SCIENTIFIC AMERICAN

FEBRUARY 1994
$3.95

Do aerosols slow climatic warming?

Halting the spread of AIDS.

Can particle physics come back?

***Digital forgery** can create photographic evidence for events that never happened.*

SCIENTIFIC AMERICAN
February 1994 Volume 270 Number 2

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Second-class postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian GST No. R 127387652. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111, or fax: (212) 355-0408.

DEPARTMENTS

THE COVER image was created in a computer by blending an 1863 photograph of Abraham Lincoln with a publicity shot of Marilyn Monroe made in 1955. Both images were scanned and then digitally manipulated; a description of the process appears on page 72 of the article “When Is Seeing Believing?” by William J. Mitchell. The ability to transform photographs in this way has brought to an end the 150-year period during which photography seemed unassailable. And it has left us with the task of learning to view photographs with a new wariness.

THE ILLUSTRATIONS

Digital image by Jack Harris/Visual Logic; original photograph of Abraham Lincoln by Alexander Gardner, Bettmann Archive; original photograph of Marilyn Monroe courtesy of Personality Photos, Inc.

Page	Source
39	J. Langevin/Sygma
42	Thomas Hatwell/Sygma (*left*), Artin/Sygma (*right*)
43	M. Shandiz/Sygma (*top*)
45	Les Stone/Sygma
48–49	Roberto Osti
50–51	Jeffrey T. Kiehl and Bruce P. Briegleb, National Center for Atmospheric Research; color manipulations by Jason Küffer
52	Runk/Schoenberger, Grant Heilman Photography, Inc.
53	Tad Anderson, University of Washington
59	Tomo Narashima
60	William McGinnis (*top*), Tomo Narashima (*bottom*)
61	Tomo Narashima
64	William McGinnis (*bottom*)
65	Jared Schneidman/JSD
66	Tomo Narashima
68–69	Original photograph by AP/World Wide Photos (*left*), image from Paul Higdon/*New York Times* (*center*), image from Angela Perkins (*right*), color manipulations by Jack Harris/Visual Logic (*top center, top right* and *bottom*)
70	Reuters/Bettmann (*top*), Department of Defense Photo (*bottom*)
71	Original photograph by NASA; digitally manipulated version courtesy of Time, Inc., Picture Collection
72	Original photograph of Abraham Lincoln by Alexander Gardner, Bettmann Archive; original photograph of Marilyn Monroe courtesy of Personality Photos, Inc.; digital images by Jack Harris/Visual Logic
73	Image by Wade Hokoda
77	Robert J. Sica, University of Western Ontario
78	Boris Starosta
79	Guy Plante, Laval University
80	Boris Starosta (*top*), Terry Byers, Lockheed Corporation (*left*), Guy Plante (*right*)
81	Robert J. Sica
82	Andrew Lichtenstein/Impact Visuals
83	Peter Haley/*Morning News Tribune,* Tacoma, Wash.
84–86	Guilbert Gates/JSD
87	Andrew Lichtenstein/Impact Visuals
88	Guilbert Gates/JSD (*left*), courtesy of Directie Secretariaat GG&GD, Amsterdam (*right*)
90–91	Roberto Osti
92	Patricia J. Wynne
93	Ronald Orenstein/Animals/Animals (*left*), Francisco Erize/Bruce Coleman, Inc. (*right*)
94	Michael O. Woodburne, University of California, Riverside (*bottom*)
95	Patricia J. Wynne
96–97	John Bird, Superconducting Super Collider Laboratory
102	David Sams/Texas Inprint (*top*), Fermilab Visual Media Services (*bottom*)
103	CERN Media Services (*top* and *bottom*)
104	Frank Veronsky
105	Argonne National Laboratory
116	Kathy Konkle

SCIENTIFIC AMERICAN®

Established 1845

SCIENTIFIC AMERICAN, INC.

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

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

A Sunken Treasure

In "Raising the *Vasa*" [SCIENTIFIC AMERICAN, October 1993], Lars-Åke Kvarning makes the point that many of the items on board the *Vasa* were still intact when the ship's hull was revealed. It is interesting that the sails, though water damaged, had not entirely decomposed. What materials were used to make them? Were they protected by the sail locker?

JEFFREY ENDY
Dauphin, Pa.

Kvarning says the helmsman steered the *Vasa* by tilting the whip staff and moving it up and down. He rightly describes this practice as strangely awkward. According to the references I have seen, during the 17th century the whip staff was used this way to steer ships in fine weather, but the tiller was also controlled by tackles attached to the sides of the ship. Is there any evidence that on the *Vasa* the whip staff was not used primarily as a telltale for indicating the position of the tiller?

R. B. ELLIOTT
Dublin, Ireland

Kvarning replies:

The sail locker on board the *Vasa* did little to protect the sails during their centuries under water, but the condition of the cloth improved the deeper we went into the folded layers. The sailcloth was of two types. The one used in the smaller sails was close-textured. The other was coarser. The condition of the fibers was so poor that it was not possible to determine their material, but the coarser cloth was probably hemp and the finer one linen.

There are no signs that below the helmsman's deck the tiller had been attached to the ship's sides by tackles. On the other hand, this was the *Vasa*'s maiden voyage, and supporting tackles could easily have been attached later—had there been an opportunity.

AIDS and Heterosexuals

Warner C. Greene, author of "AIDS and the Immune System" [SCIENTIFIC AMERICAN, September 1993], should be commended for his remarkably clear, concise description of viral mechanics. But his reference to "new infections—the majority now from heterosexual contact" is misleading. The majority of new infections have always been from heterosexual contact, as far back as the AIDS epidemic can be traced. Although homosexuals, hemophiliacs and people who inject drugs have borne the brunt of the epidemic in industrial countries, they have always constituted a minority of the world's HIV infections. Thus, AIDS was, is and will continue to be primarily a heterosexual disease.

RUSSELL MILLS
San Francisco, Calif.

Overdue Credit

"Sentries and Saboteurs," by W. Wayt Gibbs ["Science and the Citizen," SCIENTIFIC AMERICAN, October 1993], is an excellent review of new tumor therapies. I would like to point out, however, that the idea of inserting a herpesvirus gene into tumor cells and killing them with ganciclovir, which was attributed to Kenneth W. Culver, was generated several years earlier by our group. Although Culver may have had the idea independently, at least one of his collaborators attended a meeting where our work was presented in 1990, at least a year before Culver claims to have had the idea.

XANDRA O. BREAKEFIELD
Department of Neurology
Massachusetts General Hospital
Boston, Mass.

Energetic Thinker

Let me add a historical note to the interesting and informative article by Narain G. Hingorani and Karl E. Stahlkopf ["High-Power Electronics," SCIENTIFIC AMERICAN, November 1993]. In 1967, very early in the history of thyristors, Richard Cassel (now at the Stanford Linear Accelerator Center) proposed the thyristor-based magnet power-supply system used by the Fermilab Main Ring.

At its peak capacity of 400 gigaelectron volts, that system can store more than 100 megajoules of energy and has a power dissipation of more than 20 megawatts—a giant step beyond traditional synchrotron power supplies. By using the utility grid for energy storage, Cassel's system avoided all the maintenance and safety problems of traditional systems. It also had greater operational flexibility (which made it possible to tune the particle oscillations over a wide range) and a learning capability (so the power regulation improved from pulse to pulse).

Cassel's magnificent power supply was a great help in the initial operation of an accelerator more than 10 times larger than any of its predecessors. It has been copied for almost every synchrotron built since then, so it must be doing something right.

FRANCIS T. COLE
Naperville, Ill.

Here We Go Again

I now see that Fermat has played the greatest practical joke of all time on us for 350 years! But even with his hasty retreat at the end of "Fermat's Last Time-Trip," by Ian Stewart ["Mathematical Recreations," SCIENTIFIC AMERICAN, November 1993], he could not get back to the 17th century fast enough to write the proof in the book margins!

P.S. I have found a truly remarkable method for time travel, but this postcard is too small for its description.

VIKTORS BERSTIS
Austin, Tex.

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

ERRATA

The credits for the September 1993 issue neglected to mention that the illustration on page 69 was based in part on work by Karen Jacobsen and Dennis G. Osmond of McGill University.

"Never Give a Sucker an Even Break" ["Science and the Citizen," October 1993] should have attributed the game strategy "Pavlov" to David Kraines of Duke University and Vivian Kraines of Meredith College, who coined that name to refer to a class of learning rules.

50 AND 100 YEARS AGO

FEBRUARY 1944

"If your tire treads are wearing thin and you think something should be done about it, you are dead right. And something is being done. Synthetic tires are good now, but will be excellent. After performing the astounding miracle of creating in little more than two years a totally new complex industry able to produce synthetic rubber at a rate faster than Americans have ever used the product of rubber trees, American enterprise and ingenuity are now busy with the next task: That of making synthetics so good and so cheap that we shall never wish to return to Nature's rubber again."

"The automatic pilot has deservedly earned a great reputation for itself. But there has always been the feeling that it would not quite do the job in very rough weather. Now Wright Field has permitted the announcement to be made of a new electronically controlled automatic pilot developed by the Minneapolis-Honeywell Company. The sensitivity of the electronic mechanism is such that it returns the plane almost immediately to its course despite cross currents, wind variations, and air blasts from exploding anti-aircraft shells."

"The recent decline in the rate of discovery of new petroleum fields in this country has given rise to the question of what we can do to meet the demands of an air-minded and automotive postwar age. Great Britain, Germany, and Japan are making synthetic oil and gasoline. Now is the time to conduct a rigorous research program so that methods will be available to supply necessary liquid fuels from American coals when the petroleum supply begins to fail."

SCIENTIFIC AMERICAN

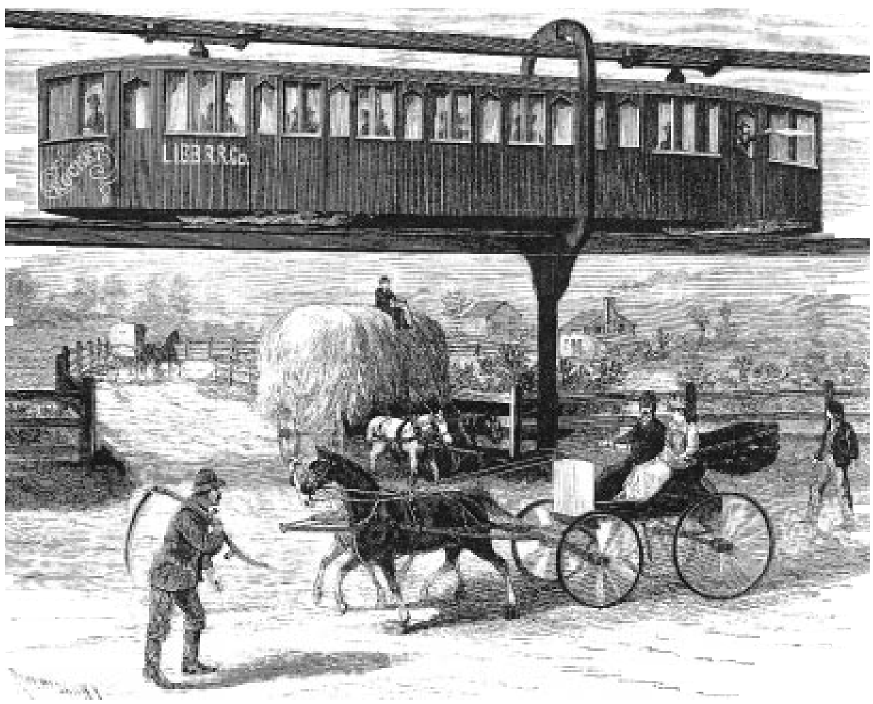
FEBRUARY 1894

"On the 30th day of January, 1894, the Bell telephone patent expired and the invention became the property of the public; so that whoever desires to do so can make, buy or sell telephones without fear of infringing on the rights of any one. This applies only to the hand instrument now used as a receiver. Patents for other telephone apparatus still remain in force; but enough is available for actual service. With two hand instruments and a suitable call, telephone communication may be maintained, under favorable conditions, over a line eight or ten miles long, no battery being required."

"A solution to the problem of connecting the European continent with England by railway seems to be meeting with favor in England. It consists in the establishment, under water, of one or more metallic tubes capable of giving passage to a railway. According to calculations, the total cost of the tubular railway ought not to exceed 375 millions at a maximum, and the construction of it might be effected in five years."

"We now know that the cholera germ is found in the human body only in the intestines; that it is not communicated directly from person to person, but the alvine evacuations of the victims find their way, generally through water, into the bowels of susceptible persons, who then become additional victims; that this germ finds a breeding place in damp soil and in stagnant pools and in running streams containing organic matter; that it is quickly destroyed by the official germicides, by drying, by acids, and by temperature below 56 degrees or above 126 degrees F. It is the application of exact knowledge that has confined the cholera to the quarantine dominions at New York, thus preventing its diffusion in the United States."

"The need of the day is rapid transit. The illustration (*below*) shows one of the last developments in true rapid transit—the Boynton Electric Bicycle Railroad—of which a line is now in process of erection across Long Island, from Bellport to the Sound. The idea of the bicycle railroad is to provide a system of transit whose speed may be from seventy-five to one hundred or more miles an hour. In the railroad in question, a narrow car with sharpened ends is employed, and is mounted upon two wheels, one at each end, and travels upon a single rail. It has the equilibrium of a bicycle, and like the latter disposes at once of the violent transverse wrenching strains which affect four-wheeled vehicles of everyday type. It is peculiarly well adapted for electric propulsion, the overhead rail giving a place for the current main."



The Boynton elevated bicycle railroad

Nobel Notes

Our man in Stockholm reports on the ceremonies

In early December the city of Stockholm enjoys only about six hours of daylight. But the concentration of scientific, economic and literary luminaries that descends on its charming 19th-century precincts to celebrate the awarding of the Nobel Prizes renders solar radiation superfluous.

A prize as famous as the Nobel carries with it the power of celebrity, whether the winners like it or not. Each laureate must adapt to the significance and implications of that power. Richard J. Roberts of New England Biolabs, who shared a Nobel with Phillip A. Sharp of the Massachusetts Institute of Technology for discovering that the genes of higher animals are split into active and inactive parts, referred to himself as a "prize virgin" and expressed a quiet joy at being selected as a Nobel winner. But when asked if the award gave new impetus to his work, he was quite emphatic: "Oh, no, no. The research is its own reward."

Douglass C. North of Washington University took the Nobel as a particularly personal triumph. He and his co-winner, Robert W. Fogel of the University of Chicago, have followed an untraditional methodology of applying quantitative methods to economic history. "At the press conference after the Nobels were announced, people asked me, 'Does this prize validate your approach?' and I told them, 'You bet it does!' " he exclaimed, his eyes glowing and fist clenched.

PRESSENS BILD AB

JOSEPH TAYLOR receives a Nobel Prize in Physics from King Carl XVI Gustaf, an honor he shared with Russell Hulse of Princeton University, his former graduate student.

Joseph H. Taylor of Princeton University, who co-discovered an unusual binary pulsar that has proved to be a valuable laboratory for studying Einstein's theory of relativity, has been quick to share credit with his many collaborators. He also made a point of inviting Jocelyn Bell Burnett of Britain's Open University to attend the Nobel festivities. In 1967, as Jocelyn Bell, she detected the first pulsars in collaboration with her thesis adviser at the University of Cambridge, Antony Hewish. She did not share in the subsequent Nobel Prize, however—a sharp reminder that the Nobel Foundation's power to elevate also confers the power to exclude.

The personalities of the laureates showed up strongly in their Nobel lectures as well. Michael Smith of the University of British Columbia began his chemistry prize lecture with a methodical, technically phrased survey of the history of genetics. He gradually focused on his own work in site-directed mutagenesis, a process that allows the study and manipulation of proteins by specific alteration of the DNA that codes their structure.

The preceding lecture, by Kary B. Mullis, a biotechnology consultant, could hardly have struck a more different tone. He presented a resoundingly personal story of his discovery of the polymerase chain reaction (invariably shortened to PCR). The process provides a fast and easy way for biologists to make billions of copies of a single strand of DNA. PCR has tremendously facilitated work in virtually all aspects of molecular biology, from DNA fingerprinting to the diagnosis of genetic disease.

Mullis described the research that led to PCR as just one component of his life. He recounted that after graduate

COREY S. POWELL

LATE-NIGHT FESTIVITIES following the Nobel banquet brought out the laureates' less serious side. Here Taylor treats the medical students to some fancy pickin'.

school he had hoped to become a writer, "but my characters were flat, so I had to get a job as a scientist." Above all, Mullis ran against the grain by relating that, in the end, the thrill of the discovery of PCR could not compensate for the emotional devastation produced by the disintegration of his relationship with his girlfriend.

Mullis's lecture aroused strong reaction from the audience, especially among the group of young students—mostly female—who mobbed him afterward. Is this kind of adulation more satisfying than winning the Nobel Prize? "The two go hand in hand," he said, grinning, "but I had groupies even before the Nobel Prize."

The formal awarding of the Nobels took place on December 10, the anniversary of Alfred Nobel's death, in the Stockholm Concert Hall. There the laureates joked nervously with one another as they awaited their turn to receive their diploma and medal from King Carl XVI Gustaf of Sweden.

A feast in the Stockholm City Hall sealed the celebration. Torches illuminating the path to the entryway reflected ecstatically off the surface of Lake Mälaren. Inside, 1,300 guests sat at 63 tables distributed through the vast Blue Hall. At one point, the stewards pouring the wine abruptly drew back and began to sing; they turned out to be Orphei Drängar, the renowned men's choir from Uppsala. Just before dessert, soprano Barbara Hendricks performed amid an artificial snowfall, beneath a convincing canopy of stars.

After dinner, five of the laureates gave the traditional speech of thanks, among them writer Toni Morrison, who poetically conjured up the spirits of literature winners yet to come. When the banquet dispersed at midnight, students and some of the Nobelists found their way to the Medical Students' reception—a traditional but unofficial event at the Karolinska Institute. In a crowded room vaguely resembling a medieval beer hall, students entertained the laureates with, among other things, fire juggling, a beer-bottle orchestra and a skit explaining the possible significance of split genes.

Two of the laureates returned the favor. Taylor donned a funny nose, glasses and a guitar; his wife joined him for a spirited if slightly wobbly rendition of "This Land Is Your Land." Mullis later contributed some hoarse singing of his own, along with a bit of free-form stand-up comedy that included a brief parody of the King of Sweden.

A light snow was falling as the Nobel laureates and their families gathered in the lobby of the Grand Hotel to depart from Stockholm. The quietly familiar conversation and warm smiles attested to the shared intensity of the past week's events. But representatives from the South African government and the African National Congress had begun to fill the Grand Hotel, and the mood of the lobby had started to change. The time had come for the Nobel cycle to begin anew. *—Corey S. Powell*

Reflections in a Quantum Well

Like matchmaking relatives, physicists have for many years been trying to marry superconductors with semiconductors, in the hope of having resistanceless electronic circuits as offspring. Although they will not be sending out birth announcements soon, the mating attempts themselves are proving to be a fascinating study. Recent results by Herbert Kroemer, Chanh Nguyen and Evelyn L. Hu of the University of California at Santa Barbara have demonstrated that an unexpected mechanism mediates superconductivity across a thin piece of semiconductor. The mechanism, called multiple Andreev reflections, also offers researchers a bonus mystery: the reflections behave inexplicably when exposed to a magnetic field.

Superconductors carry electricity without resistance because the electrons in them combine in twos to form so-called Cooper pairs. By dancing in step, the members of a pair manage to avoid bumping into each other and thus to move without resistance. The Cooper pairs can also "leak" through the superconductor, penetrating an ordinary conductor to some extent. This leakage, referred to as the proximity effect, enables two superconductors to transmit the resistanceless flow of current across an intervening substance.

Kroemer and his colleagues decided to see what would happen if they stretched the distance between superconducting contacts beyond that at which the proximity effect can happen. To do so, they created a "super-semi-super double heterostructure." That is, they sandwiched an indium arsenide semiconductor between two superconducting niobium contacts spaced a few tenths of a micron apart. The indium arsenide was structured as a quantum well—essentially a thin channel that confines electrons to two dimensions of movement. The quantum well permitted high concentrations of mobile electrons (in effect, creating a "sea" of negative charge).

The workers expected to see a certain level of resistance. Instead they found an unusual conductance peak that could not have been caused by Cooper pairs entering the quantum well. "Our contact resistance data are incompatible with the proximity effect as currently understood," Kroemer says. "The idea that the Cooper pairs penetrate into the semiconductor itself is suddenly in question and needs to be reexamined."

Rather what may mediate the superconductivity are multiple Andreev reflections, a

Too Little, Too Late?

A treatment for heart attack may be dangerously underused

A thrombolytic agent can save your life if you suffer a heart attack. But in the U.S., if you are old or slow in getting to the hospital, your chances of getting one may be disturbingly worse than you'd like. Surveys show that only about a third of all heart attack patients receive a thrombolytic—roughly half of those who may be eligible and far below the 85 percent mark attained in parts of the U.K.

Moreover, even patients who do get a thrombolytic must often wait almost an hour and a half for it, a delay that significantly reduces the drug's effectiveness. By one estimate, 14,000 more lives might be saved annually if physicians used thrombolytics sooner and more liberally. "I think the situation is improving, but it's woefully inadequate," remarks Andrew J. Doorey of the Medical Center of Delaware.

Streptokinase, tissue plasminogen activator (TPA) and other thrombolytics work by dissolving the blood clots that block coronary arteries and cause heart attacks. At least one study found that administering these agents within an hour of the onset of chest pain cut mortality by 90 percent, although most estimates put the benefit at a more modest 50 percent. Unfortunately, that gain decreases when treatment is postponed, and most patients do not reach an emergency room until at least four hours after their heart attack begins. Still, thrombolytics reduce mortality by 30 percent when given within the first six hours and by about 15 percent between the sixth and 12th hours.

The drawback of the drugs is that they promote bleeding and raise the odds of a potentially fatal stroke from an intracranial hemorrhage. Physicians have therefore tended to prescribe clot-busters only for the minority of patients who offered the best ratio of benefits to risks. "Interfering with the body's blood-clotting mechanism is a serious business," cautions H. Vernon Anderson of the University of Texas Health Science Center. "You want to be very, very careful."

Last fall in the *New England Journal of Medicine,* Anderson and James T. Willerson of the Texas Heart Institute

phenomenon the existence of which Aleksander F. Andreev of the Institute for Physical Problems in Moscow proposed in 1964. At the super-semi interface, an electron from the well enters a superconductor to form a Cooper pair. As it does so, it leaves behind a positively charged "hole" in the sea of electrons in the well. The hole is a kind of mirror image of the electron. According to theory, the hole moves along a time-reversed path of the original electron—that is, the hole travels to the other side of the well.

Once the hole reaches the other interface, it breaks up a Cooper pair in the other superconducting contact. One of the Cooper electrons destroys the hole; the other takes up this annihilation energy and shoots across the well back to the other side. The process can repeat once this electron moves across the interface and forms a Cooper pair. In theory, the cycle can go on forever.

More startling was the effect's dependence on an external magnetic field. Kroemer found that a rising magnetic field caused resistance to increase episodically instead of smoothly. The jerkiness or bumpiness of the increasing resistance should involve a fundamental parameter—the flux quantum. The flux quantum dictates that bundles of magnetic-field lines penetrating a sample must take on a particular, discrete value. Instead, Kroemer reports, the measured value is smaller than the predicted one by a factor of four to five.

So far no good explanation exists for the oscillations. One speculation is that the magnetic-flux lines assume the form of a lattice as they penetrate the semiconductor. When the magnetic field is increased, the entire lattice shifts suddenly to accommodate the new flux bundles. Kroemer plans to look for the effect in new samples before submitting his results for publication.

Multiple Andreev reflections may be more common than previously observed. For instance, Alan W. Kleinsasser of the IBM Thomas J. Watson Research Center and his colleagues will report their observations of the reflections in a quantum structure known as a tunnel junction. So whereas the birth of superconducting computers remains distant, investigators are finding plenty of excitement during the courtship period. Kroemer explains: "The physics for now takes precedence over the hypothetical applications."

—Philip Yam

in Houston pointed out just how careful physicians have been. Thirty percent of all patients with heart attacks do not get thrombolytic therapy, because they arrive at the hospital more than six hours after pain begins—too late, in the physician's opinion. Because the elderly are at higher risk for stroke, 15 percent are considered too old. Another 25 percent are disqualified because their electrocardiograms do not suggest that thrombolytic therapy would be helpful or because they seem prone to bleeding.

Are those criteria too conservative? The mortality for patients in clinical trials of thrombolytic therapy is typically 2 to 10 percent; for those excluded from therapy, it is 15 to 30 percent. These alarming numbers suggest that unless the risks of stroke and bleeding would be far higher in the excluded groups—an assumption that is especially questionable for people who have just arrived too late—those patients, too, would benefit from thrombolytic therapy. Indeed, Anderson and Willerson note that patients older than 75 years were among the groups who most benefited in clinical trials because they suffer the most heart attacks.

Doorey believes perhaps as many as 60 percent of all heart attack victims might qualify for thrombolytic therapy. In the December 1992 issue of the *Journal of the American Medical Association,* he, Eric L. Michelson of Hahnemann University and Eric J. Topol of the Cleveland Clinic Foundation tried to estimate the potential impact of thrombolytics. They concluded that expanded use of thrombolytics could triple the number of lives saved, from 7,200 to 21,950 annually.

Some advocates insist that compelling evidence for a broader use of thrombolytics has existed since at least 1988, when the Second International Study of Infarct Survival (ISIS-2) was released. And many physicians in Europe seem to have concluded that aggressive use of thrombolytics is warranted. A report in the *Lancet* last October claims that 85 percent of the heart attack patients in some English hospitals received thrombolytics. It also points out, however, that regional hospitals varied greatly in their practices: some hospitals used them only half that often.

"It's hard to have a handle on how much underutilization there is right now," Topol argues. "It appears to be much less than it was even a couple of years ago." For patients older than 75 years, he says, the rate of treatment has jumped from 2 to 15 percent. Clinical records from the Global Utilization of Streptokinase and TPA for Occluded Coronary Arteries (GUSTO) trial, which he supervised, suggested that "we're treating well over 80 percent of the appropriate patients."

RON COPPOCK *Liaison International*

HEART ATTACK PATIENTS can often benefit from getting clot-busting drugs, but many who should be eligible may still not be receiving them.

But Rory Collins of Radcliffe Infirmary at the Clinical Trial Service Unit of the University of Oxford dissents from that view. A leader of the ISIS-4 trial released last November, Collins states that "the U.S. was down in the lower end and the U.K. was up in the top end" in frequency of thrombolytic use. "I think a lot of people are still uncertain about whether they should be treating beyond six hours," he ventures. "That is changing, but it may be changing at different rates in different places."

Resistance to a therapy that may routinely kill one or two out of every 1,000 patients is understandable in a profession trained to obey the motto *primum non nocere,* "first do no harm." Emergency room internists must make rapid decisions, on the basis of incomplete information, about the care of patients they have usually never seen before. They often weight their own experience and that of their colleagues more heavily than clinical reports—which may explain why the use of thrombolytics tends to be higher in hospitals that have participated in clinical trials. Fear of liability also haunts some U.S. doctors, Doorey observes.

Better prescriptive guidelines may soon appear in an upcoming paper in the *Lancet,* in which Collins and his colleagues make new recommendations for giving thrombolytic therapy to the elderly, people with histories of strokes and other categories of patients. "It puts together all the data on the subgroups we have from the large-scale trials, and it helps to guide treatment for individuals," he says.

Quite aside from the issue of whether more categories of patients should receive thrombolytics, most experts believe the therapy should be administered much more promptly. Studies show that from the time eligible patients in the U.S. reach an emergency room, they must wait an average of about 85 minutes before their thrombolytic therapy begins. That delay not only lowers the benefit of the thrombolytics, at some hospitals it pushes patients outside the accepted interval for treatment.

Doorey and others are convinced this "door-to-needle time" can and should be cut to 20 minutes or less. To facilitate the treatment, "most good hospitals are setting up multidisciplinary, interdepartmental teams," Doorey explains. "They're like the code-blue teams that treat trauma." Some proponents have suggested that thrombolytic therapy could be started in ambulances en route to the hospital, but the evidence for the benefit of this controversial practice is uncertain.

"We're talking about up to a 90 percent reduction in mortality from the biggest killer in the Western world," Doorey insists. "That, I think, is the biggest medical advance in this century outside of antibiotics." But it cannot live up to that potential unless physicians use it more often. *—John Rennie*

No Global Warming?

CO_2 readings on Mauna Loa show declining emissions

Since 1958, when researchers first began to measure the rate at which carbon dioxide accumulates in the atmosphere, they have seen a consistent increase, perturbed only by minor seasonal fluctuations. Then, about four years ago, the trend began to waver. First a decline set in, followed by a plateau. After that, the decline resumed—sharply. The event has left scientists, including those at the observatory on Mauna Loa in Hawaii, established by the late Harry Wexler to make the measurements, wondering what has happened.

Adding to the confusion, says Charles D. Keeling of the University of California at San Diego, who has operated a gas analyzer at the observatory since its founding, is the fact that accumulation started to slump while the atmosphere was in the throes of an El Niño, a periodic shift in the circulation of trade winds over the Pacific that affects global weather and ocean currents. During an El Niño, such as those of 1982–83 and 1986–87, atmospheric carbon dioxide levels tend to rise faster than they do at other times. Keeling suspects that plants and soils release more carbon dioxide during an El Niño because when an Asian monsoon collapses, it causes drought conditions. Whatever has been reducing contributions of carbon dioxide to the atmosphere had such an impact that it entirely overrode the effects of an El Niño.

Any number of events might have had such climatic clout. Scientists can eliminate only one explanation immediately: the amount of carbon dioxide released from burning fossil fuels has not declined. The next most obvious candidate is the June 1991 eruption of Mount Pinatubo in the Philippines. "The link to the eruption is pretty speculative, but it's an attractive thing to think about because of the coincidence in time," says Ralph F. Keeling, Charles Keeling's son and colleague at U.C.S.D. Of course, discovering whether the mystery source existed at land or at sea would narrow the search further. Unfortunately, different tests have yielded conflicting clues.

The ratio of carbon 13 to carbon 12 in the atmosphere is one such measure. Photosynthesis on land prefers the lighter isotope, whereas gas exchange at sea discriminates only slightly between the two. "We saw the ratio go up, which would imply an increased carbon dioxide uptake by the terrestrial biosphere," says Pieter P. Tans of the National Oceanic and Atmospheric Administration. "But there could be considerable error in that. It is very dependent on how good our calibration is." Indeed, researchers measuring the carbon isotope ratio have reported different results at various meetings over the past year. Charles Keeling's data initially indicated a large sink at sea. After corrections were made to his calibration, the results instead pointed to a sink predominantly on land.

Oxygen emissions, on the other hand, support yet another idea. "It's fairly clear that the land did not behave in a typical way for an El Niño, but the oxygen data suggest that maybe the oceans also behaved strangely," Ralph Keeling says. Just as different flavors of carbon isotopes are preferred by surf-and-turf reactions, so, too, varying proportions of oxygen and carbon are engaged through the formation and consumption of organic matter. In addition, carbon is quite reactive at sea, whereas oxygen is chemically neutral.

After considerable number crunching, these facts taken together imply that if the sink were primarily on land, as the carbon isotope readings suggest, the change in the growth rate of atmospheric oxygen should be nearly equivalent to the recent change for carbon dioxide. In fact, Ralph Keeling has observed oxygen emissions that rose about twice as sharply as the rate by which carbon dioxide emissions fell after the Pinatubo event. This finding indicates that significant changes took place in the oceans.

No matter where this carbon sink existed, scientists face the additional challenge of figuring out how it happened. There are several models based on the fallout from Pinatubo that could conceivably illustrate why carbon dioxide emissions plummeted. Global cooling, measured in the low troposphere via satellite, provides one compelling pathway. Such cooling could affect the balance between photosynthesis and respiration on land and could lead to an increased net uptake of carbon dioxide in the oceans. "It could cause a big, short-term jolt to the carbon balance. In 1994, if the temperature comes back to normal, we should get normal carbon dioxide growth again," Tans notes.

So, is global warming on the way out? Tans does not think so. The decline in atmospheric carbon dioxide accumulation, he believes, is temporary. Ralph Keeling agrees. "That the carbon dioxide growth will stay low is doubtful," he says. "But this is relevant at least in the sense that it shows we don't really know what's happening with respect to the most important man-made greenhouse gas." *—Kristin Leutwyler*

ALBERTO GARCIA SABA

ERUPTION of Mount Pinatubo in June 1991 may be responsible for lower carbon dioxide emissions measured in the atmosphere since then.

Fertile Ground

IVF researchers pioneer the bioethical frontier

When researchers at George Washington University cloned 17 dysfunctional human embryos last summer, they were testing a possible new tool for in vitro fertilization (IVF). Their experiment opened a Pandora's box of hypothetical moral concerns—an increasingly familiar experience in IVF research—among them the prospect that many identical copies of an individual might someday be created. But the public uproar that followed has obscured a much more immediate ethical issue: how and when to test embryos for genetic disorders.

"We have developed tests for cystic fibrosis, Duchenne's muscular dystrophy, myotonic dystrophy, Lesch-Nyhan syndrome, which is a vicious neurological disorder, Tay-Sachs disease, and hemophilia A, which is a clotting deficiency. And we are working on fragile X, an inherited mental retardation syndrome," reports Mark R. Hughes, director of the Baylor College of Medicine's prenatal genetics center.

From a technological perspective, this is a remarkable feat. The technique involves retrieving eggs from a woman's ovary, fertilizing them in vitro and letting them grow to the eight-cell stage. One or two of the cells are then removed from the embryo and analyzed by making millions of copies of one bit of the gene of interest or by injecting fluorescent DNA probes that can be made to home in on certain mutations. If the embryo is judged acceptable, it is transferred (often with several others) back into the woman. The analysis is typically performed in a single day to maximize the odds that at least one embryo will attach itself to the uterine wall and launch a pregnancy.

As of December, researchers could boast of at least 18 such pregnancies. Despite worries about accuracy—the tests have reportedly failed to diagnose afflicted embryos in three cases—the screening techniques are moving rapidly toward clinical use in IVF centers. There are more than 300 such centers in the U.S., most of them private and largely unregulated. Carlene W. Elsner, a clinician at Reproductive Biology Associates in Atlanta, plans to offer genetic embryo testing to patients later this year. "We could do it next treatment cycle if we wanted to," she says.

But some clinical researchers think the tests are not yet ready for commercial use. "The way they are done in the academic setting wouldn't work in the clinic," says Donald S. Wood, vice president of science and technology for IVF America, which operates six fertility clinics. Having to produce conclusive results in a single day places a lot of pressure on doctors—"This is not something you want to rush," he argues.

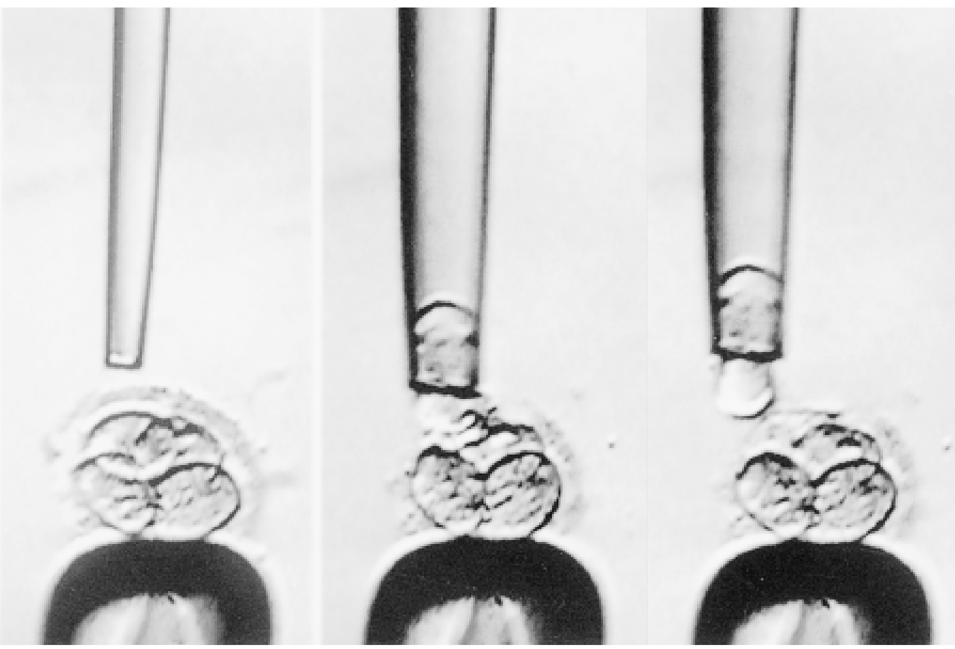
Wood says IVF America has worked out a way to freeze embryos after a cell has been removed so that researchers can do their analysis at a more leisurely pace. Acceptable embryos can then be thawed and returned to their mother during a later menstrual cycle. The company is planning three clinical trials to test the idea.

Meanwhile Hughes is collaborating with Alan H. Handyside of Hammersmith Hospital in London to increase the number of disorders for which a single embryonic cell can be screened simultaneously. "We can currently examine 10 different genetic locations from one cell," Hughes says. "It looks as though you may be able to do as many as 28" using DNA amplification.

In the private sector, research is proceeding along slightly different lines toward a similar goal. Wood thinks the polymerase chain reaction currently used to search for mutations is too slow and destructive. "It would be far better to be able to drop the sampled cell in a cocktail of probes and have the results in 30 minutes," he says, adding that IVF America is developing probes for chromosomal defects that could do just that. Although they could not identify mutations in individual genes, Wood claims that the probes could be removed without damaging the cell, so that another battery of different tests could be run. "We're still three to five years away from clinical use," he says, "but we're far enough along that we know it's going to happen."

The emerging capability to test for many different genetic and chromosomal disorders has some ethicists worried that the technology might be used for screening embryos regardless of any known risk of inherited disease. This is particularly disturbing in the absence of a consensus about what is and is not a disorder. A 1990 survey conducted by Dorothy C. Wertz, a senior scientist with the Shriver Center for Mental Retardation, found that 12 percent of those polled would terminate a pregnancy if they discovered that the fetus possessed a gene for untreatable obesity. (No such gene is known.) Most physicians would consider that unethical. Yet, Wertz reports, in a 1985 poll the "vast majority" of practitioners said they would perform amniocentesis and chromosome analysis of a fetus at a patient's request without any medical reason.

A few IVF clinics have reportedly begun offering to select embryos of a particular gender even for those who have no history of sex-linked



JUAN COTA *Baylor College of Medicine*

GENETIC MUTATIONS can be identified before pregnancy begins by sucking one cell out of an eight-cell embryo and amplifying bits of DNA. The remainder can grow into a healthy baby.

disease. “This is inevitable,” Wood warns. “You’re going to see sex selection become more widespread.” Wertz, who is tallying the results of a recent survey, claims it indicates that “perhaps half of the geneticists in the U.S. have had a request for sex selection.” John C. Fletcher, a bioethicist at the University of Virginia, worries about “selecting embryos for traits that don’t have anything to do with disease. Society has an interest in trying to help people sober up rather than entertaining fantasies about the ideal child,” he says.

Hughes dismisses the notion that embryo testing might lead to an increase in unethical reproductive choices. “Right now you can terminate any pregnancy for almost any reason up until 20 weeks,” he argues. “We think of testing as an alternative to abortion, because it allows you to make the decision before a pregnancy even begins.”

If the debate were limited to test-tube babies, many ethical questions might be moot. After all, of the more than three million couples in the U.S. thought to be unable to reproduce without IVF, only about 20,000 go to clinics every year. Costs of $6,000 to $10,000 per treatment keep many away. Others recoil when they learn that 85 percent of IVF treatments fail to produce babies.

Rapid advances in the art of fertilization promise to improve these odds, primarily in the 50 percent of cases with male factor infertility. By cutting holes in the egg’s tough coating or by injecting sperm directly into the ovum, researchers can now fertilize eggs with even the weakest sperm. If such tricks improve IVF’s dismal success rate, demand for the procedure and for embryo testing could increase.

But John E. Buster of the University of Tennessee–Memphis College of Medicine, among others, is working on a technique that might have a far greater impact. Called uterine lavage, the idea is to wash a naturally conceived embryo from the uterus before it has a chance to become implanted, then test it for genetic defects and return it only if it is healthy. Unfortunately, the chances that just one returned embryo will develop into a pregnancy are slim. If this technical hurdle can be cleared, however, uterine lavage might allow parents to look for mutations in embryos produced the old-fashioned way—at a fraction of the cost of IVF.

As the pace of innovation continues to accelerate, concern is growing among researchers, clinicians and regulators alike that there needs to be more discussion of the issues raised by genetic screening and more oversight of its development. Wood reports that leading

IVF clinics have drawn up guidelines for the industry and are searching for a professional society to enforce them. Self-policing has not worked in the past, however. The American Fertility Society's guidelines tend to lag behind research and have been given no teeth.

According to recent reports by the Office of Technology Assessment (OTA) and the Institute of Medicine, all the necessary rules and laws are already on the books—they are simply not being followed. The Food and Drug Administration has the authority to review all genetic tests before they can be sold for use by doctors. But most genetic testing is offered as a service by research labs and so falls outside FDA purview. Nevertheless, Steven I. Gutman, director of the division of clinical laboratory devices, says the FDA is "considering the possibility of involvement."

Many researchers and clinicians would like to see a permanent national ethics advisory board set up to promote public debate and offer ethical guidelines for research. Four such boards have been set up in the past 20 years. Even the most durable operated for just three years, however.

According to the OTA report, federal regulations have required since 1978 that an advisory board exist in order to review funding requests for research on human IVF. But the Reagan and Bush administrations refused to approve the board's charter, thus imposing a de facto ban on federal funding for human embryo research that stood until it was repealed last June.

In December the National Institutes of Health set up a Human Embryo Research Panel to work out ethical guidelines in a series of public meetings this spring. If the recent cloning controversy and the growing use of genetic testing spawn wider public education and discussion about the proper use of genetic technology, IVF will have been a true pioneer indeed.

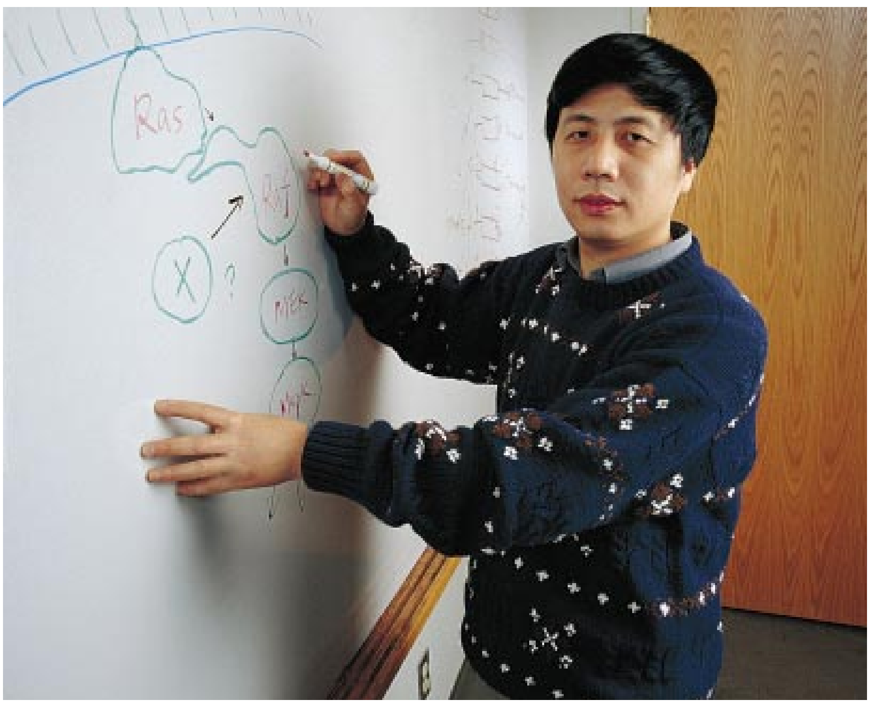
—W. Wayt Gibbs and Tim Beardsley

Design for Living

A signaling pathway found in many species is mapped

An ancient tale describes how three blind men try to identify a curious object using their sense of touch. The object is an elephant, but the men fail to recognize it because each feels a different part of the animal. Cell biologists working in such disparate areas as oncogenesis, cell division in yeast and the sexual anatomy of *Caenorhabditis elegans* (a microscopic worm) have for the past decade or so been groping their way around a molecular elephant. And like the blind men of the fable, they have been unable to agree on what they are examining. Until now, that is.

A remarkable convergence of experimental results has suddenly made the investigators realize they are all looking at different parts of the same thing—and it is a coveted prize. What they have started to discern is one of the cell's principal control mechanisms: a chain of molecular reactions that conveys signals from the cell's surface into the depths of the nucleus. There the signals empower the genes, which change a cell's shape, its activity or its growth. "We are starting to understand the molecular circuitry of the cell," comments Robert A. Weinberg of the Massachusetts Institute of Technology.



STANLEY ROWIN

XIAN-FENG ZHANG of Harvard University sketches part of the Ras pathway, a molecular cascade that controls growth and division in a wide variety of cells.

The essential molecule in the process is the Ras protein, which normally lies just under the cell membrane. The sequence of events is therefore being called the Ras pathway. The key event that starts the dominoes falling is the binding of an extracellular signaling molecule to its surface receptor. That binding causes the receptors to aggregate. The parts of the receptors that lie inside the cell then assume a distinctive enzymatic role: they attach phosphate groups to themselves and to one another at the sites of the amino acid tyrosine. That chemical modification makes the phosphorylated molecules attractive to proteins that carry a specific motif of amino acids called SH2. Some also carry another motif, SH3, that attracts a third set of molecules, which in turn activates Ras.

Until last year, the chain ended there: nobody knew what Ras did. But an array of elaborate experiments by Xianfeng Zhang and Joseph Avruch of Harvard University and Ulf R. Rapp of the National Cancer Institute and others has caught Ras red-handed. The molecule binds to and excites another protein called Raf. And to cell biologists that is a remarkable finding, because Raf in its excited form triggers a cascade of enzymes that ends up in the nucleus. Enzymes called MAP kinases are among this alphabet soup of players, and they, it is thought, can activate DNA binding proteins in the cell nucleus that in turn switch on genes.

An unexpected feature is that the molecules of this bucket brigade apparently physically aggregate to form a macromolecular complex. Tony Pawson of the Mount Sinai Hospital in Toronto speculates that making such complexes "is the only way the cell has resolved how it is going to respond to and integrate a large variety of signals with effects on cell growth."

At the bottom end of the Ras pathway, there is a good deal more hand waving over details than in the well-worked-out earlier steps. "The pathway through which MAP kinase affects transcription is poorly defined—it may work in some general manner, but we certainly don't know at the moment," cautions James E. Darnell, Jr., of the Rockefeller University.

There are still question marks—actually more of them than ever before—but there is also "the backbone of a story" to connect receptors to genes, in the words of H. Robert Horvitz, a researcher at M.I.T. who has been one of the contributors to the recent insights. "The big news is that the pathway is conserved" over the eons of evolution that separate man from yeast, Horvitz says, who studies mutations that affect development in *Caenorhabditis:* "What has emerged is a consensus pathway."

Other pathways to the nucleus besides the one that features Ras are certainly important. But what has impressed many researchers is that independent lines of work using very different organisms have all stumbled onto what seems to be the same biochemical contrivance.

Ras was originally identified more than a decade ago as the product of an oncogene, a gene that in mutated form can cause a cell to become cancerous. Perhaps unsurprisingly, many proteins produced by oncogenes have turned out to be involved in conveying signals, including those that tell cells when to grow and divide. Several oncogenes produce proteins suspected of being connected to the Ras pathway, and biologists are now becoming comfortable with the notion that a fault in a signaling pathway could cause a cell to divide incessantly, as in cancer, or to malfunction in other ways.

Many big pharmaceutical companies, including Ciba, Sandoz, Pfizer and Glaxo, are now working to find simple molecules that interfere with one or more steps in the Ras pathway. And, unlike some drug discovery efforts, evidence exists that the principle can work. The important immunosuppressant cyclosporine is a natural product that blocks an intracellular signaling pathway, as is the experimental immunosuppressant FK506. Ariad Pharmaceuticals in Cambridge, Mass., a company specializing in drugs aimed at disrupting intracellular signaling, is trying to develop an allergy medicine aimed at a Ras pathway target. Peter L. Myers of Onyx Pharmaceuticals in Richmond, Calif., says he is "starting to see hits"—that is, chemicals that seem to block interactions in the Ras chain.

Products that work by interfering with steps in the Ras pathway are probably some years away. But, as Weinberg and Sean E. Egan of the Imperial Cancer Research Fund in London noted in a recent issue of *Nature,* "the satisfaction to be had from the recent discoveries will endure—that of reducing extraordinarily complex phenomenology down to simple, apparently universally relevant, truths." —*Tim Beardsley*

Time-Trippers Beware

Instructions for building a time machine: Take two cosmic strings. Throw them together so that each moves at a speed close to that of light. Fly around both of them, and you will return to the time and place from which you started. (Suggested by J. Richard Gott of Princeton University.) Sounds simple enough. But don't try it, warns Gerard 't Hooft of the University of Utrecht in the Netherlands. You won't just fail—you might destroy the entire universe.

That Nature somehow protects herself from the contradictions of time travel has been conjectured by many physicists. Just how far she will go in self-defense and what means she will employ are questions that now have answers—answers that would make even the most optimistic time traveler cash in his or her ticket.

In principle, the Gott time machine is quite straightforward. The traveler need only induce two infinitely long, parallel cosmic strings, presumed threadlike relics of the big bang, to sweep by each other at speeds near that of light. The strings' center of mass then moves *faster* than light. The vast amounts of energy entailed are equivalent to an intense concentration of mass (remember $E = mc^2$?). The mass warps space-time so acutely that a path looping around the strings can take one back in time.

The problem arises when one attempts to build a Gott time machine in the context of an actual universe. Universes appear to come in two varieties: open and closed. Sean M. Carroll, Edward Farhi, Alan H. Guth and Ken D. Olum of the Massachusetts Institute of Technology have tried to create Gott pairs in an open universe. First, they say, find two slowly moving cosmic strings. Then split each by an explosion. The explosion serves to accelerate two of the resulting four fragments toward one another at relativistic velocity. Unfortunately, in the open universe, which is unbounded, as would be an infinite sheet of paper, there is not enough energy available to push the strings to the needed velocities.

All right, the time-machine builder says, why not start out in a closed universe? In a series of papers in *Classical and Quantum Gravity,* 't Hooft attempts just that—and finds a truly violent outcome. The members of the Gott pair do not pass each other, forming time loops. Instead they trace out chaotic orbits as they move closer and closer together at increasing speed. The tremendous gravitational stresses generated make the universe crumple and fall in toward the strings. As their kinetic energy grows infinite, the universe finally collapses in a catastrophic big crunch. A time-machine ticket-holder will see massive walls closing in while being shredded to spaghetti by the strings speeding through.

Hard times, indeed, but 't Hooft offers some consolation. "Quantum effects," he says, "will probably dilute the big crunch to a big bounce." Out of the shreds of the last universe may be born a new one, albeit a bit late for the time traveler.

So the Gott time machine can never be built. For those hopefuls now looking to wormholes, the chutes connecting distant regions of space-time—oops, there they go, too. A wormhole can be sustained only by negative energy—a no-no. Thus, it will squeeze in and collapse like a punctured balloon, probably forming a black hole, before anything—even light—can traverse it.

Any other new designs that may be dreamed up for a time machine will have to contend with powerful theorems propounded by Frank J. Tipler of Tulane University and Stephen W. Hawking of the University of Cambridge. These theorems attest that within finite regions of space-time, time loops are always accompanied by negative energy—disallowed as unphysical—or by violent objects such as black holes and imploding universes. The scene sketched by 't Hooft shows how such objects can act as Nature's dragons, guarding time machines from fools who would rush in. —*Madhusree Mukerjee*

Laid-Back Leader Rattles the Academy

The National Academy of Sciences in Washington, D.C., is indeed a temple of science. The Great Hall, a domed chamber decorated with filigreed murals and inspirational quotations, houses conclaves of the nation's greatest scientific talents. From the meeting rooms and offices opening off the Great Hall pour reports about the state of the scientific enterprise and how it affects the society that sustains it. Recently the academy, an institution chartered when Abraham Lincoln was in the White House, has been rattled by a new and decidedly unstuffy presence.

A professor of biochemistry and biophysics from the University of California at San Francisco, Bruce M. Alberts, who has been the academy's president since last summer, has brought to the hallowed halls an activist agenda, West Coast informality and a penchant for self-deprecatory humor that has mortified the institution's more straitlaced officials. His public affairs staff has yet to recover from the occasion last fall when Alberts told a group of reporters how he had explained to a senator: "The academy is 1,600 scientists who elect each other. We have a party in April. Otherwise, we don't do anything."

Alberts, who is 54 years old and has been an academy member since 1981, is recognized for his research on proteins instrumental in the replication of chromosomes. He was something of an unknown quantity when he arrived in Washington, because unlike his immediate predecessors at the academy's helm he had not previously occupied any top science post. Indeed, Alberts was not the first choice of the search committee charged with finding a successor to Frank Press, a geophysicist who was president for two consecutive six-year terms (the maximum) starting in 1981.

Despite the $250,000-plus salary and the Watergate apartment that come with the job, several other scientific luminaries with more administrative experience than Alberts declined to let their names go forward as candidates. Among them were Ralph E. Gomory, president of the Alfred P. Sloan Foundation in New York City, and Maxine F. Singer, president of the Carnegie Institution of Washington. It would seem that questions about funding and a changing relationship with the federal government have begun to create a series of problems daunting even to the most politically adept mandarin.

According to Donald D. Brown, a researcher at the Carnegie Institution's department of embryology, who served as co-chairman of the search committee, Alberts filled the bill because he is an active scientist with broad expertise who had also earned high marks as an administrator when he was chairman of his department at San Francisco. And Alberts was, Brown points out, an energetic chairman of the commission on life sciences of the National Research Council, which is the organization, formally distinct from the academy, that performs its scientific assessments. Moreover, he had been a force behind the federal Human Genome Project. "What he says is exactly what he thinks," Brown comments.

If Alberts was not the academy's first choice, then neither was the academy his. Alberts was, he says, "very happy with what I was doing," leading a laboratory. "My image of management was negative," he explains, still looking, in his open-necked shirt, like a bench scientist even as he sits in his high-ceilinged presidential suite. Whatever the high marks he had received, he had not enjoyed his stint as chairman of his department, and he thought the academy would be more of the same.

But friends persuaded him to reconsider. He capitulated, he remarks, when he realized that if he had the institution's top job there were "four or five things I might be able to do that I couldn't do" otherwise. Most of those things turn out to be variations on Alberts's principal preoccupation: education.

Alberts is a man with a mission. His research career had gotten off to a flying start when he made an important discovery about the structure of ribonucleic acid as he worked on his undergraduate thesis at Harvard College in the 1950s. The early success persuaded him to become a scientist rather than a physician, and after a spell at Princeton University he found his way to California.

It was there that he developed his consuming interest in improving scientific literacy, a passion he says was inspired

CHRIS USHER *Black Star*

BRUCE M. ALBERTS holds forth in the National Academy of Sciences's Great Hall. Can a hip lip-shooter find his way through the corridors of power?

mostly by his wife Betty's leadership of the Parent-Teacher Association in San Francisco. Alberts is the principal author of a noted textbook on molecular biology, but he prefers to talk now about another of his achievements, cofounding the University of California at San Francisco Science/Health Education Partnership. The program is a collaboration in which university scientists and public school teachers work to implement fresh approaches to science instruction.

Now Alberts would like to launch a similar scheme on a national level. "A strong motivation for me in taking this job was to see if the academy could fill a real void in leadership nationally" in science education, he states. Alberts's vision is breathtaking in its audacity. He wants to harness not just the academy but much of the country's scientific workforce to his campaign. "I'm not talking about nudging the system; I'm talking about very dramatic change," he declares. "We can use science education as a wedge to change the system, to empower teachers to change the nature of the public school experience."

Heady stuff. But Alberts is confident that scientists will rally to his cause and devote their "tremendous amount of energy and focus and skill" to enhancing education. Within a couple of months of taking office last July, after an uncontested election in February, Alberts had met with the superintendent of schools in Washington with an eye to setting up a demonstration project in the city. And in November he summoned reporters to hear about the launch of Project RISE (Regional Initiatives in Science Education). RISE is a pilot project in which the National Research Council will support regional collaborations between scientists and elementary teachers to promote hands-on science.

The academy is also developing formal national standards for science education. Alberts has talked to Richard W. Riley, the Clinton administration's secretary of education, about initiating legislation to give school districts incentives for adopting them. But there is no shortage of science curricula, Alberts notes, at least in the elementary grades: "It's just a question of having the expertise to do it and the will to do it." He hopes to persuade biotechnology companies, in particular, to follow the lead of other industries by providing challenge grants for schools and districts that adopt new approaches.

"I'm not going to have any problem finding enough scientists and doctors," Alberts asserts. Still, he acknowledges that not everyone is cut out for teaching partnerships. "There are some scientists who alienate me," he concedes. And Alberts says he is aware of the dangers of letting well-meaning researchers loose to lecture teachers or students about scientific specialities. He recalls a soil scientist who insisted there were "eight facts everyone should know about soil science." "Well," Alberts chortles, "I didn't know them myself, and I'm president of the academy!"

The jest is typical of Alberts's style. He is also famous for being absent-minded in true professorial tradition. He once startled his staff by wondering out loud how he was going to run the academy when he is too disorganized to find his glasses. On another occasion he managed to lock himself inside his office and had to be rescued. Whether such endearing foibles will serve him well in a hostile congressional hearing room is a question yet to be answered.

Despite his relaxed attitude, Alberts does not pull many punches. The academy and the research council, together with the academy's sister institutions, the National Academy of Engineering and the Institute of Medicine, employ more than 1,000 policy analysts in Washington. "It's grown to be very large, which is a problem," Alberts states bluntly. In doing so, he echoes rumors that the academy may have to lose large numbers of staff. But Alberts will not confirm suggestions that cuts of more than 20 percent over the next five years are in prospect.

Like any institutional leader, Alberts must address the problem of revenues. Although Press built up the academy's endowment substantially during his 12 years in office, it is still, at $120 million, small for an organization of the academy's size, Alberts points out. The endowment is important because only by using its own funds can the academy initiate studies in areas where the government may not want to hear advice (the academy was founded to supply advice to the government, but demand for studies is down). The complex produced more than 200 reports last year, most, though not all, commissioned by federal departments and agencies.

On another front, the new president has also taken action in his first six months to address concerns about the impartiality and independence of the research council's studies. Because the reports are independently produced, government agencies like to have the academy's "blessing." But academy members have complained, Alberts confides, that research council staff have sometimes allowed government officials to influence reports, which runs counter to the rules of the academy.

A congressional source gives the example of a research council study of the Earth Observing System, a federal satellite remote-sensing program. Criticisms of the system in a research council study were muted after government officials who saw prepublication drafts objected. Studies for the Coast Guard and the Department of the Navy and studies on agricultural policy are also said to have been influenced through cozy relations with external parties.

Alberts has undertaken "an extensive review of the proposal review process" to strengthen high-level deliberations about the rationales for studies and to underscore the responsibilities of research council officials. Finally, he has inaugurated efforts to make the research council more collegial and efficient by breaking down "institutional barriers that prevent people from working with one another."

In his role as a statesman of science, Alberts expounds on a theme initiated by Press: that scientists must realize that their demands for more funds may have made them vulnerable on Capitol Hill to accusations of selfishness. To deal with the problem, Alberts is trying to build bridges with members of Congress and their staffs.

He has also consulted with the White House on its recently announced plan for a National Science and Technology Council, the administration's initiative to get a firmer grip on federally supported research. "We have to be more adept at disposing students to a wider range" of scientific careers, Alberts says. Otherwise, "we should not be training the number of Ph.D.'s in physics and chemistry that we are turning out."

A major challenge, he observes, is to remove the barriers that prevent people with scientific skills—including unemployed defense workers—from entering new fields, such as teaching, the topic that most of Alberts's thoughts come back to. "We have got to try, at least, to make it an attractive pathway," he declares. One approach he favors is the development of specific courses for teaching science pedagogy.

Alberts is sympathetic to the burdens of teaching. He has a daughter who teaches science, and like many of her colleagues she has to buy supplies for classes out of her own salary. "We've made teaching a profession in which it is impossible to do well unless you're some kind of martyr," he says. If he gets his way, that trend could start to change. Educators, as well as researchers, might yet be beneficiaries of the new irreverence in the sanctuary of science. *—Tim Beardsley*

SCIENTIFIC AMERICAN

The Future of American Defense

U.S. forces were shaped for conflict with a superpower. The emerging multilateral world calls for a smaller, more flexible and far less expensive military

by Philip Morrison, Kosta Tsipis and Jerome Wiesner

In October 1981 President Ronald Reagan announced the beginning of the biggest military buildup ever undertaken by a nation in peacetime. Over the next decade the U.S. spent more than $3 trillion (three quarters of the current national debt) on its military. Fully 60 percent of those costs were devoted to countering the threat of communism. That tremendous expenditure marked the culmination of the 50-year competition with the Soviet Union, a period during which escalation on one side provoked reciprocal moves on the other, even after both nations had long passed the point of both mutual intimidation and overkill.

When the Soviet Union imploded in 1991, the U.S. was still spending more than $300 billion a year for a military that included 530 ships, 16 active army divisions, more than 3,000 planes and more than 25,000 nuclear warheads. Such massive forces place an unacceptable burden on the American economy and saddle the nation with a military built around an unrealistic scenario of vast global conflict. American forces therefore require prompt reduction and reform.

Prudence implies that such changes in U.S. forces cannot be too sudden, nor should they go beyond the possibility of reversal. Nevertheless, judicious American military cutbacks could save half a trillion dollars by the end of the 1990s. That money could be far more productively targeted toward rebuilding infrastructure, expanding health care, upgrading education or otherwise improving the nation's economic and social well-being.

Such reductions should not be severe enough to threaten the ability of the U.S. to maintain a strong nuclear deterrent or to field conventional forces large enough to prevail against any foe the nation might plausibly face in the near future. At the same time, the nation must continue along new paths to encourage the just resolution of international conflicts through multilateral security arrangements.

In 1990 President George Bush made tentative moves in the direction of streamlining the excessive U.S. military capability. The Pentagon presented a five-year plan to create a downsized "base force" 25 percent smaller than the one that existed at the end of the Reagan years. Despite much talk about additional military cuts under the administration of President Bill Clinton, similar thinking appears so far to guide U.S. policy.

That approach is unsatisfactory for two reasons. First, the proposed forces are in many ways as oversized as they would be if they were still structured around containment of the Soviet Union, a powerful adversary that no longer exists. Second, the plans do not yet fully address the need for restructuring our military to reflect the changed world. The 1990 "two-war" strategy remains: the U.S. needs to retain the capability to fight two major regional wars at the same time without significant aid from allies. That strategy seems more a justification for a large American military than a plausible scenario for future conflicts. We expect to see the two-war guideline fade quietly away under the Clinton administration.

To arrive at an adequate yet affordable scale for America's armed forces, we attempted to answer two basic questions: Who are our most probable ad-

PHILIP MORRISON, KOSTA TSIPIS and JEROME WIESNER have been long-standing advocates of arms control. Morrison is an emeritus professor of physics at the Massachusetts Institute of Technology. He has spoken and written on military strategy and nuclear disarmament ever since he completed four wartime years working on the American atomic bomb project. Tsipis is the director of the Program in Science and Technology for International Security at M.I.T. His background lies in experimental particle physics; he has also written extensively on the physics and technology of nuclear weapons and nuclear war. Wiesner is president emeritus of M.I.T. He has served as science adviser to presidents John F. Kennedy and Lyndon B. Johnson. Wiesner helped to establish the Arms Control and Disarmament Agency and played a pivotal role in achieving a partial nuclear test-ban treaty and in limiting the deployment of antiballistic missile systems.

GULF WAR brought together a remarkable international coalition dedicated to neutralizing Iraq's military occupation of Kuwait. Such multilateral collaboration is likely to become more common in the post–cold war world. The U.S. can now drastically cut its military spending yet remain strong enough to face down any likely aggressor.

Declared Nuclear Nations

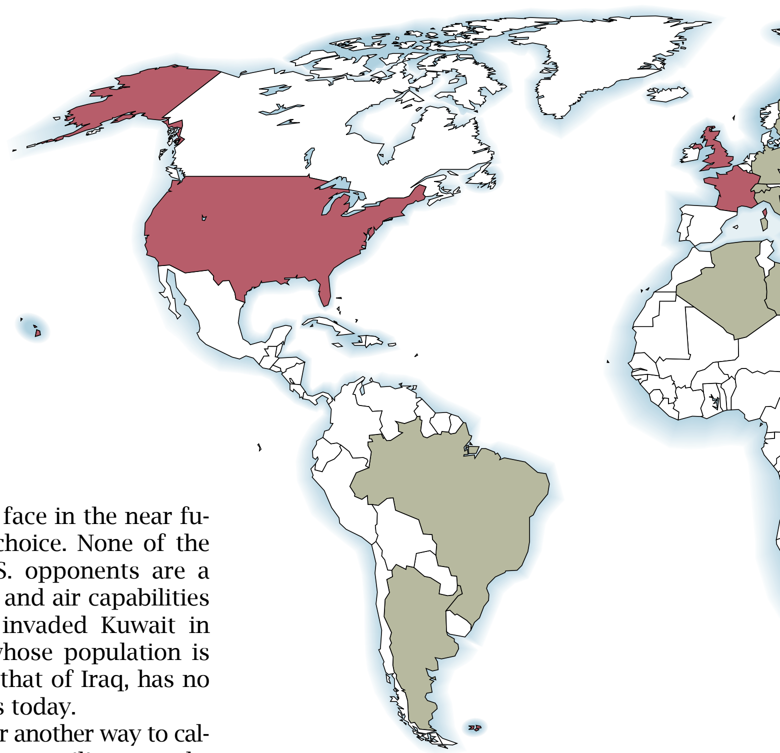
U.S.: 9,860 warheads, including 2,370 on intercontinental ballistic missiles (ICBMs). START II will reduce the total number to 3,500 by the year 2002.

Former Soviet Union: 10,920 warheads, including 6,630 on ICBMs, located in Russia, Ukraine, Belarus and Kazakhstan. START II will limit the total to 3,000 by the year 2002.

U.K.: Approximately 300 warheads, none of them on ICBMs.

France: 426 warheads, 18 of them on ICBMs.

China: About 270 warheads, roughly 100 of them on ICBMs.



versaries in the next two decades and beyond, and what missions and functions do we expect our military to perform in that time?

During the height of the cold war, potential enemies—primarily the Soviet Union, Eastern Europe and China—included some highly industrial countries that wielded nuclear weapons. Those nations contained about 1.5 billion people, about one third of the world's population; they could collectively field armies seven million strong. In contrast, the nations that might conceivably confront the U.S. in the foreseeable future (Iran, Iraq, North Korea and Libya) all lack strong industrial bases. Together these hostile countries have a population of just 110 million, from which they draw armies totaling no more than two million men and women. Even after a substantial reduction in American forces, the U.S. and its allies would outspend the rest of the world in defense.

Large-scale aggression against Central Europe is a fading vision. If the huge, powerful and militarily honed Soviet Union did not attempt it, its impoverished, fragmented successors, many of which are seeking aid from the West, surely will not. Regional conflicts in Eastern Europe or the former Soviet states might make necessary international peacekeeping missions, but they would not occasion a frontal assault by American forces. In the post–cold war world, it seems equally unlikely that American armies will be sent to fight in Asia against China or India. The U.S. certainly will not be warring against other countries in North or South America.

The most demanding military tasks the U.S. might plausibly face in the next decade or two would be countering aggression in the Middle East or on the Korean Peninsula. Congressman Les Aspin of Wisconsin, later the U.S. secretary of defense, selected Iraq as an example of the mightiest military force the U.S. is likely to face in the near future. It is a good choice. None of the other plausible U.S. opponents are a match for the land and air capabilities Iraq had when it invaded Kuwait in 1990. Even Iran, whose population is nearly three times that of Iraq, has no such military assets today.

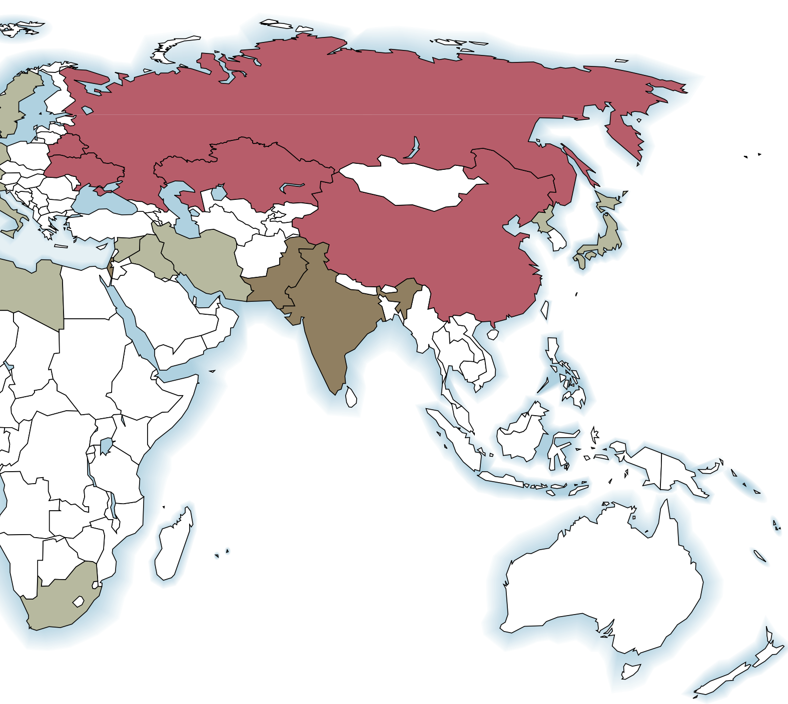
Past conflicts offer another way to calibrate America's future military needs. At present, the U.S. uses about 50 aircraft to monitor Iraq, and maybe twice as many planes (not all of them from the U.S.) monitor the airspace over Bosnia. During Operation Desert Storm, nearly 1,700 aircraft of all types flew 100,000 sorties in 45 days. The U.S. used a total of about 1,000 planes during both the Vietnam and Korean wars. Humanitarian assistance and peacekeeping efforts in Bangladesh, Lebanon and Somalia have engaged mainly Marine Expeditionary Units containing up to a total of 20,000 men and women.

During the next 10 or 20 years, the U.S. is far more likely to find itself engaged in multinational humanitarian, peacekeeping and counterterrorism activities than in a major armed conflict without allies. The U.S. should prepare to contribute to future common security forces assembled by the United Nations or other groupings of nations for operations such as the Gulf War or Somalia. For that responsibility, our country will need the means to keep sealanes open and to impose blockades as part of collective sanctions against nations that have taken up arms. We note that any force powerful enough to prevail in a large-scale battle will be more than sufficient for international policing missions.

In developing our proposal, we have tried to look beyond the present division of U.S. forces along service lines. A complex military establishment seems more amenable to timely and logical restructuring when it is defined by mission rather than along service lines. We have grouped our recommendations around the six distinct functions of the American military: nuclear deterrence, air-land battle, control of the seas, land-sea operations, intelligence gathering, and research and development.

America's long experience with joint operations offers a blueprint for such a revised military structure. For example, in the 1970s the U.S. developed an air-land battle doctrine for the defense of Europe. During the Gulf War, the U.S. effectively combined its land, air, naval and space-borne weapons and systems. The success of that venture suggests that it may be to America's advantage to reorganize its forces under six commands corresponding to the functions we have described.

One of the most dramatic and welcome results of the end of the cold war is the virtual disappearance of the possibility of nuclear conflict with the former Soviet states. As long as nuclear weapons continue to exist, however, they pose a risk to U.S. and world security. We set as a clear requirement that the U.S. preserve unquestioned deterrence against any nuclear threat. Nuclear war between two nations that possess substantial nuclear arsenals will almost surely result in mirror devastation of both combatants. A few hundred or a few dozen assuredly deliverable nuclear weapons will ensure symmetrical obliteration of bases, cities and industry even in large countries such as Russia or the U.S.



Probable Nuclear Nations

- **Israel** is believed to have 50 to 100 nearly ready warheads.
- **India** has 20 to 50 unassembled warheads.
- **Pakistan** may have up to 10 unassembled warheads, possibly of foreign design.

ALSO PLAUSIBLE

- **South Africa** says it will dismantle its six Hiroshima-type bombs.
- **North Korea** is reported to be in the final stages of warhead fabrication.
- **Iraq** was well along in development of warheads before the Gulf War.
- **Libya** and **Syria** and perhaps **Iran** and **Algeria** have long been interested in acquiring nuclear weapons.
- **Germany, Italy, Japan** and **Switzerland** could produce nuclear weapons within a year of deciding to do so.
- **Argentina, Brazil** and **Sweden** have abandoned their nuclear weapons programs.

Such modest nuclear holdings therefore can suffice to dissuade an aggressor from a nuclear attack. But superpower nuclear forces long ago grew beyond that size.

In 1990 the U.S. held more than 12,000 strategic nuclear warheads. The Soviet Union had 11,000. Under the Strategic Arms Reduction Treaty (START), signed by the two nations in 1991, the U.S. reduced its arsenal to 8,500 and the U.S.S.R. to 6,000. In June 1992 Bush and President Boris Yeltsin agreed to additional nuclear arms cuts: by the year 2003 the U.S. would have 3,500 warheads and Russia 3,000. At that time, China, France and Britain will, we expect, hold a few hundred nuclear weapons apiece. The newer undeclared nuclear powers may collectively possess an additional 200 or so warheads.

We recommend instead that by the year 2000 the U.S. possess a deterrent force of approximately 800 assuredly deliverable warheads. Present U.S. nuclear forces wield such immense destructive power that even after drastic cutbacks, the country will be able effectively to face down threats from any present or imminent nuclear nation. American holdings of nuclear warheads should be cut further as the nation strives to bring about a nuclear-free world. We would advocate even faster reductions in nuclear arms but for our awareness that more profound change will come only slowly as long as caution guides national leaders.

Tactical nuclear weapons—usually small-yield weapons meant for use against engaged forces on the battlefield or against airfields, bases and forward transport—seem to be of little military use. In September 1991 Bush wisely ordered the elimination of all sea-based tactical nukes but spared a few hundred bombs located on some carrier aircraft. This lingering ability to initiate nuclear attack on nonnuclear states is intolerable and dangerous; the complete end of all shipboard nuclear weapons by the year 2000 is a wiser stance. We also recommend eliminating the nearly 1,000 U.S. tactical nuclear weapons deployed on aircraft in western Europe, first by negotiating with Britain and France to end their deployment of such weapons. That initiative could set a splendid tone for the 1995 Review Conference on the Nuclear Non-Proliferation Treaty.

Nuclear weapons development, testing and manufacture are unnecessary and in fact are ending, both in the U.S. and in Russia. Clinton has initiated diplomatic efforts aimed at a formal ban of nuclear testing. The $8 billion our nation spends annually at some 16 specialized plants dedicated to the production of U.S. nuclear weapons can now be devoted entirely to the disassembly, safe storage and eventual disposal of nuclear cores and for the cleanup of radioactive pollution left in a dozen states.

The biggest nuclear threat to the security of the U.S. and to the rest of the world now comes from the proliferation of nuclear weapons and related technology. Attempts to combat proliferation only by stanching the flow of materials, equipment and know-how cannot prevent the appearance of new nuclear powers. Even the most thorough systems for safeguarding nuclear knowledge eventually leak.

To reduce the risk of proliferation, the nations of the world must lower the demand for, as well as the supply of, nuclear weapons. Wider international sharing of economic progress and political decision making could help ease tensions between nations and lessen the demand for nuclear prestige or protection.

Crucial though the nuclear cutbacks will be, most of the effort at military restructuring will involve conventional forces, which account for 80 percent of the U.S. defense budget. We turn our attention first to the portion of the military dedicated to air-land battle: ground forces and the aircraft that both precede and support those forces in battle. The strength of an army is often measured in terms of divisions of troops. A division is the standard unit large enough to include all the elements of ground war: infantry, armored vehicles, artillery, antiaircraft and engineers. During wartime, a U.S. Army division contains about 17,000 men and women.

The present base force plan anticipates only a modest reduction in American ground forces, from the 16 army

HOSTILE NATIONS still abound, but they present more localized challenges than did the Soviet Union or China. Iraq (*left*) has repeatedly defied the United Nations and is attempting to rebuild its forces. North Korea continues to pursue nuclear weapons. In Libya (*center*), a repressive regime sponsors terrorism. And Iran's brand of Islamic fundamentalism (*right*) may threaten Middle Eastern nations, including U.S. allies.

divisions that existed in 1991 to 12 divisions by 1995. We see no productive use for such extensive standing ground forces. Only China, India, Russia, Vietnam and North Korea have bigger armies, and they lack America's modern equipment and first-rate air and sea support. Moreover, the U.S. Department of Defense (and some defense analysts) continue to calculate American needs on the assumption that we will be fighting alone. But America has fought with allies in the past and almost certainly will do so in the future.

The massive, modern forces of western Europe and Japan are far more likely to fight beside those of the U.S. than against them. Nevertheless, there is the remote possibility that one nation with modern forces might change from ally to adversary. U.S. conventional forces that equal but do not exceed the overall strength of any two of the nation's strong allies seem a reasonable standard in a world moving toward common security.

That gauge leads us to suggest that by the year 2000 the U.S. should keep the equivalent of about five active army division equivalents, including heavy armor, airborne, helicopter-borne and light infantry units. This flexible, mobile force would retain a core of almost two full-armored divisions. The army would comprise about 180,000 full-time members, augmented by significant army reserves.

How many modern main battle tanks ought the U.S. to possess as the decade ends? German preparations for their defense against the Red Army (and now its still unformed successors) in Central Europe offer a guide to appropriate American tank strength. Germany deploys almost 4,000 top-of-the-line tanks, which are as good or better than the finest American tanks. France and the U.K., two of Germany's NATO partners, have another 2,500 good modern tanks between them. The U.S. Army has placed 6,500 of its most advanced tanks in Europe.

Because of geographic, economic and historical factors, however, U.S. interests extend beyond Europe. We therefore need to consider the tanks and other armored forces elsewhere around the world. In the Middle East, Iraq had the most extensive array of up-to-date tanks before the Gulf War dramatically reduced its intact weaponry. Today the strongest Middle Eastern tank forces are those of Israel (more than 3,000 main battle tanks) and of Syria (close to 4,000 tanks, many of them of older design). Iran has fewer than 500 tanks, just one tenth of the pre-war Iraqi force.

China has a force of 8,000 tanks, and India holds about 3,000 older tanks. Both countries have substantial armies and vast territories, but it is highly unlikely that either will engage American ground forces in the next decade or two. Aside from a few American allies (Germany, France, the U.K. and possibly Israel and South Korea), no country now possesses a tank force that could match 500 of the newer U.S. tanks.

In Europe, Germany is well prepared for land defense. We propose that, in addition, the U.S. should deploy about 1,200 of its best heavy tanks in Europe. Half of those tanks could be kept in storage, the other half on active duty. Another 800 U.S. tanks should be maintained for rapid deployment in any battles that might erupt in other theaters around the world. America's powerful combat helicopters will remain an important antiarmor weapons system that enhances the efficacy of the nation's tanks.

A reduction in tactical air strength should accompany the deep cut in ground forces. The Bush base force plan provided 56 combat squadrons of fighter and ground-attack, fixed-wing aircraft in 1993, which adds up to nearly 1,200 primary airplanes; spares, training craft and the like nearly double that number. We recommend a diminution of tactical air strength, roughly in proportion to the cuts in ground forces, to about one third the 1993 number. The reconfigured aerial forces will consist of 18 squadrons armed with the newest aircraft types: the F-117 Stealth fighters, F-15s and F-16s—America's top-of-the-line fighters—and the sturdy A-10 ground-attack airplane. Marine and navy aviation units will augment this total.

Control of the seas has long been a central function of American armed forces; this historical emphasis is sure to continue. Such capability will support both American and multilateral interventions in overseas crises of any kind. Among the armed services, the U.S. Navy now receives the largest share of the military budget. Aircraft carriers account for a hefty portion of that expenditure. The Department of Defense still intends to maintain an amazing 12 carrier battle groups in 1995, even though Soviet naval forces are no longer a threat.

U.S. Navy attack carriers are the biggest, most costly, complex warships ever invented. A ship of the nuclear-powered Nimitz class stretches 1,000 feet long. It carries its crew of 5,000 men and women anywhere in the world at a cruising speed of 35 miles per hour. One Nimitz-class carrier can transport to any coast an aerial strike force comparable to the entire air power of countries such as Canada, Denmark or Iran.

Navy carriers make it possible to project American air power a few hun-

dred miles inland, a considerable advantage when there are no friendly airbases onshore. The carriers also retain an old function: to “show the flag” spectacularly, giving warning of impending U.S. hostile action and taking control of straits and ports from the sea.

In the world we hope to enter, there is no great reason for the U.S. to act so determinedly alone. If Americans can accept a more cooperative view of security, this country will not need so much power against distant states. Five active carriers seem plenty; they would remain an unmatched force for the projection of American power. Two carriers might normally be devoted to nonviolent missions, for instance, to rescue, relief supply and evacuation duties, carried out under international sponsorship or even unilaterally. The navy might benefit from building several smaller carriers as cheaper, more versatile replacements for the current behemoths.

While at sea, each carrier currently receives support from six warships. As the number of carriers dwindles, so would this support. Twelve antiair and antiship missile cruisers, along with 18 of the most up-to-date antisubmarine ships, would furnish ample surface escorts for the remaining carriers. Two nuclear-powered attack submarines for each battle group would add undersea protection.

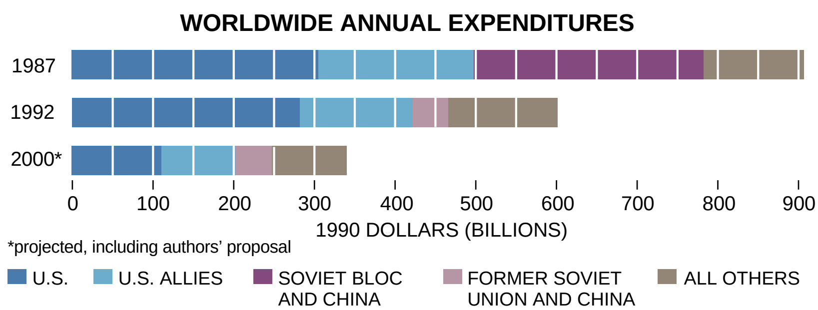
The 1992 report of the secretary of defense calls for 145 combat ships in 1995. We recommend instead a stiff drawdown in the navy comparable in magnitude to that we found reasonable for the army. Once again, we base our recommendation on the principle that the U.S. military resources contributing to the collective security of our allies should constitute a force second to none but no greater than any other two combined. That principle allows for a generous U.S. fleet containing more than 30 destroyers and frigates, over and above the 18 assigned to the carrier battle groups. The total U.S. surface naval strength (12 cruisers, 50 destroyers and frigates and five big carriers) would continue to be the biggest naval force at sea.

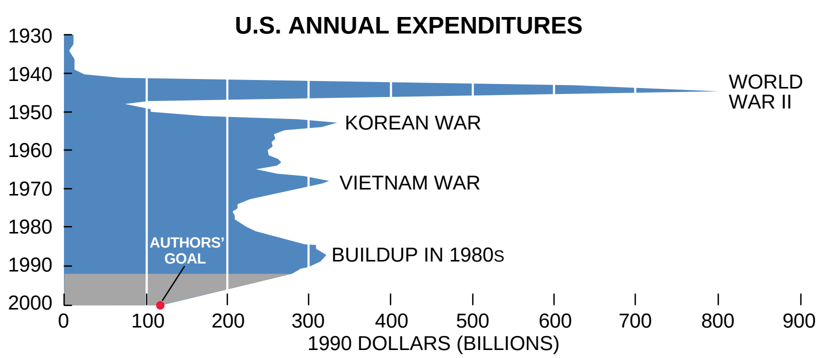
Naval air strength might be kept at a level somewhat higher than the number of aircraft that can be delivered by the reduced fleet of carriers. Such airplanes can operate from runways on land in addition to those at sea. The present 620 active carrier-borne naval fighters, bombers and attack aircraft can be pared to about 210 planes. Land-based, long-range maritime patrol aircraft can perform a wide variety of surveillance tasks on the high seas. In a world full of surprises, we recommend retaining nearly the full present number of these craft. About 250 Orion P-3 turboprop patrol airplanes can survey the most important ocean areas in conjunction with the aircraft capable of performing the same missions that belong to American maritime allies.

The U.S. Marines, by origin and by tradition, operate at the interface between sea and land. They are the most versatile component of the American armed forces. One Marine Expeditionary Unit numbers about 2,500 men and women, along with 10 tanks and artillery batteries, half a dozen vertical-takeoff attack aircraft and 30 helicopters and their crews. The marines now have a dozen or more helicopter carriers that are able to land an expeditionary unit and its armor on any beach.

Declines in Military Threats to the U.S.

Plummeting world military expenditures offer an opportunity for commensurate cuts in U.S. defense spending (*top*). Some changes seen here reflect shifts in the value of currencies and a reclassification of former Warsaw Pact nations into the “rest of the world” category. Our projection assumes that by the year 2000 the U.S. and its allies will account for 60 percent of world military spending, a higher percentage than in 1987. American military cutbacks will be steep but no more so than during postwar transitions earlier in this century (*bottom*).





SOURCE: International Institute for Strategic Studies

These units allow the flexible, speedy, small-scale response that is likely to be necessary for future missions, whether sent across a beach for an interventionary attack or for less belligerent missions, such as the U.S. entry into Somalia.

The unique capabilities of the marines seem crucial to preserve. Even if forcible entry onto foreign soil becomes as rare as it should, many humanitarian and peacekeeping missions will continue to demand quick action, often through inadequate seaports. But the existing marine forces could be substantially trimmed. We propose that the U.S. sustain a dozen distinct expeditionary units, which would be grouped as needed. This change would bring about an overdue reduction in marine forces from almost 200,000 persons to one augmented division and its air support, or about 50,000 persons.

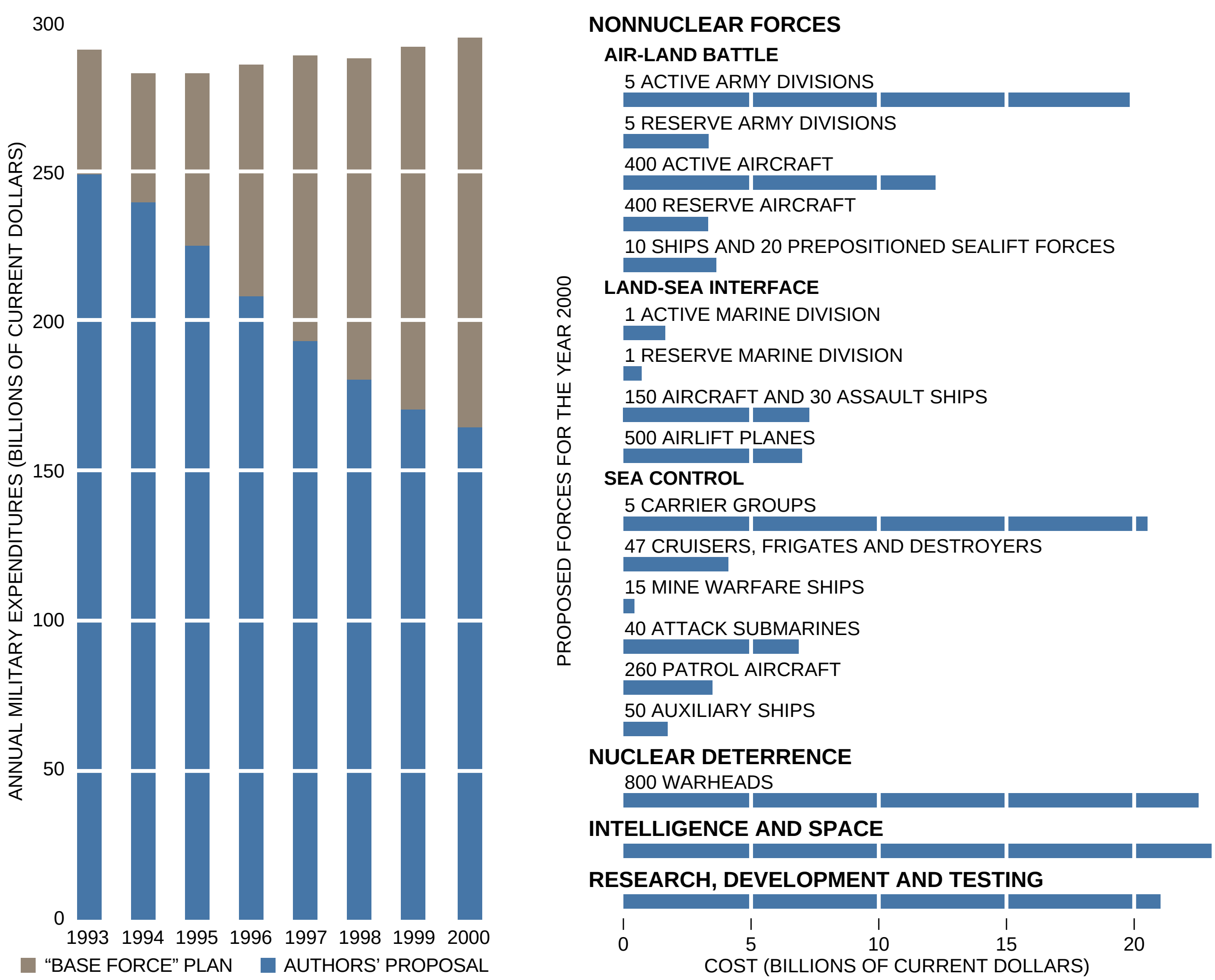
The marines would retain about 25 of the present 65 amphibious warfare ships, mainly the newer ones. The trimmed marine forces would receive air support comparable to that of one of the present Marine Air Wings: 150 fixed-wing combat aircraft (including 40 of the unique vertical-takeoff Harriers) and a similar number of gunships and transport helicopters.

Two very important components of the U.S. military—intelligence and military R&D—lie outside the boundaries of either conventional or nuclear forces. We envision these two sectors eventually falling under their own, separate commands. Only some of their functions can be reduced in proportion with the deep cuts outlined so far. The U.S. mostly keeps secret the costs of its information gathering, including that conducted from spy satellites. Defense analysts have made informal estimates, however, which are becoming increasingly credible. We use those estimates in our proposal.

The largest single information-gathering item in the U.S. defense budget ($11 billion in 1990 dollars) pays for decentralized tactical intelligence: reconnaissance aircraft, radio monitoring

A Strategy for U.S. Arms Reductions

The military reorganization outlined in this article would maintain U.S. security while leading to enormous cost reductions. A breakdown of our proposal (*below, right*) shows the size and cost of the restructured forces for the year 2000. The "base force" plan developed under President George Bush still reflects the Pentagon's spending goals through 1995. Our approach would save $670 billion in current dollars by the end of the decade (*below, left*).



and a dozen or more other activities that deliver short-term battlefield information to combat forces wherever they are. Reductions in the size of America's armed forces decrease the need for such intelligence. These activities can be pared to $3.5 billion, a somewhat smaller cut than those we suggest in the combat forces themselves.

America's diverse military satellites and its signal-decrypting organization, the National Security Agency, provide a valuable stream of hard information on global affairs. We would retain this useful flow nearly in full at $9 billion per year, a slight drop from the current $11 billion. The Central Intelligence Agency—estimated to cost $3 billion a year, about one tenth of the intelligence total—can sustain a 50 percent funding cut, in part to curtail covert operations of dubious value all over the globe. The total costs of intelligence can decline from the surmised 1992 sum of $29 billion to $18 billion in current dollars by the year 2000. The U.S. should seriously consider sharing substantial chunks of the costs and results of its intelligence activities with American allies and with the U.N.

Research and development of new nuclear and conventional weapons has provided the American military with an unsurpassed technological edge on the battlefield and a redoubtable deterrence to nuclear aggression from afar. In fiscal 1993 weapons research, at slightly more than $42 billion, took 59 percent of all federal R&D funds for the year. Of that total, nearly $38 billion was spent on the development, testing and evaluation of new weapons.

Forty years of such investments have paid off. In many cases, U.S. arms are now technologically superior to those that an opponent could deploy even a decade from now. Moreover, defense R&D spending in the former Soviet Union has plummeted since its dissolution in 1991. The race to stay ahead can now slow down.

Even at $15 billion a year, the U.S. would spend nearly 10 times as much on military R&D as Germany and Japan combined. Key savings will result from canceling further work on inessential projects such as tilt-rotor aircraft and new attack submarines and from ending other big R&D programs aimed at the nonexistent Soviet threat. Such savings will allow full funding for the communications, electronic countermeasures and surveillance-and-attack systems needed to safeguard the qualitative superiority of many American weapons systems. Basic and applied, dual-use military research can stay at the present $4 billion a year while leaving ample funds for preserving the defense technology base and for pressing forward with advanced technology programs.

SOMALIA MISSION heralds the international humanitarian actions that may become an increasingly important function of the U.S. military in the future. Despite some missteps, the mission brought food and supplies to a devastated population.

Recent American military engagements offer a glimpse of a future defined not by massive preparations for a superpower war but by a versatile mix of forces suited to a world in transition. Television has brought the American public vivid images of air deliveries to besieged towns in Bosnia, helicopters rocketing strongpoints in Mogadishu and Iraqi radar sites knocked out by missiles from American ships speeding at sea faraway. What those powerful but piecemeal images cannot show is how preposterously disproportionate U.S. military forces have become in comparison with the challenges they face.

The military restructuring that we propose attempts to respond realistically to the historical changes in the geopolitical map that have taken place in the past few years. It is a cautious plan that anticipates a welcome embrace of the notion of peace through cooperation while remaining alert to new risks of war. The reduced forces we advocate should be adequate to undertake six to eight simultaneous Somalia-like operations or to mount a force somewhat larger than the American component of Desert Storm. By the year 2000 U.S. forces would remain more capable and more versatile than any other in the world, at a cost of $164 billion per year in current dollars, a 60 percent reduction from real expenditures in 1992.

Such cutbacks will place a definite, though localized, burden on the American economy; indeed, the apprehension of that stress has slowed the inevitable reduction in military spending. But the U.S. has weathered similar transitions in the past. Moreover, each billion dollars of military savings, if spent in the civilian economy, can create more than twice as many jobs as those lost in the weapons industry.

American leadership will be vital in bringing about a more peaceful, less militarized world. At the beginning of the 1990s, nations squandered nearly a trillion dollars a year in military programs worldwide. In the near future, we hope much of the resources so misdirected can be used to tackle environmental and human problems in the U.S. and around the globe.

FURTHER READING

CUTTING CONVENTIONAL FORCES: AN ANALYSIS OF THE OFFICIAL MANDATE, STATISTICS AND PROPOSALS IN THE NATO-WTO TALKS ON REDUCING CONVENTIONAL FORCES IN EUROPE. East-West Conventional Force Study. Institute for Defense and Disarmament Studies, July 1989.

DECISIONS FOR DEFENSE: PROSPECTS FOR A NEW ORDER. William W. Kaufmann and John Steinbruner. Brookings Institution, 1991.

DEFENSE CUTS AND COOPERATIVE SECURITY IN THE POST–COLD WAR WORLD. Randall Forsberg in *Boston Review,* Vol. 17, Nos. 3–4, pages 5–9; May/July 1992 (reprinted by the Institute for Defense and Disarmament Studies).

FACING REALITY: THE FUTURE OF THE U.S. NUCLEAR WEAPONS COMPLEX. Edited by Peter Gray. Tides Foundation, 1992.

THE MILITARY BALANCE 1992–1993. Col. Andrew Duncan et al. International Institute for Strategic Studies, 1992.

Sulfate Aerosol and Climatic Change

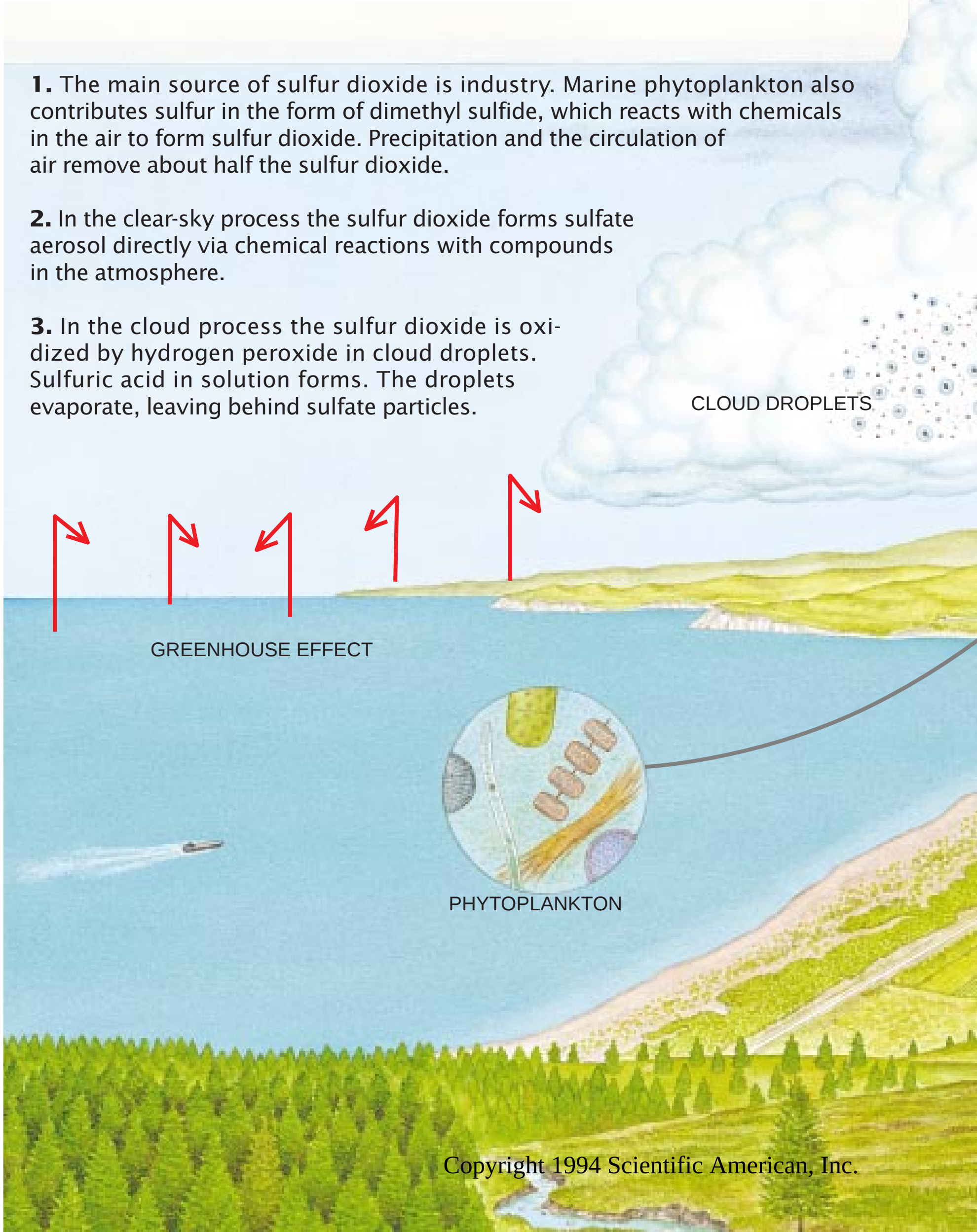
Industrial emissions of sulfur form particles that may be reflecting solar radiation back into space, thereby masking the greenhouse effect over some parts of the earth

by Robert J. Charlson and Tom M. L. Wigley

The greenhouse effect is a geophysical fact of life. Atmospheric gases such as carbon dioxide and methane trap and hold heat, enabling the earth's biota to survive. Such gases warm the surface of this planet by about 33 degrees Celsius, from below freezing to a current average of about 17 degrees C. Models and analyses of global warming generally agree that most of the long-lived gases that human economic activity adds to the atmosphere make the earth warmer than it would otherwise be. Yet discrepancies between theory and observation persist. The predicted warming based on recent increases in concentrations of greenhouse gases is slightly more than the observed warming of the atmosphere. In addition, the warming trend in North America does not appear to follow the global pattern. What might account for these and other deviations of fact from theory?

