MAY 1994
$3.95

SCIENTIFIC AMERICAN

Biology and homosexuality: a debate.

The crisis in air-traffic control.

How our species came to be.

***Quantum cat** in Schrödinger's thought experiment gets a reprieve from a rival theory.*

SCIENTIFIC AMERICAN

May 1994 Volume 270 Number 5

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Second-class postage paid at New York, N.Y., and at additional mailing offices. 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 features Tyrone (Toonces on *Saturday Night Live*) in the role of Schrödinger's cat. A chance exists that radioactive material will emit a particle that (through an unseen mechanism) tips over a bottle of poison. The standard view in quantum physics holds that an event needs an observer, so the cat is neither dead nor alive until the box is opened. A different view rejects such ambiguity (see "Bohm's Alternative to Quantum Mechanics," by David Z Albert, page 58). The photo was staged (the Editor lives happily with five cats).

THE ILLUSTRATIONS

Cover photograph by Dan Wagner; electronic manipulations by Jack Harris/ Visual Logic

Page	Source
43–45	Joyce Tenneson
46	Patricia J. Wynne (*top*), Jared Schneidman/JSD (*bottom*)
47	Courtesy of Simon LeVay (*top*), Jared Schneidman/ JSD (*bottom*)
48–49	Edward Bell
51	Joyce Tenneson
52–54	Patricia J. Wynne
55	*Platones Symposion Erklärt,* by Arnold Hug (Leipzig, 1876)
59	Dan Wagner
60–61	Michael Goodman
62	UPI/Bettmann
63	Michael Goodman
67	Mark Edwards/Still Pictures
69	William M. Carson, University of Alabama at Birmingham
70	Johnny Johnson
71	Howard M. Johnson
72–73	Jared Schneidman/JSD
74	James Holmes/Science Photo Library, Photo Researchers, Inc. (*top*), Johnny Johnson (*bottom*)
75	R Jonathan Rehg
76–81	Courtesy of Fred Durant, Space Art International
82–84	Ian Worpole
85	Roberto Osti (*top*), Ian Worpole (*bottom*)
86	Ian Worpole
87	Roberto Osti
89	Boyd Norton
90	Patricia J. Wynne (*top*), Johnny Johnson (*bottom*)
91	Patricia J. Wynne (*top*), Yves Coppens (*bottom*)
92	Patricia J. Wynne
93	Patricia J. Wynne (*top*), Des Bartlett/Photo Researchers, Inc. (*bottom*)
94–95	Patricia J. Wynne
96–97	Robert Prochnow
98–99	Ian Worpole
101	Federal Aviation Administration
102	Robert Prochnow
103	Ian Worpole
118	Patricia J. Wynne
119	Johnny Johnson (*top*), Patricia J. Wynne (*bottom*)
120	Johnny Johnson

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

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

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

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

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

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

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

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Mary Sadlier, *Research Manager;* Ethel D. Little, *Advertising Coordinator*

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

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

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue, New York, NY 10017-1111
(212) 754-0550

CHAIRMAN AND CHIEF EXECUTIVE OFFICER: John J. Hanley

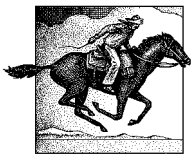
CO-CHAIRMAN: Dr. Pierre Gerckens

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

DIRECTOR, ELECTRONIC PUBLISHING: Martin Paul

CHAIRMAN EMERITUS: Gerard Piel

PRINTED IN U.S.A.



LETTERS TO THE EDITORS

A Commons Error

In his essay "The Tragedy of Enclosure" [SCIENTIFIC AMERICAN, January], George Monbiot makes some excellent points about the effects of various kinds of property rights on the sustainability of resource usage, but he introduced nomenclature that may cause confusion in the future. I feel that I am partly responsible for this unwelcome turn.

When, in 1968, I described "the tragedy of the commons," I used the word "commons" as it had been used by my only predecessor, William Forster Lloyd in 1833. He and I meant any resource that was shared on a "help yourself" basis, that is, an *unmanaged* commons. Although neither Lloyd nor I originally used that adjective, it is implicit in both of our essays.

In the early stages of the exploitation of a natural resource, when there is no real shortage, an unmanaged commons is the most economical mode of distribution. But human demands increase faster than resources, so there comes a time when a help-yourself policy becomes contrary to the interest of all. Two alternatives are then available: privatism, in which a resource becomes private property and is managed by its owner, or socialism, in which a manager (or bureaucrat) is injected into the system. Both privatism and socialism may fail or succeed, depending on many factors. But commonism—living by the Marxist ideal of an unmanaged commons—cannot possibly succeed in a world of shortages. When the needs asserted by individuals are given priority over preserving the carrying capacity of the environment, tragedy is the inescapable result.

The survival of the commons among the Turkana people does not contradict what I have described because (in Monbiot's words) every significant resource in their environment "is controlled by a committee of elders, who decide who should be allowed to use them and for how long." The Turkana elders refused to make individual need paramount in the ethics of distribution. Group survival came first.

I think the title of Monbiot's essay is oriented 180 degrees the wrong way. What he describes among the Turkana is not a tragedy of enclosure, but the tragedy that follows the breakdown of enclosure. The people of the world still have to find ways to introduce acceptable management into the global commons of air and oceans. Idealists who urge us to create a world without borders would create a new unmanaged commons and globalize poverty.

GARRETT HARDIN
Department of Biological Sciences
University of California, Santa Barbara

Monbiot replies:

Hardin confuses communities and nation-states. Local commoners are not confined to the state's narrow range of political options. Their representatives are not third parties but part of the community, acting at its behest. When commons are privatized or nationalized, owners or outsiders take over the decision making, with resultant losses of accountability and flexibility. Most members of the community are excluded not only from the land (privatization of a common typically reduces the number of participants) but also from the political process. Both resources have, in other words, been enclosed.

The Cancer War

Significant declines in the mortality rates for several major cancers (breast, stomach, cervical, testicular and colorectal) between 1973 and 1986 argue strongly that there have been major successes in the war on cancer ["A War Not Won," by Tim Beardsley; SCIENTIFIC AMERICAN, January]. Of course, there is still a long way to go. Thus far that war has primarily focused on basic science and on the development of new anticancer therapies. Until recently, there was little funding for studies of cancer prevention: curing cancer was seen as a quicker fix. Now that we know how effectively an established cancer can resist treatment, prevention is widely perceived as preferable. That it took scientists 23 years to reach this perception is unfortunate but understandable.

The lack of greater progress is also caused by society's reluctance to use the findings that have emerged from the laboratory. For example, we know that tobacco is responsible for about 400,000 deaths every year, many as a result of lung and other cancers. Yet we have not taken the tobacco industry to task for this. It is ironic that our society pays a subsidy to farmers who grow tobacco, then pays a subsidy to scientists who study the resultant cancers and finally pays a subsidy to physicians who treat those people who are dying from tobacco use.

R. GRANT STEEN
St. Jude Children's Research Hospital
Memphis, Tenn.

The only way we will make major advances against cancer will be through early detection and, more important, prevention. We should realize that many major scourges, such as polio and smallpox, were never "cured"; rather an effective means of preventing them, the vaccine, was developed. Similarly, we should seek ways of reducing our population's exposure to cigarette smoke, high-fat diets and excessive sun exposure. Given the ever increasing constraints on our health care budget, I do not foresee a day when we will be able to repair everybody's damaged DNA—especially if that damage has been self-induced.

JOSEPH P. IMPERATO
Lake Forest Hospital
Lake Forest, Ill.

Revenge in a Flash

Gerard J. Holzmann and Björn Pehrson's article on optical telegraphs ["The First Data Networks," SCIENTIFIC AMERICAN, January] has a fascinating literary echo. In *The Count of Monte Cristo,* by Alexandre Dumas (published in 1884 but set during 1815–1840), the Count revenges himself on his enemy Baron Danglar by bribing a telegraph operator to send a false message along the Bayonne line. The chapter gives a delightful description of the life of those unsung operators, and your readers may enjoy a revisit to this classic after perusing the article.

JOHN GRABINAR
Bromley, England

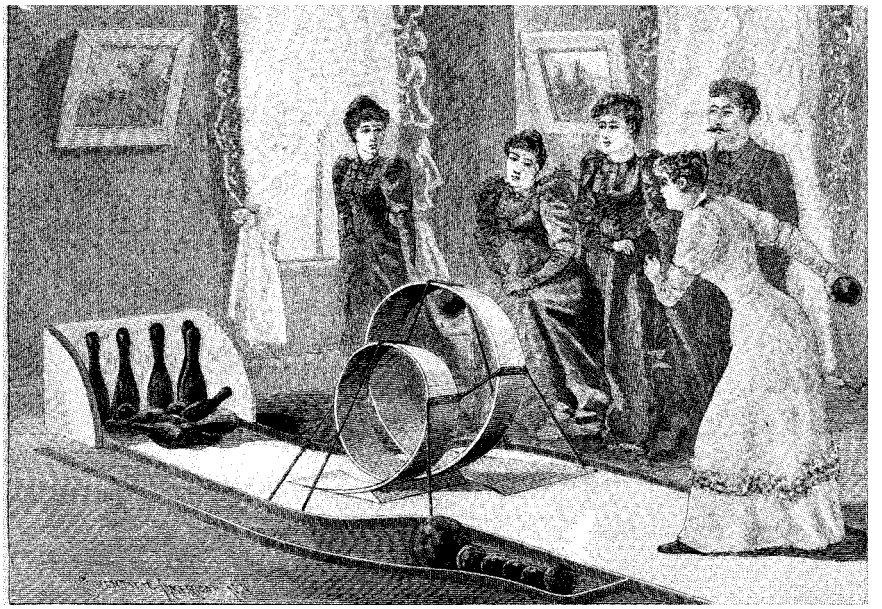
Letters selected for publication may be edited for length and clarity. Manuscripts will not be returned or acknowledged without a stamped, self-addressed envelope.

50 AND 100 YEARS AGO

MAY 1944

"A two-way electronic train telephone system, permitting continuous voice communication with moving trains, between trains, and between the head and rear of trains was recently put into operation by the Pennsylvania Railroad. This system, the only one of its kind so far established, has been applied to the freight service on the 67-mile Belvidere-Delaware branch, running northward from Trenton, New Jersey."

"Pilots can be trained for night flying in the daytime by a new technique which consists of covering the windshields and windows of the plane with green acetate and having the pilot-trainee wear red goggles. The combination of filters cuts down on the pilot's vision beyond the cockpit so that it approximates night vision. The instructor does not wear red goggles and so he can see beyond the cockpit and enjoy daylight."



The centrifugal bowling alley

"Recent agitation for grade labeling of various commodities has raised many questions in the minds of industry and consumer alike. With purely *informative* labeling we have no quarrel. We *are* seriously concerned with the governmental project under which removal of all brand names from packaged goods is contemplated and which contemplates placing a ceiling price on each grade.—*A. P. Peck, Managing Editor"*

"A plastic covering that removes all the hairs from a pig when it is peeled off has given the housewife more high-grade pork cuts, ham, bacon, lard, and other pork products, according to Hercules Powder Company. The chemical shaving method, which has been adopted by meat packers, also saves processing time and reduces de-hairing costs."

SCIENTIFIC AMERICAN

MAY 1894

"The present very active and enlightening agitation over the question of woman's suffrage calls up again the many now established facts about the physiological differences in the nervous system of the sexes. There is no doubt, we believe, in the minds of physiologists that the mental characteristics of women have a structural basis in the conformation and amount of her nervous tissues. This does not in the least prove that woman's mind is not adapted to the demands of suffrage, or of political, jury, and militia duty. It only shows that the result of conferring suffrage cannot be positively predicted either one way or the other. It is an experiment which may or may not turn out wisely.—*Med. Record"*

"The fertilization of the trumpet creeper had never been satisfactorily explained until Mr. Hamilton Gibson discovered that it is the work of humming birds. They thrust their long bills down into the nectaries at the base of the blossom, and come out with their backs covered with pollen, which they give to another flower when they seek the same sweets there."

"Mr. Henry Wilde, F.R.S., of Manchester, has a theory that the exterior of our earth is permanently magnetic; also that an interior sphere is movable and magnetic, rotating in the plane of the ecliptic, 23 $^1/_2$ degrees, and loses one revolution in 960 years; he assumes, also, that the internal sphere is electro-dynamic. At a recent conversazione of the Royal Society he exhibited two globes, one within the other, and each containing a coil of insulated wire, through which currents of electricity could be sent, and mounted so that their motions should agree with his hypothesis. By placing a compass over different parts of the outer globe, he obtains the same variations and dip as are found in nature."

"The preservation of certain species of American animals, now nearly extinct, was the primary object for which Congress was asked to establish a national zoological park at Washington. The appropriations asked for were very moderate, but in all cases they were cut down. Thus: $36,850 was asked for to erect the necessary buildings; Congress reduced this estimate to $18,000. Nevertheless, the results achieved are highly satisfactory."

"A first-class single bowling alley costs $250 and requires a flooring 85 feet long and 6 feet wide. The practice of bowling at home in ordinary dwellings is, therefore, out of the question. A new system, illustrated in our engraving, has been designed to adapt to the requirements of domestic life. Instead of the long straight floor, a circular cycloidal pathway for the balls is provided. The ball is projected in the usual manner, rolls up and down through the spiral path, and proceeds straight toward the pins at the opposite end of the room."

When Treaties Fail

If a rogue state attacks, what are the technological options?

SANDIA NATIONAL LABORATORIES

"KINETIC KILL" WARHEAD, shown here being accelerated on a rocket sled before releasing metal projectiles, is being considered as a counterproliferation weapon.

The world is not a safe neighborhood. At least a score of countries are believed to possess weapons of mass destruction or to be seeking them, and a dozen nations have ballistic missiles. The dissolution of the Soviet Union has fragmented its nuclear arsenal and raised concerns that weapons might fall into the hands of terrorists or "rogue" states. On the Korean peninsula, in the Middle East and on the Asian subcontinent, the nuclear nonproliferation treaty regime has been cracking under the strain of regional tensions. A proposed ban on chemical weapons, even its strongest supporters acknowledge, will depend heavily on the goodwill of those who sign it.

Although the Clinton administration supports diplomatic efforts to contain the spread of strategic weaponry, it has quietly begun building a technological firebreak. It is the Counterproliferation Initiative, a Department of Defense program designed to equip U.S. forces to fight adversaries armed with "special" weapons—Pentagonese for nuclear, biological and chemical arms. The initiative was the brainchild of Les Aspin, the former secretary of defense. One of Aspin's first acts as secretary was to create a Pentagon position called assistant secretary of defense for nuclear security and counterproliferation. Aspin appointed Ashton B. Carter, a professor at Harvard University's Kennedy School of Government, to the post.

According to Carter, the initiative, which has survived Aspin's departure, could spur the development of various technologies: miniature sensors that, disguised as twigs or pebbles, keep tabs on a suspected arms-manufacturing site; devices that provide warning of a chemical or biological attack, as well as vaccines, antidotes and protective gear to minimize an attack's effects; bombs that disable electronics by emitting an electromagnetic pulse; and "hypervelocity missiles" that can penetrate deep underground to destroy buried weapons facilities or bunkers—such as the one in which Saddam Hussein reportedly hid during the Persian Gulf War.

By far the most expensive component of the plan involves theater defenses against ballistic missiles. These would consist of mobile, ground-based interceptors designed to shoot down missiles launched against U.S. troops or allied populations in a regional conflict. The Clinton administration is seeking $3.25 billion for research on missile defenses in 1995, up from $2.7 billion this year. It has also suggested that the 1972 Anti-Ballistic Missile Treaty between the U.S. and the former Soviet Union be modified to permit theater defenses.

Carter emphasizes that the Counterproliferation Initiative requires not only new technologies but also new military strategies, ones focused on adversaries such as Iraq or North Korea rather than the former Soviet Union. These strategies should complement rather than supplant prevention policies, he explains, by convincing potential proliferators that special weapons will have little military value. The plans considered so far do not advocate preemptive strikes to destroy enemy arsenals, Carter declares, nor do they call for using nuclear weapons. "We are trying to look for nonnuclear responses to special weapons," he remarks.

But Laura S. H. Holgate, a member of Carter's staff, says a new class of low-

yield nuclear weapons has not been "categorically" ruled out as a counterproliferation tool. Dubbed "mininukes," such weapons would have yields ranging from five kilotons to 100 tons. (The bomb that destroyed Hiroshima had a yield of 10 kilotons.) Indeed, some officials in the weapons laboratories, notably Edward Teller of Lawrence Livermore National Laboratory, have suggested that mininukes deployed on high-precision missiles would make effective counterproliferation weapons. In addition to serving as "bunker-busters" or electromagnetic-pulse bombs, mininukes would be ideal for destroying biological-weapons stockpiles, researchers say, because the thermonuclear heat and radiation would kill and not just disperse pathogens.

In last year's Defense Authorization Bill, Congress specifically prohibited the Department of Energy from conducting "research and development which could

Neural Eavesdropping

How does the brain establish and store memories? Neuroscientists have traditionally addressed this question by focusing on neurons and the synapses that link them—just as engineers might try to understand how homes in a town communicate simply by tracing the telephone lines linking them. Now investigators must consider the possibility that each home's windows are wide open and that everyone is eavesdropping on his or her neighbors.

The new findings involve a neural mechanism called long-term potentiation (LTP), which is thought to play a crucial role in memory and learning. The effect occurs when repeated transmissions of impulses across the synapse linking two neurons result in a positive feedback effect, making future transmissions still easier. Donald O. Hebb, the great Canadian psychologist, first proposed the existence of an LTP-type mechanism 40 years ago.

Many researchers assumed that for LTP to occur, a signal must be sent from the postsynaptic neuron (which receives the impulse) back to the presynaptic neuron (which originally sent the impulse) to strengthen the bond between them. The hypothetical signal has been named the retrograde messenger.

Several years ago a number of workers, notably Daniel V. Madison and Erin M. Schuman of Stanford University, performed experiments suggesting that the retrograde messenger might be nitric oxide. Nitric oxide is a soluble gas so reactive that it generally does not last for more than a few seconds. Moreover, it slips easily in and out of cellular membranes, traveling without regard for normal channels of communication.

Yet Madison and Schuman wondered: If nitric oxide diffused from the postsynaptic cell to the presynaptic one, might it affect neighboring synapses as well? If true, that hypothesis would disprove the presumption—key to Hebbian learning—that LTP occurring at one synapse has no effect on neighboring ones. The view that LTP would be synapse specific "was the dogma," Schuman says.

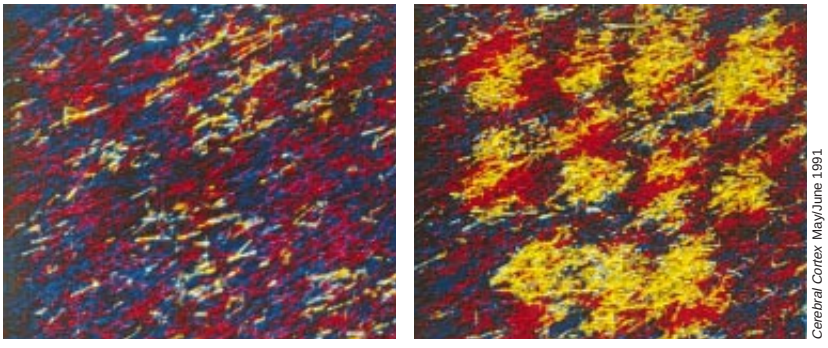
Actually, evidence that LTP might not be entirely localized had been presented in 1989 by workers at the Max Planck Institute for Biological Cybernetics in Tübingen. When they evoked potentiation in one synapse of a neuron taken from the hippocampus of a rat, potentiation was reinforced in nearby synapses. At that time, many investigators felt the results were not robust enough to overturn the accepted view.

Madison and Schuman repeated the Germans' experiments using better instrumentation and tighter controls and got the same results. They then showed that nitric oxide contributed to the effect. When they applied a chemical that blocked the production of nitric oxide in a postsynaptic neuron, long-term potentiation failed to occur either in that synapse or in neighboring ones.

Madison and Schuman emphasize in *Science* that nitric oxide may be just a link in a chain generating the diffusion effect. Yet signs of nitric oxide's extracellular activity continue to accumulate. Three weeks after Madison and Schuman's paper appeared, a group led by P. Read Montague and Michael J. Friedlander of the University of Alabama announced in *Science* that nitric oxide can mediate interactions between neural pathways that are not directly connected.

To Eric R. Kandel of Columbia University, an authority on the molecular basis of memory, the message of all this work is clear: the old Hebbian model of learning, which ignores the extracellular context of communication, is too simple. "This is the most oversold idea in neuroscience," Kandel proclaims. Indeed, the new findings lend weight to a non-Hebbian theory proposed in 1990 by Montague and two colleagues, Joseph A. Gally and Gerald M. Edelman of the Neurosciences Institute in New York City. In the model, called volume learning, the diffusion of nitric oxide contributes not only to memory formation but also to the spatial organization of the brain during embryonic development. The putative effects would stem from nitric oxide's well-known ability to dilate blood vessels as well as its role as an enhancer of neural connections.

Terrence J. Sejnowski of the Salk Institute for Biological Studies in San Diego thinks volume learning may explain why neurons gather into onionlike sheets and columnar structures in certain regions of the brain. The width of the layers and the columns are roughly equal to the effective range of a diffusing nitric oxide signal. Yet Sejnowski expects that upcoming experiments will reveal that nitric oxide is only one actor in a complex drama. "It seems," he observes, "that there is a whole family of diffusion messengers, each with its own targets and time course, that contribute to plasticity and development." —*John Horgan*



Cerebral Cortex May/June 1991

VOLUME-LEARNING simulation shows how the diffusion of nitric oxide might result in enhanced neural connections in localized regions of the brain.

lead to the production... of a new low-yield nuclear weapon." But the restriction lasts only one year, and it has a rather large loophole: it allows such research "to address proliferation concerns." In response to a query from SCIENTIFIC AMERICAN, Secretary of Energy Hazel R. O'Leary did not rule out the possibility that such weapons could be developed. "This is an issue we're grappling with," she said.

The possibility that the Counterproliferation Initiative may allow the development of mininukes alarms arms-control advocates. That step would make it much more difficult for the U.S. to negotiate an extension of the Nonproliferation Treaty when it comes up for review in April 1995, contends William M. Arkin, a national security consultant.

Even in its nonnuclear form, the initiative arouses concern. Natalie J. Goldring of the British American Security Information Council fears that by trying to increase its lead in conventional capabilities, the U.S. may actually encourage other countries to acquire weapons of mass destruction as "equalizers." She also faults the implicit premise of the initiative that traditional efforts to stop proliferation are inadequate. After all, she points out, diplomatic pressure recently helped to convince Argentina and Brazil to renounce all ambitions to build nuclear weapons.

The ballistic-missile defenses advocated by the initiative suffer from the same liabilities as the much more ambitious "Star Wars" plan proposed by President Ronald Reagan, remarks Spurgeon M. Keeny, Jr., of the Arms Control Association. The defenses could easily be circumvented by inexpensive, low-tech means of delivery, such as airplanes, boats or trucks. "I'm extremely concerned that the proposals being made to accommodate this system would undermine the Anti-Ballistic Treaty," Keeny adds.

"Counterproliferation is really an answer to the question, What if Saddam Hussein had had nuclear weapons? And I think it's the wrong answer," comments Barry R. Posen, a political scientist at the Massachusetts Institute of Technology. Posen thinks that even with improved offensive and defensive systems, the U.S. could never be sure of its ability to destroy all of an enemy's nuclear arsenal and thereby prevent a retaliatory strike.

That raises an unpleasant question: How should the U.S. respond if an enemy successfully delivers a nuclear weapon? "Our firm preference is to use nonnuclear means and to win the war anyway," Carter replies. Some conservative analysts, such as Seth Cropsey of the Heritage Foundation, a former Pentagon official in the Reagan and Bush administrations, have been taking the same position, arguing that the U.S. would thereby reinforce the "taboo" against nuclear weapons.

Ironically, John A. Pike of the relatively dovish Federation of American Scientists insists that the threat of massive retaliation—rather than missile defenses and other weapons being considered under the Counterproliferation Initiative—is still the best defense the U.S. has against an adversary armed with nuclear weapons. If such an enemy initiates a nuclear attack, the U.S. should "turn them into a sea of radioactive glass and be done with it," Pike declares. He notes approvingly that last July, President Bill Clinton warned that if North Korea used nuclear weapons, "it would be the end of their country."

But what if terrorists or breakaway elements within a country launch the attack? And what if the attack strikes an ally of the U.S. but not U.S. troops or citizens? When is massive retaliation—conventional or nuclear—appropriate? Pike acknowledges that "this is a very interesting and complicated set of questions that people are just beginning to examine."

Michael Krepon of the Henry L. Stimson Center in Washington, D.C., thinks the Counterproliferation Initiative—at least if it eschews mininukes—is a reasonable first step toward dealing with these questions. Krepon is particularly supportive of the program's call for antimissile defenses. He also suggests that the somewhat "fractured" response of experts to proliferation issues lately is a healthy sign. "That means people are taking a new look at a world that's changed," Krepon says. "But it's also confusing." —*John Horgan*

The Lucky Ones

How do some HIV-positive people avoid getting AIDS?

In most people who contract the human immunodeficiency virus (HIV), AIDS develops within a decade. Yet a few have normal immune systems more than 10 years after infection. Do these healthy long-term HIV survivors harbor clues to immune responses or shifts in the biochemistry of the virus that could be vital to efforts to combat AIDS?

Susan P. Buchbinder of the San Francisco Department of Public Health and her colleagues are finding that about one in 12 men who have been infected

CAROLYN JONES from *Living Proof* © 1994 Abbeville Press

ZOE LORENZ* (right) *of Louisville, Ky., acquired HIV, the virus that causes AIDS, almost 11 years ago, but her immune system is still working well. Lorenz's daughter, Candice, who is four years old, tests negative for the virus.

for more than 10 years has counts of CD4 *T* cells that are still above half the average value. These cells, known as helper lymphocytes, usually decline inexorably in people infected by HIV, often reaching a twentieth of their normal level in those with full-blown AIDS.

The long-term survivors in San Francisco seem to use recreational drugs just as much as those who experience a more typical course of infection, and they have the same incidence of sexually transmitted diseases. The survivors differ in one intriguing respect: they have larger numbers of the immune system agents known as CD8 *T* cells in their blood than do either their less fortunate counterparts who develop AIDS or people who are uninfected with HIV. This class of cells includes so-called cytotoxic *T* cells, which attack and kill cells infected by viruses.

John L. Sullivan, a researcher at the University of Massachusetts Medical School in Worcester, Mass., and his coworkers have noted a similar phenomenon. Sullivan observes that 80 percent of children who are five years of age or older and were infected by HIV before birth have cytotoxic *T* cells that recognize HIV proteins. In the first year of life, in contrast, only a third of infected children have such cells. Sullivan suspects that the children who lack cytotoxic *T* cells that recognize HIV may be the ones who do not survive for long.

Other observations pointing to a crucial role for cytotoxic *T* cells have been made by Andrew J. McMichael of the University of Oxford. He notes that patients with little cytotoxic *T* cell response to HIV proteins tend to do worse than those whose cells react in a more normal way. Moreover, he points out, tests to detect anti-HIV responses by cytotoxic *T* cells usually show less activity the longer someone has been infected; indeed, they often approach zero in someone who has AIDS.

Bruce D. Walker, a researcher at Massachusetts General Hospital, also sees vigorous cytotoxic *T* cell activity in long-term survivors. He and his colleagues have preliminary data indicating that this response seems to be directed at critical functional sites on the virus.

Investigators are still not sure what to make of any of these observations. Earlier studies had not consistently linked a high CD8 *T* cell count with slow disease progression. "At the moment, it's all anecdotal," McMichael cautions.

But he is willing to speculate that the new data on CD8 *T* cells could fit in with a controversial line of work pioneered several years ago by Gene M. Shearer and Mario Clerici of the National Cancer Institute.

Those workers evoked skepticism when they proposed that there are two distinct patterns of immune system response to HIV—*T* helper 1 (TH1) and 2 (TH2). TH1-type responses, which involve the cellular arm of the immune system, might, they suggested, sometimes defeat the virus. TH1 cells stimulate such cytotoxic lymphocytes as CD8 *T* cells. TH2 responses, which lead to the production of antibodies, are coupled with progressive deterioration.

Shearer and Clerici have found that macaque monkeys can be protected against the HIV-like simian immunodeficiency virus (SIV) by exposure to doses of the virus too small to cause infection. These successful immunizations prompted TH1-type cellular responses but no antibody production. Large doses of SIV given to unimmunized animals, in contrast, caused the production of antibodies and development of disease. Recently Shearer and Clerici have shown that a substantial proportion of medical workers and babies who have been exposed to HIV but who have not become infected do show HIV-specific *T* cell responses. Their immune systems recognize the virus, even though they do not have antibodies to it.

Most researchers agree it is too early to have any confidence that the CD8 findings can be put to practical use. Shearer does have a suggestion, though, based on his work: he thinks vaccine trials should focus on very low doses of vaccine that might prompt TH1 responses, rather than trying to stimulate the production of antibodies.

Even if the CD8 link holds, the question of why some patients survive longer than others is not going to be settled soon. Warner C. Greene, director of the Gladstone Institute of Virology and Immunology at the University of California at San Francisco, says long-term survivors are not all alike. Some have high levels of virus, others low levels.

Then there are tantalizing indications that the virus can sometimes change to a more benign form: for example, certain groups of patients in which every member is known to have been infected from the same source are doing unusually well, presumably because they have a relatively harmless strain. Preliminary evidence from prostitutes in Kenya who have been studied by J. Neil Simonsen of the University of Manitoba suggests that some people have a genetic resistance to infection. Some of the women remained uninfected for two years despite many exposures.

Greene is hopeful that the phenomenon of long-term survival is a biologically meaningful one. But, he notes, "after only 14 to 15 years of the epidemic, we cannot be certain these patients are not statistical outliers."—*Tim Beardsley*

Take the Pain?

Lidocaine comes under suspicion as a carcinogen

Lidocaine, the world's most widely accepted local anesthetic, is used not only in dentistry but also in the treatment of heart arrhythmias. Astra Pharmaceutical Products in Westborough, Mass., the drug's principal manufacturer in the U.S., sells more than $4 million of lidocaine every year. The drug's popularity is not hard to understand: it is effective, and complications are uncommon. Or are they?

The National Toxicology Program, a multiagency collaboration, recognizes one of lidocaine's breakdown products in humans—2,6-dimethylaniline (2,6-

DMA)—as "clearly carcinogenic" in experimental animals. Almost half a group of 112 rats exposed to the substance in utero and then fed three milligrams of it per kilogram of food developed cancers in the nasal cavity and in other tissues. The International Agency for Research on Cancer (IARC) turned up the heat late last year when it ruled that dimethylaniline was "possibly carcinogenic to humans."

That verdict has increased the pressure on the Food and Drug Administration to reach some decisions about lidocaine. The FDA's anesthetic and life support drugs advisory committee reviewed the relevant data last August but decided not to require package inserts warning physicians and patients that lidocaine metabolites are carcinogenic in animals. The problem is that nobody is sure how much 2,6-DMA the body makes. Studies on volunteers show that most of the lidocaine in an injection winds up in the urine as 4-hydroxy-DMA, a derivative of 2,6-DMA. Some of that end product is probably formed from lidocaine through a route that does not involve 2,6-DMA. Still, the IARC concluded that lidocaine is "metabolized principally to 2,6-DMA." And 4-hydroxy-DMA is not certifiably benign: the compound is known to cause mutations in *Salmonella* bacteria.

The anesthetic and life support committee did decide to ask its parent agency to do experiments on lidocaine metabolism in slices of human liver. That research has yet to yield results. One critic of the FDA committee's decision to defer action is Jack A. Hinson, professor of toxicology at the University of Arkansas at Little Rock. Two years ago Hinson provided data to support an unsuccessful attempt by the National Cancer Institute to nominate lidocaine for evaluation by the National Toxicology Program. Hinson and others have shown that another derivative of 2,6-DMA—namely, *N*-hydroxy-DMA—binds to hemoglobin in the blood of patients who have been administered lidocaine. Hinson believes the *N*-hydroxy form could cause DNA damage. "I don't think the data should be kept from the general public," Hinson comments.

Hinson thinks research on whole animals would be more likely to produce clear-cut results about any possible danger than would the work with liver slices that the FDA has in progress. Alfred A. Nickel, a dental surgeon in Walnut Creek, Calif., who has done research on the action of lidocaine, is also skeptical about the studies, because he suspects the drug's metabolites might cause problems at ion channels found mainly in nerve tissue.

Astra's senior scientific adviser, Stig Agurell, told the FDA that metabolism in an individual exposed to between 20 and 50 typical doses of lidocaine over a lifetime would produce less than 1.7 grams of 2,6-DMA. Whether that is reassuring depends on one's point of view. According to David W. Gaylor of the National Center for Toxicological Research, if the effects of 2,6-DMA on humans are similar to those on rats, then normal usage of lidocaine might plausibly result in "one additional cancer per 100,000 persons."

Whether humans are like rats is a big if. Still, should the FDA's studies confirm that humans indeed produce 2,6-DMA and related compounds in substantial amounts, the agency may be unable to avoid some action. The danger is that rats may be unusually sensitive to the substance, so package inserts would be warning physicians away from a valuable drug without cause.

The decision may be made easier by the fact that other local anesthetics have not been called into question. Astra is already putting warnings about the possible carcinogenicity of metabolites on one of its lidocaine-containing products. —*Tim Beardsley*

Unbearable Lightness

A new theory may explain why objects tend to stay put

Suffering from inertia? Gravity got you down? You are not alone. Gravity and inertia are among the most fundamental attributes of anything possessing mass. But researchers have never attained a satisfactory understanding of the fundamental nature of gravity. Inertia has proved an even more elusive problem. Ever since Isaac Newton articulated his three laws of motion, scientists have simply accepted the existence of inertia as a given: bodies in motion remain in motion, and those at rest stay at rest, unless acted on by an outside force.

Bernhard M. Haisch of the Lockheed Palo Alto Research Laboratory, Alfonso Rueda of California State University at Long Beach and Harold E. Puthoff of the Institute for Advanced Studies in Austin, Tex., think they may at last have a clue to the process that gives rise to inertia. That process, Haisch argues, must be connected to gravitation as well, neatly unifying inertial and gravitational mass, the two ways that physicists define the mass of an object.

Writing in the February issue of *Physical Review A,* the three researchers describe inertia as the consequence of the bizarre subatomic happenings that take place in ostensibly empty space. Quantum theory predicts that, on such tiny scales, random quantum fluctuations roil the vacuum, creating a soup of virtual particles. Those particles continuously pop in and out of existence before they can be directly detected.

Haisch and his collaborators started by assuming the existence of such small-scale electromagnetic fluctuations, known as the zero-point field. They then examined the effects of the field on normal matter. In the mid-1970s several researchers showed that an object accelerating through the zero-point field should be exposed to a glow of radiation stirred up from the vacuum. Haisch, whose background is in astrophysics, wondered whether that radiation would exert a "pressure" opposing the acceleration; such a pressure exactly fits the description of inertia.

Rueda cast those ideas in mathematical form and became convinced that Haisch was on to something. "Intuitively, it made a lot of sense," he says. "The only thing that can resist the accelerating agent is the vacuum—what else is there?" He notes that the zero-point field is present at all times and in all places, which would explain the instantaneous, universal nature of inertia.

The two scientists soon teamed up with Puthoff, who had been exploring possible connections between gravity and the zero-point field. Although theorists have had considerable success understanding the other three forces of nature (electromagnetism and the two nuclear forces), "gravity has always been the oddball," Haisch reflects. Puthoff, drawing on earlier work by the late Russian physicist Andrei Sakharov, seeks to explain gravity as a long-range effect of zero-point electromagnetic fluctuations. Linking gravity to the zero-point field automatically draws inertia into the explanation and so naturally accounts for the equivalence of inertial and gravitational mass.

The ambitious, unconventional theory of inertia immediately faces a dubious audience. "I like the philosophical idea of what they are trying to do," says astrophysicist Paul S. Wesson of the University of Waterloo, "but I'm skeptical about the details." He points out, for example, that the zero-point field contains a great deal of energy. Because energy is equivalent to matter (according to Einstein's famous equation), the zero-point field might be expected to gener-

ate an intense gravitational tug, in blatant conflict with the observed structure of the cosmos. Haisch suggests that if the zero-point field gives rise to gravity, as Sakharov proposed, the energy within that field would not itself produce gravitational effects.

Peter W. Milonni of Los Alamos National Laboratory voices far more serious reservations. He worries that the theory ascribes real significance to a term describing the mass of particles, one that is normally considered to have no physical meaning and so is subtracted out of quantum-mechanical equations. And he sees "many inconsistencies" in the theory resulting from idealized or ad hoc assumptions. Nevertheless, he admits the appeal of Haisch's approach. "Sometimes wrong ideas lead people to the right one," he comments.

Haisch and his co-authors plan to reformulate their results in more conventional, quantum-mechanical terminology, which may make them more appealing. "This is the first step in a new way to look at things," Haisch explains. "You can't expect us to solve everything in one fell swoop." The three researchers also look to observational support from an upcoming experiment at the Stanford Linear Collider, which will measure the effect of electromagnetic radiation on the apparent mass of the electron.

That phenomenon raises the highly speculative prospect that the proper electromagnetic field could eliminate the inertia of an object, thereby permitting levitation. Controlling inertia may be possible, Haisch reluctantly concedes, but "God knows if it's ever going to become a reality." Still, for those people trying to make their lives a little lighter, it is nice to know that science may be able—someday—to lend a hand. *—Corey S. Powell*

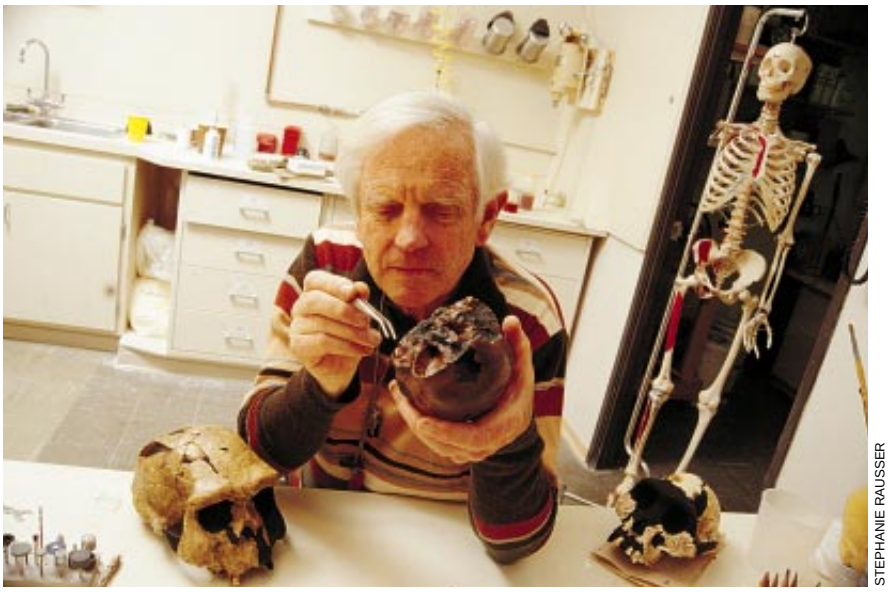
Family Matters

Revised dates invigorate debate on human origins

Dating anyone—whether romantically or paleoanthropologically—can be a tricky affair. Which is why scientists specializing in human evolution are constantly haggling over dates, contrasting one record of the ages, such as ancient DNA, with another, such as a geologic formation. The latest wrinkle in time comes from Javan fossils. The specimens, petrified bones of *Homo erectus,* have been found to be much older than many experts previously thought. Although the revised dates do not resolve controversies about the origins of humankind, they pose hard questions for one group of theorists while mollifying another.

The fossils in question were discovered in Java, one in 1936, the others in the late 1970s. Researchers determined that the age of the Mojokerto skull was about one million years old. The Sangiran series—which includes the face and cranial fragments of two hominids—was thought to be as much as 900,000 years old.

These pieces of the past seemed to fall into place in an evolutionary puzzle that emerged during the 1960s and 1970s. At that time, archaeologists and paleoanthropologists working in Africa found fossils, those of *H. erectus* among them, that were about two million years old. Bone hunters also discovered stone hand axes from 1.4 million years ago. The existence of the Acheulean tool kit, as it came to be called, suggested that after originating in Africa, *H. erectus* had become equipped enough to wander off the continent, around a million or more years ago. Although stone hand axes have never been found alongside Asian remnants, the relative youth of such fossils, including those from Java, supported the



STEPHANIE RAUSSER

GARNISS H. CURTIS holds a model of the Mojokerto skull. Thirty years ago Curtis placed the fossil's age at 1.9 million years, but many paleoanthropologists argued that it was a million years younger. Curtis and his colleagues at the Institute of Human Origins recently determined that the fossil is indeed 1.8 million years old.

view that *H. erectus* reached Indonesia well after it had arisen in Africa.

Some paleoanthropologists, however, never liked this idea. Yves Coppens of the College of France notes that the earliest humans—including *H. habilis,* which evolved before *H. erectus*—lie in the African fossil record sometime between three and 2.5 million years ago. "As soon as *Homo* appeared, it had its brain and its behaviors, and among those behaviors was mobility," Coppens explains. Even at the pace of a mere mile per year, it would take only 8,000 or so years to reach Indonesia from, say, Ethiopia. Alan Walker, an anatomist at the Johns Hopkins University School of Medicine, agrees: "I suspect that once it developed these behaviors, *Homo erectus* got out of Africa right away."

Another expert, one intimate with the Mojokerto fossil, was convinced that *H. erectus* had been peripatetic from the outset; he finally seems to have proved it. Garniss H. Curtis, a geochronologist now at the Institute of Human Origins in Berkeley, Calif., dated the Javan skull in the 1960s. Using a technique that measured the rate of decay of potassium into argon, Curtis concluded that the fossils were about 1.9 million years old, nearly a million years older than most of his colleagues considered them to be. The reaction among anthropologists to both date and technique was not favorable. "It really bugged Curtis," notes Carl C. Swisher III, another geochronologist at the institute and first author on the *Science* article that reported the update.

Following his 30-year-old hunch, Curtis, along with Swisher and their colleagues in Indonesia, recently applied a more sophisticated technique to sediments gathered near and on the fossils. "Much like *Homo erectus* was, we were equipped with new tools," Swisher comments. The newer technique, a refined version of the old one, also measures the decay of potassium into argon. The researchers revealed that the Mojokerto fossil was, in fact, 1.8 million years old and that the Sangiran assemblage was 1.6 million years old.

That is, the Mojokerto sediment is 1.8 million years old. As Francis Brown, a geologist at the University of Utah, notes, there is a possibility that older sediments could have intruded into a younger formation. Unlike many of the fossil finds in Africa, the Javan fossils pose a paleontological difficulty. Many dates in Africa are relatively clear and can be compared because they lie in stratigraphic order. "The sequences give you much more certainty," Walker says. Because Java is a volcanic island and is not as old, it is very hard to see the stratigraphy clearly.

Swisher and his team plan to return to Java soon to take more samples and verify their findings. In response to the publicity surrounding the new date, Swisher says the Indonesian government facilitated access to the fossils, "which are usually treated like the crown jewels." With greater access, Swisher intends to collect more ancient grains and do more analyses.

The revised dates may ultimately have implications for two theories that seek to explain the emergence of modern humans. Both of these views agree that a wave of early humans belonging to *H. erectus* left Africa about one and a half million years ago and spread throughout Eurasia and the Western Pacific. But from this point on, the theories diverge. According to the out-of-Africa theory—also called the Eve hypothesis and Noah's Ark—these ancestors were replaced by a later wave of *H. sapiens.* These modern humans left Africa some 100,000 years ago and traveled across the planet, differentiating along the way and giving rise to the various races that we are today. The other side, called the multiregional theory, argues that today's *H. sapiens* evolved simultaneously from the original *H. erectus* populations.

The new date for the Javan fossils does not clarify which, if either, of these hypotheses is correct. But to some specialists, including Swisher, the date could strengthen the out-of-Africa model. If human ancestors were in Asia about two million years ago and were not replaced by a relatively recent wave of *H. erectus,* they would have had time to evolve quite a different genetic makeup than we see today. (The genetic proximity of all modern humans has been the foundation of the out-of-Africa theory.)

A resolution or, perhaps more likely, another twist in the tale will have to wait for more fossils and more dates. For now, "this offers a slightly different slant on things. People are used to hearing about new fossils from Africa," Swisher remarks. "We are just shifting the focus." Brown adds: "We were all attached to Africa, but soon we'll get used to this idea, and it'll be fine. Just as we did to the ideas of continental drift and plate tectonics."—*Marguerite Holloway*

Vishnu Violated

Jetties threaten world's largest gathering of Olive Ridley turtles

One week this February, 520,000 Olive Ridley turtles emerged from the surf after dark to lay their eggs on a six-mile-long strip of beach in Orissa, in eastern India. By dawn they had left; their hatchlings followed in March. Where the turtles came from, how they spent their 30-odd years from infancy to adulthood and why they then found their way across thousands of miles of ocean to this beach, no one knows. Now we may never know.

SCIENTIFIC AMERICAN

CORRESPONDENCE

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

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

Index of articles since 1948 available in electronic format. Write SciDex®, Scientific American, 415 Madison Avenue, New York, NY 10017-1111, fax (212) 980-8175 or call (800) 777-0444.

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

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

Advertising correspondence should be addressed to Advertising Manager, Scientific American, 415 Madison Avenue, New York, NY 10017-1111, or fax (212) 754-1138.

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

The government of Orissa is completing a fishing jetty within eight miles of the turtle rookery. Three more jetties are planned. The beach, hosting the world's largest congregation of nesting turtles—610,000 were counted in 1991—borders the Bhitarkanika mangrove sanctuary and is itself relatively free of human predation. Local inhabitants eat neither the creatures nor their eggs: to Hindus, turtles are one of the 10 incarnations of Vishnu. But while mating offshore in the two months before they nest, thousands of turtles are snagged and drowned in gill and drag nets. Each jetty will add at least 500 trawlers, devastating this population of already endangered Olive Ridleys.

But there is a human side to this story. Orissa, one of the poorest of Indian states, has recently initiated several projects to develop its untapped mineral and biological wealth. A few of these efforts—such as a shrimp fishery that would have poisoned Chilka Lake, home to 150 species of native and migratory birds—have been stopped. Ironically, that accommodation appears to have hardened the government's resolve to complete the current construction. The government is also building a connecting road through—and relocating villages into—the sanctuary, in defiance of national laws. According to Banka Behary Das, a local activist, the politicians are intent on gratifying their "vote bank"—immigrant fishermen from Bangladesh, who do not share the natives' respect for turtles.

"The Olive Ridleys seem to arrive from the Indian Ocean," says Binod C. Chaudhury of the Wildlife Institute of India in Dehra Dun. "Ideally, we would radio-collar them and track them via satellite." In reality, plastic nooses await. *—Madhusree Mukerjee in Orissa*

Amazing Grace

Physicists find evidence for halos around nuclei

At the center of an atom lies the nucleus, a bundle of protons and neutrons less than one millionth of a millionth of a centimeter in size. Because the nucleus is so small and the forces that hold it together are so complex, physicists have long wondered how its constituent nucleons are arranged. In recent years, researchers have found complementary clues that hint at a surprising structure: certain radioactive nuclei exhibit a halo of nucleons orbiting a stable core. As a result, these nuclei take up much more space than do ordinary ones having the same mass.

Unfortunately, explanations for this halo structure have been as hard to pin down as the nuclei in which they occur. "Binding energies are key to the whole issue," says Maria J. G. Borge of CERN, the European laboratory for particle physics near Geneva. In those nuclei that have a high binding energy, the attractive strong force, which draws both protons and neutrons together, entirely overwhelms the strength of the repulsive electrical force, which drives the like-charged protons apart.

Light nuclei achieve this balance when they have an equal number of protons and neutrons. In radioactive nuclei, however, which have too many or too few neutrons, the scale of nuclear forces tips. Consequently, the nucleons are more loosely bound to one another and scramble to find a stable configuration. In some cases, as Borge and others have seen, these nuclei find stability only when a few of their extra nucleons part with the nuclear core and begin to orbit, forming a halo.

The first experimental evidence of the phenomenon came in 1985. A team at Lawrence Berkeley Laboratory conducted experiments in which they bombarded aluminum and copper targets using a beam of lithium 11 nuclei, each of which has eight neutrons and three protons. To determine the radius of lithium 11 nuclei, the investigators measured the rate at which they reacted with the targets: a larger nuclear ball would be more likely to make contact. Their results indicated that the nucleus of lithium 11 had a radius of 3.16 femtometers, comparable to the size of a nucleus having nearly three times as many nucleons. Recently Borge and her colleagues discovered further proof for neutron halos by examining the decay of helium 6, a neutron-rich isotope.

Borge's group and others have found that proton halos may exist as well. Proton halos are even more unusual than neutron halos. First, they can arise only in light nuclei for which the total electrostatic repulsion between the protons is relatively small. In addition, a proton halo must hover much closer to the core than does a neutron halo, well in reach of the short-range strong force.

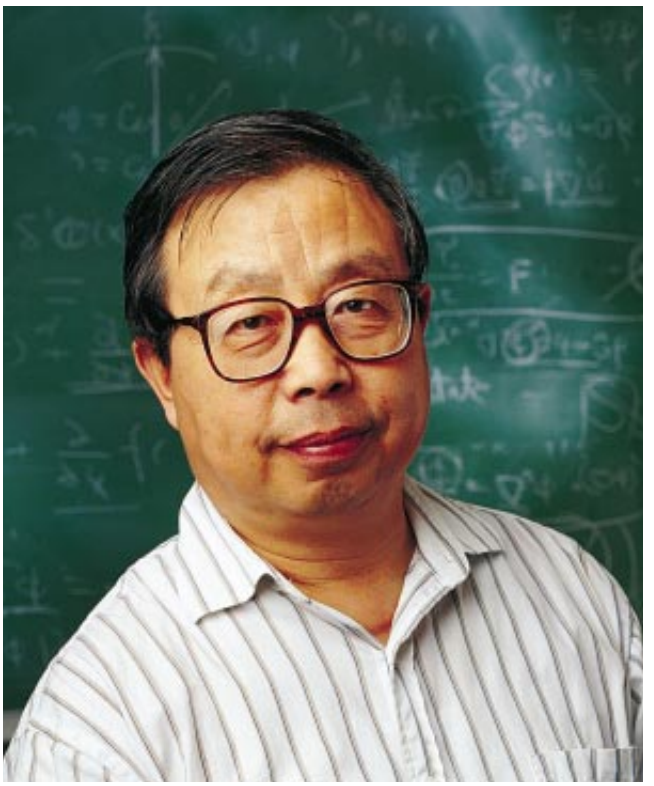
The search for more hallowed nuclei continues, but Borge doubts that many exist. "We may have to wait for a new generation of results before we know whether what we are seeing is a reflection of what was really there in the beginning or if our reactions change the picture somehow," she adds. When that question is answered, some nuclei now purported to hold a halo may well fall from grace. *—Kristin Leutwyler*

Fundamental Rights, Fundamental Physics

A few days after I had made arrangements to interview Fang Lizhi, I was surprised to see a paper on cosmology he had written in that journal of very serious science, *Physical Review Letters*. Doesn't a celebrated, exiled political dissident have more pressing concerns than stochastic fluctuations in the universe? The question, I later realized, is rather naive: Fang is a scientist first. His rise as a passionate advocate of democracy in China, which culminated in his exile in 1990, stems from being a physicist. "If you study anything, you must raise questions," he says. "But authorities in communist countries do not like you to raise questions."

Of course, Fang Lizhi (pronounced Fong Lee-JUR) is not raising any radical points; he is merely stating facts. But by openly accusing China's ruling party of suppression, incompetence and corruption, he has infuriated the authorities in a culture where appearances, order and national pride are everything. Fang indicates that he has even incurred the personal wrath of China's rapidly failing but still paramount leader, Deng Xiaoping. The physicist's penchant for saying publicly what other intellectuals are thinking may be threatening the statesman's place in history as the great reformer who led China out of the abyss of the Cultural Revolution into the modern age. "I think Deng sees me as his personal enemy, not only as the party's enemy," the 58-year-old Fang concludes. "All my personal issues are decided by Deng." Those issues, Fang maintains, even include his Beijing apartment, which remains untouched because Deng has not given orders to act on it. "I still pay the rent," chuckles Fang, who is noted for his keen wit and ready laugh.

Paying rent on an apartment to which he cannot return is only one example of how Fang turns the commonplaces of daily life into acts that are full of meaning. In a plain shirt and slacks, as he sits in his nondescript office at the University of Arizona at Tucson, Fang could be any professor of astrophysics. Indeed, he projects an air of utter normality. His immediate family all live in the U.S. His older son works as a computer systems engineer at Chrysler Corporation. His other son teaches physics at a community college in Wisconsin. But China is never far away from his thoughts. Deng decides the big policy matters, like U.S.-China relations, Fang says ominously at one point. Suddenly, he bursts out laughing. "And other big issues, like my apartment!"



TIM FULLER

EXILED ASTROPHYSICIST Fang Lizhi believes democracy will ultimately take hold in the People's Republic of China.

The Tiananmen Square demonstrations in 1989 and the brutal massacre that ensued were only the final straws that led to banishment for Fang and his wife, Li Shuxian. Since assuming power after World War II, China's communist government has engaged in a kind of love-hate relationship with intellectuals. The Communist Party officials recognize the need for scientists to modernize the country. But the habit of intellectual independence in scientists seems to threaten the party leadership. For years, they have alternately cultivated and suppressed the intelligentsia.

The first swing began one year after Fang, who developed his passion for science in his middle school years after building a radio, graduated from Beijing University with a degree in physics. In 1957 Mao Zedong invited intellectuals to air their thoughts on reform. As the chairman had predicted, many flowers did indeed bloom, among them Fang, who argued that the educational system needed to be changed so that politics would not interfere with science. Stunned by the extent of the criticism from the intellectuals, Mao abruptly launched the anti-rightist campaign, designed to bring the detractors back in line. Unwilling to recant, Fang was expelled from the party and forced to work in a labor camp for eight months. "Even then, I still believed in communism," Fang recalls. "I believed in Marxism, that Mao's thought would be good for Chinese society."

After doing his penance, Fang returned to academic life, conducting research on quantum mechanics, solid-state physics and lasers. Young and talented, he advanced quickly, becoming a lecturer in Beijing at the University of Science and Technology. But Fang was growing disillusioned with Maoism. "I lost my belief in the authorities in the 1960s," he says. "Step by step, I found that Mao's group just wanted to keep power." The Cultural Revolution began to crystallize Fang's disenchantment. Fearing that alien, capitalist influences were sapping China's revolutionary fervor, Mao ordered the reformation of anyone and the extirpation of anything remotely considered to be bourgeois. Fang was sent down for "reeducation" in the

southern countryside, where he toiled in a coal mine and on a railroad for several months.

The forced labor actually helped to turn Fang's mind to astrophysics. "During the Cultural Revolution, you were forbidden to read any books, except for Mao's small red book," Fang states. He did, however, manage to smuggle along a copy of *Classical Theory of Fields,* by the Russian physicist Lev Landau. "At that time, I had no other book, so I read it several times." The second half of the book, devoted to general relativity, inspired Fang to tackle cosmology.

As fate would have it, astrophysics made Fang stand out even more. "My first article on cosmology appeared in 1972," he recollects. Technical and obscure, only a few copies were ever circulated. It would easily have been ignored but for the ideological fixation of the government. Because mainstream modern cosmology regards the universe as having an origin, Maoist ideologues believed it therefore demanded the existence of a deity—a presumption antithetical to the Marxist worldview. The clash led to a public denouncement of Fang's piece. "The authorities wrote a very long article in a newspaper," Fang says, relishing the irony: "It was the first time I was getting famous."

Although Fang was given an academic post after working in the labor camp, the Communist Party did not fully accept him again until 1978, after the Cultural Revolution had come to an end. To help modernize China, Mao's successor, Deng Xiaoping, invited many persecuted intellectuals to rejoin the party. Fang did so, because he felt that change was best accomplished from within: "If you wanted to do good, you did not refuse." He became China's youngest full professor and in 1984 achieved the position of vice president at the University of Science and Technology.

That decade saw Fang's ascendance to political prominence begin its steepest climb. "I think my contribution to the Chinese democracy movement is the speeches I gave during the 1980s," Fang reflects. "The ideas in those speeches were very common, very basic ideas of human rights." They galvanized a generation of young intellectuals. (Many of his writings appear in his collection of essays *Bringing Down the Great Wall*.)

The party at first tolerated Fang's dissenting voice. "During the 1980s, I was quite famous, not only politically but also as a professor of physics, so the Chinese authorities were afraid of how to treat me," Fang explains. But as Fang became increasingly influential, the party's patience eroded. By the end of 1986, student demonstrations had rocked several campuses, and the authorities retaliated by rooting out those whom they believed to be responsible. They expelled Fang from the party again and reassigned him to the Beijing Observatory. A blistering series of articles in state-run papers blamed Fang for much of the unrest.

Fang believes Deng himself dismissed him from the party. "If you want to expel a party member, there must first be a vote by local authorities. Normally, it should take several months to vote, approve, vote. In the bylaws, sometimes you can immediately expel a member, but only in emergency situations. Which emergency situations? During a war or an earthquake," Fang utters evenly and then adds, "But I'm not a war or an earthquake!" He cannot resist a guffaw: "Maybe political earthquake."

The party's frustration with Fang peaked in 1989. In January, Fang wrote to Deng demanding the release of Wei Jingsheng, a jailed political dissident sometimes compared to Lech Walesa because of his labor background. (Wei was released last year.) "The letter got a strong response from the whole country," Fang remarks. "Many famous intellectuals joined and supported me." In February, Fang and his wife were prevented from attending a banquet in honor of President George Bush, who was then visiting the country.

Fang declares that he played no direct role in the Tiananmen Square demonstrations that began in April, yet he believed the party had identified him as a key instigator: "Even on television, they showed my picture." After the massacre on June 4, Fang and Li were encouraged by friends to seek refuge in the U.S. Embassy, where they remained as the Deng regime and the U.S. State Department looked for a face-saving means to free Fang.

"That was a difficult time. But I was still productive. I published several scientific papers." Fang digs up an article he wrote on the distribution of redshifts of quasars. "This was totally done in the embassy." The affiliation lists the University of Rome. "This address is not the real address," Fang explains. "I wanted to put in 'U.S. Embassy in Beijing,' but the State Department people didn't like that."

The opportunity to leave the embassy came a year later, when Fang complained of minor heart palpitations he attributes to excessive coffee drinking. U.S. officials exaggerated the magnitude of the condition, leaking a story that Fang had suffered a heart attack.

Although the exiled Fang does not exert the kind of influence he wielded a decade ago, his plight has helped spotlight China's poor human-rights record. Letters smuggled out of labor camps and accounts given by former prisoners now in the U.S. have told of beatings, starvation and poor medical care. "Before 1989 very few Westerners paid attention to human-rights abuses in China," Fang observes. The matter has now become a major point of friction. In June, President Bill Clinton must decide whether to renew China's trade status as most-favored nation. Such status confers low tariffs on imported goods, giving the U.S. cheap consumer items and providing China with hard currency and about one million jobs. The amount of trade is estimated to be worth about $40 billion.

But as a condition for renewal, the White House has demanded that China improve its human-rights record. "In my view, the conditions proposed by Clinton are not very tough," Fang insists. They include inspections of prisons, freedom for some political prisoners and an end to the jamming of American radio broadcasts. In response, China has recently taken a belligerent stance on what it perceives as meddling.

So much is at stake that some U.S. leaders have called for a separation of human rights and trade agreements. Fang scoffs at the suggestion. "Yes, I hear that some congressmen want business as usual. I don't think this is a good idea." He argues that separating the issues may only serve to lessen pressures that push the political system in a democratic direction, thereby ultimately harming U.S. interests.

Asked when he thinks he might return to his country, Fang merely states a fact: he cannot go back so long as the 89-year-old Deng lives. Fang keeps in touch with friends and relatives in China, where his brother, an aerospace engineer, is a party member. ("Of course, we have different opinions.") He can call his mother in Beijing, but he admits to being careful how he writes to other people. ("They check my letters.") The Internet also keeps him in close contact. "I have e-mail, so I can communicate with China. I even published a paper together with some colleagues."

Fang says he will go back if the situation allows. Many party members, he feels, do not really subscribe to Maoism anymore. Moreover, he states, if you ask the peasants in the remote areas, who are often described as apolitical, they admit their desire to choose their local leaders directly, thus demonstrating they have the capability for democracy. "In the long term," the exiled physicist asserts, "it will definitely come." —*Philip Yam*

Is Homosexuality Biologically Influenced?

It had four hands and an equal number of legs…. It had four ears and two organs of generation and everything else to correspond. —Plato, *Symposium*

For more than two decades, researchers have presented evidence for biological mechanisms underlying male homosexuality. Simon LeVay of the Institute of Gay and Lesbian Education and Dean H. Hamer of the National Institutes of Health summarize their work in the first of the following two articles. William Byne of the Albert Einstein College of Medicine argues that current data lack substance.

LeVay, Hamer and Byne join distinguished company. Almost 2,400 years ago Plato told in his *Symposium* how humanity once consisted of three sexes rather than two: its members were joined in pairs consisting of two men, two women, or a man and a woman. Zeus cut each pair apart to diminish their power and to teach them to fear the gods. Humans thus spend their time on the earth searching for their other half, with whom they can merge in love.

The disposition of these people differed, Plato wrote, according to original pairing: those whose sex had once been mixed were obsessed by coupling and often became adulterers, whereas people sprung from single-sex pairs were more fitted for the everyday business of the world. In particular, men whose bond was with another man were most suited for government and leadership. This story is perhaps the earliest "biological" explanation of sexual orientation; it also demonstrates how attitudes toward homosexual behavior vary from culture to culture. Plato's myth furnishes the theme and text for the photographs by Joyce Tenneson that accompany the two articles.

Evidence for a Biological Influence in Male Homosexuality

Two pieces of evidence, a structure within the human brain and a genetic link, point to a biological component for male homosexuality

by Simon LeVay and Dean H. Hamer

Most men are sexually attracted to women, most women to men. To many people, this seems only the natural order of things—the appropriate manifestation of biological instinct, reinforced by education, religion and the law. Yet a significant minority of men and women—estimates range from 1 to 5 percent—are attracted exclusively to members of their own sex. Many others are drawn, in varying degrees, to both men and women.

How are we to understand such diversity in sexual orientation? Does it derive from variations in our genes or our physiology, from the intricacies of our personal history or from some confluence of these? Is it for that matter a choice rather than a compulsion?

Probably no one factor alone can elucidate so complex and variable a trait as sexual orientation. But recent laboratory studies, including our own, indicate that genes and brain development play a significant role. How, we do not yet know. It may be that genes influence the sexual differentiation of the brain and its interaction with the outside world, thus diversifying its already vast range of responses to sexual stimuli.

The search for biological roots of sexual orientation has run along two broad lines. The first draws on observations made in yet another hunt—that for physical differences between men's and women's brains. As we shall see, "gay" and "straight" brains may be differentiated in curiously analogous fashion. The second approach is to scout out genes by studying the patterns in which homosexuality occurs in families and by directly examining the hereditary material, DNA.

Researchers have long sought within the human brain some manifestation of the most obvious classes into which we are divided—male and female. Such sex differentiation of the brain's structure, called sexual dimorphism, proved hard to establish. On average, a man's brain has a slightly larger size that goes along with his larger body; other than that, casual inspection does not reveal any obvious dissimilarity between the sexes. Even under a microscope, the architecture of men's and women's brains is very similar. Not surprisingly, the first significant observations of sexual dimorphism were made in laboratory animals.

Of particular importance is a study of rats conducted by Roger A. Gorski of the University of California at Los Angeles. In 1978 Gorski was inspecting the rat's hypothalamus, a region at the base of its brain that is involved in instinctive behaviors and the regulation of metabolism. He found that one group of cells near the front of the hypothalamus is several times larger in male than in female rats. Although this cell group is very small, less than a millimeter across even in males, the difference between the sexes is quite visible in appropriately stained slices of tissue, even without the aid of a microscope.

Gorski's finding was especially interesting because the general region of the hypothalamus in which this cell group occurs, known as the medial preoptic area, has been implicated in the generation of sexual behavior—in particular, behaviors typically displayed by males. For example, male monkeys with damaged medial preoptic areas are apparently indifferent to sex with female monkeys, and electrical stimulation of this region can make an inactive male monkey approach and mount a female. It should be said, however, that we have yet to find in monkeys a cell group analogous to the sexually dimorphic one occurring in rats.

Nor is the exact function of the rat's sexually dimorphic cell group known. What is known, from a study by Gorski and his co-workers, is that androgens—typical male hormones—play a key role in bringing about the dimorphism during development. Neurons within the cell group are rich in receptors for sex hormones, both for androgens—testosterone is the main representative—and for female hormones known as estrogens. Although male and female rats initially have about the same numbers of neurons in the medial preoptic area, a surge of testosterone secreted by the testes of male fetuses around the time of birth acts to stabilize their neuronal population. In females the lack of such a surge allows many neurons in this cell group to die, leading to the typically smaller structure. Interestingly, it is only for a few days before and after birth that the medial preoptic neurons are sensitive to androgen; removing an-

SIMON LEVAY and DEAN H. HAMER investigate the biological roots of homosexuality. LeVay earned a doctorate in neuroanatomy at the University of Göttingen in Germany. In 1971 he went to Harvard University to work with David Hubel and Torsten Wiesel on the brain's visual system. He moved to the Salk Institute for Biological Studies in San Diego in 1984 to head the vision laboratory. In 1992 he left Salk to found the Institute of Gay and Lesbian Education. Hamer received his Ph.D. in biological chemistry from Harvard in 1977. For the past 17 years, he has been at the National Institutes of Health, where he is now chief of the section on gene structure and regulation at the National Cancer Institute. He studies the role of genes both in sexual orientation and in complex medical conditions, including progression of HIV and Kaposi's sarcoma.

drogens in an adult rat by castration does not cause the neurons to die.

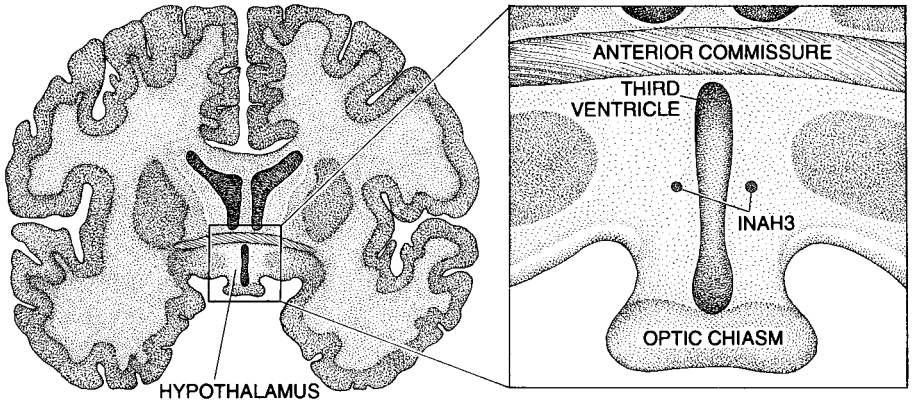
Gorski and his colleagues at U.C.L.A., especially his student Laura S. Allen, have also found dimorphic structures in the human brain. A cell group named INAH3 (derived from "third interstitial nucleus of the anterior hypothalamus") in the medial preoptic region of the hypothalamus is about three times larger in men than in women. (Notably, however, size varies considerably even within one sex.)

In 1990 one of us (LeVay) decided to check whether INAH3 or some other cell group in the medial preoptic area varies in size with sexual orientation as well as with sex. This hypothesis was something of a long shot, given the prevailing notion that sexual orientation is a "high-level" aspect of personality molded by environment and culture. Information from such elevated sources is thought to be processed primarily by the cerebral cortex and not by "lower" centers such as the hypothalamus.

LeVay examined the hypothalamus in autopsy specimens from 19 homosexual men, all of whom had died of complications of AIDS, and 16 heterosexual men, six of whom had also died of AIDS. (The sexual orientation of those who had died of non-AIDS causes was not determined. But assuming a distribution similar to that of the general populace, no more than one or two of them were likely to have been gay.) LeVay also included specimens from six women whose sexual orientation was unknown.

[*Zeus*] *cut the members of the human race in half, like fruit that is to be dried and preserved, or like eggs that are cut with a hair.* —Plato, *Symposium*

HYPOTHALAMUS of the human brain was examined for differences related to sexual orientation. The hypothalamus of each of the 41 subjects was stained to mark neuronal cell groups. The cell group termed INAH3 in the medial preoptic area was more than twice as large in the men as it was in the women. INAH3 also turned out to be two to three times larger in straight men than it was in gay men (*micrographs at far right*). This finding suggests a difference related to male sexual orientation about as great as that related to sex.



After encoding the specimens to eliminate subjective bias, LeVay cut each hypothalamus into serial slices, stained these to mark the neuronal cell groups and measured their cross-sectional areas under a microscope. Armed with information about the areas, plus the thickness of the slices, he could readily calculate the volumes of each cell group. In addition to Allen and Gorski's sexually dimorphic nucleus INAH3, LeVay examined three other nearby groups—INAH1, INAH2 and INAH4.

Like Allen and Gorski, LeVay observed that INAH3 was more than twice as large in the men as in the women. But INAH3 was also between two and three times larger in the straight men than in the gay men. In some gay men, as in the example shown at the top of the opposite page, the cell group was altogether absent. Statistical analysis indicated that the probability of this result's being attributed to chance was about one in 1,000. In fact, there was no significant difference between volumes of INAH3 in the gay men and in the women. So the investigation suggested a dimorphism related to male sexual orientation about as great as that related to sex.

A primary concern in such a study is whether the observed structural differences are caused by some variable other than the one of interest. A major suspect here was AIDS. The AIDS virus itself, as well as other infectious agents that take advantage of a weakened immune system, can cause serious damage to brain cells. Was this the reason for the small size of INAH3 in the gay men, all of whom had died of AIDS?

Several lines of evidence indicate otherwise. First, the heterosexual men who died of AIDS had INAH3 volumes no different from those who died of other causes. Second, the AIDS victims with small INAH3s did not have case histories distinct from those with large INAH3s; for instance, they had not been ill longer before they died. Third, the other three cell groups in the medial preoptic area—INAH1, INAH2 and INAH4—turned out to be no smaller in the AIDS victims. If the disease were having a nonspecific destructive effect, one would have suspected otherwise. Finally, after completing the main study, LeVay obtained the hypothalamus of one gay man who had died of non-AIDS causes. This specimen, processed "blind" along with several specimens from heterosexual men of similar age, confirmed the main study: the volume of INAH3 in the gay man was less than half that of INAH3 in the heterosexual men.

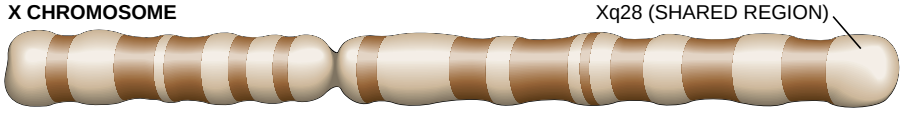
One other feature in brains that is related to sexual orientation has been reported by Allen and Gorski. They found that the anterior commissure, a bundle of fibers running across the midline of the brain, is smallest in heterosexual men, larger in women and largest in gay men. After correcting for overall brain size, the anterior commissure in women and in gay men were comparable in size.

What might lie behind these apparent correlations between sexual orientation and brain structure? Logically, three possibilities exist. One is that the structural differences were present early in life—perhaps even before birth—and helped to establish the men's sexual orientation. The second is that the differences arose in adult life as a result of the men's sexual feelings or behavior. The third possibility is that there is no causal connection, but both sexual orientation and the brain structures in question are linked to some third variable, such as a developmental event during uterine or early postnatal life.

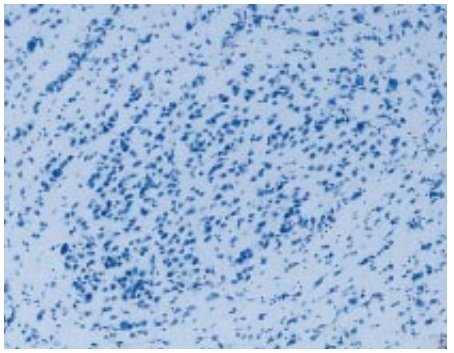
We cannot decide among these possibilities with any certainty. On the basis of animal research, however, we find the second scenario, that the structural differences came about in adulthood, unlikely. In rats, for example, the sexually dimorphic cell group in the medial preoptic area appears plastic in its response to androgens during early brain development but later is largely resistant to change. We favor the first possibility, that the structural differences arose during the period of brain development and consequently contributed to sexual behavior. Because the medial preoptic region of the hypothalamus is implicated in sexual behavior in monkeys, the size of INAH3 in men may in-

Family Tree Studies and the X Chromosome

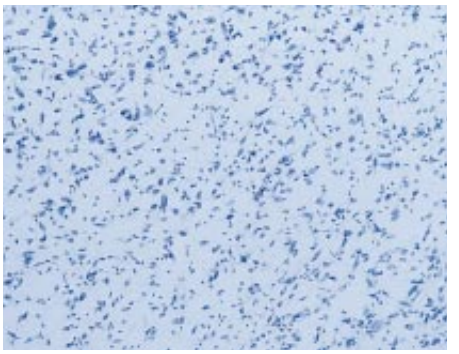
Family trees of male sexual orientation show the rates of homosexuality (*darker brown*) in maternally related males. Rates in paternal relatives were not significantly above the average population rate of 2 percent. This finding raised the possibility of involvement of the X chromosome (*shown below*). Males have two sex chromosomes—a Y inherited from the father and an X from the mother. Thus, a trait inherited through the mother's side might logically be influenced by a gene on one of her X chromosomes (*indicated here in red*). In fact, further experiments showed that one small area at the tip of the X chromosome—Xq28—was shared by a large percentage of gay brothers.



HETEROSEXUAL MAN



HOMOSEXUAL MAN



deed influence sexual orientation. But such a causal connection is speculative at this point.

Assuming that some of the structural differences related to sexual orientation were present at birth in certain individuals, how did they arise? One candidate is the interaction between gonadal steroids and the developing brain; this interaction is responsible for differences in the structure of male and female brains. A number of scientists have speculated that atypical levels of circulating androgens in some fetuses cause them to grow into homosexual adults. Specifically, they suggest that androgen levels are unusually low in male fetuses that become gay and unusually high in female fetuses that become lesbian.

A more likely possibility is that there are intrinsic differences in the way individual brains respond to androgens during development, even when the hormone levels are themselves no different. This response requires a complex molecular machinery, starting with the androgen receptors but presumably including a variety of proteins and genes whose identity and roles are still unknown.

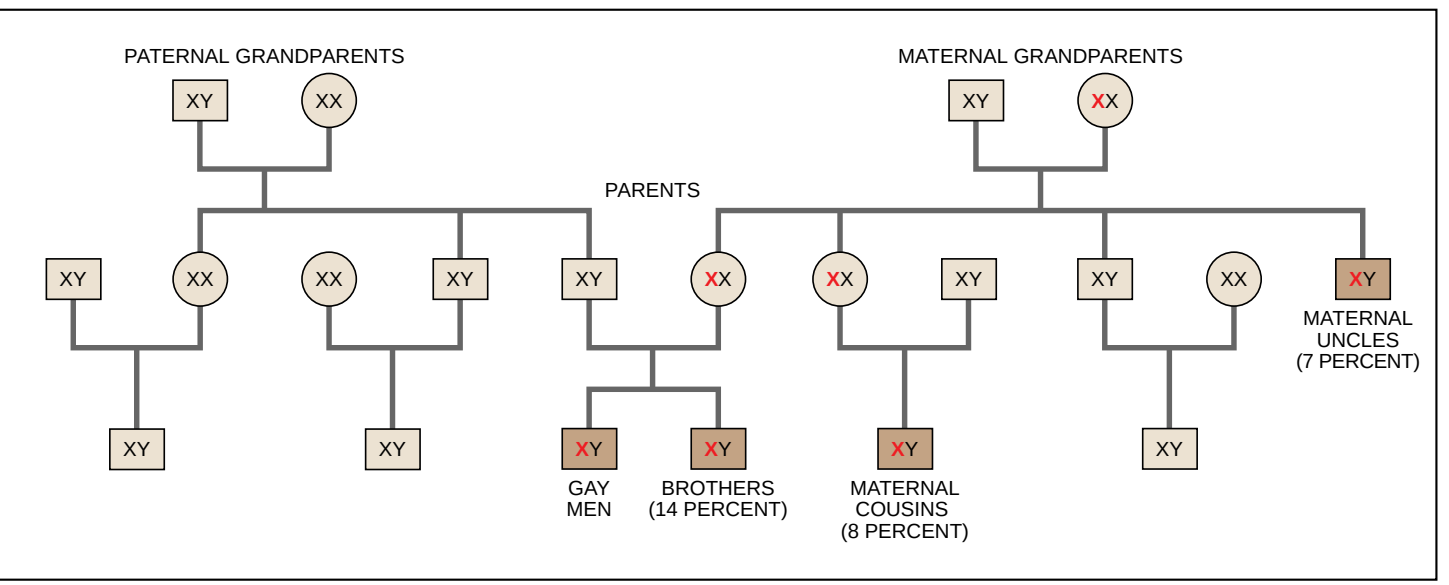
At first glance, the very notion of gay genes might seem absurd. How could genes that draw men or women to members of the same sex survive the Darwinian screening for reproductive fitness? Surely the parents of most gay men and lesbians are heterosexual? In view of such apparent incongruities, research focuses on genes that sway rather than determine sexual orientation. The two main approaches to seeking such genes are twin and family studies and DNA linkage analysis.

Twin and family tree studies are based on the principle that genetically influenced traits run in families. The first modern study on the patterns of homosexuality within families was published in 1985 by Richard C. Pillard and James D. Weinrich of Boston University. Since then, five other systematic studies on the twins and siblings of gay men and lesbians have been reported.

The pooled data for men show that about 57 percent of identical twins, 24 percent of fraternal twins and 13 percent of brothers of gay men are also gay. For women, approximately 50 percent of identical twins, 16 percent of fraternal twins and 13 percent of sisters of lesbians are also lesbian. When these data are compared with baseline rates of homosexuality, a good amount of family clustering of sexual orientation becomes evident for both sexes. In fact, J. Michael Bailey of Northwestern University and his co-workers estimate that the overall heritability of sexual orientation—that proportion of the variance in a trait that comes from genes—is about 53 percent for men and 52 percent for women. (The family clustering is most obvious for relatives of the same sex, less so for male-female pairs.)

To evaluate the genetic component of sexual orientation and to clarify its mode of inheritance, we need a systematic survey of the extended families of gay men and lesbians. One of us (Hamer), Stella Hu, Victoria L. Magnuson, Nan Hu and Angela M. L. Pattatucci of the National Institutes of Health have initiated such a study. It is part of a larger one by the National Cancer Institute to investigate risk factors for certain cancers that are more frequent in some segments of the gay population.

Hamer and his colleagues' initial survey of males confirmed the sibling results of Pillard and Weinrich. A brother of a gay man had a 14 percent likelihood of being gay as compared with 2 percent for the men without gay brothers. (The study used an unusually stringent definition of homosexuality, leading to the low average rate.) Among more distant relatives, an unexpected pattern showed up: maternal uncles had a 7 percent chance of being gay, whereas sons of maternal aunts had an 8 percent chance. Fathers, paternal uncles and the three other types of cousins showed no correlation at all.



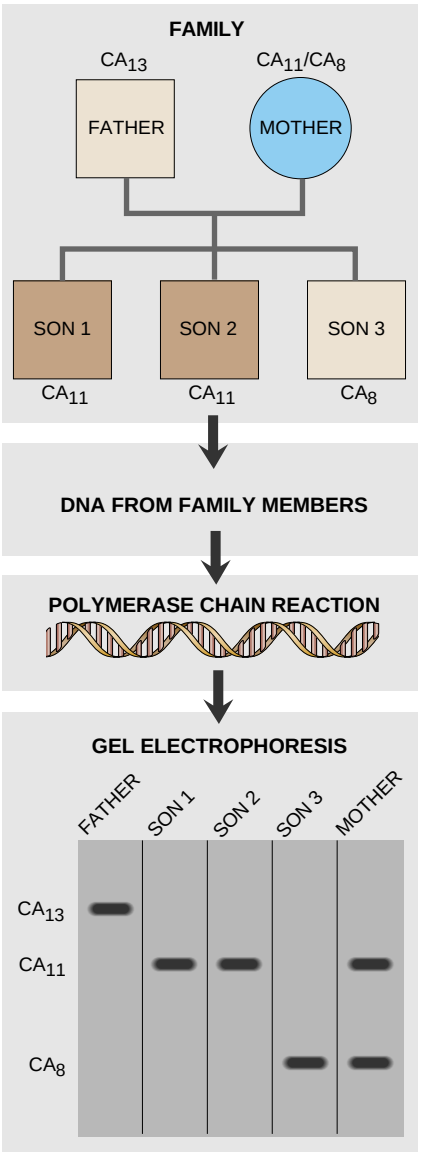
Although this study pointed to a genetic component, homosexuality occurred much less frequently than a single gene inherited in simple Mendelian fashion would suggest. One interpretation, that genes are more important in some families than in others, is borne out by looking at families having two gay brothers. Compared with randomly chosen families, rates of homosexuality in maternal uncles increased from 7 to 10 percent and in maternal cousins from 8 to 13 percent. This familial clustering, even in relatives outside the nuclear family, presents an additional argument for a genetic root to sexual orientation.

Why are most gay male relatives of gay men on the mother's side of the family? One possibility—that the subjects somehow knew more about their maternal relatives—seems unlikely because opposite-sex gay relatives of gay males and lesbians were equally distributed between both sides of the family. Another explanation is that homosexuality, while being transmitted by both parents, is expressed only in one sex—in this case, males. When expressed, the trait reduces the reproductive rate and must therefore be disproportionately passed on by the mother. Such an effect may partially account for the concentration of gay men's gay relatives on the maternal side of the family. But proof of this hypothesis will require finding an appropriate gene on an autosomal chromosome, which is inherited from either parent.

A third possibility is X chromosome linkage. A man has two sex chromosomes: a Y, inherited from his father, and an X, cut and pasted from the two X chromosomes carried by his mother. Therefore, any trait that is influenced by a gene on the X chromosome will tend to be inherited through the mother's side and will be preferentially observed in brothers, maternal uncles and maternal cousins, which is exactly the observed pattern.

To test this hypothesis, Hamer and his colleagues embarked on a linkage study of the X chromosome in gay men. Linkage analysis is based on two principles of genetics. If a trait is genetically influenced, then relatives who share the trait will share the gene more often than is expected by chance—this is true even if the gene plays only a small part. Also, genes that are close together on a chromosome are almost always inherited together. Therefore, if there is a gene that influences sexual orientation, it should be "linked" to a nearby DNA marker that tends to travel along with it in families. For traits affected by only one gene, linkage can precisely locate the gene on a chromosome. But for complex traits such as sexual orientation, linkage also helps to determine whether a genetic component really exists.

To initiate a linkage analysis of male sexual orientation, the first requirement was to find informative markers, segments of DNA that flag locations on a chromosome. Fortunately, the Human Genome Project has already generated a large catalogue of markers spanning all of the X chromosomes. The most useful ones are short, repeated DNA sequences that have slightly different lengths in different persons. To detect the markers, the researchers used the polymerase chain reaction to make several billion copies of specific regions of the chromosome and then separated the different fragments by the method of gel electrophoresis.



PINPOINTING GENES shared by gay brothers (*darker brown*) first involved taking DNA from subjects. Several billion copies of specific regions of the X chromosome were then made using the polymerase chain reaction, and the different fragments were separated by gel electrophoresis. Gay brothers shared a marker, in this hypothetical example CA_{11}, in the Xq28 region at rates far greater than predicted by chance.

The second step in the linkage analysis was to locate suitable families. When scientists study simple traits such as color blindness or sickle cell anemia—which involve a single gene—they tend to analyze large, multigenerational families in which each member clearly either has or does not have the trait. Such an approach was unsuited for studying sexual orientation. First, identifying someone as not homosexual is tricky; the person may be concealing his or her true orientation or may not be aware of it. Because homosexuality was even more stigmatized in the past, multigenerational families are especially problematic in this regard. Moreover, genetic modeling shows that for traits that involve several different genes expressed at varying levels, studying large families can actually decrease the chances of finding a linked gene: too many exceptions are included.

For these reasons, Hamer and his coworkers decided to focus on nuclear families with two gay sons. One advantage of this approach is that individuals who say they are homosexual are unlikely to be mistaken. Furthermore, the approach can detect a single linked gene even if other genes or noninherited factors are required for its expression. For instance, suppose that being gay requires an X chromosome gene together with another gene on an autosome, plus some set of environmental circumstances. Studying gay brothers would give a clear-cut result because both would have the X chromosome gene. In contrast, heterosexual brothers of gay men would sometimes share the X chromosome gene and sometimes not, leading to confusing results.

Genetic analysts now believe that studying siblings is the key to traits that are affected by many elements. Because Hamer and his colleagues were most interested in finding a gene that expresses itself only in men but is transmitted through women, they restricted their search to families with gay men but no gay father–gay son pairs.

Forty such families were recruited. DNA samples were prepared from the gay brothers and, where possible, from their mothers or sisters. The samples were typed for 22 markers that span the X chromosome from the tip of the short arm to the end of the long arm.

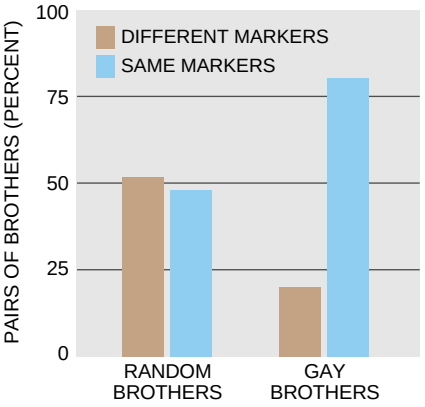
At each marker, a pair of gay brothers was scored as concordant if they inherited identical markers from their mother or as discordant if they inherited different ones. Fifty percent of the markers were expected to be identical by chance. Corrections were also made for the possibility of the mother's having two copies of the same marker.

The results of this study were striking. Over most of the X chromosome the markers were randomly distributed between the gay brothers. But at the tip of the long arm of the X chromosome, in a region known as Xq28, there was a considerable excess of concordant brothers: 33 pairs shared the same marker, whereas only seven pairs did not. Although the sample size was not large, the result was statistically significant: the probability of such a skewed ratio occurring by chance alone is less than one in 200. In a control group of 314 randomly selected pairs of brothers, most of whom can be presumed to be heterosexual, Xq28 markers were randomly distributed.

The most straightforward interpretation of the finding is that chromosomal region Xq28 contains a gene that influences male sexual orientation. The study provides the strongest evidence to date that human sexuality is influenced by heredity because it directly examines the genetic information, the DNA. But as with all initial studies, there are some caveats.

First, the result needs to be replicated: several other claims of finding genes related to personality traits have proved controversial. Second, the gene itself has not yet been isolated. The study locates it within a region of the X chromosome that is about four million base pairs in length. This region represents less than 0.2 percent of the total human genome, but it is still large enough to contain several hundred genes. Finding the needle in this haystack will require either large numbers of families or more complete information about the DNA sequence to identify all possible coding regions. As it happens, Xq28 is extraordinarily rich in genetic loci and will probably be one of the first regions of the human genome to be sequenced in its entirety.

A third caveat is that researchers do not know quantitatively how important a role Xq28 plays in male sexual orientation. Within the population of gay brothers studied, seven of 40 brothers did not share markers. Assuming that 20 siblings should inherit identical markers by chance alone, 36 percent of the gay brothers show no link between homosexuality and Xq28. Perhaps these men inherited different genes or were influenced by nongenetic physiological factors or by the environment. Among all gay men—most of whom do not have gay brothers—the influence of Xq28 is even less clear. Also unknown is the role of Xq28, and other genetic loci, in female sexual orientation.



GENE SHARING in the Xq28 region is significantly greater in gay brothers than in the general population. Of 40 pairs of gay brothers studied, 33 pairs shared the Xq28 region. In a control group of 314 randomly selected pairs of brothers, Xq28 markers were found to be almost equally distributed.

How might a genetic locus at Xq28 affect sexuality? One idea is that the hypothetical gene affects hormone synthesis or metabolism. A candidate for such a gene was the androgen receptor locus, which encodes a protein essential for masculinization of the human brain and is, moreover, located on the X chromosome. To test this idea, Jeremy Nathans, Jennifer P. Macke, Van L. King and Terry R. Brown of Johns Hopkins University teamed up with Bailey of Northwestern and Hamer, Hu and Hu of the NIH. They compared the molecular structure of the androgen receptor gene in 197 homosexual men and 213 predominantly heterosexual men. But no significant variations in the protein coding sequences were found. Also, linkage studies showed no correlation between homosexuality in brothers and inheritance of the androgen receptor locus. Most significant of all, the locus turned out to be at Xq11, far from the Xq28 region. This study excludes the androgen receptor from playing a significant role in male sexual orientation.

A second idea is that the hypothetical gene acts indirectly, through personality or temperament, rather than directly on sexual-object choice. For example, people who are genetically self-reliant might be more likely to acknowledge and act on same-sex feelings than are people who are dependent on the approval of others.

Finally, the intriguing possibility arises that the Xq28 gene product bears directly on the development of sexually dimorphic brain regions such as INAH3. At the simplest level, such an agent could act autonomously, perhaps in the womb, by stimulating the survival of specific neurons in preheterosexual males or by promoting their death in females and prehomosexual men. In a more complex model, the gene product could change the sensitivity of a neuronal circuit in the hypothalamus to stimulation by environmental cues—perhaps in the first few years of life. Here the genes serve to predispose rather than to predetermine. Whether this fanciful notion contains a grain of truth remains to be seen. It is in fact experimentally testable, using current tools of molecular genetics and neurobiology.

Our research has attracted an extraordinary degree of public attention, not so much because of any conceptual breakthrough—the idea that genes and the brain are involved in human behavior is hardly new—but because it touches on a deep conflict in contemporary American society. We believe scientific research can help dispel some of the myths about homosexuality that in the past have clouded the image of lesbians and gay men. We also recognize, however, that increasing knowledge of biology may eventually bring with it the power to infringe on the natural rights of individuals and to impoverish the world of its human diversity. It is important that our society expand discussions of how new scientific information should be used to benefit the human race in its entirety.

FURTHER READING

A DIFFERENCE IN HYPOTHALAMIC STRUCTURE BETWEEN HETEROSEXUAL AND HOMOSEXUAL MEN. Simon LeVay in *Science,* Vol. 253, pages 1034–1037; August 30, 1991.

A LINKAGE BETWEEN DNA MARKERS ON THE X CHROMOSOME AND MALE SEXUAL ORIENTATION. Dean H. Hamer, Stella Hu, Victoria L. Magnuson, Nan Hu and Angela M. L. Pattatucci in *Science,* Vol. 261, pages 321–327; July 16, 1993.

SEQUENCE VARIATION IN THE ANDROGEN RECEPTOR GENE IS NOT A COMMON DETERMINANT OF MALE SEXUAL ORIENTATION. Jennifer P. Macke, Nan Hu, Stella Hu, J. Michael Bailey, Van L. King, Terry R. Brown, Dean H. Hamer and Jeremy Nathans in *American Journal of Human Genetics,* Vol. 53, No. 4, pages 844–852; October 1993.

THE SEXUAL BRAIN. Simon LeVay. MIT Press, 1993.

The Biological Evidence Challenged

Even if genetic and neuroanatomical traits turn out to be correlated with sexual orientation, causation is far from proved

by William Byne

Human-rights activists, religious organizations and all three branches of the U.S. government are debating whether sexual orientation is biological. The discussion has grabbed headlines, but behavioral scientists find it passé. The salient question about biology and sexual orientation is not whether biology is involved but how it is involved. All psychological phenomena are ultimately biological.

Even if the public debate were more precisely framed, it would still be misguided. Most of the links in the chain of reasoning from biology to sexual orientation and social policy do not hold up under scrutiny. At the political level, a requirement that an unconventional trait be inborn or immutable is an inhumane criterion for a society to use in deciding which of its nonconformists it will grant tolerance. Even if homosexuality were entirely a matter of choice, attempts to extirpate it by social and criminal sanctions devalue basic human freedoms and diversity.

Furthermore, the notion that homosexuality must be either inborn and immutable or freely chosen is in turn misinformed. Consider the white-crowned sparrow, a bird that learns its native song during a limited period of development. Most sparrows exposed to a variety of songs, including that of their own species, will learn their species's song, but some do not. After a bird has learned a song, it can neither unlearn that song nor acquire a new one. Although sexual orientation is not a matter of mimicry, it is clear that learned behavior can nonetheless be immutable.

Finally, what evidence exists thus far of innate biological traits underlying homosexuality is flawed. Genetic studies suffer from the inevitable confounding of nature and nurture that plagues attempts to study heritability of psychological traits. Investigations of the brain rely on doubtful hypotheses about differences between the brains of men and women. Biological mechanisms that have been proposed to explain the existence of gay men often cannot be generalized to explain the existence of lesbians (whom studies have largely neglected). And the continuously graded nature of most biological variables is at odds with the paucity of adult bisexuals suggested by most surveys.

To understand how biological factors influence sexual orientation, one must first define orientation. Many researchers, most conspicuously Simon LeVay, treat it as a sexually dimorphic trait: men are generally "programmed" for attraction to women, and women are generally programmed for attraction to men. Male homosexuals, according to this framework, have female programming, and lesbians have male programming. Some researchers suggest that this programming is accomplished by biological agents, perhaps even before birth; others believe it occurs after birth in response to social factors and subjective experiences. As the function of the brain is undoubtedly linked to its structure and physiology, it follows that homosexuals' brains might exhibit some features typical of the opposite sex.

The validity of this "intersex" expectation is questionable. For one, sexual orientation is not dimorphic; it has many forms. The conscious and unconscious motivations associated with sexual attraction are diverse even among people of the same sex and orientation. Myriad experiences (and subjective interpretations of those experiences) could interact to lead different people to the same relative degree of sexual attraction to men or to women. Different people could be sexually attracted to men for different reasons; for example, there is no a priori reason that everyone attracted to men should share some particular brain structure.

Indeed, the notion that gay men are feminized and lesbians masculinized may tell us more about our culture than about the biology of erotic responsiveness. Some Greek myths held that heterosexual rather than homosexual desire had intersex origins: those with predominately same-sex desires were considered the most manly of men and womanly of women. In contrast, those who desired the opposite sex supposedly mixed masculine and feminine in their being. Classical culture celebrated the homosexual exploits of archetypally masculine heroes such as Zeus, Hercules and Julius Caesar. Until a decade ago (when missionaries repudiated the practice), boys among the Sambia of New Guinea would form attachments to men and fellate them; no one considered that behavior a female trait. Indeed, the Sambia believed ingesting semen to be necessary for attaining strength and virility.

But there is a more tangible problem for this intersex assumption: the traits

WILLIAM BYNE studies the ways that biological and social factors interact to influence behavior. He is a research associate at the Albert Einstein College of Medicine of Yeshiva University in New York City, where he investigates the brain structure of humans and other primates, as well as an attending psychiatrist at the New York State Psychiatric Institute. He received his Ph.D. in 1985 from the Neurosciences Training Program at the University of Wisconsin–Madison and his M.D. in 1989 from Einstein. He is also a psychiatrist in private practice.

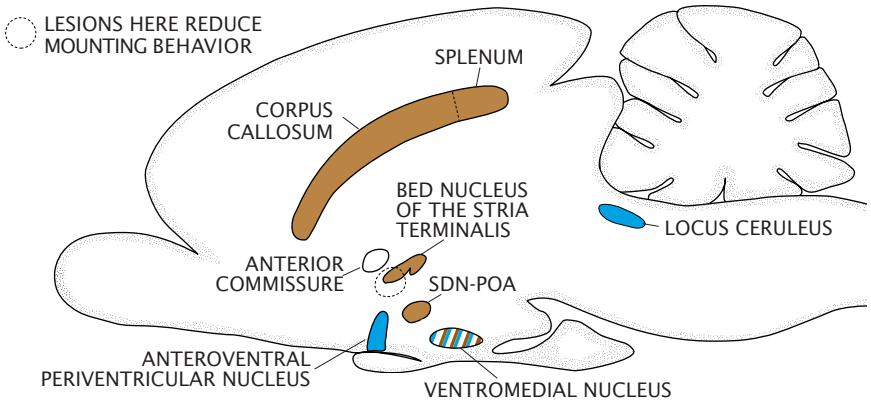
of which homosexuals ostensibly have opposite-sex versions have not been conclusively shown to differ between men and women. Of the many supposed sex differences in the human brain reported over the past century, only one has proved consistently replicable: brain size varies with body size. Thus, men tend to have slightly larger brains than women. This situation contrasts sharply with that for other animals, where many researchers have consistently demonstrated a variety of sex differences.

If brains are indeed wired or otherwise programmed for sexual orientation, what forces are responsible? Three possibilities come into play: The direct model of biological causation asserts that genes, hormones or other factors act directly on the developing brain, probably before birth, to wire it for sexual orientation. Alternatively, the social learning model suggests that biology provides a blank slate of neural circuitry on which experience inscribes orientation. In the indirect model, biological factors do not wire the brain for orientation; instead they predispose individuals toward certain personality traits that influence the relationships and experiences that ultimately shape sexuality.

During past decades, much of the speculation about biology and orientation focused on the role of hormones. Workers once thought an adult's androgen and estrogen levels determined orientation, but this hypothesis withered for lack of support. Researchers have since pursued the notion that hormones wire the brain for sexual orientation during the prenatal period.

According to this hypothesis, high

When they met, they threw their arms round one another and embraced in their longing to grow together again. —Plato, *Symposium*



SEXUALLY DIMORPHIC NUCLEUS of the preoptic area (SDN-POA) in the rat brain is among the regions whose size varies between males and females. Attempts to find an analogous cell group in humans have met with varying success [*see table below*]. Some nuclei have not even been confirmed to exist in other rodents. Regions larger in males are shaded in brown, and those larger in females are shaded in blue.

prenatal androgen levels during the appropriate critical period cause heterosexuality in men and homosexuality in women. Conversely, low fetal androgen levels lead to homosexuality in men and heterosexuality in women. This hypothesis rests largely on the observation that in rodents early exposure to hormones determines the balance between male and female patterns of mating behaviors displayed by adults. Female rodents that were exposed to androgens early in development show more male-typical mounting behavior than do normal adult females. Males deprived of androgens by castration during the same critical period display a female mating posture called lordosis (bending of the back) when they are mounted.

Many researchers consider the castrated male rat that shows lordosis when mounted by another male to be homosexual (as is the female rat that mounts others). Lordosis, however, is little more than a reflex: the male will take the same posture when a handler strokes its back. Furthermore, the male that mounts another male is considered to be heterosexual, as is the female that displays lordosis when mounted by another female. Applying such logic to humans would imply that of two people of the same sex engaged in intercourse only one is homosexual—and which member of the couple it is depends on the positions they assume.

In addition to determining rodent mating patterns, early hormonal exposure determines whether an animal's brain can regulate normal ovarian function. A male rat's brain cannot respond to estrogen by triggering a chain of events, called positive feedback, that culminates in the abrupt increase of luteinizing hormone in the bloodstream, which in turn triggers ovulation. Some researchers reasoned from this fact to the idea that homosexual men (whose brains they allege to be insufficiently masculinized) might have a stronger positive-feedback reaction than do heterosexual men.

Two laboratories reported that this was the case, but carefully designed and executed studies, most notably those of Luis J. G. Gooren of the Free University in Amsterdam, disproved those findings. Furthermore, the feedback mechanism turns out to be irrelevant to human sexual orientation: workers have since found that the positive-feedback mechanism is not sexually dimorphic in primates, including humans. If this mechanism is indistinguishable in men and women, it is illogical to suggest that it should be "feminized" in gay men.

Moreover, a corollary of the expectation that luteinizing hormone responses should be feminized in homosexual men is that they should be "masculinized" in lesbians. If that were true, homosexual women would neither menstruate nor bear children. The overwhelming proportion of lesbians with normal menstrual cycles and the growing number of openly lesbian mothers attest to the fallacy of that idea.

If the prenatal hormonal hypothesis were correct, one might expect that a large proportion of men with medical conditions known to involve prenatal androgen deficiency would be homosexual, as would be women exposed prenatally to excess androgens. That is not the case.

Because androgens are necessary for development of normal external genitals in males, the sex of affected individuals may not be apparent at birth. Males may be born with female-appearing genitals, and females with male-appearing ones. These individuals often require plastic surgery to construct normal-appearing genitals, and the decision to raise them as boys or as girls is sometimes based not on genetic sex but on the possibilities for genital reconstruction.

Research into the sexual orientation of such individuals tends to support the social learning model. Regardless of their genetic sex or the nature of their prenatal hormonal exposure, they usually become heterosexual with respect to the sex their parents raise them as, provided the sex assignment is made unambiguously before the age of three.

Nevertheless, some studies report an increase in homosexual fantasies or behavior among women who were exposed to androgens as fetuses. In accordance with the notion of direct biological effects, these studies are often interpreted as evidence that prenatal androgen exposure wires the brain for sexual attraction to women. The neurobiologist and feminist scholar Ruth H.

RESEARCHERS	BRAIN REGION			
	INAH1	INAH2	INAH3	INAH4
Swaab and Fliers, 1985	Larger in men	Not studied	Not studied	Not studied
Allen et al., 1989	No sex difference	Larger in men than in some women	Larger in men	No sex difference
LeVay, 1991	No sex difference	No sex difference	Larger in heterosexual men than in women or homosexual men	No sex difference

HYPOTHALAMIC NUCLEI are reported to be sites of sexual differences in humans. Yet speculations about the possible contribution of those nuclei to sexual orientation are premature because no differences between men and women have been conclusively demonstrated in these regions.

Bleier has offered an alternative interpretation. Rather than reflecting an effect of masculinizing hormones on the sexual differentiation of the brain, the adaptations of prenatally masculinized women may reflect the impact of having been born with masculinized genitalia or the knowledge that they had been exposed to aberrant levels of sex hormones during development. "Gender must seem a fragile and arbitrary construct," Bleier concluded, "if it depends upon plastic surgery."

Stephen Jay Gould of Harvard University has written of the way that the search for brain differences related to sex and other social categories was for the most part discredited during the past century by anatomists who deluded themselves into believing that their brain measurements justified the social prejudices of their day. The search for sex differences in the human brain was revitalized in the late 1970s, when Roger A. Gorski's team at the University of California at Los Angeles discovered a group of cells in the preoptic part of the rat hypothalamus that was much larger in males than in females. The researchers designated this cell group the sexually dimorphic nucleus of the preoptic area (SDN-POA). The preoptic area has long been implicated in the regulation of sexual behavior.

Like the sex differences in mating behaviors and luteinizing hormone regulatory mechanisms, the difference in the size of the SDN-POA was found to result from differences in early exposure to androgens. Shortly thereafter, Bleier and I, working at the University of Wisconsin at Madison, examined the hypothalamus of several rodent species and found that the SDN-POA is only one part of a sexual dimorphism involving several additional hypothalamic nuclei.

Three laboratories have recently sought sexually dimorphic nuclei in the human hypothalamus. Laura S. Allen, working in Gorski's lab, identified four possible candidates as potential homologues of the rat's SDN-POA and designated them as the interstitial nuclei of the anterior hypothalamus (INAH1–INAH4). Different laboratories that have measured these nuclei, however, have produced conflicting results: Dick F. Swaab's group at the Netherlands Institute for Brain Research in Amsterdam, for example, found INAH1 to be larger in men than in women, whereas Allen found no difference in that nucleus but reported that INAH2 and INAH3 were larger in men. Most recently, LeVay found no sex difference in either INAH1 or INAH2 but corroborated Allen's finding of a larger INAH3 in men. LeVay also reported that INAH3 in homosexual men tends to be small, like that of women. (Neurologist Clifford Saper of Harvard and I are in the process of measuring the interstitial nuclei; at present, we have no definitive results.)

LeVay's study has been widely interpreted as strong evidence that biological factors directly wire the brain for sexual orientation. Several considerations militate against that conclusion. First, his work has not been replicated, and human neuroanatomical studies of this kind have a very poor track record for reproducibility. Indeed, procedures similar to those LeVay used to identify the nuclei have previously led researchers astray.

Manfred Gahr, now at the Max Planck Institute for Animal Physiology in Seewiesen, Germany, used a cell-staining technique similar to LeVay's to observe what appeared to be seasonal variations in the size of a nucleus involved in singing in canaries. Two more specific staining methods, however, revealed that the size of the nucleus did not change. Gahr suggested that the less specific method might have been influenced by seasonal hormonal variations that altered the properties of the cells in the nucleus.

Furthermore, in LeVay's published study, all the brains of gay men came from AIDS patients. His inclusion of a few brains from heterosexual men with AIDS did not adequately address the fact that at the time of death virtually all men with AIDS have decreased testosterone levels as the result of the disease itself or the side effects of particular treatments. To date, LeVay has examined the brain of only one gay man who did not die of AIDS. Thus, it is possible that the effects on the size of INAH3 that he attributed to sexual orientation were actually caused by the hormonal abnormalities associated with AIDS. Work by Deborah Commins and Pauline I. Yahr of the University of California at Irvine supports precisely this hypothesis. The two found that the size of a structure in mongolian gerbils apparently comparable to the SDN-POA varies with the amount of testosterone in the bloodstream.

A final problem with the popular interpretation of LeVay's study is that it is founded on an imprecise analysis of the relevant animal research. LeVay has suggested that INAH3, like the rat's SDN-POA, is situated in a region of the hypothalamus known to participate in the generation of male sexual behavior. Yet studies in a variety of species have consistently shown that the precise hypothalamic region involved in male sexual behavior is not the one occupied by these nuclei. Indeed, Gorski and Gary W. Arendash, now at the University of South Florida, found that destroying the SDN-POA on both sides of a male rat's brain did not impair sexual behavior.

Jefferson C. Slimp performed experiments in Robert W. Goy's laboratory at the Wisconsin Regional Primate Research Center (shortly before I joined that group) that suggested that the

Hormonal Exposure and Mating Behavior in Rats

Mating behavior of rats is affected by exposure to hormones before birth. Males that receive insufficient androgens display stereotypically female postures, whereas females that receive an excess engage in stereotypically male behaviors. Extrapolating such data to sexual orientation, however, is difficult at best.

MALE MOUNTS FEMALE
Male rat is considered heterosexual
Female rat is considered heterosexual

FEMALE MOUNTS FEMALE
Top female rat is considered homosexual
Bottom female rat is considered heterosexual

MALE MOUNTS MALE
Top male rat is considered heterosexual
Bottom male rat is considered homosexual

FEMALE MOUNTS MALE
Female rat would be considered homosexual
Male rat would be considered homosexual

(Not studied in experiments)

precise region involved in sexual behavior in male rhesus monkeys is located above the area comparable to that occupied by INAH3 in humans. Males with lesions in that region mounted females less frequently than they did before being operated on, but their frequency of masturbation did not change. Although some have taken these observations to mean that the lesions selectively decreased heterosexual drive, their conclusion is unwarranted; male monkeys pressed a lever for access to females more often after their operations than before. Unfortunately, these males had no opportunity to interact with other males, and so the study tells us nothing about effects on homosexual as opposed to heterosexual motivation or behavior.

Interstitial hypothalamic nuclei are not the only parts of the brain to have come under scrutiny for links to sexual orientation. Neuroanatomists have also reported potentially interesting differences in regions not directly involved in sexual behaviors. Swaab and his co-worker Michel A. Hofman found that another hypothalamic nucleus, the suprachiasmatic nucleus, is larger in homosexual than in heterosexual men. The size of this structure, however, does not vary with sex, and so even if this finding can be replicated it would not support the assumption that homosexuals have intersexed brains.

Allen of U.C.L.A., meanwhile, has reported that the anterior commissure, a structure that participates in relaying information from one side of the brain to the other, is larger in women than in men. More recently, she concluded that the anterior commissure of gay men is feminized—that is, larger than in heterosexual men. Steven Demeter, Robert W. Doty and James L. Ringo of the University of Rochester, however, found just the opposite: anterior commissures larger in men than in women. Furthermore, even if Allen's findings are correct, the size of the anterior commissure alone would say nothing about an individual's sexual orientation. Although she found a statistically significant difference in the average size of the commissures of gay men and heterosexual men, 27 of the 30 homosexual men in her study had anterior commissures within the same size range as the 30 heterosexual men with whom she compared them.

Some researchers have turned to genetics instead of brain structure in the search for a biological link to sexual orientation. Several recent studies suggest that the brothers of homosexual men are more likely to be homosexual than are men without gay brothers. Of these, only the study by J. Michael Bailey of Northwestern University and Richard C. Pillard of Boston University included both nontwin biological brothers and adopted (unrelated) brothers in addition to identical and fraternal twins.

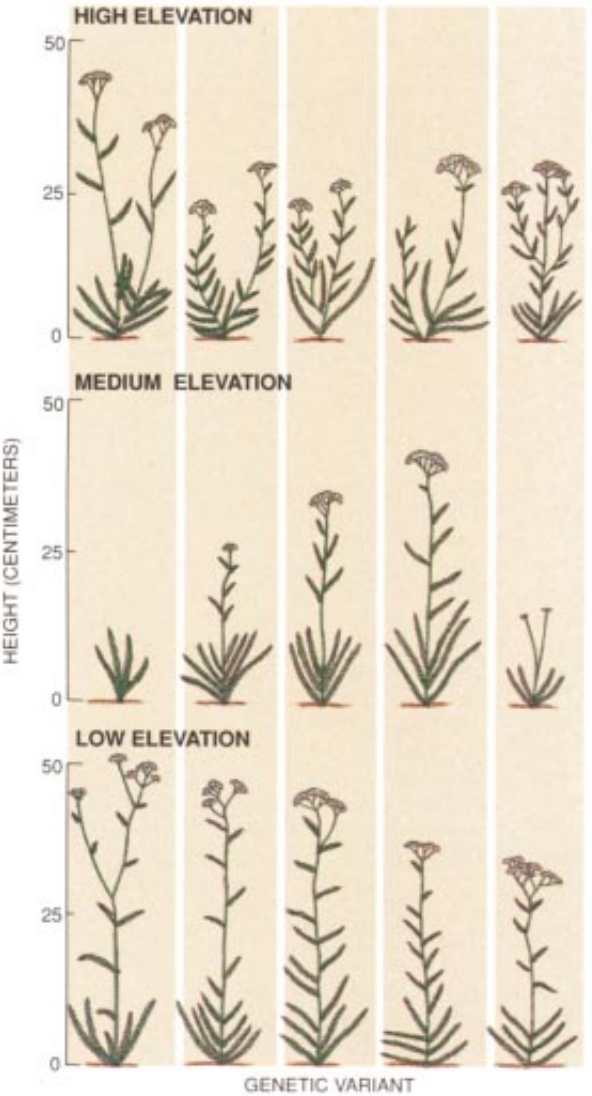
Their investigation yielded paradoxical results: some statistics support a genetic hypothesis, and others refute it. Identical twins were most likely to both be gay; 52 percent were concordant for homosexuality, as compared with 22 percent of fraternal twins. This result would support a genetic interpretation because identical twins share all of their genes, whereas fraternal twins share only half of theirs. Nontwin brothers of homosexuals, however, share the same proportion of genes as fraternal twins; however, only 9 percent of them were concordant for homosexuality. The genetic hypothesis predicts that their rates should be equal.

Moreover, Bailey and Pillard found that the incidence of homosexuality in the adopted brothers of homosexuals (11 percent) was much higher than recent estimates for the rate of homosexuality in the population (1 to 5 percent). In fact, it was equal to the rate for nontwin biological brothers. This study clearly challenges a simple genetic hypothesis and strongly suggests that environment contributes significantly to sexual orientation.

Two of three other recent studies also detected an increased rate of homosexuality among the identical as opposed to fraternal twins of homosexuals. In every case, however, the twins were reared together. Without knowing what developmental experiences contribute to sexual orientation—and whether those experiences are more similar between identical twins than between fraternal twins—the effects of common genes and common environments are difficult to disentangle. Resolving this issue requires studies of twins raised apart.

Indeed, perhaps the major finding of these heritability studies is that despite having all of their genes in common and having prenatal and postnatal environments as close to identical as possible, approximately half of the identical twins were nonetheless discordant for orientation. This finding underscores just how little is known about the origins of sexual orientation.

Dean H. Hamer's team at the



CUTTINGS FROM *Achillea* plants have the same genes, yet they develop in significantly different ways depending on their environment. Furthermore, knowing how the five genetic variants above differ in one environment does not help predict their traits in another one. That plants can display such a complex response to their surroundings makes clear the illogic of expecting direct, easily predictable links between human genes and as diffuse a trait as sexual orientation.

National Institutes of Health has found the most direct evidence that sexual orientation may be influenced by specific genes. The team focused on a small part of the X chromosome known as the Xq28 region, which contains hundreds of genes. Women have two X chromosomes and so two Xq28 regions, but they pass a copy of only one to a son (who has a single X chromosome). The theoretical probability of two sons receiving a copy of the same Xq28 from their mother is thus 50 percent. Hamer found that of his 40 pairs of gay siblings, 33 instead of the expected 20 had received the same Xq28 region from their mother.

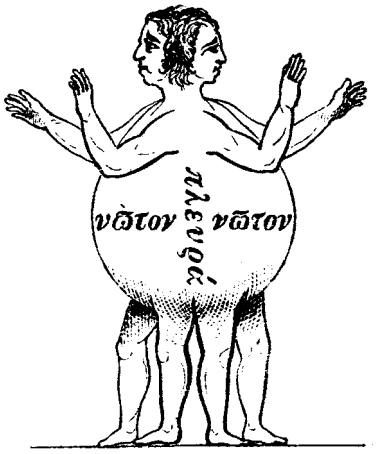
Hamer's finding is often misinterpreted as showing that all 66 men from these 33 pairs shared the same Xq28 sequence. That is quite different from what the study showed: Each member of the 33 concordant pairs shared his Xq28 region only with his brother—not with any of the other 32 pairs. No single, specific Xq28 sequence (a putative "gay gene") was identified in all 66 men.

Unfortunately, Hamer's team did not examine the Xq28 region of its gay subjects' heterosexual brothers to see how many shared the same sequence. Hamer suggests that inclusion of heterosexual siblings would have confounded his analysis because the gene associated with homosexuality might be "incompletely penetrant"—that is to say, heterosexual men could carry the gene without expressing it. In other words, inclusion of heterosexual brothers might have revealed that something other than genes is responsible for sexual orientation.

Finally, Neil J. Risch of Yale University, one of the developers of the statistical techniques that Hamer used, has questioned whether Hamer's results are statistically significant. Risch has argued that until we have more details about the familial clustering of homosexuality, the implications of studies such as Hamer's will remain unclear.

Studies that mark homosexuality as a heritable trait (assuming that they can be replicated) do not say anything about how that heritability might operate. Genes in themselves specify proteins, not behavior or psychological phenomena. Although we know virtually nothing about how complex psychological phenomena are embodied in the brain, it is conceivable that particular DNA sequences might somehow cause the brain to be wired specifically for homosexual orientation. Significantly, however, heritability requires no such mechanism.

Instead particular genes might influence personality traits that could in turn influence the relationships and subjective experiences that contribute to the social learning of sexual orientation. One can imagine many ways in which a temperamental difference could give rise to different orientations in different environments.



MALE-MALE creature from Plato's early race appeared in a 19th-century commentary on the *Symposium*.

The *Achillea* plant serves as a useful metaphor: genetic variations yield disparate phenotypes depending on elevation. The altitude at which a cutting of *Achillea* grows does not have a linear effect on the plant's growth, however, nor is the impact limited to a single attribute. Height, number of leaves and stems, and branching pattern are all affected [*see illustration on opposite page*]. If a plant can display such a complex response to its environment, then what of a far more complex organism that can modify its surroundings at will?

The possible interaction between genes and environment in the development of sexual orientation can be sketched here only in the most oversimplified of ways. For example, many researchers believe aversion to rough-and-tumble play in boys is moderately predictive of homosexual development. (Direct-model theorists argue this aversion is merely the childhood expression of a brain that has been wired for homosexuality.) Meanwhile psychoanalysts have noted that of those gay men who seek therapy, many report having had poor rapport with their fathers. They thus suggest that an impaired father-son relationship leads to homosexuality.

One could combine these observations to speculate that a genetically based aversion to rough-and-tumble play in boys could impair rapport with fathers who demand that they adhere to rigid sex-role stereotypes. Fathers who made no such demands would maintain a rapport with their sons. As a result, the hypothetical gene in question could affect sexual orientation in some cases but not in others. Even such a reductionist example (based on traits that reflect cultural stereotypes rather than biology) shows how neither temperament nor family environment might be decisive. Studies focusing on either one or the other would yield inconclusive results.

These speculations reemphasize how far researchers must go before they understand the factors—both biological and experiential—that contribute to sexual orientation. Even if the size of certain brain structures does turn out to be correlated with sexual orientation, current understanding of the brain is inadequate to explain how such quantitative differences could generate qualitative differences in a psychological phenomenon as complex as sexual orientation. Similarly, confirmation of genetic research purporting to show that homosexuality is heritable makes clear neither what is inherited nor how it influences sexual orientation. For the foreseeable future, then, interpretation of these results will continue to hinge on assumptions of questionable validity.

While attempts to replicate these preliminary findings continue, researchers and the public must resist the temptation to consider them in any but the most tentative fashion. Perhaps more important, we should also be asking ourselves why we as a society are so emotionally invested in this research. Will it—or should it—make any difference in the way we perceive ourselves and others or how we live our lives and allow others to live theirs? Perhaps the answers to the most salient questions in this debate lie not within the biology of human brains but rather in the cultures those brains have created.

FURTHER READING

NOT IN OUR GENES: BIOLOGY, IDEOLOGY, AND HUMAN NATURE. R. C. Lewontin, Steven Rose and Leon J. Kamin. Pantheon Books, 1984.

THEORIES OF ORIGINS OF MALE HOMOSEXUALITY: A CROSS-CULTURAL LOOK. Robert J. Stoller and Gilbert H. Herdt in *Archives of General Psychiatry,* Vol. 42, No. 4, pages 399–404; April 1985.

MYTHS OF GENDER: BIOLOGICAL THEORIES ABOUT WOMEN AND MEN. Anne Fausto-Sterling. BasicBooks, 1992.

HUMAN SEXUAL ORIENTATION: THE BIOLOGIC THEORIES REAPPRAISED. William Byne and Bruce Parsons in *Archives of General Psychiatry,* Vol. 50, No. 3, pages 228–239; March 1993.

Bohm's Alternative to Quantum Mechanics

This theory, ignored for most of the past four decades, challenges the probabilistic, subjectivist picture of reality implicit in the standard formulation of quantum mechanics

by David Z Albert

The study of the behavior of subatomic particles in this century is supposed to have established at least three exceedingly curious facts about the physical world. First, pure chance governs the innermost workings of nature. Second, although material objects always occupy space, situations exist in which they occupy no particular region of space. Third and perhaps most surprising, the fundamental laws that govern the behaviors of "ordinary" physical objects somehow radically fail to apply to objects that happen to be functioning as "measuring instruments" or "observers." That at any rate is what the founders of quantum mechanics decided; that is what has since become the more or less official dogma of theoretical physics; and that is what it says, to this day, in all the standard textbooks on that subject.

But it is now emerging that those conclusions were settled on somewhat too quickly. As a matter of fact, a radically different, fully worked-out theory exists that accounts for all known behaviors of subatomic particles. In this theory, chance plays no role at all, and every material object invariably does occupy some particular region of space. Moreover, this theory takes the form of a single set of basic physical laws that apply in exactly the same way to every physical object that exists.

That theory is principally the work of the late David J. Bohm of Birkbeck College, London. Although his formulation has existed in the scientific literature for more than 40 years, it has until quite recently been mostly ignored. Throughout that period, the thinking about such matters has been dominated by the standard dogma, usually referred to as the Copenhagen interpretation of quantum mechanics because it can more or less be traced back to the Danish physicist Niels Bohr and his circle.

I will begin this article with an outline of the main arguments for the standard dogma. I will then indicate briefly how Bohm's theory manages to get around some of those arguments. Finally, I will say a little about how and where Bohm's theory fits into contemporary speculation about the foundations of quantum mechanics.

Perhaps the simplest way of formulating the arguments for the standard dogma is in the context of certain experiments with electrons. The experiments all involve measurements of two components of what are usually called the spins of electrons. For simplicity's sake, I will refer to them as the horizontal spin and the vertical spin.

It happens to be an empirical fact (as far as we know) that the horizontal spins of electrons can assume only one of two possible values. The same applies for vertical spins. I will call the values of the horizontal spin right and left and those of the vertical spin up and down.

Physicists can measure the horizontal and vertical spins of electrons easily and accurately with currently available technologies. Spin-measuring devices typically work by altering the direction of motion of the electron fed into the device based on the value of its measured spin component. In this way, the value of that spin component can be determined later by a simple measurement of the electron's position. I will refer to these measuring devices as horizontal and vertical boxes [*see illustration on page 60*].

Another empirical fact about electrons is that as a rule there are no correlations between their horizontal spin values and their vertical spin values. For example, of any large collection of right-spinning electrons fed into the entry aperture of a vertical box, precisely half (statistically speaking) will emerge through the "up" aperture and half through the "down" aperture. The same applies for left-spinning electrons fed into the entry aperture of a vertical box and for up- and down-spinning electrons fed into horizontal boxes.

Another experimental truth about electrons, and an extremely important one for our purposes, is that a measurement of the horizontal spin of an electron can disrupt the value of its vertical spin, and vice versa, in what appears to be a completely uncontrollable way. If, for example, one carries out measurements of the vertical spins of any large collection of electrons in-between two measurements of their horizontal spins [*see top illustration on page 61*], what always happens is that the vertical spin measurement changes the horizontal spin values of half of the electrons that pass through it, leaving those of the other half unchanged.

No one has ever been able to design a measurement of vertical spin that avoids such disruptions. Moreover, no

DAVID Z ALBERT has done scientific and philosophical work on various aspects of the foundations of quantum mechanics, with special emphasis on the quantum measurement problem. Recently he has also been thinking about the relation between that problem and the direction of time. In 1981 he received his Ph.D. in theoretical physics from the Rockefeller University. Before taking his current position as professor of philosophy at Columbia University, he served on the faculty of the physics department at the University of South Carolina at Columbia and was a postdoctoral fellow at Tel Aviv University. His book *Quantum Mechanics and Experience* was published last year by Harvard University Press.

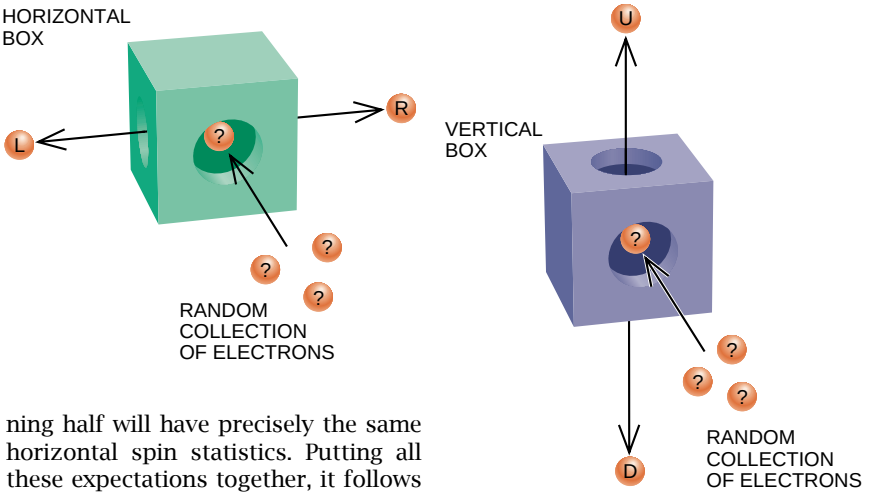
LIVE QUANTUM CAT is one possible outcome of Schrödinger's famous thought experiment, in which a radioactive substance, on emitting a particle, would trigger the release of lethal poison. The problem posed by the experiment is to reconcile the two following facts. The first is that, empirically, cats invariably appear to us either alive or dead. The second is that the linear quantum-mechanical equations of motion seem to predict that cats can be in an almost unimaginably bizarre state in which they are neither alive nor dead. In the standard formulation, sometimes called the Copenhagen interpretation, the approach to this problem involves assigning a unique and indispensable role to observers or measuring devices in bringing about a determinate outcome. Bohm's theory rejects this subjectivist picture: one of the important achievements of this theory is that it solves the problem without recourse to any special role for observers.

one has ever been able to identify any physical properties of the individual electrons in such collections that determine which of them get their horizontal spins changed in the course of having their vertical spins measured and which do not.

What the official doctrine has to say about these matters is that in principle there can be no such thing as a vertical spin measurement that has anything other than precisely that effect on horizontal spin values. Furthermore, the standard doctrine dictates that it is a matter of absolutely pure chance which electrons get their horizontal spins changed by measurements of their vertical spins and which do not; the laws governing those changes simply fail to be deterministic. And these conclusions certainly seem innocent and reasonable given the experimental data.

If measuring one type of spin indeed always uncontrollably disrupts the val-

SPIN-MEASURING BOXES change the direction of motion of electrons based on the particles' spin values. A "horizontal box" sends left-spinning electrons to the left; right-spinning electrons go to the right (*left*). A "vertical box" sends up-spinning electrons up and down-spinning ones down (*right*).



ue of the other, then there can be no way of ascertaining the values of both the horizontal and vertical spins of any particular electron at any particular moment. This phenomenon is an example of the uncertainty principle: certain pairs of measurable physical properties, such as position and momentum or, in our case, horizontal and vertical spin, are said to be incompatible with each other. Measurements of one will always uncontrollably disrupt the other. Many other known examples of incompatible pairs of physical properties exist as well.

So much for indeterminism. But there are still more puzzling features of subatomic particles. Displaying them will require a more complicated experiment. Imagine a box that measures the vertical spins of electrons [*see bottom illustration on opposite page*]. Up-spinning electrons emerge from the box along a route labeled up; down-spinning electrons exit along a route labeled down. We can then arrange a pair of "reflecting walls" to make the two paths cross at some other point. These surfaces can be designed so as not to alter the spin properties of electrons in any way. At the point where the two paths intersect, we place a "black box" that merges the paths back into one, again without altering spin values.

Suppose we feed a large collection of right-spinning electrons, one at a time, into the vertical box. The electrons travel along the paths to the black box. Then as they emerge from the exit of the black box, we measure their horizontal spins. What sorts of results should we expect? Our previous experience informs us that statistically half of such electrons will turn out to be up-spinning and will take the up route through the apparatus. The other half will turn out to be down-spinning and take the down route. Consider the first half. Nothing along the paths between the vertical box and the exit point can have any effect on the vertical spin values of the electrons. Therefore, they will all emerge from the apparatus as up-spinning electrons. In accord with our earlier data, 50 percent of them will turn out to be right-spinning and 50 percent left-spinning. The down-spinning half will have precisely the same horizontal spin statistics. Putting all these expectations together, it follows that for any large set of right-spinning electrons fed into this apparatus, half should be found at the end to be right-spinning and half to be left-spinning.

These conclusions seem absolutely cut-and-dried. But a funny thing happens when you actually try this experiment. Exactly 100 percent of the right-spinning electrons initially fed into this apparatus (one at a time, mind you) come out right-spinning at the end.

It is no exaggeration to describe this result as one of the strangest in modern physics. Perhaps modifying the experiment somewhat will clarify matters. Suppose that we rig up a small, movable, electron-stopping wall that can be slid at will in and out of, say, the up route [*see bottom illustration on opposite page*]. When the wall is out, we have precisely our earlier apparatus. But when the wall is in, all electrons moving along the up route are stopped, and only those moving along the down route get through to the black box.

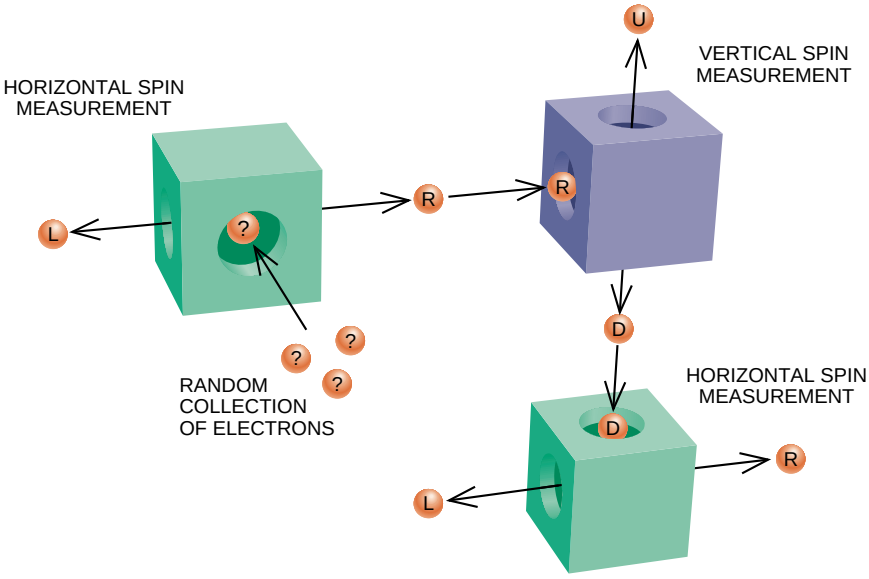
What should we expect to happen when we slide the wall in? To begin with, the overall output of electrons at the exit of the black box ought to drop by 50 percent, because one path is blocked. What about the horizontal spin statistics of the remaining 50 percent? When the wall was out, 100 percent of the right-spinning electrons initially fed in ended up as right-spinning electrons. That is, all those electrons ended up as right-spinning whether they took the up or the down route. Thus, because the presence or absence of the wall on the up route cannot affect electrons on the down route, the remaining 50 percent should all be right-spinning.

As you may have guessed, what actually happens in the experiment runs contrary to our expectations. The output is down by 50 percent, as predicted. But the remaining 50 percent are not all right-spinning. Half are right-spinning, and half are left-spinning. And the same thing happens if we insert a wall in the down path instead. (Readers familiar with quantum mechanics may recognize that this experiment is a logically streamlined version of the famous double-slit experiment.)

How can one understand the discrepancy between the results of these experiments and our expectations about them? Consider an electron that passes through the apparatus when the wall is out. Consider the possibilities as to which route it could have taken. Could it have taken the down route? Apparently not, because electrons taking that route (as the experiment with the wall in reveals) are known to have horizontal spin statistics of 50-50, whereas an electron passing through our apparatus without the wall is known with certainty to be right-spinning at the apparatus exit. Can it have taken the up path, then? No, for the same reasons.

Could it somehow have taken both routes? No: suppose that when a certain electron is passing through this apparatus, we stop the experiment and look to see where it is. It turns out that half the time we find it on the up path and locate nothing at all on the down path, and half the time we find it on the down path and see nothing at all on the up path. Could it have taken neither route? Certainly not. If we wall up both routes, nothing gets through at all.

Something breathtakingly deep, it would seem, has got to give. And indeed, something does—at least according to what has become one of the central tenets of theoretical physics over the past half-century (it is the second of the three official dogmas to which I alluded in the opening paragraph, the one about the indefiniteness of position). That doctrine stipulates that these experiments leave us no alternative but to deny that the very question of which route such an electron takes through



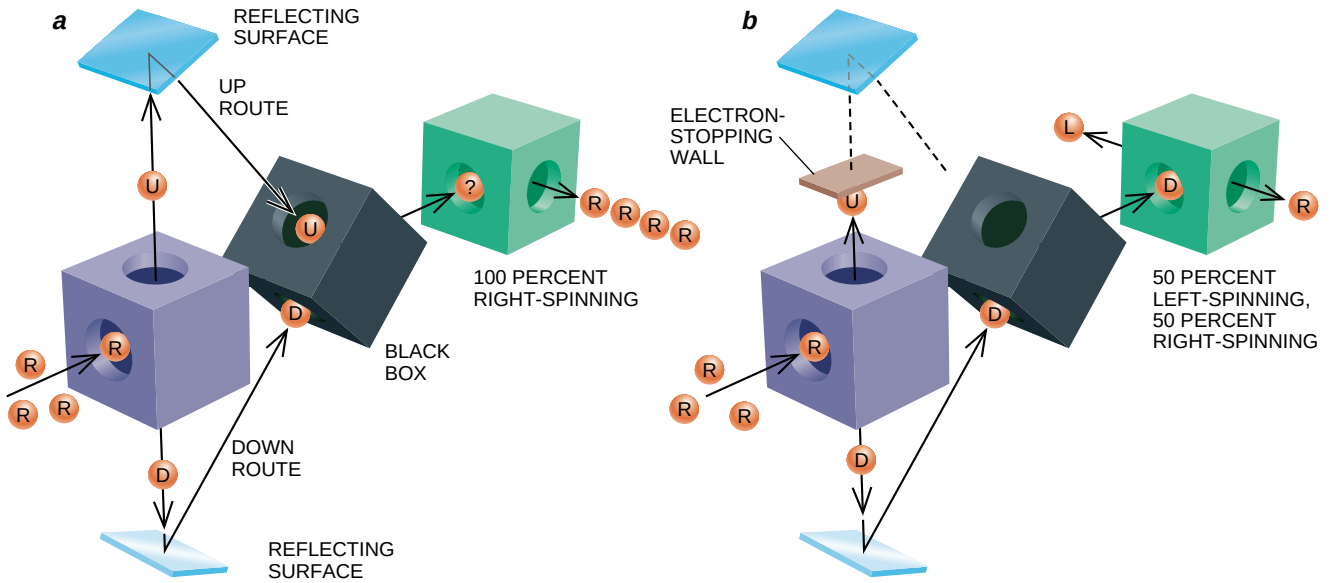
SPIN BEHAVIOR is disrupted in a sequence of three measurements. Electrons are measured one at a time for their horizontal spins (*left*), then for their vertical spins (*right*), and again for their horizontal spins (*bottom*). The vertical box disrupts the spins of half those electrons, so that half emerge from the second horizontal box with right spin, and half emerge with left spin.

such a contraption makes any sense. Asking what route such an electron takes is supposed to be like asking about, say, the political convictions of a tuna sandwich or about the marital status of the number 5. The idea is that asking such questions amounts to a misapplication of language, to what philosophers call a category mistake.

Hence, what physics textbooks typically declare about such electrons is emphatically not that the particles take either the up route or the down route or both routes or neither route through the apparatus. Rather there is simply not any fact about which route they take—not merely no known fact, but no fact at all. They are in what the textbooks term a superposition of taking the up route and the down route through the apparatus.

Notwithstanding the profound violence these ideas do to our intuitive picture of the world, to the very notion of what it is to be material, to be a particle, a compact set of rules has been cooked up that has proved extraordinarily successful at predicting all the observed behaviors of electrons under these circumstances. Moreover, these rules—known of course as quantum mechanics—have proved extraordinarily successful at predicting all the observed behaviors of all physical systems under all circumstances. Indeed, quantum mechanics has functioned for more than 70 years as the framework within which virtually the entirety of theoretical physics is carried out.

The mathematical object with which quantum mechanics represents the states of physical systems is referred to as the wave function. In the simple case of a single-particle system of the kind I have been discussing, the quantum-mechanical wave function takes the form of a straightforward function of position. The wave function of a particle located in some region A, for example, will have the value zero everywhere in space except in A and will have a nonzero value in A. Similarly, the wave function of a particle located in some region B will have the value zero everywhere in space except in B and will have a nonzero value in B. And the wave function of a particle in a superposition of being in region A and in region B—the wave function, for example, of an initially right-spinning electron that has just passed through a vertical box—will have nonzero values



TWO-PATH CONTRAPTION depicts the unusual spin behavior of electrons. In panel *a*, right-spinning electrons fed into a vertical box are sent along the up route or the down route. Reflecting surfaces cause the two paths to converge at a "black box," after which all the electrons are found to be right-spinning. In panel *b*, a wall blocks one of the paths, so that only half the electrons make it to the end. Half these electrons are left-spinning, and half are right-spinning.

Creator of a Brave, New Quantum World

David Joseph Bohm was born in 1917 in Wilkes-Barre, Pa. After studying physics at Pennsylvania State College, he pursued graduate studies at the University of California at Berkeley. There, during World War II, he investigated the scattering of nuclear particles under the supervision of J. Robert Oppenheimer. After receiving his degree from Berkeley, Bohm became an assistant professor at Princeton University in 1946.

It was during those years that Bohm wrote his now classic defense of the Copenhagen interpretation, *Quantum Theory.* At the same time, however, Bohm's doubts about the adequacy of that interpretation were becoming more acute. His own alternative emerged in published form shortly thereafter, in 1952.

By then, Princeton had forced him from its faculty**.** During the McCarthy era, Bohm had been called before the House Un-American Activities Committee in connection with completely unsubstantiated allegations that he and some former colleagues at the radiation laboratory at Berkeley were communist sympathizers. (During World War II, Oppenheimer began turning in to the Federal Bureau of Investigation names of friends and acquaintances who he thought might be communist agents. Bohm apparently was one of the accused.) A passionate believer in liberty, Bohm refused to testify as a matter of principle. As a result, the committee found him to be in contempt of Congress.

The incident proved disastrous to Bohm's professional career in the U.S. Princeton refused to renew his contract and told him not to set foot on the campus. Unable to find employment at any other university, Bohm left the country in 1951 to take a position at the University of São Paulo in Brazil. There he was asked by U.S. officials to give up his passport, effectively stripping him of his American citizenship.

After teaching in Brazil, Bohm went to the Technion in Israel and to Bristol University in England. Although he was later cleared of the contempt charges and was eventually allowed to travel back to the U.S., Bohm settled permanently at Birkbeck College, London, in 1961.

In addition to his interpretation of quantum mechanics, he contributed to mainstream physics, working on plasmas, metals and liquid helium. In 1959 he and his student Yakir Aharonov discovered what is now known as the Aharonov-Bohm effect. They showed that quantum mechanics predicts that the motions of charged particles can be influenced by the presence of magnetic fields even if those particles never enter the regions to which those fields are confined. Subsequent experiments have amply confirmed the effect [see "Quantum Interference and the Aharonov-Bohm Effect," by Yoseph Imry and Richard A. Webb; SCIENTIFIC AMERICAN, April 1989].

Later in life Bohm became interested in broader philosophical questions. He developed a picture of the universe as an interconnectedness of all things, a notion he called "implicate order." He wrote several books on physics, philosophy and the nature of consciousness. He was in the middle of a collaborative effort on another quantum mechanics book when he died of a heart attack in October 1992. Friends and colleagues remember Bohm not only as brilliant and daring but also as extraordinarily honest, gentle and generous.

DAVID J. BOHM (*center*) is escorted to the House Un-American Activities Committee hearing room by Donald Appel, a staff investigator, on May 25, 1949.

in both of those regions and a zero value everywhere else.

And it is a cardinal rule of quantum mechanics (a rule that Bohm's theory will explicitly break) that representing physical objects by a wave function represents them completely. It states that absolutely everything there is to be said about any given physical system at any given instant can be read from its wave function.

What the laws of physics are about—indeed, all that the laws of physics could be about, all that there is for the laws of physics to be about, according to quantum mechanics—is how the wave functions of physical systems evolve in time. The textbook version of quantum mechanics refers to two categories of such laws. And what is particularly peculiar about this formulation is that one of those categories applies when the physical systems in question are not being directly observed, and the other applies when they are.

The laws in the first category are usually written down in the form of linear differential "equations of motion." They are designed to entail, for example, that an initially right-spinning electron fed into a vertical box will emerge from that box in a superposition of traveling along the up route and traveling along the down route. Moreover, all available experimental evidence suggests that those laws govern the evolutions of the wave functions of every single isolated microscopic physical system under all circumstances. So, because microscopic systems are the constituents of everything that exists, there would on the face of it seem to be good reason to suppose that those linear differential equations are the true equations of motion of the entire physical universe.

Yet that conclusion cannot possibly be quite right if wave functions are indeed complete descriptions of physical systems, as quantum mechanics maintains. To begin with, the laws expressed by those equations are completely deterministic, whereas an element of pure chance seems to play a role in the outcomes, for example, of experiments with the spin boxes.

Consider the outcome of a measurement of the position of an electron that is initially in a superposition of being in region A and being in region B. Straightforward calculations reveal that the linear differential equations of motion offer a definite prediction about the end of such a measuring process. Those equations, however, do not predict that the measuring device would either indicate that the electron was found in A or that the electron was found in B (which is what happens when you actually

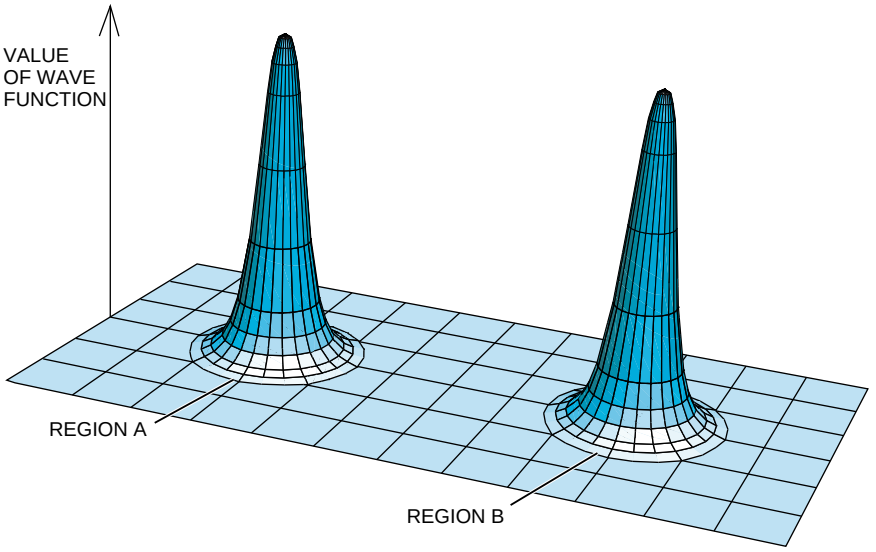
make measurements like that). Rather those equations say the measuring device would with certainty end up in a superposition of indicating that the electron was found in A and indicating that the electron was found in B. To put it slightly differently, those equations predict that the measuring device would end up in a physical state in which there is simply no fact about what it is indicating. It hardly needs mentioning that such superpositions (whatever they are, precisely) do not correctly describe how things end up when you actually make such a measurement.

As a result, according to the official reasoning, the first category of laws needs to be supplemented with a second, which will be explicitly probabilistic. It demands, for example, that if the position of an electron that is initially in a superposition of being in region A and region B were to be measured, there would be a 50 percent chance of finding that electron in region A and a 50 percent chance of finding it in region B. In other words, if the position of the electron were measured, there would be a 50 percent chance that the electron's wave function will be altered in the course of the measurement to one whose value is zero everywhere other than in region A and a 50 percent chance that its wave function will be altered to one whose value is zero everywhere except in region B. (This alteration is sometimes called a "collapse" of the wave function.)

How does one distinguish those conditions in which the first category of laws applies from those in which the second category does? All the founders of quantum mechanics had to say was that it has something to do with the distinction between a "measurement" and an "ordinary physical process," or between what observes and what is observed, or between subject and object.

For some time, many physicists and philosophers have viewed this state of affairs as profoundly unsatisfactory. It has seemed absurd that the best existing formulation of the most fundamental laws of nature should depend on such imprecise and elusive distinctions. The challenge of either eliminating or repairing that imprecision has emerged over the past 30 years as the central task of the foundations of quantum mechanics. It has gone by a number of names: the problem of Schrödinger's cat, for example, or of Wigner's friend, or of quantum state-reduction. I will refer to it by its most common contemporary name: the measurement problem.

One particularly striking solution to the measurement problem was invented by the American-born physicist David J. Bohm. The French physicist Louis de Broglie had devised a related scheme some years earlier, but de Broglie's formulation was much less general and powerful than was Bohm's. More recently, the late physicist John Bell recast Bohm's original theory into a very simple and compelling form.



PARTICLE WAVE FUNCTIONS have nonzero values in those areas of space in which a position measurement might ultimately find the particle. In the standard dogma, the observation "collapses" the wave function onto either region A or region B.

Notwithstanding all the evidence to the contrary presented above, Bohm's theory presumes that particles are the sorts of things that are invariably located in one or another particular place. In addition, Bohm's theory is a great deal clearer than is the Copenhagen interpretation about what the world is made of. In Bohm's account, wave functions are not merely mathematical objects but physical ones, physical things. Bohm treats them somewhat like classical force fields, such as gravitational and magnetic fields. What wave functions do in Bohm's theory (just as classical force fields do) is to in effect push the particles around, to guide them, as it were, along their proper courses.

The laws that govern the evolutions of those wave functions in time are stipulated to be precisely the standard linear differential quantum-mechanical equations of motion—but this time with no exceptions whatever. There are other laws in Bohm's theory as well that dictate how those wave functions push their respective particles around. All those laws are fully deterministic. Therefore, the positions of all the particles in the world at any time, and the world's complete quantum-mechanical wave function at that time, can be calculated with certainty from the positions of all the particles in the world and the world's complete quantum-mechanical wave function at any earlier time.

Any incapacity to carry out those calculations, any uncertainty in the results of those calculations, is necessarily in this theory an epistemic uncertainty. It is a matter of ignorance and not a matter of the operations of any irreducible element of chance in the fundamental laws of the world. Nevertheless, this theory entails that some such ignorance exists for us, as a matter of principle. The laws of motion of Bohm's theory literally force this kind of ignorance on us. And this ignorance turns out to be precisely enough, and of precisely the right kind, to reproduce the familiar statistical predictions of quantum mechanics. That happens by means of a kind of averaging over what one does not know, which is exactly the kind of averaging that goes on in classical statistical mechanics.

The theory describes a real, concrete and deterministic physical process—a process that can be followed out in exact mathematical detail—whereby the act of measurement unavoidably gets in the way of what is being measured. In other words, Bohm's theory entails that this ignorance—although it is merely ignorance of perfectly definite facts about the world—cannot be eliminated without a violation of physical law (without, that is, a violation of one or the other of the two laws of motion described in the box on page 66, from which everything else about Bohm's theory follows).

Bohm's theory can fully account for

the outcomes of the experiments with the two-path contraption—the experiments that seemed to imply that electrons can be in states in which there fails to be any fact about where they are. In the case of an initially right-spinning electron fed into the apparatus, Bohm's theory entails that the electron will take either the up or the down route, period. Which of those two routes it takes will be fully determined by the particle's initial conditions, more specifically by its initial wave function and its initial position. Of course, certain details of those conditions will prove impossible, as a matter of law, to ascertain by measurement. But the crucial point here is that whichever route the electron happens to take, its wave function will split up and take both. It will do so in accordance with the linear differential equations of motion.

So, in the event that the electron in question takes, say, the up route, it will nonetheless be reunited at the black box with the part of its wave function that took the down route. How the down-route part of the wave function ends up pushing the electron around once the two are reunited will depend on the physical conditions encountered along the down path. To put it a bit more suggestively, once the two parts of the electron's wave function are reunited, the part that took the route that the electron itself did not take can "inform" the electron of what things were like along the way. For example, if a wall is inserted in the down route, the down component of the wave function will be missing at the exit of the black box. This absence in itself can constitute decisive information. Thus, the motion that such an electron executes, even if it took the up path through the apparatus, can depend quite dramatically on whether or not such a wall was inserted.

Moreover, Bohm's theory entails that the "empty" part of the wave function—the part that travels along the route the electron itself does not take—is completely undetectable. One of the consequences of the second equation in the box below is that only the part of any given particle's wave function that is currently occupied by the particle itself can have any effect on the motions of other particles. So the empty part of the wave function—notwithstanding the fact that it is really, physically, there—is completely incapable of leaving any observable trace of itself on detectors or anything else.

Hence, Bohm's theory accounts for all the unfathomable-looking behaviors of electrons discussed earlier every bit as well as the standard interpretation does. Moreover, and this point is important, it is free of any of the metaphysical perplexities associated with quantum-mechanical superposition.

As to the measurement problem, it can be persuasively argued that Bohm's theory can suffer from nothing of the kind. Bohm's theory holds that the linear differential equations of motion truly and completely describe the evolution of the wave function of the entire universe—measuring devices, observers and all. But it also stipulates that there are invariably definite matters of fact about the positions of particles and, consequently, about the positions of pointers on measuring devices and about the positions of ink molecules in laboratory notebooks and about the positions of ions in the brains of human observers and thus, presumably, about the outcomes of experiments.

Despite all the rather spectacular advantages of Bohm's theory, an almost universal refusal even to consider it, and an almost universal allegiance to the standard formulation of quantum mechanics, has persisted in physics, astonishingly, throughout most of the past 40 years. Many researchers have perennially dismissed Bohm's theory on the grounds that it granted a privileged mathematical role to the positions of particles. The complaint was that this assignment would ruin the symmetry between position and momentum, which had been implicit in the mathematics of quantum theory up until then—as if ruining that symmetry somehow amounted to a more serious affront to scientific reason than the radical undermining, in the Copenhagen formulation, of the very idea of an objective physical reality. Others dismissed Bohm's theory because it made no empirical predictions (no obvious ones, that is) that differed from those of the standard interpretation—as if the fact that those two formulations had much in common on that score some-

The Exact Mathematical Formulation of Bohm's Theory

Bohm's theory in its entirety consists of three elements. The first is a deterministic law (namely, Schrödinger's equation) that describes how the wave functions of physical systems evolve over time. It is:

$$i\frac{h}{2\pi}\frac{\partial}{\partial t}\psi(x_1 \ldots x_{3N},t) = H\psi(x_1 \ldots x_{3N},t)$$

where i is the imaginary number $\sqrt{-1}$, h is Planck's constant, ψ is the wave function, H is a mathematical object called the Hamiltonian operator, N is the number of particles in the system, $x_1 \ldots x_{3N}$ represent the spatial coordinates of those particles, and t is the time. Loosely speaking, the Hamiltonian operator describes the energy in the system.

The second element is a deterministic law of the motions of the particles:

$$\frac{dX_i(t)}{dt} = \frac{j_i(X_1 \ldots X_{3N},t)}{|\psi(X_1 \ldots X_{3N},t)|^2}$$

where $X_1 \ldots X_{3N}$ represent the actual coordinate values of the particles, $dX_i(t)/dt$ is the rate of change of X_i at time t, and j_i represents the components of the standard quantum-mechanical probability current. The subscript i ranges from 1 to $3N$.

The third element is a statistical rule analogous to one used in classical statistical mechanics. It stipulates precisely how one goes about "averaging over" one's inevitable ignorance of the exact states of physical systems. It runs as follows. Assume one is given the wave function of a certain system but no information about the positions of its particles. To calculate the motions of those particles in the future, what one ought to suppose is that the probability that those particles are currently located at some position $(X_1 \ldots X_{3N})$ is equal to $|\psi(X_1 \ldots X_{3N})|^2$. If information about the positions of the particles becomes available (as during a measurement), the rule indicates that that information ought to be used to "update" the probabilities through a mathematical procedure called straightforward conditionalization.

That is literally all there is to Bohm's theory. Whatever else we know about it—everything presented in this article, for example—derives strictly from these three elements.

how transparently favored one of them over the other. Still others cited "proofs" in the literature—the most famous of which was devised by the American mathematician John von Neumann, and all of which were wrong—that no deterministic replacement for quantum mechanics of the kind that Bohm had already accomplished was even possible.

Fortunately, those discussions are mostly in the past now. Although the Copenhagen interpretation probably remains the guiding dogma of the average working physicist, serious students of the foundations of quantum mechanics rarely defend the standard formulation anymore. A number of interesting new proposals now exist for solving the measurement problem. (There are, for example, attempts at resuscitating in a more precise language the idea of a collapse of the wave function, which I mentioned earlier.) It is against those, against other proposals yet to be invented and, of course, against the experimental facts that Bohm's theory will ultimately have to be judged. The jury on all that is still very much out.

QUANTUM APOSTATE David J. Bohm, shown here three years before his death in 1992, formulated his interpretation of quantum mechanics in the 1950s.

Bohm's theory is the only serious proposal around just now that is fully deterministic. It is also the only one that denies there are any such things as superpositions, even for microscopic systems. But it is certainly not free of transgressions against what one might call common physical sense. Perhaps the most flagrant of those transgressions is nonlocality. The theory allows for the possibility that something that occurs in region A can have a physical effect in region B, instantaneously, no matter how far apart regions A and B may happen to be. The influence is also completely independent of the conditions existing in the space between A and B [see "Faster than Light?" by Raymond Y. Chiao, Paul G. Kwiat and Aephraim M. Steinberg; SCIENTIFIC AMERICAN, August 1993].

But nonlocality may be something we need to learn to live with, something that may simply turn out to be a fact of nature. The standard formulation of quantum mechanics is also nonlocal and so are most of the recently proposed solutions to the measurement problem. Indeed, according to a famous argument of Bell's, any theory that can reproduce those statistical predictions of quantum mechanics already known to be correct and that satisfies a few extremely reasonable assumptions about the physical nature of the world must necessarily be nonlocal. The only schemes that have been imagined for denying those assumptions and so avoid nonlocality are the "many worlds" and "many minds" interpretations of quantum mechanics. They suggest that in some sense all possible experimental outcomes, and not simply one or another of those outcomes, actually occur. And they are (maybe) too bizarre to be taken seriously.

Workers have raised various other concerns as well. What is the exact philosophical status of the probabilities in Bohm's theory? Does guaranteeing that every particle in the world invariably has a determinate position really amount to ensuring that every imaginable measurement has a determinate outcome and that everything that we intuitively take to be determinate is really determinate? Those questions continue to be the subject of active debate and investigation.

Finally, and most important, I must stress that all of what has been said in this article applies, at least for the moment, only to nonrelativistic physical systems. That is, it pertains just to systems whose energies are not very high, that are not moving close to the speed of light and that are not exposed to intense gravitational fields. The development of a Bohmian replacement for relativistic quantum field theory is still under way, and the ultimate success of that enterprise is by no means guaranteed. If such a replacement were somehow found to be impossible, then Bohm's theory would have to be abandoned, and that would be that.

But as it happens, most other proposals for solving the measurement problem are in a similar predicament. The exceptions, once again, are the many-worlds and many-minds interpretations, whose relativistic generalizations are quite straightforward but whose metaphysical claims are difficult to believe. Much of the future course of the foundations of quantum mechanics will hinge on how attempts at relativization come out.

In the meantime, the news is that a great deal more than has previously been acknowledged about the foundations of our picture of the physical world turns out to be radically unsettled. In particular, the possibilities that the laws of physics are fully deterministic and that what they describe are the motions of particles (or some analogue of those motions in relativistic quantum field theory) are both, finally and definitively, back on the table.

FURTHER READING

A SUGGESTED INTERPRETATION OF THE QUANTUM THEORY IN TERMS OF "HIDDEN" VARIABLES, I AND II. David Bohm in *Quantum Theory and Measurement.* Edited by J. A. Wheeler and W. H. Zurek. Princeton University Press, 1983.

ON THE IMPOSSIBLE PILOT WAVE. In *Speakable and Unspeakable in Quantum Mechanics,* by John S. Bell. Cambridge University Press, 1987.

BOHM'S THEORY. In *Quantum Mechanics and Experience,* by David Z Albert. Harvard University Press, 1992.

QUANTUM EQUILIBRIUM AND THE ORIGIN OF ABSOLUTE UNCERTAINTY. Detlef Dürr, Sheldon Goldstein and Nino Zanghi in *Journal of Statistical Physics,* Vol. 67, Nos. 5/6, pages 843–908; June 1992.

How Interferons Fight Disease

They are not the cure-alls researchers once hoped they would be, but they are providing therapy for a variety of infectious illnesses and for some cancers

by Howard M. Johnson, Fuller W. Bazer, Brian E. Szente and Michael A. Jarpe

In 1957 two investigators at the National Institute for Medical Research in London made a landmark discovery. They knew that when one kind of virus colonized cells in animals or in a culture dish, the invasion interfered with the ability of other, unrelated viruses to establish infections at the same time. One would expect an intact immune system in the body to repel subsequent attacks by the original virus. But how would cells in the body or in culture acquire resistance immediately to unrelated viruses?

The British researchers, Alick Isaacs and Jean Lindenmann, showed that a substance secreted by the infected cells was responsible. Logically enough, they named it interferon. Isaacs and Lindenmann also demonstrated that the agent, a protein, did not interact with viruses directly; rather it induced diseased cells—and their neighbors—to make still other proteins that could prevent invading viruses from replicating.

Since then, countless investigators have devoted study to interferon. Their efforts have led to a series of surprising findings, such as the discovery that interferon is not a single molecule. It comes in various forms, all of which to some degree are able to interfere with viral infection. These molecules also turn out to belong to the ever growing superfamily of cytokines (small proteins that carry signals locally from one cell to another). Further, as a group, they are more versatile and important to health than anyone could have anticipated in the 1950s.

Indeed, interferons modulate the activity of virtually every component of the immune system. In so doing, they enhance the body's ability to quell attacks by most disease-causing agents, including bacteria and parasites as well as viruses. Interferons can promote or impede the differentiation (specialization) of certain cells. They can inhibit cell division, which may partly explain why they can often impair the proliferation of cancer cells. And, recently, one form has been proved essential to the maintenance of early pregnancy in several species of animals.

Naturally, with all these impressive properties, interferons have come under close scrutiny as potential therapeutic agents. Early on, it was hoped they would serve as magic bullets for a broad range of viruses and cancers, curing these disorders without harming healthy cells. Those hopes, it now seems, were unrealistic. Nevertheless, the proteins have been approved by the Food and Drug Administration for treating seven diseases, among them chronic hepatitis, genital warts and Kaposi's sarcoma. Last July one variant became, with much fanfare, the only approved drug therapy for a form of multiple sclerosis. And the proteins are under continuing study for treating many other conditions, notably such cancers as non-Hodgkin's lymphoma and malignant melanoma.

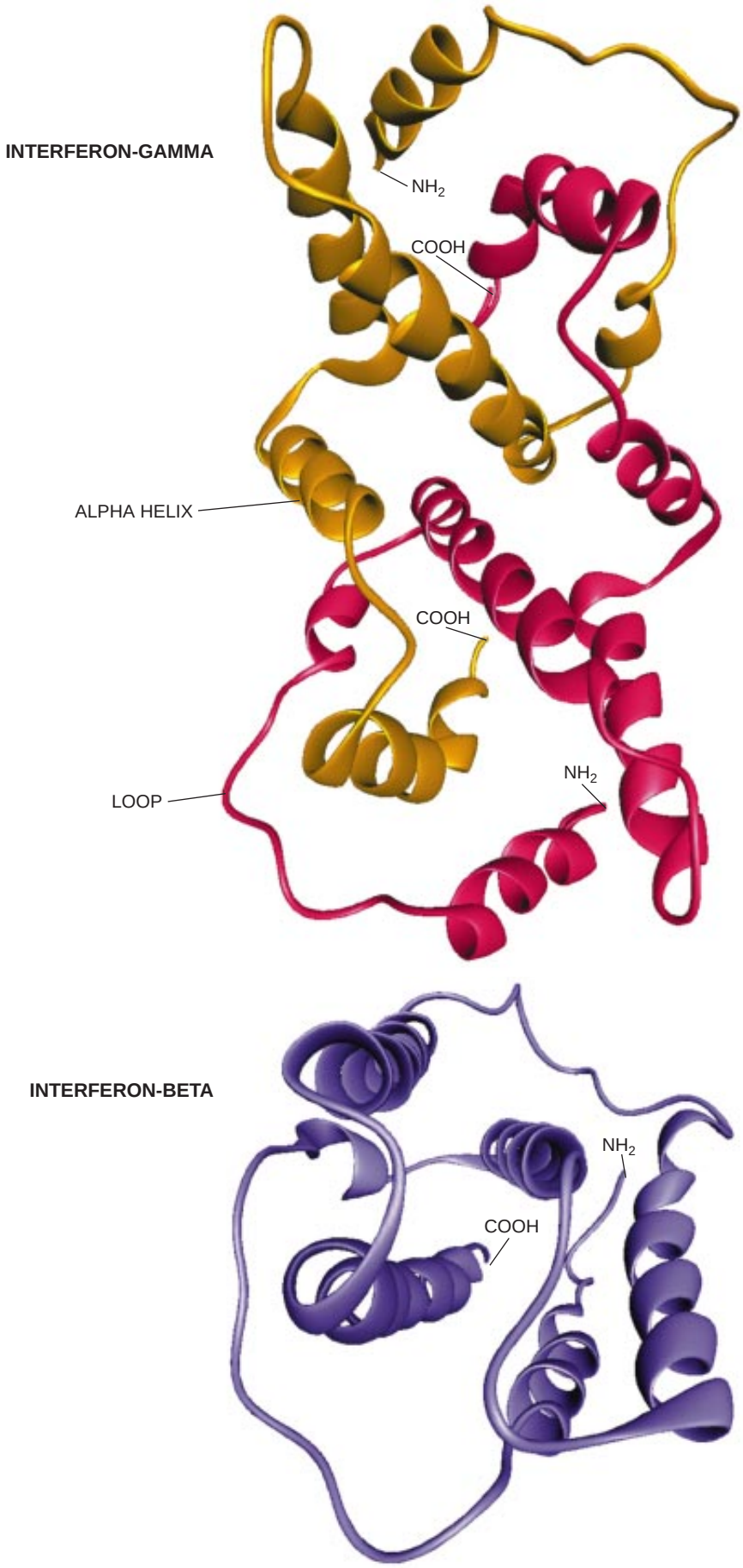
Equally significant, scientists are beginning to elucidate the three-dimensional structure of the interferons and to identify the central steps by which they and their relatives work their myriad effects. Such insights should ultimately lead to production of more potent, less toxic drugs. For cell biologists, the research offers an added bonus: it is clarifying the molecular interactions through which other kinds of cytokines regulate the activities of cells.

Most interferons are regarded as belonging to one of three classes—alpha, beta or gamma. Initially the molecules were grouped mainly on the basis of their reactivity to particular antibodies. Today classification is based on the amino acid sequence of the proteins. The alpha family is the largest and most variable group, comprising more than 20 members. So far only one beta and one gamma protein have been identified conclusively. Workers have discovered two additional classes—omega and tau—as well. They closely resemble the alpha variants but are slightly larger.

Interferon-alpha and interferon-beta have more in common structurally and functionally with one another than with gamma, and so they (along with the alphalike omega and tau molecules) are often classified as type I interferons. Interferon-gamma stands by itself as the sole type II molecule. The type I interferons are, as a rule, better able to induce viral resistance in cells. Interferon-

HOWARD M. JOHNSON, FULLER W. BAZER, BRIAN E. SZENTE and MICHAEL A. JARPE all contributed to the current understanding of interferons while working at the University of Florida at Gainesville. Johnson is graduate research professor in the department of microbiology and cell science. He is also a consultant to Pepgen Corporation, a biotechnology investment firm in Huntington Beach, Calif. Bazer, formerly graduate research professor in the department of animal science at Gainesville, now holds the O. D. Butler Chair in the department of animal science at Texas A&M University. Szente is a graduate student in Johnson's group. Jarpe, who was a postdoctoral associate in Johnson's laboratory from 1990 to 1992, is staff scientist at Cambridge Neurosciences in Massachusetts.

STRUCTURES OF TWO INTERFERONS—human gamma (*top*) and mouse beta (*bottom*)—have been elucidated in the past few years. The ribbons represent the backbones of the constituent proteins. Interferon-gamma consists of two identical molecules (*distinguished by color*) that are intertwined; interferon-beta consists of a single, distinct protein. Structural analyses are helping to clarify the steps by which interferons exert effects on cells. The gamma structure was determined by Charles E. Bugg and his colleagues at the University of Alabama at Birmingham; beta by Yukio Mitsui and his colleagues at Nagaoka University of Technology in Japan.



gamma is more prominent in orchestrating the defensive maneuvers of the immune system.

Type I and type II interferons differ in another way. If infected by a virus, essentially all cells in the body can produce one or another type I interferon (most typically alpha). But only two cell types—*T* lymphocytes and natural killer cells (both members of the immune system)—release interferon-gamma. And they do so not when they are infected themselves but when they are alerted to the presence of viruses, bacteria or parasites in other cells or to the emergence of cancer. (Superantigens, which overexcite the immune system, and mitogens, chemicals that induce cells to divide, can also cause *T* cells and natural killer cells to secrete interferon-gamma.) Finally, the two types of interferons bind to different receptors on cells.

For decades, investigators interested in interferons have struggled to answer a basic question: How does the binding of these proteins to their respective receptors enable cells to combat viruses and execute other tasks? It has long been known that interferons, in common with other proteins that carry directives from cell to cell, exert their effects by activating so-called signal transduction pathways. A signal transduction pathway consists of a cascade of reactions that begins when a signaling molecule binds to a receptor that spans the cell membrane. The binding causes the part of the receptor that lies within the cell to convey commands to other molecules in the cell. Then these molecules, often acting through a string of intracellular intermediaries, relay the commands to the agents that ultimately carry out the orders.

Interferons activate pathways that cause cells to transcribe, or copy, certain genes into molecules of messenger RNA. The RNA transcripts, in turn, are translated into proteins that interfere with viral replication or produce other effects. But what, precisely, goes on between receptor binding and gene transcription? Many of the intervening events have recently been revealed. James E. Darnell, Jr., and his colleagues at the Rockefeller University and, separately, a team led by Michael David and Andrew C. Larner of the FDA are largely responsible for the discoveries.

The signaling pathways of the interferons turn out to be startling in their directness. This property contrasts strongly with the complexity of the pathways activated by many other molecules, such as the hormone epinephrine. Binding of these other molecules to their receptors leads, often

circuitously, to elevations in the intracellular levels of substances, called second messengers, that can migrate through the cytoplasm. Typical second messengers include ringed, or cyclic, nucleotides and certain by-products of the breakdown of phospholipids in the cell membrane. Rising levels of these substances can set in motion potentially lengthy cascades of enzymatic and other molecular interactions that finally end in a change in the cell's behavior.

The signal transduction pathways of the interferons do not depend on second messengers. In a further departure from the standard pathways, those initiated by interferons involve activation of a newly identified class of enzymes known as Janus kinases. These proteins, discovered in 1990, are a subgroup of enzymes (tyrosine kinases) that add phosphate groups to tyrosine amino acids in proteins. The subgroup is named after the Roman god Janus, who had two faces, because its members carry two sites potentially able to add phosphate groups to proteins.

The pathways activated by the type I interferons differ slightly from those activated by type II interferon. All type I interferons are thought to recognize the same receptor (or similar receptors). Darnell's group showed in 1992 that binding of type I interferons to their receptor results in activation of an enzyme called tyrosine kinase 2 (Tyk2) that is apparently attached to the intracellular domain of the receptor. More recently, Ian M. Kerr of the Imperial Cancer Research Fund Laboratories in London demonstrated that binding to the receptor also leads to activation of Janus kinase 1 (JAK1).

These kinases do not trigger complicated cascades of reactions. Instead, separately or cooperatively, they apparently phosphorylate three proteins named Stat113, 91 and 84. The term "Stat" refers to the fact that, as will be seen, the proteins are *s*ignal *t*ransducers as well as *a*ctivators of *t*ranscription (transcription factors). The numerical part of the designation refers to molecular weight, measured in kilodaltons. (Stat84 turns out to be little more than a truncated version of Stat91.)

Phosphorylation of tyrosine induces these factors to associate with one another and with a 48-kilodalton protein. Together these four proteins promptly form a complex on certain genes in the nucleus. Formation of the complex, in turn, directs the gene transcription machinery of the cell to make messenger RNA copies of the bound genes. The genes that become bound and activated are ones that contain a specific sequence of nucleotides in their promoter region. (The promoter of a gene acts as a kind of switch that controls the initiation of transcription.) For type I interferons, the target sequence is known as the interferon-stimulated response element.

Stimulation of gene transcription by interferon-gamma apparently involves even fewer transcription factors. Binding of interferon-gamma to its receptor results in activation of two Janus kinases: JAK1 (again) and JAK2, which seems to be bound to the receptor. These enzymes then phosphorylate the same Stat91 protein that participates in the type I pathway. So decorated, Stat91 combines with another protein and probably with a duplicate of itself. This complex attaches to genes whose promoters contain a response element called the interferon-gamma activation site. Transcription ensues.

As researchers identified the components of these novel signaling pathways, they began to explore the possibility that similar short strings of molecular interactions could mediate the effects of other cytokines. Sure enough, a recent spate of studies by a few groups, among them Darnell's, indicates that binding of various cytokines to their respective receptors triggers activation of Janus kinases and the phosphorylation of molecules related to the Stat proteins. The phosphorylated transcription factors bind to response elements on genes that resemble the interferon-stimulated response element and the interferon-gamma activation site. Hence, exploration of the interferons seems to have uncovered the basic elements of a heretofore unknown type of pathway—one through which cells respond particularly rapidly and directly to many of the signals that impinge on them from the outside.

Of course, there are gaps in understanding of the signaling pathways activated by the interferons. For instance, one interferon-alpha molecule can induce transcription of a set of genes slightly different from the set transcribed in response to another form of alpha. The emerging model does not explain how similar kinds of interferons would elicit divergent effects. Investigators would also like to know more about the process by which interferons bind to their receptors and about missing steps in the model of signal transmission: the interactions that lead from receptor binding to the activation of tyrosine kinases.

Examination of the three-dimensional structure of the interferons is help-

Interferons at a Glance

Interferons are divided into two main types, based on differences in their structure and other properties. Both types can interfere with viral replication in cells and can modulate the activities of the immune system. Nevertheless, type I interferons are more effective at bolstering the ability of cells to resist viral infections, and type II interferon is more important to the proper functioning of the immune system.

	TYPE I	TYPE II
MAIN TYPES	Alpha and beta	Gamma
OTHER TYPES	Tau and omega	None
STRUCTURE	Single chain of amino acids	Dimer of two identical proteins, unrelated in amino acid sequence to type I interferons
MAIN PRODUCERS	Almost any cell infected by a virus will make interferon-alpha; fibroblasts are the main producers of interferon-beta	*T* lymphocytes and natural killer cells
MAIN EFFECTS	Induce infected cells to produce proteins that inhibit the proliferation of viruses and cells	Promotes activity by components of the immune system that eradicate tumors and infections in cells

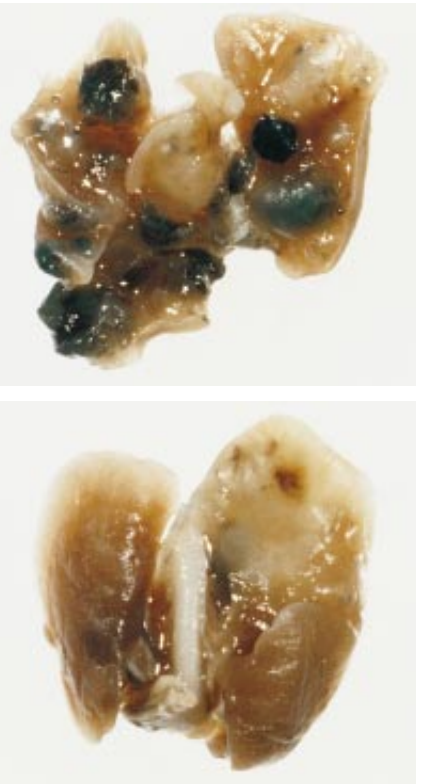
ing to fill in some of the holes. So far the conformation of two of the molecules has been determined. In 1990 Yukio Mitsui and his colleagues at Nagaoka University of Technology in Japan worked out the structure of interferon-beta produced by mice. Then, in 1991, Charles E. Bugg and his colleagues at the University of Alabama at Birmingham determined the shape of the human gamma molecule.

Interferon-beta (like its other type I relatives) consists of one chain of amino acids. In contrast, the active form of interferon-gamma is a dimer; it is composed of two, intertwined copies (monomers) of a single protein, one copy of which is oriented upside down relative to the other. To be more precise, the monomers (and all proteins) have an amino group (NH_2) at one end and a carboxyl (COOH) group at the opposite end. In the dimer, the amino terminal (N terminal) domain of each monomer interacts with the carboxyl terminal (C terminal) region of the other monomer.

Yet interferon-beta and interferon-gamma display similarities as well. The backbones of both the beta and gamma monomers twist into coils, known as alpha helixes, that are linked by simple loops. More important, certain parts of the folded beta molecule and the gamma dimer look much alike. Interferon-beta folds so that a helix in the C terminal region rests in a kind of cup formed by two helixes and their connecting loops in the N terminal region. When Bugg and his colleagues worked out the gamma structure, they quickly realized that a very similar arrangement, or structural motif, occurs in interferon-gamma. In interferon-gamma, however, the C terminal region of one monomer nestles into the N terminal region of the other.

Computer modeling conducted by many laboratories, including that of one of us (Johnson), soon indicated that members of the alpha class possess the same structural motif. The ubiquity of the motif implies that the arrangement is required for the functioning of the interferons. It might, for instance, enable them to combine with their receptors.

Johnson's group at the University of Florida has produced evidence consistent with this hypothesis. The workers found that synthetic peptides created to mimic the N and C terminal regions of mouse interferon-gamma do indeed bind to free copies of the gamma receptor. Maria van Volkenburg, one of Johnson's graduate students, has also found evidence that the docking site for at least the N terminal region is located near the center of the extracellular part of the receptor. Gianni Garotta and his colleagues at Hoffmann–La Roche Pharmaceuticals in Basel independently suggested the same receptor region as the binding site. They also discovered that the gamma dimer—with its two pairs of nested N and C terminal domains—probably associates with two receptors at once.



MELANOMA TUMORS (*black areas*) induced to form in the lungs of mice did not shrink in an untreated animal (*top*) but regressed markedly (*bottom*) in an animal treated with interferon-tau. Because this interferon seems to be less toxic than the others, it might prove to be particularly valuable in medicine.

How exactly does binding to the extracellular part of interferon receptors lead to activation of tyrosine kinases inside the cell? Taken together, several findings hint at an answer—and yield yet another surprise. Two of us (Johnson and Szente) have recently established that when receptor molecules are exposed to interferon-gamma, the C terminal domain of interferon-gamma can bind to a part of the receptor that normally lies in the cytoplasm. It might seem that such binding could not occur in intact cells (and, thus, in the body), because the cell membrane would shield the cytoplasmic part of the receptor from contact with anything located outside the cell. Yet such binding suddenly seems feasible when other discoveries are taken into account.

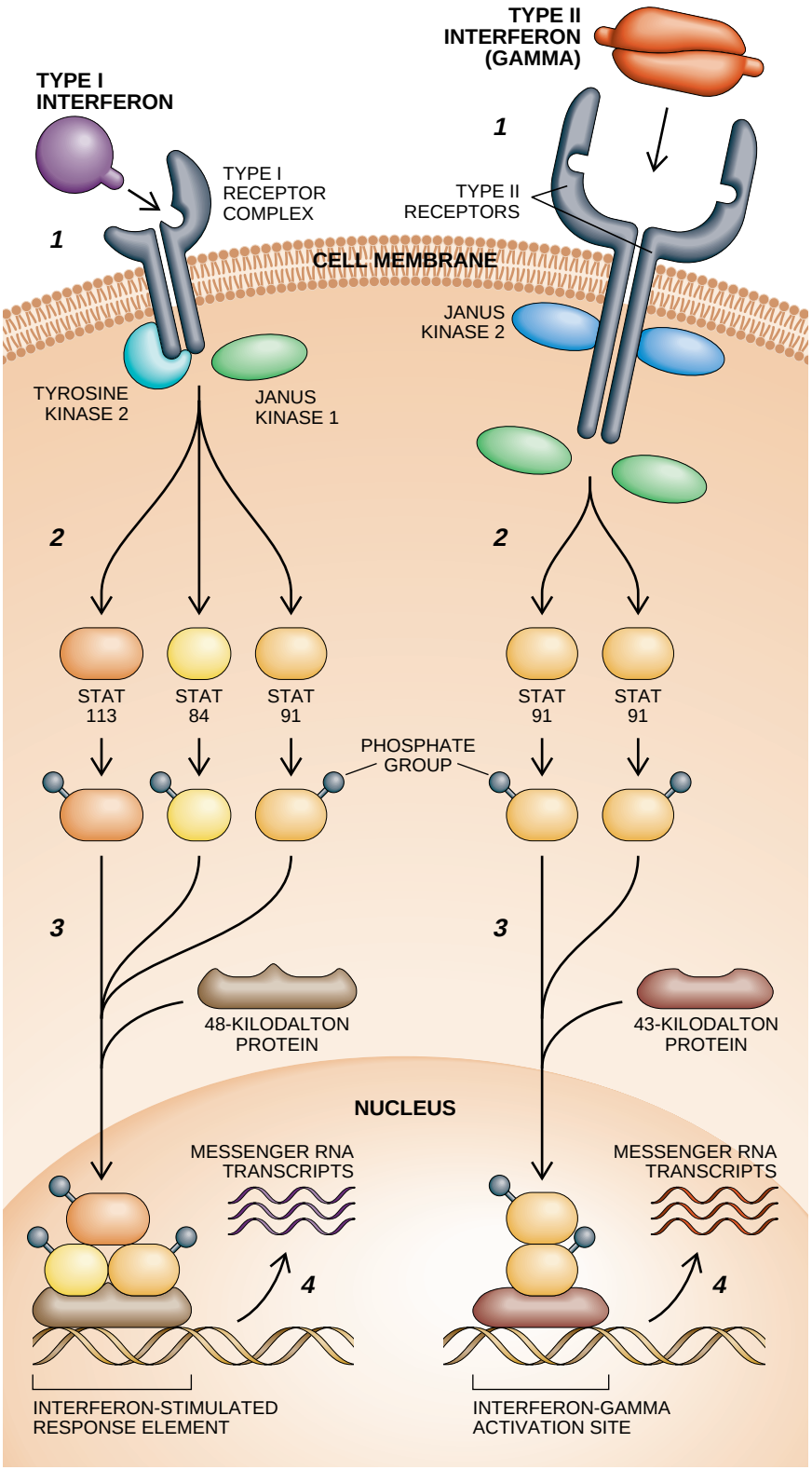
The part of the interferon receptor bound by the C terminal region of the gamma molecule has been shown to resemble an intracellular area on the receptors for two other cytokines—erythropoietin and growth hormone. When those receptors are inactive, they are bound at that precise site by the kinase JAK2.

These findings all fit together nicely if one keeps in mind that after binding to receptors many intercellular signaling proteins, including the interferons, are drawn into cells along with their receptors. Common wisdom has it that the signaling proteins and their receptors are soon degraded or recycled without interacting further. In the case of interferon-gamma, however, it seems reasonable to guess that soon after the interferon is pulled into the cytoplasm, at least one of its C terminal domains contacts its binding site on the part of the receptor that was formerly hidden under the cell membrane. In so doing, the C terminal domain may well displace JAK2 from the receptor, freeing the enzyme to take up a phosphate group needed for its own activation. A similar sequence of events may enable type I interferons to send signals down their transduction pathway.

No matter how the signal transduction process gets started, the result is the synthesis of proteins that protect us from harm. Investigators have now identified more than 30 of the proteins that are induced by interferons, many of which play key roles in inhibiting viral replication. Better still, workers have learned quite a bit about how the proteins function.

For example, one of the best-studied proteins (the eIF-2-alpha protein kinase) interferes with the cellular machinery that viruses exploit in order to reproduce themselves. Viruses trick the protein-making machinery of host cells into translating viral messenger RNA into the proteins needed to make new infectious particles. Messenger RNA, viral or otherwise, is translated by ribosomes. These structures travel down the length of the RNA strand, linking one specified amino after another to a growing protein chain. First, however, each ribosome has to be built. Several molecules join together on the RNA transcript to form the smaller of two ribosomal subunits, and then the larger subunit comes on board.

All three interferons can precipitate the production of the eIF-2-alpha protein kinase, the active form of which phosphorylates one component required for forming the smaller ribosomal subunit. Such phosphorylation blocks further construction of the sub-



INTERFERONS ACTING AT THE CELL SURFACE (*top*) are believed to switch on genes in the nucleus (*bottom*) by triggering a remarkably short chain of molecular interactions. First, the proteins bind to their respective receptors (*1*). Type I interferons (*left*) bind to a single receptor complex, and type II molecules (*right*) bind to two copies of a different receptor. In both cases, binding activates enzymes (kinases) that phosphorylate so-called Stat proteins in the cytoplasm (*2*). Those proteins and others then assemble on distinct sites on genes in the nucleus (*3*). Such assembly leads to transcription, or copying, of the bound genes into molecules of messenger RNA (*4*), the templates from which proteins are made.

unit and thus stalls protein synthesis. The newly made kinase becomes active only when it encounters double-strand RNA. Such RNA appears in a cell only when a virus replicates its genetic material. Consequently, the enzyme blocks protein synthesis in infected cells but not in healthy ones.

Among other groups of proteins induced by both type I and type II interferons is the family consisting of the 2′,5′-oligo (A) synthetases. These enzymes, too, interfere with the production of viral proteins, but they do so by activating enzymes that break down RNA before it can be translated into protein. The proteins generated in response to interferons are so diverse that they are able to inhibit the proliferation of a variety of viruses. Some of the proteins that inhibit protein synthesis may also participate in the demonstrated ability of interferons to slow the growth of certain tumors in culture and in the body. As is true of viruses, in order to reproduce, cancerous and other cells must be able to manufacture new proteins.

Studies conducted since the 1970s have detailed not only the antiviral effects of interferons but also the extraordinary breadth of their immunologic repertoire. Their immunologic effects include an ability to arouse cells of the immune system against tumors. Indeed, the discovery that ignited some of the greatest excitement over the likelihood that interferons could be useful in medicine was the observation that all interferons enhance destruction of tumor cells by natural killer cells. The finding was made independently in the late 1970s by groups led by Ronald B. Herberman and Julie Y. Djeu of the National Cancer Institute, Ion Gresser of the Pasteur Institute in Paris and Giorgio Trinchieri of the Wistar Institute in Philadelphia.

Interferon-alpha is now used routinely to treat two cancers, and collectively the three major interferons are being tested in patients suffering from many other malignancies [*see bottom illustration on page 74*]. Unfortunately, physicians are limited in the doses they can administer to cancer patients and others because interferon-alpha, and to a lesser extent interferon-beta and interferon-gamma, can produce serious side effects. They can cause flulike symptoms, such as fever and fatigue, and they can lead to decreased production of blood cells in the bone marrow. Ways to combat these effects are under study.

Even before the antitumor potential of the interferons emerged, researchers were beginning to delineate the various

other ways these proteins modulate the defensive activities of the immune system. In the early 1970s Johnson, then at the Cincinnati branch of the FDA, established with Samuel Baron of the National Institutes of Health that interferons could influence the activity of *B* lymphocytes. These lymphocytes are the white blood cells that secrete antibody molecules after receptors on their surface recognize foreign proteins, or antigens, on pathogens and their products. Antibodies neutralize invaders directly or mark them for destruction by other parts of the immune system.

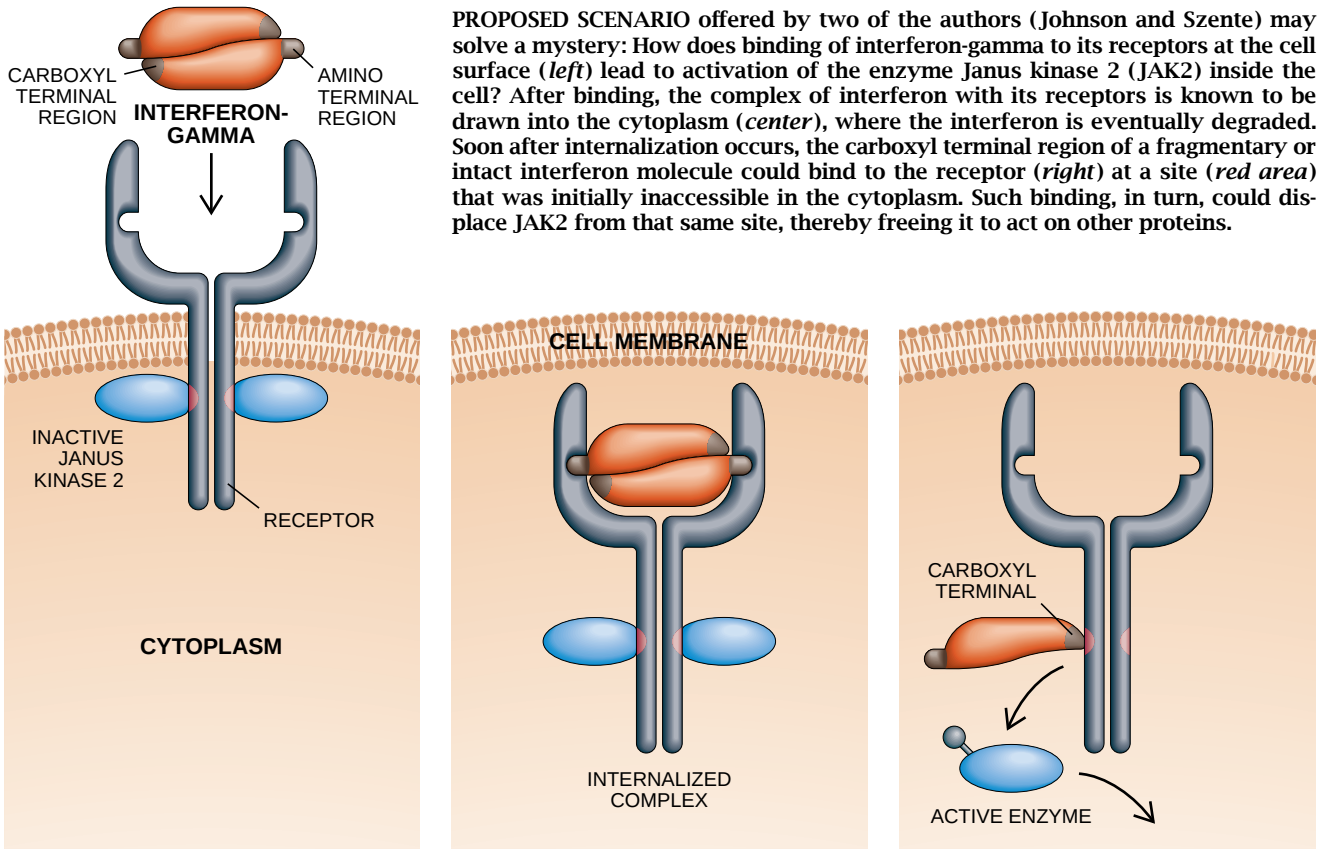
Johnson and Baron—and, separately, Gresser—found that if type I interferons act on *B* cells before the cells have begun to produce antibodies, the cytokines block such production. If, however, the cells are already producing antibodies, synthesis is enhanced. Subsequent studies carried out by Johnson's group in Florida demonstrated that interferon-gamma could also regulate the activities of another group of lymphocytes, a subset of *T* lymphocytes known as suppressor CD8 cells. After interferon-gamma stimulates these *T* cells, the *T* cells inhibit antibody production by *B* cells. They also retard the manufacture of certain cytokines by other cells.

Several discoveries made in the 1980s then revealed a major way in which interferon-gamma can amplify immune reactivity. Johnson's laboratory was among those to demonstrate early in the decade that the macrophage, another form of white blood cell, needs interferon-gamma in order to carry out many of its functions. Macrophages are scavenger cells; they engulf and degrade diseased or infected cells and microbes of every kind. They also stimulate other immune system cells.

Interferon-gamma can induce macrophages to kill tumor cells and cells infected by parasites, bacteria and viruses. It can also prod macrophages to destroy pathogens that have colonized the scavengers themselves. And interferon-gamma stimulates macrophages to produce what are called class II MHC (major histocompatibility complex) molecules. After macrophages ingest pathogens, they break up several of the microbes and fit the fragments into grooves on the MHC molecules, which are then transported to the cell surface. There they display the antigenic fragments to what are called CD4 *T* cells. (These lymphocytes can "see" antigens only if the foreign fragments are complexed with a class II MHC molecule.) Having recognized particular antigens, the CD4 cells proliferate and release chemicals that help other immune system cells to fight off infections.

The collected findings, only some of which are described here, contribute to the growing sense that interferon-gamma, which is made mainly in response to cellular infections and tumors, serves as a kind of immunologic switch. The protein helps to turn on the cell-mediated arm of the immune system, consisting of macrophages, various kinds of *T* cells and other cells that respond to microbes hidden inside the cells of other tissues. At the same time, interferon-gamma may dampen the production of antibodies. Antibodies are better suited to eradicating pathogens that establish colonies outside of cells.

The résumé of interferons is broader still and goes beyond immunity. Studies of cells in culture indicate that they can regulate the differentiation, or specialization, of certain cell types, among them fibroblasts. Fibroblasts normally synthesize the protein collagen, a major constituent of connective tissue. Yet some fibroblasts grown in culture can be forced to differentiate into fat cells if they are exposed to glucose and certain hormones, such as insulin. Livia Cioé and her colleagues at the Wistar Institute and Sidney E. Grossberg and his colleagues at the Medical College of Wisconsin have found that such differentiation can be blocked by type I interferons. This work raises an intriguing, still unanswered question: Does interferon-gamma have an influ-



PROPOSED SCENARIO offered by two of the authors (Johnson and Szente) may solve a mystery: How does binding of interferon-gamma to its receptors at the cell surface (*left*) lead to activation of the enzyme Janus kinase 2 (JAK2) inside the cell? After binding, the complex of interferon with its receptors is known to be drawn into the cytoplasm (*center*), where the interferon is eventually degraded. Soon after internalization occurs, the carboxyl terminal region of a fragmentary or intact interferon molecule could bind to the receptor (*right*) at a site (*red area*) that was initially inaccessible in the cytoplasm. Such binding, in turn, could displace JAK2 from that same site, thereby freeing it to act on other proteins.

MANUFACTURE OF INTERFERONS is now routinely accomplished through genetic engineering. Here a technician tends tanks used in the mass production of interferon-gamma.

ence on the amount of fat in the body?

Quite recently, another facet of the interferon story, involving reproduction, began to unfold. For a new pregnancy to progress in many mammals, the egg case, or corpus luteum, that envelops the ovum before fertilization must survive and secrete progesterone for an extended period. In humans, a substance called human chorionic gonadotropin, made by the nascent placenta, prevents the body from degrading the corpus luteum prematurely. In many domestic animals, such as cows, sheep and goats, the outer layer (the trophoblast) of membranes surrounding the embryo secretes copious amounts of a different molecule—called trophoblastin—to protect the corpus luteum.

In 1993 Kazuhiko Imakawa of the University of Kansas and R. Michael Roberts and their colleagues at the University of Missouri and at Upjohn Company determined the amino acid sequence of sheep trophoblastin, which one of us (Bazer) had isolated in 1982. Unexpectedly, the sequence closely resembled that of interferon-alpha. Later the protein was shown to bind to the same receptor as interferon-alpha. Both findings suggest that trophoblastin is, in fact, an interferon. But does it possess antiviral activity?

Indeed it does. Carol H. Pontzer of the University of Florida, along with Bazer and Johnson, has established that the protein inhibits viral replication as effectively as any known interferon. It is therefore now classified as an interferon and called tau, for trophoblast. Humans, too, make interferon-tau, but its role in the human body is not known.

Further studies of interferon-tau have shown that it inhibits the enzyme reverse transcriptase in cells infected by the human immunodeficiency virus (HIV), the cause of AIDS. (HIV produces reverse transcriptase in order to read its genes into the DNA of the host cell.) Interferon-tau also blocks the division of tumor cells in culture.

Good results in the laboratory do not necessarily translate into good results in the clinic. Nevertheless, these discoveries raise intriguing therapeutic possibilities. Tau's most exciting quality, however, from a medical standpoint, may be its low toxicity. Whereas high doses of other interferons can damage cultured cells, equally high doses of tau do not seem to have the same effect. The low toxicity of tau suggests that the human version might one day prove to be a safer yet still potent alternative to the other interferons. On the basis of such findings, Pepgen Corporation (to which Johnson is a consultant) has obtained the rights to develop interferon-tau as a therapeutic agent.

In spite of their current limitations, interferons, which are delivered by injection, are now well established as useful drugs. Moreover, their application is likely to expand as research continues. Interferon-alpha is the most widely used. In 1986 it became the first interferon to gain FDA approval, as a therapy for hairy-cell leukemia (a rare malignancy). Taken in low doses, interferon-alpha causes significant regression of the cancer in about 90 percent of patients, who must then take it indefinitely in order to avoid relapses.

In 1988 interferon-alpha was approved for treating a much more common disorder: sexually transmitted genital warts, which grow as a result of infection by papillomavirus. Injection of the drug into a wart or under the surrounding skin eradicates such growths in as many as 70 percent of patients who are unresponsive to other therapies. In contrast to cryotherapy and surgery, interferon-alpha is not caustic and does not cause scarring.

In that same year the FDA approved the use of interferon-alpha in the treatment of a second cancer, Kaposi's sarcoma. This cancer was once rare in

	INTERFERON-ALPHA	INTERFERON-BETA	INTERFERON-GAMMA
FDA APPROVED	CHRONIC HEPATITIS B AND C HAIRY-CELL LEUKEMIA GENITAL WARTS CAUSED BY PAPILLOMAVIRUS KAPOSI'S SARCOMA	RELAPSING-REMITTING MULTIPLE SCLEROSIS	CHRONIC GRANULOMATOUS DISEASE
CURRENTLY IN CLINICAL TRIALS	THROAT WARTS CAUSED BY PAPILLOMAVIRUS COLON TUMORS HIV INFECTION KIDNEY TUMORS CHRONIC MYELOGENOUS LEUKEMIA BLADDER CANCER NON-HODGKIN'S LYMPHOMA MALIGNANT MELANOMA	BASAL CELL CARCINOMA	KIDNEY TUMORS LEISHMANIASIS

MANY DISORDERS are now being treated by interferons. The Food and Drug Administration has approved their use in seven diseases. Clinical studies are under way for a range of other conditions, only some of which are listed here.

young people. Now it appears frequently in individuals infected with HIV. Unfortunately, some 30 percent of patients who receive interferon-alpha for this disease withdraw from treatment because of side effects resulting from the high doses required. Some evidence implies, however, that giving interferon-alpha with zidovudine (also known as AZT)—or other nucleoside analogues used to combat HIV itself—enhances the drug's effect. Perhaps this synergy will obviate the need for high doses.

Sufferers of hepatitis, too, have benefited from interferon-alpha. Before the protein was licensed in 1991 for the treatment of chronic hepatitis C (non-A, non-B hepatitis), there was no reliable therapy for this contagious viral infection of the liver. Now a six-month course of treatment can eliminate symptoms, and longer treatment periods are being adopted to prevent relapse. In 1992 interferon-alpha also became the first approved treatment for chronic hepatitis B, which affects about 300 million people in the world. Some form of therapy for these diseases was needed urgently because, unchecked, they can lead to cirrhosis and cancer of the liver.

Interferon-gamma and interferon-beta have joined the pharmaceutical arsenal as well. Interferon-beta became the newest addition last year, when the FDA, under its accelerated approvals program, licensed the drug as a treatment for the relapsing-remitting form of multiple sclerosis, in which symptoms come and go.

Multiple sclerosis is an autoimmune disease. Macrophages, *T* cells and *B* cells are all thought to collaborate in the destruction of the myelin sheath that protects nerve fibers in the central nervous system. Degradation of myelin leads to many symptoms, among them double vision, loss of balance, muscle weakness and paralysis. Investigators have not yet elucidated the precise way in which interferon-beta eases symptoms, but it probably helps by directly or indirectly suppressing cells of the immune system. In any case, it is clear that in many patients therapy decreases the frequency of flare-ups.

The gamma form of interferon is used to combat chronic granulomatous disease, a hereditary disorder. In this condition, white blood cells called granulocytes ingest but fail to kill bacteria. As a consequence, patients endure severe, repeated infections in many parts of the body, such as the skin, liver, lymph nodes, lungs and bone. The interferon, which was approved in 1990, is given prophylactically, often with antibiotics. It is thought to act in part by boosting the ability of macrophages to destroy bacteria.

LELA RECTOR of Marion, Ill., was a winner in an unusual lottery last year. After a form of interferon-beta gained FDA approval for treating multiple sclerosis in July 1993, demand far exceeded supply. So the distributor, Berlex Laboratories, held a lottery for some 67,000 qualified patients. Rector, who drew number 109, was among some 30,000 people reportedly in treatment by March 1, 1994. The other registrants are expected to have access to the drug by the end of this year.

Of the many other applications that are in clinical trials, one that is especially exciting is the use of interferon-gamma to combat leishmaniasis, a serious problem in many parts of the world. Leishmaniasis is an ulcer-causing infection of macrophages (often those in the skin) by the leishmania parasite and is common in parts of Africa, North and South America, eastern and southern Europe, and Asia. Clinical studies carried out in the past few years provide strong evidence that interferon-gamma alone or in combination with antiparasitic chemicals can control and, at times, cure the disease.

Even the mechanism of action is known. Juana Wietzerbin and her colleagues at the Curie Institute in Paris have determined that interferon-gamma stimulates parasite-infected macrophages to produce the chemical nitric oxide. Nitric oxide turns out to be very toxic to leishmania; it also plays a major role in eradicating some viruses.

Interferons, then, have clearly proved their value in medicine. Yet they may turn out to be most helpful when administered with other therapies. For example, there are indications that interferon-alpha in combination with tamoxifen may be more effective against breast cancer than either drug alone. The potential benefit of combination therapies using interferons has yet to be explored fully.

All in all, research into interferons has progressed impressively since 1957. We now know the structure of the molecules and have a good handle on how they work. We have learned that inhibition of viral growth is but one of many essential functions of interferons. And we are moving at full speed to apply this knowledge in the treatment of a host of diseases.

FURTHER READING

MECHANISM OF INTERFERON-γ PRODUCTION AND ASSESSMENT OF IMMUNOREGULATORY PROPERTIES. Howard M. Johnson in *Lymphokines,* Vol. 11. Edited by Edgar Pick. Academic Press, 1985.

INTERFERONS AND THEIR ACTIONS. Sidney Pestka, Jerome A. Langer, Kathryn C. Zoon and Charles E. Samuel in *Annual Review of Biochemistry,* Vol. 56, pages 727–777; 1987.

THREE-DIMENSIONAL STRUCTURE OF RECOMBINANT HUMAN INTERFERON-γ. Steven E. Ealick, William J. Cook, Senadhi Vijay-Kumar, Mike Carson, Tattanahalli L. Nagabhushan, Paul P. Trotta and Charles E. Bugg in *Science,* Vol. 252, pages 698–701; May 3, 1991.

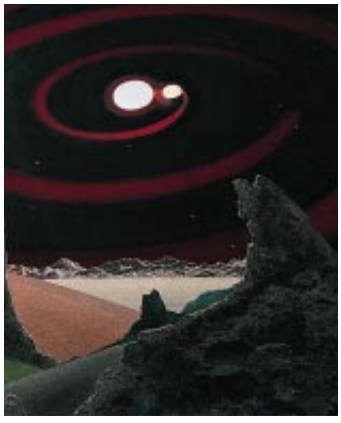
EARLY EVENTS IN SIGNALLING BY INTERFERONS. Sandra Pellegrini and Chris Schindler in *Trends in Biological Sciences,* Vol. 18, pages 338–342; September 1993.

THE MOLECULAR CELL BIOLOGY OF INTERFERON-γ AND ITS RECEPTOR. Michael A. Farrar and Robert D. Schreiber in *Annual Review of Immunology,* Vol. 11, pages 571–611; 1993.

Chesley Bonestell's Astronomical Visions

This artist's unique combination of technical knowledge and graphic prowess brought astronomy alive and helped to advance the manned spaceflight program

by Ron Miller




Artist and illustrator Chesley Bonestell was born on January 1, 1888, 15 years before the maiden voyage of the Wright brothers' airplane and 38 years before the launch of the first tiny liquid-fuel rocket. When he died 98 years later, men had walked on the moon, and robotic craft had toured most of the planets of the solar system. Bonestell's paintings of astronomical vistas and space-faring explorers not only anticipated the great technological triumphs of the 20th century, they helped to bring them about. His attention to pictorial and technical realism transformed the popular perception of spaceflight from fantasy to an immediate possibility, if not a near certainty.

Bonestell came from a moderately wealthy family in San Francisco, where he and his two older sisters were raised by his father and grandfather (his mother died when he was an infant). Chesley displayed an acute interest in both art and astronomy early on. He started drawing at age five and began formal art instruction by the time he was 12. When he was 17, he visited Lick Observatory; he was excited and inspired by the view through the observatory's 12-inch and giant, 36-inch refractors. As soon as he returned home, Bonestell sketched a picture of Saturn as he had observed it—probably his first attempt at space art.

Despite his attraction to painting and drawing, Bonestell initially applied his visual skills toward architecture—spurred, in part, by his grandfather's disdain for the "bohemian reputation of artists." He took architecture courses at Columbia University, and although he never attained a degree, he passed the California State Board of Architects' examination. By 1911 he was working as a renderer for the innovative architect Willis Polk, who quickly made Bonestell his chief designer.

RON MILLER has worked extensively as an illustrator specializing in scientific subjects. He earned his bachelor's degree in illustration from Columbus College of Art and Design in Ohio in 1970. He is the author or co-author of nearly two dozen books, among them *The History of Earth,* a new translation of Jules Verne's *20,000 Leagues under the Sea,* and *The Dream Machines,* a history of 2,000 years of real and imagined spaceship design. Miller has worked as a production illustrator on motion pictures, including *Dune;* in 1992 he designed the 10-stamp Space Exploration set for the U.S. Postal Service. He is also a lecturer at the International Space University in Cambridge, Mass.

CHESLEY BONESTELL'S IMAGINATION gave visual life to faraway worlds and unborn technologies. Shown above (*left to right*): a trail of gas streams from the double star Beta Lyrae; the moon rises over the young, molten earth; a winged rocket sheds its third stage; and Bonestell confers with Wernher von Braun (*right*), who helped to realize America's manned space program. Bonestell's cinematic experience aided him in depicting dramatic astronomical settings, such as this view of astronauts exploring Saturn's satellite Mimas (*opposite page*).

During his time with Polk, Bonestell produced renderings and decorative design work for the Seventeen-Mile Drive at Pebble Beach, Calif., part of the ambitious real estate development on the Monterey Peninsula. In 1927 he assisted William van Alen in the design of New York City's renowned Chrysler Building. Later, Bonestell created detail drawings for another American landmark—the Golden Gate Bridge in San Francisco. Yet even as his architectural career flourished, he occasionally made drawings of Martian and lunar landscapes, which he gave away to friends and family.

In 1938, at the age of 50, Bonestell moved to Hollywood and began a very different, highly lucrative career as a special-effects matte painter. The first film he worked on was Orson Welles's *Citizen Kane.* All the views of turn-of-the-century New York and of Charles Foster Kane's mansion, Xanadu, are Bonestell's artwork. Over the next decade he contributed to movies from every major studio; he ended up as the film industry's highest-paid matte artist. In the 1950s Bonestell briefly returned to Hollywood to provide the special-effects art for several science-fiction films produced by George Pal, including *Destination Moon* and *War of the Worlds.*

Through careful investments, Bonestell achieved a comfortable financial independence, which allowed his mind to return to his old love, astronomical illustration. He took inspiration from the skills he had acquired through his film work. In the introduction to his 1949 book *The Conquest of Space,* he recalled, "As my knowledge of the technical side of the motion picture industry broadened I realized I could apply camera angles as used in the motion picture studio to illustrate 'travel' from satellite to satellite, showing Saturn exactly as it would look, and at the same time I could add interest by showing the inner satellites or outer ones on the far side of Saturn, as well as the planet itself in different phases." Those thoughts took visible form in a series of paintings, published in the May 29, 1944, issue of *Life,* that astonished and delighted the public.

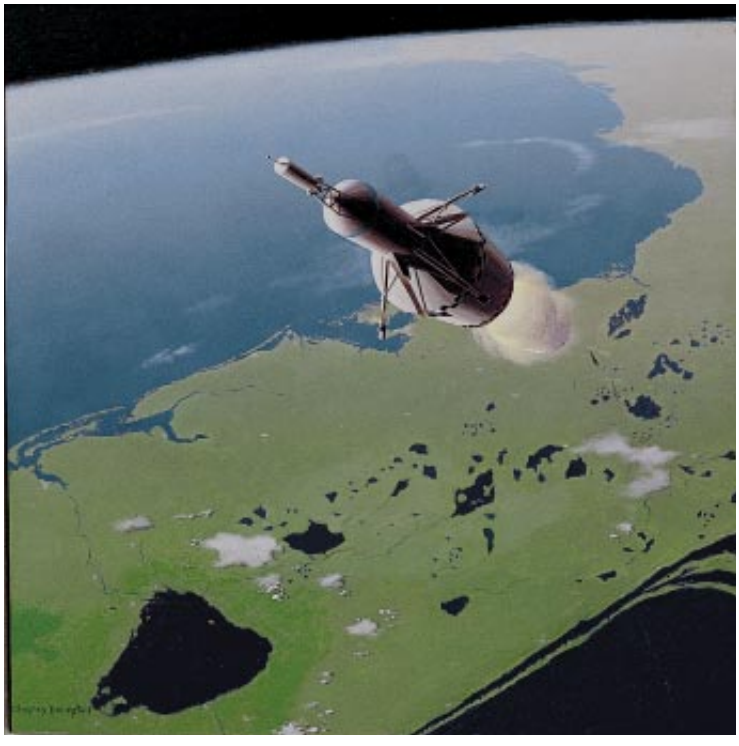
Baby satellite (1953)

Martian colony (1964)

Construction of a manned space station (1949)

Lunar expedition (1953)

Mars base (1956)



Rocket launch over Florida (1962)

Lunar landing (1953)

SPACEFLIGHT figured prominently in Bonestell's art, especially during the 1950s. Some of these images, such as the launch over Florida, the unmanned satellite and the lunar landers, closely anticipate real spacecraft. Others, including the bases on the moon and on Mars, still lie in the future. The realism of Bonestell's paintings made the human exploration of space seem tangible to the American public.

At about this time, Bonestell began a long-term collaboration with Willy Ley, an expatriate German historian and science popularizer who had been a member of the pioneering German Rocket Society (Verein für Raumschiffahrt). Taking advantage of Ley's advice, Bonestell began adding spacecraft to his paintings. In 1946 *Life* published another set of his illustrations, this time depicting a manned flight to the moon.

Bonestell's imagined spacescapes soon showed up in a growing number of magazines, including *Look, Coronet, Pic Magazine, Mechanix Illustrated, Air Trails* and *Astounding Science Fiction.* The November 1948 issue of *Scientific American* featured a Bonestell cover painting of the sun as seen from Mercury. So popular had his art become that the overworked painter once inadvertently sent the cover painting for a science-fiction magazine to the wrong publication. The editor of that magazine promptly ran it.

Bonestell's first book, *The Conquest of Space,* featured 48 of his stunning paintings and an accompanying text by Ley. In a review for *The Aeroplane,* the prominent science and science-fiction writer Arthur C. Clarke wrote, "Mr. Bonestell's remarkable technique produces an effect of realism so striking that his paintings have sometimes been mistaken for actual colour photographs by those slightly unacquainted with the present status of interplanetary flight.... In the years to come it is probably destined to fire many imaginations, and thereby to change many lives."

During the first half of the 1950s, Bonestell entered a period of unprecedented creative activity that showed just how prescient Clarke had been. In 1951 Cornelius Ryan, the associate editor of *Collier's* magazine, invited Bonestell to illustrate a series of five articles on the future of spaceflight. The prime author was Wernher von Braun, America's most influential rocket scientist. Bonestell and von Braun became good friends; over the years, they collaborated closely in developing concepts for manned spacecraft.

Von Braun found himself awed by Bonestell's sharp eye for scientific and engineering accuracy. Some years later von Braun wrote that "Chesley Bonestell's pictures...are far more than reproductions of beautiful ethereal paintings of Worlds Beyond. They present the most accurate portrayal of those faraway heavenly bodies that modern science can offer. I do not say this lightly. In my many years of association with Chesley I have learned to respect, nay fear, this wonderful artist's obsession with perfection. My file cabinet is filled with sketches of rocket ships I had prepared to help him in

Mars from Phobos (1949)

Surface of Venus (1949)

Surface of Pluto (1949)

Surface of Mercury (1948)

Jupiter from Europa (1948)

Saturn from Titan (1952)

WORLDS OF THE SOLAR SYSTEM turned into craggy landscapes, reminiscent of the paintings of the Hudson River School, in Bonestell's hands. Although many particulars in these works—the clear skies on Titan and improbably rugged mountains, for example—clash with recent scientific discoveries, Bonestell effectively evoked the essential details and exotic alienness of other planets and moons.

his art work—only to have them returned to me with penetrating detailed questions or blistering criticism of some inconsistency or oversight."

The articles in *Collier's*—published between 1952 and 1954—took America by storm. The country turned space-happy; reproductions and knockoffs of Bonestell's paintings appeared in settings ranging from commercial advertisements to television programs to school lunch boxes. After the *Collier's* series completed its run, it was recycled into a triad of highly popular books: *Across the Space Frontier, Conquest of the Moon* and *Exploration of Mars.*

Bonestell's artwork strongly influenced the American public and, in turn, the government to support an investment in space exploration. Although space travel was a familiar topic by the 1950s, writers had almost always relegated it to the distant future. But the central tenet of the *Collier's* series was that humans could venture into space using only existing technology and materials. The illustrations showed every nut and bolt in each stage of the evolution of a completely realized space program, from the first unmanned satellites to an expedition to Mars. Von Braun and Bonestell even presented the reader with pages of calculations and long lists of materials and costs.

As a result, scientists and politicians alike began to recognize that the successful exploration of space was not so much a matter of time or technology as it was of money and incentive. The researchers who had been urging the U.S. government to back an earth satellite program no longer seemed to be such wild visionaries. By the mid-1950s the U.S. had an artificial satellite project under way. In 1958, one year after the launch of *Sputnik 1,* the Americans (after a desperate, failure-studded scramble) lofted their own satellite, *Explorer 1.*

In the next decade Bonestell watched manned space exploration become a reality. He grumpily noticed that the soft lunar hills seen by the Apollo astronauts bore little resemblance to the craggy, romanticized landscapes in his paintings. Such inaccuracies do little to diminish the primary importance of Bonestell's work. His illustrations gave immediacy and verisimilitude to dry astronomical data and engineering calculations. Countless scientists, engineers and astronauts were inspired in their choice of careers by Bonestell's images.

Bonestell continued to work until he died in 1986, an unfinished painting still on his easel. Asteroid number 3129 was recently given the name "Bonestell"—a fitting honor for the man whose art contributed to the birth of the space age.

FURTHER READING

SPACE ART. Edited by Ron Miller. O'Quinn Studios, 1978.

WORLDS BEYOND: THE ART OF CHESLEY BONESTELL. Frederick C. Durant III and Ron Miller. Donning, 1983.

VISIONS OF SPACE: ARTISTS JOURNEY THROUGH THE COSMOS. David A. Hardy. Paper Tiger, 1989.

BLUEPRINT FOR SPACE: SCIENCE FICTION TO SCIENCE FACT. F. I. Ordway III and R. Liebermann. Smithsonian Institution Press, 1992.

Directional Drilling

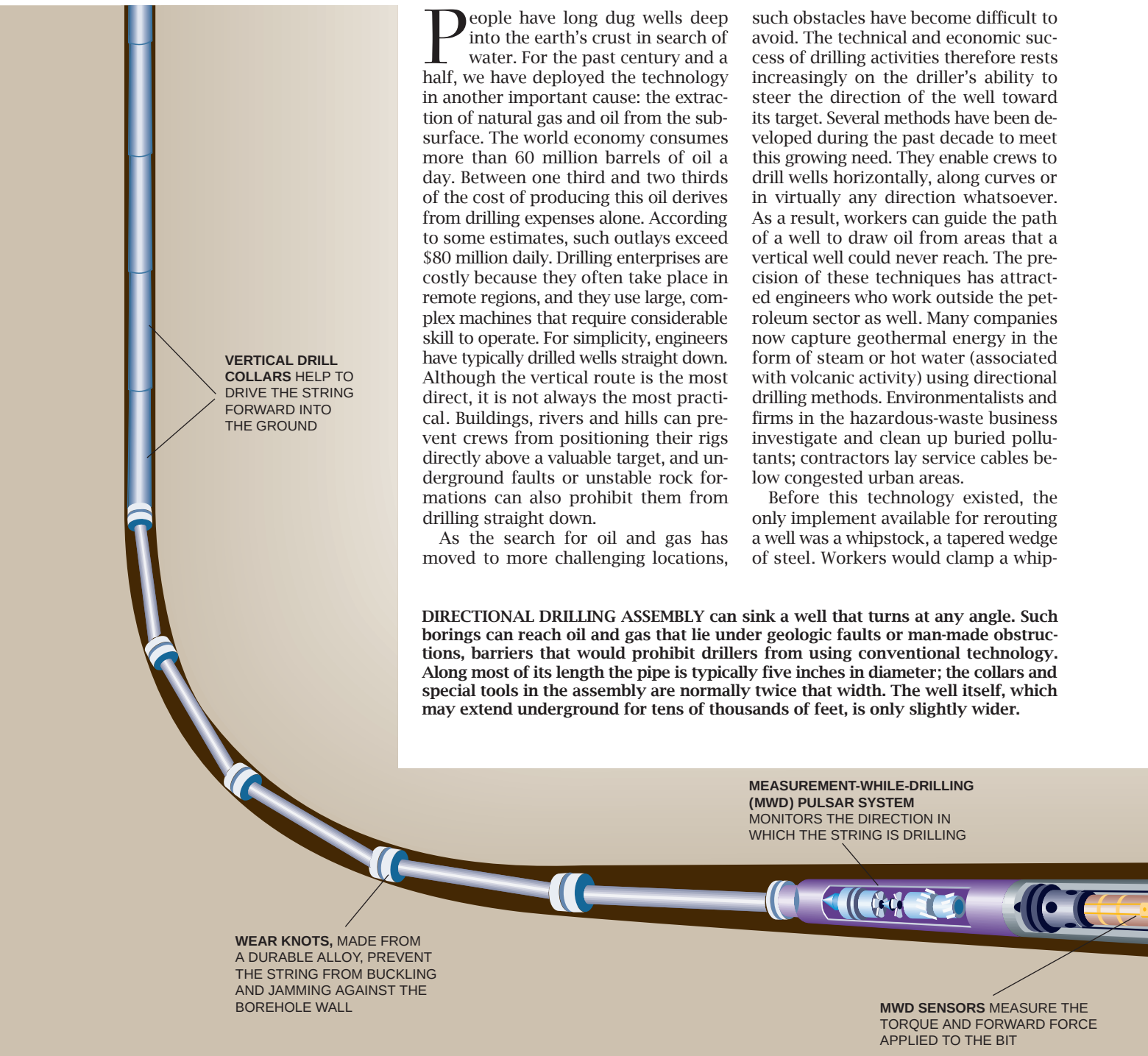
New techniques enable crews to drill around natural or man-made obstructions for oil and gas. These same methods are used to sample underground pollutants and bury service lines

by George A. Cooper

People have long dug wells deep into the earth's crust in search of water. For the past century and a half, we have deployed the technology in another important cause: the extraction of natural gas and oil from the subsurface. The world economy consumes more than 60 million barrels of oil a day. Between one third and two thirds of the cost of producing this oil derives from drilling expenses alone. According to some estimates, such outlays exceed $80 million daily. Drilling enterprises are costly because they often take place in remote regions, and they use large, complex machines that require considerable skill to operate. For simplicity, engineers have typically drilled wells straight down. Although the vertical route is the most direct, it is not always the most practical. Buildings, rivers and hills can prevent crews from positioning their rigs directly above a valuable target, and underground faults or unstable rock formations can also prohibit them from drilling straight down.

As the search for oil and gas has moved to more challenging locations, such obstacles have become difficult to avoid. The technical and economic success of drilling activities therefore rests increasingly on the driller's ability to steer the direction of the well toward its target. Several methods have been developed during the past decade to meet this growing need. They enable crews to drill wells horizontally, along curves or in virtually any direction whatsoever. As a result, workers can guide the path of a well to draw oil from areas that a vertical well could never reach. The precision of these techniques has attracted engineers who work outside the petroleum sector as well. Many companies now capture geothermal energy in the form of steam or hot water (associated with volcanic activity) using directional drilling methods. Environmentalists and firms in the hazardous-waste business investigate and clean up buried pollutants; contractors lay service cables below congested urban areas.

Before this technology existed, the only implement available for rerouting a well was a whipstock, a tapered wedge of steel. Workers would clamp a whip-



DIRECTIONAL DRILLING ASSEMBLY can sink a well that turns at any angle. Such borings can reach oil and gas that lie under geologic faults or man-made obstructions, barriers that would prohibit drillers from using conventional technology. Along most of its length the pipe is typically five inches in diameter; the collars and special tools in the assembly are normally twice that width. The well itself, which may extend underground for tens of thousands of feet, is only slightly wider.

stock, point upward, inside the borehole. When they lowered the drilling assembly into the hole again, the sloping face of the wedge would force the drill bit sideways, away from its original direction. The team could make further changes using additional whipstocks, and so, in principle, holes could be drilled in any direction. In practice, however, so much time had to be devoted to the operation that crews resorted to it only when they had to bypass equipment stuck in the hole, a maneuver known as sidetracking.

That attitude changed when oil companies became interested in obtaining oil from offshore sites. To get at oil buried below the ocean, crews needed to operate from floating platforms or, more commonly, from platforms standing on the seafloor. Such installations are enormously expensive; the Gullfaks C platform in the North Sea, for example, weighs 1.5 million tons, stands 850 feet tall and cost $2 billion to build. Its base covers four acres of the seabed, and the steel used to reinforce it could build 10 Eiffel Towers. Several of these structures are often required to drain a reservoir. Using directional drilling techniques instead, a crew could work from one platform to drill wells along different paths and reach various target sites. That strategy has more than redeemed its promise. In January 1993, Norway's national oil company set a world record by drilling a well that extended 23,917 feet horizontally from its starting point, at a depth of 9,000 feet in the North Sea.

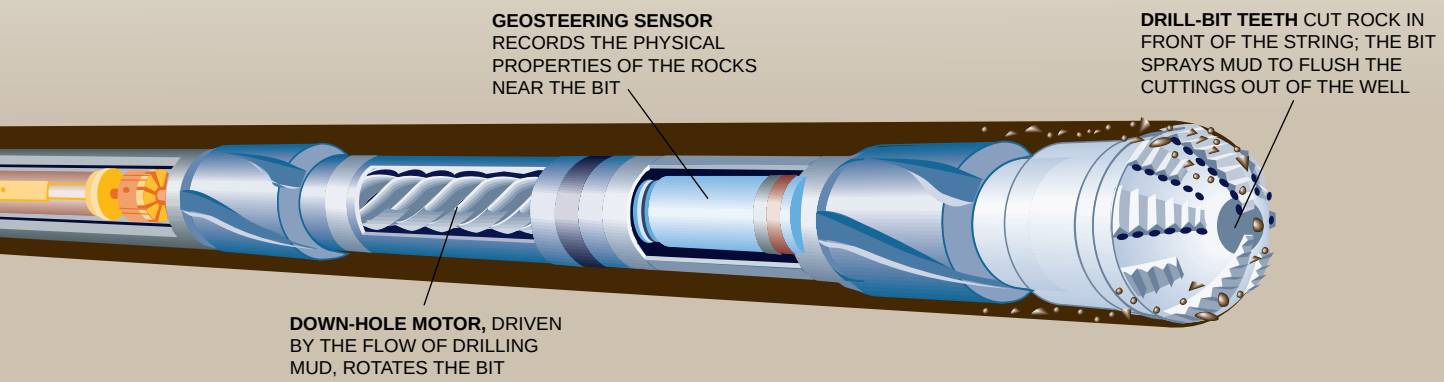
Oil companies found directional drilling capabilities attractive for a second reason: the orientation of a well strongly affects how efficiently it can be drained. Oil and gas are found in porous and permeable rocks lying below impermeable beds that seal the reservoir from above. The oil floats in a horizontal layer between gas and water. The oil layer is usually wider than it is high, and so vertical wells contact the oil over only a few feet. The production engineer aims to tap as much oil as possible, without drawing out the water that lies under the oil. The gas layer above can be collected to sell or left in place to put pressure on the more valuable oil, driving it toward the well. Despite the engineer's expertise, in a short period, as the fluids in the reservoir are drained, the water level rises and the gas level drops. At that point, the well loses value because it begins to yield water and gas as well as oil. In contrast, a well running horizontally through the oil zone would take up much less water and gas than would a vertical well because the boring in the oil zone could contact over hundreds or maybe even thousands of feet of the oil zone.

Horizontal wells offer the possibility of yielding more oil in another way. Oil frequently seeps into fractures that traverse the reservoir. These fractures usually run parallel to one another (most often vertically) in a given location. They do so because the crust normally breaks in a direction perpendicular to that in which there is the least geologic stress. Skilled operators could use directional drilling methods to place horizontal wells that would intersect and drain the predominant fractures in an oil field. Indeed, directional drillers have increased the production of their wells as much as 10-fold by using this technique.

Conventional drilling methods could not easily be modified to accomplish such feats. In these operations the drill bit is attached to a string of piping that workers rotate from the surface. The rock at the bottom of the hole is broken as the bit is forced downward by the weight of the drill pipes above. A hundred feet or more of particularly thick, heavy sections of pipe, known as drill collars, placed immediately above the bit provide the primary driving force. The rest of the string consists of thinner drill pipe, which hangs from the derrick or the drilling mast. The crew pumps drilling mud, usually made from clay and other additives in water, down the drill pipe and out through jet nozzles in the face of the bit. The mud returns to the surface through the space between the pipe and the borehole, flushing out rock cuttings.

Most of the string is made from steel drill pipe that may be five inches in diameter and perhaps 10,000 feet long. Pipe of this size may appear very strong, but if the string were scaled down uniformly, it would resemble a hypodermic needle, more than 60 feet in length and one thirty-second of an inch in diameter. To drill along a predetermined curve, the crew must be able to influence which way the flexible string will bend. Directional drillers gain this control by inserting a bent section of collar (or a "bent sub") in the string close to the bit. This bent sub usually curves by no more than a few degrees, which is sufficient to drill a hole having a turning radius between a few hundred and a few thousand feet. Because the bend points the bit sideways, however, the drillers cannot rotate the entire assembly from the surface. If they did, the bit would trace a fine spiral in the rock, which would, in effect, create a vertical boring. For directional drilling assemblies, then, technologists developed motors that are placed in the hole to rotate the bit directly when the bent sub is in place. Rig hands on the drilling floor twist the pipe at the surface to change the orientation of the bend, so that the drill bit

GEORGE A. COOPER heads the petroleum engineering program at the University of California, Berkeley. His current interests concern methods for improving the efficiency of drilling and for reducing its environmental impact in petroleum, geothermal and civil engineering applications. Before arriving at Berkeley six years ago, he spent 20 years in Europe, investigating drilling and tunneling techniques as a government and industry researcher. He received a Ph.D. in materials science from the University of Cambridge.

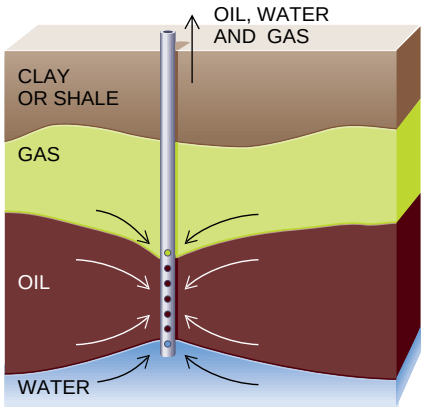


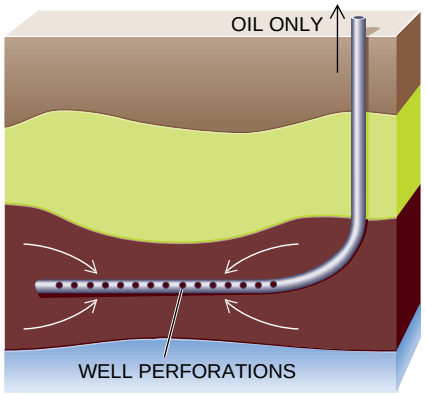
faces up, down, left or right as desired.

The success of these designs relies heavily on minimizing friction in the lower part of the hole. In a borehole that is nearly horizontal, the drill collars lean on the side of the hole. The workers must use a very slippery drilling mud so that no equipment or rock cuttings get stuck. In a truly horizontal hole the drill collars cannot lie near the bit, because they would just drag along the bottom wall of the boring instead of forcing the bit forward. Hence, in these situations the collars rest in the upper, vertical section of the hole. From this location, they push a relatively light section of pipe along the flatter part of the hole toward the bit. The lightweight pipe must be kept from buckling and jamming against the hole wall. To avoid this difficulty, engineers have created drill pipes that are both light and rigid.

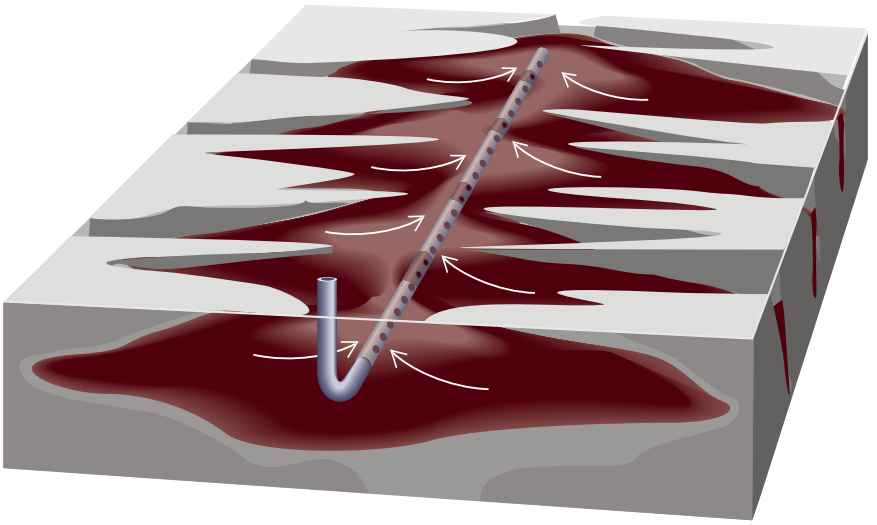
Design teams have solved other potential problems as well. The flow of the drilling mud through the pipe drives the down-hole motors used in a directional drilling assembly. At first, experimenters tested turbines, but these motors operated most efficiently at speeds of several hundred revolutions per minute—too fast for most conventional drill bits. Furthermore, because the mud can flow through the device whether it is turning or not, drillers sometimes could not tell whether the turbine was running at the correct speed, if at all.

So the industry has turned to a device known as a positive displacement motor. In these motors the mud flows in cavities between a spiral-fluted, steel rotor and a rubber-lined stator. The cavities open and close as the rotor turns, and there is a direct correlation between the mud's flow rate and the motor's speed. Manufacturers can build motors having a wide range of speed and torque by changing the number of radiating lobes on the spiral as well as by controlling the angle of those lobes. When drill strings first carried positive displacement motors, the rotor and mud flow chafed the rubber stator, wearing it out quickly. Luckily, the difficulty has since been solved.





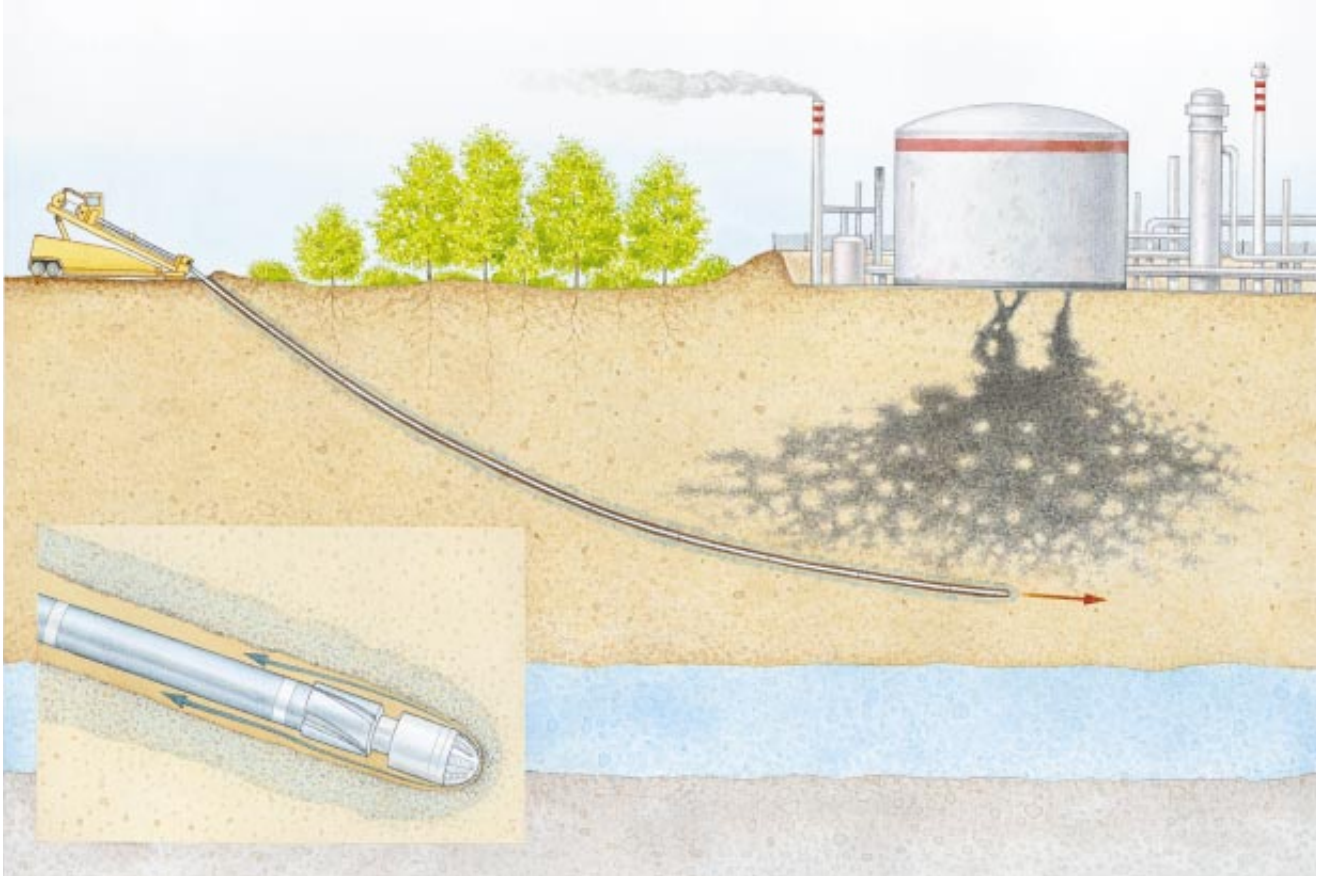
OIL PRODUCTION from a vertical well (*top*) is much less than from a horizontal well (*middle*). Horizontal wells experience very little coning, an effect that occurs when the fluids in a reservoir diminish so that a well begins to yield water and gas. A vertical well draws on a relatively narrow cross section of the oil layer; thus, as this layer thins, coning sets in quickly. A horizontal well can contact oil over hundreds or even thousands of feet deep, and so coning does not impede recovery until most of the oil has been collected. In addition, horizontal wells can intersect and drain fractures in the reservoir rock (*bottom*).



In addition to developing specialized, down-hole motors, engineers have devised novel assemblies to form holes that turn at sharp angles. For certain applications, the borehole must curve so acutely that the elastic bending of a flexible drill string is insufficient. One solution uses an external guide tube. This tube bears slots that enable it to flex between a straight position and one with a radius of curvature of 30 to 60 feet (depending on the hole size). A jointed drive tube composed of interlocking segments fits inside the guide tube. The links, like vertebrae, allow the tube to bend but still deliver torque to the bit, which sits at the end of the drive tube. Yet another flexible tube inside this backbone carries drilling mud to the bit. The steel of the external tube bends naturally into a curve. Although all three tubes must be straight in order to enter the well, once drilling starts, the guide's natural springiness directs the bit along a curved trajectory. Consequently, the hole becomes horizontal in a few tens of feet.

In certain reservoirs, a starlike arrangement of holes that extend horizontally from the bottom of a vertical well acts as an effective drain. To form these drains, specialist drilling teams lower a purpose-built guide into a well. The guide is shaped in such a way that it first bends an internal pipe into a curve and then straightens the piece horizontally within the span of one to two feet. The operator places this guide at the bottom of the main borehole and then feeds a narrower steel pipe through it. At the front end of the internal pipe, a jet nozzle fires high-pressure water that erodes the rock and thus enables the pipe to move forward. After the small-diameter pipe has passed a sufficient distance beyond the guide, workers use chemicals to sever it from the rest of the string. The guide is then removed or turned in a new direction to accommodate the placement of one or more additional holes.

The ability to steer a drill accurately is essential. At a minimum, the driller needs to know the inclination and the compass bearing (azimuth) of the borehole, as well as the tool-face angle. The angle of the tool face tells the crew which way the bit is pointing and therefore which way the borehole will travel

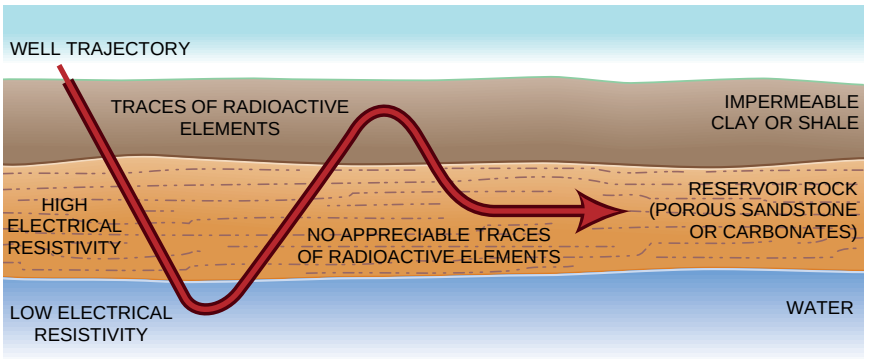


CLEANUP OPERATIONS now make use of directional drilling methods to investigate and treat pollutants that have leaked into the soil underneath industrial installations. This technique has many advantages, one being that a team does not need to work directly over the waste site or even nearby. In this way, workers run less risk of being contaminated. Also, if a low-temperature gas is used as the drilling mud, the pollutants freeze in place and can be sampled more accurately.

next. Field engineers keep track of this measurement by charting which way the drill pipe twists at the surface; a cartridge of instruments—a compass and pendulum, for example—lowered inside the drill pipe on a cable can indicate the orientation of the assembly at any moment.

Unfortunately, the extra sections of drill pipe that the crew must add to the top of the string as the hole gets deeper cannot be threaded onto the cable without disconnecting it. Hence, the cable and instrument pack must be removed while normal drilling operations are under way. To stop drilling and send the instrument pack down the hole every time a measurement is needed wastes considerable time. For the sake of efficiency, several systems have been developed that now possess self-powered measuring instruments that operate while drilling takes place. A measurement-while-drilling (MWD) device transmits information to the surface by stopping and releasing the flow of drilling mud as it passes through the tool. Changes in the flow rate cause fluctuations in the drilling mud pressure that can be detected at the surface and decoded to reveal the direction in which the assembly is drilling.

The most sophisticated MWD tools carry supplemental electronics that measure the physical properties of rocks in and around a reservoir. Such sensors open the way to "smart" drilling: Anadrill, a division of the Schlumberger group, has pioneered the technique of placing sensors at the drill bit. These sensors allow the driller to follow a particular geologic formation toward the pay zone (where the valuable oil resides) simply by monitoring the make-up of the rock in front of the bit. For example, petroleum reservoirs consist of porous, permeable sandstones or carbonates lying underneath layers of impermeable clays or shales. These layers seal the oil zone and usually contain traces of uranium, thorium and potassium. In contrast, the reservoir rocks are composed of relatively pure sand (silica) or calcium carbonate, neither of which contains appreciable quantities of radioactive elements. Instruments on the MWD tool can detect the radioactivity of the clays and shales lying above the reservoir and report their findings



WELL TRAJECTORY can stay in the productive reservoir rock if the drill string contains geosteering sensors. These sensors tell the drillers on the surface if the resistivity or radioactivity of the rock they pass through changes. In this way, the drillers know quickly when the well strays from the reservoir.

to the surface. Drillers can then steer the hole so that it always stays in the reservoir rock and skirts the unproductive shales.

Other probes measure the electrical resistivity of the rock surrounding the borehole. The rock minerals themselves have a very high resistivity, as do oil and gas. Water at this depth, on the other hand, is usually somewhat salty and gives low resistivity values. From this distinction, the driller can tell whether the bit is still in the hydrocarbon-bearing zone or has dipped below the oil-water contact. Finally, instruments can be included that gauge the torque and the forward force applied to the drill bit. By comparing down-hole measurements with those made at the surface, engineers can tell if some part of the equipment is getting stuck in the hole and thus predict a borehole collapse or other difficulties. The workers can also measure the level of torque registered for a given drilling force, information that will indicate whether the bit is clogged or has worn out and should be replaced.

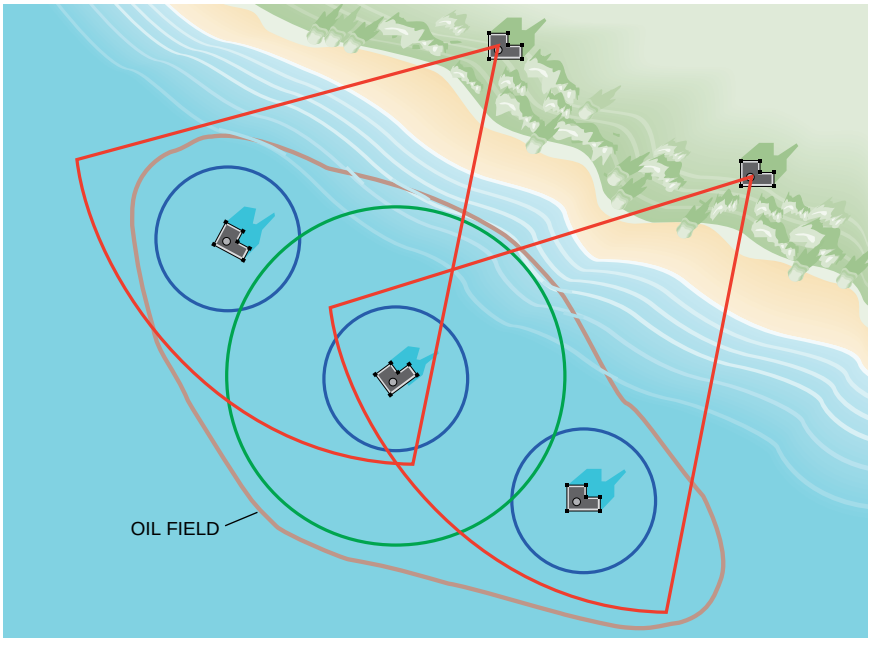
Directional drilling technology has found many uses outside the petroleum sector. Suppliers of water, gas, electricity and communications services are rapidly adopting the technology to pass cables or pipelines under roads, buildings and rivers. Directional drilling is ideal for such projects because, unlike conventional methods, it does not require access to the surface above the hole. Only 50 river crossings, having a total length of 60,000 feet, were drilled during the 1970s. In 1988 alone, however, various companies completed 200 installations, having a total length of 200,000 feet. Cherrington Corporation set a world record in early 1993 by installing 4,150 feet of 42-inch gas pipeline under the Sacramento River in California for the Pacific Gas Transmission–Pacific Gas & Electric Company. In Europe, 2,000 feet of 48-inch gas pipeline now lies under the Noord Holland canal in the Netherlands, and approximately 6,000 feet of eight-inch line runs under the Saint Lawrence River near Trois-Rivières in Quebec.

During these crossings, the hole must slope very gently so that the horizontal pipe passing under the river experiences as little stress from bending as possible. Therefore, these projects use equipment to drill holes that dip just below the river, instead of following a longer, U-shaped path. As a consequence, the hole has no vertical section. Powerful jacks on the drilling rig must push the drilling assembly into the ground without help from drill collars. A self-contained sensing system near the drill head can steer the equipment. If the hole is shallow and the surface above its path is clear, an electrical cable on the ground can transmit signals to underground sensing units placed near the bit. In this way, the driller can aim the direction of the bit very accurately; often the hole emerges within a foot of its goal after extending several hundred feet underground.

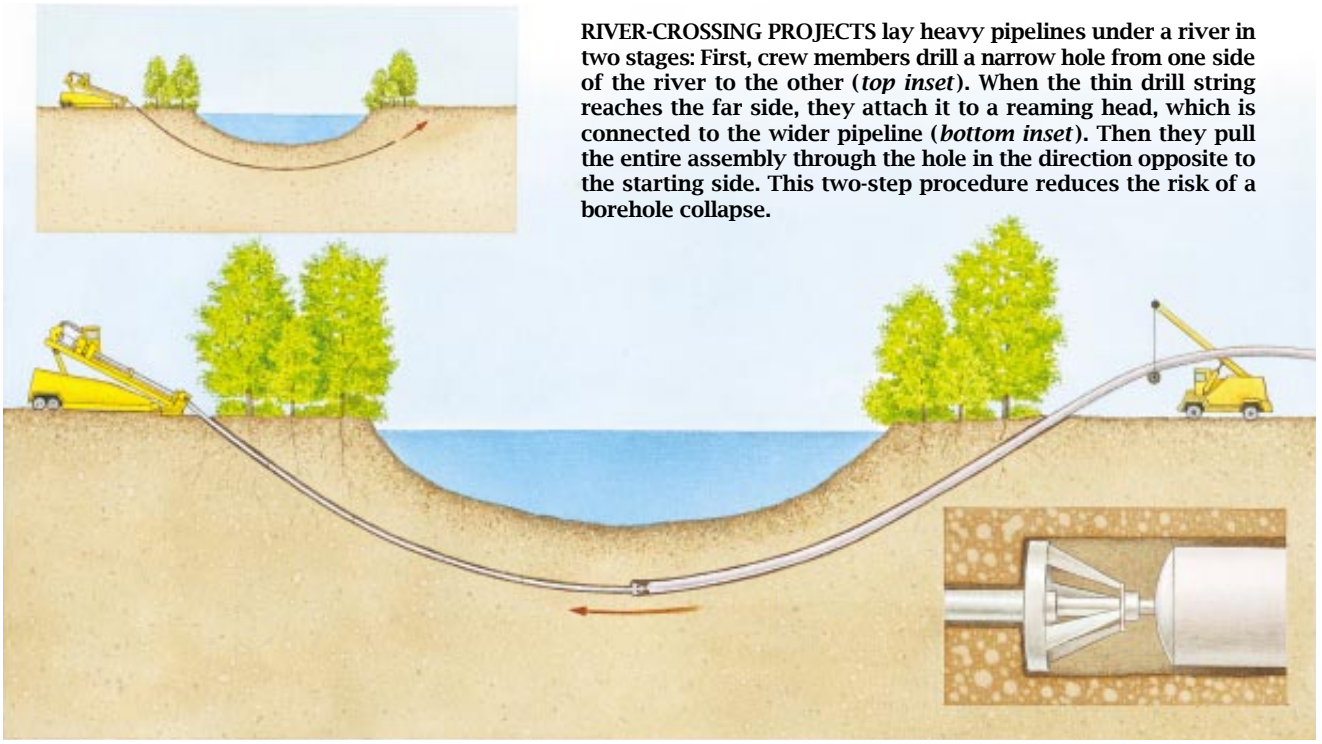
The soil under a riverbed is generally soft. Therefore, river-crossing operations must include special measures to avoid borehole collapses. As one precaution, drilling is frequently carried out in stages. First, workers drill a pilot hole a few inches in diameter under the river, running the full length of the proposed hole. A "washover pipe," the same diameter as the borehole, is then fed down over the drill pipe behind the assembly. This pipe protects the soft borehole wall from abrasion created by the rotating drill pipe and from erosion caused by the returning flow of the drilling mud. Once the drilling assembly has reached the surface on the far side of the crossing, workers remove the bit and attach a reaming head. A swivel links the reaming head to a pipeline on the other side. The swivel allows the head to rotate without forcing the pipeline to revolve as well. The rig then pulls the drill pipe, reaming head and pipeline back through the hole in the opposite direction to the starting side of the crossing, and the passage is complete.

Directional drilling has proved useful in efforts to stop pollution. Leaky storage tanks and other industrial installations often pollute the ground underneath them. Such sites can be difficult or impossible to approach from above. Directional drilling offers a way to diagnose and clean up such areas. Moreover, because the rig and crew need not be directly over the site, or even nearby, they run less risk of being contaminated. At the Department of Energy's Savannah River Site, for example, a cleanup crew used a pair of horizontal boreholes to remove volatile, organic contaminants from around a leaking sewer line. The workers pumped air down one hole into the ground and then extracted vapor from the other hole. The team reported that this approach saved some $125 million in comparison with an alternative scheme that involved digging a series of vertical wells.

We are developing another interesting technique in my laboratory at the University of California at Berkeley. In this method, very low temperature air or nitrogen serves as the drilling fluid rather than mud. The approach has several advantages. First, under some circumstances even water can upset the composition of underground pollutants or wash them into previously uncontaminated regions. Gas, on the other hand,



EXTENDED REACH OPERATIONS use directional drilling rigs to place wells in an offshore oil field. A directed well has a much longer reach (*green*) into the oil zone than does a vertical well (*blue*). Therefore, such operations need fewer expensive, offshore platforms to tap the reservoir. These costly, offshore platforms can be avoided altogether if the crew can drill wells that are long enough to reach the reservoir from land-based drilling sites (*red*).



RIVER-CROSSING PROJECTS lay heavy pipelines under a river in two stages: First, crew members drill a narrow hole from one side of the river to the other (*top inset*). When the thin drill string reaches the far side, they attach it to a reaming head, which is connected to the wider pipeline (*bottom inset*). Then they pull the entire assembly through the hole in the direction opposite to the starting side. This two-step procedure reduces the risk of a borehole collapse.

is much less invasive. In addition, because the gas has a low temperature, pollutants freeze in place and can be sampled more accurately. Furthermore, the annulus of frozen ground supports the borehole and prevents it from collapsing. This effect is particularly strong in loose, sandy soils where pollutants are commonly found. Directional drilling and ground freezing may find other interesting applications. For example, cryogenic barriers might be created in order to block the spread of pollutants in the subsoil.

Directional drilling technology has also given contractors a new way to bring domestic services, such as cable, gas and water lines, from the street to consumers in cities. Traditionally, workers bury the lines in shallow trenches, the digging of which disrupts roads, sidewalks and the normal flow of traffic. In contrast, directional drilling has no effect on the surface. One directional drilling service uses a novel percussive device that hammers the drilling head forward. The head has an inclined face that digs a hole in the direction of the slant. Behind the head, spiral fins rotate the hammer body as it moves forward through the earth. The drilling head may be left free to turn on its own, or it can be clamped down so that it rotates on the hammer body. In the former case, the head holds a constant orientation and drills an inclined hole. If the head is clamped to the hammer body, it drills a straight hole. An operator walks over the ground above the hole holding a sensing unit. This device can be used to detect the position of the drilling head and to control its connection to the hammer body. In this way, the driller can steer the direction of the hole foot by foot.

Another device developed to create conduits for service lines uses jet nozzles that shoot liquid whose pressure reaches 250 bar (3,750 pounds per square inch). The instrument fires a stream of either fresh water or drilling mud. The mud coats and seals the borehole wall, an effect that reduces the risk of a collapse. Water and mud jets at this pressure penetrate most soils but leave concrete, pipes and cables intact. As a result, the jet-drilling method runs little risk of damaging other underground service lines, especially in crowded environments. The head can be equipped with tungsten carbide teeth to supplement the force of the jets if particularly solid terrain lies ahead. The drill bit is steered by angling the jet nozzles; a portable detector carried above the hole determines the position of the tool. The detector can sense and direct the position of the head at depths as great as 30 feet. Because workers can steer the drill at this depth, they can lay new lines well below a network of existing services at a cost comparable to that of boring a shallow hole.

In the near future we can expect to see systems in which a driller may steer the bit in real time, without having to survey what lies below beforehand. Already, though, directional drilling capabilities have changed the oil industry dramatically. In California, for example, industry leaders, local governments and environmentalists alike support a recent proposal to launch an extended reach operation that would rely on directional drilling equipment. The plan is to demolish an existing platform that stands off the coast of Santa Barbara, Calif., and tap the field it reaches, the South Elwood field, from land-based drilling sites instead. Oil companies could save money on this project because their wells will be more productive. In addition, they will need no platforms, tankers or service vessels. Their efforts will bring added revenue to the local community. Furthermore, extended reach operations run less risk of causing an offshore spill, generate less air pollution and leave lucrative fishing territories and marine sanctuaries unharmed. Not only are directional drilling techniques more efficient and less expensive for a range of uses, they are proving themselves safer as well.

FURTHER READING

DIRECTIONAL DRILLING. Society of Petroleum Engineers Reprint Series No. 30, 1990.

HORIZONTAL DRILLING. Society of Petroleum Engineers Reprint Series No. 33, 1991.

THE PRIZE: THE EPIC QUEST FOR OIL, MONEY AND POWER. Daniel Yergin. Simon & Schuster, 1993.

East Side Story: The Origin of Humankind

The Rift Valley in Africa holds the secret to the divergence of hominids from the great apes and to the emergence of human beings

by Yves Coppens

Humans are creatures whose roots lie in the animals. Accordingly, we find ourselves at the tip of one of the branches of an immense tree of life, a tree that has been developing and growing ever more diverse over a period of four billion years. From an evolutionary standpoint, it is important to locate the place and the time that our branch separated from the rest of the tree. It is these questions that the present article attempts to answer. When, where and why did the branch that led to us, the genus *Homo,* diverge from the branch that led to our closest cousin, the genus *Pan,* or the chimpanzee? Because this parting of the ways seems to unfold several million years before *Homo,* properly speaking, was born, the issue of our precise origin also needs to be addressed. When, where and why did *Homo* appear in the bosom of a family, Hominidae, that was well planted in its ecosystem and well adapted to its environment?

I first realized in 1981 that it might be possible to find answers to these questions. The occasion was an international conference in Paris organized by UNESCO to celebrate the 100-year anniversary of the birth of Pierre Teilhard de Chardin. As an invited speaker, I gave a talk on the French paleontologist and philosopher's scientific work. Although this aspect of Teilhard's writing is often forgotten by biographers, who are essentially interested in his philosophical texts, he produced more than 250 scientific reports over the course of 40 years. His opus includes articles on the structural geology of Jersey, Somalia, Ethiopia and China; on the Paleocene and Eocene mammals of Europe; on the Tertiary and Quaternary mammals of the Far East; on the fossil men of China and Java; on the southern African australopithecines (a kind of prehuman, one that was already hominid, but not yet *Homo*); as well as on the Paleolithic and Neolithic tools of all those countries.

A member of the audience, whom I did not know at the time, came up to me after my talk and congratulated me very courteously, admitting that he had not known about this technical aspect of Father Teilhard's work. He asked me several questions about this science of evolution that I practiced and about its state of development. My visitor ended this short interview with a precise question: Is there at present an important issue that is still being debated in your field?

Yes, I responded, there is a problem of chronology, as is often the case in historical sciences. Biochemists, struck by the great molecular proximity between humans and chimpanzees, place the beginning of the divergence of these two groups some three million years ago. This discipline also assigns a strictly African origin to humanity. In contrast, the field of paleontology describes a divergence that dates as far back as 15 million years ago. Paleontologists also postulate a broad origin, that is, one radiating from both the Asian and the African tropics.

The gentleman seemed interested, thanked me and left. Several months later I received a letter of invitation to a conference in Rome that he proposed to hold in May 1982. My questioner had been none other than Carlos Chagas, president of the Papal Academy of Sciences! In search of subjects that would have both current interest and important philosophical implications, he had considered what I had said and had organized, under the aegis of his institution, a confrontation between paleontologists and biochemists.

That meeting did take place and, although discreet, its influence on scientific thought was considerable. Two significant facts, one paleontological and one biochemical, were presented to the participants. The first was the announcement by David Pilbeam, professor of paleontology at Harvard University, that his research group had discovered, in the Upper Miocene levels of the Potwar Plateau in Pakistan, the first known face of a ramapithecid. This face resembles an orangutan's much more closely than it does a chimpanzee's face. Pilbeam's data were particularly important because the ramapithecids had for many years been considered by some paleoanthropologists to be the first members of the human family.

The second fact presented was a statement by Jerold M. Lowenstein of the University of California at San Francisco that active proteins had been discovered in the dental material of a ramapithecid. He had determined that activity by injecting extract from the ramapithecid teeth into a rabbit, where it brought on the formation of antibodies. Lowenstein then told us of the indisputable reaction of these antibodies

YVES COPPENS specializes in the study of human evolution and prehistory. He received his degrees from the Sorbonne, where he studied vertebrate and human paleontology. A member of many organizations, including the French Academy of Sciences and the National Academy of Medicine, Coppens is currently chair of paleoanthropology and prehistory at the College of France in Paris. He is also known for having done 20 years of extensive fieldwork in Africa, particularly in Chad and Ethiopia.

RIFT VALLEY cuts across eastern Africa from north to south, created by tectonic forces eight million years ago. The changing landscape and mountain boundaries divided an ancestor of ours into two groups. The western party thrived in forests and became our closest cousins, the chimpanzees. The eastern population evolved on the savanna and became human.

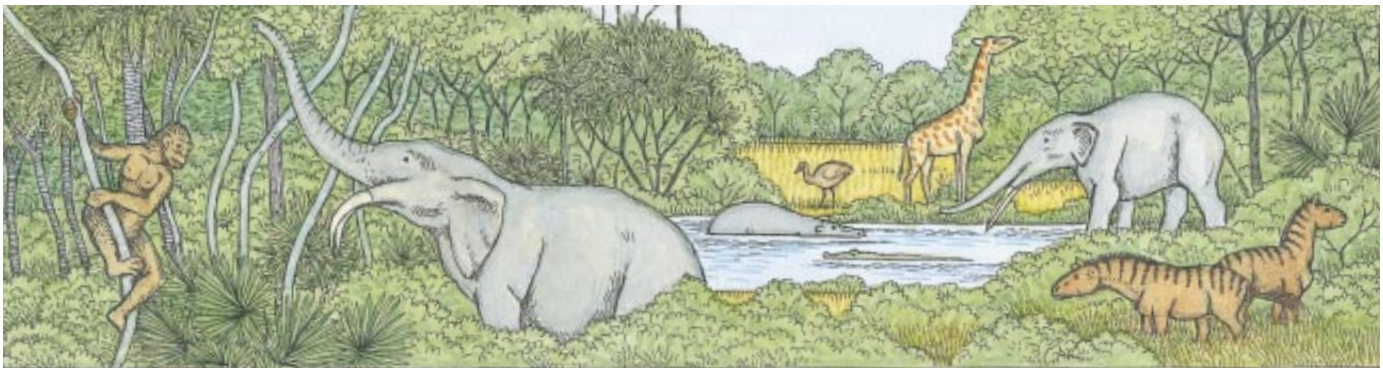
to the antigens of orangutans. This strong reaction made it clear that some of the ramapithecid proteins were still preserved and that the creature seemed related to orangutans.

Before the discovery of the ramapithecid face, scientists had procured only some of this genus's teeth and jaw fragments. Although these features were certainly interesting, it is necessary to know that all the bones of a skeleton do not carry information of equal value. These pieces were less significant than the orbit area and the nose and upper jaw region found in the new Pakistani piece. Paleontologists use such facial fossils to draw anatomical comparisons with similar or contemporary fossils. A simple comparison of the face of this ramapithecid, an orangutan and a chimpanzee clearly revealed the similarities between the ramapithecid and the orangutan.

Rather than comparing anatomical attributes, biochemists examine molecular details. They look at DNA, at the proteins and chromosomal maps of current species—elements that are not usually conserved in fossils. Their work helps paleontologists, who can then arrange species in order of complexity and compare their protein maps. The progression from simple to complex and the sequence that emerges reproduces, in some fashion, the evolution of creatures in the fossil record. In the case of the ramapithecid, however, biochemistry had made, as never before, a foray back in time by examining fossil proteins.

Circumstances had come together in such a way that we could finally put the ramapithecid in his place. This hominoid had been known to be Eurasiatic, and he remained so. Now that his relationship to the great ape of Asia, the orangutan, had been brought to light, the geographic picture became clear. Indeed, it made complete sense, as so often happens when one has found the solution to a problem. The origin of humanity, as the molecular biologists had suspected, appeared to be Africa, and Africa alone. The question of our family's place of birth seemed settled.

But the question of the date of this birth remained to be addressed. Several paleontologists present at this congress continued to defend the great antiquity of the hominids, whereas the molecular biologists extolled the extraordinary brevity of the independent part of our branch. The most generous of the paleontologists had arrived in Rome convinced of the 15-million-year history of our family. The most extreme of the molecular biologists were sure that three million years, at most, would measure the length of existence of the human family. Both sides came to the conclusion—made, of course, with only the most serious considerations possi-



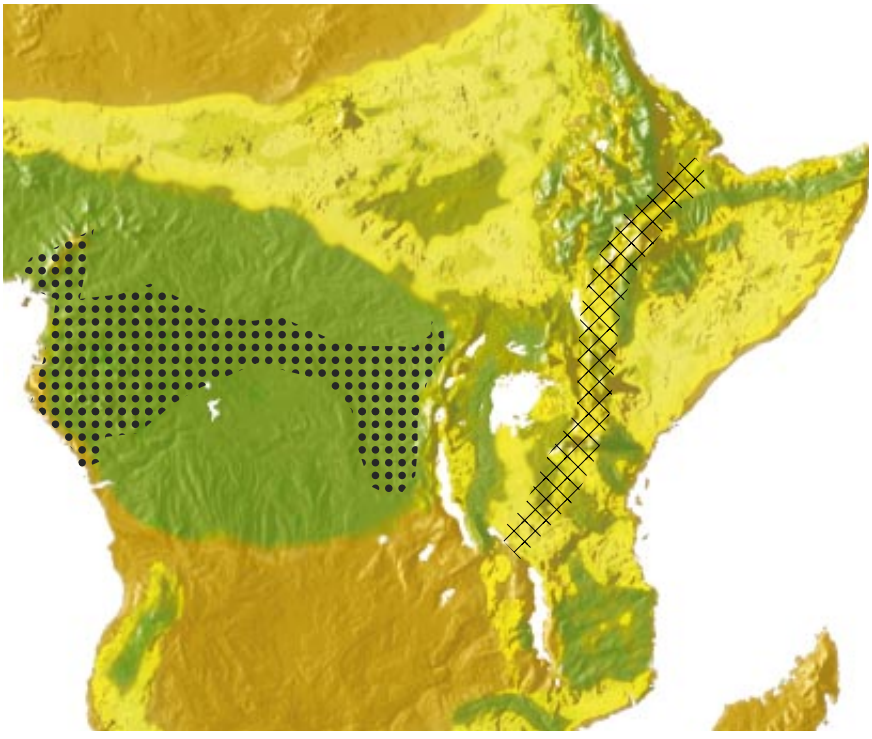
The Omo River Sequence

LATE MIOCENE — AROUND EIGHT MILLION YEARS AGO

ble—that seven and a half million years was a good span. I dubbed this conclusion "the prehistoric compromise."

The two paleontological and biochemical announcements of the Rome meeting were not the only crucial items that came to light in the early 1980s. Another set of results further clarified our understanding of human origins. Twenty years of excavations in eastern Africa (between 1960 and 1980) had finally yielded a mass of information in which could be sought evolutionary sequences and patterns. This extensive material had not been looked at in such a way before because it takes time to study and identify fossils. Its implications were vast, particularly when coupled with the information from the ramapithecid and the newfound consensus on dates.

The entry of paleoanthropologists into eastern Africa was actually an ancient affair. In 1935 Louis Leakey's expedition to Olduvai Gorge in Tanzania discovered remains attributed to *Homo erectus.* In 1939 the German team of Ludwig Kohl-Larsen found fossils that were named *Praeanthropus africanus*—later considered to be *Australopithecus*—near Lake Garusi, an area also called Laetoli, in Tanzania. In 1955 another Olduvai expedition led by Leakey revealed a single australopithecine tooth. These modest discoveries, however, did not command much interest.

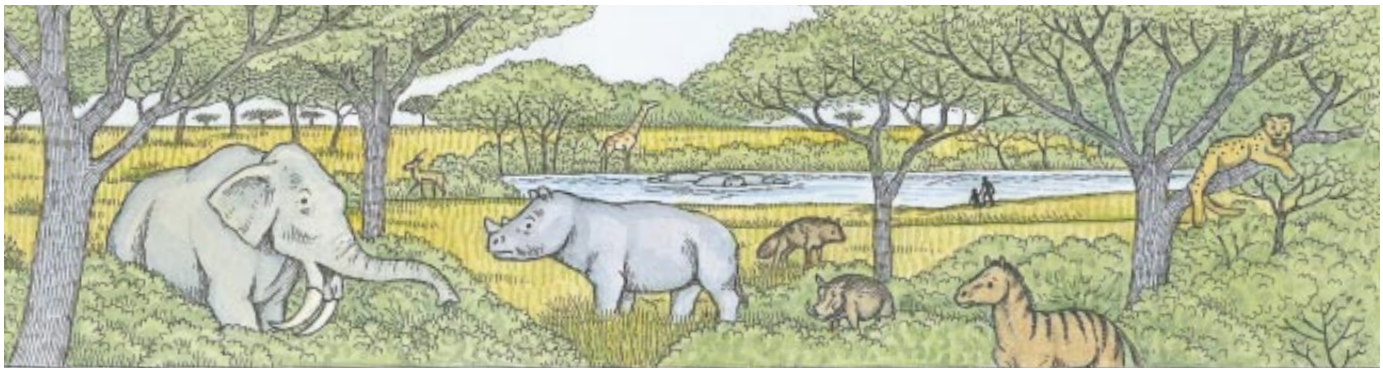


VEGETATION AND CLIMATE vary dramatically on either side of the Rift Valley: wet western woods (*dark green*) give way to eastern grasslands (*yellow*). Reflecting these ecological differences, which arose millions of years ago, chimpanzees are distributed only to the west (*stippling*), whereas hominid fossils are found only to the east (*cross-hatching*).

It was not until the 1960s that the world eagerly turned its attention to eastern Africa. In 1959 Mary Leakey found at Olduvai an australopithecine skull equipped with all its upper teeth. This skull could be absolutely dated to about two million years ago by the volcanic tuff below which it had been enveloped. The new hominid was named *Zinjanthropus;* it was a small-brained bipedal hominid species that went extinct about one million years ago. After that significant finding, expeditions started to arrive in abundance: a new team came each year for the first 12 years, and each one excavated for 10 or 20 seasons. Never before had such an effort been deployed by paleontologists or paleoanthropologists.

The results reflected the investment. Hundreds of thousands of fossils were discovered, of which about 2,000 were hominid remains. Yet, despite the constant work of preparation, analysis and identification of these fossils as they were unearthed, it is understandable that it was not until the 1980s that the first complete inventory of these thousands of finds was published. It is precisely this new information that, when added to the data received at the Rome conference, became essential to solving the mystery.

What emerged so clearly was that there was absolutely no sign of *Pan,* or one of its direct ancestors, in eastern Africa during the time of the australopithecines. Molecular biology, biochemistry and cytogenetics continued to demonstrate that humans and chimpanzees were molecularly extremely close, which meant, in evolutionary terms,



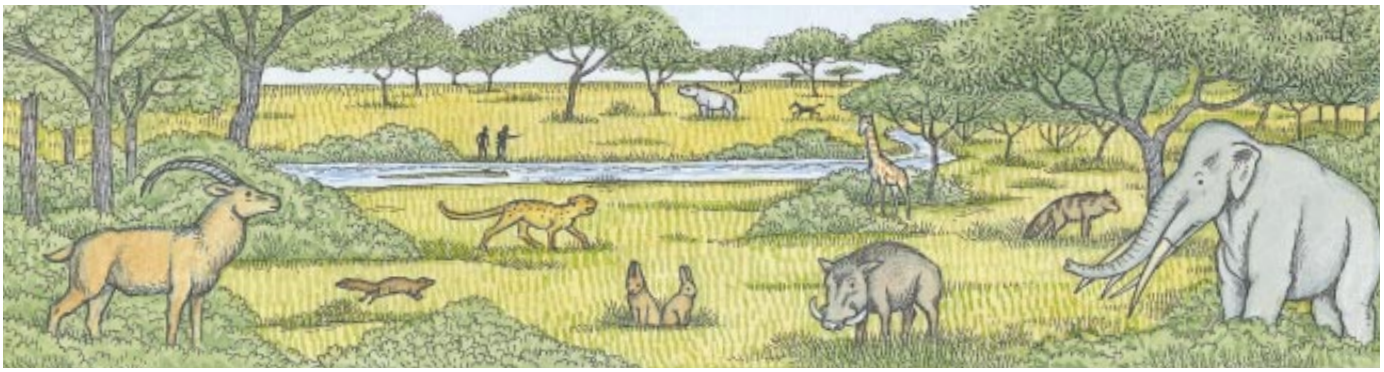
that they had shared a common ancestor not very far back in time, geologically speaking. And field-workers had just revealed that Hominidae, as of seven or eight million years ago, were present in Ethiopia, Kenya and Tanzania. But during the same period, this region had not seen the least sign of the family Panidae, no precursor of the chimpanzee and no precursor of the gorilla. Even though one cannot base a hypothesis on a lack of evidence, the striking absence of these Panidae where Hominidae were abundant represented a sufficient contrast to cause concern—all the more so because the 200,000 to 250,000 vertebrate fossils that had been collected constituted a statistical base with a certain authority.

I had been thinking about this puzzle during the conference in Rome. A quite simple explanation came to mind when I opened an atlas marking the distribution of vertebrates. The map devoted to chimpanzees and gorillas showed a significant group of territories, including all the large forested regions of tropical Africa, but stopped, almost without overflow, at the great furrow that cuts perpendicularly across the equator from north to south: the Rift Valley. All the hominid sites that dated to more than three million years ago were found, without exception, on the eastern side of this furrow. Only one solution could explain how, at one and the same time, Hominidae and Panidae were close in molecular terms but never side by side in the fossil record. Hominidae and Panidae had never been together.

I therefore suggested the following model. Before Hominidae and Panidae had separated, the Rift Valley did not constitute an irregularity sufficient to divide equatorial Africa. From the Atlantic to the Indian Ocean, the African continent constituted one homogeneous biogeographical province in which the common ancestors of the future Hominidae and Panidae lived. Then, about eight million years ago, a tectonic crisis arose that entailed two distinct movements: sinking produced the Rift Valley, and rising gave birth to the line of peaks forming the western rim of the valley.

The breach and the barrier obviously disturbed the circulation of air. The air masses of the west maintained, thanks to the Atlantic, a generous amount of precipitation. Those of the east, coming into collision with the barrier of the western rim of the Tibetan plateau, which also was rising, became organized into a seasonal system, today called the monsoon. Thus, the original extensive region was divided into two, each possessed of a different climate and vegetation. The west remained humid; the east became ever less so. The west kept its forests and its woodlands;

COMPARISON OF THREE HOMINOID SKULLS illustrates the proximity between two of the creatures. The ramapithecid (*center*) found in Pakistan resembles the great ape of Asia, the orangutan (*left*), much more closely than it does one of the African apes, the chimpanzee (*right*). Indeed, this very comparison led paleontologists to reject the Eurasiatic ramapithecids as close ancestors of humans and to focus on an African origin.



the east evolved into open savanna.

By force of circumstance, the population of the common ancestor of the Hominidae and the Panidae families also found itself divided. A large western population existed, as did a smaller eastern one. It is extremely tempting to imagine that we have here, quite simply, the reason for the divergence. The western descendants of these common ancestors pursued their adaptation to life in a humid, arboreal milieu: these are the Panidae. The eastern descendants of these same common ancestors, in contrast, invented a completely new repertoire in order to adapt to their new life in an open environment: these are the Hominidae.

This uncomplicated model has the advantage of explaining why Hominidae and Panidae are so close in a genetic sense and yet never together geographically. It also has the advantage of offering, by means of a situation that is at first tectonic and then ecological, a variant of the situation found on islands. Compared to complex solutions about the movements of Hominidae from the forest to the savanna or about the movements of Panidae from the savanna to the forest, the Rift Valley theory is quite straightforward.

It was only later, when I was reading the work of geophysicists, that I learned that the activity of the Rift Valley some eight million years ago was well known. Reading the studies of paleoclimatologists fortified me with the knowledge that the progressive desiccation of eastern Africa was also a well-known event, whose starting point had been placed at about eight million years ago. Finally, reading the declarations of paleontologists further reassured me, because they placed the emergence of eastern African animal life—a fauna labeled Ethiopian, to which the australopithecines belong—at about eight or 10 million years ago. Each discipline knew this date and in one way or another was familiar with the event or its consequences, but no interdisciplinary effort had brought them all into a synthesis. Adrian Kortlandt, a famous ethologist from the University of Amsterdam, had thought about such a possible scenario, but without any paleontological support, some years before.

The hypothesis lacked only a name. Three years later I was invited by the American Museum of Natural History in New York City to present the 55th James Arthur Lecture on the Evolution of the Human Brain. I also assumed a visiting professorship at the Mount Sinai School of Medicine of the City University of New York. The idea of giving this model a title that would be easy to remember and that would honor my hosts came to me then. I called it the East Side Story.

It is possible that the East Side Story has answered the first volley of questions: the when, where and why of our divergence from Panidae. Our phyletic branch, the one that now bears us, was marked off from the rest of the genealogical tree of living creatures eight million years ago in eastern Africa by reason of geographic isolation. The need for adaptation to the new habitat of the savanna, one that was drier and more bare than the preceding one, promoted further genetic divergence.

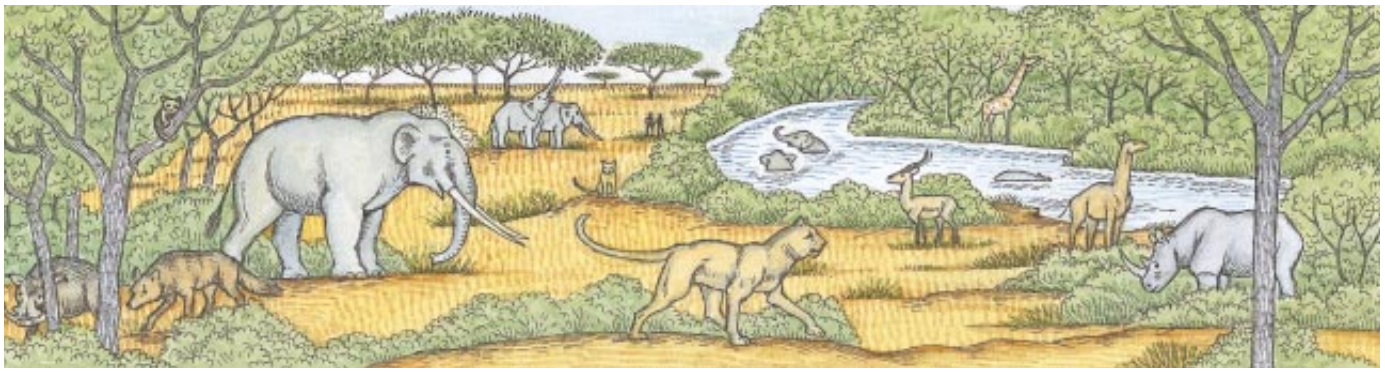
The second series of questions is more intricate: the when, where and why of the appearance of the genus *Homo* in the family Hominidae. The past eight million years during which our branch of the tree has grown have revealed themselves to be more complex than one might have imagined. The story begins with the diversification of a subfamily, the australopithecines. These creatures made very modest movements from eastern Africa to southern Africa. The story then continues from about three million years ago to today, with the emergence of another subfamily, the hominines. The hominines moved extensively, from eastern Africa across the entire planet. The last of the australopithecines coexisted for about two million years with the first of these hominines, which have only one genus, *Homo.*

The emergence of this hominine subfamily can be seen in a remarkable series of geologic beds and fossils found along the banks of the Omo River in Ethiopia. And, not surprisingly, because this is the second part of the East Side Story, the role of climate proves to be as powerful a force for change three million years ago as it did eight million years ago.

The Omo River tale began at the turn of this century, when a French geographic expedition proposed to cross Africa diagonally, from the Red Sea to the Atlantic. The Viscount du Bourg de Bozas directed the expedition. Having departed from Djibouti in 1901, the exploration was to end dramatically in the death of its leader from malaria on the banks of the Congo. The team nonetheless brought back from the journey, which followed the original itinerary, a fine harvest of fossils. Among the collection was a group of vertebrate remains gathered in what was then Abyssinia, on the eastern bank of the lower valley of the Omo River. The Omo lies on the eastern side of the Rift Valley.

Intrigued by this yield, which was described in two or three articles and in Émile Haug's geologic treatise in 1911, Camille Arambourg decided at the beginning of the 1930s to conduct a new expedition. Arambourg, future professor of paleontology at the National Museum of Natural History in Paris, reached the Omo and stayed eight months in 1932. He returned to Paris with four tons of vertebrate fossils.

The next major operation—the Omo



Research Expedition—was undertaken between 1967 and 1977. It was catalyzed, in part, by the bone rush of the 1960s and 1970s, described earlier, which had followed the 1959 find by Mary Leakey at Olduvai. A series of researchers conducted the 10-year Omo expedition in stages. In 1967 Arambourg and I worked on the site with Louis and Richard Leakey and Francis Clark Howell. Between 1968 and 1969 Richard Leakey left the expedition, and Arambourg, Howell and I continued the work. Finally, from 1970 until 1976, