive voltage is imposed on the *p* side and a negative one on the *n* side. (Current in *n* regions is carried by electrons, which are negatively charged, and in *p* regions by positively charged "holes" that represent the absence of an electron where one would normally be.) A single *p*-*n* junction, a diode, permits current to pass in only one direction, from the *p* to the *n* region. Such devices are also called rectifiers because they can convert alternating current to direct current.

Thyristors, the workhorses of power electronics, have four layers, arranged in a *p*-*n*-*p*-*n* configuration. The first *p*-*n* junction acts as a diode rectifier that controls the direction of current. The second is controlled by a gate and so acts as a switch: the voltage applied to the gate (hence to the *p* layer) determines whether the second junction—and thus the device as a whole—will conduct. The simplest form of thyris-



will flow across a junction between *p*- and *n*-type semiconductors whenever the potential on the *p* side is higher than the potential on the *n* side.) The gate turn-off (GTO) thyristor (*center*) can be turned on or off at any point in the AC cycle; applying a voltage to the gate diverts current from the junction between the *p* and *n* regions and so prevents the device from conducting. The MOS-controlled thyristor (MCT) is more efficient than are other kinds of thyristors but is significantly more complex (*right*). When the device is conducting, current flows through the thin *p* region immediately below the anode. To turn the device off, one applies a voltage to the gate; the resulting electric field makes a small part of the *p*-type material act like *n*-type material; the current thus flows around the *p* region, disrupting the thyristor's conductivity (*far right, top*). To turn the MCT back on, a voltage of opposite polarity is applied to the gate; this causes a thin *n* layer to act like *p*-type material. Current flows through this layer to the *p* region under the anode and establishes the conditions for normal conduction (*far right, bottom*).

tor, the silicon-controlled rectifier (SCR), can be turned on by applying a positive voltage to its gate (and so to the buried *p* layer) when the voltage on the surface *p* layer is also positive. Current then flows through the entire device. Once a rectifier has been turned on, it will continue to conduct, regardless of the gate voltage, until the voltage on the surface *p* layer returns to zero.

One step up in sophistication is the gate turn-off (GTO) thyristor, which can interrupt current at any point in the AC cycle. The second *n* layer in this device consists of multiple channels, all embedded in the adjacent *p* layer. This *p* layer is controlled by a gate electrode, also divided into multiple channels: when there is no voltage on the gate, current flows from *p* to *n*. A negative voltage on the gate sidetracks current from the *n* channels and thereby prevents current from flowing through the device. GTO thyristors are not yet in use; the Tennessee Valley Authority expects to install units capable of controlling 100 megawatts on a transmission line in 1995.

Because they require that current pass through many narrow channels, GTO devices are inefficient. They are also relatively expensive to build. Engineers have been working on other switch designs that will require much less control current and be able to turn current on or off more quickly. One such device is the MOS-controlled thyristor (MCT). It consists of conventional integrated circuits ("metal oxide semiconductor" devices) etched into the top surface of the silicon that forms the main part of the thyristor. When a voltage is applied to the gate of the MOS circuit, the resulting electric field enhances the conductivity of the buried *p* layer so that current can flow. MCTs are at a fairly early stage of development. Devices now reaching the market can control only 120 kilowatts and are being sold for industrial applications. Utility-scale MCTs are expected to become available around 1997.

Although FACTS devices have clear theoretical advantages over the slow electromechanical switches now used to operate high-current, high-voltage lines, many utilities are waiting for solid-state technology to become available and cost-effective before adopting it on a wider scale. EPRI is now working to demonstrate FACTS controllers in several key applications. First comes the solid-state subsynchronous damping scheme. Next is the thyristor-controlled variable capacitor system, which can reduce impedance on transmission lines and thus control and increase power flow. Fast-acting voltage regulators will be able to prevent voltage fluctuations on heavily loaded lines. Finally, EPRI is promoting the use of phase angle regulators to reduce power on transmission lines overloaded by loop flow. These devices can also increase flow on underutilized lines.

Although power companies in western states have employed fixed series of capacitors for years to reduce impedance on long transmission lines, utilities in the rest of the country have avoided them because a line with too much capacitance is vulnerable to subsynchronous resonance. Such low-frequency vibrations, once triggered, can damage generators and other equipment. Thyristor-controlled series capacitors (TCSCs) avoid this danger. If power fluctuates, the thyristor can change the period for which the capacitor is in the circuit. In this way, the device controls the impedance and damps the unwanted resonance.

EPRI first demonstrated thyristor-based systems for voltage control and damping of power system oscillations in 1978. In 1985 the institute installed another system for damping subsynchronous oscillations on a 500-kilovolt transmission line belonging to Southern California Edison. Each of these in turn allowed some increase in transmission capacity of these lines.

In 1991, as part of the FACTS program, the American Electric Power Service Corporation tested a thyristor-based switch in part of a series capacitor bank installed on a 345-kilovolt line near Charleston, W.Va. Since then, the Western Area Power Administration has installed a similar capacitor system for testing at the midpoint of a 200-mile-long 230-kilovolt line in Arizona. The system increases power transfer on the line from 300 to 400 megawatts. The Bonneville Power Administration, with EPRI sponsorship, has now completed a much larger TCSC system, on a 500-kilovolt, 2,500-megawatt transmission line in north-central Oregon. This installation, built by General Electric, consists of multiple small capacitor sections; each section contains its own thyristor-based controller so that engineers can tune capacitance to the precise level desired. The Bonneville Power Administration expects that the system will increase transmission capacity during springtime peaks. Indeed, preliminary studies indicate that an installation of this type could pay for itself in less than a year. Furthermore, the controller is designed in a highly modular fashion that will allow it to be adapted readily to other transmission lines.

When the thyristor-controlled capacitor system begins operation late in 1993, the Bonneville Power Administration will conduct tests to determine whether precise control over capacitance current and power can in fact damp subsynchronous resonance. The Portland General Electric Company, which owns a power plant near the TCSC installation, will join in these tests. Other utilities are studying the benefits of installing TCSCs. Southern Company Services of Georgia, for example, has concluded that it could save $120 million by installing a thyristor-controlled series inductor system rather than building a new transmission line where its facilities join those of the Tennessee Valley Authority.

EPRI's next transmission project is an outgrowth of technology that has been used to regulate voltage since the 1970s. Adding a capacitor or inductor to a line can keep voltage within acceptable limits when sags or surges threaten stability. The devices now used, called SVCs, consist, in effect, of a conductor running from a transmission line to ground through a capacitor or an inductor; thyristor switches determine whether current can flow through the capacitor or inductor. When voltages on transmission lines go below about 80 percent of normal, these simple shunts can no longer compensate enough to bring it up to proper levels.

A new type of shunt, based on GTO thyristors, can do a better job of compensating for voltage variations than can the current SCR-based devices. Instead of simply changing the amount of capacitance and inductance in the shunt, this so-called static condenser employs a direct-current capacitor that feeds precisely timed pulses of voltage into the transmission line to raise its potential to the proper level. The GTO device connects the capacitor to the transmission line and then disconnects it—producing a pulse—as many as 48 times in a single alternating-current cycle. The new device can also lower the voltage on a line by injecting pulses that oppose the normal cycle of alternating-current sinusoids.

With EPRI's support, Westinghouse and the Tennessee Valley Authority are building a GTO-based static condenser near Johnson City, Tenn. The system will regulate voltage on a 500-kilovolt transmission line and two 161-kilovolt lines, which are connected by large transformers. The device is scheduled to go into operation in 1995. It will hold down voltage on the 500-kilovolt line, which tends to rise when loads are light. It will also prevent voltage on the 161-kilovolt lines from sagging under peak loads. In addition, it will help damp voltage oscillations that have begun to appear on the lines of a neighboring utility. The Tennessee Valley Authority might otherwise have to construct another 161-kilovolt line to provide adequate operating margin for this area of its power system, at considerably greater cost.

The static condenser technology is also expected to find use in distribution systems, where it will provide reliable, high-quality power for commercial and industrial customers. Today a utility cannot effectively supply a customer with two independent, redundant distribution lines. A mechanical circuit breaker would take too long to switch to the second line if the first failed. A solid-state circuit breaker, however, can make the transfer in a single AC cycle. A GTO-based regulator can hold the voltage constant during this transition so that equipment will not be affected. Such regulators are already working in the laboratory; they could be on the market as soon as 1995.

Even more complex than the static condensers are thyristor-controlled phase angle regulators, which may also be available in the future. These devices consist of a large transformer equipped with an extra coil that can, in effect, bleed off voltage and then reinject it at a different phase angle with respect to the current. Depending on the precise angle at which the voltage is reinjected, the overall voltage waveform will lead or lag current by varying amounts. By varying the phase angle of the voltage, engineers can control the amount of power that flows from one end of a transmission line to the other. Thyristors control the flow of electricity through this extra coil and so determine the phase relation between voltage and current for power passing through the device.

One demonstration project is now under study: a thyristor-controlled phase angle regulator (TCPR) at International Falls, Minn., would help control power flow on a 115-kilovolt line that connects Ontario Hydro and Minnesota Power and Light. The direction of this flow typically shifts with the seasons—energy flows north during winter and south during summer. The two systems are only weakly connected, however, and voltages on one power system may become widely out of phase with those on the other. Increasing the 100-megawatt carrying capacity of the line requires direct control of the phase angle; otherwise, phase mismatch caused by a voltage disturbance on either system could rapidly lead to an overload.

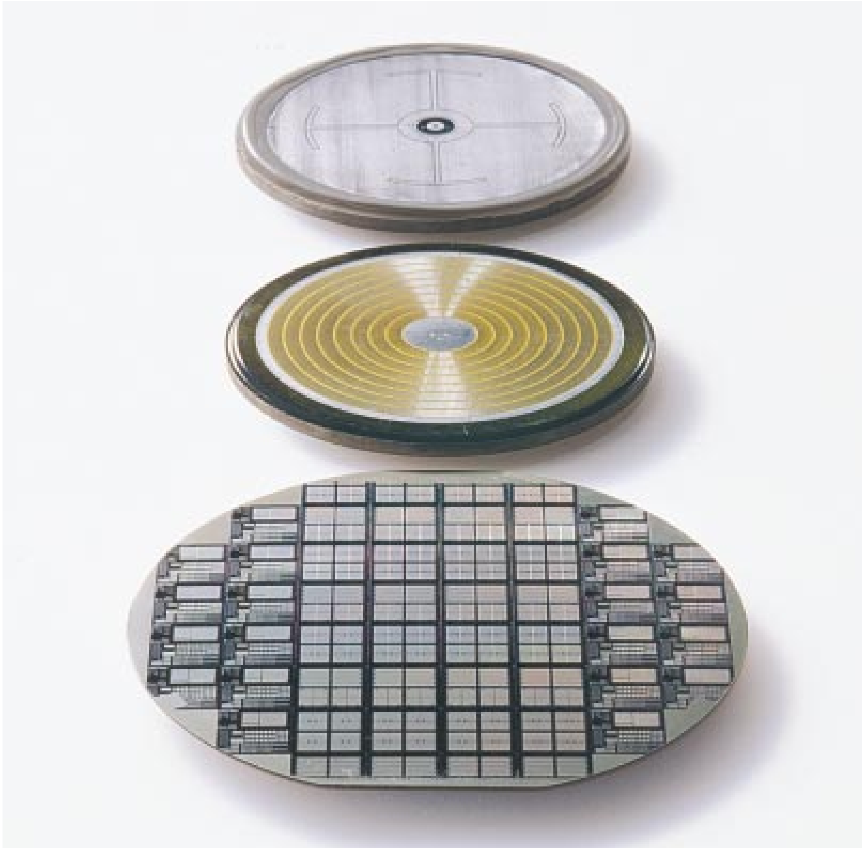
Specifically, the TCPR at International Falls would shield the interconnection from a power surge that could be caused by the inadvertent shutdown of a nearby 500-kilovolt line in Canada. This protection will enable utility managers to up the safe transfer limit to 150 megawatts. The phase angle regulator may also help damp low-frequency oscillations in the area. EPRI is directing detailed design studies, but implementation is at least two years away.

Whereas EPRI is sponsoring work primarily on a few key controllers, manufacturers inspired by potential market opportunities are developing various other thyristor-based controllers. Perhaps the most prosaic of these is the solid-state circuit breaker. Because these devices are not yet suited for high-voltage transmission, they will appear (in contrast to the other controllers) first in local power distribution systems, where they need only carry hundreds of kilowatts rather than the hundreds of megawatts of power flowing in transmission systems. These solid-state devices are much faster than their mechanical counterparts and last far longer because they do not have contacts that are damaged by arcs when they turn off.

An example of the more sophisticated controllers is the thyristor-controlled series reactor (TCSR), the inductive counterpart of the thyristor-controlled series capacitor. It adds impedance to reduce the load on a line or to limit current in the case of a sudden fault. Like its sibling, the TCSR can be placed into operation on a few microseconds' notice. Together the TCSR and TCSC may provide a lower-cost alternative to phase angle regulators. TCSR technology should be ready for demonstration by the late 1990s.

A thyristor-controlled braking resistor, meanwhile, could protect generators supplying power to long transmission lines from a sudden loss of load. Without such protection, a generator feeding an unloaded line would start gaining speed until it was damaged or had to be disconnected. Although mechanically controlled resistors can also brake generators, these devices have never gained acceptance among utility engineers because they can only be slammed on or turned off. They are useful in major emergencies but could cause more trouble than they avert. Thyristor-controlled units, in contrast, can be applied gently when only small braking actions are needed or sharply in an emergency. Manufacturers are expected to demonstrate a prototype before the end of the decade.

Another kind of device that shows promise is the thyristor-controlled active filter. Such filters address problems of power quality rather than reliability. Heavy electrical equipment often produces higher harmonics—voltage fluctuations at multiples of 60 cycles per second—that feed back into the power grid and cause disruptions. Utilities now use passive filters, which permit only specific frequencies to pass, on some lines subject to harmonics. Unfortunately, passive filters add impedance and so waste energy. Active filters, in contrast, would detect harmonics and apply a precisely timed counter voltage to neutralize them. Such devices have yet to leave the laboratory.



THREE GENERATIONS of thyristors show how these devices have increased in complexity and capability. The first-generation silicon-controlled rectifier (*top*) can switch power off only at the beginning of an AC cycle. Gate turn-off devices (*middle*) can switch at any point in the cycle, and MOS-controlled thyristors (*bottom*) switch more quickly while consuming significantly less power.

Taken together, the various kinds of high-power silicon devices offer utilities unprecedented control over power transmission and distribution. The economic benefits of the technology should be substantial for both utilities and their customers. Reduction of sags, momentary outages and subharmonic resonances will save some of the billions that those problems cost every year, as well as the billions now spent on uninterruptible power supplies. Moreover, "smart" power devices in utility lines consume only about 1 percent of the power they deliver, instead of the 10 to 20 percent that existing devices take.

In addition, solid-state controllers could increase the overall capacity of today's transmission network by 20 percent or more. Engineers have estimated, for example, that adding thyristor-controlled series capacitors to two 500-kilovolt lines serving Florida could increase the power transfer capacity from 2,000 megawatts to 3,000 megawatts. Such an investment could pay for itself in less than a year. Nationwide, utilities could save $6 billion compared with the cost of adding the same capacity with new lines.

Even more important, increased flexibility in distribution will make it easier to transfer power from one region to another on demand, thus permitting individual utilities to reduce their generating reserve margins. Smaller reserve margins, in turn, will enable utilities to defer construction of new power plants. If thyristor-based devices make it possible to reduce margins from the current 20 percent to 15 percent, $50-billion worth of power plants could be left unbuilt over the next 25 years.

FURTHER READING

ADVANCED STATIC VAR COMPENSATOR EMPLOYING GATE TURN-OFF THYRISTORS FOR UTILITY APPLICATIONS. L. Gyugyi, N. G. Hingorani, P. R. Nannery and N. Tai in *International Conference on Large High Voltage Electric Systems: Proceedings of the 33rd Session.* CIGRE, Paris, 1990.

UNIFIED POWER-FLOW CONTROL CONCEPT FOR FLEXIBLE AC TRANSMISSION SYSTEMS. L. Gyugyi in *IEE Proceedings–C*, Vol. 139, No. 4, pages 323–331; July 1992.

FACTS: THE DELIVERY SYSTEM OF THE FUTURE. J. Douglas in *EPRI Journal,* Vol. 17, No. 7, pages 4–11; October 1992.

FLEXIBLE AC TRANSMISSION. N. Hingorani in *IEEE Spectrum,* Vol. 30, No. 4, pages 40–45; April 1993.

Ancient DNA

Genetic information that had seemed lost forever turns out to linger in the remains of long-dead plants and animals. Evolutionary change can at last be observed directly

by Svante Pääbo

Most of our knowledge of the molecular processes that underlie evolutionary change is based on the comparison of the genes of living species. From such differences, molecular evolutionists infer the historical changes that gave rise to present-day DNA sequences. Yet these studies are tentative in nature. Unlike the remains of animals and plants, DNA molecules do not leave impressions in rock. Biologists therefore despaired of ever being able to check their conclusions against the historical record, as paleontologists do.

But in the past decade scientists have learned that ancient DNA, though degraded, sometimes survives the ravages of time, and molecular biologists have perfected methods of amplifying these trace amounts of ancient DNA. Workers have so far used DNA from bone and soft tissues to establish reliable sequences for seven extinct mammals. The oldest was the woolly mammoth—a frozen carcass that was found in the permafrost of Siberia. More such studies are under way, including efforts to decode DNA extracted from insects entombed in amber millions of years ago. We can thus look forward to learning much more about the genetic relations among extinct species.

Encouraged by such work, the Natural History Museum in London established a molecular biology research laboratory. The Smithsonian Institution in Washington, D.C., the American Museum of Natural History in New York City and other museums have since emulated the British example.

I first became aware of the spectacular ease with which molecular techniques allow us to isolate and study DNA sequences from humans and other organisms when I was a graduate student in the early 1980s in Uppsala, Sweden. I began to wonder whether the same techniques could retrieve meaningful data from the skins of animals and the mummies of humans that abound in museums. When a search of the literature suggested that no one else had considered the possibility, I embarked on a quest for suitable samples to test.

It was no easy task to get materials, for an archaeological sample is valuable, and inevitably one will have to destroy a part of it to extract the DNA. Yet my proposal intrigued curators of the University of Uppsala's Egyptian antiquities, and they provided me with small samples of skin and muscle from their mummies. Even more important, they had excellent connections with a large museum in what was at that time East Berlin.

I spent four days with the chief conservator of the State Museums in Berlin, looking through the mummies, some of which had been partly destroyed during World War II. With a sterile scalpel I removed samples of a few grams of tissues from the 23 mummies that appeared best preserved.

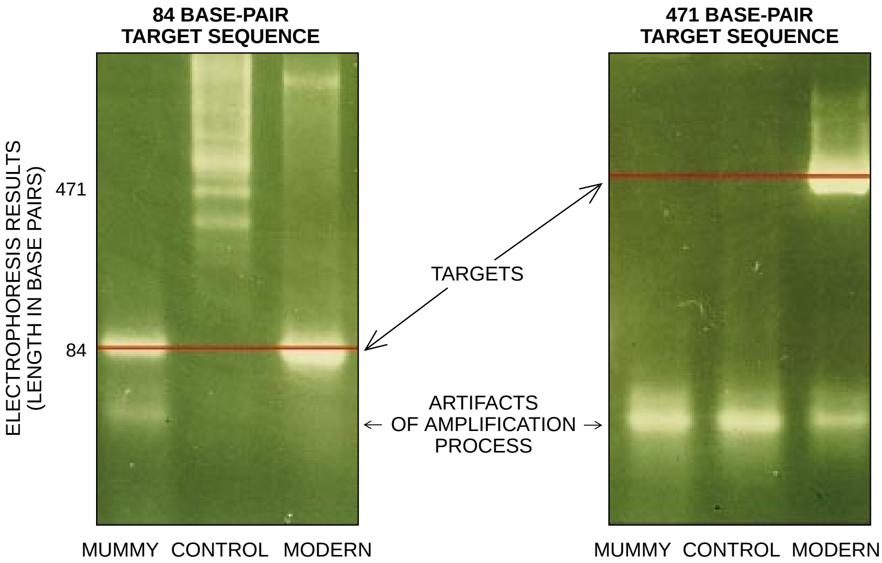
Back in Uppsala I set out to study the tissue, working nights and weekends so as not to impede my research in molecular virology, which was supposed to constitute my future thesis. I began by examining the specimens under a microscope. I found that they varied dramatically in their state of preservation. Most of the tissues were badly degraded, but there were exceptions. Superficial and peripheral parts, such as the skin on fingers and toes, were much better preserved [*see illustration on page 92*]. These tissues quite often retained nuclei that accepted stains specific for DNA. The reason appeared to be that the enzymes that cause dying tissue to digest itself, a process called autolysis, require water. Superficial tissues, however, may dry out before the postmortem degradation of DNA is complete.

With great excitement I proceeded to extract the DNA from the tissues that contained cell nuclei. I worked as one would with any modern tissue, digesting away the proteins with enzymes and using solvents to extract DNA from sugars, proteins and lipids. I then analyzed the DNA by electrophoresis, a technique that uses an electric field to separate DNA fragments as they migrate through a gel. The smaller the fragments are, the farther they migrate, a correlation from which one can calculate their size. The results showed that the DNA had been degraded to fragments of only 100 to 200 base pairs (the four nucleotides that are the building blocks of DNA). Fresh tissues, in contrast, yield DNA fragments more than 10,000 bases in length.

To study the genetic information of such DNA fragments, one must amplify them into large quantities. At that time, the only available way to do so was molecular cloning, in which the old DNA is fused to a carrier DNA molecule that confers on it the ability to copy itself in bacteria. These hybrid molecules are then introduced into bacteria, and thousands of bacterial colonies are grown, each one multiplying a single piece of the original DNA. The bacterial colonies are then analyzed for

SVANTE PÄÄBO is professor of biology at the University of Munich. He was born in Stockholm in 1955 and studied Egyptology, medicine and the history of science at Uppsala University, where he took a doctorate in molecular immunology in 1986. He then worked in Zurich and, for four years, in the laboratory of the late Allan C. Wilson of the University of California, Berkeley. His current research into ancient and modern DNA focuses on the elucidation of evolutionary history and the molecular processes underlying genetic change. In 1992 the German Science Foundation awarded him the Leibniz Prize.

SKIN AND BONES have yielded priceless genetic data from such extinct species as quagga (*zebralike animal, center*), moa (*skeleton, foreground*) and thylacine, or marsupial wolf (*front*).



MUMMY DNA allows amplification of only short fragments, whereas modern DNA yields longer ones. The control lanes on the electrophoretic gel show the absence of contaminating human DNA but show amplification artifacts.

sequences of interest. In my experiment it was crucial to demonstrate that at least some of the cloning products were of human origin and could not have come from bacteria or fungi that may have lived in the mummy.

For reasons that became clear only later, I obtained far fewer clones carrying human DNA than I had expected. The paucity of clones precluded an attempt to isolate the most interesting genes—those that code for specific proteins—because these generally occur in only two copies per cell. Therefore, to show that I had indeed replicated ancient human DNA, I had to resort to isolating a so-called Alu-repeat, a sequence that occurs in nearly a million copies in the genome of humans. One bacterial clone containing two such Alu-repeats was isolated from tissues of a mummy reliably dated to between 2,310 and 2,550 years ago. This finding made it clear that DNA can survive for long periods after the death of an individual.

In 1984, some months before I had my results ready for publication, Russell Higuchi and the late Allan C. Wilson of the University of California at Berkeley published work that also showed that DNA could survive after the death of an organism. They studied the quagga, a member of the horse family that lived in southern Africa until going extinct at the end of the past century. Samples from a 140-year-old quagga skin in a German museum yielded bacterial clones containing sequences from the mitochondrial DNA, which exists in many copies outside the cell nucleus. By comparing these clones with sequences of modem zebras, Higuchi and Wilson showed that the quagga was closely related to the zebra and much more distantly related to other equids. These were the first sequences determined from an extinct species.

Despite these successes, it soon became clear that there were severe limitations to our ability to study ancient DNA. The dearth of clones made it extremely difficult to repeat experiments and made many experiments impossible to perform in the first place. The reason for the low efficiency of cloning was discovered when the ancient DNA was analyzed biochemically. The average length of the old molecules (about 100 base pairs) was the same whether the material came from a 13,000-year-old ground sloth of southern Chile or a piece of dried pork just four years old. Most of this fragmentation occurs in the first hours after death, before the tissues dry out. Other types of damage appear to come from oxygen radicals, which react with bases or the sugar backbone of the DNA. These reactions cause modifications and loss of bases, breakage of the strands and cross-linking of DNA molecules to one another. Such changes degrade the DNA and destroy its genetic information.

When ancient DNA damaged in these ways is introduced into bacteria, the organisms try but generally fail to make copies. Even when they succeed, they often introduce errors. For example, the quagga sequences included two positions that had bases different from those that occupy the same position in the genomes of other vertebrates. It seemed likely that the isolated clones had errors in them at these positions, but because it was so difficult to obtain clones from the quag-

Ancient DNA Milestones

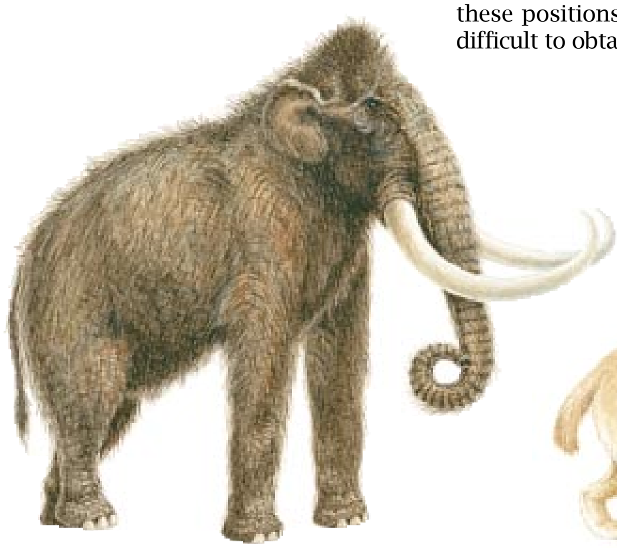
These extinct organisms have yielded meaningful genetic sequences.

AMBER

40 MILLION YEARS OLD

FOSSIL LEAVES

17 MILLION



MAMMOTH

40,000

13,000

AMBER TIME CAPSULE for a prehistoric scorpion may provide an ideal environment for the preservation of DNA. Amber, the solidified resin of ancient trees, generally protects its contents from the effects of air and water.

ga skin, there was no practical way to go back and repeat the experiments. Similarly, I could not again isolate the same Alu-repeats from the mummy in order to verify their sequence. Molecular evolutionists who were keen on time travel therefore found themselves in a depressing situation. Because they could not verify their results by repeating an experiment, the study of ancient DNA could not qualify as a fully respectable science.

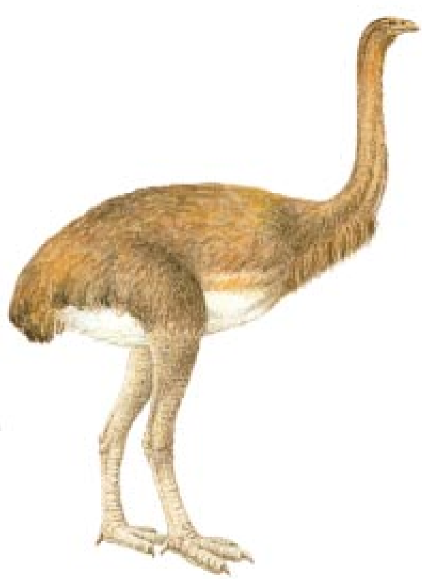
A dramatic change came in 1985, when Kary B. Mullis described a test-tube cloning technique of extraordinary sensitivity and power [see "The Unusual Origin of the Polymerase Chain Reaction," by Kary B. Mullis; SCIENTIFIC AMERICAN, April 1990]. This method, the polymerase chain reaction (PCR), allows the multiplication in the test tube of a particular piece of DNA. The piece is specified by two short DNA fragments, or primers, which are chemically synthesized to match the base sequences flanking the sequence one wishes to study. The primers initiate the test-tube amplification, which proceeds in successive cycles that each double the number of copies of the specified sequence.

A cycle begins with the melting of the double-strand template into single strands. This separation enables the two primers to sit down on their respective target sequences—one on one strand, the other on the complementary strand. Next, an enzyme begins adding bases at the ends of the primers, extending the double helices that the primers had initiated. Because each single strand produces a new double helix, the desired DNA sequence doubles in quantity with each cycle.

It immediately became clear that PCR could open great possibilities for the study of ancient DNA. The main reason is that the technique is extremely sensitive. One can literally generate billions of copies starting from as little as a single DNA molecule. Even one or a few intact molecules that survive in an ancient tissue might be amplified by PCR, whereas the damaged molecules, which may be present in an excess of several 1,000-fold, would not disturb the experiment. In the hope that this potential would be realized, I joined Wilson's group at Berkeley, which was one of the first to use PCR.

Our efforts were soon rewarded. When we designed primers for the segments of quagga DNA that Higuchi had cloned in bacteria a few years before, we were able to amplify the pieces of DNA from extracts of the skin that he had used. Our experiment revealed that the positions in which the cloned quagga sequences had differed from other vertebrates had in fact carried the expected general vertebrate sequence. Thus, the bacterial clones at these positions carried errors, presumably induced by damage in the ancient DNA that the bacteria had incorrectly repaired. PCR, however, was able to retrieve the correct sequence.

We have since shown that PCR brings back the correct ancient sequence because its errors, though frequent, are rarely serious. It is accurate because, in practice, PCR starts the amplification from at least a few dozen or a few hundred ancient DNA fragments. When the end product of an amplification is analyzed all at once, any error in a particular position in the sequence of one initial template will be canceled by the

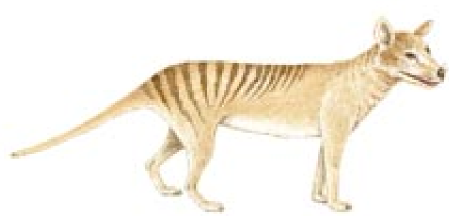


MOA

4,300

QUAGGA

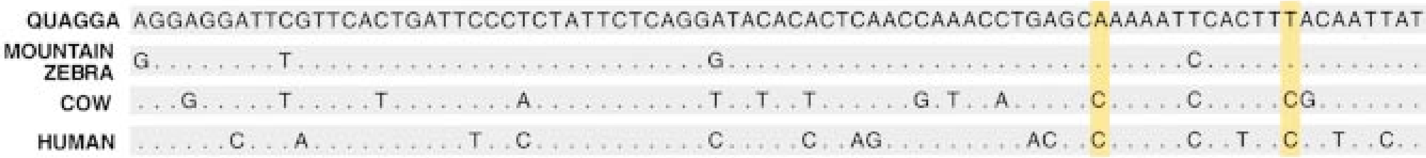
140



THYLACINE

80

PRESENT



EVOLUTIONARY TREE OF THE QUAGGA is deduced from differences between DNA from the extinct equid and from three modern species. The quagga shares much more sequence similarity with the zebra than with the others. The evolutionary tree (*far right*) reconstructs genetic changes, such as two inferred in the ancestor of the quagga and the zebra (*yellow*). The quagga dif-

other templates that do not carry mistakes in that position. Bacterial cloning, in contrast, starts from a single molecule and is so more prone to make errors of consequence.

Further studies have shown that PCR can also reconstruct intact DNA from several partially degraded ancient molecules. In such cases, the two primers sit down on partial molecules. They are extended up to a damaged site or to an end of the template molecule. In the next cycle the resulting molecules separate into single strands; these primers are longer than the original primers were. They can therefore bind to other template molecules that are not damaged where the previous templates were.

This process of jumping from template to template can piece together information from many partially degraded pieces. Under favorable conditions, it can enable PCR to amplify sequences that are longer than any single fragment in the original sample. We are attempting to use this property of PCR to develop procedures that would allow the reconstruction of even more extensively damaged DNA than can now be analyzed. Other experiments are directed at repairing the damage in the old DNA before it can be amplified by PCR.

Yet it can be argued that the most important advantage of PCR is its technical simplicity, which made it possible to amplify the same sequence from different quagga extracts in just a few days. We and other laboratories could thus reproduce and verify the results: molecular archaeology could for the first time claim to be a respectable branch of science.

Could we extend the technique to materials that were thousands of years old and were retrieved from archaeological excavations? To find out, we tested the new technique on a human brain that, amazingly, had survived intact for 7,000 years at Little Salt Spring, a sinkhole in Florida. William W. Hauswirth and his collaborators at the University of Florida at Gainesville had already shown by other techniques that DNA was preserved in a brain found at a similar site, presumably because the water was neutral in pH and contained very little oxygen. We extracted and attempted to amplify pieces of mitochondrial DNA, whose rapid rate of evolution makes it very useful for studying relations among populations.

Our first attempts failed because the extracts contained unknown factors that somehow inhibited the enzyme used in PCR to replicate DNA. This problem annoyed us for several months, until we found out that albumin, a protein found in the blood, appears to relax the inhibition by binding to impurities that disturb the polymerase. The poor state of the old DNA allowed only short pieces to be amplified, but even so, the fragments sufficed to reveal a mitochondrial type that had not been found in contemporary Native Americans.

We could not be certain that this genetic variant had gone extinct in the Americas, because we did not know much about the mitochondrial sequence variations in contemporary Native Americans. Fortunately, several groups are now engaged in studies to provide such data, not only for Native Americans but for other populations in the Old and New Worlds as well.

The objects of such investigations are many. Comparisons of DNA sequences in different populations help workers not only to detect shared variants but also to estimate the degree to which the sequences are related. This information can be used to gauge how populations are related to one another and the time that has elapsed since they shared an ancestor. One can even find signs of ancient expansions and contractions in population size, such as might have followed epidemics, famines, wars or migrations.

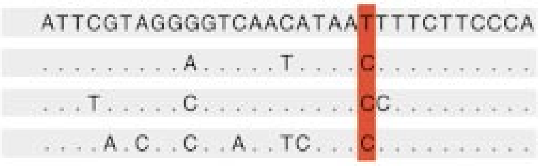
Genetic studies strongly imply an Asian origin for living Native Americans. They also suggest that a fairly small group of Asians colonized the New World, perhaps in just a few waves of immigration. For the first time, we can begin to test these theories by analyzing the DNA of ancient populations in the New World. This possibility is extremely exciting in itself. It may also enable us to determine whether a population was replaced by another one or simply adopted a new material culture.

Unfortunately, the study of old human remains is still impeded by technical difficulties that are to some extent inherent in PCR: its exquisite sensitivity can be a weakness as well as a strength. Any trace of modern DNA that makes it into the experiment will be amplified if it carries sequences to which the primers can bind. One must wonder whether a similarity between oneself and one's putative ancestors is real or merely the result of sloppy laboratory technique.

Contaminating DNA can come from many sources: skin cells shed by the archaeologists or museum keepers as they handle the samples; dust particles; or minute amounts of DNA from earlier experiments conducted in the same room. This problem has forced workers to establish cumbersome procedures. All reagents must be treated with the utmost care. Extractions and amplifications must be performed in a room separate from all other work. Ventilation systems must not connect one PCR laboratory with another using similar material. Yet despite such precautions, contamination continues to be a serious concern, especially when one tries to amplify DNA from ancient human remains in which contamination from modern humans cannot easily be ruled out.

For this reason, the first study of old DNA from populations focused not on humans but on a remotely related species—the kangaroo rat. My colleagues Kelley Thomas, Francis Villablanca and I studied museum collections containing 48 skins collected during the early part of this century from three populations in the Mojave Desert. All the skins contained mitochondrial DNA good enough to yield products 200 base pairs long.

After we had determined the sequences from the museum skins, we made use of old field maps and collected samples from kangaroo rats that live at the same localities today. We discovered that the contemporary rats carry sequences identical or closely similar to the rats that lived in the same places 60 to 80 years ago and that the populations contain as many different sequences now as then. Thus, the contemporary kangaroo rats are the direct descendants of the previous populations, and no ma-



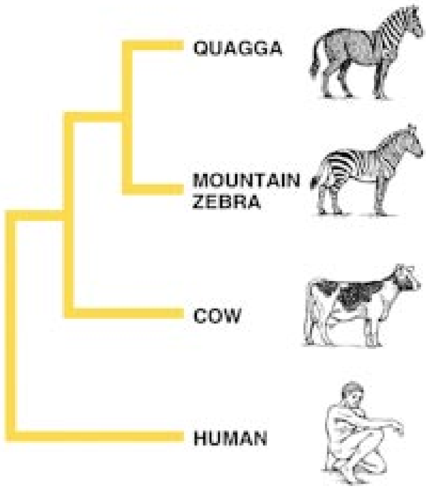
fers from all other vertebrates at one position (*red*), an error introduced by cloning and corrected by the polymerase chain reaction.

jor migrations or other disturbances have occurred in the intervening decades. This work demonstrated that museum collections could be used to monitor the genetics of a population over time.

The urgency of these developments stems from the realization that the study of both extinct and living populations is important now that so many species are in danger of extinction. It is already known, for instance, that as a population shrinks its genetic diversity declines, making it more susceptible to infections and other stress, as in the case of the cheetah [see "The Cheetah in Genetic Peril," by Stephen J. O'Brien, David E. Wildt and Mitchell Bush; Scientific American, May 1986]. Studies of ancient DNA may deepen our understanding of how a gene pool changes over time and may show whether a reduction in genetic diversity generally precedes extinctions.

At first, workers dared to seek ancient DNA only in soft tissues, which form but a small fraction of the remains turned up in archaeological excavations. Then, in 1989, Erika Hagelberg and Bryan Sykes of the University of Oxford succeeded in extracting usable DNA from ancient bones. Such DNA presumably stems from osteoblasts and osteoclasts, cells that reside within the compact bone and continuously remodel it throughout life. Perhaps because DNA binds to the minerals that make up bone, skeletal remains may actually surpass soft tissue as a repository for DNA.

A good example is provided by bones of moas, an extinct group of flightless birds of New Zealand. Some species grew to 3.5 meters in height and 200 kilograms in weight; they lived by browsing until humans arrived on the island about 1,000 years ago and hunted them to extinction. Alan Cooper of

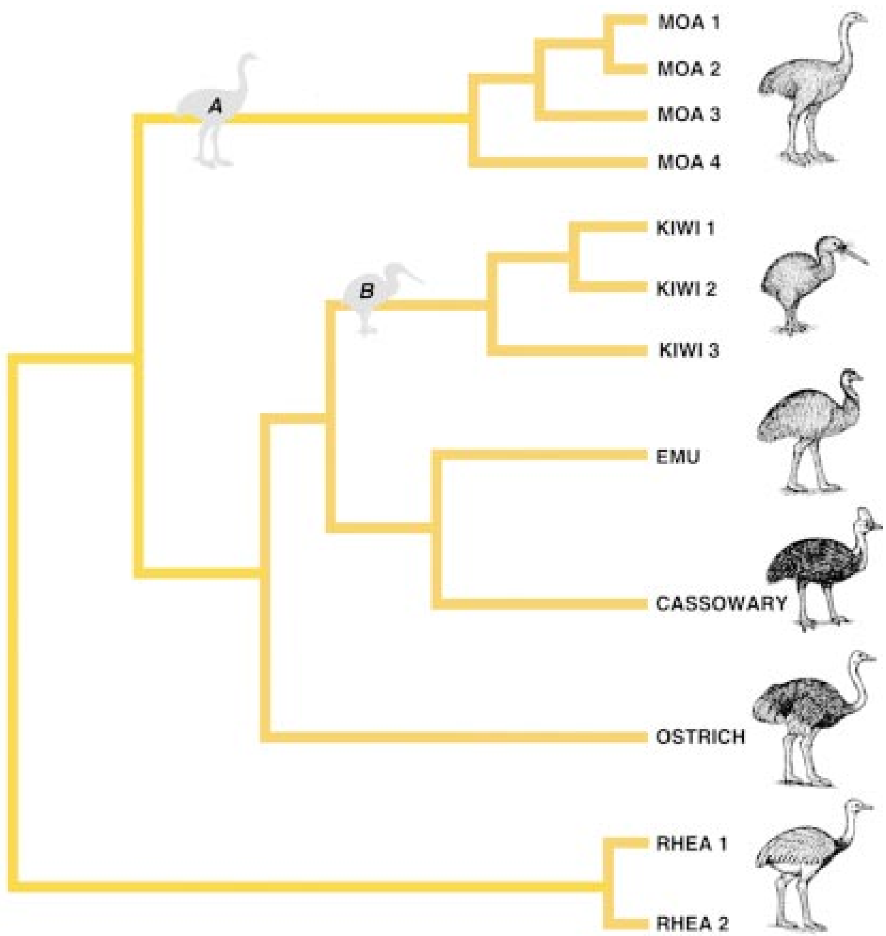


Victoria University in Wellington analyzed both the bones and the mummified soft tissues of moas found in caves and swamps. He managed to amplify mitochondrial DNA from four of the six genera that existed. The sequences reveal that the moas were all closely related to one another but not to the kiwis, a group of flightless birds that still exist on New Zealand. Kiwis are instead more closely related to the emus and cassowaries of Australia.

This discovery indicates that the kiwis came to New Zealand later than the moas did—probably after the island separated from Australia some 80 million years ago. Was the kiwi, like its Australian cousins, flightless when it came? If so, it must have swum to New Zealand, perhaps following a chain of islands that has since disappeared. Alternatively, it may have come by air, losing the power of flight only after arriving on a landmass where no ground-dwelling carnivorous animals lived.

Another insight emerging from the DNA sequences of the different genera of moas is that they diverged from one another rather recently. Cooper has suggested that this timing can be explained by the geologic history of New Zealand. Much of the island sank beneath the waves 30 million years ago, apparently driving most of the ancestral moas to extinction. Then, the rise of mountain ranges restored the island's surface area and provoked a burst of speciation in the survivors, which quickly occupied the new niches.

From how far back in time can we expect to retrieve DNA—from the progenitors of modern humans some 200,000 years ago, from the founders of the hominid line some



FLIGHTLESS BIRDS of New Zealand, the kiwi and the extinct moa, are not directly related. The moa ancestor probably arrived first (*A*), possibly when the island separated from Gondwana some 80 million years ago. The kiwi ancestor seems to have arrived later (*B*), perhaps by flying.

GENETIC TREASURE TROVE is provided by mummies from the Americas, Europe and Egypt. The author is shown holding the foot of an Egyptian mummy similar to that which inspired his first foray into the study of ancient DNA.

three million years ago or from the dinosaurs themselves? Certain physical limits seem inescapable. In approximately 50,000 years, water alone strips bases from the DNA and leads to the breakage of strands into pieces so small that no information can be retrieved from them. Oxygen also contributes to the destruction of DNA. Even in ideal conditions—in the absence of water and oxygen and at low temperature—background radiation must finally erase all genetic information.

Despite these considerations, recent results have encouraged hopes of going further into the past than many believed possible. In 1990 Edward M. Golenberg and Michael T. Clegg of the University of California at Riverside published a DNA sequence from a magnolia leaf deposited in clay on the bottom of a lake in northern Idaho some 17 million years ago. They were able to amplify a fragment as long as 800 base pairs. Pamela S. Soltis and Douglas E. Soltis of Washington State University repeated this feat with specimens of another plant species found at the site, but extracts from many other leaves from the same site have failed to yield DNA. The clay was wet, however, and one wonders how DNA could have survived the damaging influence of water for so long.

Perhaps more promising are the results of George O. Poinar, Jr., of Berkeley and Rob DeSalle of the American Museum of Natural History. They and their collaborators study insects that have been entombed in amber, a dry environment that may also be protected from oxygen if the amber is deposited under suitable conditions. They have extracted DNA from insects imprisoned in amber about 40 million years ago. DeSalle and his colleagues have shown that their sequences, from termites, are compatible with those from a modern termite.

The next few years will undoubtedly show much activity in this field. Above all, it will be crucial to find out whether other laboratories can confirm the results reported for the very old DNA sequences. If DNA sequences millions of years old can indeed be determined reproducibly, one of the most important opportunities that would result is the ability to measure the rate of molecular evolution directly. Yet some problems will remain. It will always be hard to tell whether an old specimen is truly an ancestor of a present-day species or whether the ancient and modern sequences merely shared an ancestor much further back in time. The practitioners of molecular evolution will thus still be able to disagree fruitfully.

If DNA provides a recipe for building an organism, and ancient DNA stores the recipe with fidelity, then a question naturally arises: Can we hope to reverse extinction by resuscitating vanished species? Could we clone identical twins of our recent and more distant ancestors, recall the moas or quaggas to life or even establish breeding farms for dinosaurs?

It is my firm conviction that such dreams (or nightmares) will never be realized. We have no idea how to piece together the millions of DNA fragments that we extract from an animal into chromosomes in a functional cell, nor can we set in motion the thousands of genes that regulate development. If we cannot even take a fresh cell from an adult vertebrate and use it to clone another individual, how can we imagine cloning an extinct species from the flotsam and jetsam of ancient DNA?

The resurrection of lost species will remain beyond our power. The furthest step toward "reanimation" that one can imagine would be the isolation of a single ancient gene. Such a gene, introduced into existing species, might create animals that mimic an aspect of an extinct species. In this way, one could test the function of an ancient gene. Yet such experiments do not in any sense preserve or re-create lost species or ecosystems. Extinction is and will always be forever.

What we can hope is that the study of ancient DNA will help us learn more about the dynamics of genetic change in populations over time. With such knowledge we may be able not only to understand better the history of ourselves and other species but also to frame more rational strategies to limit the ongoing erosion of biodiversity.

FURTHER READING

DNA SEQUENCES FROM THE QUAGGA, AN EXTINCT MEMBER OF THE HORSE FAMILY. R. Higuchi, B. Bowman, M. Freiberger, O. A. Ryder and A. C Wilson in *Nature,* Vol. 312, No. 5991, pages 282–284; November 15, 1984.

ANCIENT DNA: EXTRACTION, CHARACTERIZATION, MOLECULAR CLONING AND ENZYMATIC AMPLIFICATION. Svante Pääbo in *Proceedings of the National Academy of Sciences,* Vol. 86, No. 6, pages 1939–1943, March 1989.

INDEPENDENT ORIGINS OF NEW ZEALAND MOAS AND KIWIS. A. Cooper, C. Mourer-Chauviré, G. K. Chambers, A. von Haeseler, A. C. Wilson and S. Pääbo in *Proceedings of the National Academy of Sciences,* Vol. 89, No. 18, pages 8741–8744; September 15, 1992.

DNA SEQUENCES FROM A FOSSIL TERMITE IN OLIGO-MIOCENE AMBER AND THEIR PHYLOGENETIC IMPLICATIONS. R. DeSalle, J. Gatesy, W. Wheeler and D. Grimaldi in *Science,* Vol. 257, pages 1933–1936; September 25, 1992.

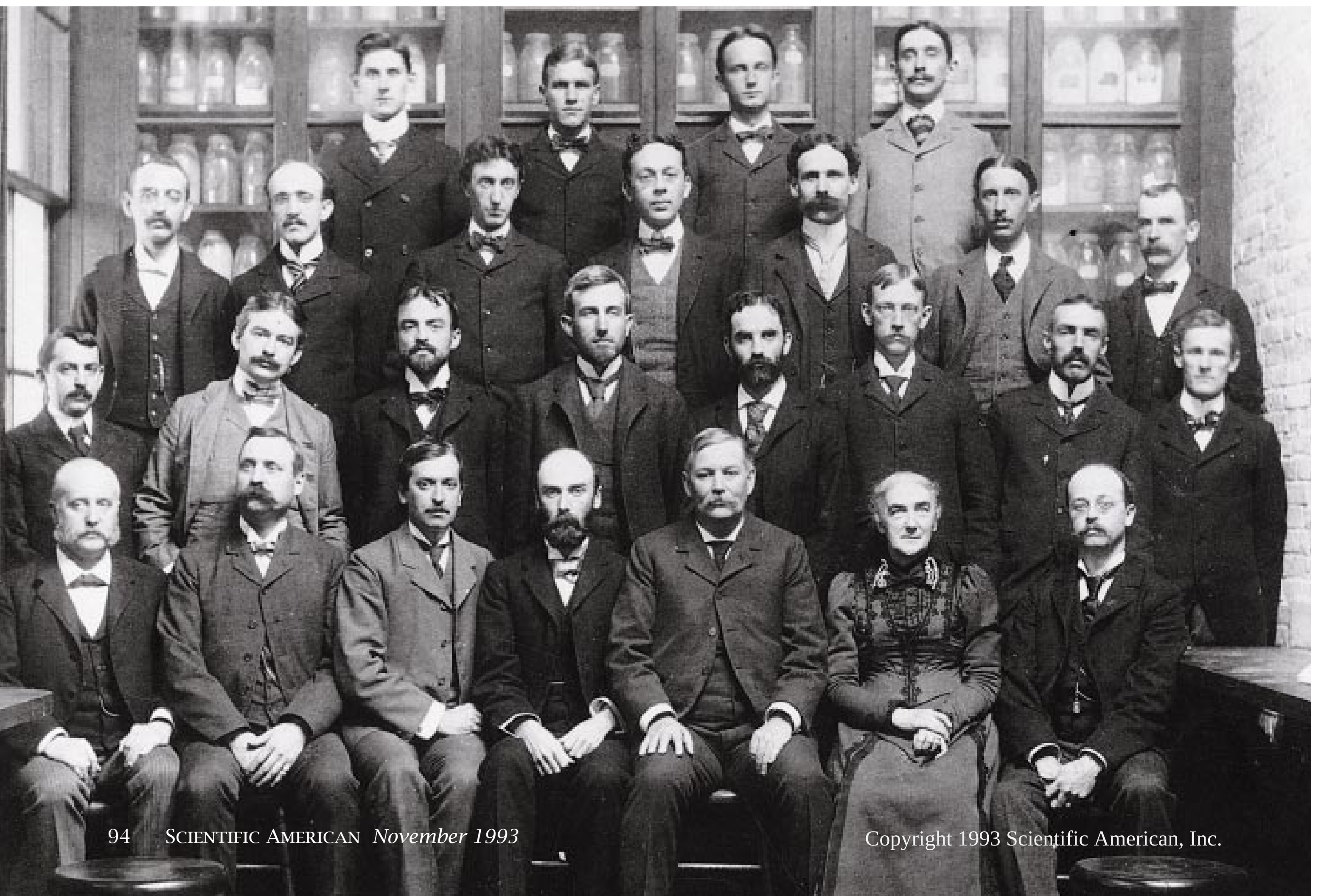
A Lab of Her Own

by Marguerite Holloway, *staff writer*

There is no one story to tell about women in science. C. Dominique Toran-Allerand just received tenure—after 20 years at Columbia University, after watching male peers enjoy promotion, after listening to colleagues laugh when she requested recommendations. "They thought I was joking," explains the neuroscientist, who studies the role of hormones in brain development, with light bitterness in her voice. "People generally did not believe I did not have tenure."

Cheryl Ann Butman, a tenured biological oceanographer at the Woods Hole Oceanographic Institution, moves clothes at midnight from a backpack, unemptied since her return from a Gordon Conference, into a canvas bag. In four hours she will leave on a cruise to place research equipment on the ocean floor. Downstairs, Bradford Butman, branch chief of the U.S. Geological Survey in Woods Hole, washes dishes. He is just back from a meeting. The race against time is interrupted when Dylan, their two-year-old, has a nightmare. The evening is remarkable only in that both scientists are home. "Once Brad met me at the airport, handed Dylan to me and then got on a plane himself," Butman recalls.

Kay Redfield Jamison, a psychiatrist at the Johns Hopkins University School of Medicine who studies creativity and manic-depressive illness, would rather not talk about problems that women may encounter. "The system is a harsh one, but it is for men as well," she asserts. "In the end, you just have to get your work done. How many women really spend much time thinking about these things?"



Despite decades of struggle, women remain a small minority in the scientific community

Some find it hard to avoid doing so. A researcher at a prestigious women's college describes being told to "go knit or do whatever it is you women do" when she asked for comments on her grant application to the National Institutes of Health.

The experiences of these scientists and the challenges they face are as varied as the women themselves and as the research they do. Which is perhaps why the fight that women wage so that they and their daughters can practice science remains unfinished. Although their struggle to enter and to advance in this overwhelmingly male-dominated field parallels the struggles of women in other professions, science seems a uniquely well fortified bastion of sexism. "How shocking it is that there are any women in science at all," remarks Sandra Harding, a philosopher at the University of Delaware.

Despite speeches, panels and other efforts at consciousness-raising, women remain dramatically absent from the membership of the informal communities and clubs that constitute the scientific establishment. Only 16 percent of the employed scientists and engineers in this country are female. At a finer level of detail, the numbers of women in dif-

A FACE IN A CROWD characterizes the situation of many female scientists. Ellen Swallow Richards was the first woman on the faculty of the Massachusetts Institute of Technology (opposite page). *She is shown in 1900 with her chemistry department colleagues. Today there are more women in science, as these students on the steps of the M.I.T. library illustrate* (this page). *But they make up only 16 percent of U.S. working scientists.*

Biological oceanographer at the Woods Hole Oceanographic Institution in Massachusetts. Butman studies the physical dynamics of organisms—such as examining how the formation of stacks of mussels improves their ability to feed.

CHERYL ANN BUTMAN

Pediatrician who recently became U.S. surgeon general. Elders was formerly the director of the Arkansas Department of Health.

JOYCELYN ELDERS

Biochemist at Burroughs Wellcome Company. Elion won the 1988 Nobel Prize in Physiology or Medicine with her colleague George H. Hitchings for their work on compounds that led to the development of drugs to treat leukemia, organ transplant rejection, malaria, gout and herpesvirus infection.

GERTRUDE BELLE ELION

Some of the Women in Science Today

ferent disciplines and positions are so low that a recitation of the statistics sounds like a warped version of "The Twelve Days of Christmas": 1 percent of working environmental scientists, 2 percent of mechanical engineers, 3 percent of electrical engineers, 4 percent of medical school department directors, 5 percent of physics Ph.D.'s, 6 of close to 300 tenured professors in the country's top 10 mathematics departments, and so on.

"There is still so much to be done," rues Jane Z. Daniels, director of women's programs at the National Science Foundation (NSF). "The traditional areas of science for women are still those areas where there is the most growth. There is not a lot of change in physics, geology and engineering. Those are the ones where the stereotypes have been preserved." Other fields are not quite so male heavy. Forty-one percent of working biologists and life scientists are women. Nearly half of all psychology and neuroscience graduate students are female. According to the American Chemical Society, women constituted 17 percent of their members in 1991, up from 8 percent in 1975.

Regardless of their field, women scientists typically earn salaries that are about 25 percent lower than those paid to men in the same positions, they are twice as likely to be unemployed and they are rarely promoted to high positions (in 1989, 7 percent of tenured faculty in the sciences were female). Women report less encouragement from their peers and supervisors, less mentoring and help with professional advancement as well as greater isolation and harassment.

These conditions persist despite more than two decades of efforts to redress an imbalance that was brought to light in large part by the women's movement. In the past 20 years an array of federal and other educational programs have sought to attract women into science. These attempts gained some momentum in 1988, when a congressional study announced that the U.S. would need more than half a million scientists and engineers by the year 2010. As men were dropping out of science, women and members of minority groups were seen as possible replacements.

The cumulative attention has brought about some gains. In 1989 women received 27.8 percent of the doctorates in science and engineering, whereas in 1966 only 8 percent of such degrees were awarded to women. The NSF recently found that differences in science scores between girls and boys on some standardized tests had decreased. The U.S. Equal Employment Opportunity Commission has also documented an increase in the number of female full professors.

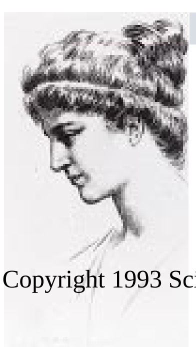
"I have never seen a period in history where they are trying to encourage women so much," notes Londa Schiebinger, a historian of science at Pennsylvania State University. "But I think what is extremely interesting is that there is all this funding and this goodwill, and they are still dropping like flies." Attrition has increasingly led many observers to examine the culture of science for clues about why so few women stay in the field. What, if anything, ask the researchers, is it about science that continues to exclude or deter women from remaining in research?

"They have been attempting to get more women into science, trying to fix the women, give them enough science courses, prevent them from falling behind," Schiebinger, who wrote *The Mind Has No Sex? Women in the Origins of Modern Science.* "But we can't fix the girls, we have to fix science, get it to be something they want to do. We have to look deeply into the culture of science and see what is turning women off."

Peering into the scientific establishment to pinpoint the origins of the problem—why so few women?—reveals both the mysterious and the obvious. Throughout the centuries, for no cogent reason, women have been excluded from most aspects of professional and political life. And the majority of fields have until recently remained male. Within this larger tradition of sexism, there are some clear explanations for the absence of women in science. From the moment they begin to be socialized, most girls are directed away from science. This subtle and overt deterrence can been seen in the educational system and is fortified by the perceptions of many male scientists that women simply should not be scientists.

It is not that there have historically been no women in science. Only nine women may have been awarded a Nobel Prize as opposed to more than 300 men, but there are many unsung women who have made vital contributions in all fields. In the past decade or so, historians have increasingly begun to describe these mostly invisible participants. In 1982 Margaret W. Rossiter, a historian of science at Cornell University, published a lengthy account of American women who did science before 1940. "People said the book would not be very long, because there were no women of consequence. They were wrong," says Rossiter, who is working on her next tome: women in science from 1945 to 1972. Although many workers were tucked away as assistants and technicians, their

Some of the Women in the History of Science



HYPATIA

Circa 370–415
Egyptian mathematician, teacher and philosopher who was murdered by a group of monks. One legend has it that these holy men resented the fact that a woman was lecturing.

MARIA SIBYLLA MERIAN

1647–1717
German biologist who extended the field of entomology through her observations and illustrations of the life cycle of caterpillars and butterflies. She supported herself by publishing books and by designing fabrics.

SOPHIE GERMAIN

1776–1831
Self-taught French mathematician and physicist who produced original work in number theory and the theory of elasticity. Germain was excluded from the male scientific community and received recognition for her work only late in life.

ANNE FAUSTO-STERLING

Professor of biology and medicine at Brown University and feminist scholar. Fausto-Sterling has written extensively about the biology of sex differences and is currently doing research on *Planaria.*

MARGARET GELLER

Astronomer at the Harvard-Smithsonian Center for Astrophysics. Geller and her colleague John P. Huchra discovered the Great Wall of galaxies, a structure that runs for three billion trillion miles and contains 1,700 galaxies.

DEBORAH M. GORDON

Entomologist at Stanford University. Gordon, who does research on ants, is one of the few women studying social insects. She is currently observing how information is passed from generation to generation in harvester ant colonies in Arizona.

contributions were invaluable. She found many of them hidden in footnotes in books about male scientists.

Other researchers have traced the roots of the scientific establishment's attitude toward women. Each period of history and each culture are, of course, characterized by a different prevailing view, but there is no shortage of "documentation" by males of the physical and mental inferiority of women. In the late 1880s, following a series of studies on the small size of women's brains—and, not insignificantly, their enormous pelvic bones, all the better to bear children with—a friend of Charles Darwin's summed up that illustrious scientist's view of women's intellectual powers: "It must take many centuries for heredity to produce the missing five ounces of the female brain."

The emergence of the modern scientific establishment appears to have institutionalized many of these perceptions. Historian David F. Noble of York University in Toronto argues that the first universities were monastic, organized by the Christian church, and thus excluded women. In his book *A World without Women: The Christian Clerical Culture of Western Science,* he discusses how this segregation persisted in the academies and institutions that arose with modern science. The Royal Society was established in 1662 and did not admit women until 1945. Before then, as Schiebinger notes, the only woman in the Royal Society was a skeleton in the anatomy collection. Today 2.9 percent of the "fellows" are female.

Some institutions have better records, but by and large, women were not made to feel at home in the inner sanctum of science and were denied access to traditional training. Beatrix Potter, for instance, was an accomplished mycologist—in fact, she was the first person to report on the symbiotic aspects of lichen and to catalogue the fungi of the British Isles. But Potter was not allowed to join any professional scientific societies because of her sex. So, fortunately for English-speaking children and their parents, she turned to writing and illustrating children's books. The 1880 official minutes from "The Misogynist Dinner of the American Chemical Society," unearthed by Rossiter, are part of the same tradition.

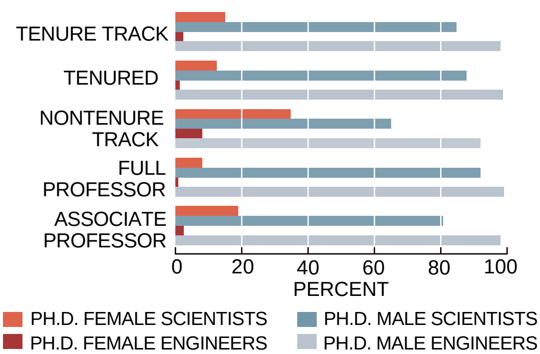
It is the vestiges of these attitudes and the impenetrability of the elite social institutions that most frustrate female scientists still. The National Academy of Sciences currently has only 70 female members, out of 1,750 living scientists. "There is still resentment between the old guard and women," says Betty M. Vetter, executive director of the Commission on Professionals in Science and Technology. She adds bluntly, "It will change when they die."

By maintaining a male majority, many institutions perpetuate the status quo, preventing women from participating in forums where important contacts are made. Women "don't get invited to write as many book chapters, and they don't get a chance to network as much. It is not a question of a more or less collaborative style," comments Christina L. Williams, a neuroscientist at Barnard College. "You do what you can do. You can't get yourself invited to things if you don't get invited."

Studies have found that meetings organized by men usually have a male majority—no matter what the percentage of women in the field. Only 24 percent of the speakers at past meetings of the American Society for Cell Biology, which is roughly 50 percent female, were women, even when the conferences were organized by women, notes Susan Gerbi, president of the society. When men organized the conferences, less than 10 percent were female. "It is not men sitting around saying, 'Don't invite women,'" Williams explains. "It is done blindly, and it is just that there is no concerted effort. A lot of the people in power in science are still men."

Another place where similar discrimination may occur is on editorial boards. Staff at many scientific publications remains mostly male and has shown a tendency to accept more male-authored papers or to invite men to do review articles. It is not clear, however, that selection of papers would change

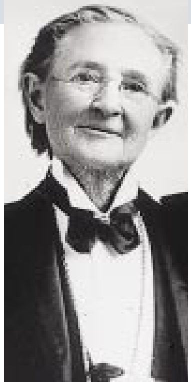
Academic Rank of Doctoral Scientists 1990



SOURCE: National Science Foundation; the statistics are for U.S.

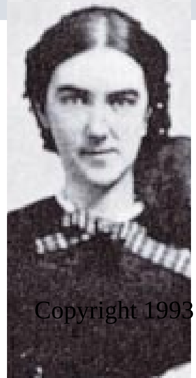
MARIA MITCHELL

1818–1889
Established the Vassar College Observatory in the U.S., one of the earliest and most important astronomy programs for women. In 1847 Mitchell, who learned astronomy from her father and her own reading, received widespread acclaim for the discovery of a comet.



MARY EDWARDS WALKER

1832–1919
Surgeon and feminist who worked as a nurse and, later, as the first female assistant surgeon in the American Civil War. Walker adopted male dress for her work in the field.

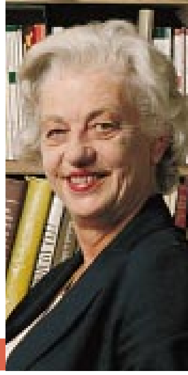


ELLEN SWALLOW RICHARDS

1842–1911
Engineer lauded as the "woman who founded ecology." Richards, denied a deserved Ph.D. in chemistry at the Massachusetts Institute of Technology, was the first woman to be elected to the American Institute of Mining and Metallurgical Engineers.

Professor of computer science at Harvard University. A pioneer in the subdiscipline of artificial intelligence known as natural language processing, Grosz works on ways to make computers easier for people to use by incorporating features of human dialogue into computer systems.

BARBARA GROSZ



Chairperson of the department of anatomy at Harvard Medical School. Hay studies the regeneration of cells and tissues. She made some of the first electron micrograph autoradiographs, a substantial contribution to the study

ELIZABETH DEXTER HAY

Microbiologist at Harvard University. Huang studies the replication of RNA animal viruses.

ALICE S. HUANG

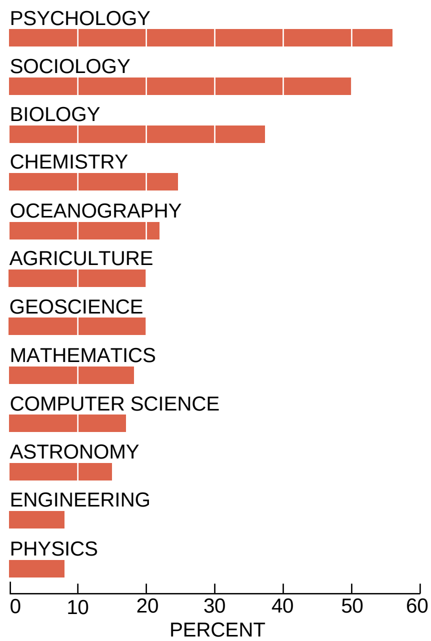
if editorial boards were more sexually balanced. A study conducted in the early 1980s asked 180 men and 180 women to rate comparable papers. One third of the papers was supposedly written by John T. McKay, another third by Joan T. McKay and the final series by J. T. McKay. Both the women and men gave the "John T." papers the highest score. Whatever the cause, Harriet Zuckerman of the Andrew W. Mellon Foundation and Jonathan R. Cole of Columbia have found that women tend to publish 30 percent fewer papers than do their male colleagues in the first 12 or so years of their careers. The disparity increases over time.

One controversial solution to making meetings more reflective of the work force, thereby spreading the wealth of information and contacts, is affirmative action. Last year the NSF announced it would not fund conferences unless a number of women proportionate to the number in the field were invited. "You hope it is not going to lead to less qualified women being asked," Williams says. "But there is no reason that it should. There are plenty of good women out there in all fields."

Opinions about affirmative action are, inevitably, mixed. An editorial in *Nature* bemoaned the NSF's new "quota" policy. "There is no evidence that sex is related to success in scientific research," the editors wrote, "and no inherent justification for holding women out for special treatment as part of a formal policy carrying the bludgeon of budgets." Many female scientists also view legislative remedies with some skepticism. "I personally do not want any favors because I am a woman. I want to be competitive on a gender-free basis," Toran-Allerand says. Her view echoes that of many female scientists, in particular those who struggled through the system before it was subjected to feminist scrutiny. Many of those who succeeded, including Nobel laureates Gertrude Belle Elion and Rita Levi-Montalcini, did not want special attention as women in science—they just desperately wanted to do their science.

Toran-Allerand, who was the only woman in both her medical school class and residency, attributes some of this to a double standard for women. "In the past, women were really an intellectual elite. You had to be slightly crazy if you wanted to go do that in that kind of environment," she comments. A woman "who interviewed me at Yale said I had to realize that the women had to be perfect. There were so few women; they could not tolerate any imperfection. The imperfections in the men would be accepted because there were so many of them that they would even out over the population."

Doctorates Awarded to Women, by Field in 1989



SOURCE: National Science Foundation; the statistics are for U.S.

With more women in science, such pressures have been alleviated—to a point. Many scientists and educators have noted that scientific institutions are not the only source of discouragement for women: the educational system does not foster a love of science in girls (for that matter, however, it has not been wildly successful in recent years with boys either). Most teachers of kindergarten through eighth grade are women, and many are not well versed in science. They do not serve as effective role models for young girls interested in science. In addition, many stereotypes—of scientists as nerds, as mad and as male—persist. "The basic idea is that if you are a woman interested in science, you are gender confused," notes Catherine J. Didion, executive director of the Association for Women in Science.

Research by the American Association of University Women has found that at all educational levels, boys receive more attention than do girls in the classroom. The effect is independent of the teacher's sex. Adults also encourage boys to be assertive in answering questions and expressing opinions. Therefore, a young woman who pursues a career in science needs a particularly strong endowment of mettle.

SOFIA KOVALEVSKAIA

1850–1891
Russian mathematician who did work on partial differential equations. She is thought to be the first woman to receive a doctorate in mathematics, from the University of Göttingen in 1874.

MARIE S. CURIE

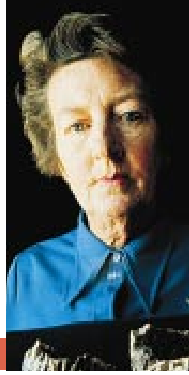
1867–1934
French scientist who discovered radium and polonium. She shared the 1903 Nobel Prize in Physics with her husband, Pierre Curie, and Henri Becquerel. In 1911 she won the Nobel Prize in Chemistry.

FLORENCE RENA SABIN

1871–1953
Medical researcher who studied the development of the lymphatic system and, later, tuberculosis. She fought to modernize public health laws in the U.S. Sabin was the first woman to be elected to the National Academy of Sciences.

Professor emerita of biology at Harvard University. Hubbard did research on the photochemistry of vIsion early in her career and, more recently, has written extensively about women in science, women's health issues and genetics.

RUTH HUBBARD



Paleoanthropologist who, with her late husband, Louis Leakey, unearthed and analyzed fossil hominids in Africa. Leakey now lives in Nairobi, Kenya.

MARY LEAKEY

Neuroscientist who won the Nobel Prize in Physiology or Medicine in 1986 for her work on nerve growth factor. Levi-Montalcini teaches and continues her research at the Institute of Neurobiology at the National Research Council in Rome.

RITA LEVI-MONTALCINI

Girls are told in myriad ways that they are not as good at mathematics as boys are. This social myth has no foundation in reality. Researchers have found that girls often do as well as boys in math in elementary and junior high school. Yet girls hear "quite early that higher math is for boys," Vetter notes. "Girls are not taught to put themselves forward to get into that group of precocious math kids. You have to push yourself forward, but girls are not encouraged to do that."

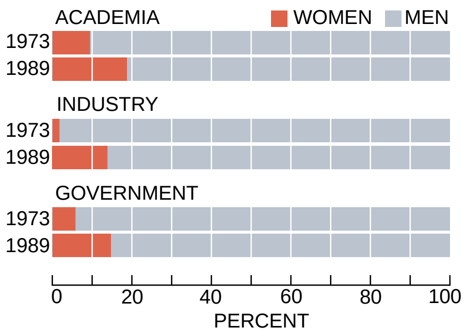
Exploiting this perception of feminine math anxiety, the toy manufacturer Mattel last year made a Barbie doll that said, "Math class is tough." The company deleted the statement from the doll's voice track after several women's groups protested. (The NSF's Daniels points out that the pink Lego building blocks designed for girls do not send the right message either.)

A distinct irony surrounds the issue of women and mathematics. Mathematics was at times considered a woman's subject. Schiebinger describes the English *Ladies' Diary,* published between 1704 and 1841, which encouraged women to perfect their "Arithmetick, Geometry, Trigonometry… Algebra… and all other Mathematical Sciences." It goes to show that "when the rules of society change, the girls perform just as well as the boys," remarks Mildred S. Dresselhaus, professor of electrical engineering and physics at the Massachusetts Institute of Technology. "If they act as though they are interested, they get very discouraging signals. I got my share of those, too, I suppose. But I went to an all-girls school, and there I did not know that girls were not supposed to study math."

In addition to discouragement, women cite boredom as the reason that they stopped studying science. Many experts are trying to find new ways of teaching girls and women to maintain interest. Sue V. Rosser, director of women's studies and professor of family and preventive medicine at the University of South Carolina, has found that women tend to be interested in a problem or a question if it has some context or social relevance or the solution produces some benefit. They also respond to a challenge better if the process of meeting it is framed as a collaboration rather than a competition. "Men, in general, find that a technological fix in and of itself is enough," she explains. Rosser and many others have designed successful teaching methods that harness these insights. Re-forming questions and experiments appears to have an unexpected boon: it captures the imagination of male students as well.

An NSF study of questions for National Assessment of Educational Progress tests reached the same conclusion. When math problems have some social implication, girls do better. On the other hand, boys' scores on tests of verbal ability, which are traditionally lower than those of girls, improve if the excerpts describe sports or science. Ellen Spertus, a graduate student in computer science at M.I.T., observes that computer games in which the objective is, say, to prevent a meteor from hitting the planet are often more likely to interest girls than are games in which the players are supposed to slaughter invading aliens.

Science Ph.D's, by Employment Sector



SOURCE: National Science Foundation; the statistics are for U.S.

Changes in testing and in school curriculums, however, may not be sufficient to hold women in science. Sometimes as many as half of the first-year female college students are interested in science and engineering, yet at some point in their academic careers, their attrition rate exceeds that of the male students. Certain universities and colleges as well as the Association for Women in Science have sought to combat this tendency by establishing mentoring programs.

In 1990, for instance, Dartmouth College set up internships to give as many as 75 female students experience in a laboratory, to demystify science and to introduce them to scientists. "They get to see what is going on in science firsthand, that scientists are not all geeky, that they are very regular people who make mistakes and have to do things over again," according to Mary Pavone, director of the women in science project.

Initially, many of the participants do not think the program is necessary. "They come here freshman year and see that the numbers of men and women in introductory science courses

LISE MEITNER

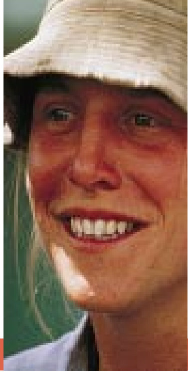
1878–1968
Austrian-Swedish physicist and mathematician who studied the decay of radioactive elements. She was the first person to calculate the energy released during nuclear fission and thereby contributed to the development of the atomic bomb.

MARGARET SANGER

1883–1966
Nurse who became the leader of the campaign for family planning and birth control in the U.S. Her efforts to establish birth-control clinics and to disseminate information were the subject of great controversy.

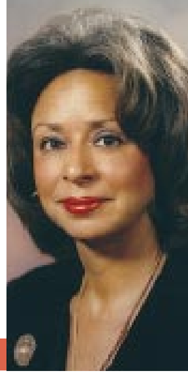
ETHEL BROWNE HARVEY

1885–1965
American biologist and embryologist whose studies of induction preceded those of Nobel laureate Hans Spemann and Hilde Mangold by more than 10 years. An investigator at Princeton University for 25 years, she was never made a full professor.



Behavioral ecologist with NYZS The Wildlife Conservation Society. Moehlman, who works mostly in the field in Africa, has studied and filmed free-roaming wild animals for over two decades. In recent years, she has turned her attention increasingly to advising about wildlife management.

PATRICIA MOEHLMAN



First director of the Office of Research on Women's Health at the National Institutes of Health. Pinn is a renal pathologist and was professor and chairperson of the department of pathology at Howard University before she joined the NIH.

VIVIAN PINN

Astronomer at the Carnegie Institution of Terrestrial Magnetism. Rubin has worked for more than 25 years with collaborator W. Kent Ford on the existence of dark matter and galactic rotation.

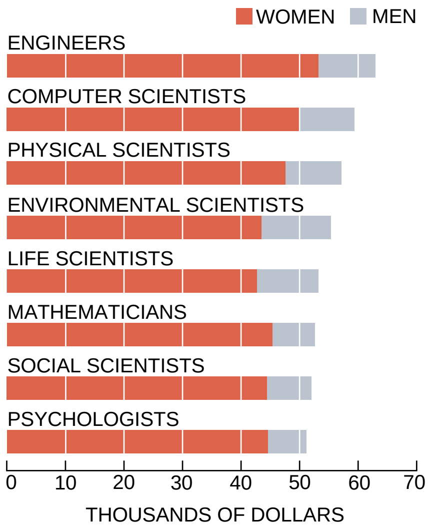
VERA RUBIN

are fairly even—they don't see what happens; they don't see the filter. By junior year, they look around their classes, and all of a sudden a light goes on," Pavone says. The number of women majoring in science at Dartmouth was up in 1993, but it is not clear that the project is responsible.

At the doctoral level the situation becomes more difficult. Women have a higher attrition rate than do men before they enter Ph.D. programs; they are about 15 percent less likely to finish their degrees. "You have to have someone on the faculty who wants you," says C. Megan Urry, chief of the research support branch at the National Aeronautics and Space Administration's Space Telescope Science Institute. Science is ultimately a guild, in which a master passes on skills and professional touch to apprentices. For reasons of ancient tradition and contemporary culture, those apprentices are predominantly male. "No one ever told me what was going on. The men are getting a lot of help, and the male advisers are helping them write. The women don't get it much," Urry says.

A combination of institutional changes, including mentoring programs, educational reforms and affirmative action strategies, has traditionally been perceived as the means for bringing women into science and keeping them there. These approaches address the problem illustrated by the often strange metaphors that have been used to explain why there are so few women in science or in any other field: the pipeline is leaking, the glass ceiling has not cracked, women are stuck on the bottom rung of the ladder.

Median Annual Salaries at the Doctoral Level, 1989



SOURCE: National Science Foundation; the statistics are for U.S.

But a growing number of observers are questioning the fundamental and long-term success of these efforts. Feminist thinkers, including Schiebinger and Rosser as well as Brown University biologist Anne Fausto-Sterling and Harvard University professor emerita Ruth Hubbard, take a more radical position. They believe the whole edifice—plumbing, ceiling and ladder—has to be reconstructed. "My view is that getting women into science is not just being nicer to them at younger ages, although that is important. But we really have to rethink our whole notion of what science is and how it functions," says Fausto-Sterling, author of *Myths of Gender: Biological Theories about Women and Men.*

Fausto-Sterling and others are examining how scientific knowledge in the West has been shaped by social mores and by the white male culture that has directed it. The scrutiny is not well received by many in the scientific community. "Scientists think this is not very important," says Harding, who wrote *Whose Science? Whose Knowledge?* "But our conceptions of how we think about the history of science shape how we are doing science now. We want to learn from the past. If we have distorted views, we should understand them."

Perhaps the most prickly issue that some of these thinkers have raised is whether women and men approach science differently and, if they do, whether differences in style account for the low numbers of women attracted to science. Most of the discussion was initiated by Evelyn Fox Keller's 1983 book about Nobel laureate Barbara McClintock. In *A Feeling for the Organism: The Life and Work of Barbara McClintock,* Keller suggests that McClintock's unusual insights into genetics were shaped by intuition, by a more stereotypically "female" approach. The fires have been stoked by other researchers, among them Doreen Kimura, a psychologist at the University of Western Ontario. Her work shows distinctly different patterns of male and female mentation with respect to solving problems and framing intellectual challenges [see "Sex Differences in the Brain," by Doreen Kimura; SCIENTIFIC AMERICAN, September 1992].

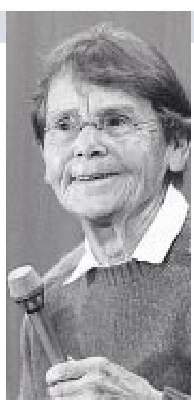
Many scientists think this idea of difference has some validity in the biological sciences in particular. "There is a strong argument that when you bring women in, they look at what the female [subjects] are doing," Schiebinger notes. "So far we have found these examples only for sciences

GERTY RADNITZ CORI

1896–1957
Biochemist who won the Nobel Prize in Physiology or Medicine in 1947 with her husband, Carl Cori, for their work on how cells use and convert food into energy—a process now called the Cori cycle.

IRÈNE JOLIOT-CURIE

1897–1956
Won the 1935 Nobel Prize in Chemistry with her husband, Frédéric Joliot-Curie, for their synthesis of new radioactive elements.



BARBARA McCLINTOCK

1902–1992
Geneticist who revolutionized the field through her observations of jumping genes. McClintock's novel ideas were not accepted for many years. In 1983, however, she won the Nobel Prize in Physiology or Medicine.

Mathematician at Rutgers University and the University of Minnesota's Geometry Center. Taylor studies soap bubbles and crystals and simulates them on computers in order to understand their underlying mathematical properties.

JEAN E. TAYLOR

Secretary of the Air Force. Widnall is on leave from M.I.T., where she is associate provost and professor of aeronautics and astronautics. She has served as president of the American Association for the Advancement of Science.

SHEILA WIDNALL

Mathematician at Wesleyan University in Connecticut, who recently served as the president of the Association for Women in Mathematics. She does work on logic, specifically model algebraic theory.

CAROL WOOD

where there is sex involved." Perhaps the best example of this view is the work of Jane Goodall, Dian Fossey and Birute Galdikas, anthropologists who revolutionized understanding of the primates by changing the way animals were observed, by following individuals. "They looked at female-female interactions and saw new behaviors," Rosser explains.

Rosser has her own example. She recalls that when she first taught animal behavior she asked the class to examine Siamese fighting fish: What were the reactions of males to males, to self and to females? The exercise never included female reactions to females or to males. Sandra Steingraber, a Bunting Fellow at Radcliffe and Harvard, studied dioramas of white-tailed deer in natural history museums and found that the males were always depicted in a warriorlike stance, about to defend a doe and fawn. In reality, Steingraber says, does and bucks unite only to mate. Does and fawns stay together only until they begin to compete for food. The dioramas, an educational tool, were shaped by the anthrocentric and anthropomorphic social vision of the men who designed them.

If it is true that women can bring a different perspective, feminist scholars argue, that is all the more reason to encourage women, minorities and people from diverse cultures to practice science. "I think there is a lot of validity to the idea that women do things differently, not from a biological basis but from a sociological perspective. There is a clash between women's and men's cultures," Schiebinger says. "Women and men are not interchangeable parts. They act in very different ways, and it seems to me that that carries over into the professional world. It brings an enriching perspective."

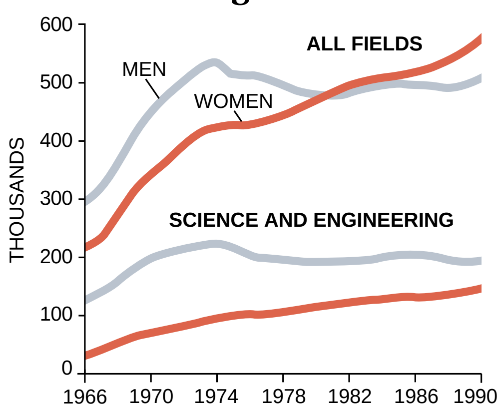
Anecdotal reports suggest that many women organize their laboratories differently, in a less hierarchical fashion, than do their male colleagues. "I don't think I think differently in terms of questions, because I have been trained," says Kathie L. Olsen, a program director at the NSF. "The difference is in the daily operation in the laboratory and in terms of how I interact with people." Ruth Ginzberg, a philosopher at Wesleyan University in Connecticut, has observed the same phenomenon in other fields, such as business. "For a long time, women were not thought of as good managers. Then somebody decided that perhaps women might have a different management style. Women were not rising in the ranks, because they were doing things differently—not because they were doing it less well."

Other differences have also been found. Studies of men and women interacting in groups suggest that women are interrupted more frequently, that their contributions are more often attributed to men in the group and that they are less comfortable with antagonistic discussions. "The problem is that women are being judged by men in a system set up by men that basically reflects their standards and criteria," Urry maintains. "Some of that has not to do with excellence in science but with style."

Whereas many female scientists agree that women and men may act differently, the idea that this variation translates into a different way of doing science remains sticky. "I find this topic a bit difficult," Dresselhaus admits. "Spending a lot of time on this doesn't do credit to women in fields with few women. In medicine, say, gynecology, women may have a different approach. But if you are solving a flow equation, there is not a woman's way or a man's way: there is the way the air flows around an airplane wing—it just flows around the wing."

There is a lack of very solid evidence for the proposition, Zuckerman concurs. In an unpublished study the sociologist and her colleagues found that sex differences were minor with respect to how scientists think about and describe their work. The criteria were how they chose their research topics and the significance of the research they were doing. "Gender is not a good predictor of difference," Zuckerman says. "Science is supposed to be attentive to evidence, and there is a lack of it here. These are matters about which people feel very strongly."

Bachelor's Degree: 1966–1990



SOURCE: National Science Foundation; the statistics are for U.S.

MARIA GOEPPERT MAYER

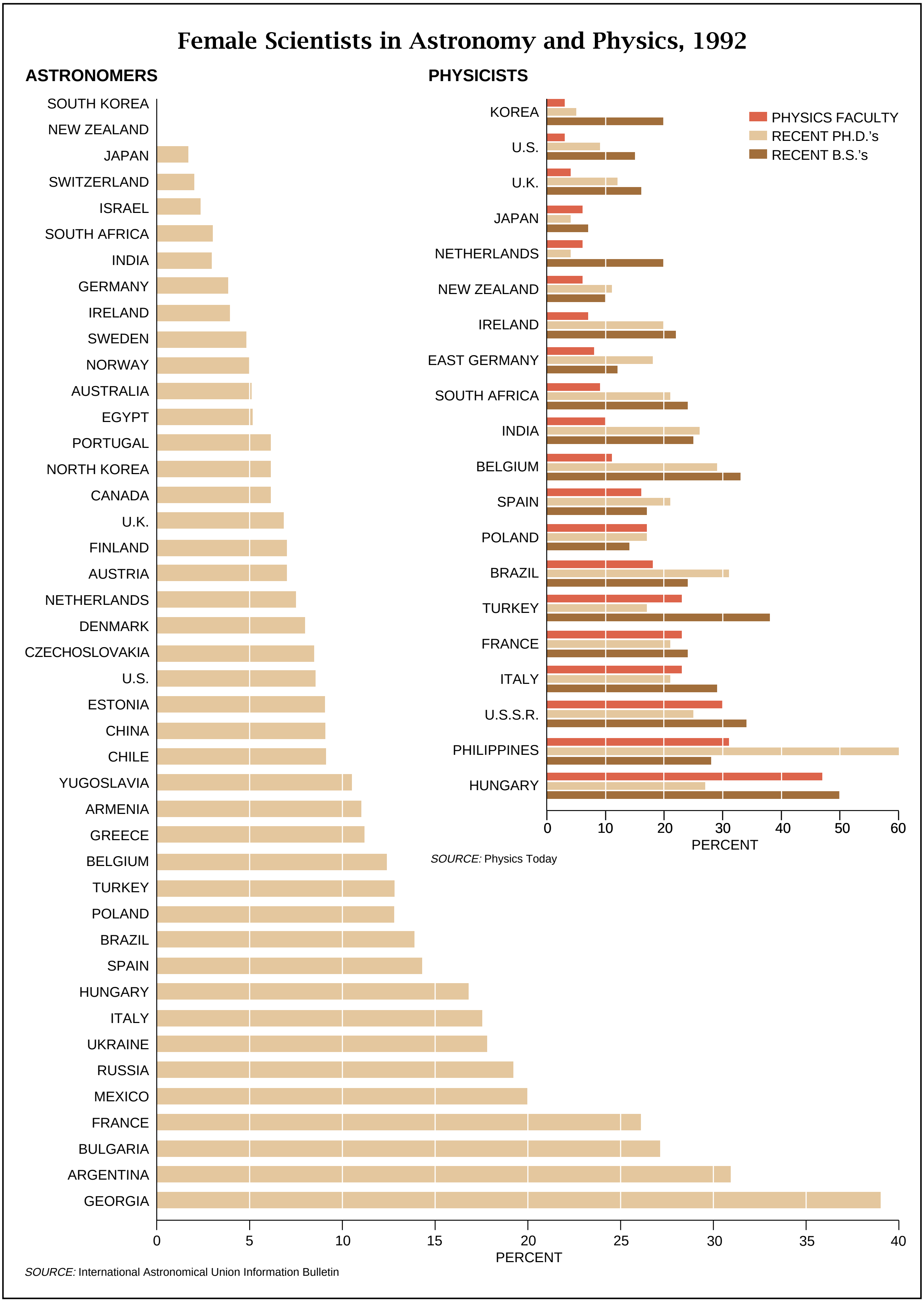
1906–1972
Mathematical physicist from Germany who won the 1963 Nobel Prize in Physics for her discovery of nuclear shells—the discrete energy levels that neutrons occupy— which Mayer described as similar to layers of onion skin.

RACHEL LOUISE CARSON

1907–1964
Marine biologist and author of several books, including *Silent Spring.* Carson's work alerted the scientific community and the public to the dangers of pesticides and to potentially destructive interactions between people and the environment.

ROSALIND ELSIE FRANKLIN

1920–1958
X-ray crystallographer. Franklin studied the structure of DNA and provided information needed by James Watson and Francis Crick, who later won a Nobel Prize, to describe the molecule as a double helix.


Female Scientists in Astronomy and Physics, 1992
ASTRONOMERS
SOUTH KOREA
NEW ZEALAND
JAPAN
SWITZERLAND
ISRAEL
SOUTH AFRICA
INDIA
GERMANY
IRELAND
SWEDEN
NORWAY
AUSTRALIA
EGYPT
PORTUGAL
NORTH KOREA
CANADA
U.K.
FINLAND
AUSTRIA
NETHERLANDS
DENMARK
CZECHOSLOVAKIA
U.S.
ESTONIA
CHINA
CHILE
YUGOSLAVIA
ARMENIA
GREECE
BELGIUM
TURKEY
POLAND
BRAZIL
SPAIN
HUNGARY
ITALY
UKRAINE
RUSSIA
MEXICO
FRANCE
BULGARIA
ARGENTINA
GEORGIA
0
5
10
15
20
25
30
35
40
PERCENT
SOURCE: International Astronomical Union Information Bulletin
PHYSICISTS
KOREA
U.S.
U.K.
JAPAN
NETHERLANDS
NEW ZEALAND
IRELAND
EAST GERMANY
SOUTH AFRICA
INDIA
BELGIUM
SPAIN
POLAND
BRAZIL
TURKEY
FRANCE
ITALY
U.S.S.R.
PHILIPPINES
HUNGARY
PHYSICS FACULTY
RECENT PH.D.'s
RECENT B.S.'s
0
10
20
30
40
50
60
PERCENT
SOURCE: Physics Today

Some female scientists also see such perceptions of difference as potentially dangerous. “We would all be better off if we could forget about gender altogether,” says Deborah M. Gordon, an animal behaviorist at Stanford University and one of very few women studying social insects. “It is hard for young women starting out to hear that they are different. They should hear that everyone will have to work hard to be a good scientist. They should be thinking about how to do their work as best they can, not how their work is channeled by gender.”

Many scientists believe individual variations are so great that they outweigh those between men and women. “I am an adamant feminist about a lot of these issues, but this one I get sort of riled by,” Williams states. “People approach science differently, so to categorize a woman’s versus a man’s approach would be difficult. It is not that it does not exist, it is just that there are many styles of doing science.”

Yet by questioning the culture of science, many feminist scientists and scholars claim to be broadening this repertoire of styles. In their view, such enrichment will ultimately lead to a more thorough science and a better society. Some such thinkers have suggested that research priorities may shift as a consequence. “A lot of physics has been defense related, and many women left it for that reason,” Didion explains. “At the minimum, the way that science would be communicated would be different, not necessarily the science itself.”

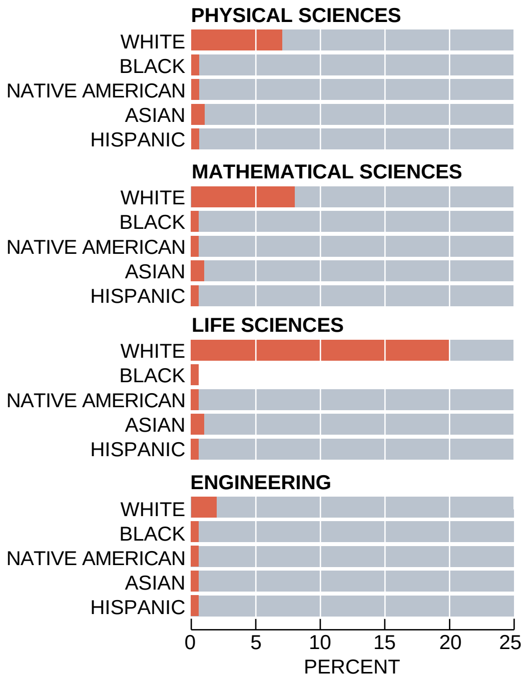
Medicine has already changed in some ways as a result of the growing number of women researchers and practitioners. Women’s health is receiving more funding and attention. Conclusions based on white male patients are no longer being blindly applied to female or minority patients. A recent study in the *New England Journal of Medicine* concluded that female patients who have female doctors were twice as likely to receive Pap smears and mammograms.

“These are not just women’s issues,” Didion declares. The contemporary culture of science “is not only not good for women, it is not good for many men.” In particular, raising a family has been seen as incompatible with a successful scientific career. Women are often perceived to be less committed if they want to have children—although Zuckerman and Cole found that married women and mothers publish as many papers as single and childless women [see “Marriage, Motherhood and Research Performance in Science,” by J. R. Cole and H. Zuckerman; SCIENTIFIC AMERICAN, February 1987].

Nevertheless, without support at the institution where they work or from their spouses, women are more likely to drop out of science to have children. “Early career time is when women raise children, and organizations have to make it doable,” Dresselhaus says. “I know in my own career it was awfully hard in those years. I had four children. I got help from my husband and hired a baby-sitter. But many people made totally unreasonable demands on me—it was almost humanly impossible to do what I was being asked to do.”

Dresselhaus’s success at maintaining her career and family is unusual. “The fact remains that science, like professional life in general, has been organized around the assumption that society need not reproduce itself—or that scientists are not among those involved in reproduction,” Schiebinger notes wryly. The American Chemical Society, for example, found that 37 percent of female chemists older than 50 years had no children; only 9 percent of the men older than 50 were childless.

Employed Female Ph.D. Scientists and Engineers by Racial Group, 1989



SOURCE: National Science Foundation; the statistics are for U.S.

Moreover, 43 percent of women had relocated because of a change in the employment of a spouse. Only 7 percent of the men had relocated. For that reason, many female and male scientists—and people in all fields—support more flexible work time, family leave and child care legislation.

“These professions have evolved around the lives of men who could be professionals around the clock and spend little time on anything else,” Hubbard says. “In science, it is certainly ridiculous to the extent that there is this notion that if you don’t work 24 hours a day, nature is going to run away. It won’t. It will still be there next year, unless we louse it up.”

Gordon of Stanford would like to see universities organized in such a way that they could provide jobs to both partners of a married couple. “They are set up for men whose wives went with them, but the world is not like that anymore,” Gordon observes.

All these considerations, from legislation to more subtle changes in the culture of science, will not ensure that more women or minorities will study or stay in science. Many scientists describe the situation as a catch-22: more women will enter the field only when there are more women in it. And, they say, the only way out of the conundrum is to change society’s attitudes toward women—and men.

Such changes require continued vigilance, especially now that the anticipated job shortage has not materialized. To some, the existing economic malaise portends a loss of ground that women and members of minorities have gained in science. For others, it bodes well. “We are being forced to recognize that people do not have the money to throw at the problems, but that may be a blessing in disguise because people are going to have to do something,” Didion asserts.

As for the old guard, the best thing that could happen is “that they get terribly fond of a granddaughter who is very interested in science,” Vetter says. “It is possible to change people for their granddaughters.”

FURTHER READING

THE POLITICS OF WOMEN’S BIOLOGY. Ruth Hubbard. Rutgers University Press, 1990.

THE OUTER CIRCLE: WOMEN IN THE SCIENTIFIC COMMUNITY. Edited by Harriet Zuckerman, Jonathan R. Cole and John T. Bruer. W. W. Norton, 1991.

BIOLOGY AND FEMINISM: A DYNAMIC INTERACTION. Sue V. Rosser. Twayne Publishers (Macmillan), 1992.

NOBEL PRIZE WOMEN IN SCIENCE: THEIR LIVES, STRUGGLES, AND MOMENTOUS DISCOVERIES. Sharon Bertsch McGrayne. Birch Lane Press, 1993.

Turning Green

Can industrial chemistry trade benzene for sugar?

JOHN STARKEY *Black Star*

JOHN W. FROST of Purdue University holds a flask that contains microbes that can break down a cornstarch-derived sugar solution into industrial chemicals.

Life for the synthetic chemist used to be simple. He (or very occasionally she) simply had to know how to mix inexpensive chemical feedstocks to achieve high yields of one of the thousands of industrial ingredients that make up the accoutrements of modern living, from Tupperware and roller-blade wheels to Valium. Now the industrial chemist's world has been complicated, even darkened, by the fact that many of the ingredients in time-honored recipes are featured on the most-wanted lists of the ecological watchdogs at the Environmental Protection Agency.

For its part, the EPA has done more than just growl. A year ago the agency began a small grants program to encourage the discovery of environmentally kind alternatives for the billions of pounds of toxic substances that are released into the environment every year.

The funding, supplied under the agency's Design for the Environment Program, has spurred research chemists to look at cornstarch, sunlight and other replacements for heavy-metal catalysts, benzene solvents and a host of other bad actors. The program represents something of a departure by the agency from its usual approach. "The EPA's idea of environmental chemistry has traditionally been bulldozers rather than round-bottomed flasks," says Paul T. Anastas, chief of the new chemicals section of the agency's industrial chemistry branch. "But source reduction is the only option that truly prevents pollution."

In late August the EPA put the initial results of the effort on display. It helped to set up the first symposium on alternative chemical synthesis at the biannual meeting of the American Chemical Society. In addition to the work backed by the agency, the conference attendees heard about different approaches to the same types of problems presented by industrial colleagues.

The academic work sponsored by the agency marks a radical shift away from conventional chemical-process technologies. That could be seen in the accomplishments of John W. Frost. Frost is a chemist who believes metabolic products from microbes may one day replace such carcinogenic solvents as benzene. Frost began his talk at the symposium by showing a slide of a midwestern cornfield like the ones near the Purdue University campus in West Lafayette, Ind., where he teaches and does his research. Frost and his co-workers there have proposed using nontoxic glucose, a component of cornstarch and ordinary table sugar, as an alternative to processes that employ petroleum-derived benzene.

To concoct this biocatalyst, the Purdue team has tinkered with the genetic makeup of the ubiquitous intestinal bacterium *Escherichia coli.* Normally, *E. coli* metabolizes most of the glucose available to it in order to provide energy. The Purdue group has genetically engineered the bacterium so that more glucose is diverted toward the other main chemical pathway within the organism that utilizes the sugar. This pathway manufactures amino acids, such as tryptophan, the building blocks for proteins. Before the amino acids can be made, an enzyme breaks down a chemical derived from glucose to produce catechol. Genes for this non-native enzyme were imported into *E. coli* from other microbes. The catechol is a feedstock for vanilla flavoring. It also goes into the making of adipic acid, a key ingredient in nylon manufacturing. Both chemicals would otherwise require cancer-causing benzene as a starting material.

Microbes that chomp on glucose will have trouble making a substance that can match the 65-cents-a-pound price of adipic acid that comes from petrochemicals. That figure, however, does not reflect the externality of environmental impact. "If you take into account the damage to the ozone from making petroleum, the adipic acid would not cost 65 cents," Frost noted at the meeting, which took place in Chicago.

Furthermore, the Purdue process does

not emit ozone-depleting nitrous oxide as a by-product, and the various reactions can be carried out at a frugal 37 degrees Celsius. A biotechnology firm, Genencor International, is considering setting up a pilot plant that will use glucose to make quinic acid, a starting material for the feedstock hydroquinone, in a process that employs a similar biocatalytic reaction. Frost recently formed his own company to provide microorganisms for chemical or grain processors interested in these methods.

Chemists funded by the EPA have turned to plants, as well as bacteria in the intestinal tract, for inspiration. Gary A. Epling, a chemistry professor at the University of Connecticut, has used food dyes and light from a lamp bought on sale at Caldor to create a catalyst that replaces heavy metals in commonplace oxidation reactions. George A. Kraus of Iowa State University has reported how a sunlamp could initiate a chemical re-

More Fun than a Root Canal

At last, biotechnology for the masses. Less noble perhaps than gene therapy for a rare disease, less impressive than a cure for cancer, Creative BioMolecules's latest product, if it works, may nonetheless make many people smile. The Hopkinton, Mass., company is now testing in humans a protein-laced compound that in monkeys seems to rescue teeth doomed to certain root canal.

The death of a tooth is an all-too-familiar experience. When decay or a blow punches a hole through both the hard, mineralized layers of enamel and dentin, the exposed pulp, home to blood vessels and exquisitely sensitive nerves, can become infected. Badly abscessed teeth, still beyond the help of medicine, must be pulled. But when naked pulp is caught before infection has become too deeply entrenched, traditional dentistry can save the tooth—for a price reckoned in dollars and winces. Sucking the pulp out of the root canal and replacing it with gutta-percha or something similar kills the tooth but preserves its function. The American Dental Association estimates that this safe but rather ghastly procedure was performed 13 million times in 1991.

Creative BioMolecules thinks it has a less invasive, less expensive and—mirabile dictu—less painful alternative for about three million of those cases, the 20 to 25 percent with the least infection. The treatment uses a cultured version of the human osteogenic protein OP-1. Research in animals has shown that when applied to broken bones, OP-1 stimulates bone growth. Stryker, Creative BioMolecules's corporate partner in Kalamazoo, Mich., is already testing that application in human clinical trials.

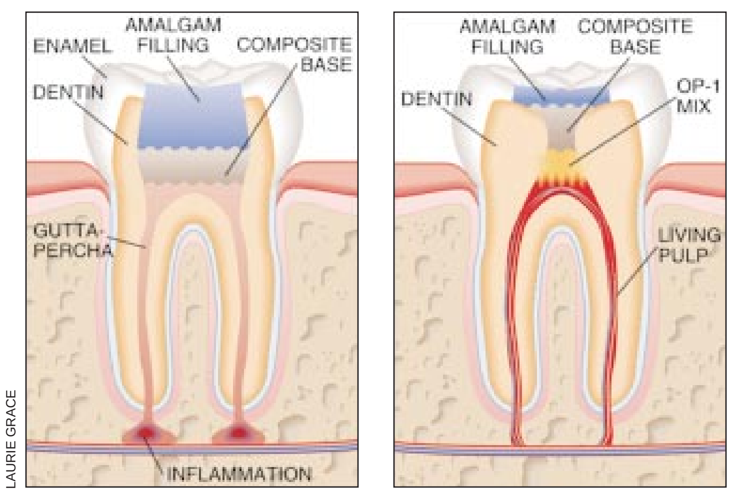
R. Bruce Rutherford, now at the University of Michigan School of Dentistry, wondered whether OP-1 might also repair damaged teeth. Late last year, with scientists from Creative BioMolecules, he tried an experiment. The researchers drilled 30 deep holes into the healthy teeth of four adult macaques, then packed five of the artificial cavities with calcium hydroxide, a common dental filler. Another five they plugged with a collagen mush made from cow bones. The remainder they either left empty or jammed with various amounts of collagen to which a dash of OP-1 had been added. All the cavities were then capped.

After six weeks, the researchers harvested the teeth and found that all those filled with OP-1 had grown new layers of dentin thick enough to protect the pulp from infection. Three of the calcium hydroxide–packed teeth also showed some dentin repair, but it was fragile and invaded the pulp chamber. None of the other controls healed.

"We don't profess to know how this protein works," admits Victor A. Jegede, a vice president of Creative BioMolecules. The tissue that grows to seal the cavity "does not look like normal dentin," Rutherford points out. "It resembles what we call atubular dentin." Whereas normal dentin is permeated by fluid-filled tubules, reparative dentin is more amorphous. That may actually be a good thing. "It might make the pulp less sensitive and less susceptible to decay," Rutherford speculates. He plans to ask the National Institutes of Health to fund further research into how OP-1 works.

Enticed by the potential market for the agent—three million treatments a year at "somewhere under $1,000" apiece, Jegede says—Creative BioMolecules is not waiting for an answer to the "how" question. In July the Food and Drug Administration approved the company's plans to test the compound (which the FDA classifies as a "device") on 40 human patients whose teeth have minimal inflammation and need to be extracted anyway. In each patient a bit of the exposed pulp will be cut away and the cavity filled with either the OP-1/collagen mixture or the collagen alone as a control. The firm hopes to have results by January and—presuming the drug works—to move quickly into a bigger trial with patients who would like to keep their teeth.

Biotech companies have largely avoided dentistry so far. If Creative BioMolecules succeeds in bringing a profitable product to market, it may reverse that trend. "We believe there is significant commercial opportunity in dental therapeutics," says Charles Cohen, the firm's president. Investors appear to agree. They snapped up $16 million worth of stock floated by the company in late August. *—W. Wayt Gibbs*



ROOT CANAL replaces the entire pulp of a tooth with rubber **(left).** ***A less painful alternative may be to remove only infected pulp and cover with OP-1, which regenerates dentin to seal the wound*** **(right).**

action that normally requires toxic reagents. And James Tanko of the Virginia Polytechnic Institute and State University has tested supercritical carbon dioxide (carbon dioxide in a state between a liquid and a gas) as a solvent to replace two environmentally malign carcinogens, carbon tetrachloride and benzene, as well as freon, an ozone-depleting chlorofluorocarbon.

Industrial chemists who have to justify their results to hardened chemical engineers down the hall are less than likely to be enthralled with the department-store paraphernalia of their academic colleagues. The hundreds of millions of dollars spent on researching the best way to make a commodity petrochemical have in many cases been amortized for decades. It will be some time before the industry decides to replace reaction vessels for processing petrochemicals with huge vats of bioreactive bacteria. "When you think of the size of the tanks needed to make 500 million pounds a year of adipic acid, you run into some very practical problems that have to be addressed. I don't know if they can be," says Michael K. Stern, an organic chemist at Monsanto.

Yet mainstream chemical makers have become advocates for a more moderate brand of "green" synthesis—one that takes into account the color of money. Patents on many chemicals expired long ago. Replacing a skull and crossbones on a 50-gallon drum with a green label may bring not only a public relations coup but also new revenues from products replacing tired, generic commodities. Monsanto has invented a process that eliminates phosgene, the redoubtable chemical warfare agent, in the manufacture of isocyanate, an intermediate chemical that goes into urethane foams and resins. A pilot plant at Monsanto headquarters in St. Louis is also testing replacements for chlorine in a widely used process that yields materials for such products as fibers and tires. Benzene is still present in the latter formulation, but the new recipe avoids having to get rid of mountains of chloride salts. "Monsanto could potentially eliminate 50 million to 60 million pounds of salt," Stern remarks.

Going much beyond mere environmental centrism may require a change in thinking that detaches industrial chemistry from the precursors of oil slicks. Synthetic chemists have come to view environmental chemistry as more about cleaning up a waste repository than about designing wholly novel molecules. The EPA and the National Science Foundation, which launched another grants program with the agency earlier this year, want to carve out a new role for chemists that takes into account pollution prevention when designing chemicals. "Chemists have as fundamental an effect on a new chemical product as an architect does on how a building looks," says the EPA's Anastas.

To move in this direction, the EPA and the National Science Foundation want to get them young. Officials from the two agencies are to meet next spring with a number of publishers to try to convince them to emphasize a "benign by design" agenda in new textbooks—a theme that will allow book editors to indulge in the green-marketing fad. Separately, the EPA is also examining expert system software that could flag different chemical routes for arriving at a target molecule.

The agency may get some assistance from industries not wedded to the catalytic cracking tower. Mammoth grain companies in search of fresh markets, such as Archer Daniels Midland (ADM), have already begun large-scale processing of biochemicals. ADM has started to convert cornstarch into the amino acid tryptophan for animal feed—just a chemical step or two away from concocting industrial products such as indigo dye. The reactions needed to make faded blue jeans are not much different from those that Frost coaxes from his microbes. "If you look at the sheer size of some of these grain-processing companies, they can give commodity chemical manufacturers a run for their money," Frost says.

Still, a move to green synthesis may proceed slowly, since it will require volunteers. The EPA only makes recommendations to companies on alternative ways of making chemicals. It does not mandate such changes.

These toxic crusaders must also become fund-raisers. The EPA doled out about $330,000 in grants during the federal fiscal year that ended September 30. The National Science Foundation's Environmentally Benign Chemical Synthesis and Processing Program, whose grant applications the EPA will jointly evaluate, will distribute $2 million or more for the current funding year.

More help may be on the way. Almost $2 billion in funding has been proposed for a multiyear environmental technologies program that the Clinton administration put forward in mid-February as part of its economic-stimulus plan. Congress is considering a spate of bills intended to promote green technologies. The legislative hyperkinesis may be appropriate. Transmuting base metals into Krugerrands may be easier than switching from benzene to cornstarch. —*Gary Stix*

Power Play

A bioentrepreneur does some expert broken-field running

Should entrepreneurial conviction be able to influence how a product is tested and approved? Does deft manipulation of Congress and executive agencies amount to cutting red tape, or is it harnessing the machinery of government to corporate interests? Listen to the tale of MicroGeneSys in Meriden, Conn., and its therapeutic vaccine for the AIDS virus, VaxSyn.

In principle at least, MicroGeneSys's product appears to represent a plausible approach. It is administered to individuals already infected with HIV, the human immunodeficiency virus, which causes AIDS. VaxSyn, its proponents maintain, stimulates the immune system in such a way that the progression of AIDS may be slowed. Several other companies are also developing therapeutic AIDS vaccines. They include Genentech, Immuno AG, Biocine (a collaboration between Chiron and Ciba-Geigy) and Immunization Products. The MicroGeneSys vaccine incorporates the envelope protein of the AIDS virus, known as gp-160.

All the products are currently in the early stages of testing with small numbers of volunteers. But because MicroGeneSys's product has been in development the longest, the company's chairman, Franklin Volvovitz, decided last year to go all out for a full-scale trial to prove that VaxSyn benefits patients. He retained the former senator Russell B. Long of Louisiana to represent MicroGeneSys's interests to the world's most influential deliberative body.

Lo and behold, in October 1992, when Congress passed the 1993 Defense Appropriation Act, it provided $20 million to pay for "a large-scale Phase III clinical investigation of the gp-160 vaccine." AIDS researchers at the National Institutes of Health and regulators at the Food and Drug Administration were incensed that one product was being pushed ahead of the crowd. Although VaxSyn appears to be safe and to have some effects on the immune system, the importance of those effects is hotly disputed. "We have no idea what MicroGeneSys's data mean," declares Steven Schnittman, a researcher at the NIH. It did not help that the U.S. Army was investigating its own chief AIDS researcher, Robert R. Redfield, for scientific misconduct, after he was accused of exaggerating the effects of VaxSyn in a public presentation. (Redfield was cleared of the charge.)

The heat coming from the kitchen began to reach the upper floor. Bernadine P. Healy, then director of the NIH, and FDA commissioner David A. Kessler questioned the plan for the one-drug study. Healy convened a special blue-ribbon committee to come up with a plan for a comparison trial of several of the available products. Patients in a specific stage of HIV infection—about 12,000 in all—would be assigned to receive one of three therapeutic vaccines or a placebo. Defense Secretary Les Aspin agreed to transfer the $20-million appropriation for the trial to the Department of Health and Human Services, the NIH's parent agency.

Testing a number of different products to see which might work has a certain commonsense appeal. But Volvovitz says because such a test spreads resources more thinly, it "would significantly reduce your chances of proving that one of them works." Volvovitz, who accuses the NIH of being biased against therapeutic vaccines, also contends that the participation of other companies threatened to slow the trial.

Volvovitz's next move effectively torpedoed the NIH plan. Although companies normally supply products they have developed free for trials organized by the government—and although the other potential participants agreed to do so—Volvovitz wrote the health department that he would have to charge cost price for VaxSyn: some $10 million. The department, not wishing to set the precedent, declined to go forward, and the $20 million earmarked for the test stayed with the army.

The military, however, found the funds insufficient to conduct a multivaccine study, even had the law allowed it to. According to Captain Bill Buckner, an army spokesperson, the $20 million is sufficient to test only one product—VaxSyn. The trial will study about 6,000 volunteers infected with HIV. The army—like the Department of Health and Human Services—needs the manufacturer to donate the vaccine. And for the one-vaccine trial, MicroGeneSys has been able to find money to supply the product. Wyeth-Ayerst Laboratories, a commercial partner in the development of VaxSyn, is reimbursing MicroGeneSys for supplies used in the first year of the trial.

Some observers have sympathy with the plight of a small company that is effectively obliged to supply large amounts of product without charge. Manufacturers have to assume full liability should anything go wrong, they point out. "The government should at least provide sovereign immunity to developers of vaccines," says Robert W. Rubin, vice provost and deputy dean for research at the University of Miami School of Medicine, who has run several clinical studies. Volvovitz cites previous instances where the NIH has purchased vaccines or products for use in clinical trials. But Rubin insists MicroGeneSys's willingness to provide its product for the army's single-vaccine trial makes its previous plea of penury "harder to defend ethically."

"It is rather disconcerting," comments Victor Zonona, a spokesperson for the health department. "We would still like to go forward with a multicandidate trial, but that has been scuttled by one company." According to Zonona, cabinet-level discussions are under way to ascertain whether a multivaccine trial can be rescued.

Other manufacturers are also dismayed. "We still stand behind the recommendations of the NIH that there should be a comparative trial of several vaccines," asserts Geoffrey Teeter of Genentech. But Larry Kurtz of Chiron believes not one of the AIDS vaccines now in development has enough supportive evidence to justify a full-scale efficacy trial. "It is just a shame with AIDS that this kind of money should be devoted to one product on the basis of a good lobbyist," he says. —*Tim Beardsley*

Software Skipper

A convoy steams through the icy waters of the North Atlantic, much the way convoys did 50 years ago carrying needed goods to overseas friends and allies. But this convoy, except for the flagship, is totally unmanned. All navigation functions have been automated. Inputs from radar, Doppler sonar and satellite navigation systems are analyzed by artificial intelligence, compared with digitized chart and course data, and any necessary course corrections are made by the ship. If other vessels are met on the high seas or in restricted channels, evasive maneuvers are carried out automatically, and when the danger is past, the convoy is placed back on its optimized course.

The skeleton crew on the flagship has little to do. Its presence is mostly a concession to out-of-date maritime regulations that require human control of critical functions. No such convoys ply the seas today, but according to Mitsubishi Heavy Industries (MHI) in Tokyo, much of the technology that could make such a scenario possible is already in place.

MHI, one of the largest shipbuilders in the world, has already launched one ship, a 258,000-ton very large crude carrier supertanker, with much of the onboard technology that an unmanned ship would require. Built at MHI's Nagasaki Shipyard and Engine Works for Shinwa Kaiun Kaisha Ltd., the first of this new generation of computer-managed and computer-guided ships has already successfully completed a voyage to the oil fields of the Middle East with fewer than two dozen hands and officers on board—far less than a ship of this size normally requires. Computer systems assist with virtually every aspect of the ship's operation, from power plant maintenance and cargo handling to navigation and collision avoidance.

The nerve center of the ship, called Super Bridge by MHI, has the edge on any human helmsman because it never sleeps, and it monitors inputs from all instruments and navigation systems simultaneously. Without skipping a beat, the system also surveys its human counterpart on the bridge by periodically flashing a signal on its CRT touchscreen that calls for a response. If the watchkeeper is napping, is otherwise incapacitated or fails to respond in the requisite time for any reason, the system will alert the ship's officers in their quarters or anywhere else they may be. This feature alone might have averted an *Exxon Valdez* disaster, but Super Bridge does much more.

Based on digitized chart data for the chosen course, the system displays the planned route on a monitor that can also report on the movements of up to 40 other nearby vessels. If any one of those other ships is found to be on a collision course, the systems sound a warning, calculate an optimized evasive course and present it on the monitor. The suggested evasive maneuver takes digitized chart data into consideration, avoiding reefs and shallows to prevent groundings and strandings.

In the absence of imminent danger, the system collects and monitors ship data, plans optimum routes that take into consideration present and predicted weather conditions and acts as an autopilot. Of course, ships have had autopi-

ECM for 747s

Should commercial airliners carry high-tech protection?

For the white-knuckle flier—or even the passenger who sleeps through takeoff and landing—the case for protecting an airliner from Stinger missile attacks and other paramilitary assault would seem open and shut. After all, the dispersion of advanced weaponry from the arsenals of democracy, not to mention the bargain basement of the former Soviet Union, has put formidable antiaircraft power in the hands of terrorists.

The U.S. government and the aerospace industry would appear to agree. The American Defense Preparedness Association (ADPA), a de facto brain trust of the military-industrial complex, along with the Federal Aviation Administration (FAA) and the Department of Defense, has sponsored a three-day conference in mid-October. It will touch on how to protect both commercial and military aircraft against the predations of heat-seeking missiles by employing electronic countermeasures (ECM). The conference on aircraft survivability will also report on structural hardening and other continuing work toward bombproofing of passenger and cargo aircraft.

Such considerations are not hypothetical. Since 1970 a surface-to-air missile attack on a civilian aircraft has become an almost once- or twice-yearly occurrence, mostly in world trouble spots, according to Marvin B. Schaffer, a senior physical scientist with RAND, the Santa Monica, Calif., research group. One might even wonder why the carnage has not been greater. For the price of a top-of-the-line Mercedes—$50,000 to $100,000—the enterprising paramilitarist can buy an SA-14, the Russian equivalent of a Stinger missile launcher. And the real thing is out there—courtesy of U.S. efforts to support its allies and surrogates around the world. The Clinton administration has reportedly proposed allocating $55 million to buy back Stingers from the Afghan freedom fighters, the mujahedeen.

What to do while such plans are developed and implemented? One approach might be to adapt the protective hardware used on board military transports and special aircraft such as *Air Force One*. They are equipped to send out flares or modulated infrared pulses. Heat signals from an infrared jammer, for example, may confuse the missile's heat-seeking detectors as to the actual location of the targeted aircraft. Attendees at the conference will now consider the notion of equipping even commercial aircraft with this technology.

So, the average passenger may think, what are we waiting for? In fact, an FAA task force was set up last year to study the threat from shoulder-fired missiles. But the carriers themselves may regard such threats as far less real than their immediate economic challenges. American Airlines and United Airlines, it would seem, might view the fares of budget carrier Southwest Airlines as a more imminent menace than Stingers in the airspace above U.S. airports.

No major air carrier had accepted the ADPA's invitation to attend the meeting. Joseph J. Shallcross, executive vice president for operations at World Airways, a Virginia-based international cargo and charter carrier, is uncertain whether he will speak, despite a place on the agenda. "This industry is in a deep and serious depression," Shallcross says. "We're losing jobs on a daily basis. We have no spare cash for anything, especially for hardening the top of an aircraft in case a meteor hits it."

Shallcross and other industry executives may well have a point. Furthermore, attempts to turn civilian aircraft into missile-impervious fortresses may produce their own set of technical pitfalls. "As bad as it would be for an airliner to get hit by a missile while wait-

lots for a long time, but the system that seamen dubbed "Iron Mike" in the past did nothing but maintain a compass heading. Super Bridge, thanks to the Global Positioning System, can record actual position and maintain the vessel on a true course without errors because of winds or currents.

At present, the digitized chart data that Super Bridge requires must be produced chart by chart for each ship and route. There is no source of digitized charts for the world's oceans that is approved by all maritime nations or even by one nation. That should change soon. The International Maritime Organization is now working on a standardized Electronic Chart Display Information System. But for now, MHI will digitize charts for shipowners who, under Japanese maritime law, are not allowed to do so on their own. MHI itself must get permission from the Japanese coast guard.

Whether the prospect of unmanned supertankers plying the trade routes of the world's oceans will ever be acceptable from a geopolitical point of view is an open question. But the concept of backing up human fallibility with a fully computer-operated ship is certainly a welcome development at a time when fragile ecosystems are increasingly threatened by marine disasters. The *Exxon Valdez* is arguably more famous than Captain Bligh's *Bounty*. Such events, absent an improvement in the control of ocean traffic, are likely to become even more common. One estimate has it that within 10 years, the Bosporus Strait could see 1,200 supertanker passages every year—each bearing more than 100,000 tons of crude from cash-starved former Soviet republics to energy-hungry industrial economies.

—Robert Patton, Tokyo

MITSUBISHI

AUTOMATED CAPTAINS are presaged by this 258,000-ton oil tanker, built by Mitsubishi Heavy Industries, which contains a computerized guidance system.

ing for takeoff at Washington National, it would be almost as bad for the airliner to all of a sudden start spitting flares out its tail," remarks Russ Stanton, the army's assistant program manager for infrared countermeasures.

Infrared countermeasures may make army toilet seats seem inexpensive. Army helicopters currently fly with bargain-basement infrared jammers that cost $20,000 to $30,000, Stanton says. Prices could escalate dramatically for a wide-body aircraft.

Ultimately, some experts think, electronic countermeasures may preserve more industry jobs than aircraft. "This is an attempt to hype the problem to establish a market where none exists," comments Vincent M. Cannistraro, a former chief of counterterrorism operations for the Central Intelligence Agency. Despite the availability of these weapons for 20 years from international arms dealers, most incidents have involved civilian aircraft entering war zones. For the anxious, he says, the tools of terror lurk everywhere: radioactive isotopes or biological agents could have made the World Trade Center uninhabitable for more than just a few weeks. "Does that mean we ought to inoculate everybody against anthrax? I don't think so."

Bombs that get past preflight security are also getting attention, although the number of bombings has diminished, perhaps because of better security after the downing of Pan Am Flight 103 over Lockerbie, Scotland, in 1988. The FAA's research and development staff is contemplating how to fortify the commercial airliners of the future with hard bodies. The agency sponsors a more than $4-million-a-year program to do computer modeling of blast effects and live bomb testing on scrapped airframes. It has been working with the Department of Defense and Boeing, McDonnell Douglas and Lockheed to determine the dynamics of a bomb blast in a civilian liner. These efforts may eventually yield a bomb-resistant design that might come into service after the year 2000.

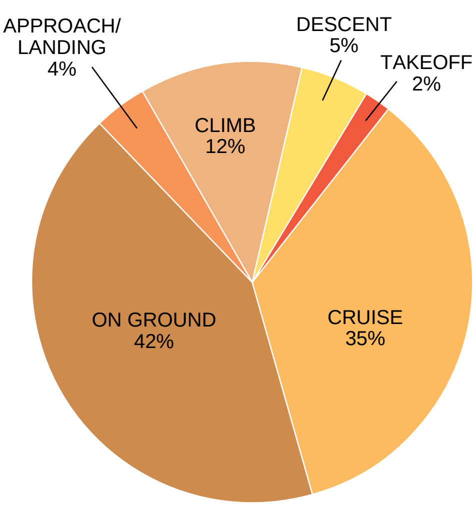
In both Europe and the U.S., specialists in structural design have built prototype baggage containers that will absorb or attenuate the impact of the blast of a bomb weighing up to three pounds (the plastic explosive that destroyed Pan Am 103 may have weighed less than a pound). Baggage containers, designed with finite-element analyses to calculate structural loading from a blast, have been built from polymer composites bolstered with reinforcing fibers. The boxlike structures would replace the removable compartments now in wide-body aircraft (narrow-body airplanes do not carry them) toward the end of their three- to four-year lifetimes. "They would be a lightweight modular structure that might resemble a Fisher-Price toy," says Kenneth J. Hacker, the chief of the FAA's aircraft-hardening program. "They might cost more, but they would last four to five times longer than a container made of aluminum."

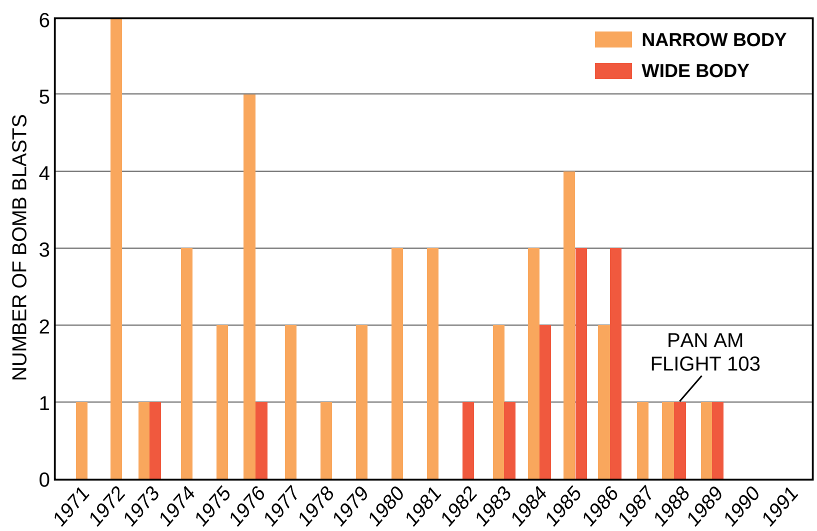
One container, designed by Jaycor, a San Diego company, has been built to absorb the entire force of the blast. Another prototype, from Royal Ordnance, a division of British Aerospace, is lighter in weight than is the Jaycor baggage hold. It is intended to lessen the impact, although it may not fully contain the blast effects. If the receptacle does rupture, it does so slowly, which diminishes blast damage to the fuselage.

The metal of the fuselage in future aircraft might be fitted so that if the container could not hold the forces generated by the bomb explosion, the energy would be discharged through the release of a side panel in the fuselage. This hole, adjacent to surrounding reinforced areas, would impede the formation of loose metal flaps that may have contributed to the Pan Am disaster by causing long, seamlike rips in the aircraft body, which could lead to the loss of aerodynamic stability.

Yet even such relatively inexpensive replacements for today's baggage containers do not thrill financially strapped airlines. It costs American Airlines more than $5 million a year to repair and maintain existing baggage compartments, an expenditure that might increase for a more elaborately constructed receptacle.

Homer Boynton, who recently retired as managing director for security at American Airlines, opposes forcing carriers to absorb security costs. Outside the U.S., he notes, other governments often pick up the check for such protective measures. "This is another example where our government has come up with ways to spend the airlines' money," Boynton remarks. Cries of poverty may yet grow louder. In coming years the Clinton administration may find what keeps the airplanes on the ground are airline creditors, not Stinger missiles. *—Gary Stix*





LAURIE GRACE

***BOMBINGS** on commercial aircraft are diminishing* (right); *many incidents over the 20-year period occurred while a plane was on the ground* (left). *The data, compiled by McDonnell Douglas, exclude incidents in which a bomb did not explode.*

Ivy League Bonus Babies

Should elite universities such as Harvard and Cornell be allowed to consult with one another (some may say collude) on their offers of financial aid in order to avoid competing for the most desirable applicants? Can this "price fixing" be justified if it stretches aid budgets further? Does it mean that more minority and low-income kids can attend? Do we all benefit from having a diverse mix of students at the most selective private universities? Is such an arrangement fair if it reduces the size of the scholarships going to individual students at the top of the heap—the ones for which the schools might otherwise have a bidding war?

In 1991 eight Ivy League colleges signed a consent decree ending their practice of sharing information on financial aid, faculty salaries and tuition, after the U.S. Department of Justice brought a civil antitrust suit against them. The Massachusetts Institute of Technology alone refused to sign and was tried and found guilty of price fixing. In September of 1993, M.I.T. won an appeal of this ruling. The Department of Justice has not decided whether to pursue the matter further.

When sellers compete with one another, according to economic theory, consumer welfare is maximized. This proposition forms the basis of our antitrust laws. They explicitly forbid "joint pricing" or "price fixing." Economists know that in any market where the number of sellers is small enough to organize and monitor and it is difficult for new sellers to enter and compete, there is incentive to form a cartel.

Such economists as Lawrence J. White, professor at the New York University Stern School of Business, believe this theory applies to colleges as much as it does to profit-maximizing firms: "The nonprofit, cooperative status of colleges does not make them unique. Even nonprofits are concerned with revenue needs and competition, so they have an economic incentive to try to fix prices," he says. If collusion results in a greater diversity of students and that improves the overall quality of education at the school for everyone, White still sees no justification for exemption from the law. "The college should be able internalize these returns," he asserts. "If they really go after the most attractive guy or girl [in a bidding war], it must be that he or she contributes more to the rest of the university, and if this is the case, then the college should be able to charge the other students more because it is a more worthwhile educational experience for them. If this is not the case, then why should Harvard get into a bidding war? Why doesn't it say: let Princeton get the top guy or girl, we'll go for the next worthy student."

"I think it's a phony argument," declares Daniel Steiner, former vice president and general counsel at Harvard, now at the John F. Kennedy School of Government there. "Colleges are not like dealers selling automobiles. Who else goes out and seeks people who cannot pay and tries to attract them as the colleges do?" he asks.

Richard R. Spies, chief financial officer of Princeton University, also disagrees with the antitrust position. "I find the argument against the colleges unpersuasive," Spies says. "There is an institutional self-interest in having a diversity of students, but our actions are also driven by a larger social objective. There is a return to the country as a whole."

If having the Ivy League educate more poor and minority kids has a payoff for the country as a whole, the colleges are producing what economists call "externalities." These are incidental by-products of production or competition that benefit (or hurt) others. Acid rain is a negative externality.

No one doubts the existence of positive externalities from education, but as Ronald G. Ehrenberg, professor of economics at Cornell, points out, there is not much hard evidence to prove that coordinated aid actually produces external benefits. He would rather see government policies increase attendance of needy students at selective public institutions. "It does not satisfy broad national goals to facilitate access of low-income and minority students to elite private colleges," Ehrenberg says.

On the other side, Spies makes a compelling point when he offers an argument much like the theory recently advanced by the former president of Princeton, William G. Bowen, now president of the Andrew W. Mellon Foundation, and David W. Breneman, the former president of Kalamazoo College, now at Harvard. Spies observes that the rationale for the laws against joint pricing should not apply to these particular, highly selective colleges since their joint action was not designed to maximize their profits or even to increase their revenues. "This is not price discrimination," contends Spies, referring to a practice in which sellers charge high prices to some buyers and lower the price to those who cannot or will not pay more. "The colleges that were the target of the antitrust action could easily sell all the places in their incoming freshman classes at full tuition. They don't offer financial aid to increase revenue; in fact, they lose money by doing so. They are performing a vital public function," he argues.

In July of 1992, right after the M.I.T. loss in Federal District Court, Congress passed an amendment to the Higher Education Act of 1965 giving colleges limited permission to cooperate in setting standards for financial aid. William R. Cotter, an attorney and the president of Colby College, sees this action as a vindication of the colleges' position: "Congress, too, has recognized that need-based financial aid is really the most sensible way to allocate scarce scholarship resources," Cotter says.

Even so, as Spies emphasizes, "there is still a lot of room for a bidding war to develop, and there is a general climate of inhibition. The colleges feel you have to have three lawyers present defending every conversation. They have spent a fortune in legal fees defending the antitrust charges—and to what social purpose, I just don't know."

Whatever the eventual result of the legal process, financial constraints alone may force the final outcome. According to Ehrenberg, "We have likely reached the point when need-blind admissions policies and need-based financial aid have become too expensive for the colleges to maintain." —*Judith Fields*

JUDITH FIELDS teaches economics at Herbert H. Lehman College, City University of New York, in the Bronx.

Fermat's Last Time-Trip

On a warm evening late in June, I settled into a garden recliner to read the newspaper. A front-page story reported that Andrew J. Wiles of Princeton University had solved one of the great problems in mathematics—Fermat's Last Theorem. Wiles announced the result at the end of three lectures, which were innocently entitled "Modular Forms, Elliptic Curves and Galois Representations." The packed audience strongly suspected that Wiles had Fermat up his sleeve, and they were right. I wondered what the French mathematician Pierre de Fermat would make of all this were he to be transported suddenly 350 years into the future. The night air was still quite warm, and I decided to shut my eyes for a bit. I was not asleep very long when I was disturbed by a curious high-pitched whine. I looked up, and there on the lawn was a strange contraption made of shiny metal and glass. A man was sitting inside the machine. He wore a black cloak, and his long hair was tied up with a ribbon. He climbed out and introduced himself as Pierre de Fermat. "Excuse me," I stammered, "but I thought you died three centuries ago."

"Oh, heavens no. I just met this interesting gentleman, H. G. Wells—perhaps you know him."

"I know of him," I said in a strangled voice.

"He claimed to be from the future. I did not believe this was possible, so he loaned me his machine. I now see that he was telling the truth."

"You've arrived at a propitious moment," I told him. "Your last theorem has just been proved."

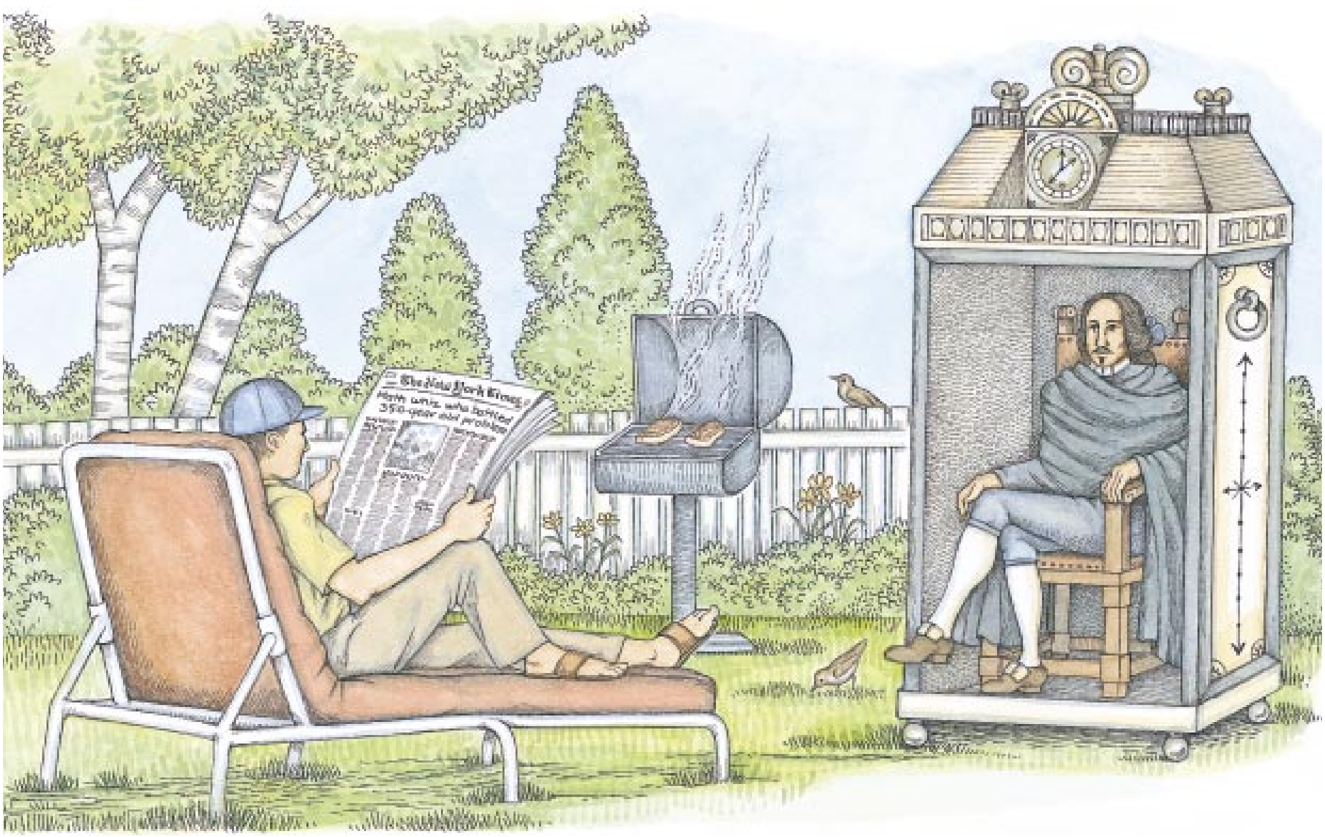
"My what?"

"Do you remember writing a note in the margin of your copy of the *Arithmetica?* You wrote: 'To resolve a cube into the sum of two cubes, a fourth power into two fourth powers, or in general any power higher than the second into two of the same kind, is impossible; of which fact I have found a remarkable proof. The margin is too small to contain it.' "

"Ah, yes."

"In algebraic symbolism, you said that if n is a whole number greater than two and if x, y and z are nonzero whole numbers, then $x^n + y^n = z^n$ has no solutions. This conjecture has come to be known as your last theorem, because it was for many years the sole remaining assertion of yours that had neither been proved nor disproved by your successors. Nobody could reconstruct your alleged 'remarkable proof,' and most people doubted that you ever possessed one. For more than three centuries, the best mathematical minds in the world grappled with your last theorem—occasionally making progress but never quite polishing it off—until its dramatic solution by Andrew Wiles."

I showed Fermat the paper and wait-



"MATH WHIZ Who Battled 350-Year-Old Problem" runs the headline of a story about the recent proof of Fermat's Last Theorem. The great French mathematician himself arrives in a Time Machine to visit Ian Stewart as he ponders the news.

ed while he digested the article. "They are calling it the 'Holy Grail' of mathematics? 'Theorem of the century'? Good gracious!"

"Did you have a proof?" I asked.

He smiled. "Not so fast. Before I reveal my secret, I must ask you to tell me of the remarkable mathematical developments that have culminated in the proof of my modest conjecture."

"Fine," I said, "but first you must at least tell me how you came up with your last theorem."

"Certainly. My first love has always been number theory: the study of ordinary whole numbers. As you may know, a Greek named Diophantus had set the subject in motion by inventing the idea of a 'Diophantine equation'—an equation that must be solved in whole numbers—and writing a book about them, the *Arithmetica*. An important example is the Pythagorean equation $x^2 + y^2 = z^2$: two squares whose sum is a square. This equation has whole number solutions, such as $3^2 + 4^2 = 5^2$ and $5^2 + 12^2 = 13^2$. In fact, there are infinitely many whole number solutions.

"Some time around 1637, I was reading Diophantus and thinking about the Pythagorean equation. It occurred to me to ask what happens if instead of squares you try cubes. At first I experimented with some simple cases: is $1^3 + 2^3$ a cube? No, it equals nine. Is $2^3 + 3^3$ a cube? No. I came close: for instance, $9^3 + 10^3 = 1{,}729$ and $12^3 = 1{,}728$, but I could find no solutions at all, except for uninteresting ones, such as $0^3 + 1^3 = 1^3$, in which one number is zero. My repeated failures to find any solutions led me to my modest conjecture."

"Okay, let me begin to explain how your conjecture was proved," I said. "As you know, the only cases that need be treated are when $n = 4$ or n is prime."

"Yes. Because, for example, every perfect 15th power is also a perfect cube, so any solution for the 15th power is automatically a solution for cubes."

"Right. In symbols, $x^{15} = (x^5)^3$, so if $x^{15} + y^{15} = z^{15}$ then $(x^5)^3 + (y^5)^3 = (z^5)^3$. Because any integer greater than two is divisible either by four or by an odd prime, the same argument shows that four and odd primes are the only powers that need be investigated. You yourself made the first inroad, with a proof for fourth powers, $n = 4$."

"Yes, I am proud of that—my method of 'infinite descent.' I assumed for tactical reasons that there was a solution to a slightly more general equation, $x^4 + y^4 = z^2$. It is more general because any fourth power is also a square. I noticed that such a solution would give a Pythagorean triangle with two sides that were themselves squares [*see illustration on next page*]. I applied the standard formula for Pythagorean triangles, made a few simple deductions and found I could construct another solution of the equation $x^4 + y^4 = z^2$ with smaller nonzero values of x, y and z. Continuing in this manner, I was led to the inevitable conclusion that the existence of any solution at all would force the smallest possible nonzero numbers—$x = 1$, $y = 1$—to be a solution, but it does not. Therefore, no solution to $x^4 + y^4 = z^2$ can exist, and in particular no solution to $x^4 + y^4 = z^4$ can."

"Cunning," I said. "So then the only cases left were when n is an odd prime. You also proved the case $n = 3$, sums of two cubes. Independently, the Swiss mathematician Leonhard Euler proved the same two cases $n = 3$, 4. Peter G. Lejeune Dirichlet proved $n = 5$ in 1828, and so did Adrien-Marie Legendre in 1830. In 1839 Gabriel Lamé tried to prove $n = 7$; he made some mistakes, but these were corrected by Henri Lebesgue in 1840."

"So after two centuries, only the special cases $n = 3$, 4, 5, 7 were proved? Did nobody have a general idea?"

"Lamé, in 1847. He claimed to have a proof for all powers n. But Ernst E. Kummer discovered a mistake—a very interesting mistake that pointed the way for future progress. Lamé's basic strategy turned out to be very fruitful, but his tactics were bad."

"And what was that strategy?"

"To introduce new kinds of numbers called algebraic numbers—more general numbers than just whole numbers. They are referred to as algebraic because they are solutions of algebraic equations, but the details aren't important. The expression $x^n + y^n$ can be written as the product of two other expressions. For instance, when $n = 5$ we have

$$x^5 + y^5 = (x + y)(x^4 + x^3 y + x^2 y^2 + xy^3 + y^4).$$

The $x + y$ factor is nice and simple, but the other one is a mess. Lamé noticed that by using his algebraic numbers he could express the complicated factor as a product of four much nicer factors. More generally, he could write $x^n + y^n$ as a product of n simple terms. Furthermore, the product of those terms was a perfect nth power, because $x^n + y^n = z^n$. He also noticed that no two of those terms possessed a common divisor. Now, with ordinary whole numbers, if a product of terms without any common divisors is an nth power, then each term individually is an nth power. That statement depends only on the fact that every number can be expressed as a product of primes in a *unique* way.

"Lamé assumed that the same property held for algebraic numbers. Then, instead of just having your single equation, he had n different equations, each saying that one of the n terms is an nth power. And all of them had to be true simultaneously. This is asking an awful lot, and not surprisingly he was able to deduce that no solution existed."

"Hmm," Fermat said. "I had some similar ideas myself. But—"

"But it's not that simple, is it?" I interjected. "Kummer and others pointed out that for $n = 23$, Lamé's algebraic numbers can be written as a product of primes in more than one way. It's very curious, and the calculations are rather hairy—I could show you if you want."

He waved me on. "No, I will think about that for myself. Fascinating."

"Kummer asked why algebraic numbers could have more than one prime factorization, and eventually he discovered that he could sort the whole thing out by introducing a new kind of gadget altogether, which he called ideal numbers. They provided some 'extra' prime factors to make everything work out right. You can see it's getting rather complicated and more and more abstract now, but that's the way the mathematics inevitably led.

"By 1847 Kummer had used his theory of ideals to dispose of your conjecture for all n up to 100, except $n = 37$, 59 and 67. By adding extra bells and whistles to the mathematical machinery, Kummer and Dimitri Mirimanoff disposed of those cases, too, in 1857. By 1992 similar methods had proved all cases n less than or equal to a million."

"But almost all numbers are larger than a million," Fermat observed. "Such a case-by-case approach could never succeed in solving the full problem."

"No. Of course, at any point somebody might have noticed something that opened up the entire game. But it didn't go that way. Just more and more intricate special cases. A new idea was needed. And that came by a rather different route. People started asking how many solutions a Diophantine equation might have. Some Diophantine equations have infinitely many solutions—such as the Pythagorean equation. Some have none—your equation for $3 \le n \le$ one million, if we ignore trivial solutions. Some have finitely many—like $y^2 + 2 = x^3$, whose only solutions in positive integers are $x = 3$, $y = 5$.

"In 1922 the English mathematician Louis J. Mordell was trying to work out what distinguished these possibilities, and he started to see a likely pattern. He noticed that if you look at all solutions of such an equation in complex numbers—getting as general as you can, no assumptions about whole numbers at all—then those solutions form a topological surface. The surface has a finite number of 'holes,' like a doughnut or a pretzel. What struck him as remarkable was that equations with infinitely many whole number solutions always had no holes, or just one, when solved in complex numbers. There seemed to be a connection between the topology and the arithmetic.

"This was wild stuff—nobody could see any way to get a solid connection between two such disparate branches of mathematics. But Mordell was sufficiently convinced that he published what is now called the Mordell conjecture, which states that equations that give rise to surfaces with two or more holes have only finitely many integer solutions."

Fermat looked puzzled. "How does this link to my conjecture?"

"The number of holes in the surface corresponding to your equation $x^n + y^n = z^n$ is $(n-1)(n-2)/2$, and for $n \leq 3$ this is at least two. So the Mordell conjecture implies that if your equation has any integer solutions at all, then it must have only finitely many."

Fermat looked even more perplexed. "But if x, y and z form a solution, then so do $2x$, $2y$ and $2z$ or $3x$, $3y$ and $3z$, and so on. Infinitely many."

"Oops. I should have made it clear that Mordell was talking about solutions without any common factor."

"I see."

"Right. Now comes the first big breakthrough. In 1962 Igor R. Shafarevich came up with a new, rather technical conjecture about Diophantine equations. In 1968 A. N. Parshin proved that the Shafarevich conjecture implies the Mordell conjecture. Finally, in 1983, the young German mathematician Gerd Faltings proved Parshin's conjecture and therefore also Mordell's. Which means that your conjecture is nearly true: if for any n there are exceptions, there can be only finitely many of them. You'd like his proof: it uses a version of your method of infinite descent."

"Ah."

"But applied to very abstract things called abelian varieties."

"Oh. Still, it is gratifying to see that my simple guess has given rise to so many deep and powerful new mathematical concepts."

"Indeed. Now, you may object that finitely many solutions is not the same as none, and that's quite right. But you will appreciate that it's a big step to get a potentially infinite number of solutions down to a finite number. Soon afterward D. R. Heath-Brown modified Faltings's approach to prove that the proportion of integers n for which your conjecture is true approaches 100 percent as n becomes very large. Your last theorem is 'almost always true.'"

Fermat seemed pleased. "That I would consider a general result, though perhaps a rather unspecific one, in that it does not state which values of n are exceptional."

"Correct. A more specific idea was still lacking. That came from a very beautiful theory that lies at the heart of the modern approach to Diophantine equations. It is the theory of elliptic curves."

"What are those?"

"Equations of the form $y^2 = ax^3 + bx^2 + cx + d$—a perfect square equal to a cubic polynomial. They are called 'elliptic' because of a rather vague connection with the problem of finding a formula for the perimeter of an ellipse and 'curves' because every equation defines a geometric curve by way of coordinate geometry. One of the striking properties of elliptic curves is that given a few integer solutions of the equation, you can combine them to get other solutions. There is a geometric construction to build new solutions out of old ones [*see illustration on opposite page*].

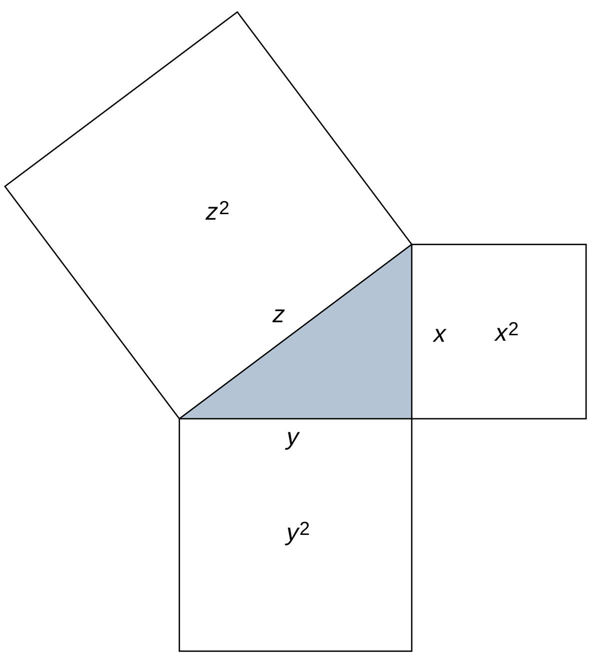
"Elliptic curves were one of the things that stimulated Mordell to his conjecture, because the surfaces associated with them have only one hole or none. Over the years a very strong theory of elliptic curves has been developed. You could say they are the one area of Diophantine equations that people really understand pretty well. But the area has its own big unsolved problems, and the biggest of all is the Taniyama-Weil conjecture. This theorem says that every elliptic curve can be represented in terms of modular functions—which are a kind of broad generalization of the usual trigonometric functions, sines and cosines. It means that every elliptic curve has a kind of nice coordinate system.

"Now we're into the home stretch. Early in the 1980s Gerhard Frey of the University of the Saarlands made a crucial connection between your last theorem and elliptic curves. He related the least understood area in number theory to the best understood area. Frey's line of thought was this: suppose there is a solution $X^n + Y^n = Z^n$ of your equation. I've used capital letters to show we're thinking of some specific solution. You want to show that no such solution exists, so you just have to assume one does and then fight your way to some contradictory conclusion. It doesn't matter what the contradiction is."

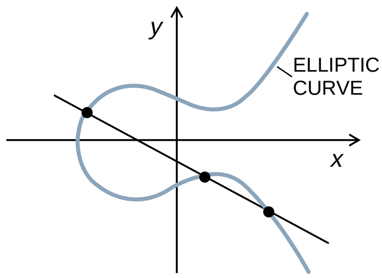
"Reductio ad absurdum," Fermat declared.

Pythagorean Triangle

To solve the Pythagorean equation $x^2 + y^2 = z^2$, pick any whole numbers k, u, v. Let $x = k(u^2 - v^2)$, $y = 2kuv$ and $z = k(u^2 + v^2)$. Then you have a solution. For example, let $k = 1$, $u = 2$ and $v = 1$. Then $x = 3$, $y = 4$ and $z = 5$. Or let $k = 1$, $u = 3$ and $v = 2$, so that $x = 5$, $y = 12$ and $z = 13$. This method generates all solutions.



Points on Elliptic Curve



A straight line intersects a typical elliptic curve at three points. If the coordinates of two of those points correspond to whole number solutions of the associated Diophantine equation, then so do the coordinates of the third point. To construct new solutions from old ones, you take two solutions, draw the line through the corresponding points and calculate the coordinates of the third point at which this line hits this curve.

"We call it 'proof by contradiction' nowadays, but that's it. Frey followed a suggestion of Jean-Pierre Serre of the College of France in Paris and looked at the elliptic curve $y^2 = x(x - X^n)(x - Y^n)$. By hammering it with the general theory of elliptic curves, he discovered that what is now known as Frey's elliptic curve is a very curious beast. It has a strange combination of properties—so strange that it looks highly improbable that such a beast could exist at all, which of course is exactly what you want to prove. In 1986 Kenneth A. Ribet of the University of California at Berkeley made Frey's idea precise. He proved that if the Taniyama-Weil conjecture is true, then Frey's elliptic curve cannot exist. That would prove your conjecture completely, by contradiction.

"This is a major, major discovery. It says your conjecture is not just an isolated curiosity. Instead it lies at the heart of modern number theory—the Taniyama-Weil conjecture. As a child, Andrew Wiles had wanted to prove Fermat's Last Theorem. When he became a professional mathematician, he decided that it was just an isolated, difficult problem—nice to prove but not really important beyond its notoriety. But when he learned of Ribet's work, he decided to devote all of his research effort to a proof. He realized that you don't need the full force of the Taniyama-Weil conjecture to make the approach work; you just need one particular special case of it, valid for so-called semistable elliptic curves. He broke the problem down into six pieces, and one by one he solved them until only one held out. Then a lecture by Barry C. Mazur of Harvard University on something totally different sparked an idea that gave him the final clue. In a 200-page paper he marshaled enough machinery to prove this special case of the Taniyama-Weil conjecture. So Wiles proved the semistable case of Taniyama-Weil, which implies that Ribet's argument proves that Frey's elliptic curve doesn't exist—and you were right all along." I took a deep breath. "Wiles had a huge battery of powerful techniques to work with, but it took him seven years of hard effort to see how to fit the pieces together. He knew the broad strategy, but he had to get his tactics right. It's an amazing achievement."

Fermat nodded wisely. "And has this proof been verified by others? I know from my own work that it is so easy to make mistakes—"

"Not in every tiny detail. But a surprisingly large number of experts are willing to say right now that they believe it. Mazur has summed up the consensus view: 'It has the ring of truth.' "

I looked at Fermat. "I've done my bit. Now it's your turn. Did you really have a proof? The one that the margin was too small to contain? Do we need all this beautiful but complicated apparatus, or is there still a simple proof to be found?"

"Well," Fermat began, "it's like this. I—" At that moment the Time Machine began to pulsate, alternately fading from view and reappearing. "Oh," Fermat exclaimed. "The Time Traveler warned me of this. Unless I return to my own time immediately, I will remain stranded here forever. Farewell!" And before I could stop him, he had leaped on board, pulled the lever and vanished. I lay back in my recliner, stunned by the experience, disappointed not to have been given the answer to my final question. I felt tired, wiped out. The bees buzzed—

"Wake up, lazybones!" It was my wife, with a plate of charbroiled steak. "Did you see a Time Machine on the lawn?" I asked, still fuzzy. The look on her face was peculiar. "No, maybe not," I said hastily. "Just dreaming."

FURTHER READING

THE LAST PROBLEM. Eric Temple Bell. Edited and updated by Underwood Dudley. Mathematical Association of America, 1990.

THE PROBLEMS OF MATHEMATICS. Ian Stewart. Oxford University Press, 1992.

FERMAT'S LAST THEOREM FINALLY YIELDS. Barry Cipra in *Science*, Vol. 261, pages 32–33; July 2, 1993.

BOOK REVIEWS by Philip Morrison

The Biology of the Mind

Molecules and Mental Illness, by Samuel H. Barondes. Scientific American Library, 1993 ($32.95).

For frontispiece we see Freud himself, brow deeply furrowed in an intricate character that must be cogent in Hebrew—or is it Chinese?—the vision of Ben Shahn. The text ends with a fierce Bosch oil, mocking the medicos in a scene both comic and tragic, the *Extraction of the Stone of Folly* from the bloodied top of a bemused patient's scalp. Between the two are many other paintings, colored diagrams of the suite of human chromosomes and of molecules, a portrait of Santiago Ramón y Cajal at the microscope along with Salvador Dali's impression of one of Cajal's real stained micronets, even a snapshot of four little girls in matching white ruffles, identical quadruplets who would all develop symptoms of schizophrenia by the age of 24.

SCALA/Art Resource

EXTRACTION OF THE STONE OF FOLLY, ***1475–1480, by Hieronymus Bosch***

These signs are diagnostic. The author is a gifted expositor and a most cultivated man who has given us a serious, exciting, up-to-date and compact introduction to biological psychiatry, a path along which Freud tried to walk before even the biggest fallen logs were cleared away. Dr. Barondes is a distinguished psychiatrist (at U.C. San Francisco), "hooked" on molecules since a chance year in the very lab where and when the genetic code for proteins was broken. He aims to show readers "unfamiliar with both fields...how they are coming together" as the pace begins to pick up, and his aim is steady.

The book cleaves in half. Each part deserves a careful report. The molecular half is so much more familiar to readers and reviewer that it will be lightly touched. It is a pleasure to remark that the chapters devoted to the genetics of behavior, both simple and complex, to the molecular mechanism of the genes themselves and in particular to neurons and their circuitry are as readable and concise accounts as any a reviewer can cite. They owe their excellence to the considered examples and the clarity of logical sequence, aided by first-rate diagrams. A few pages on the polymerase chain reaction and its utility as an amplifier for gene identification are striking, as indeed are those on the ion channels that penetrate neural membranes and the processes of neuromodulation they allow.

The depth of our molecular understanding is not yet matched by its specificity or applicability. Only lately has clinical understanding gone beyond the jigsaw-puzzle logic of molecular fit and blockage. It has long been known that atropine, the principle of belladonna, blocks the neural receptor site that binds the neurotransmitter acetylcholine and thus inhibits so straightforward a response as the reflex contraction of the iris of the eye. Heart muscle and peripheral nerve were early, easy victories for the biological approach.

Dozens of neurotransmitters are now known; even a simple gas, nitric oxide, can enter the neuron interior to affect protein. About 40 years ago the neurotransmitter dopamine was found to be reduced in patients with the rigid paralysis of Parkinson's disease, especially low within certain motor-related ganglia. Strangely, a successful antischizophrenic drug caused a similar disorder of motion. The new science had begun.

It was no big step to seek and find related dopamine receptors of graded binding strength and then to find that these receptors populate differently various distinct regions of the brain. The count now is half a dozen dopamine receptors. The newer calming drug clozapine (alas, it has its own cheerless side effects on the white cells) binds well to dopamine receptor sites of a type uncommon in the basal ganglia but frequent in brain regions more important for emotion and thought. The once "unbreakable coupling" between welcome calming action and alarming motion disorders has been broken. Clozapine has no motion effects, and we can say why. It is heartening to learn that cloning by recombinant DNA techniques allowed expression of the receptor protein in human cell cultures, where it could be assayed in detail for drug-binding efficacy. Although no safe bullet is yet known, here is rational hope for the relief of schizophrenia even before full understanding of its subtleties.

The "talking cure," as the celebrated patient Anna O. called her analysis in the 1890s, was close to the source of Freud's psychotherapy of neuroses. In 1987 the very term "neurosis" was dropped from official international diagnostic guidelines on anxiety disorders. The method of psychotherapy proved of little use against the bizarre disturbances of the genetically linked ailment. Indeed, the schizophrenic patient seldom even acknowledges any disorder but concentrates on the "malevolence" of hidden plotters, just as the manic accepts the intensity of sheer excellence and the depressive concedes personal worthlessness.

It is the sufferers from minor anxiety disorders who typically note and dislike their distressing symptoms and seek help in discussion. That is the true domain of psychotherapeutic insight, symbols in dream and in memory given meaning by close analysis of past personal relationships. Yet the more severe anxiety disorders, the obsessive-compulsive ritual repetitions and panic attacks, all made vivid here by the lucid accounts of patients, pass more and more to the side of biology. The talking cure has become less like inquiry and more like reassurance. A clue comes from animal studies. Some dogs of two well-controlled breeds compulsively lick their paws even to the point of ulceration. A champion pointer, Alleghany Sue, "became exceptionally fearful after a normal life as hunting dog." As puppies, her progeny are normal but soon fall prey to panic triggered by specific situations. The parallel to uncontrollable human anxieties is telling. Symbolic analysis is surely not what these facts suggest.

This fine book, aimed at health, does not attempt to assess the wider legacy of Sigmund Freud. He gave his century much more—and less—than a therapy, a philosophical view generalizing hero and myth, will and conscience, from Moses to Oedipus to common slips of the tongue. Even if biology is ascendant for hereditary illness, history still colors whatever we become. One-eared Vincent van Gogh looks steadily at us out of his moving self-portrait, a reminder that mental health is not the only spade that is able to unearth the treasure of human accomplishment.

Virtual Protection?

Jane's NBC Protection Equipment 1992–93, edited by Terry J. Gander. Jane's Information Group, Inc., 1340 Braddock Place, Alexandria, VA 22314–1651 (paperbound, $225).

Herein is seen no panoply of rakish warships or angular jet fighters. This example of Jane's celebrated compilations of military technology worldwide, grim as it is, seems almost homey, for its central topic has the intimacy of breath itself. The form of the work is familiar: country by country and model by model, compact paragraphs catalogue what is currently made and used in each category of protective equipment. The descriptions are cool and specific, usually with a photograph or two. (NBC is the service jargon for nuclear, biological and chemical warfare. Those destructive weapons themselves stay rather offstage, outlined generically in a few tight pages.)

The editor's judgment is firm and clear: "Biological warfare is an uncertain weapon. It takes time...is not always...effective and has a nasty tendency to be harmful to perpetrators as well as victims." Chemical warfare, too, "is no longer a way of inflicting casualties" against a prepared military force. Yet any present toxic threat plainly "reduces the fighting efficiency" of anxious men and women within these suits and shelters. The book, illustrating masks and clothes and other practicalities from all major powers and a dozen or so of the lesser, Belgium to Romania to Taiwan, supports his conclusions.

Some colorful pages are set apart for commercial advertisements. "Prepare for the worst" admonishes a gloomy Swedish firm whose impermeable, disposable five-layered plastic coveralls are modeled in a text photograph by a file of filmy troops, masked, booted, gloved and armed. They are safeguarded for hours against nerve and blister gases and droplets, but surely they are uncomfortable and sweat-drenched long before that.

More alarming still is mass toxic attack on civilians. The Iraq-Iran war was the first in 70 years to revive the chemical warfare of World War I at serious scale. Iraqi mustard gas was released against Iranian infantry, at first with more effect against Iraqi troops themselves, though later with decisive results against unprotected Iranian units. Users understand the needs. Protective products of the Iraqi state factories receive four listings here (standard Iraqi military masks are of Romanian design); Iran does not appear. No chemical attack against urban centers is recorded from that long conflict, although more than 100 Iraqi Scud missiles fell on Teheran and other cities, and the Iranians fired a few Libyan-supplied Scuds against Baghdad. Rumors of toxic warheads were part of the strong psychological result. It was real, not virtual, gas, however, that killed many people in a few Kurdish villages on Iraqi soil; with what agent and by whom sent remain in some doubt.

Collective protection is widespread here for the military—filtration for aircraft and tank crews, command posts and the like. Only one listing is designed for general issue to a civil population: a Tel Aviv firm offers an inexpensive long plastic hood with face window and a belt-mounted, battery-powered, filtered ventilation system. "A special version is available for children." In Helsinki, enough public underground shelter is at hand to receive 120 percent of the city's people, providing carefully filtered air behind blast doors. A few other bystander states are not far behind. Clearly, their goal was in part protection against fallout made in America but powerfully delivered well to the east.

A small sample of this disparate technology may amaze in its variety. A Netherlands chemical firm makes "The Button" for individual soldiers. Take 15 breaths through the button into the mask you are wearing; when there is a safe, nontoxic level of nerve agent outside, a blue color appears on reagent-impregnated papers. The U.S. Marines shopped further upscale: for the Persian Gulf War, they purchased many American-made, versatile badgelike electrochemical detectors, read out by a plug-in electronics module. But the top of the line is the U.S.-German NBC reconnaissance vehicle the U.S. Army dubs the M93E1 Fox, a six-wheeled armored personnel carrier fitted to detect multiple threats. It is able to perform mass spectrometry of air and tire-collected ground samples and to count external radiation. Then it records, maps and indicates dangers right on the ground, ejecting smoke grenades and setting coded markers while it rolls along. The German army provided about 80 of these multimillion-dollar Vernean engines from their large stocks to the U.S. and Coalition forces for the war in Kuwait. In contrast, the former Yugoslavian army has thoughtfully developed a unique protective mask, MZK M-1, for horses and mules, offered in two sizes by muzzle circumference.

Warships, aircraft, even rolling armor, have passed their zenith within the past few years, both in numbers and diversity. Cheaper toxic technology is still growing, but this threat faces limits from the active international campaign for NBC control agreement. (Not a word is said on the matter in an otherwise comprehensive reference.) How long can we survive uncontrolled preparation for such well-engineered wars?

Gravity on Trial

The Rise and Fall of the Fifth Force: Discovery, Pursuit, and Justification in Modern Physics, by Allan Franklin. American Institute of Physics, 1993 ($29.95).

For a generation now, we have produced and observed two heavy unstable neutral mesons, called kaons, with strongly differing lifetimes and decay products but the same measured mass. The short-lived form of kaon decays into two pions, the longer one into three—just what we expect on quantum principles, if the two physical states are each a simple quantum superposition of a neutral particle and its equally neutral, yet distinct, antiparticle.

Real kaons had the temerity to mess things up. Occasional but astonishing decays *of the wrong kind* were first found 30 years ago by James W. Cronin and Val L. Fitch, demonstrating a unique failure of the deep symmetry called CP conservation. It claims that a joint exchange of particle into antiparticle and in the handedness or parity between two states must restore full equality (and the right to mix but not to cross) under the laws. When the Nobel Prize was awarded in 1980 for their still unmatched exception, long confirmed, Cronin remarked a little wistfully: "We are hopeful...that this cryptic message from nature will be deciphered."

Not yet, alas; the weak failure of CP is easily built into electroweak theory but hardly explained. That is not for want of trying; one attractive path was to invoke some novel long-range interaction with an environment that is after all never quite symmetrical. By the early 1980s only one faint hint survived: certain kaon interactions with material targets seemed to vary with energy. Maybe some small outside force would do the trick? Good old gravity, inverse-square and universal, was well tested, both on the lab bench and in space orbits. Only in between, at scales larger than any bench yet far smaller than earth orbit, there might lurk novelty, a small new force less impartial as to source and not so long in reach.

Far from the big rings, the geophysicists too had been at work. Careful measurements with gravimeters supported reasonable doubts about the expected variation of gravity. Lab and field results on its strength seemed to differ slightly; was gravity locally contaminated?

Those two threads, kaons and proposed blemishes on classical gravity, came together first in the hands of Ephraim Fischbach at Purdue in 1986. Early that year he proposed a new Fifth Force, beyond strong, weak, gravitational and electromagnetic. The main support he offered for his daring hypothesis was a surprise. It arose from his reinterpretation of the famous results of Roland, Baron Eötvös at Budapest, work begun before World War I. Eötvös used the delicate fiber-suspended torsion balance to measure tiny forces. He sought to test Einstein's postulate of strict proportionality between inertial and gravitational mass. It proved correct to high precision for a variety of materials.

But the 1986 reworking of the old data showed that the slightly discrepant accelerations of the samples gave no noisy scatter but a straight line once the listing was ordered by the computed baryon number (number of protons plus neutrons) per unit weight. Eötvös of course could know nothing about any neutrons; his posthumous paper came out in 1922. The result of "paleophysics" was plainly deniable. Was it chance? That might be if the meticulous Baron had underestimated his errors. (Likely, we say today.) The bold proposal that a simple Fifth Force could all at once fit the geophysics, the old Budapest data and the K meson was tempting but too imprecise to leave untested. Fifty papers appeared on both sides of the issue during the first busy year. The proponents were peripatetic; an overflow seminar would assemble about every fortnight at some lab somewhere. The popular press zealously followed the explorers. Paleophysics soon gave way to novel experiments: physics cannot simply order the given world but must test by extending experience.

Early in 1987 two impressive new experiments were at hand. One confirmed the original inferences from the Budapest data; the other sharply disagreed. Now there are close to 500 papers in the case, including a goodly number of relevant experiments. By the end of 1990 the Fifth Force was pronounced dead almost without dissent. No such force is seen with relevant range as strong as 1 percent of gravity, couple it to its material sources however you like. K meson mixing seems not to be energy dependent after all; intrinsic properties may yet naturally explain the failure of CP.

The geophysicists met the challenge well. Gravimeters went up tall radio masts and deep down into a borehole on the Greenland ice cap. Most reported a preliminary discrepancy. Time after time their small effects went away on meticulous remeasurement of surface matter. Standard surveys off the shelf proved unexpectedly biased. In North Carolina, for instance, surface-gravity surveys around the high tower had understandably but fatally slighted the lowest-lying places, which tended to be marshy. The original geophysical experimenters found errors, too, and tried again. Some went to hydropower storage reservoirs and canal locks where large nearby masses of water moved controllably nearer or farther. The results: Newtonian.

One positive Fifth Force result stands since early 1987 without known flaw. A hollow copper ball, adjusted to float submerged in icy water, was brought below the cliffs of the Palisades. All noise was adequately controlled, as far as anyone can now see. The float fled from the cliff face at a speed of a few inches a day, although it also moved an inch a day in the unexpected direction along the rockface. Later experiments of similar design done near Florence showed drifts 100 times slower, about the same in both directions. This work, on Galileo's own turf, led to an upper limit on the Fifth Force 50-fold smaller than that found along the Hudson. Some unknown convective drift seems present in the earlier work.

Allan Franklin is a keen physicist-turned-paladin for the entire profession in a philosophical battle he has waged now for years. This concise, intimate story of five years of inviting missteps is like the script of a compelling courtroom drama, the verdict based on the whole richness of the evidence. "To be in Error, and to be cast out," said the poet Blake, who also knew much of failure, "is also part of God's Design."

ESSAY by Bruce Russett

Peace among Democracies

A vision of peace among democratically governed nation-states has long been invoked as part of a configuration of institutions and practices that reduce war. In 1795 Immanuel Kant spoke of perpetual peace based in part on states that shared "republican constitutions." Woodrow Wilson expressed the same conviction in his 1917 war message to Congress when he asserted that "a steadfast concert of peace can never be maintained except by a partnership of democratic nations." This vision once sounded utopian. But now, at the end of the 20th century, it is newly plausible.

The research of social scientists points to an irrefutable and tantalizing observation: stable democracies are unlikely to engage in military disputes with one another or to let any such conflicts escalate into war. In fact, they rarely even skirmish. Democracies are more likely to accept third-party mediation, to reciprocate one another's behavior and to settle their disputes peacefully. At the same time, they are about as prone to violence in their relations with authoritarian states as authoritarian states are toward one another.

In the modern world, a "democracy" means a country in which nearly everyone can vote, elections are freely contested and the chief executive is chosen by popular vote or by an elected parliament. The more democratic a nation is, the more peaceful its relations with other democracies are likely to be. Since 1946, pairs of democratic states have been only one eighth as likely as other kinds of states to threaten to use force against each other and only one tenth as likely actually to do so. Established democracies fought no wars against one another during the entire 20th century. (Although Finland, for example, took the Axis side against the Soviet Union in World War II, it nevertheless engaged in no combat with the democracies.)

Moreover, the fairly tranquil relations among democracies are not spuriously caused by some other shared influence such as high levels of wealth or rapid growth or ties of alliance. Peace reigns among democracies even when these other favorable conditions do not apply. These peaceful relations are not limited just to the rich industrialized countries or to NATO members. During the cold war, the phenomenon was not maintained simply by pressure from a common adversary, and it is outlasting that threat.

Peace among democracies derives in part from cultural restraints on conflict, primarily from the belief that it would be wrong to fight another democracy. This view extends into the international arena the norms of live-and-let-live and peaceful resolution of conflict that operate within democracies. Institutional constraints also play an important role: it is a complicated procedure to persuade the people, the legislature and other independent institutions that war is necessary. Consequently, in a dispute with a democracy another nation can expect that there will be ample time for conflict-resolution processes to be effective and virtually no risk of incurring surprise attack.

Evidence supports both the cultural and the institutional explanations; they reinforce each other. The cultural explanation may be somewhat more powerful, however. A culture of democracy, manifested in the absence of violence in domestic politics and the duration of stable democratic regimes, seems to exert a somewhat stronger influence on peace among democracies than do particular institutional constraints.

Nonindustrial societies, studied by ethnographers and anthropologists, also show less war among democratically organized polities. These units lack the institutional constraints of a modern state. Nevertheless, groups that are organized democratically fight one another significantly less often than do groups that have more authoritarian rule. The relation between democracy and peace is weaker and less consistent in preindustrial societies than in the modern international system, but to find it there at all shows that the pattern of peace among democracies is not limited to contemporary Western countries.

The more democracies there are in the world, the fewer potential adversaries the U.S. and other democratic nations will have and the wider the zone of peace will spread. The emergence of new democracies with the end of the cold war presents a chance for change in international relations more fundamental even than the opportunities created by the end of the Napoleonic Wars and World Wars I and II. Freedom House, a New York City–based research organization, has for more than 20 years graded countries by their types of political systems. At the end of 1991, for the first time ever, close to a majority of states (91 of 183) approximated reasonable standards employed for judging whether or not a country is a democracy; another 35 were in some form of transition to democracy.

The real possibility for widespread peace provides a compelling reason to strengthen democracy in the successor states of the former Soviet Union. There is no massive American Marshall Plan, and outside influence will be limited in any case. Yet a favorable international environment can make a difference. Western states can offer some economic help. International organizations can promote human rights and democracy. Successful transitions to democracy in some countries can supply a model for others. A stable and less menacing international political environment can assist the emergence and consolidation of democratic governments.

Democracy in many of these nation-states may not prove stable. This global wave of democratization may crest and fall back, as earlier ones have done. But states probably can become democratic faster than they can become rich. The cold war policy of containment succeeded. The chance for wide democratization now offers a new policy of building democratic peace. If it can be grasped and consolidated, world politics might be transformed.

BRUCE RUSSETT is Dean Acheson Professor of Political Science at Yale University. His most recent book is Grasping the Democratic Peace: Principles for a Post–Cold War World *(Princeton University Press, 1993).*