The answer is ironic. In all probability, aerosols primarily composed of sulfates, themselves the result of commercial activity, enhance the ability of the atmosphere to reflect sunlight back into space before it can reach the planet's surface and participate in the warming process. The sulfate particles, about 0.1 to one micron in diameter, are particularly concentrated over the industrial areas of the Northern Hemisphere. Their roles as contributors to acid rain, as irritants and as obscurers of such splendid vistas as the Grand Canyon have been known for years. But their capacity to cool by scattering sunlight has become a recognized force in climatic change only recently. Clearly, both the cooling effects of aerosols and the warming caused by greenhouse gases must be taken into account if we are to attain accurate climate models and effective industrial policies.

In theory, industrial aerosols are not the only particles that can contribute to cooling. Several kinds of aerosols exist

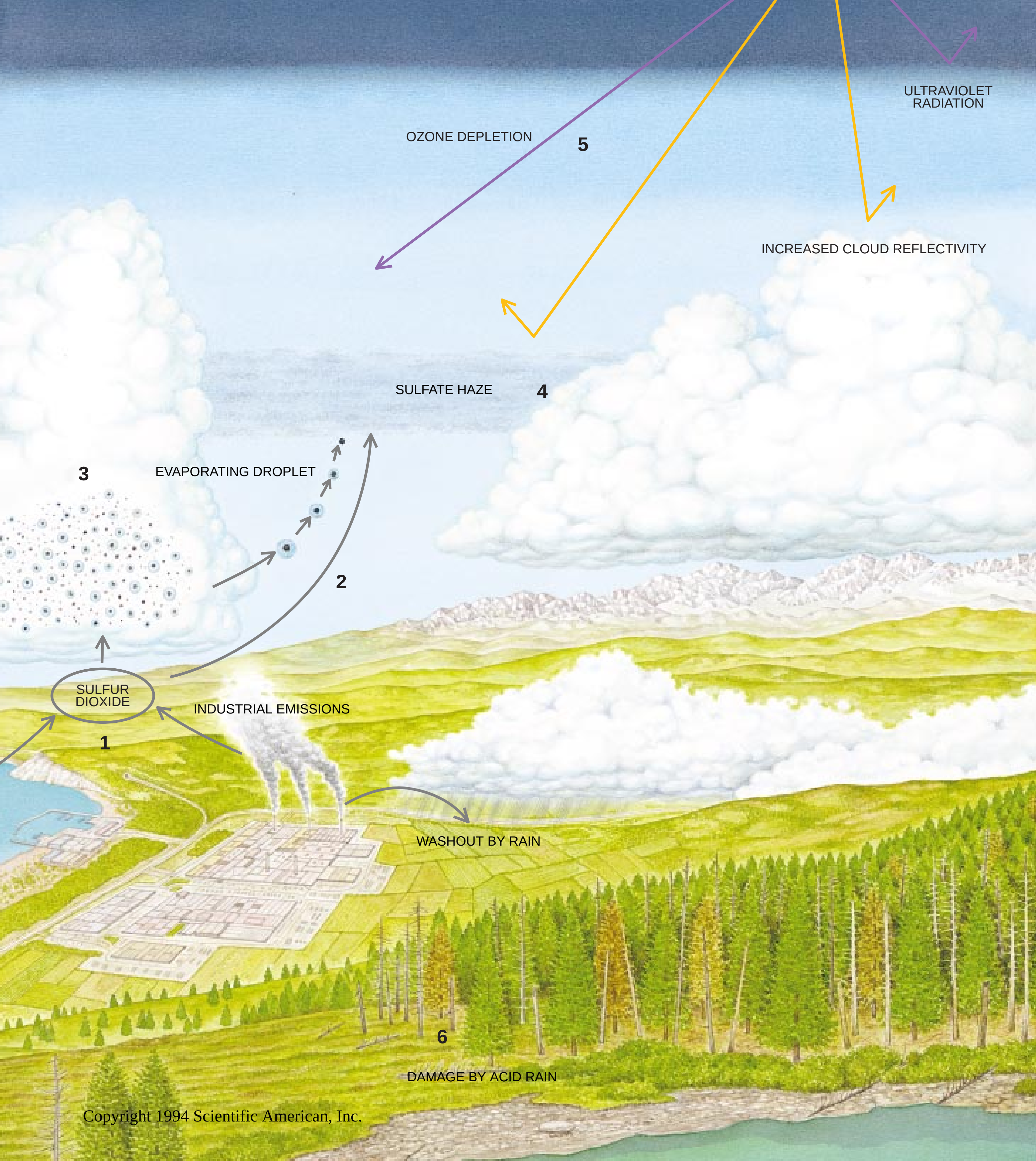


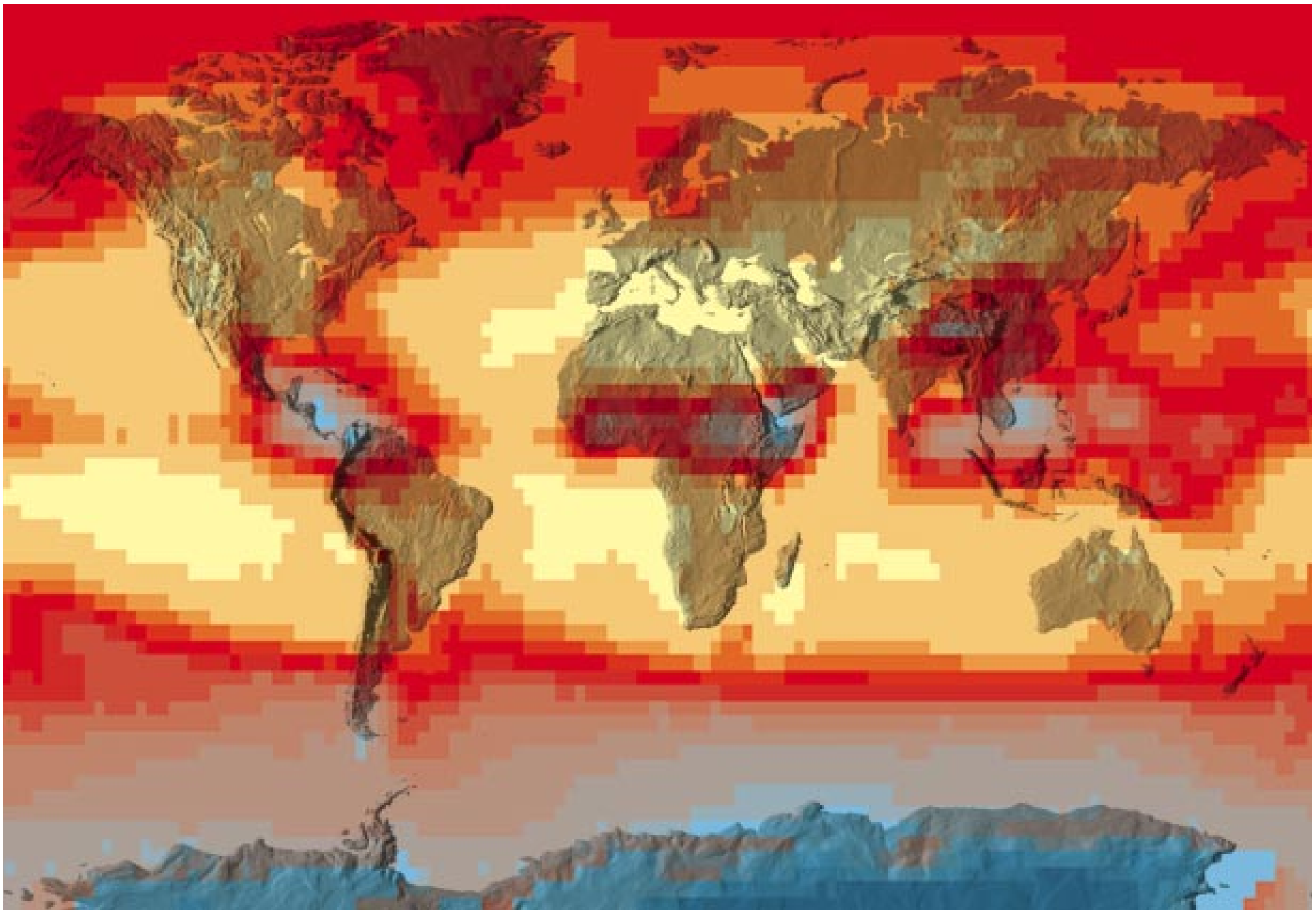
SULFUR FROM INDUSTRY and, to a lesser extent, phytoplankton exerts many environmental influences. It cools the earth by forming minute particles that scatter sunlight back into space, offsetting the greenhouse effect in part. Sulfate compounds also help to cause haziness, acid rain and ozone depletion.

4. Residing mostly in the lower troposphere, sulfate aerosol directly cools the earth by reflecting sunlight. The particles may also act as seeds for cloud condensation and increase the reflectivity, or albedo, of clouds.

5. When injected into the stratosphere by large volcanic eruptions, the sulfate compounds can help destroy the ozone layer.

6. The sulfate particles acidify rainwater, damaging lakes and plant life.





CLIMATIC FORCING by human activity is evident in calculations of global heat gain during the Northern summer. Every July greenhouse gases warm the earth by about 2.2 watts per square meter (*left*); the effect is most pronounced over the warm areas of the subtropics. When the cooling by sulfate aerosol is included, however, the forcing drops to about 1.7 watts per square meter (*right*). In fact, the cooling dominates over industrial regions in the Northern Hemisphere.

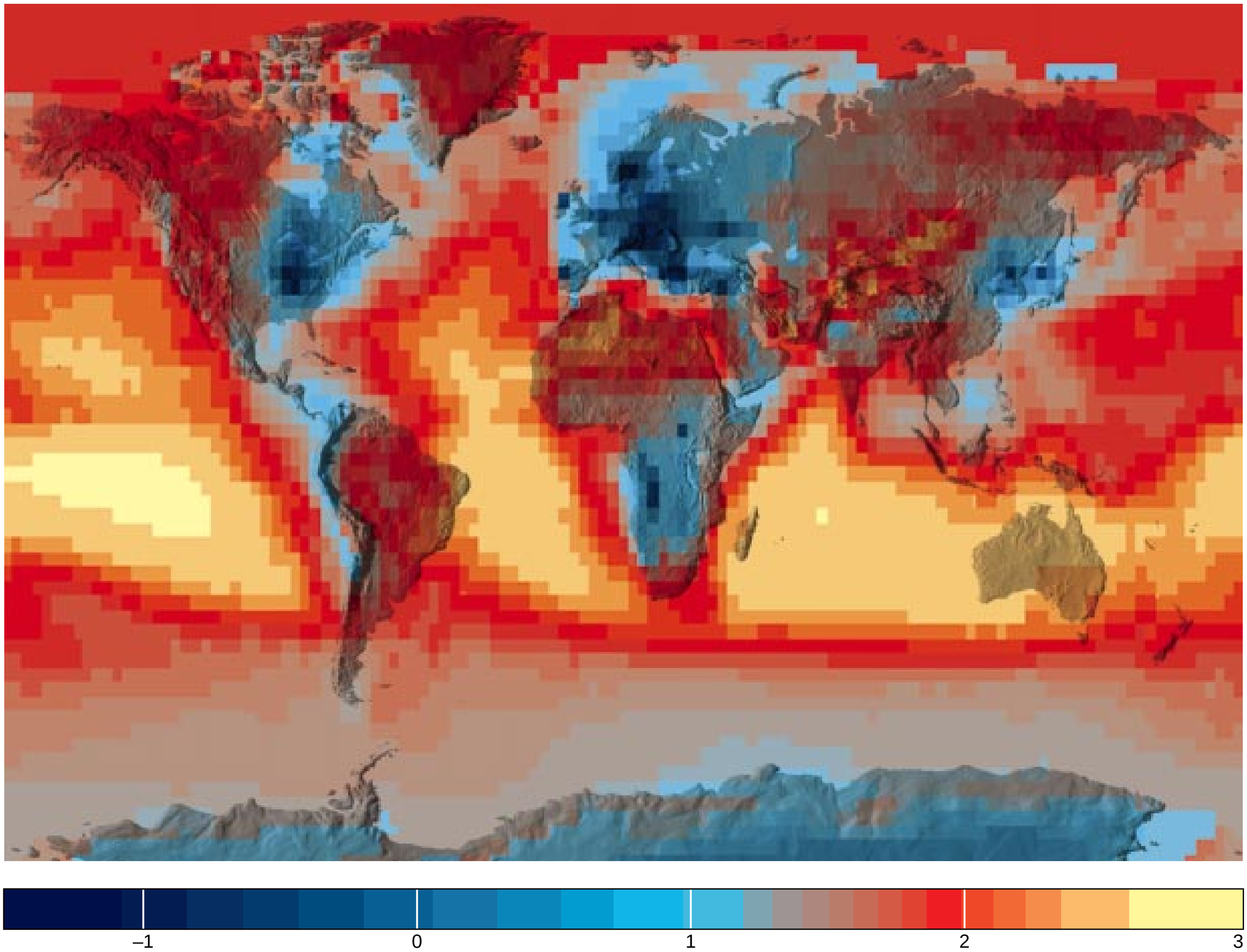
naturally. They do not, however, seem to be major factors causing change. Natural aerosols—mostly continental dust, sea salt and marine sulfate compounds—have probably remained roughly constant in their concentration, distribution and properties for at least a century. Thus, they would not have contributed to any observable alterations in climate. Volcanic aerosols have probably not added to long-term effects. The cooling trends precipitated by the gigantic eruptions of Tambora in 1815, Krakatoa in 1883 and Pinatubo in 1991 lasted only a few years.

In stark contrast, man-made aerosol compounds in the atmosphere have increased dramatically, primarily during the course of industrialization and most rapidly since about 1950. Of all the particulate pollutants humans create, climatologists have thus far focused much of their attention on sulfate compounds. That is because a large body of data, gathered in studies of acid rain, makes sulfates the best-understood aerosol. Other aerosol substances—soot from oil combustion, soil dust from desertification and smoke from slash-and-burn agriculture—may have an impact approaching the magnitude of that caused by industrial sulfur. Limited research findings, however, render the uncertainties in calculating those effects much greater.

As one might expect for complex systems such as the climate, determining the amount of cooling by sulfate aerosol is not a straightforward task. Many variables complicate the effort—among them, the amount of sulfur in the atmosphere, its distribution over the globe, the mechanism of aerosol formation, the degree of reflectivity of the particles and their effect on clouds. An accurate prediction also depends on making correct assumptions. Some early studies exploring the role of aerosols on climate did not do so. For example, one common and unsubstantiated supposition was that most of the haze outside cities was a "natural background" aerosol.

Another early, implicit assumption was that processes at the earth's surface make most aerosol particles. But that conclusion is valid for only two kinds of aerosols: those introduced into

ROBERT J. CHARLSON and TOM M. L. WIGLEY have combined their respective expertise in atmospheric chemistry and the interpretation of temperature records to study how sulfate aerosol affects the earth. Charlson is professor of atmospheric sciences at the University of Washington, where he received his Ph.D. He earned B.S. and M.S. degrees in chemistry at Stanford University. He holds six patents on instruments for atmospheric measurements and has served on numerous committees of the National Academy of Sciences. Wigley, the former director of the Climatic Research Unit at the University of East Anglia in Norwich, England, now heads the Office for Interdisciplinary Earth Studies at the University Corporation for Atmospheric Research (UCAR) in Boulder, Colo. He received his Ph.D. from the University of Adelaide in Australia.



the atmosphere by wind (such as sea salt and soil dust) and those arising directly from combustion (for example, industrial smoke or smoke particles from forest and grass fires). Studies during the past decade indicate that most sulfate aerosol originates from chemical reactions of sulfur gases discharged into the air. These reactions take place in the troposphere, that part of the atmosphere extending from the surface of the earth to an altitude of about 10 kilometers.

To calculate the increases of sulfur in the troposphere, climatologists rely on industrial emission rates. These rates act as excellent guides for estimating changes in the average concentration of atmospheric sulfate aerosol over time. Sulfur gases and the sulfate they make last only a few days in the troposphere, so that the average concentration in the atmosphere is directly proportional to the product of the emission rate and the lifetime of the substances. Consequently, the primary effects must mirror the geographic distribution of the sources of sulfur.

More than two thirds of the troposphere's supply of sulfur gases, mostly emitted in the form of sulfur dioxide (SO_2), is man-made. About 90 percent of that amount arises in the Northern Hemisphere. There human activity injects about five times the amount of sulfur gases emitted naturally. In the Southern Hemisphere, man-made emissions currently equal only about one third of natural emissions. The main natural carrier of reactive sulfur is dimethyl sulfide ($(CH_3)_2S$), or DMS, which originates from marine phytoplankton. In the absence of anthropogenic sources, DMS is thought to be the dominant source of submicron particles. A small amount of sulfur (as hydrogen sulfide or sulfur dioxide, or both) comes from volcanoes and from swamps and bogs.

The sulfur dioxide generally remains in the hemisphere in which it was produced. The thermal and chemical mixing of the two halves of the earth's atmosphere requires about a year—far longer than the average lifetime of sulfur dioxide or the sulfate aerosol it produces. Although the hemispheres are essentially decoupled regions as far as aerosol distributions are concerned, the aerosols in the Northern Hemisphere may nonetheless influence the climate worldwide, just as regional cloud cover controls the average albedo, or reflectivity, of the earth.

About half the amount of sulfur gases is lost directly from the atmosphere; it is either washed out by rain or reacts chemically with plants, soil or seawater. The remainder goes on to oxidize with compounds in the troposphere and hence to produce aerosol particles. Indeed, almost all types of sulfur-containing gases are chemically reactive in the presence of oxidizing agents. The most important such agent is the hydroxyl (OH) radical.

The reactions that create sulfate aerosol can loosely be divided into clear-sky and in-cloud processes. In clear-sky processes, sulfur dioxide and DMS in the presence of water vapor react via a complex series of steps to produce gaseous sulfuric acid (H_2SO_4). The compound forms particles a fraction of a micron in size. It does so by condensing on existing particles or by interacting with water vapor or other sulfuric acid molecules. This transformation is called gas-to-particle conversion. The sulfuric acid then reacts with minute quantities of ammonia to form varying hydrated forms of ammonium sulfate ($(NH_4)_2SO_4$) salts. In addition, DMS can react to form another condensable species, methane sulfonic acid (CH_3SO_3H), or MSA. Although MSA is an important atmospheric constituent and tracer compound (it has been measured in fossil form in ice cores), the latest research indicates that its aerosol has only a small impact.

Sulfate aerosol is also produced in clouds. This pathway begins when sulfur dioxide dissolves into existing cloud droplets. There it may be oxidized by the small concentrations of aqueous hydrogen peroxide (H_2O_2) that form when two hydroxyl molecules combine. The oxidation reaction then forms sulfuric acid and its ammonium salts in solution. In the droplet the acid sulfate exists as a strongly hydrated form, in which water molecules are bonded to the sulfate. Evaporation removes some of the moisture. Because the sulfates cling to water, the product of evaporation is a highly concentrated sulfate solution. The result is a submicron aerosol droplet that is chemically indistinguishable from the aerosol produced by gas-to-particle conversion.

The strong chemical affinity that sulfuric acid and its ammonium salts have for water is highly significant in terms of the aerosol's ability to scatter light. When the tiny solution droplets mix with humid air (such as over moist land or oceans), they absorb moisture and grow. Larger particles scatter more visible light, thus explaining the increase in haze when humidity is high. At a relative humidity of 80 percent (the global average value for air near the ground), a given amount of sulfate produces about twice as much apparent haziness as it does during a low-humidity day.

Once formed through chemical reactions, the sulfate particles in the troposphere can cool the climate in two ways: either directly, under clear skies, by reflecting away some incoming solar radiation, or indirectly, by increasing the reflectivity of clouds.

In the direct, or clear-sky, effect the sulfate aerosol particles scatter sunlight out of the atmosphere and into space; as a result, less solar radiation reaches the ground. There are two ways to estimate the fraction of incoming energy lost to space. One technique is to conduct detailed optical calculations based on particle sizes and refractive indices. An alternative and currently more reliable approximation is simply to make use of the observed correlation between the amount of aerosol in the atmosphere and the energy loss caused by scattering [*see box on page 57*]. These analyses indicate that at today's levels man-made sulfate scatters about 3 percent of the direct solar beam. About 15 to 20 percent of this amount goes back into space, for a total loss of about 0.5 percent. The average reduction of sunlight, however, is actually about half this amount, because clouds cover about half the earth at any given time. On the ground, the deprivation of sunlight is calculated to be roughly about 0.2 to 0.3 percent.

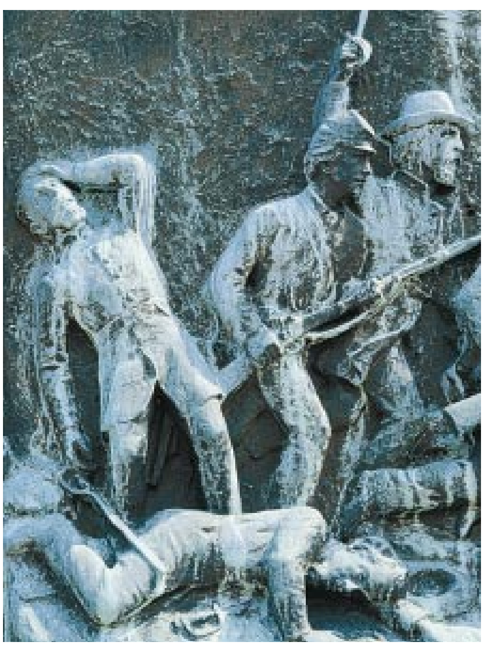
Is this loss at all significant? The solar radiation reaching the layer of sulfate haze near the ground amounts to roughly 200 watts per square meter, so the implied loss amounts to about 0.4 to 0.6 watt per square meter. Because the Northern Hemisphere contains more aerosols, the average forcing there must be greater, probably around one watt per square meter. (Climatologists use the term "forcing" to refer to the effect of factors external to the atmosphere and oceans on the changes in the planetary energy balance.)

Such a loss of incoming energy may seem small, but it is not inconsequential. The present-day increases of carbon dioxide resulting from human activity amount to a gain of 1.5 watts per square meter in the planet's heat balance. (When other greenhouse gases such as methane and nitrous oxide are considered, the increase is about two to 2.5 watts per square meter.) Hence, the cooling caused by sulfate aerosol is

Sulfur, Acid Rain and the Ozone Layer

Sulfur compounds in the air do more than just cool the earth. Investigators have known for many years that sulfur contributes to the acidity of rainwater and to the depletion of ozone in the stratosphere. Acid rain comes about in large part because of the oxidation of sulfur dioxide in the atmosphere. The oxidation process forms sulfuric acid, which creates aerosol particles. In the troposphere, these submicron-size particles attract water. Hence, they can act as nuclei for cloud condensation when the relative humidity exceeds 100 percent. The cloud droplets incorporate the acidic ingredient, which will be deposited on the earth's surface as rain or snow. Industrial acidity can spread over perhaps 1,000 kilometers (about 600 miles) from its source before the particles precipitate out.

Sulfate particles help to deplete ozone when they reside in the stratosphere, that part of the atmosphere above the troposphere. Deposited there primarily by enormous volcanic eruptions, the sulfur particles can provide surfaces on which ozone-destroying compounds act [see "Polar Stratospheric Clouds and Ozone Depletion," by Owen B. Toon and Richard P. Turco; SCIENTIFIC AMERICAN, June 1991]. Ironically, the effects of ozone depletion in the stratosphere are countered to some degree by sulfate aerosol in the lower atmosphere. The aerosol there reduces the amount of solar ultraviolet radiation that reaches the ground and offsets some of the increases expected from stratospheric ozone loss (but of course only in those areas where sulfate particles abound).



BRONZE MEMORIAL in Gettysburg National Military Park shows discoloring by acid rain.

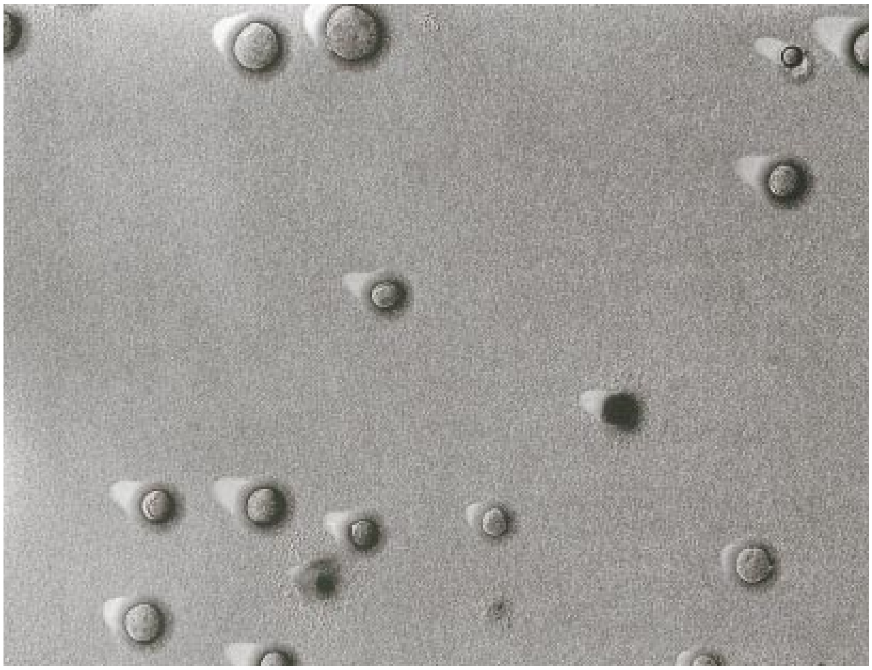
comparable in magnitude to the heating caused by carbon dioxide, at least in the patches of haze concentrated over industrial regions.

Needless to say, these calculations are crude. To quantify more precisely the aerosol effect and to describe its geographic distribution, researchers at Stockholm University and the University of Washington used a meteorologic model developed at the Max Planck Institute for Chemistry in Mainz. The description, which details the chemical production and wind-driven transport of particles generated from anthropogenic sulfur dioxide, enabled the workers to produce a map of the change in heat balance caused only by the direct effect of anthropogenic sulfate. This model showed three large masses of haze in the Northern Hemisphere. One mass, over the eastern U.S., creates losses of solar radiation of more than two watts per square meter. The two others, over Europe and the Middle East, reflect up to four watts per square meter. The average over the Northern Hemisphere, based on sulfur dioxide emissions recorded in 1980, is 1.1 watts per square meter, happily close to the crude calculation above.

The second, indirect way sulfate aerosol cools the earth is by influencing the albedo of clouds. When in clouds, some of the sulfate particles act as nuclei for condensation. The density of cloud-condensation nuclei determines the number and the size of cloud droplets. For a given amount of condensed water, the number density in turn affects the albedo of the cloud. A 30 percent rise in cloud albedo only over the world's oceans would be sufficient to counteract the average warming by anthropogenic carbon dioxide increases during this century.

Unfortunately, this indirect effect of the sulfate particles has thus far resisted reliable quantification. Although observations show that cloud-condensation nuclei are greatly enhanced over industrial regions, investigators do not know how differences in the number of nuclei relate to the changes in the amount or mass of anthropogenic aerosols. As a result, estimating the magnitude of the indirect aerosol forcing is not yet possible. Satellite observations suggest that the effect is not huge, although theoretical analyses permit it to be comparable to direct forcing.

In view of the fact that the ability to model completely the meteorologic effects of aerosols is limited, one may wonder whether aerosol cooling is real. In particular, one may ask whether the aerosol cooling is evident in the observational record. The most straightforward way to answer this question is to compare the changes in the Northern Hemisphere with those in the Southern Hemisphere. As a whole, the globe has warmed by about 0.5 degree C during the past 100 years [see "Global Warming Trends," by Philip D. Jones and Tom M. L. Wigley; SCIENTIFIC AMERICAN, August 1990]. If the enhanced greenhouse effect (that is, the additional warming caused by human activity) is the sole mechanism for climatic forcing, then the Northern Hemisphere should warm a bit more quickly than the Southern Hemisphere. The Southern Hemisphere holds most of the world's oceans and hence has more inertia with respect to thermal changes.



SULFATE AEROSOL sampled from the atmosphere was photographed through an electron microscope. The particles are about 0.1 micron in diameter.

Yet the observations show otherwise: since 1940 the Northern Hemisphere has warmed more slowly. In fact, the strong warming trend that occurred earlier this century in the Northern Hemisphere ceased around 1940 and was not renewed until the mid-1970s, even though industrial emissions of greenhouse gases continued to rise over the entire period. This reprieve in warming may have resulted from the counteracting properties of sulfate aerosol, at least to some extent. Although the changes broadly parallel the hypothesized aerosol cooling, they are not enough to prove a causal relation. (Indeed, the lack of a marked difference between the warming trends in the two hemispheres throughout the 20th century imposes an upper bound on the total magnitude of the aerosol forcing, which implies that the cloud albedo contribution has been small.)

Another piece of circumstantial evidence comes from an analysis conducted by the United Nations Intergovernmental Panel on Climate Change (IPCC). In 1990 the panel pointed out a discrepancy between the observed global mean temperature changes and the predictions made by climate models. The models suggested that the world should have warmed somewhat faster than the record indicates. Sulfate aerosol may help explain the discrepancy.

To see why, we need to introduce the concept of "climate sensitivity." In computer simulations of the climate, investigators double the atmospheric carbon dioxide concentration and then allow the climate system to adjust to the new (warmer) steady state. The change in global mean temperature is a measure of the sensitivity of the global average temperature to external forcing. The IPCC has given a "best guess" value of 2.5 degrees C for this quantity, although the sensitivity may in fact range from 1.5 to 4.5 degrees C. When observations are compared with the results from climate models designed to estimate specifically the time-dependent response to observed changes in greenhouse-gas forcing, the implied climate sensitivity is found to be a little less than 1.5 degrees C. In other words, the empirical estimate of the climate sensitivity gives a value more than a full degree below the IPCC's best guess and slightly below the expected range.

These numbers suggest that the current global warming induced by greenhouse gases may have exceeded the observed 0.5 degree C rise and been offset by some kind of cooling process. Natural variability of the climate could account for the cooling. Alternatively, external factors may be responsible. The aerosol effect is an obvious candidate. Indeed, factoring aerosol cooling into models yields a value for climate sensitivity that is a little above the IPCC's best guess but well within the expected range. Unfortunately, none of the conclusions is sufficiently convincing to allow us to jump out of the bathtub crying, "Eureka!"

Although subject to considerable quantitative uncertainty, the evidence clearly indicates that aerosols have a significant influence on the climate, comparable to that produced by greenhouse gases. In fact, from 1880 to 1970, aerosol cooling may more or less have canceled out the enhanced greenhouse effect in the Northern Hemisphere. (Since 1970 emissions of greenhouse gases have increased more rapidly than have those of aerosol particles.) The cooling caused by aerosols may even dominate in some areas. Recent work by Jeffrey T. Kiehl and Bruce P. Briegleb of the National Center for Atmospheric Research in Boulder, Colo., suggests that aerosols produce a net cooling in local regions of the eastern U.S., south central Europe and eastern China.

A crucial complication, however, is hidden in the use of the words "cancel out." The term is deceptive. Aerosol cooling and the greenhouse effect have characteristics that prevent them from neatly offsetting each other. First, the cooling and warming occur mostly over different parts of the world. As we mentioned, sulfate cooling happens primarily over industrial zones in the Northern Hemisphere. Although carbon dioxide spreads throughout the atmosphere, greenhouse forcing is more potent over the subtropical oceans and deserts.

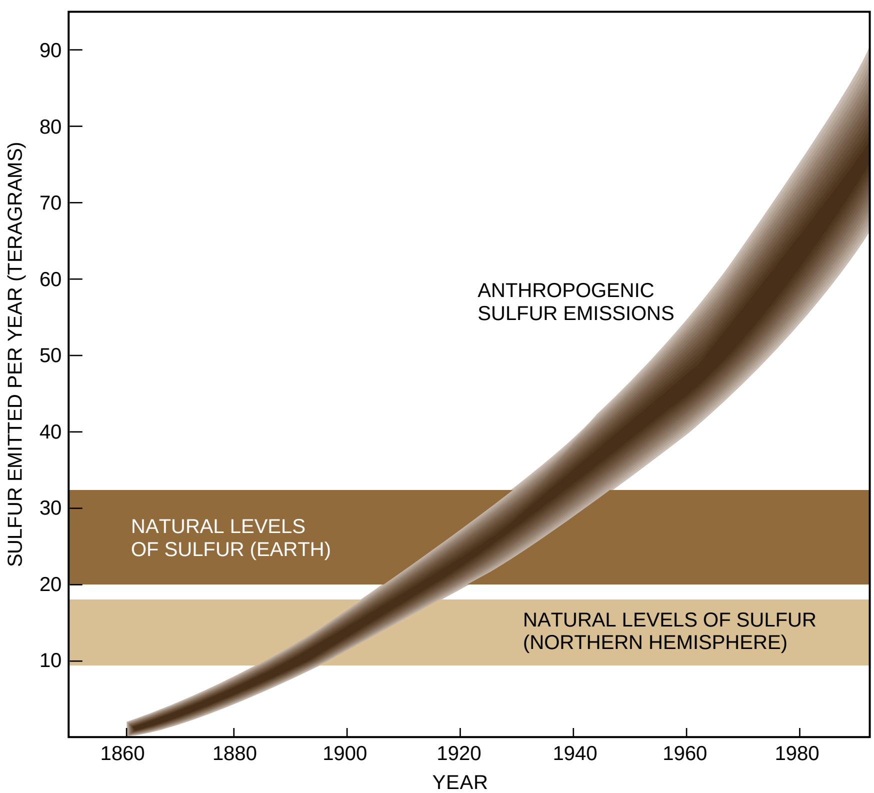
Both types of forcing also differ temporally. The heat-trapping property of carbon dioxide varies only moderately during the course of a day and throughout the year. In contrast, the aerosol effect has a distinctive diurnal and seasonal character. It acts more vigorously in the summer and, of course, operates only during daylight hours. Thomas R. Karl of the National Climatic Data Center and his co-workers have shown that the U.S., the former Soviet Union and China all have displayed increases in annual average minimum temperatures but no increases in the maximum temperatures. It is possible, therefore, that aerosols may now be canceling out greenhouse warming during the day (when temperatures are highest) but not at night (when temperatures are usually lowest).

How should one regard the evidence garnered to date for aerosol cooling? A good way to judge is to compare it with the enhanced greenhouse effect. Although the IPCC recommends large-scale cuts in emissions of carbon dioxide, it has not been able to say conclusively that changes in greenhouse-gas concentrations have caused the observed global warming. A definitive verdict is difficult because the magnitude of the effect to date is roughly comparable to the natural variability of the climate. In other words, the signal is about the same strength as the background noise.

Precisely the same situation applies to aerosol forcing of the climate. Climatologists have not yet found the "smoking gun" that would prove beyond doubt the existence of sulfate cooling. Yet the strong theoretical basis of the aerosol effect, the consistency of the data with expectations and the lack of any counterevidence give us considerable confidence in its reality. Still, two large areas of uncertainty limit our predictive capability: understanding the fundamental physics of global climatic change and forecasting the levels of future emissions of sulfur dioxide. Right now the estimated uncertainty in forcing caused by the best-understood anthropogenic aerosol—sulfates—is much greater than the uncertainty in forcing created by greenhouse gases. For sulfates, the amount of cooling ranges by a factor of two; for greenhouse gases, the degree of warming is known to within roughly one tenth to one fifth.

Nevertheless, we can make a few general predictions. Because anthropogenic sulfate aerosol is confined for the most part to specific parts of the Northern Hemisphere, greenhouse warming should proceed relatively unabated in the Southern Hemisphere (and in more rural parts of the Northern Hemisphere). The IPCC forecast of a sea-level rise of a few tens of centimeters over the next 50 years therefore remains reasonable. A substantial fraction of this rise is associated with the global-scale thermal expansion of the warming water. Other repercussions are somewhat harder to predict, because they depend on the regional details of the combined aerosol and greenhouse forcing.

Reducing the emissions of carbon dioxide and sulfur dioxide would have two contrasting outcomes. Because the



ANTHROPOGENIC SULFUR EMISSIONS now far outstrip those from natural sources, such as marine phytoplankton. It is estimated that humans currently release between 65 billion to 90 trillion grams, or teragrams, of sulfur every year.

How Much Light Do Aerosols Reflect Away?

Atmospheric sulfate aerosol scatters light in all directions. About 15 to 20 percent of the light is scattered back into space. The backscattering constitutes the direct effect of atmospheric aerosol on incoming radiation. The light-scattering efficiency of aerosol, represented by the Greek letter alpha (α), is high, even at low humidity: each gram represents an area of about five square meters. Moisture increases the scattering by making the aerosol expand. At the global average relative humidity, the efficiency doubles, to almost 10 square meters per gram. One can use this value to estimate the magnitude of the direct effect of anthropogenic sulfate.

The rate at which light is lost from the solar beam is defined by the scattering coefficient, represented by the Greek letter sigma (σ, expressed in units of per meter). This value is determined by the amount of aerosol mass, M (in grams per cubic meter), multiplied by the light-scattering efficiency: $\sigma = \alpha M$. When both sides of this equation are integrated over altitude, z, a dimensionless quantity called the aerosol optical depth and represented by the Greek letter delta (δ), results:

$$\int_0^\infty \sigma\, dz = \delta = \alpha \int_0^\infty M dz = \alpha B$$

Here B is the world average burden of anthropogenic sulfate aerosol in a column of air, in grams per cubic meter. The optical depth is then used in the Beer Law (which describes the transmission of light through the entire vertical column of the atmosphere). The law yields $I/I_o = e^{-\delta}$, where I is the intensity of transmitted radiation, I_o is the incident intensity outside the atmosphere and e is the base of natural logarithms. In the simplest case, where the optical depth is much less than 1, δ is the fraction of light lost from the solar beam because of scattering. The question, then, is just how large δ is or, more properly, that part of it that results from man-made sulfate.

This global average burden of anthropogenic sulfate aerosol can be estimated by considering the entire atmospheric volume as a box. Because the lifetime of sulfate aerosol is short, the sum of all sulfate sources, Q, and its lifetime in the box, t, along with the area of the earth, determine B:

$$B = \frac{Qt}{\text{area of the earth}}$$

About half the man-made emissions of sulfur dioxide become sulfate aerosol. That implies that currently 35 teragrams (35×10^{12} grams) per year of sulfur in sulfur dioxide is converted chemically to sulfate. Because the molecular weight of sulfate is three times that of the elemental sulfur, Q is about $(3)(35 \times 10^{12})$ or 1.1×10^{14} grams per year. Studies of sulfate in acid rain have shown that sulfates persist in the air for about five days, or 0.014 year. The area of the earth is 5.1×10^{14} square meters. Substituting these values into the equation for B yields about 2.8×10^{-3} gram per square meter for the burden.

This apparently meager amount of material produces a small but significant value for the aerosol optical depth. Using the value of scattering efficiency (α) of five square meters per gram and a factor of two for the increase in scattering coefficient because of relative humidity, the estimated anthropogenic optical depth becomes $\delta \approx 5 \times 2 \times (2.8 \times 10^{-3}) \approx 0.028$. This value means that about 3 percent of the direct solar beam fails to reach the earth's surface because of man-made sulfate. A smaller amount—perhaps (0.15)(3 percent), or about 0.5 percent—is thus lost to space. This scattering operates over the noncloudy parts of the earth. About half the earth is cloudy at any given time, so that globally 0.2 to 0.3 percent is lost.

carbon cycle and the climate system are slow to respond to changes, carbon dioxide-induced warming would continue for decades. In contrast, reduced sulfur dioxide emissions would rapidly result in a cessation of the cooling because of sulfate aerosol's short atmospheric lifetime. Thus, the ironic result of curtailing fossil fuel use may initially be a warming, particularly in industrial areas.

Of course, many issues about climatic forcing and sulfur remain open. Do other sources of aerosol, such as extensive biomass combustion in the tropics, have a more substantial impact than has commonly been assumed? Even more important, how do meteorologic processes respond to forcing that does not act uniformly over the earth?

One might be tempted to conclude that the uncertainty leaves the issue of human-induced climatic change unresolved. Therefore, one might reason, no changes in policy should be contemplated. That line of thinking, we believe, is a serious mistake. Obviously, no panacea presents itself that would cure the problem of global change. For instance, decreasing emissions of sulfur dioxide to reduce acid rain might accelerate global warming. What does seem clear is that a better and more complete understanding is needed and that a cautious path be taken. Many good arguments can be made for conserving fossil fuels and reducing emissions of both carbon dioxide and sulfur dioxide. Doing so sooner would be less disruptive to the climate than waiting, because these industrial gases humans are releasing at this moment are having an impact on the weather that will persist for decades. The longer the world delays implementing reductions, the more severe will be the consequences.

FURTHER READING

THE CHANGING ATMOSPHERE. Thomas E. Graedel and Paul J. Crutzen in *Scientific American,* Vol. 261, No. 3, pages 58–68; September 1989.

CLIMATE FORCING BY ANTHROPOGENIC AEROSOLS. R. J. Charlson, S. E. Schwartz, J. M. Hales, R. D. Cess, J. A. Coakley, Jr., J. E. Hansen and D. J. Hofmann in *Science,* Vol. 255, pages 423–430; January 24, 1992.

CLIMATE CHANGE: THE IPCC SCIENTIFIC ASSESSMENT. Edited by J. T. Houghton, et al. Published for the Intergovernmental Panel on Climate Change by Cambridge University Press, 1990. And: CLIMATE CHANGE 1992: THE SUPPLEMENTARY REPORT TO THE IPCC SCIENTIFIC ASSESSMENT. Edited by J. T. Houghton et al. Cambridge University Press, 1992.

THE RELATIVE ROLES OF SULFATE AEROSOLS AND GREENHOUSE GASES IN CLIMATE FORCING. J. T. Kiehl and B. P. Briegleb in *Science,* Vol. 260, pages 311–314; April 16, 1993.

The Molecular Architects of Body Design

Putting a human gene into a fly may sound like the basis for a science fiction film, but it demonstrates that nearly identical molecular mechanisms define body shapes in all animals

by William McGinnis and Michael Kuziora

All animals develop from a single fertilized egg cell that goes through many rounds of division, often yielding millions of embryonic cells. In a dazzling and still mysterious feat of self-organization, these cells arrange themselves into a complete organism, in which bone, muscle, brain and skin integrate into a harmonious whole. The fundamental process is constant, but the results are not: humans, mice, flies and worms represent a wide range of body designs.

Noting that variation, biologists have often supposed that the molecular architects of body form—the genetic processes that control embryonic development in different species—would also be quite diverse. There is compelling evidence, however, that an interrelated group of genes, called *HOM* genes in invertebrates and *Hox* genes in vertebrates, governs similar aspects of body design in all animal embryos.

In at least some of the molecular systems that mold our form, we humans may be much more similar to our far distant worm and insect relatives than we might like to think. So similar, in fact, that—as our work has shown—curious experimenters can use some human and mouse *Hox* genes to guide the development of fruit-fly embryos.

WILLIAM McGINNIS and MICHAEL KUZIORA investigate the effects of the homeobox supergene family on animal development. McGinnis is professor of molecular biophysics and biochemistry, with an appointment in biology, at Yale University. After earning his bachelor's degree in biology from San Jose State University in 1978, he continued his studies in molecular biology at the University of California, Berkeley. While working at the University of Basel during the mid-1980s, McGinnis and Michael Levine discovered the homeobox sequence motif in genes of *Drosophila* fruit flies. Kuziora has been assistant professor of biology at the University of Pittsburgh since 1991. He received his Ph.D. from the Baylor College of Medicine and worked with McGinnis for five years at Yale as a postdoctoral fellow.

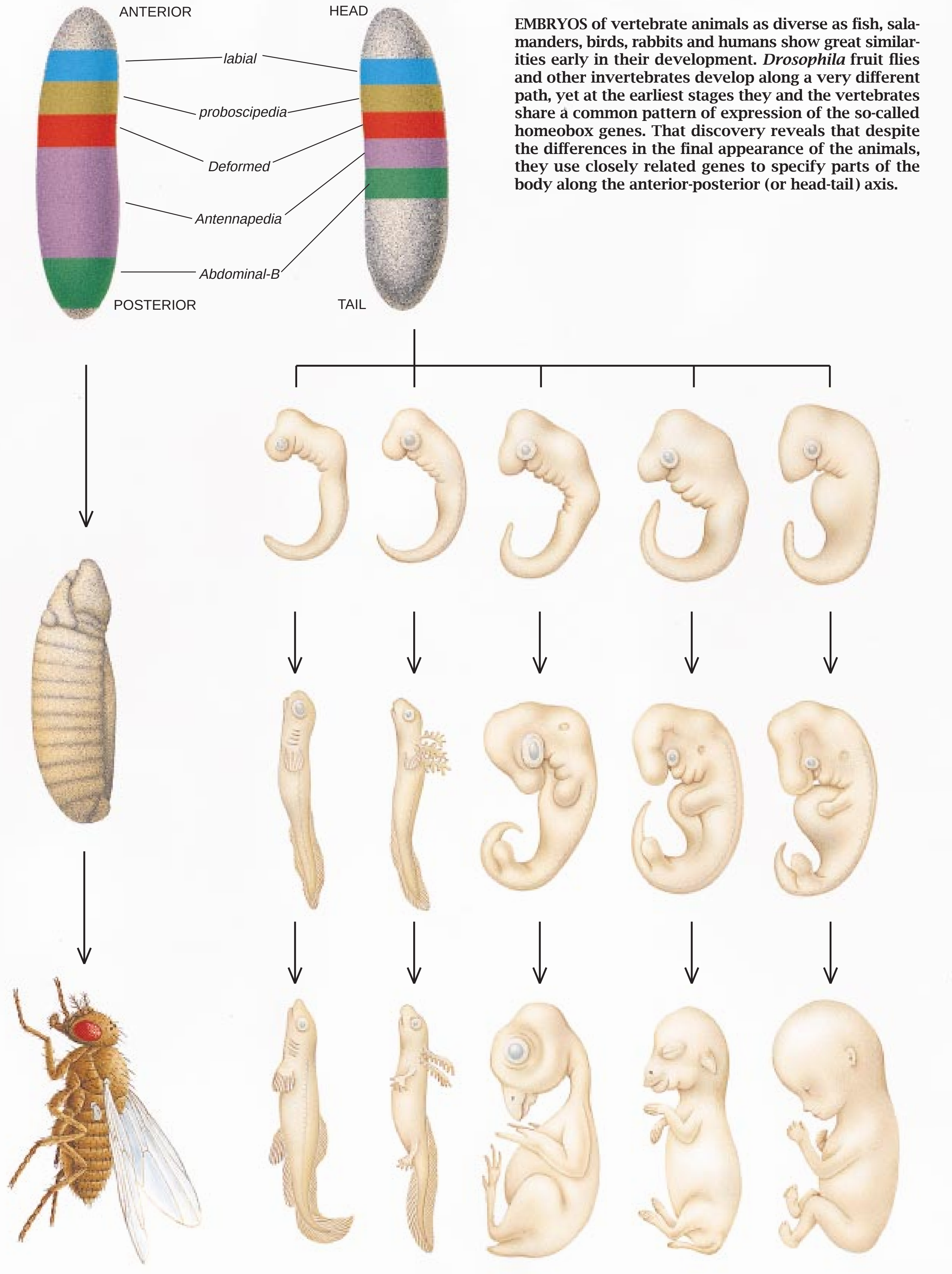
The story of these universal molecular architects actually begins with the pioneering genetic studies of Edward B. Lewis of the California Institute of Technology. Lewis has spent much of the past 40 years studying the *bithorax* complex, a small cluster of homeotic genes in the fruit fly *Drosophila melanogaster.* The Greek word *homeo* means "alike," and the fly homeotic genes are so named because of their ability, when mutated, to transform one body segment of the fruit fly into the likeness of another. Mutations in *bithorax* complex genes usually cause such developmental defects in the posterior half of the fly body plan. Thomas C. Kaufman of Indiana University and his colleagues have discovered and studied a second cluster of fly homeotic genes, the *Antennapedia* complex (named for the founder gene of the complex, *Antennapedia*). Mutations in these genes usually cause homeotic defects in the anterior half of the fly body plan.

It is often the case in biology that bizarre defects in odd organisms contain the clues to solving important problems, and few biological phenomena are more bizarre than the disruptions in body design caused by homeotic mutations. For example, some mutations in the *Antennapedia* gene can cause the antennae on the head of the fruit fly to be transformed into an extra pair of thoracic legs. Surprisingly, some of the animals that develop the extra legs survive, feed and even mate with normal flies.

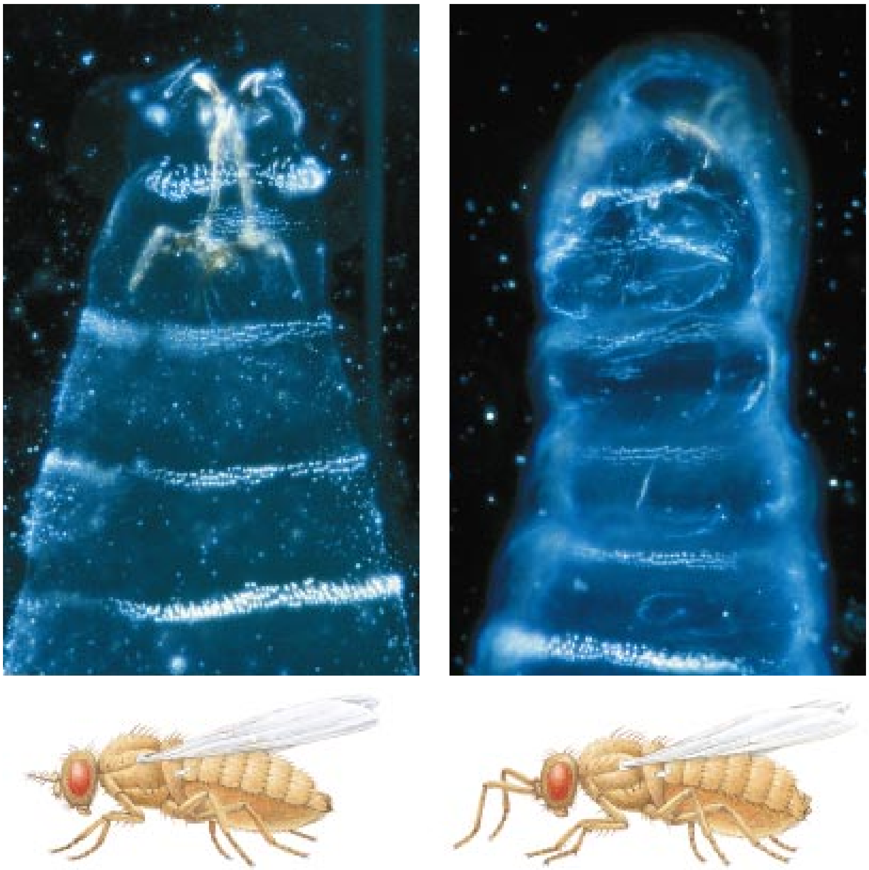
Antennapedia adults are rare exceptions, because most mutations in homeotic genes cause fatal birth defects in *Drosophila*. Nevertheless, even those dying embryos can be quite instructive. For instance, Ernesto Sanchez-Herrero and Gines Morata of the Independent University of Madrid found that elimination of three genes in the *bithorax* complex—*Ultrabithorax, abdominal-A* and *Abdominal-B*—is lethal. Yet such mutant embryos survive long enough to develop specialized structures that indicate all eight abdominal segments are replaced by thoracic segments. Most people would be unnerved by analogous birth defects in mammals, but these grotesque defects in flies can be observed with equanimity.

From his original genetic studies of the *bithorax* complex genes, Lewis derived two key insights. The first was that the normal function of these homeotic genes is to assign distinct spatial (or positional) identities to cells in different regions along the fly's anterior-posterior axis. That is, they "tell" cells that they are part of the fly's head or thorax or abdomen. These identities are to some extent abstract, in that the positional coordinates assigned by homeotic genes are interpreted in dissimilar ways in different developmental settings. *Antennapedia* assigns thoracic identity during both the embryonic and pupal stages of the fly's life cycle, even though the structures (sense organs, legs, wings and so on) that develop along the thorax differ in larvae and adults.

Lewis's second important insight was that the linear order of the *bithorax* complex genes on the fruit fly's chromosome exactly paralleled the order of the body regions they specified along the embryo's anterior-posterior axis.



EMBRYOS of vertebrate animals as diverse as fish, salamanders, birds, rabbits and humans show great similarities early in their development. *Drosophila* fruit flies and other invertebrates develop along a very different path, yet at the earliest stages they and the vertebrates share a common pattern of expression of the so-called homeobox genes. That discovery reveals that despite the differences in the final appearance of the animals, they use closely related genes to specify parts of the body along the anterior-posterior (or head-tail) axis.



NORMAL FLY

Antennapedia MUTANT FLY

HOMEOTIC TRANSFORMATIONS, in which body parts develop in the wrong positions, occur in fruit flies that have mutations in their homeobox genes. Mutations of the *Antennapedia* gene, for example, can cause belts of thoracic denticles (spikes) to appear on the heads of larvae (*top right*). Another developmental consequence of the mutation is that the mutant adults have legs growing in place of antennae (*bottom right*). A normal larva and adult are shown at the left.

The same relation also holds for the genes of the *Antennapedia* complex. United by these shared characteristics, the genes in the *bithorax* and *Antennapedia* groupings are collectively referred to as the *HOM* complex.

A partial understanding of how the *HOM* complex genes determine axial positions in the fruit-fly body plan can come from looking at where those genes are active in embryos. The *HOM* genes are present in the DNA of all of a fly's cells but are active only in some of them. When activated, the *HOM* complex genes are copied as molecules of messenger RNA, which serve as templates for the synthesis of HOM proteins. During early developmental stages, before regions of the embryo show any signs of their eventual fates, the different *HOM* complex genes are activated in successive stripes of cells along the anterior-posterior axis. Some of these stripes of activation overlap, but each *HOM* complex gene has a unique anterior boundary of activation in the body plan.

If deletion of a gene or some similar incident interferes with the expression of a HOM protein, then embryonic cells that normally contain high levels of that protein often undergo a homeotic transformation. That transformation occurs because of a backup *HOM* gene that is already active in the same cells and that can substitute its own positional information. For instance, if the function of the *Ultrabithorax* gene is eliminated from cells within a fly's anterior abdominal region, *Antennapedia* will take over the development of that region. As a result, structures normally associated only with the thorax (which *Antennapedia* helps to specify) also appear more posteriorly.

Homeotic transformations can also result from mutations that cause a homeotic gene to become active in an inappropriate position. The *Antennapedia* mutations in adult flies are caused by activity of *Antennapedia* in the head, where that gene is normally turned off. In summary, the genetic evidence indicates that each *HOM* complex gene is needed to specify the developmental fate of cells in a certain position on the anterior-posterior axis: the posterior head, anterior thorax and so on. More important (and more instructive about their biological function), the activity of *HOM* complex genes is apparently sufficient to determine the fate of at least some cells, even when those cells would not normally fall under a given gene's influence.

The genes of the *HOM* complex are virtually the only ones in *Drosophila* that have those properties. They also share an interesting resemblance at the structural level because all of them are members of the homeobox gene family. Homeoboxes are DNA sequences that carry the descriptions for making a related group of protein regions, all about 60–amino acid residues in size, called homeodomains. The *homeo*- prefix in the name of these domains stems from their initial discovery in *Drosophila* HOM proteins. Since then, however, homeodomains have been found in many other proteins with varying degrees of similarity. The homeodomains of the *Drosophila* HOM proteins are especially similar to one another, which suggests they are closely related. For that reason, they are often referred to as *Antennapedia*-class homeodomains.

What do these HOM proteins do at the biochemical level? Only a superficial answer can be given at present. They belong to a large group of proteins whose function is to bind to DNA in the regulatory elements of genes. The right combination of these bound proteins on a DNA regulatory element will signal the activation or repression of a gene—that is, to start or stop making that gene's encoded protein. Investigators have shown that the homeodomain region of the HOM proteins is the part that directly interacts with the DNA binding sites.

We are fascinated by the contrast between the structural similarity of the HOM proteins and their varied, specific effects. Here is a family of proteins that all bind to DNA and are presumably derived from a single ancestral Antennapedia-class protein. Yet their roles in development are remarkably diverse: one protein assigns cells to become parts of the head, another assigns cells to become thorax and so on. It seems likely that HOM proteins designate various positions along the anterior-posterior axis by regulating the expression of what may be large groups of subordinate genes. The functional specificity of the HOM proteins can therefore be defined by the differences between them that allow them selectively to regulate certain genes in embryos.

To learn more about this specificity, we decided in 1986 to construct chimeric HOM proteins that had components derived from different sources. (The chimera, a monster of Greek mythology, was part lion, part goat and part snake.) By testing the function of these chimeric proteins, we thought it would be possible to define which subregions of the HOM proteins determined their selective regulatory abilities.

For the subjects of our first experiments, we chose the HOM proteins Deformed, Ultrabithorax and Abdominal-B. These proteins have structurally similar homeodomains: that of the Deformed protein is identical to that of Ultrabithorax protein at 44 of its 66 amino acids—but they share no extensive resemblance in other regions. Each of these proteins also exerts an influence on other genes in the HOM family. Thus, the Deformed protein selectively activates the expression of its own gene; Ultrabithorax protein represses the expression of the *Antennapedia* gene; and Abdominal-B protein regulates its own gene and those of others in the *HOM* complex, including *Antennapedia, Ultrabithorax* and *abdominal-A*. We knew we could use these auto- and cross-regulatory relationships in tests of the specific functions of chimeric HOM proteins.

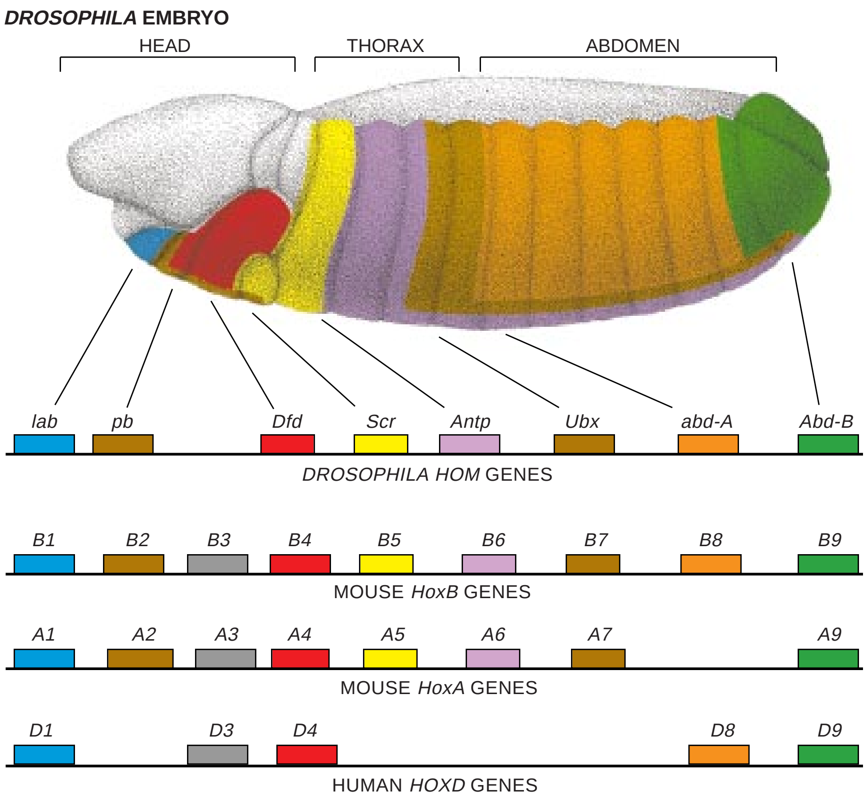
The first challenge was to create genes that would make the chimeric homeotic proteins we desired. Recombinant DNA techniques make that feat possible through the splicing of bits and pieces of genes at the DNA level. If the gene engineering is done with care, protein domains can be very precisely moved from one protein to another while retaining their functional characteristics. We then had to make sure that the chimeric genes would be active in all embryonic tissues. We therefore used a method worked out a few years previously by Gary Struhl, now at Columbia University, that involves attaching the gene to regulatory DNA sequences that can be activated by a mild heat shock. Finally, we inserted our heat-inducible *HOM* gene chimeras into *Drosophila* chromosomes by a technique called P-element transformation.

The *Drosophila* flies that we transformed in this way thereafter carried the chimeric genes in every cell of their body, and those genes would produce chimeric proteins at any stage of development if we simply raised the temperature of the flies' growth chamber to 37 degrees Celsius for a brief period. (*Drosophila* prefer to live at 25 degrees C but can tolerate 37 degrees C for an hour or two with no ill effects.) Using these animals, we could assay the ability of the chimeric proteins to act on the regulatory elements of target genes in their normal chromosomal positions and in their natural embryonic environment—a demanding test that closely mimics the usual conditions under which these proteins operate.

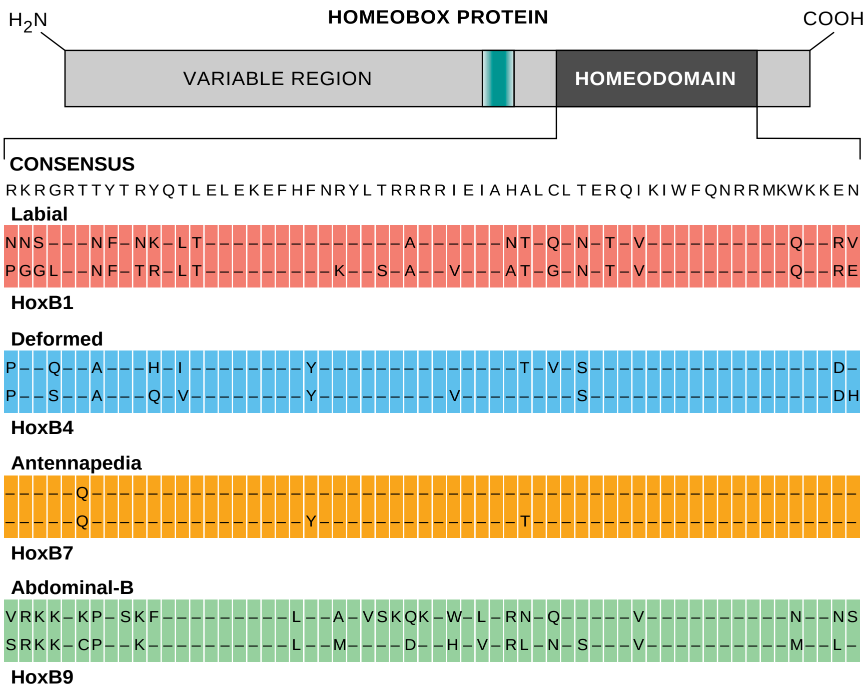
Because HOM proteins have highly similar homeodomains, they bind to nearly identical DNA sites when tested in the laboratory. It therefore initially seemed likely that the features giving each protein its functional specificity would be found outside the homeodomain—in the parts of the proteins that were most individual. Yet as often happens when simple deductive reasoning is applied to biological problems, that expectation was wrong.

We found that if we removed the native homeodomain from a Deformed protein and put an Ultrabithorax homeodomain in its place, the chimeric protein lost the ability to regulate *Deformed* gene expression in embryos. Instead the new protein acted on the expression of the *Antennapedia* gene—much as a normal Ultrabithorax protein would. By transferring the Ultrabithorax homeodomain to Deformed, we had apparently also transferred its selective regulatory abilities. Another homeodomain swap experiment gave us similar results. A Deformed protein carrying an Abdominal-B homeodomain instead of its own mimicked the regulatory specificity of an Abdominal-B protein.

The chimeric proteins did not behave exactly like the protein from which their homeodomain was derived. Both the Deformed/Ultrabithorax chimera and the Deformed/Abdominal-B chimera activated expression of their target genes, whereas the normal Ultrabithorax and Abdominal-B proteins repressed expression of the same genes. Presumably, regions of the Deformed protein outside the homeodomain region supply a strong activation function that can work with any of these HOM homeodomains. Consistent with



HOMEOBOX GENE COMPLEXES have been identified in both invertebrates and vertebrates. *Drosophila* have *HOM* genes, which occupy the same order on the fly chromosome as the anterior-to-posterior order of body regions whose development they control. Mice and humans have *Hox* genes, which are closely related to members of the *HOM* complex and show the same spatial and functional arrangement.



HOMEODOMAINS are the highly similar 60-amino acid regions of the proteins made by all homeobox genes. Each letter in the consensus string represents an amino acid; deviations from that consensus are shown for several closely related HOM and Hox proteins.

this notion, Deformed does have a few regions of protein sequence that are rich in the types of amino acids characteristic of "activation domains" in other gene regulatory proteins.

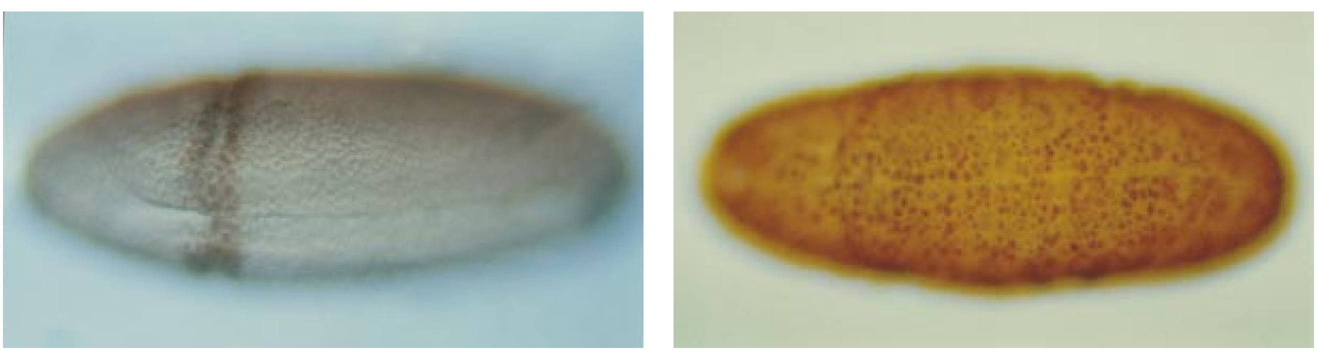
Similar experiments on the functional specificity of HOM proteins have also been carried out by Richard Mann and David S. Hogness of Stanford University and by Greg Gibson and Walter J. Gehring and their colleagues at the University of Basel. Their experiments were based on evaluations of the homeotic transformations that mutant and chimeric HOM proteins induced in developing flies. Because they were looking at the developmental effects of the HOM proteins rather than just at their effects on gene expression, those investigators were using a more demanding measure of HOM protein function than the one we applied. Yet their results, too, support the idea that much (though not all) of the functional specificity of the HOM proteins resides in the small differences within or immediately adjacent to the homeodomain regions.

To us, all those findings also suggested that certain long-shot experiments already ongoing in our laboratory, for which we had only faint hopes of success, actually had a chance of yielding interpretable results. Those experiments involved functional assays of mouse and human homeodomain proteins in *Drosophila* embryos. To convey the significance of those tests, we need to review what is known about the mammalian *Hox* genes.

During the past nine years, genes that contain *Antennapedia*-class homeoboxes have been found in the chromosomes of many animal species besides *Drosophila*. Such genes have been carefully studied in frogs, mice and humans, where they are called *Hox* (short for "homeobox") genes. In both mice and humans, *Hox* genes cluster into four large complexes that reside on different chromosomes. In their organization and patterns of embryonic expression, the genes of the *Hox* complexes share intriguing likenesses to the genes of the fly *HOM* complex. For example, one can identify *Hox* genes that structurally resemble the *HOM* genes *labial, proboscipedia, Deformed, Antennapedia* and *Abdominal-B.* The equivalent *Hox* and *HOM* genes are arranged in the same linear order within their respective complexes.

A further parallel has been observed both by Denis Duboule of the University of Geneva and Pascal Dollé at the CNRS Laboratory of Eukaryotic Molecular Genetics in Strasbourg and by Robb Krumlauf and his colleagues at the National Institute for Medical Research in London. They have assembled convincing evidence that the patterns of expression for the two types of genes are alike. That is, the *Hox* genes are activated along the head-tail axis of the early mouse embryo in the same relative order that the *HOM* genes are activated on the anterior-posterior axis of *Drosophila.*

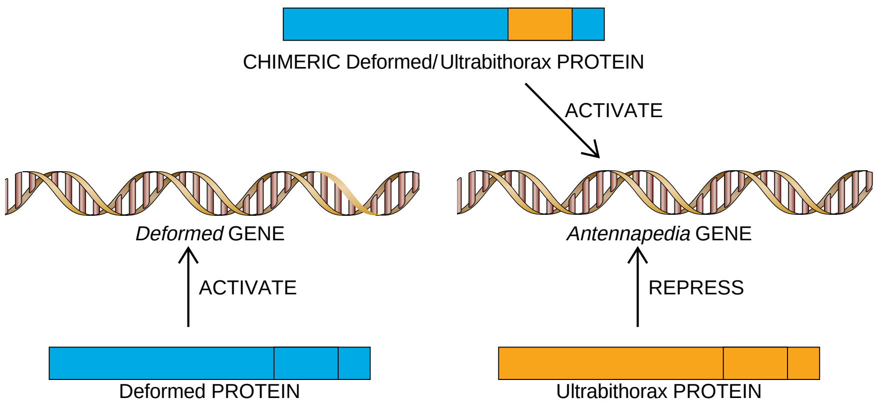
Structural similarities between the mouse and fly proteins are mainly limited to the homeodomain regions. Fly Antennapedia and mouse HoxB6 are nearly identical in the amino acid sequence of their respective homeodo-



EXPRESSION of the *Deformed* gene in fly embryos, as revealed by a brown dye, is normally confined to a band of cells that become posterior head structures (*left*). Genetically engineered embryos that carry heat-inducible *Deformed* genes, however, will produce the Deformed protein in every cell of their body after a brief exposure to heat (*right*). The developmental abnormalities found in such embryos can be used to infer *HOM* gene function.

mains (they differ at only four of 61 positions), which means that these two proteins resemble each other more than Antennapedia does any other fly HOM protein. In an evolutionary sense, this information argues that *HoxB6* and *Antennapedia* are structural homologues—that is, they descended from a common ancestral gene different from the one that gave rise to, say, *Abdominal-B* or *Deformed.*

By the same reasoning, the similarity between the entire *HOM* complex and the *Hox* complexes argues that the most recent common ancestor of *Drosophila*, mice and humans—a wormlike creature that lived about 700 million years ago, give or take a few hundred million years—had a protocomplex of *Antennapedia*-class homeobox genes. The exact type and arrangement of genes in that complex remain a mystery. Nevertheless, we can be confident, using the modern *HOM* and *Hox* complexes as guides, that the ancient protocomplex contained structural homologues of *labial, proboscipedia, Deformed, Antennapedia* and *Abdominal-B.* This overall view of *HOM* and *Hox* gene evolution is strongly supported by research on beetle homeotic genes by Richard W. Beeman of the U.S. Department of Agriculture and Rob E. Denell of Kansas State University and by recent reports from many laboratories that the primitive roundworm *Caenorhabditis elegans* also has a *HOM* complex distantly but recognizably related to the *Drosophila HOM* and vertebrate *Hox* complexes.



GENETIC TARGETS of homeodomain proteins are largely determined by the homeodomain regions of those proteins. For example, a chimeric Deformed protein carrying an Ultrabithorax homeodomain acts on the same genes as does Ultrabithorax. Yet the regulatory effect of the chimera—activation—is more like that of Deformed because of protein regions outside the homeodomain.

All this structural evidence, though suggestive, still does not directly tell us whether HOM and Hox proteins do serve the same developmental function in embryos. After all, the mouse and fly gene complexes have been in different evolutionary lineages for hundreds of millions of years, with plenty of time to evolve new or divergent abilities. So the similarities in structure and expression might be historical quirks and not trustworthy indicators of functional resemblance between present-day HOM and Hox proteins.

One approach to the problem is to explore the biological effects of *Hox* genes in vertebrate embryos and to compare them with what is known about the effects of *HOM* genes in invertebrates. For instance, does the inappropriate activation or specific inhibition of *Hox* gene function during mouse development cause homeotic transformations? In one effort to answer this question, Peter Gruss and his colleagues at the Max Planck Institute for Biophysical Chemistry in Göttingen created strains of mice whose embryos produce HoxA7 protein in the head and anterior cervical region. Normally, HoxA7 protein (which is similar to the Antennapedia and Ultrabithorax proteins of the *HOM* complex) is most abundant in the posterior cervical and anterior thoracic regions and is excluded from more anterior parts. Some mice in which *HoxA7* is expressed inappropriately develop deformities of the ear and palate and occasionally have homeotic transformations of the cervical vertebrae.

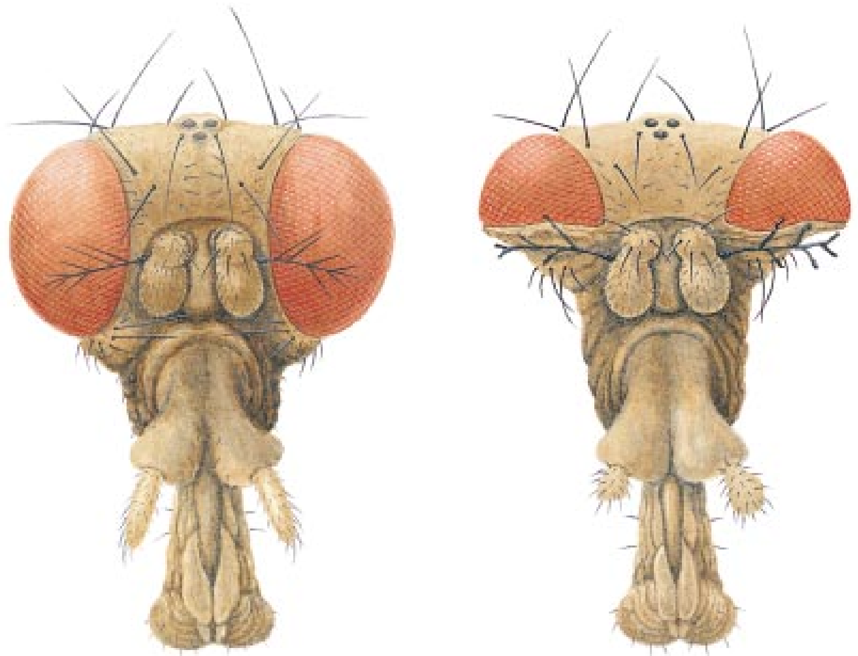
The difficult converse experiment—knocking out *Hox* gene function—has been accomplished for *HoxA3* by Osamu Chisaka and Mario R. Capecchi of the University of Utah and for *HoxA1* by Thomas Lufkin and Pierre Chambon and their collaborators at CNRS in Strasbourg. Their work has shown that some structures in the anterior regions of mouse embryos do depend on those genes. Mutation of the *HoxA3* gene results in mice that die just after birth with a complicated set of head and neck deformities, including abnormally shaped bones in the inner ear and face and the absence of a thymus.

Such deformities are reminiscent of a human congenital disorder called DiGeorge's syndrome, raising hopes that the study of *HOM* and *Hox* genes will be of practical benefit in explaining some human birth defects. Much more research needs to be done before biologists have a good understanding of how *Hox* genes participate in the developmental design of mice and humans, but these and other initial experiments certainly suggest that the *Hox* and *HOM* genes serve comparable purposes.

In our own work, we have tried to make a direct comparison by testing whether Hox proteins can take the place of HOM proteins in developing *Drosophila* embryos. Ideally, one would accomplish this swap by completely replacing the *HOM* gene of a fly with its *Hox* homologue; the *Hox* gene might then be expressed only when and where the *HOM* gene would be ordinarily. Unfortunately, such an experiment is not yet feasible, because the genes in their entirety are too big to be manipulated by current technology. Still, we could do the next-best possible thing: by using *Hox* DNA sequences linked to heat-inducible regulatory elements, we could make all the cells of a developing fly express a Hox protein.

The first protein that we and our laboratory colleague Nadine McGinnis tested in this way was the human HOXD4 protein, the equivalent of a mouse HoxD4. (When referring specifically to human genes, the *HOX* label is capitalized to conform with standard genetic nomenclature.) The gene for this human protein, which has a homeodomain like that of the fly Deformed protein, was isolated and characterized in 1986 by Fulvio Mavilio and Edoardo Boncinelli and their colleagues at the Institute for Genetics and Biophysics in Naples.

In *Drosophila,* when the *Deformed* gene is expressed outside its normal anterior-posterior limits, the adult flies suffer a variety of head abnormalities, such as the absence of a ventral eye. We were amazed to find that the human HOXD4 protein, when expressed in developing fly cells, caused the same deformities. We could not attribute these changes entirely to the human protein, however: our experiments indicated that the human protein was



***DEFORMED* MUTANT FLIES have a variety of head abnormalities, including the absence of the lower half of the compound eyes. The same abnormalities can be induced by making the immature flies express the human HOXD4 protein. This human protein resembles the Deformed protein and seems to have a similar function.**

promoting the expression of the fly's *Deformed* gene as well. (Remember that one normal effect of the Deformed protein is that it activates its own gene, in a cycle of positive feedback.) The human HOXD4 protein was therefore mimicking the effects of inappropriate *Deformed* expression because—at least in part—it *was* causing inappropriate *Deformed* expression. Nevertheless, we could see that *HOXD4* did act like a weak but specific replica of its *Drosophila* homologue.

Encouraged by this result, Jarema Malicki, a graduate student in our laboratory, tested the function of the mouse HoxB6 protein in developing flies. HoxB6, which Klaus Schughart and Frank H. Ruddle identified and characterized a few floors away from us at Yale University, has a homeodomain that is highly similar to Antennapedia protein. The effects of HoxB6 protein expression in developing fly cells was spectacular and unmistakably homeotic. In *Drosophila* larvae the HoxB6 protein caused much of the head region to develop as if it were thoracic: instead of a larval head skeleton, the transformed flies produced denticle belts, rows of spikes that are usually arranged on the bellies of *Drosophila*. In *Drosophila* adults, HoxB6 caused a homeotic transformation of the antennae into thoracic legs. Both the larval and adult homeotic transformations were much like those caused by the inappropriate expression of Antennapedia protein throughout the body.

What can one make of these evolutionary swap experiments? First of all, they reinforced our conclusions that the homeodomains themselves determined much of the regulatory specificity of the proteins: the homologous fly and vertebrate proteins have little in common outside the homeodomain region. In addition, the experiments suggested that from a functional standpoint, the homologous proteins are at least somewhat interchangeable and have similar "meanings" for early embryos. The system for determining anterior-posterior axial positions has evidently changed little in the past 700 million years.

If one were to imagine the complicated network of interactions between gene regulatory proteins inside an organism as a jigsaw puzzle, then the homologous fly and mammal proteins are pieces that can fit in the same places. Looking at the *HOM/Hox* system in this way also highlights how much we still have to learn: the other puzzle pieces that enable the HOM and Hox proteins to regulate genes and to have a specific function have yet to be identified.

In a way, these experiments also hark back to the classical observations of Karl Ernst von Baer, who in the 1820s concluded that if one examined early embryonic morphologies, all vertebrate forms seemed to converge toward a common design. The story, which sounds too good to be true, is that von Baer came to this epiphany after the labels fell off some of his bottled specimens of early embryos, and he realized with some chagrin that he could not be sure whether the embryos were lizards, birds or mammals. The structure and function of the *HOM* and *Hox* gene systems suggest that this developmental convergence embraces the early development of a great many animal species. But only at the level of molecular pattern can the developmental convergence of such different embryos be "seen."

Sometime between 600 million and a billion years ago, the *HOM/Hox* system evolved; it has proved so useful that many animals have since relied on its fundamental abilities to determine axial position during development. Is it the only developmental genetic system that has been so conserved? That seems unlikely. Researchers have found hints that some other regulatory genes in flies and mice are highly similar in structure and are activated in the same or homologous tissues. Exploring the functions of those novel genes, and how they interact with the *HOM/Hox* system, promises to reveal many more fascinating insights into the evolution and mechanism of the ancient genetic systems that serve as the molecular architects of animal body plans.

FURTHER READING

HOMEOBOX GENES AND THE VERTEBRATE BODY PLAN. Eddy M. De Robertis, Guillermo Oliver and Christopher V. E. Wright in *Scientific American,* Vol. 263, No. 1, pages 46–52; July 1990.

MOUSE *HOX-2.2* SPECIFIES THORACIC SEGMENTAL IDENTITY IN DROSOPHILA EMBRYOS AND LARVAE. Jarema Malicki, Klaus Schughart and William McGinnis in *Cell,* Vol. 63, No. 5, pages 961–967; November 30, 1990.

HOMEOBOX GENES AND AXIAL PATTERNING. William McGinnis and Robb Krumlauf in *Cell,* Vol. 68, No. 2, pages 283–302; January 24, 1992.

THE MAKING OF A FLY: THE GENETICS OF ANIMAL DESIGN. Peter A. Lawrence. Blackwell Scientific Publications, 1992.

THE MOUSE *HOX-1.3* GENE IS FUNCTIONALLY EQUIVALENT TO THE DROSOPHILA SEX COMBS REDUCED GENE. J. J. Zhao, R. A. Lazzarini and L. Pick in *Genes and Development,* Vol. 7, No. 3, pages 343–354; March 1993.

When Is Seeing Believing?

Digital technology for manipulating images has subverted the certainty of photographic evidence

by William J. Mitchell

The Manipulation Process

To create the "angry" image (*above, center*) from the original photograph (*above, left*), the computer artist first silhouetted the flowers out of the background. Next he cloned foliage from the upper right of the original and used it as a "paintbrush" to stipple greenery that replaces the porch behind the flowers in the original photograph. He softened, or blurred, Bush's outline and surround, a process called vignetting, and slid them under the silhouetted flowers. All the shrubbery behind the pair was blurred slightly.

The artist then silhouetted the rock in front of Thatcher and flipped it. The ferns that conceal the end of the hose in the

In September 1993 newspapers around the world published an astonishing—almost unbelievable—photograph of Israeli prime minister Yitzhak Rabin shaking hands with Palestine Liberation Organization chairman Yasser Arafat on the White House lawn while President Bill Clinton looked on. In 1988 *Life* magazine published an equally striking picture of Chairman Arafat warmly greeting then Prime Minister Yitzhak Shamir under the approving gaze of President Ronald Reagan. One of these images recorded an actual event and provided reliable evidence that peace was perhaps at hand in the Middle East. The other was a computer-manipulated composite, a tongue-in-cheek fiction.

Such fake "photographs" can now be produced by using widely available "paint" and image-processing software to rearrange, recolor and otherwise transform the elements of a scene. The same software can combine fragments of different images into one new image. Other software can generate completely synthetic photorealistic pictures by applying sophisticated perspective projection and shading to digital models of three-dimensional scenes—a technique commonly employed to present architectural projects and to create Hollywood special effects.

Unlike drawings and paintings, which we regard as inherently untrustworthy products of human intention, these fakes can easily trick us into false beliefs. Like fingerprints

DIGITAL MANIPULATION of news photographs can become a novel form of spin doctoring. Two of these pictures were manipulated to suggest different emotions. What actually took place between George Bush and Margaret Thatcher—a chat, a quarrel or an intimate whisper?

original had to be duplicated and expanded to give the hose some place to go in the new image. And to the right of Bush's extended leg is a piece of the wall duplicated from that just to the right of the hose in the original. The new fern softens the transition between the two pieces of wall as well as hiding the hose.

In the "intimate whisper" image (*above, right*), Bush's vignetted surround was slid over the leafy shrub between him and Thatcher: Bush replaces bush.

Cut-and-Paste Photomontage

Traditional cut-and-paste methods transformed a 1923 photograph showing three Soviet farmers (*bottom*) into "evidence" that surfaced in July 1991 of the continued imprisonment of three lost fliers in Vietnam (*top*). Conventional photomontages, except when they are produced with great care by highly skilled artists, are relatively easy to detect as fakes. This one has many imperfections that give it away.

Look closely at the Stalinesque mustaches—especially the one on the left. Do they look fully convincing? Can you detect the handwork?

Should we read the sign as a foreshortened rectangle (with the right edge nearer the camera than the left) or as an irregular shape? Is it in front of the figure on the left, or is it behind? What is holding it up? (Note the hand at the top left in the original.) The spatial ambiguities suggest that it was pasted in.

Notice the fuzziness and poor tonal quality. This blurring may be deliberate, to conceal some of the imperfections. It also suggests generational loss resulting from rephotographing rather than printing from an original negative.

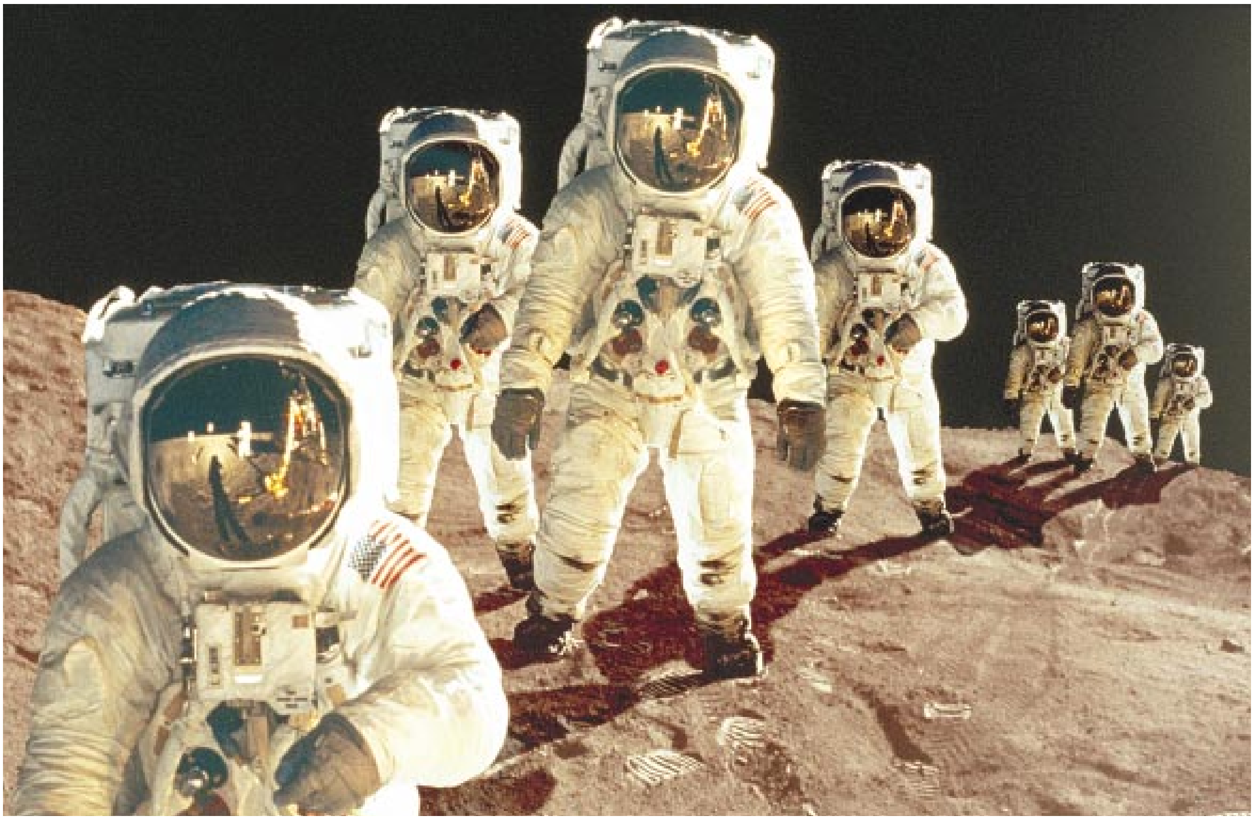
left at the scene of a crime or lipstick traces on a collar, they apparently result from causal rather than intentional processes and therefore seem accurate and dependable evidence of what actually took place.

Faking photographs to manipulate belief is not new. Multiple exposure and printing, cutting-and-pasting and retouching have frequently been used for hoaxes and for political propaganda. In the 19th century, for example, "spirit photographs" were produced by double exposure, and in the Stalinist era propagandists airbrushed the politically inconvenient Trotsky out of a famous photograph depicting Lenin addressing a crowd on May 5, 1920. But digital images make production of photographic falsehoods quicker and easier—and often much more difficult to trace. The question of how to distinguish visual fact from fiction is becoming increasingly urgent as we witness the explosive proliferation of digital-imaging technology. We are approaching the point at which most of the images that we see in our daily lives, and that form our understanding of the world, will have been digitally recorded, transmitted and processed.

Close inspection of traditional photomontages often reveals clear physical evidence of the doctoring that has taken place [*see illustration above*]. Printing masks and knife cuts may produce implausibly sharp edges; pencil marks and paint dabs may stand out against surrounding grainy textures; blends and reconstructions of surface textures may be imperfect; colors may not quite match. But digital images are manipulated by altering pixel values stored in computer memory rather than by mechanically altering surfaces. Therefore, digital photomontages typically show far fewer traces of the artist's hand. Furthermore, software for producing digital composites often simplifies the artist's task to such an extent that it may no longer take much time or much craftsmanship. Such software offers effective tools for tracing and blending edges, matching tones and colors and replicating textures. Thus, the fact that an image presents a seamless surface no longer provides strong grounds for concluding that it has not been manipulated. We must look for other kinds of clues—such as internal consistency, documentable provenance and consistency with existing beliefs.

Digital Photomontage

Digital photomontage of seven astronauts on the surface of the moon was produced from an original photograph made in 1969 by NASA of a single astronaut, Edwin F. Aldrin, Jr. (*bottom*). This montage is of high technical quality; it is carefully contrived to seem spatially consistent, and sophisticated digital technology has eliminated any obvious signs of cutting and pasting. So how can we tell it is a fake?



There is an obvious internal inconsistency. Because the composition was produced from scaled-down replicas of the original figure, the reflections in the visors are incorrect. Each shows the image of just one other astronaut, not the several we would expect.

There is implausible repetition—good grounds for suspecting the image was produced by replication operations. Is it likely that all the astronauts except one would be holding their left arms in precisely the same position?

There are some questionable cast shadows. The artist had to insert these with a digital "paintbrush" and faced the difficult task of making them consistent with the rough terrain. Do those at the back stand up to close, critical inspection?

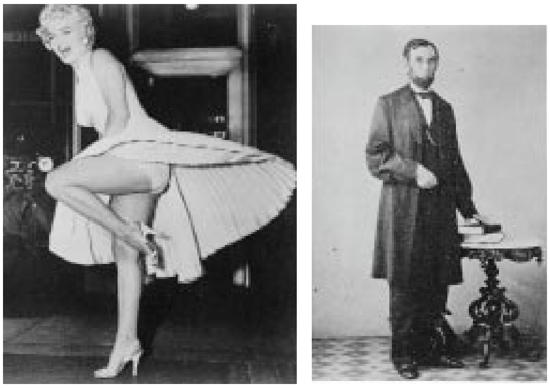
Such flaws and inconsistencies are very obvious once they have been pointed out, but they often go unnoticed at first glance. A skilled forger attempts to anticipate the types of cross-checking that a suspicious viewer will perform, then adjusts the visual evidence accordingly.

Freedom from internal inconsistencies does not demonstrate the veracity of an image. The existence of such flaws may, however, serve to refute claims that an image is a photographic transcription of physical reality [*see illustration above*]. Here are a few of the more obvious questions to ask when looking for inconsistencies. Do all the objects in the scene seem to be in correct perspective? Does the perspective foreshortening of a surface seem consistent with the spatial orientation suggested by its shading? Do indicators of time, such as clocks and shadows, all show the same moment of exposure? Do some objects seem surprisingly light or dark in relation to their surroundings? Are inserted objects betrayed by lack of shadows or by shadows cast at angles different from those cast by other objects? Are there shadows that do not seem to be cast by any object? Are shadows and specular highlights consistent with the assumptions about locations of light sources? Do unexpected discontinuities in the background of the scene suggest that objects have been deleted from the foreground? Do shiny surfaces display the expected reflections of other parts of the scene? Are surface intensities appropriately modified by diffuse interreflection effects? Are there plausible texture

Creating the Cover Image

For base images, the computer artist chose a publicity still of Marilyn Monroe made to promote *The Seven-Year Itch* in 1955 and a formal portrait taken of Abraham Lincoln in 1863 (*1*). To yield a richer final image, he scanned these black-and-white pictures into the computer as if they were in color. Once he had the images digitized, he flipped the Monroe photograph across its vertical axis and silhouetted her out of the background (*2*). In the Lincoln image, he extended the left side of the floor (*3*); the piece shown floating in space was copied and duplicated several times to provide the additional pattern needed.

1



2

3

Next he filled the wall in roughly (*4*) in preparation for the silhouetted figure of Monroe, which he placed on top of the Lincoln image (*5*). He masked the two figures from the surrounding area so that the background could be evened out without affecting them (*6*). Close-ups (*7*) show Monroe before and after “noise” was added to degrade her image to match that of the older Lincoln photograph. (The noise function, which is built into the software, adds random pixels.) To provide Monroe with a hand and arm that she could tuck under Lincoln’s elbow, the artist made a video of his wife holding the arm of his assistant (*8*), using lighting that simulated that in the original photographs. He then froze a frame of the video and captured it in his computer. The video was converted from color to gray scale and composited into the Lincoln-Monroe image (*9*).

4

5

6

7

8

9

COMPUTER ART BY JACK HARRIS/*VISUAL LOGIC*

WHICH VERSION IS A TRUTHFUL DEPICTION? Most people have no way to judge which of these two images produced for low-resolution video portrays the actual building. No internal clues reveal the fake. Scholars of Italian Renaissance architecture, however, can immediately recognize that the version with the tower (*right*) is Andrea Palladio's published but unbuilt scheme for the Villa Godi, whereas the version without a tower (*left*) shows the design that was actually constructed. The image of the unbuilt project was produced by shifting windows and chimneys, extending the narrow steps and replicating fragments of wall and roof surface to delineate the tower. Increasingly, our capacity to sort visual facts from falsehoods will rest on our ability to cross-check the visual evidence against established knowledge and beliefs.

gradients as surfaces recede? Do geometric perspective (size diminution with depth into the scene) and atmospheric perspective (color shift with depth) convey the same depth information? Is there a consistent gradient of sharpness from some focus plane?

In general, the more information there is in an image, the more difficult it is to change without introducing detectable inconsistencies. It is, for example, much easier to alter a fuzzy, low-resolution, dimly lit, black-and-white image than it is to do the same with a sharp, high-resolution, full-color picture. A photographic manipulator, like a dissembler who weaves a tangled web of lies and eventually trips himself up, runs the risk of being caught out by some subtle inconsistency that shows up when the visual evidence is carefully cross-checked.

Fully synthetic images produced by three-dimensional visualization software may, however, be free of such defects. Carefully modeled three-dimensional scenes, rendered in shaded perspective by ray tracing or radiosity (which calculate the interreflection of light within a scene), can simulate all the complex effects of light and shade that even the most demanding observer expects to see. Some fictional "photographs" may readily pass the internal consistency test.

We must then turn to such other criteria as the provenance of the picture. Because a photograph is exposed at a specific time and place, we can always ask to hear the story of how the photographer came to be at that very spot at that exact moment. Furthermore, we can ask for an account of how the picture subsequently made its way from the point of exposure to its present location. Sometimes the general credibility of a source—whether, for example, a purported picture of the surface of Mars is published in *Nature* or in a supermarket tabloid—can stand as a surrogate for an explicit narrative.

The provenance of a traditional photograph is often easy to trace, because exposed films, negatives and prints must be carried physically from place to place and because developing and printing must be performed in suitably equipped darkrooms. Digital-imaging technology makes this job much tougher: it eliminates negatives, it can replicate files in seconds, and digital images can be transmitted rapidly and invisibly through computer and telephone networks.

The subtlest challenge is posed by images that show no detectable signs of tampering and no obvious internal inconsistencies and yet contradict our established beliefs. Let us say, for example, we do not believe Elvis Presley still lives but are presented by an apparently reputable source with a sharp, detailed "photograph" of him in a recognizably contemporary setting. We can either maintain our confidence in the reliability of prior evidence that Elvis is dead and reject the image as a visual falsehood, or we can accept the new evidence before us and correspondingly modify our beliefs. Such intellectual judgments will be increasingly crucial and increasingly difficult to make with confidence in a world where convincing visual evidence can be faked with ease.

For a century and a half, photographic evidence has seemed unassailably probative. Chemical photography's temporary standardization and stabilization of the process of image making effectively served the purposes of an era dominated by science, exploration and industrialization. Photographs appeared to be reliably manufactured commodities, readily distinguishable from other types of depictions. They were generally regarded as causally generated, truthful reports about things in the real world, unlike more traditionally handcrafted images, which seemed notoriously ambiguous and uncertain human constructions. The emergence of digital imaging has irrevocably subverted these certainties, forcing us all to adopt a far more wary and vigilant interpretive stance. The information superhighway will bring us a growing flood of visual information in digital format, but we will have to take great care to sift the facts from the fictions and the falsehoods.

WILLIAM J. MITCHELL is author of *The Reconfigured Eye: Visual Truth in the Post-Photographic Era* (MIT Press, 1992). He is professor of architecture and media arts and sciences and dean of the School of Architecture and Planning at the Massachusetts Institute of Technology. His research focuses on computational techniques and digital media in design and the visual arts. His previous publications include *The Logic of Architecture* (MIT Press, 1990) and, with Malcolm McCullough, *Digital Design Media* (Van Nostrand Reinhold, 1991).

Liquid Mirrors

Light, liquid-mercury mirrors, which can potentially be made much larger than glass mirrors, may enable astronomers to construct enormous telescopes and see farther than ever before

by Ermanno F. Borra

For almost four centuries, reflecting telescopes have gathered the shimmering light from billions of stars and galaxies. The observations they made possible have revealed a universe vast and complex. In doing so, these instruments have rescaled our world and our imaginations. And yet, for all their success, classical reflector technology has some severe limitations. It is very expensive, and often nearly impossible, to grind and polish a massive glass surface into a perfect parabola, the ideal contour for a device that focuses parallel rays of light into a single point. Such mirrors can deform during temperature changes, and beyond a certain size they are prone to buckle under their own weight.

Astronomers and optical experts have therefore sometimes toyed with an old, even bizarre technological alternative: a mirror made from liquid. A liquid mirror could never sag and so conceivably could be made as large as purpose requires. Moreover, coaxing a liquid into the shape of a parabola is trivial. The pull of gravitational and centrifugal forces shapes the surface of a rotating mass of reflecting liquid, such as mercury, into a perfect parabola. This phenomenon, which also occurs when you stir your coffee, provides a superb optical surface that needs no polishing. As a result, liquid mirrors could be far less expensive than glass mirrors. The potential reduction in cost could render large optics affordable even for the amateur astronomer [see "Making a Mirror by Spinning a Liquid," by Mark Dragovan and Don Alvarez, "Amateur Scientist," page 116]. In addition, because optics are so central to many scientific measurements, liquid mirrors could prove useful in many research fields and engineering as well.

No one knows who first conceived of a liquid mirror. Newton himself may have originated the idea. The great natural philosopher understood that the surface of a rotating bucket of water takes a parabolic shape, and it was he who invented the first reflecting telescope. Still, the concept was never pursued seriously until the beginning of this century, when Robert Williams Wood, an eccentric physicist and writer at Johns Hopkins University, tried to construct a liquid-mirror telescope.

Wood's interests ranged from researching optics and spectroscopy to writing nonsense verse for children. He also wrote fiction and co-authored *The Man Who Rocked the Earth* with Arthur Train. Wood may be best known as the man who rocked France in 1904, by denouncing the discovery of N-rays, a form of radiation proffered by René Blondlot of the University of Nancy. At Wood's request, Blondlot demonstrated in his darkened laboratory how N-rays, which he believed to be similar to x-rays, could be detected once filtered through a prism. Wood was skeptical and so slipped the prism, central to the experiment, into his pocket. Blondlot did not notice, and neither did his results. The incident was widely celebrated in the American press.

Wood's liquid mirror, however, received less acclaim. Technical challenges await anyone who attempts to construct such a device, and Wood had little luck in solving them. Although he did photograph passing stars, the images were blurred. The bearing supporting his mirror was inadequate. Consequently, the rotational speed of the mercury varied, causing the focal length to shift. Moreover, vibrations and drafts made the surface of the mercury ripple. If that was not enough, yet another major difficulty dogged the astronomer: a liquid mirror cannot be tilted as can a glass mirror to compensate for the rotation of the earth. As a result, stars registered on his photographic film as streaks. Wood delightfully chronicled all these problems in an account, written in the leisurely manner of his time, for the *Astrophysical Journal*.

There the technology rested until January 1982. During that month, a team of scientists demonstrated an extraordinary method that overcomes the tilting problem. All night atop Mount Palomar, James E. Gunn and Peter Schneider of Princeton University and Maarten Schmidt of the California Institute of Technology surveyed a strip of sky through the Hale five-meter telescope while the instrument sat parked in a fixed position. Charge-coupled devices (CCDs), sophisticated solid-state light sensors, enabled the investigators to capture an image accurately without moving the telescope's mirror.

The CCD detector compensated for the earth's rotation by moving its light sensors electronically from the east to the west, at a rate matching the drift of images in view of the telescope. This is equivalent to taking a picture with a photographic film that moves in a camera at the same speed as the image of a moving object. An object crosses the narrow width of the detector typically in only a couple of minutes, limiting the amount of light that can be gathered. As the same regions of the sky are observed night after night, however, it is possible to create increasingly intense images by digitally adding subsequent exposures on a computer.

ERMANNO F. BORRA has worked to develop liquid-mirror telescopes since early 1982, when he became interested in cosmology during a sabbatical at the University of Arizona. He received a physics degree in 1967 from the University of Torino in Italy and a Ph.D. in astronomy in 1972 from the University of Western Ontario. For the next two years, he served as a Carnegie postdoctoral fellow at the Hale Observatory (now the Mount Wilson, Palomar and Las Campanas observatories). Since 1975 he has been at Laval University in Quebec, where he is now a professor of physics and a member of the Center for Optics, Photonics and Lasers. Borra is a member of the Canadian Network for Space Research and an associate member of Mont Mégantic Observatory.

The achievement of Gunn, Schneider and Schmidt reawakened my own interest in liquid mirrors. I had first learned about them in high school. The concept had intrigued me, but even as an astronomer I never thought seriously about employing them in research. At the time of the Palomar work, I was on sabbatical at the University of Arizona, where two investigators hoped to take advantage of the capabilities of CCD recorders. J. Roger P. Angel and John McGraw were planning to build a novel telescope, now in routine operation, to search for distant supernovae. The telescope is permanently fixed, thus having no need for an expensive movable frame and dome.

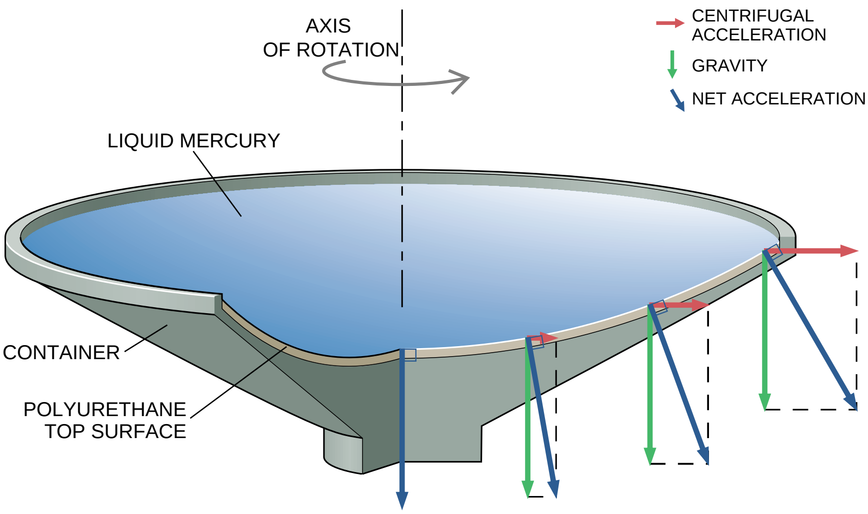
By the time my sabbatical ended, I realized that if a CCD camera could compile precise images from a fixed glass mirror, then the same should be true for a liquid mirror. Of course, such precision would be wasted if a liquid mirror could not generate a high-quality image. So when I returned to Laval University, I determined to find out just how well a liquid mirror could perform. First I ordered an air-lubricated bearing as well as an electric motor. Our shops built all the other components I needed to mount the device.

Within a few months I examined my first mercury mirror, 50 centimeters in diameter, using the knife-edge test. To perform it, one brings a sharp blade toward the reflected image beam of a point of light. An ideal mirror should cast an image equal in size to the point source. The image then can be eclipsed in an instant as the thin razor edge interrupts the path of reflected light. A bad mirror diffuses light, creating a larger image that shines around the edge of the blade. The test resulted in a clean block—the unmistakable signature of a parabola.

I quickly built a one-meter mirror to study the liquid-mercury technology more thoroughly. I enlisted the help of Laval graduate students Robin Arsenault, now at the CFHT Corporation in Kamuela, Hawaii, and Mario Beauchemin. Further analysis confirmed that the mirror's surface was perfectly parabolic and reasonably smooth. In particular, the machinery supporting the mirror was so stable that surface ripples were trivial and the focal length was steady.

Given these encouraging results, I de-

LIQUID-MIRROR TELESCOPE at the University of Western Ontario's Light Detection and Ranging observatory collects light after atmospheric molecules have been excited by a powerful laser beam.



ROTATING LIQUID SURFACE assumes the shape of a parabola under the constant pull of gravity and a centrifugal acceleration, which grows stronger at distances farther from the central axis. The parabolic curve occurs because a liquid surface must always be locally perpendicular to the net acceleration it experiences, which in this case becomes increasingly steep with distance from the central axis.

cided to embark on serious development. First we had to build a facility suitable for making exact optical measurements and supply it with state-of-the-art equipment. This preparation took far more time and energy than the actual experiments. Control of subtle optical alignments, building vibrations and air turbulence had to be considered during construction. Such disturbances could do great harm: defects as small as 1/40 of the wavelength of optical light on the surface of any mirror must be minimized to achieve a high degree of optical accuracy.

After the painstaking construction was completed, we were rewarded with a spectacular sight. The event happened in the course of our early work with a mercury mirror, 1.5 meters in diameter. Stanislaw Szapiel, now at the National Institute of Optics, had obtained an unresolved image of an artificial star on a television monitor. We magnified that image through a microscope lens, and to our surprise the monitor showed a disk surrounded by alternating rings of darkness and dim light: the image looked like the diffraction pattern of the mirror! This shading occurs because even when the quality of an optical surface nears perfection, the wave nature of light imposes a fundamental limitation. Light waves reflected from the mirror overlap and either eliminate or reinforce one another, much as ripples on the surface of a pond can generate complex patterns. A point of light then appears as a disk surrounded by shaded rings, corresponding to the constructive or destructive interplay of reflected light. As a result, the sharpness of an image is set not so much by the quality of the mirror but by its diameter: the broader the mirror, the sharper the image. We were skeptical at first that what we saw on the monitor was indeed the mirror's signature diffraction pattern. But after much arguing, computing and experimenting, we hesitantly—but happily—accepted the likelihood that our mirror was almost perfect.

Confirmation of this conclusion required still more rigorous testing. Robert Content, then a Laval doctoral candidate, surveyed the 1.5-meter mirror extensively using a scatter plate interferometer. This instrument maps the contour of a surface with hair-splitting precision by recording interference patterns of light. We processed hundreds of interferograms recorded by a CCD camera to satisfy ourselves that our results were representative.

Liquid mirrors are more challenging to survey than glass mirrors because the surface of a liquid can change shape rapidly. The measurements cannot be averaged together, as is done with glass mirrors, because averaging minimizes the aberrations introduced by air turbulence and vibrations. In this stage of testing, our careful preparations paid off. The interferometry ultimately showed that a well-tuned liquid mirror maintains a parabolic surface accurate to at least 1/30 the size of a wave of light, which is close to the original accuracy specified for the *Hubble Space Telescope.*

Although our indoor laboratory tests delivered results far better than our expectations for two years, we needed to evaluate how a liquid mirror would behave in an outdoor setting, exposed to the elements. For this purpose, we built an observatory housing a 1.0-meter liquid mirror in 1986 and a 1.2-meter liquid mirror the next year. Students operated the observatory for 63 clear nights, searching for unconfirmed stellar flashes that had been previously reported elsewhere. The detector was a programmable 35-millimeter camera that registered star trails lasting no longer than two minutes. We had little difficulty assembling this observatory at a reasonable cost [*see box on page 80*]. In the end, the entire setup performed well. We found no flashes but concluded that if they exist, they are rare. More important, this work led to a milestone publication, describing the first successful research done using a liquid-mirror telescope.

But what observations can be made with a telescope that cannot be pointed at will, no matter how easily built, inexpensive and accurate? I maintain that almost any kind of sensing and recording system could be adapted to a fixed telescope. Even if we consider only the proved technique—recording observations using a CCD—liquid-mirror telescopes can be expected to aid survey work significantly.

Cosmologists, who chart the universe in particular, should benefit. They must observe extremely faint objects in their research and therefore need considerable time observing with large telescopes. Because conventional glass-mirror telescopes are so costly, no individual scientist can obtain this kind of access. The demand on sizable telescopes is so great that even collaborative teams secure at most a dozen nights a year for a particular project. Consequently, it takes years to complete these surveys.

Inexpensive liquid-mirror telescopes could be made available to cosmologists more often than glass-mirror telescopes. This affordable technology could speed the rate of progress in many specialized undertakings, ranging from the search for distant supernovae and quasars to the study of the evolution and topology of the universe. Already Paul Hickson of the University of British Columbia, in a joint project with Laval, has built a 2.7-meter liquid-mirror telescope to carry out spectroscopic surveys. The device uses a CCD recorder and interference filters. It saw first light in late 1992. Andrew E. Potter, Jr., of the National Aeronautics and Space Administration and Terry Byers of Lockheed Corporation have built a liquid-mirror telescope three meters in diameter to search for space debris as small as half an inch. Such detritus can threaten satellites, spacecrafts and space stations.

This new technology offers interesting properties in addition to large sizes: high surface quality, low-scattering, very fast apertures and a variable focus that can be controlled with fine precision. As a result, liquid mirrors can enhance research in many fields of science. We have worked with a group led by Robert J. Sica of the University of Western Ontario to use a liquid mirror as the receiver for a Light Detection and Ranging (LIDAR) system. Such instruments monitor the atmosphere from 30 to 110 kilometers above the earth. First the device fires a powerful laser beam into the sky to excite atmospheric molecules. These molecules then emit light, the intensity and wavelength of which are indicative of density and temperature conditions at their location. The receiver collects the telltale light for interpretation. The light-gathering power of the 2.65-meter mercury mirror we built for this system places it among the world's most sensitive atmospheric recorders.

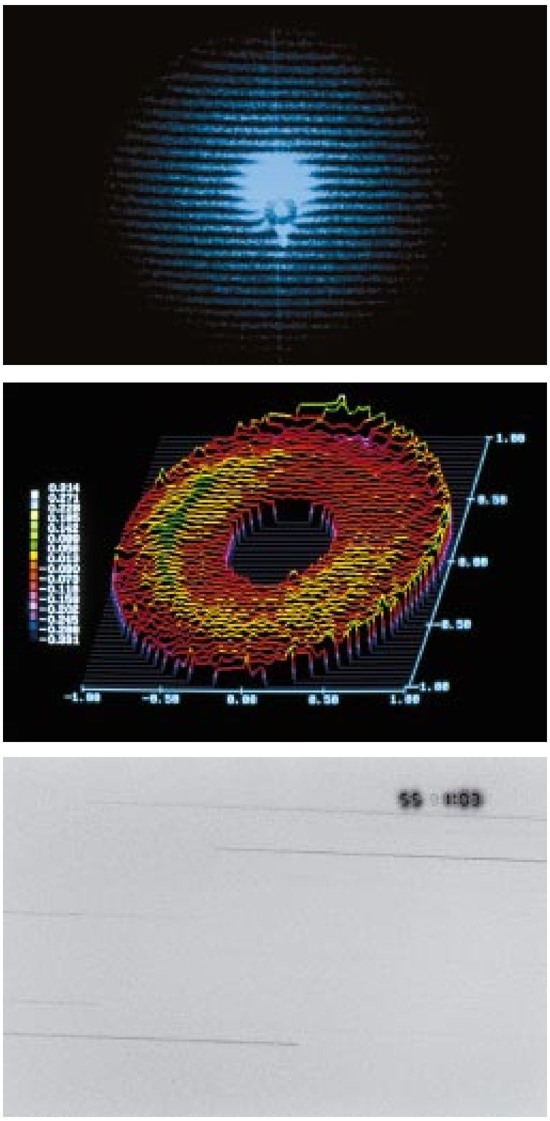
Fluid mirrors can also serve as inexpensive reference surfaces in optical-shop tests because their parabolic shape is so accurate. In the most original application of liquid mirrors so far, Nathalie Ninane of the Space Center in Liège is making a hologram of a 1.4-meter mercury mirror. When the hologram is reilluminated, it will generate an image that can be used as a guide for polishing the surface of a glass mirror into a perfect parabola.

The prospect of building mirrors with gigantic diameters, perhaps larger than 30 meters, motivated me initially. How large can we really make them? This we will know only by trying to build mirrors of increasing breadth. But the relative ease we experienced making a 2.7-meter mirror at low cost is a good omen. Still, several factors may eventually limit size.

The curvature of the earth introduces a minor focus change, but this can easily be corrected. There is one worrisome geophysical factor, however—the Coriolis effect, which gives air masses moving across the globe a spiral appearance. This effect occurs when an object, such as the surface of a rotating liquid, moves on a rotating frame of reference, such as the earth. Hickson and Brad K. Gibson, also at the University of British Columbia, and, independently, I have done calculations showing that this problem should not be serious. All the same, our early observations indicated that wind does create significant disturbances. Although a shelter can isolate a liquid mirror from outside drafts, the wind generated by its own rotation will probably set the ultimate size limit. The outer regions of a large mirror move faster than those of a smaller mirror—much as on a merry-go-round a horse near the edge moves faster than one near the center. This greater speed will generate more local air turbulence.

Fortunately, there are several tacks we can take to minimize surface swash: the use of thin mercury layers damps such disturbances quite well. Floating a single layer of organic molecules, such as oils, above the mercury can also decrease wind-driven waves. A guaranteed solution involves placing a plastic film above the rotating liquid. After testing a variety of materials, we found that thin, tough Mylar films protect the mirror's surface and do not distort light.

Other improvements can be made. Because mercury is heavy, replacing it with a lighter liquid would prove advantageous: the lighter mirror could then rest on a less expensive bearing and container. We have therefore begun experiments on gallium, a lighter liquid metal than mercury. The results have been encouraging. Although gallium solidifies at 30 degrees Celsius, the material is easy to supercool. This process keeps a substance in a liquid state below the temperature at which it normally solidifies. Two Laval graduate students, John Gauvin and Gilberto Moretto, have supercooled gallium samples to –30 degrees C by exposing them to the cold Quebec winter. They verified that the samples remained stable as a liquid for several weeks, solidifying only after the temperature dropped below –30 degrees C. Gauvin made a gallium alloy mirror, 0.5 meter in diameter, which performed reasonably well.



TEST RESULTS have shown that liquid mirrors provide superb optical surfaces. The interferogram (*top*) was obtained from a liquid mirror 2.5 meters in diameter. Computer analysis of this recording gives a false-color rendering that shows the surface contour of the 2.5-meter liquid mirror (*middle*). The stellar trails (*bottom*) were observed in 1987 using a 1.2-meter liquid mirror and a photographic camera.

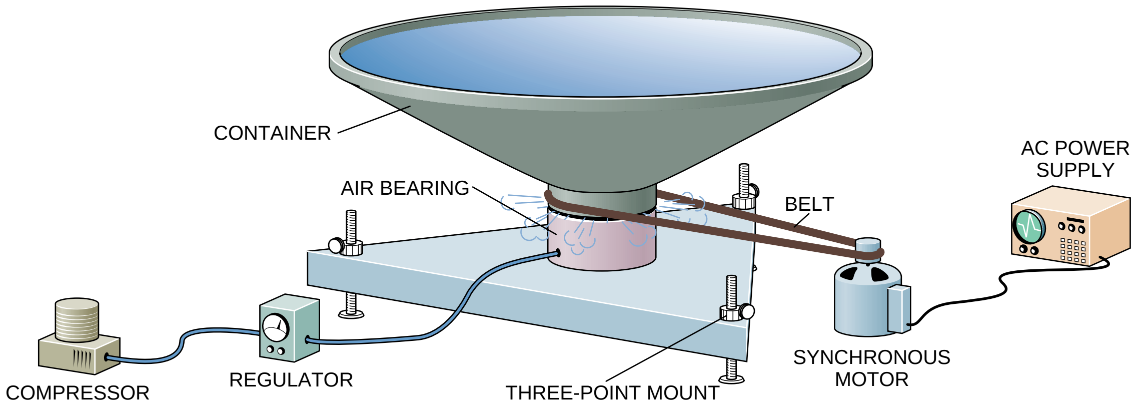
Improvements in optical correcting devices will also make liquid-mirror telescopes more useful by increasing the region of sky from which they can gather light. In reflections from a parabolic mirror, only those distant, pointlike objects that are placed directly over the parabola's center are focused to a single point. Images of those points that lie off this central axis appear as spots that grow larger in relation to their distance from the axis, giving blurred pictures. These distortions, which occur in any telescope, are minimized by auxiliary optical devices. Such apparatuses consist of coordinated lenses or mirrors that pull straying beams of reflected light in line. In doing so, they eliminate many errors from the final image. A typical arrangement of these guides allows for true pictures over a one-degree field, or twice the apparent diameter of the moon. By placing telescopes at various latitudes in order to sample different regions of the sky, this viewing region can be broadened. It would be more efficient and far less costly, however, to develop improved correcting devices so that one telescope might have a larger field of view.

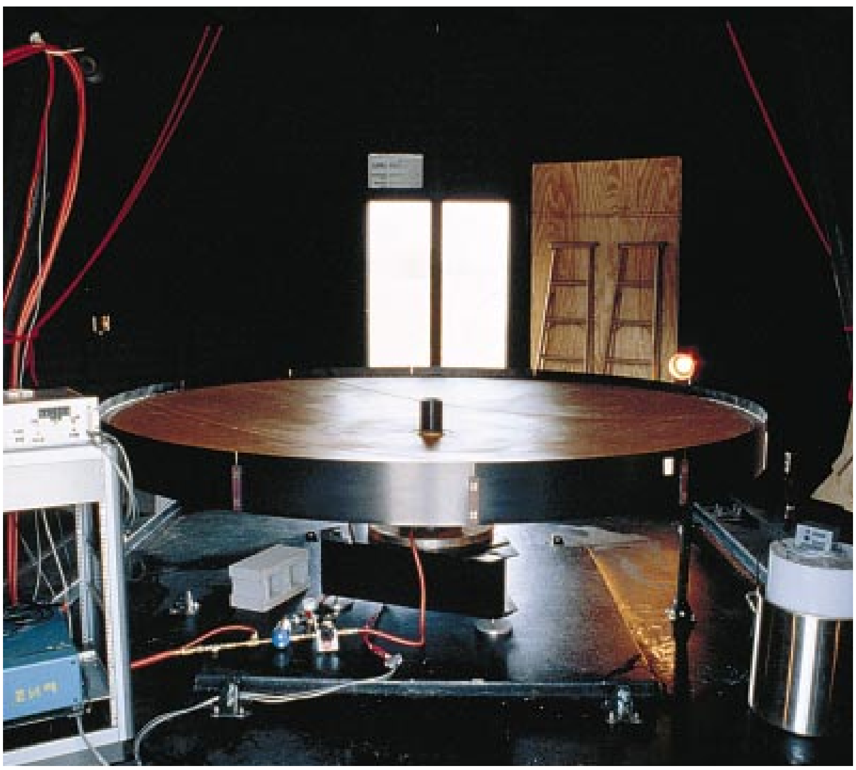
For this reason, Harvey R. Richardson of the University of Victoria and Christopher L. Morbey of the Herzberg Institute of Astrophysics in Victoria have designed a corrector on a computer to counteract the specific errors occurring in liquid-mirror reflections. The instrument is somewhat cumbersome because it relies on choreographing the movement of three individual mirrors. Still, this corrector enables liquid mirrors to produce excellent images of light traveling toward the mir-

Liquid-Mirror Telescopes at Work

A mirror, three meters in diameter, made recently by the National Aeronautics and Space Administration (*bottom left*) will search for space debris. The diagram (*top*) shows how the mirror works. The mirror and bearing rest on a three-point mount; the axis of rotation is aligned vertically using two adjustable screws. A synchronous electric motor, connected by pulleys and a belt, drives the turntable under the mirror's container. An AC power supply, stabilized with a crystal oscillator, controls the motor.

A solid parabola made from a polyurethane resin is spin-cast over the basic container. After this plastic parabola has hardened, liquid mercury can be poured into the basin. Such simple designs make liquid mirrors relatively affordable and easy to construct. The rudimentary observatory (*bottom right*), built at Laval University in 1987, supported the first astronomical research performed using a liquid mirror. A mirror 2.7 meters in diameter would cost approximately U.S.$15,000 to assemble.





ror at an angle 7.5 degrees away from the central axis. Indeed, this pioneering work shows that it is possible to correct the severe aberrations introduced when light is reflected at a sharp angle from a fluid parabolic surface.

Optimized correctors face few limitations in theory. They should give liquid-mirror telescopes access to much of the visible sky, albeit over narrow fields of view, valuable for spectroscopy or very high resolution imagery. On the practical side, Min Wang, Gilberto Moretto and I, in collaboration with Gerard Lemaître of the Marseilles Observatory, are exploring novel adaptations of conventional corrective mirrors. This optical technology, pioneered by Lemaître, is based on warping mirrors into complex shapes capable of removing reflective errors. Recently Wang, Moretto and I have designed on a computer such a high-performance corrector that uses two auxiliary mirrors. It can give excellent images in regions of the sky located as much as 22.5 degrees away from the central axis.

LIQUID MIRROR, tended by graduate student Luc Girard, acts as the receiver for an atmospheric monitor. The 2.7-meter mirror is at the University of Western Ontario.

A holographic device could at least in theory serve as the perfect mediator, reconciling the differences between reflected light and its original source. A prerecorded hologram could be cast in the path of reflected light. As light passed through the hologram, it would filter out predictable errors. Mosaics of computer-generated holograms could compensate for aberrations that occur as light travels great distances from the zenith, over large fields of view. At Laval, Guylain Lemelin, Roger A. Lessard and I are exploring the promise of holographic correctors. Unfortunately, we find that practical versions of these instruments must await technological advances that we are only beginning to glimpse.

I am often asked if a liquid-mirror telescope could be placed in space. The possibility is intriguing because liquid mirrors offer excellent optical qualities, low masses and simple packaging. Despite inhospitable temperatures, the moon could certainly host a liquid-mirror telescope. Mirrors made from light gallium alloys, or perhaps even lighter alkali alloys, would remain liquid in lunar telescopes because such alloys have low melting temperatures. Until recently, though, I thought placing a liquid-mirror telescope in orbit would be impossible. Gravity supplies the necessary acceleration to form a parabola on the earth or the moon; an orbiting telescope free-falls and thus is not affected by gravity. Using an engine for acceleration would be impractical because the engine would eventually run out of fuel.

The potential of solar sail–powered crafts has changed my mind. In 1992, I published an article in the *Astrophysical Journal* that examines the plausibility of using solar sails to propel a liquid-mirror telescope in orbit. Our sun supplies an inexhaustible source of energy that a solar sail could harness to accelerate a liquid surface into a parabola. Although this concept may appear to belong more to science fiction than to science, it rests on reasonable assumptions. No solar sail–powered craft has ever been launched successfully, but a study carried out by NASA in the late 1970s showed that solar sail–powered crafts are feasible.

A mirror accelerated by solar sails would not gain momentum and leave the solar system, provided the vessel traveled more slowly than its proper orbital speed. The solar sail would then counteract just enough gravitational pull to keep the telescope in orbit. It could possibly replace all gravitation, yielding a stationary instrument capable of long integration times. Colin McInnes of the University of Glasgow has shown that crafts rigged to solar sails could navigate through a rich variety of paths and switch orbits midcourse. In this case, an orbiting liquid-mirror telescope could be pointed like a conventional, glass-mirror telescope. I have fearlessly considered placing mirrors having diameters as vast as one kilometer in space. It is staggering to imagine what scientists could discover with such gigantic mirrors.

In the June 1987 issue of *Physics Today,* associate editor Per H. Andersen wrote a news note entitled: "Will Future Astronomers Observe with Liquid Mirrors?" Now, nearly seven years later, a handful of liquid-mirror telescopes have been built for research purposes. The next question is how many astronomers will make observations using liquid-mirror telescopes. Only the future knows, but at the very least, I expect that liquid-mirror telescopes will perform specialized astronomical tasks such as surveys. At the other extreme, I dream that one day liquid-mirror telescopes will support most astronomical research, relegating tiltable telescopes to specialized niches. That may seem preposterous, but I believe the simplicity and cost of liquid-mirror telescopes will be persuasive. In the meantime, amateur astronomers, engineers and all scientists alike should be aware of the capabilities of liquid mirrors. They may yet reveal greater wonders than did Lewis Carroll's looking glass.

FURTHER READING

LIQUID MIRROR TELESCOPES: HISTORY. B. K. Gibson in *Journal of the Royal Astronomical Society of Canada,* Vol. 85, No. 4, pages 158–171; August 1991.

THE CASE FOR LIQUID MIRRORS IN ORBITING TELESCOPES. E. F. Borra in *Astrophysical Journal,* Vol. 392, No. 1, pages 375–383; June 10, 1992.

LIQUID MIRRORS: OPTICAL SHOP TESTS AND CONTRIBUTIONS TO THE TECHNOLOGY. E. F. Borra, R. Content, L. Girard, S. Szapiel, L. M. Tremblay and E. Boily in *Astrophysical Journal,* Vol. 393, No. 2, pages 829–847; July 10, 1992.

AIDS and the Use of Injected Drugs

The AIDS epidemic continues to grow among drug users who inject. It could be curbed if governments more readily adopted effective prevention programs

by Don C. Des Jarlais and Samuel R. Friedman

The tiny blood transfusions that take place when drug users share needles or syringes serve to transmit disease efficiently. Nothing has illustrated this fact more tragically than the rapid spread of HIV, the human immunodeficiency virus and the causative agent of AIDS. More than 50 countries have documented HIV infection among persons who inject illicit drugs. An additional 30 countries have reported the practice of illegal drug injection; as a result, these places have populations at high risk of contracting or transmitting the AIDS virus. Once HIV is established among such a group, it moves into the community at large through sexual contact and pregnancy.

No hard figures exist for the number of these drug users throughout the world; the best estimate is about five million people. There are, however, hard figures for the impact of AIDS on this group in the U.S.: more than one third of all AIDS cases, or about 113,000, are associated with the injection of illicit substances. Even in the absence of definitive numbers, the proliferation of HIV among these users has clearly become a medical catastrophe internationally as well. The most conservative predictions by public health institutions foresee increased use of illegal drugs via injection in many countries and increased transmission of HIV among people who take drugs in this way.

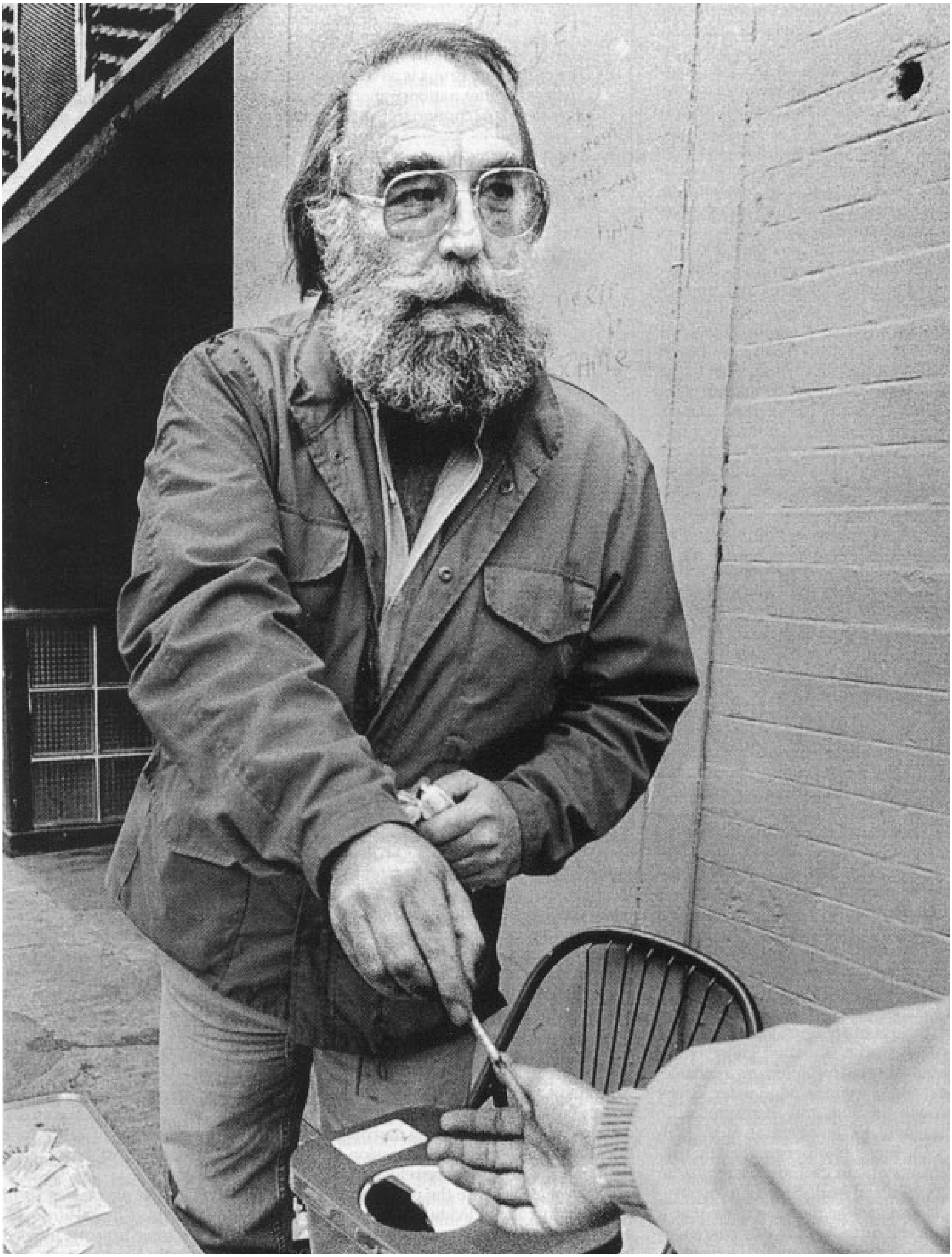
Fortunately, the efforts of some health care providers in North America, Europe and Australia show that the expansion of the AIDS epidemic among drug injectors can actually be curtailed. But a lack of political commitment often renders these valuable insights moot. Instead of designing policies that could effectively reduce the spread of HIV, many public health officials have desperately tried to avoid the controversy that surrounds drug use.

We hope that knowledge of the encouraging results of both pilot and full-scale programs may lead officials to strengthen policies to deal with this aspect of the AIDS pandemic. In this article, we review an array of strategies that should be more widely pursued and elaborated before the epidemic becomes uncontrollable among drug injectors, their sexual partners and their children. (The familiar phrase "intravenous drug user" refers only to people who inject into their veins or arteries. Here we use the terms "drug users who inject" and "injecting drug users" to include individuals who may administer the substance into the muscle or just below the skin surface as well.)

Many manifestations of AIDS in drug users who inject are quite different from those in homosexual and bisexual men who do not use drugs in this manner. Among injecting drug users the signs of the disease extend beyond the opportunistic infections and Kaposi's sarcoma that traditionally served as the basis of a diagnosis of AIDS. The difference in diagnosis derives from the fact that the U.S. Centers for Disease Control primarily tracked the epidemic in homosexual and bisexual men in developing the original definitions of AIDS.

Studies conducted by our research group and by Peter A. Selwyn and Gerald Friedland of Yale–New Haven Hospital and their colleagues at Montefiore Medical Center in New York City have shown that in drug users who inject and in their sexual partners HIV infection is associated with substantially increased morbidity and mortality from bacterial infections. Patients suffer from diseases such as pneumonia and endocarditis (an inflammation of the inner membrane of the heart). In the U.S., much of the resurgence of tuberculosis is occurring among HIV-infected drug users who live in crowded conditions without access to good medical care. There are also indications that

DON C. DES JARLAIS and SAMUEL R. FRIEDMAN study drug addiction, drug abuse treatment and human immunodeficiency virus infection among drug users. Des Jarlais is director of research for the Chemical Dependency Institute of Beth Israel Medical Center and senior research fellow at the National Development and Research Institutes, Inc. He also serves as a member of the U.S. National Commission on AIDS. Friedman is a senior principal investigator at the National Development and Research Institutes, Inc., and has worked as a consultant for the National Institute on Drug Abuse and for the World Health Organization.



NEEDLE EXCHANGE PROGRAMS, such as this one being carried out in Tacoma, Wash. (*above*), seek to curb the spread of AIDS among individuals who inject illicit drugs and among their sexual partners. Public health workers try to prevent drug users from sharing needles contaminated with the human immunodeficiency virus (HIV). These programs have been controversial at times, and demonstrators have often rallied to encourage their funding and legality (*opposite page*).

HIV infection increases the severity and frequency of gynecological illnesses, including cervical cancer and pelvic inflammatory disease. In the U.S., more than 80 percent of HIV-positive women inject drugs or are the partners of people who do.

The CDC definition of AIDS has been periodically updated to incorporate these findings. The 1993 revision includes a diagnosis of AIDS based solely on severe immunosuppression—that is, a CD4 cell count of less than 200 cells per microliter of blood. (CD4 cells are a form of white blood cell and, normally, help the body protect itself against infection.) This criterion provides an alternative to diagnosing AIDS on the basis of a continually changing list of opportunistic infections and malignancies.

There appear to be two predominant routes by which HIV is introduced into populations of drug users. The virus can enter through overlapping, or bridge, groups—for example, men who inject drugs and also have sex with men. This form of transmission appears to have occurred in New York City, Rio de Janeiro and Sydney. The second is travel. Contrary to the stereotypic view of injecting drug users, many travel frequently, for business and pleasure. For instance, a recent study found that 62 percent of drug users from Berlin had injected drugs outside their home city in the past two years; 14 percent of users from New York City had injected outside the area in the past two years.

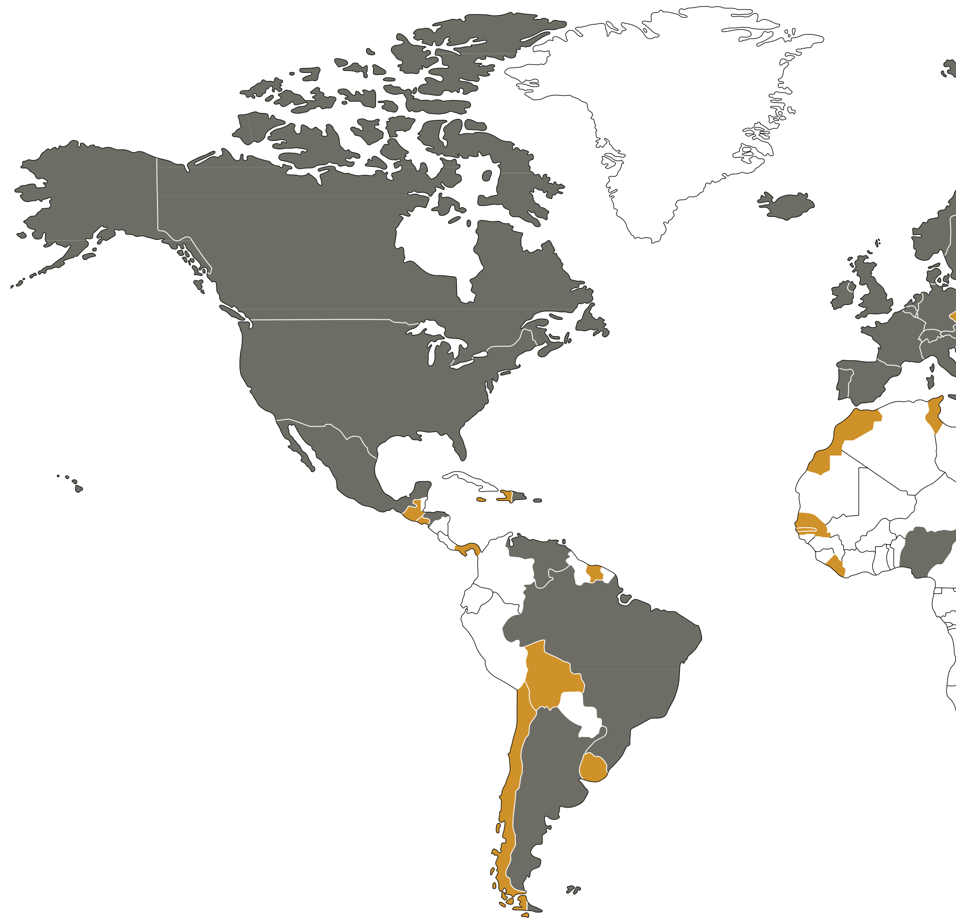
Once HIV has been introduced into a community of drug users, it can diffuse extremely quickly. Rapid increases in HIV seroprevalence, or the percent of drug users who are infected, have been observed in both developed and developing nations. In cities such as Bangkok and Edinburgh and in the Indian state of Manipur, HIV seroprevalence among drug users has risen from 0 to 40 percent within two years or less.

This fast transmission is frequently associated with certain kinds of behavior: the use of shooting galleries and of dealers' works. Shooting galleries are places where drug injectors rent needles and syringes, administer drugs through the skin and then return the equipment to the operator of the gallery for rental to other customers. Payment may be in cash or in drugs. The term "dealers' works" refers to the needles, syringes or paraphernalia that a seller will often lend to one customer, take back and then lend to the next.

Shooting galleries and dealers' equipment serve several functions within the drug subculture. They provide a source of needles and syringes, which are often scarce because of legal restrictions.

Prevalence of Injecting Drug

Use of injected drugs is associated with HIV infection in many countries (*black*), whereas other nations report illicit drug injection that is not currently linked to AIDS (*yellow*). According to the World Health Organization, some 15 million people



In the U.S., for instance, most states that have large populations of such drug users require prescriptions for the sale of needles and syringes. In many other states and countries, over-the-counter sales are permitted, but pharmacists are often unwilling to sell to people they suspect of using illegal drugs.

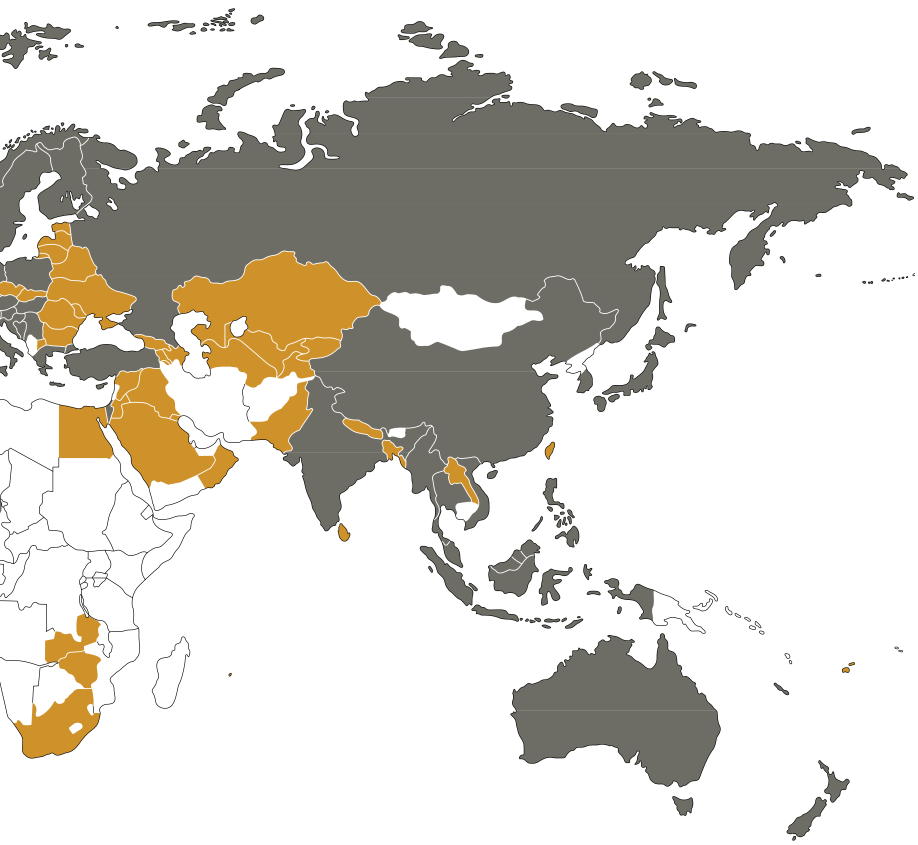
Shooting galleries allow drug users to avoid toting around equipment. In almost all parts of the U.S., it is illegal for people to possess equipment for administering narcotics. In countries where such possession is legal, many drug users are reluctant to carry syringes because the police often conclude that the individual is engaged in illegal activities. Keeping needles at home circumvents some of these difficulties, but the practice can provoke confrontations between the drug user and the people he or she lives with.

At the same time that dealers' works and shooting galleries have served a utilitarian purpose within the drug-using community, they have also served as vectors for hepatitis and other bloodborne infections. The injecting devices are often shared by dozens of drug users who do not have any social relationship with one another and, therefore, do not know one another's HIV status or about any risky behavior that they may engage in. The materials are frequently used repeatedly until the needle or syringe becomes bent or too dull or clogged with dried blood and is discarded. With the advent of AIDS, such paraphernalia and places have taken on a deadly aspect.

The reasons for the pervasiveness of injection as a form of illegal drug administration is a topic urgently in need of further research. Nevertheless, several features of its appeal are clear. During the first half of the 20th century, the consumption of illicit drugs through injection was relatively rare outside the U.S. But as with many parts

Users and HIV Infection

are infected with the human immunodeficiency virus, but it is unclear how many of them were infected through injection or through sex with a drug user. In the U.S., however, one third of all AIDS cases can be attributed to drug injection.



of American culture—from T-shirts and blue jeans to fast food—many different countries have adopted the practice.

The fact that cocaine, heroin and related drugs are illegal encourages the use of injection. Severe statutory restrictions greatly increase the cost of illicit substances to nonmedical users. Injecting provides a way to economize. Injectable forms of opiates and coca are much more concentrated than traditional forms, such as opium or coca tea. Injection provides an intense and economical effect by maximizing the amount of drug that reaches the brain. People who sniff or smoke drugs say that if they inject they need only one third of the amount of the drug to maintain a habit.

Because the injectable forms of illicit drugs are concentrated, they are relatively small in volume and easier to ship. Indeed, the movement of HIV can be traced along some drug distribution routes. Once established in the Golden Triangle of Southeast Asia—the region where Thailand, Myanmar and Laos meet—HIV radiated outward by traveling west into northeast India, south into Thailand, Malaysia and Vietnam and east across southern China toward Hong Kong. The virus has also moved along cocaine routes within Brazil.

When AIDS was first detected among drug users who inject, there was tremendous skepticism about the ability of these individuals to change their behavior. This assumption was based on another stereotype: drug users are not at all concerned about their health. Fortunately, this attitude does not hold true for a great many of them.

In some cities, such as New York, drug users altered their behavior before any formal prevention program was implemented. Based on information obtained through the mass media and through the informal communication networks of the drug subculture, they started to use illicitly obtained sterilized needles: AIDS risk reduction was reported by 80 percent of drug users in New York City, 84 percent in Glasgow, 82 percent in Lund, Sweden, 84 percent in Sydney and 73 percent in Tacoma, Wash. Anecdotal reports found similar attempts among laborers in cities and villages throughout Asia. Although these findings make it clear that drug users who inject often seek to reduce risk, their success depends on whether they can obtain clean needles.

Prevention programs have also demonstrated that it is quite possible to encourage behavioral changes on the part of drug users. Yet these services have not always received the political attention, or funding, they deserve because it is difficult to quantify their effectiveness. The lack of control groups creates measurement problems. In addition, it is hard to evaluate behavioral changes brought about by prevention programs because different studies use different units to measure such shifts in behavior. Furthermore, actual HIV transmission has only seldom been used as an end point in these studies.

Despite these methodological problems, we have found it possible to make a set of generalizations with respect to developing effective AIDS prevention programs. First, drug users who inject are quite capable of learning the basic facts about HIV transmission and about how to reduce exposure to the virus—indeed, one study found that 93 percent of the users of a particular methadone maintenance center knew about AIDS, and 59 percent had sought to protect themselves against infection. Second, programs are more effective if they involve drug users in design and implementation. Third, programs should provide the means for behavioral change, including treatment for users wishing to limit or stop their use of drugs as well as access to sterile injection equipment and condoms. Fourth, people should be given choices as to how they want to reduce the chances of HIV transmission: there is no one solution that works for everyone. Finally, it is apparent that to date no single program and no combination of programs have been able to eliminate high-risk behavior completely.

One point that emerges in an analysis of the programs is that their success appears to be a function of when they were implemented—that is, whether they were established when the HIV infection rate was minimal or after HIV was entrenched in the community. We have recently analyzed the relation between AIDS prevention and risk behavior in areas where HIV prevalence among drug users is low. Holly Hagan of the Tacoma-Pierce County Health Department, David Goldberg of the Communicable Diseases Unit of Ruchill Hospital in Glasgow, Kerstin Tunving of the University of Lund and Alex Wodak of St. Vincent's Hospital in Sydney and their many colleagues provided information for our review. In each of these cities, HIV had been introduced into the population of drug users. Yet seroprevalence among heterosexual users remained low (less than 5 percent) and stable for at least four years. In each location, AIDS prevention activities were limited, but all included community outreach and access to sterile injection equipment through syringe exchanges and over-the-counter pharmacy sales.

Eliminating risky behavior was not necessary in order to keep HIV infection rates low in these four cities. Approximately half of the drug users whom we interviewed reported that they were at least occasionally injecting with needles and syringes that had been used by others. Most of the unsafe practices appeared to be contained within small groups of friends. Low and stable rates of HIV infection have also been reported in Athens, Toronto, Seattle, Perth and Melbourne. Thus, the experience of the four cities studied may well be generalized to many other places where prevention efforts were begun early.

In addition to cities where HIV infection rates were kept low, several teams of researchers have observed declining rates of new HIV infections and the stabilization of HIV seroprevalence in cities with high seroprevalence. These findings come from John K. Watters of the University of California at San Francisco, Kachit Choopanya of the Bangkok Metropolitan Administration, Roeland A. Coutinho of the Municipal Health Service of Amsterdam, Alfredo Nicolosi of the National Research Council in Milan and our group in New York. We found, for instance, that HIV seroprevalence in New York City has remained stable since 1983. Stabilization, however, does not imply a lack of new HIV infections. Rather stabilization occurs when there is a rough balance between new infections and the loss of HIV-infected drug users through death or the discontinuance of drug injection. It is not an ideal goal, but it does limit the extent of an epidemic.

Whereas the studies suggest that it is possible to prevent or stabilize epidemics among such drug users, reversing a well-established HIV epidemic is a much more formidable task. The situation in Amsterdam offers a good example of the difficulties of fighting an entrenched HIV epidemic. HIV seroprevalence among drug users in Amsterdam had reached approximately 30 percent by the mid-1980s.

Percent of Drug Users with HIV Infection*

	1978	'79	'80	'81	'82	'83	'84	'85	'86	'87	'88	'89	'90	'91
AMSTERDAM									33	31	31	34		
BANGKOK										1	43			
BERLIN								31	49	49	45			
BILBAO							42	42		44	45	27	30	
BOLOGNA		0	0	9	7.5	39	37							
DETROIT								13		16		16		
EDINBURGH					0	14	42	37						
GENEVA				7		27		52						
HAMBURG								0	23	16	13			
LONDON								5	6	4	5			
MANIPUR									0	0	0		54	
MILAN				11	28	61	67	69	73					
NEW YORK CITY	9	26	38		50		57		55				50	
PADUA						20	28	65	50					
RIO DE JANEIRO									4				20	37
ROME								34	42	33	31	32		
SAN FRANCISCO									7	12	12	11		
SARDINIA	0	0	0	1	10	18	32	43	57					
TOURS					0	0	15	17						
VIENNA									7	12	30	28	24	

*Figures come from myriad sources, including hospital records, stored blood samples and treatment programs. The information is incomplete because studies were not undertaken every year.

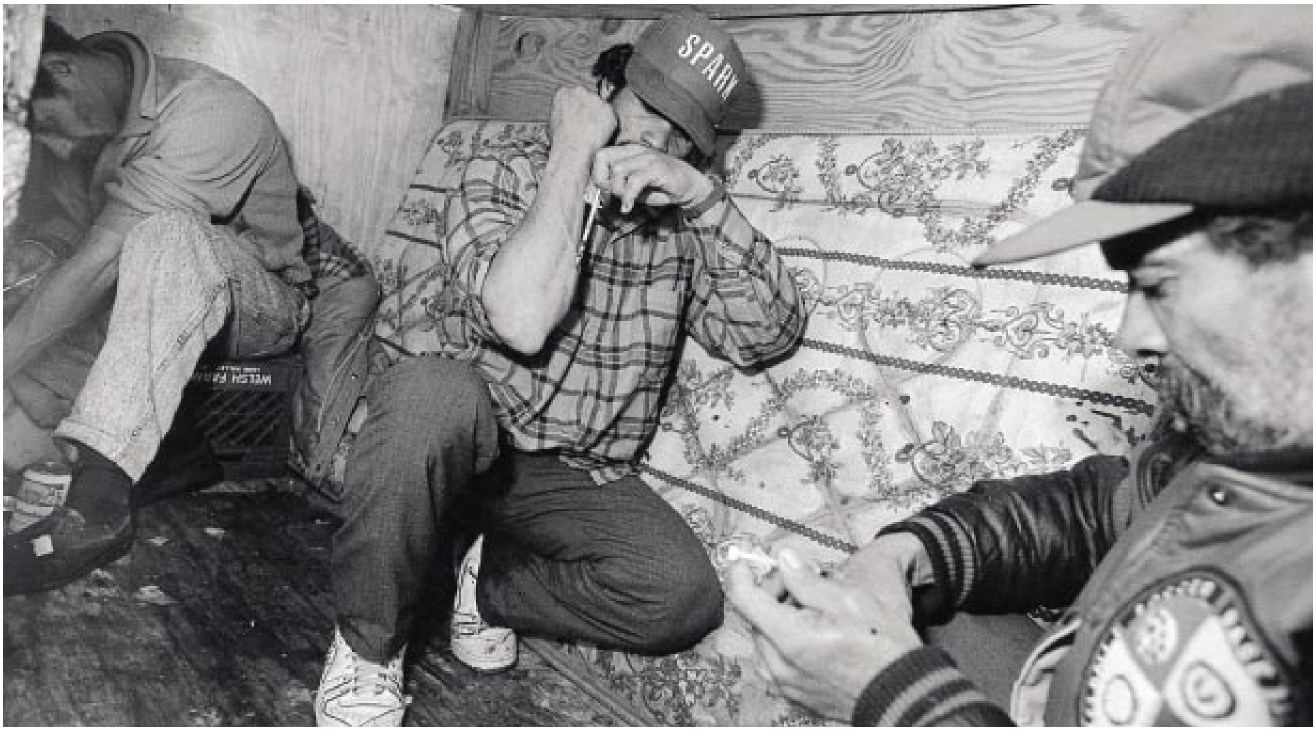
Public health officials quickly mounted a large AIDS prevention effort. This activity included enlisting and funding the local drug users' group to do prevention work, syringe exchange, over-the-counter sales of needles and syringes and drug abuse treatment. (It is necessary to keep in mind one important criticism of the Amsterdam AIDS prevention program: many of the methadone maintenance clinics give methadone at dosages that are too low to alleviate the symptoms of drug withdrawal or craving.)

Coutinho and his colleagues found that the AIDS prevention programs in Amsterdam led to a reduction in the frequency of unsafe injections. The programs also brought about significant but smaller reductions in unsafe sexual activities. In addition, the rate of new HIV infections among drug users who inject declined from a peak of approximately eight per 100 person-years at risk during the early 1980s to a current rate of approximately four per 100 person-years at risk. This reduction in the rate of new infections is impressive, but incidence is still much too high.

A high prevalence of HIV infection among users who inject drugs increases the risk for everyone in a community of drug users. An individual who is usually careful is more likely to be infected if, for example, he or she forgets and injects with a needle that has been used by someone else. Sexual activity also becomes more dangerous, even though sex is less effective in transmitting HIV than is injection.

If addiction can be reduced by treatment programs, and if individuals can be discouraged from starting to inject drugs, the number of AIDS cases might be reduced even in areas where seroprevalence is high. The city of Edinburgh may have made the most progress in both these respects. The Community Drug Prevention Project there works with local general practitioners to administer drug abuse treatment—mostly in the form of methadone. Recent research by Sally J. Haw of City Hospital and her colleagues has shown that a large percentage of addicts who were previously injecting drugs are now in treatment. Many others are maintaining themselves on illicitly obtained orally administered narcotics.

Strategies for preventing individuals from starting to inject drugs have not been well studied. Indeed, the initiation into injecting may be the least investigated aspect of halting HIV infection among drug users. Estimating the rate at which new users start to inject illicit



SHOOTING GALLERIES have facilitated the AIDS epidemic among people who inject illegal drugs. These sites, where drug users can inject heroin or cocaine, often provide the same syringes and needles to many individuals. These users are thus at great risk of transmitting or contracting HIV, which is effectively spread through even tiny transfusions.

drugs is an extremely difficult research task. One point is clear from the limited evidence available: knowledge of and concern about AIDS is not enough to deter persons from starting or returning to injection. In one study we conducted, people who were sniffing heroin and cocaine were given extensive AIDS education, including HIV counseling and testing. During a nine-month follow-up period, one quarter of the group injected drugs. Most of these individuals had a close relationship with a drug user who injected.

Despite the knowledge gathered from the studies described above as well as from other research efforts, treatment and prevention programs for drug users have not been widely implemented. Three obstacles stand in the way. In some developing countries, the lack of economic resources limits such prevention. Providing sterile injection equipment to drug users in societies that cannot make sterile equipment available to clinics and hospitals is not likely. Household bleach has been tried as a method for disinfecting, but the most recent studies fail to show any protective effect. Some of the difficulties may arise because needles and syringes need a relatively long contact time to be disinfected—that is, two rinsings, each for 30 seconds or longer.

There are also regions in which greater knowledge of the dangers of drug injection and sexual behavior is needed for prevention of transmission. Present programs appear adequate for controlling HIV transmission in low seroprevalence areas, but a new generation of risk-reduction plans may be needed to reverse the situation in locations where seroprevalence is high. Among other components, these services would need to make sterile needles available, deter people from sharing injection equipment, offer drug abuse treatment and reduce the number of people who start to inject drugs. Encouraging the use of noninjectable forms of these drugs has also been suggested as a way to prevent the transmission of HIV. But workable strategies for implementing this idea have not yet been developed.

The biggest obstacle to reducing the spread of HIV among drug users is neither a lack of resources nor a lack of knowledge but a lack of political resolve to utilize already existing information. Many of the strategies for preventing HIV infection have reasonable evidence for their effectiveness but remain highly politically charged. In Europe, programs that offer maintenance treatment have remained contested, and that approach is not widely used. In some cities, such as Amsterdam, even when methadone maintenance is permitted, it is provided at dosages that are too low to be effective and for periods that are too short. In other places, a user must already be infected with HIV before he or she is permitted to receive methadone maintenance treatment.

In the U.S., and to a lesser extent in Sweden, syringe exchanges and over-the-counter sales of sterile injection equipment have remained very controversial. Opponents of legal access to such equipment have claimed that these programs would promote illicit drug injection and would be taken as an official condoning of the behavior. Yet no study has found any evidence to support this contention. In the U.S., the debate over syringe exchange has been complicated by our historical tendency to stigmatize certain ethnic groups for their illicit drug use, adding racial hostility to the debate over AIDS prevention [see "Opium, Cocaine and Marijuana in American History," by David F. Musto; SCIENTIFIC AMERICAN, July 1991].

Some of the opposition to syringe exchange, over-the-counter sales of injection equipment and other less dangerous injection programs does not consider research or scientific evidence. In Sweden the Parliament rejected the findings of an evaluation study that called for the expansion of the existing syringe exchange plans into a national system. In the U.S., federal legislation prohibited funding of similar exchange services unless the surgeon general

Some Successful Treatment Programs

COUNSELING AND TESTING

NANTES, FRANCE
MILAN
STOCKHOLM
CONNECTICUT
NEW YORK CITY
SAN FRANCISCO
OTHER CITIES IN U.S.

EDUCATION AND INFORMATION CAMPAIGNS

CHICAGO
NEW JERSEY
BRISBANE, AUSTRALIA
GERMANY
PADUA
ROME
MEXICO
MADRID

INCREASED ILLEGAL SALE OF SYRINGES

NEW YORK CITY

LEGAL SALE OF SYRINGES

INNSBRUCK
PARIS
GLASGOW

MEDIA COVERAGE AND SOCIAL SUPPORT

NEW YORK CITY
OTHER CITIES IN U.S.

METHADONE TREATMENT

ITALY
AMSTERDAM
STOCKHOLM
NEW YORK CITY
PHILADELPHIA
BALTIMORE
SAN FRANCISCO

OUTREACH AND BLEACH DISTRIBUTION

CHICAGO
SACRAMENTO
SAN FRANCISCO
OTHER CITIES IN U.S.

OUTREACH AND INCREASED TREATMENT

NEW JERSEY
NEW YORK CITY
OTHER CITIES IN U.S.

SYRINGE EXCHANGE

AUSTRALIA
AMSTERDAM
LUND, SWEDEN
LONDON
SAN FRANCISCO
TACOMA
OTHER CITIES IN U.K.

METHADONE BUS and health workers in Amsterdam make treatment available to drug users, thereby reducing the use of illicit drugs and, consequently, the risk of spreading HIV through injection.

found such practices to be safe and effective. But until 1992, the federal government refused to fund research to examine the safety and effectiveness of the locally funded efforts.

These counterproductive attitudes are slowly changing. The debate about methods to prevent AIDS among drug users has led to a general rethinking of how best to regulate the use of licit and illicit drugs. An approach called harm reduction has developed from work done in the Netherlands, the U.K. and Australia. The harm-reduction perspective pragmatically acknowledges the difficulties of ending all misuse of psychoactive drugs. Nevertheless, it emphasizes the possibilities for reducing the individual and social harm associated with drug use.

Harm-reduction programs consider it possible to lower the chance of HIV infection among persons who continue to take drugs—just as it is possible to reduce the chances of automobile accidents among persons who continue to drink alcohol. Harm-reduction programs are particularly compatible with scientific research on behavior stemming from drug use because such data are critical to determining which social and public health policies work.

HIV infection among drug users who inject, their sexual partners and their children is now a global public health problem [see "Understanding the AIDS Pandemic," by Roy M. Anderson and Robert M. May; SCIENTIFIC AMERICAN, May 1992]. It is very likely that the situation will continue to worsen as both drug use and HIV infection among users extends to more developing countries. Present risk-reduction strategies offer a good possibility for preventing epidemics in places where HIV prevalence is low and where adequate resources for community outreach and sterile injection equipment are available. Actually reducing further HIV transmission among drug users, however, requires a political philosophy that enables officials to implement the findings of current research. So far, and at great social cost, such enlightenment has eluded many governments.

FURTHER READING

HIV-1 INFECTION AMONG INTRAVENOUS DRUG USERS IN MANHATTAN, NEW YORK CITY, FROM 1977 THROUGH 1987. D. C. Des Jarlais et al. in *Journal of the American Medical Association,* Vol. 261, No. 7, pages 1008–1012; February 17, 1989.

HIV AMONG DRUG INJECTORS: THE EPIDEMIC AND THE RESPONSE. S. R. Friedman and D. C. Des Jarlais in *AIDS Care,* Vol. 3, No. 3, pages 239–250; September 1991.

INTERNATIONAL EPIDEMIOLOGY OF HIV AND AIDS AMONG INJECTING DRUG USERS. D. C. Des Jarlais, S. R. Friedman, K. Choopanya, S. Vanichseni and T. P. Ward in *AIDS,* Vol. 6, No. 10, pages 1053–1068; 1992.

HARM REDUCTION: A PUBLIC HEALTH RESPONSE TO THE AIDS EPIDEMIC AMONG INJECTING DRUG USERS. D. C. Des Jarlais, S. R. Friedman and T. P. Ward in *Annual Review of Public Health,* Vol. 14; 1993.

The Terror Birds of South America

These huge, swift creatures were the dominant carnivores of the continent for millions of years, until competitors drove them into extinction

by Larry G. Marshall

It is a summer day on the pampas of central Argentina some five million years ago. A herd of small, horselike mammals are grazing peacefully in the warm sun. None of the animals is aware of the vigilant creature standing 50 meters away in the tall grass. Most of the watcher's trim, feathered body is concealed by the vegetation. Its eyes, set on the sides of a disproportionately large head perched on a long and powerful neck, are fixed on the herd. The head moves from side to side in short, rapid jerks, permitting a fix on the prey without the aid of stereoscopic vision.

Soon the head drops to the level of the grass, and the creature moves forward a few meters, then raises its head again to renew the surveillance. At a distance of 30 meters, the animal is almost ready to attack. In preparation, it lowers its head to a large rock close to its feet, rubbing its deep beak there to sharpen the bladelike edges.

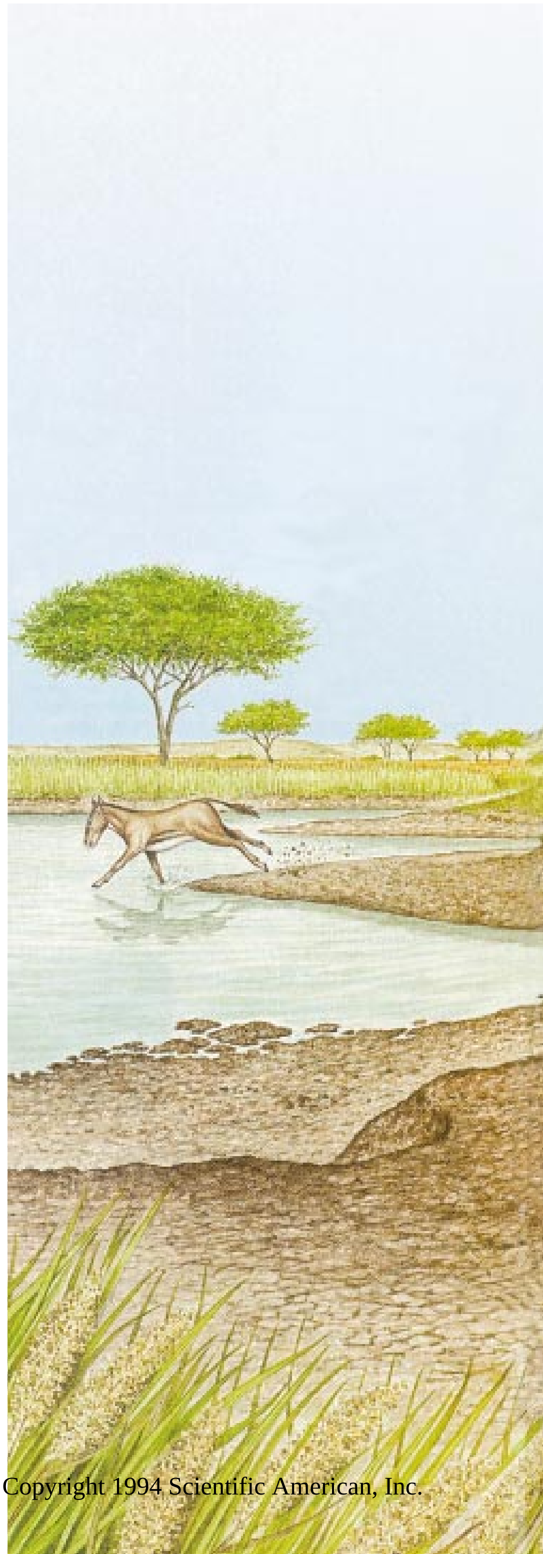
Now the terror bristles its feathers and springs. Propelled by its two long, muscular legs, it dashes toward the herd. Within seconds it is moving at 70 kilometers per hour. Its small wings, useless for flight, are extended to the sides in aid of balance and maneuverability.

The herd, stricken with fright, bolts in disarray as the predator bears down. The attacker fixes its attention on an old male lagging behind the fleeing animals and quickly gains on it. Although the old male runs desperately, the attacker is soon at its side. With a stunning sideswipe of its powerful left foot, it knocks the prey off balance, seizes it in its massive beak and, with swinging motions of its head, beats it on the ground until it is unconscious. Now the attacker can swallow the limp body whole—an easy feat, given the creature's meter-long head and half-meter gape. Content, the gorged predator returns to its round nest of twigs in the grass nearby and resumes the incubation of two eggs the size of basketballs.

LARRY G. MARSHALL is a research scientist at the Institute of Human Origins in Berkeley, Calif. He received his Ph.D. in paleontology in 1976 from the University of California, Berkeley. Since 1974 he has worked on the evolutionary history of South American land mammals. In connection with that work, he has participated in 22 expeditions to Argentina, Bolivia, Brazil, Chile, Colombia, Mexico and Peru. He has collaborated in a number of programs on dating South American rocks and animals and is currently involved in a long-term program to document the relation between the uplift of the Andes mountain range and concurrent changes in the continental faunas.

Meet the terror birds, the most spectacular and formidable group of flightless, flesh-eating birds that ever lived. They are all extinct now, but they were once to the land what sharks are to the seas: engines of destruction and awesome eating machines. In their time, from 62 million years to about 2.5 million years ago, they became the dominant carnivores of South America. The story of their rise and decline is my subject here.

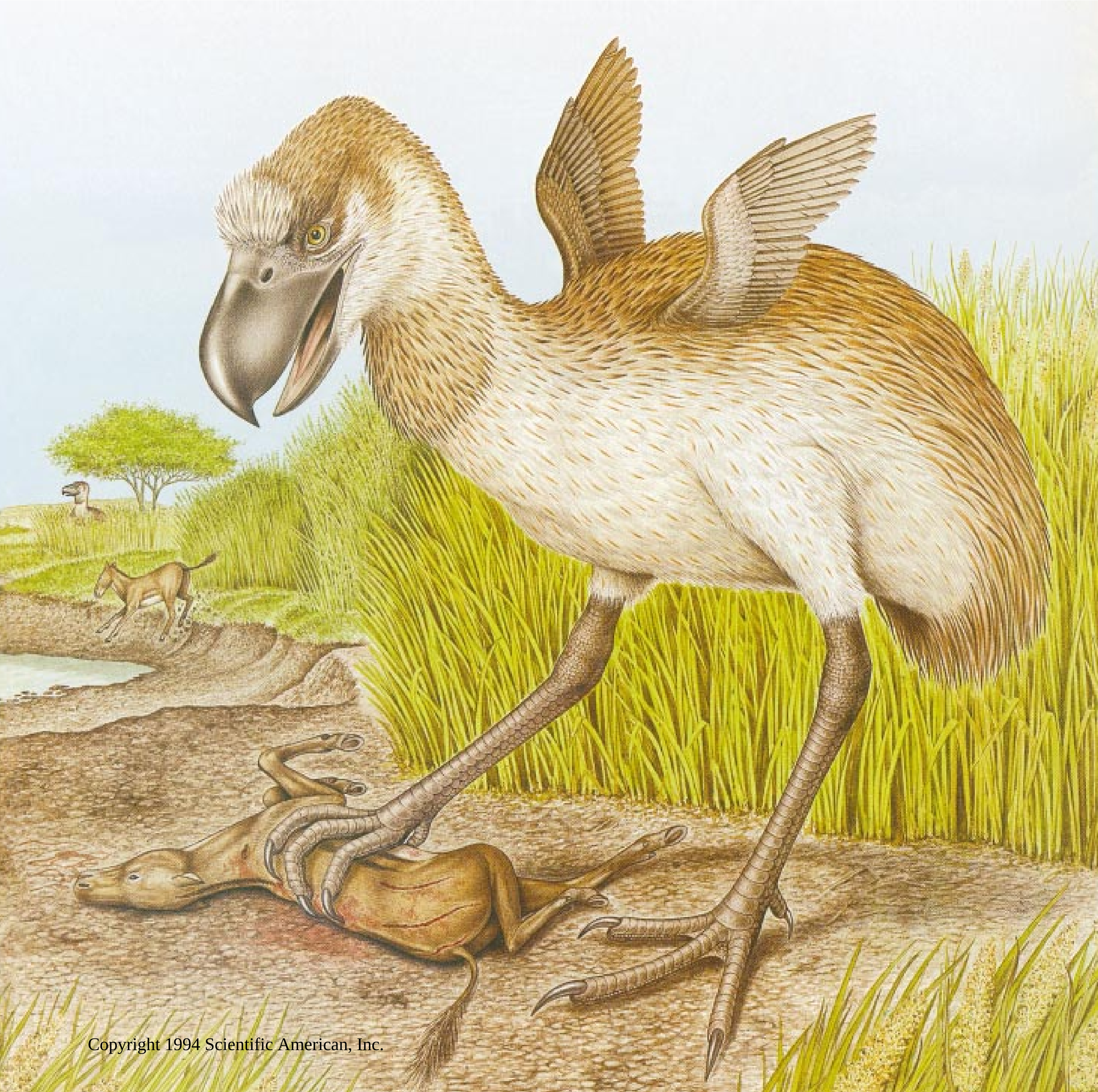
The terror birds are members of a group ornithologists call phorusrhacoids. The first phorusrhacoid to be described scientifically—in 1887 by the Argentine paleontologist Florentino Ameghino—was a fossil that he named *Phorusrhacos longissimus.* (*Longissimus* is the species, *Phorusrhacos* the genus. Taxonomists go on to classify living and extinct organisms in increasingly larger groups: family, order, class, phylum and kingdom.) The fossil came from the Santa Cruz Formation in Pata-



gonia, the southernmost region of Argentina; the formation is about 17 million years old.

Ameghino and other researchers reconstructed the appearance of the birds from their fossil remains and their behavior from what creatures that might be living relatives do. The investigators initially interpreted the flesh-eating habits of the phorusrhacoids as an indication that they were related to modern eagles and hawks. Not all paleontologists agreed, and the issue was debated over the next 12 years. Charles W. Andrews of the British Museum resolved the controversy in 1899, concluding that among all living and extinct groups, the phorusrhacoids were most closely related to the South American seriema birds, which could also be regarded as the structural ancestors of the phorusrhacoids. Seriemas live today in the grasslands of northern Argentina, eastern Bolivia, Paraguay and central and eastern Brazil. Seriemas and phorusrhacoids are classified as members of the order Gruiformes, which includes cranes and rails and their kin.

TERROR BIRD prepares to eat a small, horselike animal (*Brachytherium*) that it has caught in a swift chase and stunned by holding the prey in its beak and beating it against the ground. This bird (*Andalgalornis*), which was as tall as a human, was one of many species—all now extinct—known as phorusrhacoids. They were the dominant terrestrial carnivores of South America until about two million years ago.



There are two living seriema species, the red-legged seriema (*Cariama cristata*) and the black-legged, or Burmeister's, seriema (*Chunga burmeisteri*). These birds reach a height of 0.7 meter. They are light-bodied, long-legged and long-necked. Their wings are small relative to their body, and the birds resort to spurts of short-distance flight only when pressed. They are excellent runners, able to attain speeds in excess of 60 kilometers per hour. Seriemas build twig nests, four to six meters above the ground, in low trees. The young, usually two, mature in about two weeks, whereupon they leave the nest to live and hunt in the nearby grasslands. Like most carnivorous animals, seriemas are territorial. Their call has been described as eerie and piercing.

Like the phorusrhacoids, seriemas are carnivorous. They eat insects, reptiles, small mammals and other birds. Under favorable conditions, they will attack larger game. They seize their prey in their beaks and beat the animal on the ground until it is limp enough to be swallowed whole. This feeding strategy is also practiced today by the roadrunner (*Geococcyx californianus*) of the southwestern U.S. and the secretary bird (*Sagittarius serpentarius*) of Africa.

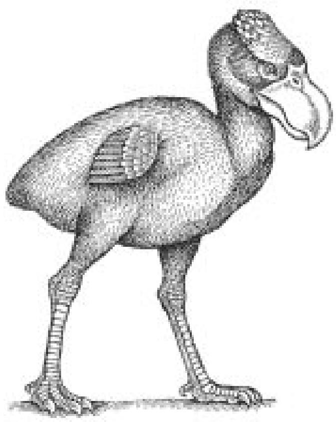
Seriemas are placed in the family Cariamidae, which now is restricted to South America. About 10 fossil species have been found there, the oldest being from the middle Paleocene epoch (some 62 million years ago) of Brazil. Relatives of this group are represented by two fossil families: the Bathornithidae, which appear in beds 40 to 20 million years old in North America, and the Idiornithidae, found in certain European rock formations 40 to 30 million years old. Some workers believe these families are so closely related that they should all be grouped in the family Cariamidae.

Most of the terror birds were considerably larger than their living relatives. The creatures ranged in height from one to three meters (just shy of 10 feet). The earliest known members are virtually as specialized as the latest, indicating that they originated before their first appearance in the fossil record.

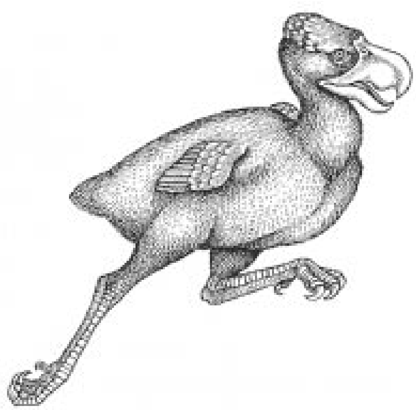
About a dozen genera and 25 species of terror birds have been recognized. The relation among them is still not clear. They were classified in 1960 by Bryan Patterson of the Museum of Comparative Zoology at Harvard University and Jorge L. Kraglievich of the Municipal Museum of Natural and Traditional Sciences of Mar del Plata, Argentina. This classification ordered the terror birds in three families that, in comparison to families of mammals, include animals of medium, large and gigantic size. Other workers, basing their view on the period of greatest diversity among terror birds, achieved between five and three million years ago, recognize two families—gigantic and medium—as well as two subfamilies. Some researchers place all the fossils in one family.

In the three-family system the gigantic forms are members of the family Brontornithidae. Fossils of this family have been found in beds ranging in age from 27 to 17 million years. A heavy, ponderous build characterized the birds; the leg bones were fairly short, the beaks massive. This evidence suggests that the birds were cumbersome runners, slower afoot than the members of the other two families.

Next comes the family Phorusrhacidae. Its members ranged between two and three meters in height. Fossils have been found in rocks ranging in age from 27 million to three million years. The third family, Psilopteridae, comprised quite small members; most of them stood no more than one meter in height. Their known fossils range from 62 million to two million years in age. Within this family is the oldest known phorusrhacoid, *Paleopsilopterus,* found in Brazil. Members of these last two

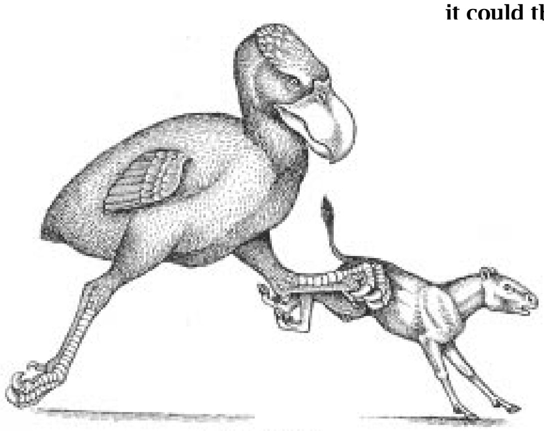


STALKING

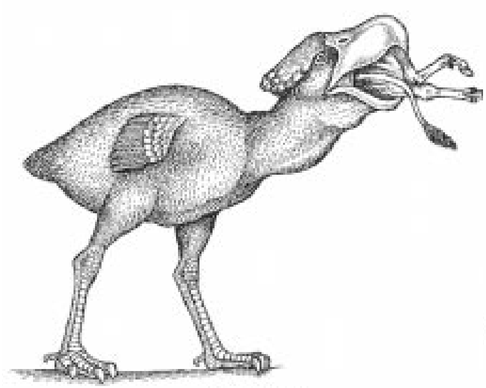


RUNNING

HUNTING TECHNIQUE of a terror bird was focused and deadly. Living on the pampas of South America, the bird could stay hidden in the grass until it had drawn close to its prey. It would then dash toward its victim at speeds close to 70 kilometers per hour, seize the catch in its beak and render it unconscious by beating it against the ground. It often ate its catch whole. Having no natural predators itself, it could then feed at leisure before returning to its nest.



STUNNING



EATING

GORGED

LIVING RELATIVES of the phorusrhacoids are the seriema birds of South America: the red-legged (*Cariama cristata*) and the black-legged, or Burmeister's, seriema (*Chunga burmeisteri*). The seriemas, which reach a height of about 0.7 meter, hunt much as the terror birds did. Seriema nests are in low trees, but the terror birds built nests on the ground.

families were lightly built, swift runners. They were the ones that became the dominant running carnivores of their time, and they held that status for millions of years.

The fact that phorusrhacoids came in several sizes indicates that the adults were capable of preying on a wide variety of animals, from rodents to large herbivores. Although some of the adult herbivores were as big as some adult phorusrhacoids, the birds could easily have preyed on the young ones. Phorusrhacoids newly out of the nest would have had different food needs because they were smaller; they probably hunted rodents and other small vertebrates, much as their living seriema relatives still do.

During much of the age of mammals (the past 66 million years), phorusrhacoids thus occupied the role of fleet-footed carnivores in South America. They were able to assume this role by giving up the greatest virtue of being a bird—the power of flight. The door to dominance as carnivores opened to the phorusrhacoids when their predecessors in that role—the small, bipedal dinosaurs known as coelurosaurs—disappeared in the dinosaur extinction 66 million years ago. Paleobiologists call such a transition an evolutionary relay.

The body forms of the terror birds and the coelurosaurs were quite similar: trim, elongated bodies; long, powerful hind limbs; long necks; large heads. Many coelurosaurs had reduced anterior limbs, indicating that the animals captured, killed and processed prey primarily with the hind limbs and mouth, as the phorusrhacoids did. Coelurosaurs apparently used their long tail as a balance while running; phorusrhacoids probably used their reduced wings for the same purpose. Different strategies and appendages were thus employed to serve the same functional purpose.

Terror birds and their relatives are also known outside South America. Their distribution is the key to the intriguing biogeographic history that accounts for the gradual ending of the terror birds' reign as South American carnivores.

In rocks from 55 million to 45 million years old in North America, Europe and Asia, large carnivorous birds are represented by the family Diatrymatidae—a family that, according to my Brazilian colleague Herculano M. F. Alvarenga, developed characteristics similar to those of the phorusrhacoids. Diatrymatidae family members attained heights of about two meters. Like the phorusrhacoids, they had massive skulls and large claws. Their legs, however, were relatively shorter and sturdier, suggesting that they were more methodical and cumbersome in their movements, much as the brontornithids were.

A reported phorusrhacoid, *Ameghinornis,* is known from the Phosphorites du Quercy rocks, 38 million to 35 million years old, in France. This animal was the size of a living seriema and was apparently capable of brief flight.

The Antarctic is also the scene of similar fossils. Two isolated footprints, 18 centimeters in length, are known in rocks about 55 million years old on the Fildes Peninsula of King George Island in West Antarctica. The three-toed bird was big, broad and elongated, either a ratite (a rhea or an ostrich or one of their relatives) or a phorusrhacoid.

The anterior part of a phorusrhacoid's beak was collected from rocks (40 million years old) of the La Meseta Formation on Seymour Island, which is on the south side of the Antarctic Peninsula. The proportions of the beak indicate that the bird was more than two meters tall.

Finally, a formidable phorusrhacoid named *Titanis walleri* is known from rocks aged 2.5 million to 1.5 million years in northern Florida. The estimated height of the bird is more than three meters. This record is the youngest yet found and represents the last of the known terror birds.

A scenario for this pattern of phorusrhacoid distribution can be constructed from the premises that these flightless birds required overland routes for dispersal and that the fossil record accurately reflects their occurrence in space and time.

Both biological and geologic evidence show that a continuously dry land bridge united North and South America about 62 million years ago. It ran by way of the Greater and Lesser Antilles, providing an opportunity for dispersal for various groups of terrestrial vertebrates. Among them were a seriema and a phorusrhacoid (probably a Psilopteridae) that dispersed north.

Forty-five million to 55 million years ago a land corridor between North America and Europe that included what is now Ellesmere Island provided another route by which the raptors could disperse. One group that appeared to have used the route was *Ameghinornis,* whose remains have been found in France. A note of caution here is in order: the supposition presumes that the phorusrhacoid group was present in North America. No fossils of that age have yet been found there.

From at least 45 million years ago, perhaps as much as 70 million, a body of land united southernmost South America and West Antarctica. The existence of a land connection at this time is supported by a group of marsupials, an armadillo and the southern beech in the same rock beds as the phorusrhacoid on Seymour Island. Together the land bridge and the cool, temperate climate of the time account for the presence of terror birds in West Antarctica 40 million years ago.

Eventually the land bridges uniting South America with North America and Antarctica disappeared. South America remained an island continent until the appearance of the Panamanian land bridge 2.5 million years ago. The bridge formed as a result of the continued tectonic uplift of the northern Andes, probably associated with a worldwide drop in sea level of as much as 50 meters resulting from the buildup of the polar ice caps. The final connection of the bridge was in the area of what has become southern Panama and northern Colombia.

A cooling of world climates at the time shrank tropical habitats and expanded the savannas. Grassland environments were established on the land bridge. After a time, a continuous corridor of savannas extended from Argentina to Florida. The reciprocal dispersal of terrestrial fauna made possible by these conditions is now known as the Great American Interchange. It represents the best-documented example in the fossil record of the intermingling of two long-separated continental biotas. Among the participants were the terror birds. One phorusrhacoid lineage survived beyond 2.5 million years ago in South America, and individuals dispersed north to give rise to *Titanis* in Florida.

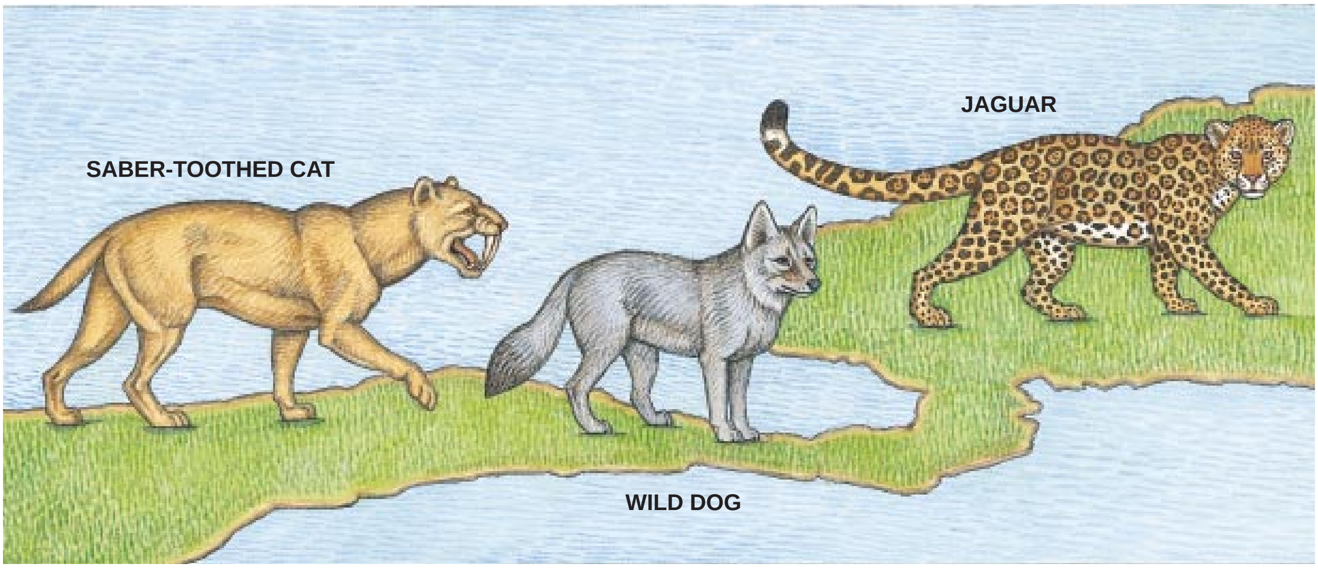
Against this background, one can begin to see why a group of large, flightless birds rose to the top of the food pyramid in South America and why they finally lost that position. The answer lies in the historical development of the terrestrial fauna of South America. Recall that for most of the past 66 million years South America was, as Australia is today, an island continent. As a consequence of the groups that inhabited each continent 66 million years ago, the role of terrestrial mammal carnivores was filled in South America by marsupials and the role of large herbivores by placentals. This marsupial-placental combination was unique among continental faunas; both roles were filled by marsupials in Australia and by placentals in North America, Europe and Asia.

The group of South American marsupials that evolved to fill the place that placental dogs and cats eventually held on the northern continents is called borhyaenoid. Its doglike members are further grouped into three families. They ranged in size from that of a skunk to that of a bear. One specialized family, the thylacosmilids, had characteristics similar to those of the placental saber-toothed cats. It is particularly significant that all these animals were relatively short-legged and that none showed marked adaptation to running. These were the mammal occupants of the carnivore niche in South America.

Also in this niche were large terrestrial or semiterrestrial crocodiles of the family Sebecidae. They had deep skulls; their limbs were positioned more under the body than those of aquatic, flat-skulled crocodiles, and their laterally compressed teeth had serrated cutting edges, much like those of carnivorous dinosaurs. The other occupants of the carnivore niche were the terror birds. Thus, from about 66 million to about 2.5 million years ago, the role of terrestrial carnivore in South America was shared at various times, but not equally, by marsupial mammals, sebecid crocodiles and phorusrhacoid birds.

From about 27 million to 2.5 million years ago, the fossil record shows a protracted decrease in the size and diversity of the doglike borhyaenoids and a concurrent increase in the size and diversity of the phorusrhacoids. Consequently, by about five million years ago, phorusrhacoids had completely replaced the large carnivorous borhyaenoids on the savannas of South America. (The smaller ones, which were not competitive with the terror birds anyway, also became extinct before the Panamanian land bridge appeared.) This transition demonstrates another relay in the evolutionary history of the phorusrhacoids

DECLINE OF TERROR BIRDS began when the Panamanian land bridge (*red*) formed between North and South America some 2.5 million years ago, allowing North American mammals that could outhunt the terror birds to enter South America. Fossils of animals that migrated south or north have been found on both continents (*green*). Terror bird fossils (*orange*) are mostly in South America. A tip of a jaw (*bottom*) was found in Antarctica.



DOWNFALL OF TERROR BIRDS was apparently caused by three examples of the many species of animals that crossed the Panamanian land bridge into South America. Greater intelligence, more speed and agility, or the ability to prey on terror bird eggs and hatchlings could explain how these migrants from North America displaced the terror birds.

whereby they successfully replaced their marsupial counterparts, the borhyaenoids. Just why the phorusrhacoids were able to do so is unclear, but their superior running ability would certainly have been an advantage for capturing prey in the savanna environments that first came into prominence about 27 million years ago.

After the emergence of the Panamanian land bridge, placental dogs and cats of the families Canidae and Felidae dispersed into South America from North America. Because all the large marsupial carnivores of South America were by then long extinct, the only competition the dogs and cats had was the phorusrhacoids. It proved to be a losing battle for the birds.

Thus it was that the phorusrhacoids reached their peak in size and diversity just before the interchange, gradually declining thereafter because of the competition with the dogs and cats. Only one lineage survived beyond 2.5 million years in South America; it is the one that dispersed to Florida, where it is represented by *Titanis.* This was the only South American carnivorous animal to disperse northward. Its success there at coexisting with the advanced placental carnivores was brief. Why that was so is a major riddle. Perhaps the resident placental carnivores were too well established for the phorusrhacoids to find a permanent niche.

The fate of the phorusrhacoid relatives in North America and Europe between 55 million and 45 million years ago is also linked to the appearance of advanced placental carnivores. During that time on the northern continents, the large mammalian carnivores were the creodonts. This primitive group of placentals resembled the marsupial borhyaenoids in that they lacked special running abilities and had rather small brains. The phorusrhacoid relatives on these continents disappeared with the appearance of advanced placental mammals beginning about 45 million years ago.

The terror birds thus flourished in the absence of advanced placental carnivores, which have repeatedly shown themselves to be better competitors. The marsupial borhyaenoids and placental creodonts were, in essence and in comparison with the terror birds, second rate.

Although plausible, this argument is speculative. One cannot identify with certainty a single factor that explains the extinction of any group of animals now found only as fossils. In the case of the terror birds, their disappearance on two occasions in time correlates directly with the appearance of advanced placental carnivores. Were the advanced placentals more intelligent than the terror birds and so better adapted to capturing the prey that the birds had had to themselves? Did the fact that they had four legs give them an advantage over the two-legged phorusrhacoids in speed or agility? Did the placentals eat the phorusrhacoids' eggs, which were readily accessible in ground nests because of the birds' large size? Did the placentals prey on the vulnerable hatchlings?

It is intriguing to think what might happen if all big carnivorous mammals were suddenly to vanish from South America. Would the seriemas again give rise to a group of giant flesh-eating birds that would rule the savannas as did the phorusrhacoids and their bygone allies?

FURTHER READING

SISTEMÁTICAS Y NOMENCLATURA DE LAS AVES FORORRACOIDEAS DEL PLIOCENO ARGENTINO. Bryan Patterson and Jorge L. Kraglievich in *Publicaciones del Museo Municipal de Ciencias Naturales y Tradicional de Mar del Plata,* Vol. 1, No. 1, pages 1–51; July 15, 1960.

A GIANT FLIGHTLESS BIRD FROM THE PLEISTOCENE OF FLORIDA. Pierce Brodkorb in *The Auk,* Vol. 80, No. 2, pages 111–115; April 1963.

THE TERROR BIRD. Larry G. Marshall in *Field Museum of Natural History Bulletin,* Vol. 49, No. 9, pages 6–15; October 1978.

LES OISEAUX FOSSILES DES PHOSPHORITES DU QUERCY (ÉOCÈNE SUPÉRIEUR À OLIGOCÈNE SUPÉRIEUR): IMPLICATIONS PALÉOBIOGÉOGRAPHIQUES. Cecile Mourer-Chauviré in *Geobios, Memoire Special,* No. 6, pages 413–426; 1982.

A GIGANTIC PHORORHACOID(?) BIRD FROM ANTARCTICA. Judd A. Case, Michael O. Woodburne and Dan S. Chaney in *Journal of Paleontology,* Vol. 61, No. 6, pages 1280–1284; November 1987.

LAND MAMMALS AND THE GREAT AMERICAN INTERCHANGE. Larry G. Marshall in *American Scientist,* Vol. 76, No. 4, pages 380–388; July/August 1988.

THE TERROR BIRD STILL SCREAMS. Kent H. Redford and Pamela Shaw in *International Wildlife,* Vol. 19, No. 3, pages 14–16; May/June 1989.

Particle Metaphysics

by John Horgan, *senior writer*

In the aftermath of the Superconducting Super Collider's death, physicists are divided over how—or even whether—they should continue their search for a unified theory of nature

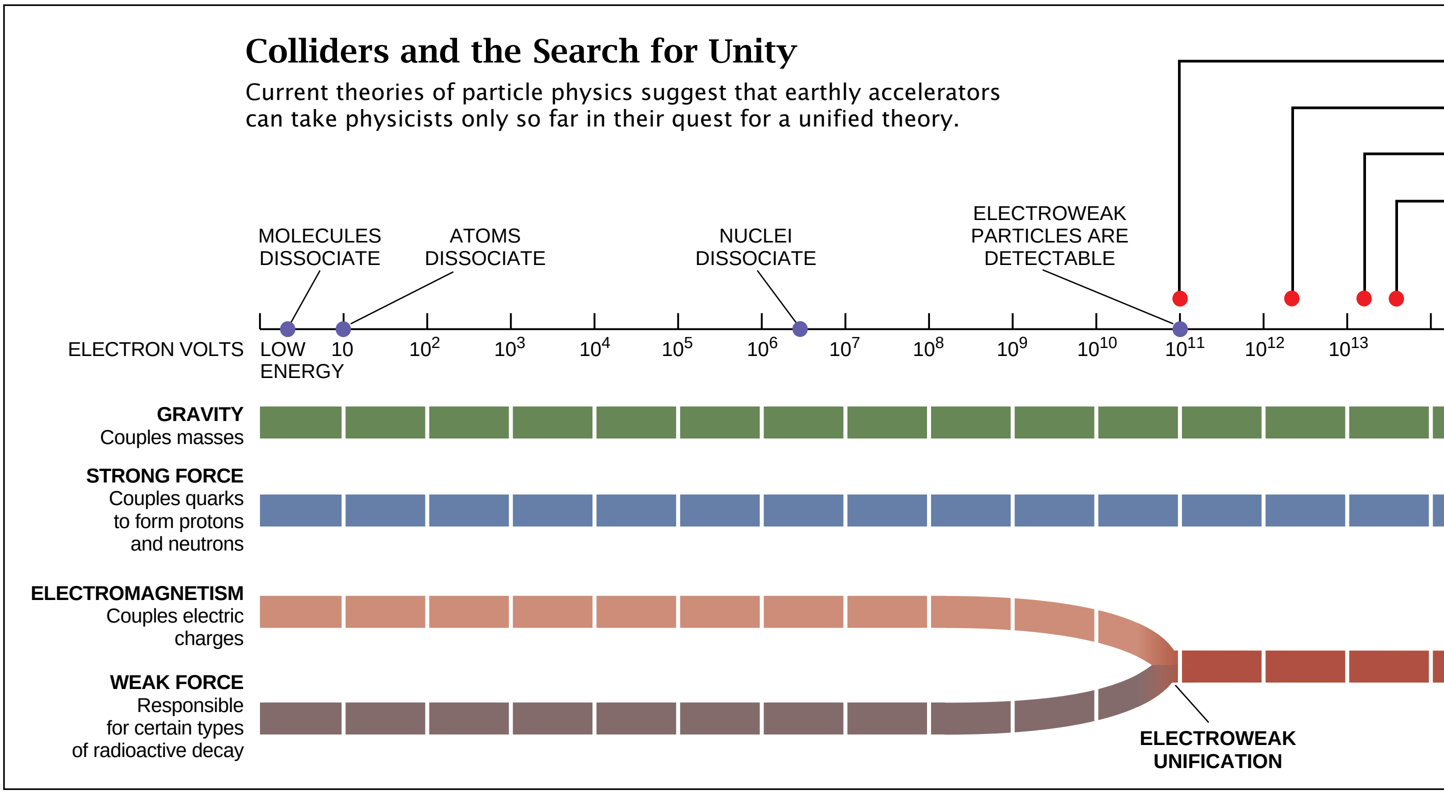
More than 150 years ago Michael Faraday revealed through a series of brilliant experiments that electricity and magnetism are manifestations of the same underlying force. Inspired by this success, Faraday sought to demonstrate that electromagnetism is similarly linked to gravity, which Newton had mathematically described some 150 years earlier. Although he failed, Faraday remained convinced that such a unified theory existed.

Many—though certainly not all—modern physicists have come to share Faraday's faith that nature's seemingly distinct forces are but facets of a single, symmetrical jewel. The quest to find this touchstone has transformed modern physics into an epic drama, one with primordial roots. In his recent book *Dreams of a Final Theory,* Nobel laureate Steven Weinberg of the University of Texas says a unified theory would bring to fruition "the ancient search for those principles that cannot be explained in terms of deeper principles." Could it be that this quest will never be completed?

Physics has certainly become more complicated since Faraday's time. Early in this century Albert Einstein replaced Newton's relatively simple account of gravity with general relativity, which turned the once rigid girders of space and time into rubber. Later, investigators discovered two other forces in addition to gravity and electromagnetism: the weak force, which gives rise to certain kinds of nuclear decay, and the strong force, which keeps protons and neutrons gripped together in atomic nuclei.

Yet until recently, physicists were advancing toward unification with what seemed like inexorable momentum. They have developed a theory—validated by accelerators just over a decade ago—depicting electromagnetism and the weak nuclear force as two aspects of the same electroweak force. Emboldened by this success, workers have invented so-called grand unified theories, or GUTs, linking the electroweak force and the strong nuclear force. They have even put forward proposals embracing all the forces, including gravity. These are sometimes called quantum gravity theories, since they attempt to fuse quantum mechanics and general relativity, which are ordinarily as compatible as oil and water. Because such theories hold out the promise of illuminating

SUPERCONDUCTING SUPER COLLIDER, which would have been 20 times more powerful than the largest existing accelerator and was expected to cost $11 billion, would still have fallen far short of the realms where all nature's forces are thought to merge. Last October, after workers had already spent $2 billion and dug one fifth of what was to be an oval, 86-kilometer-long tunnel in Waxahachie, Tex., Congress canceled the project.



the fiery birth of the universe, when a single, supreme force may have briefly reigned, they are also known as theories of everything.

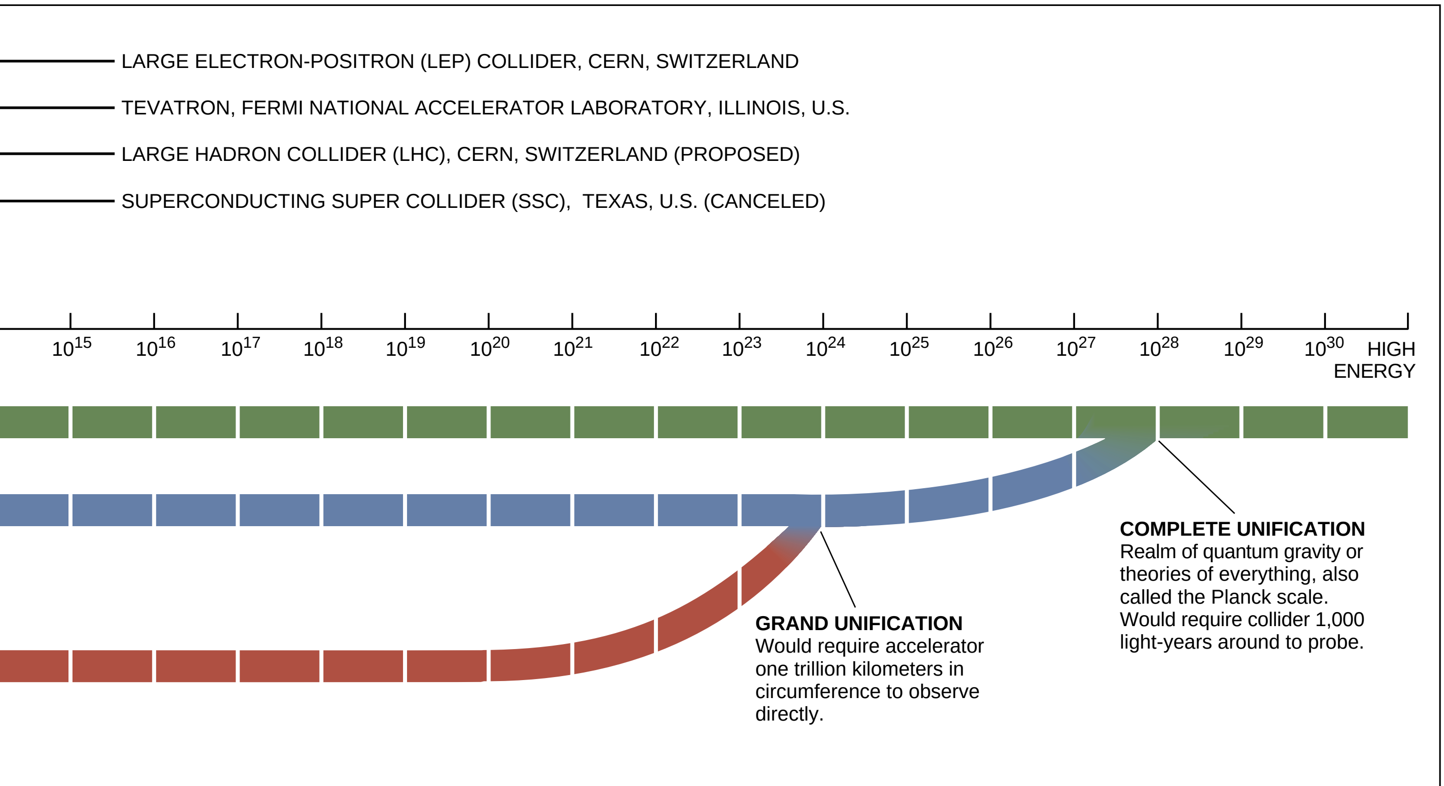
Now, just when the goal seemed to be in sight, physicists face the possibility that their journey might end short of its destination. The immediate cause of this concern is the decision of the U.S. Congress last October to cancel the $11-billion Superconducting Super Collider (SSC), four years after construction had begun. Leon M. Lederman of the Illinois Institute of Technology, a Nobel Prize–winning experimentalist, notes that particle physics has advanced through the probing of ever smaller distances and higher energies. The SSC, which would have been at least 20 times more powerful than any existing accelerator, would have brought physicists that much closer toward unification. "If we don't build the SSC or something like it," he says, "I think the field may come to an end." If particle physics stalls, so might cosmology, which in its efforts to reconstruct the universe's history has become increasingly dependent on unified theories.

But the demise of the SSC may only have brought physics to its current impasse more quickly. The rhetoric employed in defense of the SSC often implied that it would lead to an ultimate theory. The title of Lederman's recent book *The God Particle,* which eloquently argues for the SSC, is a case in point. Yet Lederman concedes that the SSC would have been too small by many orders of magnitude to achieve the energies at which unification is thought to occur. An SSC-type accelerator would have to trace a circle one trillion kilometers around to reach the grand unification realm; light would take one month to traverse that distance. A collider capable of probing the infinitesimal quantum gravity realm would need to be as much as 1,000 light-years in circumference. The entire solar system is only one light-*day* around.

Before the plug was pulled on the SSC, in fact, one prominent theorist had advocated downplaying the search for a unified theory, suggesting that it is beyond the purview of particle physics. In a recent lecture, Howard Georgi of Harvard University even expressed a kind of regret at having proposed one of the first grand unified theories some 20 years ago: "The legacy of grand unification, which in my view is very bad for the field of particle physics, is that it is considered reasonable—and even fashionable—for someone who calls him or herself a particle theorist to spend full time speculating about the world at distances much smaller than anything that we will ever be able to study in the laboratory."

Others keep the faith. Among them is John Ellis, a theorist at the European laboratory for particle physics near Geneva, known by the French acronym CERN. Ellis helped to promulgate the terms "grand unification" in the mid-1970s and "theory of everything" a decade or so later. "There are still prospects for progress in the traditional directions," he asserts. He notes that CERN intends to build an accelerator nearly as powerful as the SSC within 10 years. Until then, clues might be gleaned from high-precision experiments on existing accelerators; studies of neutrinos, which in spite of their shyness in the presence of detectors loom large in many unified theories; and observations of cosmic background radiation, thought to be a relic of an era shortly after the big bang, when nature's forces were still interwoven.

Over the longer run, innovative colliders, such as one that boosts particles on a wave of plasma, may allow physicists to probe higher energies more cheaply. And there is always the possibility of a mathematical or conceptual breakthrough that propels researchers to a deeper level of understanding in the absence of experimental guidance. But even Ellis, a self-proclaimed "dyed-in-the-wool optimist," concedes that "until someone comes up with a very good idea, I think we will be looking at only indirect evidence" of unification in the foreseeable future. Frank A. Wilczek of the Institute for Advanced Study in Princeton, N.J., agrees. "We've had the joy of making fundamental progress," he says. "The prospects for



that continuing are really diminished."

In a sense, physicists are victims of their own success. They have constructed a theory that accounts for particle interactions with extraordinary accuracy. Called the Standard Model (or, by Michael Dine of the University of California at Santa Cruz, the "theory of almost everything"), it rests on the sturdy foundation of quantum mechanics, a radical theory of matter and energy erected by such giants as Niels Bohr, Werner Heisenberg and Erwin Schrödinger in the 1920s and 1930s. In the 1950s Richard Feynman and others invented a theory of electromagnetism, called quantum electrodynamics, that accounts for virtually all chemical and electronic phenomena.

During the following decade, physicists developed a theory for the strong nuclear force: quantum chromodynamics, or QCD. It holds that protons and neutrons are manifestations of more elementary particles named quarks. Each proton and neutron is composed of three quarks, which are themselves bound together by force-carrying particles described as gluons. (The prefix "chromo" refers to the fact that quarks are categorized according to their "color," a quantum mechanical property unrelated to color in its usual sense.)

The field took a huge step toward unification in the 1960s, when Weinberg, Sheldon L. Glashow of Harvard and Abdus Salam of the International Center for Theoretical Physics in Trieste and others proposed that electromagnetism and the weak nuclear force are actually two manifestations of the same electroweak force. Both quantum chromodynamics and the electroweak theory have been validated by increasingly stringent tests at the world's major accelerator laboratories, notably the Stanford Linear Accelerator Center, or SLAC; the Fermi National Accelerator Laboratory, commonly known as Fermilab, in Batavia, Ill.; and CERN.

No GUTs, No Glory

As soon as experimentalists finished validating the Standard Model, however, they were eager to shatter it. In 1983 Carlo Rubbia announced that a team under his supervision at CERN had found not only *Z*'s and *W*'s, which are particles that carry the electroweak force, but also "monojets," a phenomenon that apparently violated the Standard Model. "Rubbia said that CERN had not only confirmed the Standard Model but had also put it in its grave," recalls Glashow of Harvard, where Rubbia held a joint appointment at the time of his announcement.

Rubbia shared a Nobel Prize for uncovering the *Z* and *W* particles, but the monojets were soon found to be accounted for by the Standard Model. Glashow jokes that as punishment for his premature announcement Rubbia has been forced to oversee ever more refined validations of the Standard Model at CERN, which he directed from 1989 through the end of last year. "It's like something out of a Greek myth," Glashow says. Unfortunately, the rest of particle physics has had to share Rubbia's punishment.

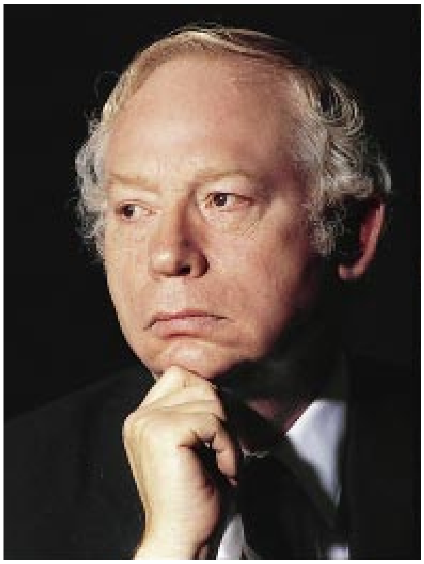
Meanwhile theorists had already forged far beyond the Standard Model in search of a deeper theory. They were encouraged by the fact that quantum chromodynamics and the electroweak theory are both gauge theories, which posit that all the elements of a system can undergo transformations—such as rotation or reflection in a mirror—without being fundamentally altered. This feature, called symmetry, has become for many particle physicists the epitome of truth and beauty.

Early in the 1970s Glashow and Georgi, his younger colleague at Harvard, invented a gauge theory, called SU(5), that could yield both electroweak interactions and strong interactions. (The term "SU(5)" refers to the number of symmetries displayed by the theory.) The grand unified theory—a term that Glashow and Georgi did not invent and insist they do not like—made what was then a startling prediction. Quarks, it seemed, could change into neutrinos, electrons and their antimatter counterparts; that meant protons (which are composed of quarks) were unstable and would eventually decay. Although the estimated half-life of any particular proton would be longer than the age of the sun (according to calculations done later by Weinberg, Georgi and Helen Quinn of SLAC), physicists could test the prediction by watching a sufficiently large number of protons.

Proton-decay detectors have now been built at more than half a dozen sites around the world. Most are placed deep underground to minimize signals from cosmic rays—high-energy particles from outer space. One of the largest experiments began operating in a salt mine near Cleveland, Ohio, some 10 years ago. It consists of a gigantic vat of water surrounded by photodetectors that watch for the minute flashes of light that should be released by the decay of a proton in the water. So far neither this nor any other detector has observed proton decay.

The lack of experimental support for the SU(5) theory merely opened the door to alternatives, notably a more general approach known as supersymmetry. Supersymmetry holds that fermions, the particles that constitute matter, and bosons, which transmit forces, share deep symmetries. The scheme requires that each known particle have a relatively massive, supersymmetrical partner, or "sparticle." One striking feature of supersymmetry is that its power can be increased if it is extended into extra dimensions. Just as an astronaut rising above the two-dimensional plane of the earth can apprehend its global symmetry, so can theorists discern the more subtle symmetries underlying particle interactions by viewing them from a higher-dimensional standpoint.

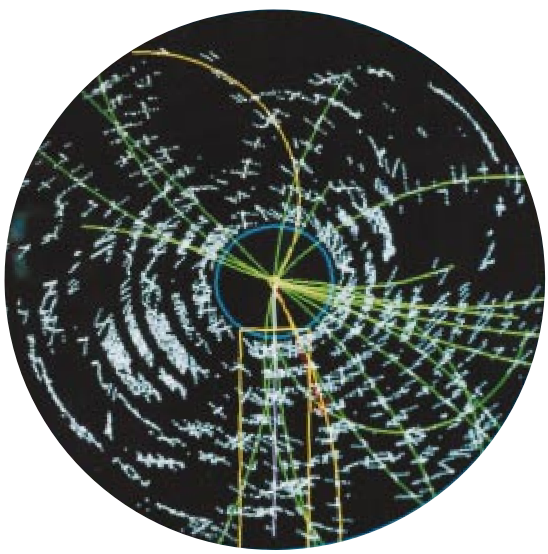
Theorists have constructed various supersymmetrical grand unified theories and even quantum gravity theories. An example of the latter is supergravity, which assumes that gravitons, the particles that transmit gravity, have supersymmetrical partners called gravitinos. In 1980 supergravity seemed so promising that Stephen W. Hawking of the University of Cambridge announced that it might represent the long-sought "complete and unified theory of physics." But supergravity soon bogged down in mathematical problems related to the definition of gravitons as points. Just as division by zero yields an infinite and hence meaningless result, so do calculations involving pointlike particles. Gauge theories had helped physicists constructing models of electromagnetism and the nuclear forces overcome this problem. But gravity, with its distortions of space and time, seemed to demand an even more radical approach.



STEVEN WEINBERG of the University of Texas, a leading proponent of unification, fears that further progress may be difficult if a higher-energy accelerator is not built.

Many physicists think superstring theory represents that approach. Superstring theory began rather modestly. In the early 1970s theorists proposed that the strong nuclear force could stem from interactions of stringlike particles. In the same way that vibrations of violin strings give rise to different notes, so could the vibrations of these strings yield the disparate particles involved in the strong force.

That route was abandoned in favor of the far more successful method utilizing quarks and gluons. But string theory was resurrected in supersymmetrical form in the late 1970s by Michael B. Green of Queen Mary College, London, and John H. Schwarz of the California Institute of Technology. Somewhat to their amazement, Green and Schwarz found that supersymmetrical strings generated all the forces of nature, including gravity. Best of all, the substitution of strings for points eliminated many of the mathematical problems arising in other quantum gravity theories.



TOP QUARK is the sole piece of the Standard Model yet to be unambiguously detected. This image shows a top-quark candidate spotted at the Fermi National Accelerator Laboratory in 1989.

The theory demands acceptance of some far-fetched assumptions about physical reality. The strings are thought to inhabit as many as 26 dimensions, and they are as small in comparison to a proton as a proton is in comparison to the solar system. This microrealm, named the Planck scale, is inaccessible to any conceivable experiment. Physicists and, increasingly, mathematicians have nonetheless become entranced by the theory's rich structure. Indeed, Edward Witten of the Institute for Advanced Study, a premier superstring theorist, has become as potent an influence in mathematics as in physics.

Yet even Witten, whose analytical skills are legendary, has struggled to link superstrings to known physical phenomena. Recently he has forged a bridge between superstrings and black holes, which have traditionally been the playground of theorists specializing in general relativity rather than particle physics. In 1991 Witten showed how superstring theory could yield black holes—albeit only highly simplified, two-dimensional ones. Witten's paper triggered a burst of theoretical activity that continues to this day.

Superstrings may also help wrap up a conundrum related to black holes pointed out by Hawking two decades ago. Hawking showed that quantum effects might cause black holes to radiate away energy—and therefore mass—until they eventually evaporate. He summed up his finding with the phrase "Black holes ain't so black." Because a black hole represents, at least in principle, a record of the processes that created it, its evaporation results in a permanent loss of information. The past, in a sense, is eradicated. Hawking proclaimed, and many theorists agreed, that he had uncovered a paradox that could be resolved only by modifying either quantum mechanics or general relativity.

In a paper published in *Physical Review Letters* last October, Leonard Susskind of Stanford University shows how superstrings might solve the puzzle. The paradox identified by Hawking, Susskind explains, arises out of the assumption, embedded in general relativity, that different observers have the same picture of how information is stored in a given region of space and time. But according to superstring theory, different observers can have different pictures. For any single observer, the past is preserved.

Critics charge that such work is not even physics, because it is so divorced from any experimentally accessible phenomena. Susskind retorts that progress in physics can no longer be achieved in traditional ways. "It's quite apparent to me that if the questions raised over the past 15 to 20 years are going to be resolved, it is not going to be through experiments proceeding in incremental steps," Susskind declares. "It would be hopeless to get to the Planck scale this way. People who ignore this are simply going to become irrelevant." Witten expresses the same view, though somewhat more mildly: "I think we could do much better with experiments, but I have faith in human perseverance."

Desperately Seeking Data

Most theorists nonetheless crave some hint from experiments that they are on the right track. Many have pinned their hopes on finding evidence of supersymmetry, which is a necessary though not sufficient test for superstrings. Both Fermilab and CERN have attempted to detect supersymmetrical particles, in vain.

Some workers claim CERN has already provided what may be tentative evidence of supersymmetry. Researchers there have been performing highly precise measurements of the so-called coupling constants of electromagnetism, the weak force and the strong force. (The coupling constant of a force is a measure of its strength.) Grand unified theories predict that the coupling constants of the three short-range forces, which have different values at low energies, should converge at high energies. CERN's data disagree with the predictions of the old SU(5) GUT, but when supersymmetry is added to the theory the predictions match "exactly," according to Ellis of CERN.

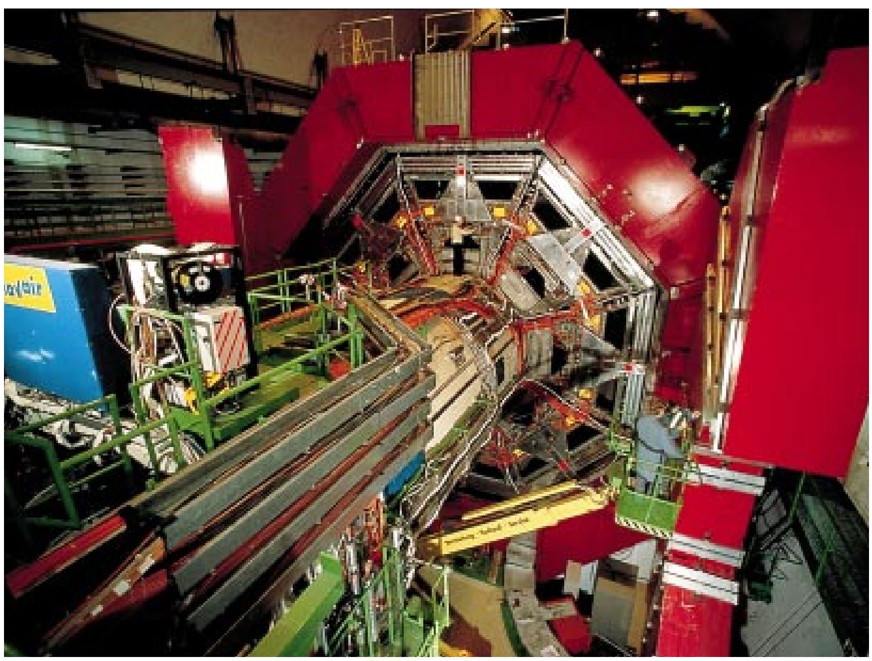
Others find these results unconvincing. "I wouldn't jump up and down yet," says Alvaro De Rújula, another CERN theorist. He notes that the trends measured at CERN support supersymmetry only if they are extrapolated by many orders of magnitude. Indeed, De Rújula suggests that some of his colleagues are a bit too enamored of supersymmetry, given the paucity of evidence. "The coincidence in sound between supersymmetry and superstition is significant," he says dryly.

Various other experiments may illuminate territory beyond the Standard Model soon. Fermilab is still searching for the top quark, the lone particle predicted by the Standard Model that has not yet been captured. The continued failure to find the top quark would contradict the Standard Model and would thus be "a major breakthrough," notes José N. Benlloch, a top-quark hunter at Fermilab. Yet workers there may already have glimpsed the particle in the expected range, and most observers anticipate that they will announce its discovery soon.

Both Fermilab and CERN are planning upgrades of their main accelerators. Fermilab hopes to increase the beam density of the Tevatron, which slams protons and antiprotons into one another, and CERN intends to double the power of its Large Electron-Positron (LEP) collider. These machines will give workers another shot—albeit a long one—at finding supersymmetrical particles. They have an even more remote chance of finding the Higgs boson. Theorists think this particle may play a crucial role in the symmetry breaking, or unraveling, of electromagnetism and the weak force at high energies. The Higgs boson may even illuminate how particles acquire such a wide range of seemingly arbitrary masses.

Accelerators may also keep theorists busy with studies of mesons, extremely short-lived particles composed of a quark and an antiquark. Experiments with so-called *K* mesons, or kaons, in the 1960s led to the discovery that matter and antimatter are not mirror images of each other but instead exhibit a subtle asymmetry. Some theorists have speculated that without this asymmetry, called charge-parity violation, the universe would not exist, because the big bang would have spawned precisely the same amounts of matter and antimatter. Matter and antimatter annihilate each other on contact.

Physicists hope to explore these ideas further with experiments on *B* mesons, which are expected to display charge-parity violation more often than *K* mesons do. Last fall the same bill that killed the SSC authorized the construction of a $237-million facility for gener-



EUROPEAN LABORATORY for particle physics near Geneva, also called CERN (*top*), represents the best hope for particle physicists now that the SSC is dead. Over the next decade, officials there hope to build the world's most powerful accelerator, the Large Hadron Collider, in the 27-kilometer tunnel now housing the Large Electron-Positron Collider (LEP). One of the LEP's massive detectors is shown (*bottom*).

Does Particle Physics Need a New Paradigm?

Although it is difficult to understand and to manipulate, superstring theory is by far the leading candidate for a quantum gravity theory. One reason may be that it represents an extension of ideas, such as symmetry, that are deeply ingrained in the culture of particle physics. Some theorists think it is time to consider fresh approaches.

One innovative quantum gravity scheme has emerged over the past half a dozen years from a group of specialists in general relativity. It is called loop-space theory, or the Ashtekar theory, after one of its originators, Abhay Ashtekar of Syracuse University. Ashtekar and his colleagues have found a way to rewrite the equations of general relativity so that they resemble the equations of quantum electrodynamics.

This method allows them to treat gravity as a quantum mechanical phenomenon without encountering the mathematical problems that have blocked other attempts. One of the implications of the theory is that space is not a seamless entity but is composed, like a sheet of chain mail, of discrete, infinitesimal loops.

An even more radical concept has been put forward by Gerard 't Hooft of the University of Utrecht in the Netherlands. 'T Hooft played a major role in the development of gauge theories, which are the language in which the Standard Model is written. Yet he has become increasingly dissatisfied with current approaches to quantum gravity, which he feels gives short shrift to such essential concepts as causality.

He suggests that physicists try constructing physical models based on cellular automata, creations of computer science that have causality as a basic feature. A cellular automaton consists of a grid of discrete units, or cells, that evolve according to specific rules. The state of each cell is determined by the states of its immediate neighbors.

Another prominent physicist advocating new modes of thinking is Roger Penrose of the University of Oxford. He questions the assumption of most particle physicists that symmetry is a fundamental feature of nature rather than of their theories. This belief, Penrose contends, rests ultimately on an aesthetic preference, one that he and other physicists do not necessarily share.

"In my view," Penrose writes in a recent essay, "if there is to be a final theory, it could only be a scheme of a very different nature. Rather than being a physical theory in the ordinary sense, it would have to be a principle—a mathematical principle whose implementation might itself involve nonmechanical subtlety."

LOOP-SPACE THEORY is modeled by a sculpture that is made of key rings.

ating copious amounts of *B* mesons at SLAC. This "*B* factory" could generate results that violate the Standard Model, which permits charge-parity violation only within certain parameters.

Not all the action is at the accelerator laboratories. Results with cosmic significance may stem from neutrino observatories. Invariably described as "elusive" because they scarcely interact with normal matter and are therefore difficult to detect, neutrinos play a crucial role in electroweak interactions and in cosmological theories. Beginning in the 1970s, underground detectors—some of them originally built to detect proton decay—have found that the sun emits fewer neutrinos than predicted by the Standard Model.

The solar-neutrino deficit has been confirmed by two new detectors, also set underground to cut down on signals from cosmic rays, that watch for neutrinos in gallium compounds rather than vats of fluids. The two observatories are the Soviet-American Gallium Experiment, or SAGE, which is buried beneath a mountain in Russia's Caucasus Mountains, and the Gallium Experiment, or GALLEX, part of the Gran Sasso laboratory in Italy's Apennine Mountains.

One explanation for the discrepancy is that neutrinos "oscillate" between various types, changing from electron neutrinos into versions more difficult to detect. This is called the MSW conjecture, after its authors Stanislaw P. Mikheyev and Aleksei Y. Smirnov, both from the Academy of Sciences in Moscow, and Lincoln Wolfenstein of Carnegie Mellon University. Such oscillations could occur only if neutrinos have mass, a result that would contradict the strictest version of the Standard Model.

John N. Bahcall of the Institute for Advanced Study, an authority on neutrinos, acknowledges that the Standard Model could be modified to accommodate massive, oscillating neutrinos. On the other hand, the values observed so far are all within the ranges predicted by grand unification theories, including those incorporating supersymmetry. More detailed studies "could be used as a test of which of many grand unification ideas are correct," Bahcall adds. The confirmation that neutrinos have masses could make them a leading candidate for the "dark" or "missing" matter that most cosmologists think pervades the universe. More data should be forthcoming within the next few years as extremely sensitive neutrino observatories begin operating in Japan and Canada. At the same time, accelerator-based experiments—including two at CERN and another at the Gran Sasso laboratory—will attempt to pin down the masses of neutrinos.

Weinberg warns against being overconfident that any of these experiments will yield substantially new physics: "There are various ways we might get lucky over the next five to 10 years, but we've already been saying that for 10 or 15 years now." Glashow is similarly pessimistic. He doubts whether low-energy experiments are ever again likely to discover phenomena that violate the Standard Model. He points out that there have been numerous claims of such phenomena in recent years—involving an entirely new "fifth force" of nature that counteracted gravity; extremely heavy (17,000 electron volts) neutrinos; and monopoles, exotic particles possessing only a single magnetic pole. None of these has held up.

"The question in my mind is how you keep together the high-energy physics establishment" until the next large accelerator is built, Glashow says. "People will do boring things. Nobody will admit that what they're doing is boring, but it'll still be boring."

The Superconducting Super Collider, Weinberg and Glashow contend, represented the best possible avenue to exciting new physics. The SSC was often marketed as a machine that would reveal the Higgs boson. But as CERN theorist De Rújula points out, the discovery of the Higgs boson would not necessarily break the field of physics wide open. Whereas such a finding would be a crucial confirmation of electroweak theory, in itself it would represent merely an extension of the Standard Model, he explains. "The books have already been written on that. It might not open a window to anything new."

Supersymmetry represented a far more significant target of the SSC, according to David J. Gross of Princeton University. "That would have been a major discovery, as major as any in this century," says Gross, who played an important role in the development of quantum chromodynamics and is now a prominent superstring advocate. "It would have enlarged our view of space and time. It would have proved the existence of other dimensions." John P. Preskill of Caltech had still another hope, that the SSC would find something utterly unanticipated. "That was what was needed to liven up particle physics," he says.

Now that the SSC is gone, physicists hope to find salvation through the Large Hadron Collider, or LHC, a similar but smaller proton smasher that may be built at CERN. CERN once intended to build the LHC before the super collider and possibly discover the Higgs boson first. The chief advantage of the LHC is that it would be built in a preexisting, 27-kilometer-long tunnel, the one now housing the LEP collider. Officials at CERN estimate that the Large Hadron Collider could be built for under $3 billion, less than a third of the estimated cost of the SSC.

Some observers have voiced concern that the European Community may follow the lead of the U.S. Congress and cancel or indefinitely delay the construction of the LHC. Yet Christopher Llewellyn Smith, who succeeded Rubbia as director general of CERN in January, is confident that the LHC will be built and operating sometime early in the next century. "I was confident anyway, but now it just has to go forward, because it will be unique," he says. While acknowledging that some CERN members fear "their voices might get lost" if the U.S. decides to contribute to the LHC project, Llewellyn Smith insists that "we have an obligation to find a way" to let American physicists participate.

Llewellyn Smith expects CERN to operate the LEP collider through the end of this century and to bring the Large Hadron Collider on-line by 2003. This schedule will permit physicists to exploit fully the LEP and to design the LHC carefully. Letting that date slip much might endanger the vitality of the field, Llewellyn Smith warns. "You can't start an experiment and expect your grandchildren to finish it."

There is no guarantee that the LHC will be built—or that it will be a success. Although cheaper than the SSC, it will be a riskier project, as American physicists were fond of pointing out before the SSC was finally killed. With an energy range only one third that of the SSC, the LHC would have a smaller chance of finding the Higgs boson or supersymmetrical particles—or some unexpected phenomenon. Moreover, to achieve that energy range in its relatively small tunnel, it must push superconducting-magnet technology to its limits. American physicists are nonetheless rallying behind the machine they once criticized. The LHC represents "our best hope now," Weinberg comments.

Plasma Surfing

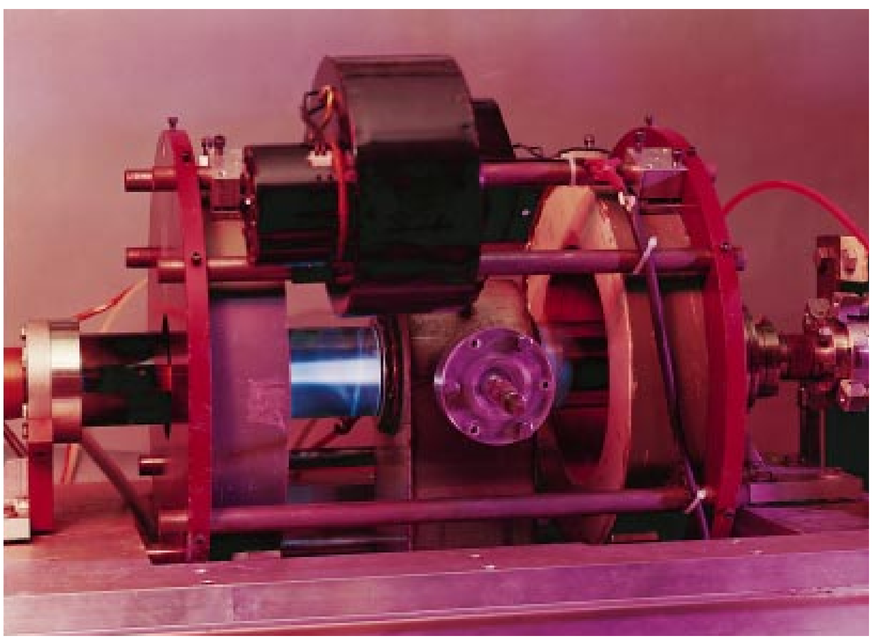
Some workers, particularly those who speculate about quantum gravity, are hopeful that their field can progress through some dramatic theoretical advance that yields greater understanding of experimentally accessible phenomena. Witten says physics may leap forward when investigators unearth what he calls the "core geometrical principles" underlying superstring theory. His fellow string enthusiast Gross adds, "I have fantasy scenarios in which we succeed without experimental input."

But even Sidney R. Coleman of Harvard, who is renowned for his work on such highly speculative phenomena as parallel universes and wormholes, which are ruptures in the fabric of space and time, finds such scenarios unlikely. "Experiment is the source of scientific imagination," he remarks. "All the philosophers in the world thinking for thousands of years couldn't come up with quantum mechanics."

Samuel C. Ting, a professor at the Massachusetts Institute of Technology and head of the largest detector at CERN's LEP collider, agrees. He points out that in this century, advances in physics "almost always come from a totally unexpected experimental result." The discoveries of antimatter (predicted by P.A.M. Dirac in 1930) and of the *Z* and *W* particles (predicted by Weinberg and others) were exceptions to this rule. More typical was the discovery in the 1950s of a subtle asymmetry in the behavior of certain particles that was not only unexpected but was thought to be prohibited by the known rules of physics.

"We need revolutionary ideas in accelerator design more than we need theory," Ting argues. "Most universities do not have accelerator courses. Without such a course, and an infusion of new ideas, the field will die." This would have been true even if the SSC had been built, he adds.

For some time, plans have been under way for a large linear accelerator that, while less powerful than the SSC, would provide more precise, "cleaner"



ELECTRON BEAM flashes through a chamber of plasma in a demonstration of the wake-field acceleration technique at Argonne National Laboratory.

The Crisis in Physics, 1904

Does the current state of physics represent the dark before the dawn? Modern physicists seeking support for this Panglossian view often turn to their field's history, which offers many examples of crises that were eventually overcome.

Perhaps the most dramatic such period occurred at the beginning of this century. The discoveries of x-rays, electrons, protons and other forms of radiation emanating from atoms had strained classical physics to the breaking point. So had experiments on light, which posed problems that commonsense notions about space and time could not resolve.

In 1904 a number of the most prominent physicists in the world gathered at the Congress of Arts and Science in St. Louis, Mo., to discuss the plight of physics. At the meeting, Ernest Rutherford, Henri Poincaré, Ludwig Boltzmann and other luminaries outlined the quandaries they faced and suggested possible resolutions.

Chen Ning Yang of the State University of New York at Stony Brook has read these speeches, which the American Institute of Physics printed in 1986 as volume five of its series *The History of Modern Physics.* "They all recognized there was a great crisis," observes Yang, who shared the 1957 Nobel Prize for having proposed an experiment that confirmed the existence of a subtle asymmetry, or "handedness," in certain particle interactions. "They lamented that everything they thought they understood was wrong."

Only a few months later, of course, Albert Einstein formulated his theory of special relativity, which resolved the paradoxes posed by the propagation of light with radical new conceptions of time and space. Within 20 years after that, physicists had developed quantum mechanics, which provided a strange but astonishingly accurate description of the behavior of atoms.

But Yang warns that the crisis of 1904 is very different from the present one. The conflict then derived from the inability of previously established ideas to explain experimental results. "That kind of conflict is good," Yang asserts. "We're in a different kind of trouble now." Although general relativity and quantum theories are in apparent conflict, no conceivable experiment can address that incompatibility directly.

"There are very few experiments" that can bring about significant progress, Yang comments, and "without checking, it becomes a shot in the dark." The field might also progress through some great advance in mathematics, but Yang notes that, historically, such advances have been rather rare. In any case, neither outcome is likely to occur within the next 20 years. The field of particle physics "is in trouble, deep trouble," Yang says.

data. The project, which involves a collaboration of physicists in the U.S., Europe and Asia, is the Next Linear Collider, or NLC. As currently conceived, the machine would consist of a tunnel some 20 kilometers long in which electrons and positrons would be rammed into one another head-on. Participants say they are trying to avoid the problems that doomed the SSC by making sure that the technological, political and financial foundations of the accelerator have been laid before construction proceeds. They emphasize, in particular, that the project will succeed only if it has international support.

Wholly different accelerator schemes are being explored at the University of California at Los Angeles, Argonne National Laboratory and elsewhere. A technique called plasma beat-wave acceleration involves sending a laser pulse through a chamber of oppositely charged particles, or plasma. The electromagnetic pulse creates a wave in the plasma on which electrons can "surf" to high energies. A similar approach, wake-field acceleration, generates the wave with a beam of electrons rather than a laser. James D. Simpson of Argonne estimates that it will take at least 15 to 20 years for these technologies to mature. "There are lots of problems to overcome," he says.

Georgi emphasizes that no conceivable machine will confirm grand or complete unification; at best, experiments can generate only circumstantial evidence. "We've been spoiled over the past few decades, because you get not just one [datum confirming a theory] but many, so you really know you're right," Georgi says. "Now I'm afraid we're going to have to be satisfied with data that convince only those who were already convinced." He adds, "It's a permanent problem; we're running up against a fundamental limit of nature."

Some physicists, particularly European ones, profess to see challenges rather than pitfalls ahead. "Crisis? No, I see no crisis," declares Ugo Amaldi of CERN. "Every day we need new approaches to build new images of nature." De Rújula believes physicists should be thankful they are not on the verge of finding a final theory. "Converting science into liturgy would be depressing," he comments.

Yet some of their American counterparts fear that the field, while awaiting some hoped-for breakthrough in either accelerators or mathematics, may disintegrate. Lederman, an indefatigable proponent of science education who still teaches undergraduate physics, recalls that since the cancellation of the SSC a number of his students have expressed concern. "They ask me, 'What does it mean?'" Lederman says. "I tell them I don't know." He broods a moment, then continues: "If I were a young guy, I'd go into brain science or chaos or computer science."

Glashow reveals that several of Harvard's most talented graduate students recently defected to Wall Street. "Goldman, Sachs loves theoretical physicists," he confides. Georgi notes that even before the SSC was terminated, the slumping economy and the influx of physicists from Eastern Europe had created a shortage of physics jobs in the U.S. "I don't understand why, but we still get fantastic young people entering the field," he comments. "Well, I do understand why, because the questions this field addresses are so interesting."

Faraday expressed as well as anyone why it is so difficult to abandon the hope that a single force rules nature. "If the hope should prove well-founded," the British scientist wrote, "how great and mighty and sublime in its hitherto unchangeable character is the force I am trying to deal with, and how large may be the new domain of knowledge that may be opened to the mind of man?"

FURTHER READING

DREAMS OF A FINAL THEORY. Steven Weinberg. Pantheon Books, 1992.

THE END OF PHYSICS: THE MYTH OF A UNIFIED THEORY. David Lindley. Basic Books, 1993.

THE GOD PARTICLE: IF THE UNIVERSE IS THE ANSWER, WHAT IS THE QUESTION? Leon Lederman and Dick Teresi. Houghton Mifflin Company, 1993.

Dr. Big Brother

Will health care reform make everyone's life an open book?

The Clinton administration's health plan should give every American access to medical care and protection from financial ruin. It would also eventually give the government some of the most intimate details of a person's private life. Without carefully crafted safeguards, the president's health security plan could become an instrument of Big Brother.

If the Clinton plan becomes law, a powerful seven-member National Health Board will sit down to work out the fine details for a medical information highway. This electronic health data network would over time enable regional health alliances, insurers and federal, state and local health officials and researchers to assemble and exchange information about each citizen covered (the plan will ultimately include everyone). Every American will also carry a plastic card with a patient identification number, which some privacy groups fear could turn into a de facto national identification card.

WALLY MAN NAMEE *Sygma*

NOT A NATIONAL ID CARD, say Clinton administration officials of the health card brandished by the president. Nevertheless, some privacy advocates have doubts.

Certainly, a health information network offers clear, even compelling, benefits. Through it clinicians could quickly retrieve information about patients from anywhere in the country. Redundant medical tests could thus be avoided. Health claims could be speeded. Corrected information could simultaneously be entered in many files dispersed throughout the network. Epidemiological studies could be performed rapidly.

Such ready access also means the information could more easily be abused. Whether authorized or not, the peruser of a data file could discover details of a patient's life history that go well beyond a list of past vaccinations. The records might include information about genetic testing, sexual orientation, drug use, sexually transmitted diseases, finances and other personal details.

"These data are important because health care information can influence decisions about a person's access to credit, admission to educational institutions and his or her ability to secure employment and obtain insurance," said Paula J. Bruening, an Office of Technology Assessment policy analyst, at a congressional hearing last November. "Inaccuracies in the information, or its improper disclosure," Bruening claimed, "can deny an individual access to these basic necessities of life, and can threaten their personal and financial well-being."

The Clinton administration is well aware of the threat to privacy that easy access to medical information poses. A provision of the health proposal—Title V of the Health Security Act—specifies that electronic records will not be available to employers or to others who are not directly involved with health care. Furthermore, during the program's early days the system would offer access only to insurance information, not patients' complete health records. Later, though, at least some clinical and administrative data would be added. The law also has teeth: civil penalties for unapproved disclosure of confidential information and criminal sanctions for misuse of the health card or a patient identification number.

Privacy advocates have applauded the broad outlines of the Clinton plan. Yet in the next breath, they worry about what has been left unwritten. Some protections would be in place at the time each person receives coverage under the new law, perhaps during 1996. But the National Health Board that will administer the plan would not then have to produce recommendations for comprehensive legislation concerning health care privacy until 1997. "The Clinton Act is a framework. It leaves everything up to the National Health Board," says Aimee Berenson, legislative counsel for the AIDS Action Council, an organization that has worked on protecting privacy of the medical records of AIDS patients. And, Bruening adds, "It's important to build safeguards and security at the beginning. It's more effective and less costly."

If a revamped health system relies on the Social Security number to identify each patient, the risk to personal privacy could be greatly amplified. To be sure, broadening the uses of these existing numbers would forestall the need to devise new personal identification codes. But an individual's Social Security number fails to meet the Clinton health plan's goal of giving every citizen a unique and private identifier.

Indeed, critics argue, employing the

Social Security number would make it that much easier to build a composite profile of how a person lives. Health information could be added to other data made accessible with the Social Security number. The number already identifies individuals in the records of many institutions, including utilities, collection agencies, credit-card issuers, stock brokerages and the Internal Revenue Service. "Why build what should be a tamper-resistant system on a system that is broken?" comments Janlori Goldman, head of the American Civil Liberties Union's Privacy and Technology Project.

"The issue is not what number do I use. The issue is what are the allowable linkages between different data sets using that number?" replies Joan Turek-Brezina, who chairs a Department of Health and Human Services task force that examines privacy issues surrounding health records. "People know the Social Security number, and it is the cheapest and fastest way to implement all of this."

The debate will become more acute as advances in technology further complicate the trade-off between privacy and efficiency. Clinton administration officials have said they will forgo the additional expense of fitting every American's wallet with a "smart" card, equipped with an electronic memory and microprocessor, or optical storage cards, which retain large amounts of data on a medium similar to that used for recording music on compact discs. In the administration's plan, a simple magnetic stripe on the back of each card can be swiped through a card reader. The health identification number encoded on the card will validate insurance coverage. That decision will avoid the inevitable debate over the privacy implications of storing the equivalent of hundreds of pages of text on a credit-card-size device.

Officials have not, however, ruled out the smart card as future health care reforms are introduced. The prospect leads David Banisar, a policy analyst at Computer Professionals for Social Responsibility, to worry about what he calls "function creep"—the temptation to add features once the national health network is in place. "It could start off as a benign ATM card," Banisar says. "If you add extra functions, it could be a real problem."

Functions may already have begun to creep. Entrepreneurs and health care providers throughout Europe and even in the U.S. are smartening cards. A California firm, LaserCard Systems, recently demonstrated to Congress a four-megabyte card with an optical memory that lets 2,000 pages of encrypted text or multiple images be written on it. In California the Loma Linda School of Medicine has used the card to store a photograph and an x-ray of a patient and the voice of a physician dictating diagnostic information.

Ultimately, social and economic pressure, the American Civil Liberties Union contends, may prove more dangerous than any lapse in a new technology. The Clinton health bill sets out general principles about patient rights to inspect and guard medical data. But most people have little choice about whether to let someone else see a record if such perusal is a condition of receiving benefits or a job. To remedy the problem, Goldman would like to see the establishment of "fire walls"—narrow definitions of the information that any corporation or institution can ask an individual to release.

The brave new world of computerized medical data has already inspired a number of legislative initiatives. Congress may try to bring order to the inconsistent mix of state laws that now govern access to medical records. One problem is that most of the state legislation enacted to date has failed to deal with access to health records by police, the courts, the media and other so-called secondary users. The Federal Privacy Act, a 1974 law, pertains only to medical information gathered by federal agencies—a Veterans Administration hospital, for example. Today access to medical records varies from state to state. "Records in one state may offer a reasonable level of protection, but if you carry them across the river you lose the protection," says Robert M. Gellman, an aide to Representative Gary Condit of California, who held hearings last November on a bill being written to cover health care privacy.

The Clinton health proposal calls for a 15-member privacy panel to advise the National Health Board. Advocacy groups want to broaden the debate to include more than just medical records and have welcomed a Privacy Protection Commission proposed by Senator Paul Simon of Illinois. Other legislators are also thinking about what a national health plan means for privacy. Besides the bill from Condit, Senator Christopher Bond of Missouri proposed a law last September that would regulate electronic health care records.

Privacy for medical records may finally have achieved the national standing it deserves. A poll conducted last year by Louis Harris & Associates for Equifax, a credit reporting agency, revealed that 85 percent of Americans attached great importance to protecting medical records under health care reform. Confidence in the Clinton plan may hinge on assuring people about the integrity of what is told to a physician in confidence. *—Gary Stix*

Extra! Extra!

Newspaper publishers reinvade cyberspace

There has never been a successful electronic newspaper. No one has been able to devise an information-age substitute for the Gutenberg product, that cellulose collage of Waco, Tex., Michael Jackson and Wall Street that serves as a semiotic footnote to Donna Karan, butterball turkeys and everyday low prices. Despite the thousands of hours spent crafting the software for graphical user interfaces, pull-down menus and keyword searching, no programming avatar has proved that the subsecond mouse click can provide random access that is as acceptable and convenient as a well-placed thumb burrowing toward last night's box scores.

No matter. The electronic newspaper is back. Publishers are once again taking a baby step toward sending the printing press the way of the Linotype and the green eyeshade. Among the big names involved in this round are Dow Jones, the Washington Post Company, Knight-Ridder and the Tribune Company.

As with previous expeditions into electronic publishing, the main motivator is sheer terror. True, electronic home delivery does not replace the paperboy. And so far newspapers have weathered the advent of radio, television and the cable channel infomercial. But now the merger frenzy in the telephone and cable television industry augurs the creation of the fiber-optic equivalent of the Channel Tunnel. Furthermore, the economic roots of the traditional newspaper are being starved by the disappearance of the downtown department store and the arrival of direct-marketing powerhouses that have whittled away at lucrative advertising inserts from supermarkets and other stores. An improvement in advertising expenditures during 1993 has not removed the apprehension that several previous years of declining or flat spending have instilled.

Circulation has weakened, too. The daily newspaper reached only 63 percent of households in 1993, compared with 98 percent in 1970. Twenty years before that, 124 percent of U.S. households took a daily newspaper, a feat accomplished by buying both a morning and afternoon publication.

Newspaper executives may not be sure what to do, but they know they can't just stand there like the proverbial deer in the headlights. Another reason publishers are looking at the on-line approach again is that the cost of such ventures is much less than it was 10 years ago. At that time, Knight-Ridder spent $50 million before it turned off the modems on its Viewtron videotext experiment. In fact, a Knight-Ridder newspaper, the *San Jose Mercury News,* may have become a role model for the current revival. Launched last May, the service employs the delivery vehicle of a commercial electronic information service, America Online.

Through the carrier, Mercury Center, as it is called, provides electronic text of current and back issues of the newspaper and a bulletin board, or electronic forum, that enables readers to talk back to the paper or with one another. Mercury Center also supplies information that cannot be found in the printed paper version—a presidential speech or a court document, for example.

With the highest per capita density of technophiles anywhere, Silicon Valley is probably a good place for Knight-Ridder to begin again in earnest. Newspaper editors, though, have trouble keeping themselves from seeing both sides of the story. "It takes longer, it's slower and it's visually unexciting; an electronic newspaper is not as good as the printed product yet," says Robert D. Ingle, the newspaper's editor. Ingle adds: "Eighty-five percent of the users are men. Many of them are nerds."

Some of those nerds have problems with the mainstream media, though. In an article in the new monthly *Wired,* a chronicle of cyberspace, Michael Crichton, the author of *Jurassic Park,* lobbed the epithet "dinosaur" at establishment bulwarks such as the *New York Times.* The piece was adapted from a speech that Crichton delivered at the National Press Club last year. In it Crichton rails: "And along with many other American industries, the American media produce a product of very poor quality. Its information is not reliable, it has too much chrome and glitz, its doors rattle, it breaks down almost immediately, and it's sold without warranty. It's flashy but it's basically junk. So people have begun to stop buying it." Crichton's corollary to this argument is that an electronic network like the Internet will furnish information without recourse to what he considers the superficiality and glibness of a Peter Jennings or a *New York Times.* These direct connections, Crichton says, already exist: C-SPAN, for one, with its gavel-to-gavel coverage of government proceedings.

"It was an interesting article, but I think he's crazy," says Mark Potts, who directs product development for the *Washington Post's* electronic media effort. Even with artificial-intelligence software to select items of personal interest, people do not wish to become research librarians. This fact should broaden the role of the news packager. "A congressional transcript is boring as hell," Potts says. "You need somebody to distill and cut through the garbage."

Yet the *Washington Post* has recently demonstrated PostCard, its idea of how to box and ribbon-tie information for the New Age reader. In the prototype, which runs on an Apple computer, a reader can call up a video clip of Supreme Court Justice Ruth Ginsburg's confirmation hearing, follow an animation of how the Mississippi flooded beyond its banks, file a letter to the editor, buy tickets to the Baltimore Orioles, search the classifieds by job title, track a personal stock portfolio or calculate how a change in the tax code might affect one's personal finances. A commercial version of the service may eventually become available over cable television as well.

By judicious use of information technology, even the product that relies on cellulose can help obtain the higher-quality information for which Crichton yearns. Frank Daniels III, the executive editor of the *News & Observer* in Raleigh, N.C., recently listed an Internet address just below the direct office telephone number that appears under his name on the masthead of the paper. It was Daniels who lobbied the publisher, his technology-averse father, for the $250,000-a-year infusion into the paper's research department. Aging stacks of paper clippings from the newspaper's morgue have been supplemented by more than 150 data bases for use by reporters. They include the Money Machine, nearly 70,000 records of campaign contributions dating back to 1987, street-by-street census maps, a legislative tracking system and a remote 24-hour connection to a selection of state and county records. Some of this information might eventually be packaged and resold to local businesses such as real estate and law firms.

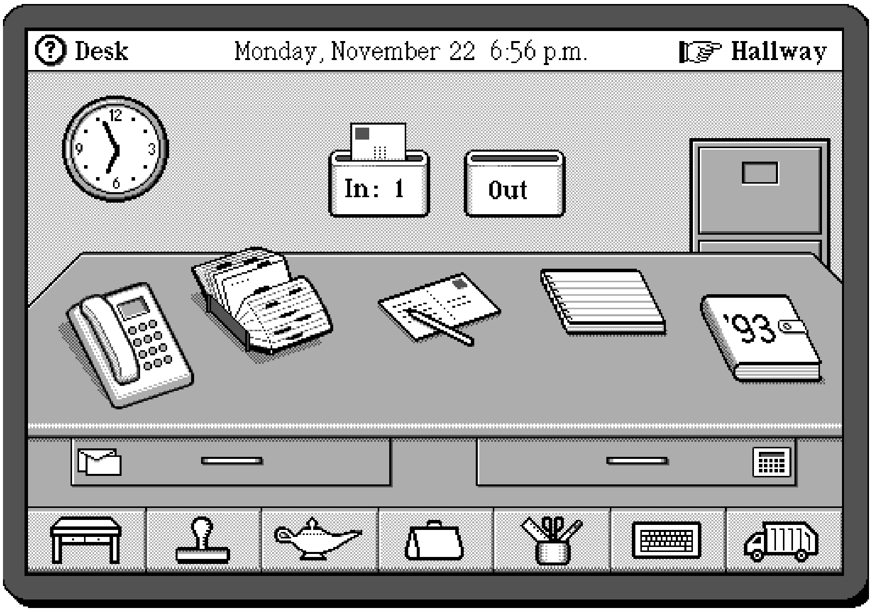
So-called computer-assisted journalism has paid off in producing a more incisive picture of government: political appointees who tried sneaking into lower-profile jobs when a new governor came into office and a bartender on the state payroll are two examples.

The *News & Observer* also sees its expanded role as an electronic gatekeeper for its readers. This year it is establishing a public access point for the Internet, and it is also beginning to provide the newspaper on-line.

Will such experiments open the path of evolution for the daily paper, or will innovation succumb to the iron rule of thumb? Watch this space for further developments. *—Gary Stix*

JOHN McGRAIL

POSTCARD, the* Washington Post*'s prototype of an electronic newspaper, is inspected by digital-media developer Mark Potts.



GENERAL MAGIC

HANDHELD COMPUTERS need friendlier software to bring cyberspace down to earth. General Magic's user interface gives applications a more mundane look.

Heads in the Cloud

Can programmers entice consumers into the net?

To the engineers who design digital superhighways, the symbol of a cloud represents far-flung, interconnected networks such as the Internet. In many ways, the metaphor is an apt one: global networks are gathering behind an emerging front that promises to liberate us from the tyranny of information yet threatens to inundate us with it. As entertainment, electronics, computer and telecommunications companies race to create and then dominate this new medium for sharing and selling information, they grope for the devices and uses that will draw customers into their clouds.

Meanwhile, back on earth, software writers and confectors of electronic gizmos are busily attempting to make the dream come true. They have invented a bewildering array of gadgets and programs to roll into market this year and next. Apple's Newton and AT&T/EO's Personal Communicator, released last year, were merely first drafts. IBM, Mitsubishi and BellSouth are readying for spring release a one-pound cellular phone/fax/modem with a computer in its handle. Motorola hopes to debut this quarter the first of a planned family of palm-top computers with built-in wireless fax, pager and electronic mail capability. Sony, Philips and Matsushita will follow with similar devices.

Handheld communicators are but one enticement. Coming soon to a television near you are a raft of new computer-based entertainment and information services that will flow through the cable and telephone lines that already penetrate most American homes. Such "interactive TV" could be a bonanza for Scientific Atlanta and for General Instruments. They have contracts to make decoder boxes that can send viewer commands back upstream.

Initially, interactive TV services will try to compete in established markets to meet proved demand. *TV Guide* and Tele-Communications, Inc. (TCI), are preparing to launch an onscreen version of the magazine sent weekly to 40 million homes. Aside from providing all the standard features of the printed guide (except the advertising), TV Guide On Screen will allow parents to automatically lock out shows based on their rating for violence, language or nudity. Voyeuristic couch potatoes can use the system to lock the shows in.

Several other interactive television systems will provide movies on demand in the hope of plugging into the $12-billion video rental market. Not all will use cable as a pipeline. Bell Atlantic is scaling up its StarGazer system for a field trial this spring. StarGazer uses special transmission equipment and sophisticated video compression techniques to store digitized movies and mail-order catalogues on computer servers and pump them through ordinary telephone lines into homes.

Program guides, rental movies and shopping channels are just the first steps in a strategy to hoist a new industry by its bootstraps. "Most new media start off recapitulating the old media," observes Nathan P. Myhrvold, senior vice president for advanced technology at Microsoft (which has its own interactive television software). "Later on they start doing the really cool stuff."

Myhrvold's wisdom is as conventional as it is true. The competitor who establishes the dominant technology wins. Bell Atlantic, *TV Guide,* Time Warner and every other contender thus seek to build the pier at which the as-yet-unbuilt fleets of video information and entertainment services will come to dock. Bell Atlantic designed StarGazer from the beginning to be open to new services. As a common carrier, it must offer "video dial tone" service to any company that asks for it. When the system is completed—probably not until 1996—customers will be able to steer their television sets through a "virtual" shopping mall. "You might go into the EchoTV store to get the six o'clock news that you happened to miss or into the MGM store to select any MGM movie," muses spokesperson Missy McTamny.

"As much as we consume media, we spend more time communicating," Myhrvold remarks, pointing to another carrot with which the industry will entice consumers into the net. Motorola has already ridden to record profits the fourfold increase in pager transmissions and 20-fold growth in cellular subscribers since 1985. This despite a general slowdown in consumer electronics sales that is estimated to have cut the profits of Sony, Sharp and Toshiba by 8 to 30 percent last year.

To computer executives, who face commoditization of their own industry, the path is clear: produce handheld, computerized communications devices and software that actually makes them useful. Apple's Newton, designed as a "digital assistant," met with chilly reviews in the trade press, which criticized it for not focusing enough on communications. While Apple recoups, two other visions—that of the giant Microsoft and that of the tiny Silicon Valley start-up General Magic—will probably rise as top contenders in what is sure to be a messy struggle.

Almost everyone in the industry agrees that if handheld devices are ever to capture the interest and dollars of the mass market, they must be substantially easier to use than desktop computers. Almost everyone but Myhrvold. "We've become inured to the fact that incredibly popular consumer devices are very hard to use," he says. "Consider a movie. You have to stand in line, it starts with or without you, and it's expensive. We go to movies because they make us laugh or cry, not because they are convenient. The trick is to make the ratio of how compelling something is to how hard it is to use greater than one."

Microsoft's idea of compelling starts in the workplace, so the company calls its new platform Microsoft At Work. "We thought the best thing is to put our software into fax machines and copiers and telephones—all in a way that integrates super-well with your desktop PC," Myhrvold explains. At Work also lies at the heart of WinPad, the operating system Microsoft is designing to run on devices built by Compaq and Motorola.

General Magic would like such evolution to look more like a revolution. It has decided to go straight to the consumer. Without benefit of Microsoft's incumbency or Goliath proportions, General Magic may seem an unlikely David. But a gang of big brothers—Apple, AT&T, Matsushita, Motorola, Philips and Sony—has invested in the company and licensed its technology. "We think of ourselves as a *keiretsu,*" says Marc Porat, General Magic's chief executive officer and chairman.

General Magic hopes it can seduce consumers as well as gadgeteers by making its Magic Cap operating system more adept at communications and much friendlier than the product's competitors. "Abstraction is the enemy," says Andy Hertzfeld, whose official title at General Magic is Software Wizard. "Magic Cap goes further down the road of making computing tangible."

Whereas on a Macintosh, say, folders lie on a plane called the desktop and open to reveal a window containing files—badly mixed metaphors—Magic Cap tries to emulate wherever possible the way things look and work in the real world. The desktop *looks* like a desk, on which sit a telephone, a rolodex, a notepad and other items that do just what you would expect them to. Away from the desk one can wander around a corridor, opening doors to other applications. Or leave the building altogether and walk down a street on which sit various buildings, each representing an information service.

Making electronic mail more like the real thing poses a larger challenge. To most of the world, e-mail means text—and nothing else. So General Magic humbly plans to change the world. It has developed an interpreted programming language, called Telescript, that allows a message to include images, sounds and even instructions as to how it should behave. A message could schedule an airline flight, for instance, and then monitor the flight and notify you of any changes. Magic Cap has Telescript built in, and all Magic Cap devices (the first of which is the Motorola product) will come bundled with AT&T's MagicMail. But that is, so far, the only network that has licensed Telescript, although Porat promises that "we will license Telescript to anyone," even Magic Cap's competitors.

The competitors may have other ideas. "The notion that their programming language and the limited features it has will be the only way to do things is completely wrong," Myhrvold fumes. Although Telescript would make it much easier for programmers to develop the "really cool stuff" that will lure customers into this new marketplace among the clouds, Myhrvold hints that Microsoft will be proposing alternatives. "I've got 400 people and $100 million a year working on this stuff. It may not generate significant revenue for three to five years. But it sure is a hell of an opportunity." —*W. Wayt Gibbs*

Pipe Dream

A consortium ponders remaking the automobile

The White House has called for a transformation of the machine that epitomizes the American dream—and also consumes billions of gallons of gasoline while spewing a mix of pollutants that generates up to half the ozone that engulfs major cities.

This government collaboration with the automobile industry, referred to informally as the clean-car initiative, would come up with a vehicle having three times better fuel efficiency. It will also focus on nonpolluting substitutes or improvements for the internal-combustion engine, advanced materials, as well as better manufacturing methods and technologies that enhance automobile safety. The roster of government participants includes the departments of energy, defense, commerce and transportation, the Environmental Protection Agency, the National Aeronautics and Space Administration and the National Science Foundation. They are joined by the Big Three automakers.

Will it fly—or roll? No, says Roger G. Noll, a professor of economics at Stanford University and co-author of a book on government-supported research and development, *The Technology Pork Barrel*. He maintains that the consortium has the hallmarks of many large-scale commercial research and development projects that have fallen short of their original goals. "The clean-car consortium strikes me as fitting right into the mold," Noll says. "It's got two fundamental flaws: it eliminates the possibility of people competing against one another for what could be the emerging market of the 21st century, and if it turns out it doesn't work, it will be very difficult to kill the program."

The White House remains undeterred by naysayers. "This is a clear and finite goal with a deadline," counters Henry C. Kelly, assistant director of technology for the White House Office of Science and Technology Policy. "If the feds are wasting money, the industry is sitting there alongside us sharing the costs." Cooperation is needed, Kelly says, because of the expense and risk involved. "It's too big for anyone to undertake on their own," he states.

Whatever the prognosis for a cornucopia of new technology, the clean-car initiative promises immediate institutional benefits for the administration, the national laboratories and the automakers. For the administration, there is reassurance to entrepreneurs, managers and policy mandarins who had begun to wonder what had become of an ambitious commitment to technology policy, unveiled a year ago.

For the federal laboratories, the clean-car initiative provides a program that, in concept, if not dollars, is comparable to the nuclear energy and weapons endeavors that occupied them for half a century. A total dollar figure has yet to be set, although the federal outlays will primarily come from shifting monies away from existing defense programs, not from new appropriations.

The decade-long program could also become a litmus test of the yet unproved boast that technology fostered in high-cost government research can transform the ultimate consumer item, the family car. Can an ultracapacitor developed for the Star Wars program be used to store energy to power a four-door sedan?

For the automobile industry, the program has served as a chance to demonstrate that Detroit executives are good corporate citizens. It did take some last-minute, one-on-one persuasion from Vice President Al Gore to garner industry executives' acceptance. Even at the White House ceremony, industry chiefs ballasted optimism with caution. "The difficulties and obstacles confronting this revolution should not be underestimated, and our expectations should be realistic," noted John F. Smith, Jr., General Motors's president, in a speech.

Detroit automakers also see no inconsistency between their commitment to the project's long-term goal and their continuing legal and lobbying efforts to soften the nation's most far-reaching clean-air laws. "We're trying to maximize the resources we have to try to solve these technology problems," says Gerald A. Esper, director of the vehicle environment department for the American Automobile Manufacturers Association (AAMA), the lobbying and trade organization for the Big Three. "At the same time, the manufacturers feel electric-vehicle technology is not available today. It would be irresponsible not to make that known."

The AAMA is contemplating a campaign in coming months to publicize its misgivings about a California regulation that is essentially a mandate for electric vehicles. The zero-emission vehicle regulation requires that 2 percent of cars sold by 1998 by major companies emit no pollutants, a figure that rises to 10 percent of cars for all manufacturers by 2003.

The automakers have also been fighting in the courts the efforts of New York State, Maine and Massachusetts to impose California-like emission regulations, which, in effect, require electric vehicles. Such rules are also being contemplated by nine other northeastern states. In early December the AAMA forwarded a counterproposal for a vehicle that would exceed federal emission standards while using reformulated gasolines. Any mention of electric vehicles was conspicuously absent.

Much of the industry's resistance to the California standard stems from the belief that no electric-car battery can be found in the short term to meet it. In a letter to the office of California governor Pete Wilson, Ford's vice chairman, Alan D. Gilmour, noted: "While we have designed and are now starting to build a good electric vehicle, the Ecostar minivan, there is no battery with acceptable cost, range and life to power it." Meanwhile the California regulation has inspired an explosion of activity among companies small and large that stretches well beyond the greater Detroit metropolitan area. "The car industry is bigger than Detroit," says Bill Sessa, a spokesman for California's Air Resources Board, which set the state's zero-emission standard. "Asia and Europe and many entrepreneurial companies are involved in this research."

A one-time Lockheed aircraft warehouse is the headquarters for Calstart, an agglomeration of 80 public agencies, utilities and other businesses that are adapting defense technology to electric vehicles as well as developing fuel cells, flywheels, automotive electronics and other components. (General Motors is a minor participant.)

Calstart has demonstrated a prototype containing an advanced lead-acid battery, a lightweight aluminum frame and other technology that consortium members believe could lead to an affordable electric vehicle. Countless garage shops are also laboring on components for electric cars. "We're just one example of literally thousands of flowers blooming," says Michael J. Gage, Calstart's president.

Similar dynamism is lacking in the U.S. Advanced Battery Consortium. The $260-million consortium, which involves the Big Three, the Department of Energy, utilities and battery manufacturers, has taken about two years longer than anticipated to complete some of the contracts to develop advanced battery technology for electric vehicles. It will not meet its original mid-1994 goal of a new-generation battery that can readily be slotted into electric vehicles able to meet the 1998 California standard.

That leaves the opportunity ripe for Calstart and the garage gang. If Adam Smith were betting, that's where his long green ones would go. —*Gary Stix*

A Blade of Grass

Social history often returns to the theme of the lawn versus crabgrass. Properly nurtured, suburban sod serves as a symbol of gentility and social superiority. Dominance over the encroachments of chickweed, partridgeberries, plantains, rushes and innumerable other undesirable forms of vegetation proclaims the orderliness, the rightness of an existence distanced from the rigors of urban living.

F. Herbert Bormann, an emeritus professor of forest ecology at the Yale University School of Forestry and Environmental Studies and a leading lawn revisionist, reads a different message. To him, the lawn is a metaphor for environmental excess. Several years ago Bormann, who, with his colleague Gene E. Likens, is credited with discovering that acid rain was falling on northeastern forests, turned the lawn into a mission for 11 Yale graduate students. The project unearthed some startling facts. "In area, grass is the largest single crop in the U.S.," Bormann says. If aggregated, American lawns would cover 25 million acres, an area equivalent to that occupied by the state of Pennsylvania.

LISA VERNEGAARD

FREEDOM LAWN replaces a grass monoculture with a herbicide-free jumble of weeds and grass.

Grass is the basis for a $25-billion-a-year lawn care industry. Suburban greensward also guzzles often scarce water resources and uses as much as 10 times more chemical pesticides per acre than does an equivalent area of cropland. U.S. lawns absorb more synthetic fertilizers than India applies to all its food crops. In the mindless hour of back-and-forth traversals, the power mower spews pollutants equivalent to a car driving 350 miles.

The particulars in this bill of indictment appear in a book published last year by Yale University Press called *Redesigning the American Lawn: A Search for Environmental Harmony.* Bormann co-authored the volume with two colleagues from the Yale School of Forestry, Diana Balmori and Gordon T. Geballe.

Bormann has a ready answer for those who would like to strike a blow for the environment or just feel virtuously green while enjoying hours not spent behind a mower: grow a Bormann Freedom Lawn. It's easy. Just stand back and let the crabgrass, chickweed, wood sorrel and plantains, or, for Texans, maybe even the buffalo grass, flourish. —*Gary Stix*

THE ANALYTICAL ECONOMIST

Paying for Light at the End of the Chunnel

Is the private sector more efficient than government in financing and executing major construction projects? The Channel Tunnel, due to open this spring, provides the most recent and prominent case in point for economists, government officials, business people—and the public.

The lessons that financiers, managers and administrators draw from the Chunnel deal will affect economies throughout the world. For better or worse, officials trying to balance shrinking tax bases against demands for infrastructure are increasingly enticed by the zesty Thatcher-Reagan tonic of private-sector mechanisms and their presumed inherent efficiencies. “Governments of all hues will continue to look for new ways of incorporating private-sector funds for public works projects,” says Graham Corbett, financial director of Eurotunnel, the London-based consortium of banks and investment firms that financed the tunnel.

Britain’s insistence on using none of its taxpayers’ money to build the undersea road and railway set up this particular experiment of nature. Without government backing, the project fell to a network of participants—bankers, contractors, operators and shareholders—a kind of metacorporation.

The decision to go private had two immediate consequences. First, private concern with profit exposed the true costs and risks of such a large-scale enterprise. Usually bureaucratic fog and public inattention obscure the outlays in projects bankrolled by taxpayers. And the government’s ability to run on less than empty can push payback horizons out to the vanishing point.

But private-sector firms do not enjoy such luxury. Once the extent of the financial exposure became clear, aversion to risk led to the second consequence: investors protected themselves by handing risk off to others. To reduce the uncertainty of the project, the British and French governments had granted the contractors a “concession” for the tunnel. A concession (a concept dating back to the Middle Ages) gives a firm or other entity exclusive rights to develop and exploit a certain geographic or economic territory. But even this did not mollify the consortium of banks and construction companies.

Instead they apparently decided early on to get out from under the project after its completion. One of their first acts was to form a separate entity—Eurotunnel—to assume all long-term responsibility for the tunnel’s financial success, while keeping the short-term risks to themselves. Once Eurotunnel had been formed and shares sold to investment firms, the construction companies, under the TransManche Link (TML) umbrella, had merely to build the tunnel in return for a paycheck in the end. The banks, with signed credit agreements in hand, had only to ensure that they could collect on the debts that Eurotunnel owed. Putting up just $75 million of the seed capital, the founding consortium moved quickly to raise an additional $300 million from investment firms and a syndication of banks. But once it had done so, its equity—and its share of the long-term risks—fell to less than 10 percent.

Eurotunnel, responsible for the ultimate financial success of the venture, had to cope with flaws in the project that made disagreements with TML inevitable. Before Eurotunnel was formed, the construction firms had drawn up their own contracts to represent the interests of investors and shareholders. Not surprisingly, the builders failed to include penalty clauses, incentive pay and other provisions that give such agreements teeth. “The builders had written themselves an open-ended contract, and construction was getting out of hand,” says an executive at one of the U.K. banks in the original consortium. Worried about the project’s credit worthiness, the banks threw their support behind Eurotunnel, which quickly renegotiated the contracts. By then, however, construction was far behind schedule and had overrun the budget. Costs ultimately more than doubled—from $6 billion to $13.4 billion.

Of even more lasting impact was the early and ambitious decision by the founding consortium to hasten construction of the tunnel in order to complete it within seven years. To accomplish this plan, they broke ground immediately after winning the concession. While most of the tunnel’s rail and signaling systems and trains were in the preliminary stages of design, seed money was used to buy digging equipment. “Contractors were too happy to rush off with their spades, and the banks were too happy to rush off with their checkbooks, because they all were eager to make profits,” says Roger Vickerman, professor of regional and transport economics at the University of Kent at Canterbury.

The repercussions of this rush to start construction were felt only after Eurotunnel and TML had gone their separate ways. Many lapses in design became evident: shafts on the English side were too narrow to carry parts for the tunnel-boring machines, tunnel-wall linings leaked, train carriages did not meet safety requirements. Such problems had enormous financial implications and precipitated confrontations between Eurotunnel and TML over who was responsible.

Design details had not been set in the beginning, and so the contracts between Eurotunnel and TML (which declined interviews) specified performance criteria rather than precise blueprints and construction methods. This oversight left gray areas that became the focus of protracted disputes. For instance, when design limitations of the tunnel were found to make achieving anything close to the 187-mile-per-hour French train speed exorbitant, both parties agreed to lower the speed to 100 mph after costly negotiations.

Who will provide financial light at the end of the Chunnel? Presumably countless carloads and trainloads of Ians and Maudes and Alains and Maries, as well as citizens of the other nations of the European Community—in other words, the public. If the Chunnel’s customers do not pay enough to make the link a going proposition, the costs will land on other members of the public—those who have invested so far and those who will buy.

Amid these tribulations, the stock market renders its running judgment on the soundness of the Chunnel. At press time, shares were trading at about $7, suggesting that the market values the company at about $3.75 billion. The viability of the investment will become clear in the next few decades. “By February we believe they’ll be worth about £5.70 [$8.50],” says an analyst at a London brokerage firm. “Beyond that is anybody’s guess.”—*Fred Guterl, London*

Making a Mirror by Spinning a Liquid

When a liquid is spun in a container, the surface of the liquid does not remain flat but takes on a concave, parabolic shape. The parabola, of course, is the basis for the primary mirrors in optical telescopes because the shape will focus parallel light rays to a point. Investigators have exploited the rotational effect in fluids to make large telescope mirrors from liquid metals, primarily mercury. The parabolic shape remains as long as the liquid is kept spinning. Liquid mirrors made from mercury have been shown to have an optical quality near their ground-glass cousins, while having the advantage of being lighter and less expensive [see "Liquid Mirrors," by Ermanno F. Borra, page 68].

A simpler variation of this procedure, which can be done at home, is to use a liquid that hardens, so that the parabolic surface is preserved after the spinning has ceased. Epoxy fits the bill. Consisting of a mixture of a resin and a hardener, epoxy slowly cures to a solid with little change in shape. The resulting mirror cannot compete against aluminized glass mirrors for image quality; it is perhaps best suited as a collector of light at the infrared and microwave end of the spectrum. We have relied on this method to produce an inexpensive mirror the size of a satellite dish to be used for observations of the cosmic microwave background radiation.

A turntable from a stereo system is the ideal device on which to spin the liquid. Be sure to remove the phonograph cartridge. The record player does not have to be of the highest quality—we have found that minor variations in turntable speed do not seriously affect the mirror. Those who have switched to compact discs should be able to find a secondhand turntable without too much trouble or expense.

The container to hold the spinning epoxy can be anything, even a common dinner plate, but we like to use food storage containers made of polyethylene or polypropylene. Such plastic containers are practical because the epoxy does not stick to the surface very well. The spun mirror pops out easily after the epoxy has cured. The container must have enough depth so the liquid does not spill over the sides during rotation. It can be any diameter so long as it can spin freely on the turntable. Be aware, however, that a majority of turntables operate only when the tonearm is cued over the edge of the platter. To provide clearance for the tonearm, you may need to choose a container that has a diameter slightly less than that of the platter.

The best epoxy is one that has the lowest viscosity and longest setting time. (Viscosity is the thickness or runniness of a liquid.) The epoxy should flow at least as well as warm maple syrup. The epoxies we tried were type CRS-107 from Custom Resin Systems in Carver, Mass., and Stycast 1266 clear epoxy from Emerson & Cuming in Canton, Mass. These two varieties are by no means the best or only choices. The number of epoxies available is truly astounding, so check the local hardware store or look under "epoxy" or "resins" in the yellow pages for a suitable one. You will probably need no more than about 0.1 liter (several fluid ounces) for each mirror.

As for setting time, we recommend choosing epoxies that have a one- or two-day cure period if you plan to spin mirrors larger than about 30 centimeters in diameter. Smaller mirrors should be made from epoxies that cure in no less than four to six hours. Do not use the quick-drying, "five-minute" epoxy.

Epoxy usually comes in two separate containers: one for the resin and the other for the hardener. They must be thoroughly mixed so that the resulting liquid is uniform. Follow the package instructions to determine the right proportions.

Epoxy is generally safe to use, but some caution should be exercised. The chemicals involved are rather nasty; gloves and good ventilation are required. Also, you may feel some warmth coming from the epoxy because the chemicals give off heat as they cure. If you intend to spin several mirrors and want to prepare large amounts, mix the epoxy in batches of no more than a liter at a time and keep the batches separate. It is possible that larger batches could smoke or even catch fire. (The fast-drying epoxies are the most susceptible types, one of the reasons to avoid them.)

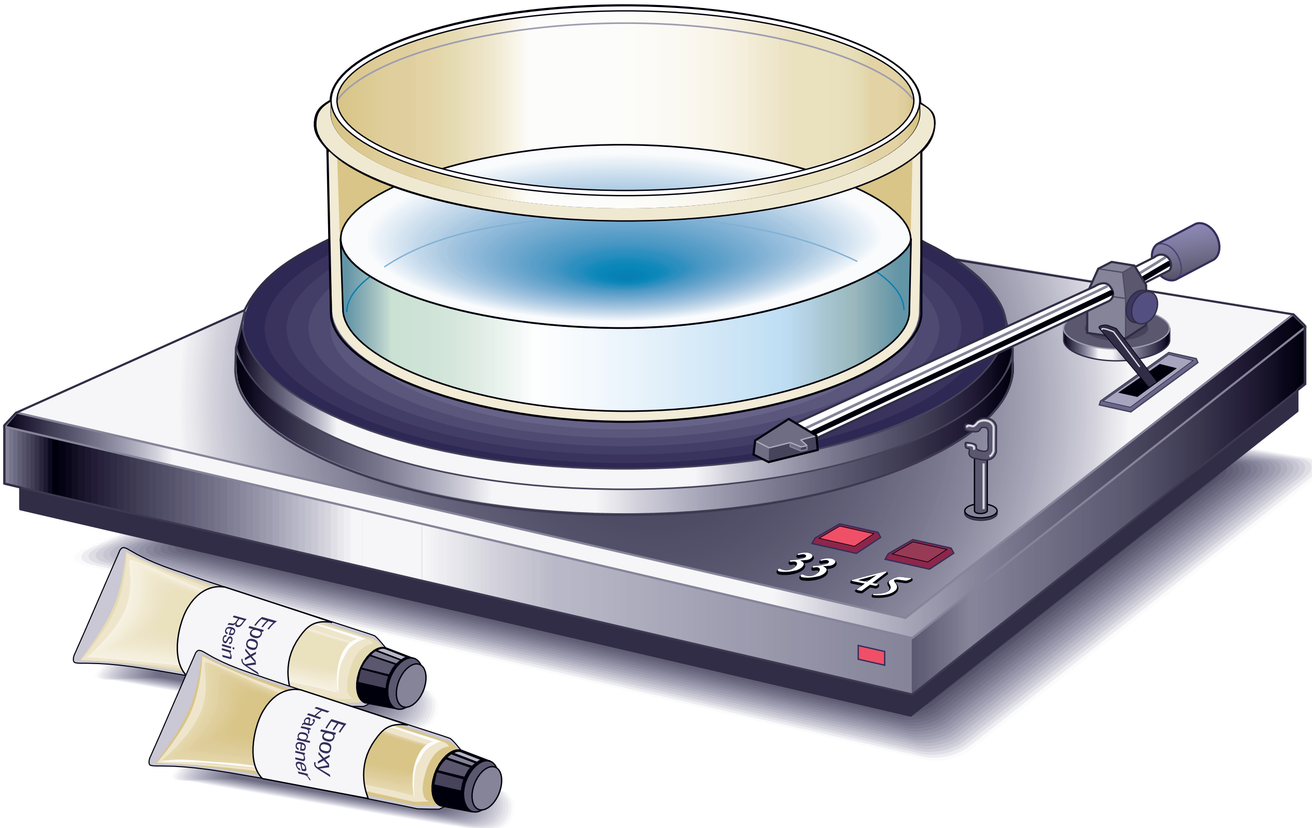
To determine exactly how much epoxy to use, spin some water in the container first. Pour in just enough so that it wets the container but does not spill out during rotation. Make sure the container is centered on the platter. If necessary, adjust the tilt of the turntable so that the water spreads uniformly across the surface. Measure the amount of water used and mix that much epoxy. Be sure to dry the inside of the container completely before pouring in the epoxy, because moisture prevents many epoxies from curing properly.

The turntable must spin until the epoxy has cured. It is important to use a "windscreen," that is, a plastic wrap over the spinning liquid. The breeze generated by the rotation can cause ripples to form on the surface. If you do find defects in the dried epoxy, just repeat the procedure with another coat. In fact, we have found that two thin coats produce a smoother surface than does one thick coat of the same volume.

The cured epoxy has the stiffness of a hard piece of plastic, like an acrylic. The focal length will be proportional to the square of the rotation period. A turntable speed of 45 revolutions per minute (rpm) will yield a focal length of about 22.1 centimeters. A speed of 33 rpm gives 41.1 centimeters. A heavy container may make the turntable speed somewhat inconsistent and affect the focal length.

The surface can be made reflective through chemical silvering, a technique popularized by amateur telescope makers. You can ship your cured epoxy to a silver-plating company. The prices such shops charge vary greatly, ranging from a few dollars to about $70 per square foot. Alternatively, you can do it your-

MARK DRAGOVAN and DON ALVAREZ collaborate at Princeton University. An assistant professor of physics there, Dragovan received a Presidential Young Investigator Award in 1989. His main interest is in observing anisotropy in the cosmic microwave background radiation. Alvarez is a graduate student in physics who currently holds a University Research Board Fellowship. He was a National Defense Science and Engineering Graduate Fellow from 1991 to 1993.



EPOXY MIRROR is created by spinning a mixture of resin and hardener in a plastic food storage container. The rotation causes the surface of the liquid to take on a concave, parabolic shape. Once cured, the epoxy can be made reflective through chemical silvering. The speed of the turntable determines the focal length: the faster the rotation, the shorter the focus.

self. Peacock Laboratories in Philadelphia offers the chemicals and detailed instructions, although the kits may be more appropriate to group projects. Each kit sells for just under $300 and provides enough material to coat 300 to 500 square feet of mirror. Like epoxy, these chemicals are somewhat noxious, so gloves and good ventilation are again recommended. The single most important requirement in obtaining a good silvering is a clean surface. Follow the cleaning instructions given in the kit; you will need distilled and demineralized water for the final rinse.

Note that some epoxies are easier to silverplate than others. If at first you don't succeed, try another epoxy. Of the two epoxies we mentioned, the Stycast was the most amenable. If you do not wish to take the trouble of silvering, you can try aluminized tape, which is quite reflective and is easy to apply.

Your silvered epoxy dish can serve as the primary mirror in a Newtonian reflector telescope. You will need a small, flat secondary mirror to angle the light rays out of the tube. The images formed by the spun mirror are reasonably good, although an amateur telescope maker would find them barely acceptable. The spun mirror works better as an infrared light collector. With a moderately sized spun mirror pointed at the sun, paper or wood can be ignited instantly at the focal point. You can record infrared images by coupling the system to an infrared camera [see "Seeing Infrared," by Donald G. Mooney, "The Amateur Scientist"; SCIENTIFIC AMERICAN, March 1992] or to an ordinary camera with infrared film.

The spun-mirror technique lends itself to several variations. You can create an "off-axis" parabolic mirror by placing the container slightly off the center of the turntable platter. The result is a mirror that is sensitive to microwave radiation. Materials other than epoxy, including urethanes, acrylics and silicone compounds, can yield good parabolic surfaces as well. To vary the focal length, slow down the turntable or experiment with other types of rotational devices.

Suppliers

For epoxy:

Custom Resin Systems
66 Main Street
Carver, MA 02330

Emerson & Cuming
869 Washington Street
Canton, MA 02021

For silver-plating chemicals:

Peacock Laboratories
54th Street and Paschall Avenue
Philadelphia, PA 19143

FURTHER READING

AMATEUR TELESCOPE MAKING. Fourth edition. Albert G. Ingalls. Scientific American Publishing Company, 1935.

PROCEDURES IN EXPERIMENTAL PHYSICS. John D. Strong. Prentice-Hall, 1938.

OPTICAL SHOP TESTING. Edited by D. Malacara. John Wiley & Sons, 1992.

LARGE OFF-AXIS EPOXY PARABOLOIDS FOR MILLIMETRIC TELESCOPES AND OPTICAL LIGHT COLLECTORS. Donald L. Alvarez, Mark Dragovan and Giles Novak in *Review of Scientific Instruments*, Vol. 64, No. 1, pages 261–262; January 1993.

BOOK REVIEWS by Philip Morrison

Old Blue Eyes

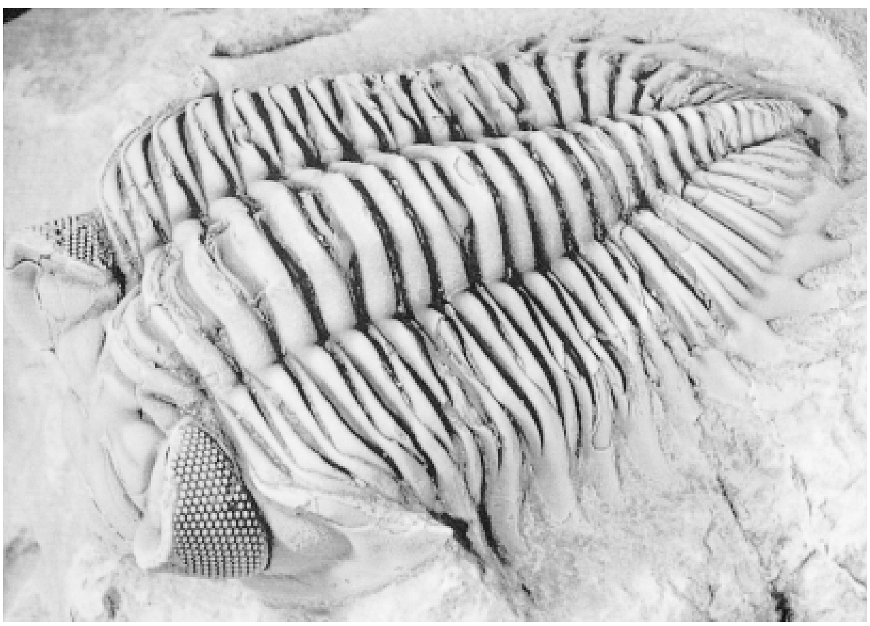
TRILOBITES, by Riccardo Levi-Setti. Second edition. University of Chicago Press, 1993 ($45).

It is nearly 20 years since this book first appeared. It then bore the subtitle: *A Photographic Atlas.* Most pages show in thrilling black-and-white images the indelible forms of trifid creatures, sampled by chance from ancient shallow seas, whether single or in heaps, molted carapaces or entire little beasts. Some are eyeless; most are Argus-eyed. Some are shown by x-ray or under the SEM. All are relics of an abundant, complex and enduring kind of life. The bulk of the photographs are the author's own, although he has been helped gloriously by his collector friends. (The original surface colors are lost; the bright yellows and reds reported for some specimens are only coatings of iron minerals.)

In 1975 this reviewer was won by the whole spectacle and wrote that the book was "above all to delight the eye." In this new edition the subtitle has gone, but delight is enhanced; there are half again as many plates as before. Many of them are recent Moroccan finds, brilliant with detail disclosed by modern air-abrasive techniques of preparation.

Professor Levi-Setti is both an active experimental physicist and a lover of trilobites. In the early edition he announced his remarkable finding: each of the multiple eyes of certain trilobites contains a tiny optical correction plate, its section shaped like an Asian compound bow. That particular fourth-order curve was first computed by Descartes and by Huygens centuries back, in total ignorance of trilobites but in an eagerness to correct the intrinsic focusing errors for fast lenses with spherical surfaces. The effect is a contrast increase of about fivefold.

The trilobites evolved just the right correctors. The curves now measured in those minuscule fossil eyes have the needed form, and bend just as they should, to fit the refractive index of the optical materials—calcite for the outer cornealike surface and protein-loaded chitin for the inner corrector. Both materials are designed to work in a watery medium. It was that early biochemical patent, the ability to load their tough chitin integument with hard calcite crystals, that made the trilobites the first common hard-shelled animals. Hence, their conspicuous mark in the rocks from the early Cambrian to the time of the reptiles, 300 million years later.



LARGE COMPOSITE EYES allowed trilobites to see their immediate environment. This feat marks the first use of optics coupled with sensory perception in nature. Shown is a **Traveropyge** *species, with prominent eyes, from Devonian deposits in Morocco. (The specimen is three times actual size.)*

This physicist with a geologist's hammer joyfully reports an amazing new result. Genetic design of fast lenses did not end with the last of the trilobite stock. Marine biologists have found in the cold Antarctic waters a number of living species of isopods, small crustaceans that also sport Descartes-Huygens correction plates behind each one of their compound eyes. The whole animal, no close relative of the trilobite, rather resembles one. Witness the fact that the discoverer of the first of these, James Eights, collecting in the South Shetlands in the 1830s, named his isopod *Brongniartia trilobitoides.*

Levi-Setti has looked for calcite within those eyes; it is absent. The optical materials of these more modern eyes are as proteinaceous as our own. But the tale is richer yet. The familiar (and delicious) scallop swims fleetly, its two valves aflutter. In every scallop's mantle, 50 or 60 simple eyes are embedded, like so many diminutive, bright sapphires. The scallop sees with top-drawer fast optics, at a hard-to-match f-number of about 0.6! Its eye is a corrected catadioptric camera, something similar to a Schmidt camera. A recurved Huygensian external correcting lens feeds through an iris into an interferometric spherical mirror within, made of spaced layers of guanine crystals. The mirror image is focused on a close pair of concentric spherical retinal surfaces. One of the retinas responds only to an "off" signal; the other, only to an "on." This dedicated detector gives the scallop extraordinary sensitivity to any dimming of ambient light by a predator's moving shadow. The corrected optics does not contribute to any well-formed overall image. Rather its concentrated light spot provides a high signal-to-noise ratio for photometry in dim light. Probably the computing power of these small brains is too small to justify detailed imaging but is quite enough to repay close attention to moving shadows.

The wary scallop is widely known for its bright blue eyes. Naturalist Eights also commented on his isopod's eyes: they were "elevated and prominent… an infinite number of facets, distinctly visible to the naked eye: color blue."

What color do you suppose the corrected fast eyes of trilobites were?

The conjecture is irresistible. The physicist has given his collecting self a most precious gift, evidence for the living color of turreted eyes long turned to stone. A reader shares that gift, and many another, like the transient match between the giant trilobites of Newfoundland and those of the Anti-Atlas Mountains overseas, a pulse of evolutionary change while continents recede.

What's in Store

NUCLEAR RENEWAL: COMMON SENSE ABOUT ENERGY, by Richard Rhodes. Viking Penguin, 1993 ($17.50).

This traveler is informed and articulate, his eye on the telling detail. "A shielded robot-transfer cask, painted bright red," swallows a canned radioactive black glass slug, about the size of a shot glass, and moves it off to the storage hall, without exciting the monitor that clicks away at background rate. The wide storage floor, as Cartesian as its planners, is "patterned with red and green manhole-size concrete plugs," each sealing one deep borehole wherein vitrified high-level nuclear waste is buried. This test site stores the reduced wastes from the first decade of French reactors; a single vitrified slug holds all the fission product (long-lived plutonium is removed) of a lifetime of nuclear electric power for a French family of four. A much larger permanent storage site has just been opened overlooking the English Channel. The French, at least, know how to store nuclear waste over the centuries of its decay time, though not on the cheap.

Rhodes visits next one of the cluster of 15 power reactors that dot the little bays of western Japan. He dons a yellow hard hat and enters, setting off the entry alarms. The American visitor at 190 pounds is too much for the weight-sensitive trigger meant to disclose a second entrant sneaking past on a single legitimate pass card. The account evokes the culture of Japanese industry, that tireless attention to detail that pays off in the touchy task of generating fission power.

The reactor building is spotless, everywhere color codes and labels. A printout of personal radiation dosage given monthly to every employee shows an average exposure that declines year by year even as the number of reactors doubles. High morale and high training, scrupulous maintenance and disciplined management allow operation of Japanese reactors of American design with only one tenth as many unplanned shutdowns as in the less economical American experience.

The U.S. has more than 100 power reactors. They provide a fifth of our electric power, twice the share provided by hydropower. Not one new reactor has been ordered for use in the U.S. since 1978, although most of the hundreds of power reactors overseas are of American design. Of the other major users of fission power, the former Soviet states operate Soviet designs, some of which, like the reactors at Chernobyl, could not even be licensed in America because of design deficiencies; most relevant is their lack of containment vessels. In both France and Japan, as the saying goes, they had "no coal, no oil, no gas, no choice." French electric power is now three quarters nuclear, Japan's, about 27 percent. Successfully operating improved versions of licensed American designs, those two national systems make plain that reactor type does not fully determine either economy or safety.

The prevalent judgment that American manufacturers after 1960 rather rashly scaled up naval reactor design by 10-fold in power during only a few years' time seems valid. A bandwagon began to roll before design experience matured. But design is not where most trouble arose. Often operating utilities were not up to their tasks. "If management doesn't pay attention, then you end up with people sleeping at the controls," one American critic concludes. When instead it was a bold Soviet absolutism that held sway—"Reactors are regular furnaces, and the operators who run them are stokers," said one high official—the result was worse still in both design and operation. The Pripyat Valley would have been better off had the Chernobyl operators been asleep rather than hard at work on their ill-considered postmidnight experiment.

Certainly there are doubts not addressed here. Are new reactor designs manifestly safe against accident? When will we have full confidence in the low public health risks of low-level doses of fission products? The Manhattan Project made embryonic nuclear war with less than 6,000 tons of uranium in sight. Now the world mines annually four or five times that total: the price is low, and the reserves are large. When will we need the clever Integral Fast Reactor of Argonne that burns up uranium so completely and contains within itself both its fission products and its plutonium? Can we expect instead an affordable large-scale source of solar power, say, silicon photovoltaics with the total area of many counties?

Rhodes has made a concise and attractive case for a new beginning with fission power. "The real world...needs more energy," with less of the greenhouse gases. A simpler, energy-poor world is not a solution; rather it is neglect of the resource-poor majority of our fellow humans. Frugality is not the same as affluent indifference to the well-being of others.

Perhaps Enrico Fermi was right when he mused once that whatever the designs and precautions, perhaps people simply cannot accept the widespread presence of stores of radioactivity, each like 100 tons of Marie Curie's glowing radium. But we will surely gain from seriously renewed discussion.

Molecules and Reactions

THE NORTON HISTORY OF CHEMISTRY, by William H. Brock. W. W. Norton & Company, 1993 ($35; paperbound, $15.95).

This welcome, chunky little volume was published in 1992 by a London firm, and the entire series of 10 science histories, of which it is the second, will appear in the U.S. under the Norton imprint. Dr. Brock studied chemistry in London before he turned to historiography; this thoughtful and lively overview is his latest among half a dozen more specialized studies. Most of the book is nontechnical; he includes human interest enough "to propel the non-chemical reader toward the final pages." Anyone who recalls high school chemistry enough to read a string of chemical equations (hardly any algebraic ones), to recognize such terms of art as ions, isomers and ligands and to grasp modest structural formulas will travel a smooth path, if here and there a rather steep one.

Chemistry, Brock says, "deals with the properties and reactions of different kinds of matter." A literal reading of that phrase would set an impossible task. In fact, this fresh and quite comprehensive book does not treat so broad a swath across science. Its 16 chapters are nicely organized by theme, some restricted in time, others extending across a century or two. This is the history of laboratory synthesis and analysis, of structures and reaction mechanisms, of radicals, valences, complexes, of Mendeleev's periodic system, of pure reagents, hands-on teaching, publishing results, even of organizing research teams.

A reader sees this history as one defined by a living profession and its celebrated issues and leaders. It does not much treat biochemistry, molecular biology, metallurgy or atomic physics, whatever those disciplines may say about properties and reactions. Among all 700 of these rather small pages, at most a few lines speak of enzymes and DNA. Quantum chemistry is certainly present; it is electronic charge clouds in clustered atoms that retain the forms and make the moves chemists have envisioned and constructed since the 20th century began. The entire tale is imbued with the rich flavors of material and intellectual diversity, so distinct from the narrowness of much of physics, for "theoretical chemistry is still a quirky empirical science based upon a Schrödinger equation that can hardly ever be solved."

Applied chemistry is here, both in the giant industries of the chemical engineers and in the empirical technologies of old and subtle crafts. Those men and women who were "metallurgists, brewers, dyers, calciners and pharmacists" are not forgotten. Their vessels, stills and furnaces drew in philosophers East and West, from the Taoists to Newton himself. That long arcane symbiosis brought us much that lasts, from chemical terms and symbolism to phosphorus and gunpowder, even if the "spiritual quest" that would perfect gold and the soul together now seems less than successful.

In engineer Vannoccio Biringuccio's wonderful little treatise of 1540 on metals, weapons and waterpower, he draws from experience the conclusion that would bury alchemy: "I could still discourse concerning the art of transmutation, or alchemy…but I am drawn…to follow the path of mining more willingly…even though mining is a harder task, both physical and mental, is more expensive, and promises less…; it has as its scope…seeing what really exists rather than what one thinks exists."

Given the complexity of practice and the obscurity of ambitious philosophers, the rise of modern chemistry from the 16th century was slow. Inspired and influential medical thinkers, such as Joannes Baptista van Helmont, led the way. That van Helmont was denounced as heretical for his advocacy of "weapon salve"—treating the weapon instead of the wound—provides support for the historian's refusal to accept the naive tale of simple progress that ends in our own flawlessness. Van Helmont's own benchwork showed him many airlike substances he called gases, even inventing the generalizing word from a Greek term for chaos. One of his disciples saw in the hidden processes of digestion the duality of acid and alkali (bile), still a main chemical theme. But it is purity, along with its particulate implications, that is the "fundamental concept of chemistry." Although Antoine-Laurent Lavoisier was first to define a pure substance sharply, the assayers and the pharmacists had long recognized the idea. The noun and verb "test" itself comes from a Latin word for the little cup used by the expert assayer, also called the cupel. The war against adulterants and counterfeits is old indeed.

"Chemistry is a French science; it was founded by Lavoisier of immortal fame," wrote one historian in 1869. Such nationalism is plainly hyperbolic. But it is not without its kernel. The chemical literature left to us from before Lavoisier is mostly incomprehensible to a modern chemist; after him, today's reader finds "little difficulty." This text makes Lavoisier real, both on the human side and in his chemical work. We see even that the theory of phlogiston with all its crotchets was not at once absurd. The attention, money and effort Lavoisier devoted to his apparatus present a second surprise. His laboratory was instrumented enough to challenge the state of the art with its supply of special bulbs, gasometers, balances and calorimeters. Some have even questioned whether he was a chemist at all. Possibly we should see the chemical revolution of the 1790s as "the result of a brief and useful invasion of chemistry by French physicists." Our author rather audaciously outlines "necessary and sufficient" conditions for Lavoisier's revolution. The recognition of the gas phase as substance and participant, always to be held to account, was perhaps crucial. His views on acidity led to the development of a language that reflected composition.

Atomism came in strength with John Dalton, but it remained split until the last Victorian years. The chemists knew their atoms first, from atomic weights, the periodic change of properties with weight, and many rich molecular formulas, but they knew of no evident size or structure. A paper of 1850 by Alexander Williamson of University College, London, gave a picture of a single reaction that demanded two clear steps of rearrangement and implied a constant state of motion at the atomic level. Chemists' atoms were happily arranged in hexagons, nets and chains. They left the plane of the paper for good with the chemical logic of Jacobus van't Hoff in 1874; no other geometry could explain all the distinct isomers of a given atomic cluster. Physical atoms, scaled and arrayed in jingling molecules milling about in space, were not commonly known about until the early years of this century.

The electron within the nuclear atom, ionic states, quantum-chemical bonding, all described for a time with little mathematics by one or another fruitful yet limited approximation, become the main themes of the later chapters. That the physics of hydrogen will hardly do, even P.A.M. Dirac's physics, is clear enough, once you consider that the chemists have by now characterized a few million molecular species. (This holds even if we set aside the molecular libraries written in the nucleic acids of life.) From Gilbert N. Lewis and his paired electrons to quantum chemists of deeper, more dynamic outlook, such as Robert S. Mulliken, molecules and reactions have come to demand more than the logical and geometric arguments of the past. The art of employing approximate descriptions of all those interacting atoms in space, even in energy and in time, has been taken to dizzying heights by creative weavers of molecular tapestry, such as Robert Woodward, Roald Hoffmann, Kenichi Fukui and Donald Cram.

A few final pages treat the environmental problem of the chemical industry, by which in large part we are clothed, housed and fed. In one telling citation from 1899, a critic describes a town near Liverpool dominated by the alkali industry: "The sky is a . . . roof of smeary smoke. The atmosphere is a blend of railway tunnel, hospital ward, gas works and open sewer. . . . The products are pills, coal, glass, chemicals, cripples, millionaires and paupers." By 1999 some matters will have changed there, but not all.

"Is Linus Pauling this century's greatest chemist?" Few of all the personal stories here told are better than this little one about our splendid sage in California. His wife, reading about Watson and Crick's Nobel Prize for DNA, very reasonably asked: "Linus, if this structure is so important, how come you didn't solve it?" Pauling had left a little room for others, though not much; after all, he had to compete for the Nobel Peace Prize, too, without a passport.

AIDS and Population "Control"

The now worldwide AIDS pandemic finds its ugliest manifestation in the proposition that AIDS has arrived in time to stop the population explosion. One hears it voiced by otherwise blameless people. Some see AIDS as the solution, in particular, for the "problem" of Africa. There the rate of population growth is highest and poverty deepest. Epidemiologists of the World Health Organization estimate that Africans constitute about 10 million of the 15 million people infected worldwide with the human immunodeficiency virus (HIV) and so fated to die of AIDS.

The AIDS proposition scants history and grossly underestimates the durability of the human species, Africans included. At its present rate of transmission, HIV will infect some 200 million people by 2010. The African share of the casualties might then approach 100 million. That, as a disciple of Thomas Malthus observed of the million Irish who perished in the 1845–50 potato famine, would scarcely be enough.

The Black Death, to which proponents of this cure for population growth hopefully compare the AIDS pandemic, carried off more than half the people of 14th-century Europe. By the middle of the 17th century, the European population had arrived at the point on the growth curve to which it would have increased by that time without deflection by the Black Death.

The paroxysm of violence that seized the industrial world through the three decades from the start of World War I to the end of World War II killed 200 million people. That was more than 10 percent of all the people who lived in those years. Their absence was not remarked in 1970, when the rate of world population growth reached its all-time peak, at around 2 percent.

It was Malthus who made economics the dismal science, but he also made this branch of moral philosophy a science. He rooted economics in what had been the unrelieved experience of humankind from the time of the agricultural revolution and the opening of the first village markets. "Apart from short, exceptional periods," Alfred North Whitehead observed, "the normal structure of society was that of a comparatively affluent minority subsisting on the labors of a teeming population checked by starvation and other discomforts."

At the very time Malthus set out his baleful equation, however, industrial revolution had begun to make the growth of production outrun population. Within 20 years of Malthus's death, moreover, John Stuart Mill discovered from inspection of baptismal records that the birth rate of England had begun to decline. This discovery did not shake the conviction, which Mill shared with Malthus, that population growth was the "dynamics of political economy," for the population of prospering England was exploding.

The population of all the European countries undergoing industrial revolution was exploding during this period. Now, after this gigantic increase, which multiplied the number of Europeans 20 times over that in 1600 and avalanched them onto all continents, the populations of all the industrial countries are at or approaching zero growth. These lucky 1.25 billion people—counting the Japanese, the first non-Europeans, in their number—are completing the so-called demographic transition. From near-zero growth in 1600 at high death rates and high birth rates with life expectancy at 25 years, they are arriving at near-zero growth again but at low death rates and low birth rates and with life expectancy at 75 years.

Recent history gives every reason to expect that the other three quarters of the world population will make the demographic transition. The leading edge of industrial revolution—mass education, sanitation and primary medicine, and the green revolution—has brought down death rates and lengthened life expectancy throughout the preindustrial world. The rest of humankind has entered the first phase of the demographic transition. Hence the ongoing swelling of the population.

Entrance into the second phase is marked for some few developing countries by decline in their birth rates. These are countries where industrial revolution has proceeded furthest and where its increasing product is most widely shared—small countries like Costa Rica and Sri Lanka and also the biggest countries, India and China.

Whitehead cited India and China as "instances of civilized societies which for a very long period in their later histories maintained themselves with arrested technology.... They provided the exact conditions for the importance of the Malthusian Law." By the turn of the 19th century, when Malthus published his *Essay on Population,* they were the world's most populous countries.

Since the end of World War II, India and China have been engaged in industrial revolution, China leading. With life expectancy lengthening to 60 years in India and to 70 years in China, their huge populations have more than doubled. In both countries, calories per capita now meet the daily requirement, and potable water is available to three quarters of their citizens. India has reduced its child death rate to 142 per 1,000 live births, and China to 42. In India, 27 percent of the population lives in cities, in China, 33 percent. Literacy among the female population is 34 percent in India and 62 percent in China. Contraceptives are in use in 43 and 71 percent of their households, respectively. As these statistics suggest, both nations have entered the second phase of the demographic transition. The fertility rate in India has declined from more than six (infants per female reproductive lifetime) to four; in China it is 2.3, close to the zero growth rate of 2.1.

How to hasten the passage of the preindustrial world through the demographic transition was the principal topic before the United Nations Conference on Environment and Development, the Earth Summit, held in Rio de Janeiro in the summer of 1992. The major product of the conference, Agenda 21, is now the agenda of the United Nations. It sets out a program of "sustainable development" to bring human numbers and appetites into accord with the finite resources of the earth before the end of the next century. The AIDS pandemic can only divert physical resources and human energy from this hopeful and urgent enterprise. Delay portends a larger ultimate world population. Industrial revolution has set the terms of a morality different from that implied by the Malthusian equation. It is people living, not dying, who bring population growth to a stop.

GERARD PIEL is chairman emeritus of Scientific American.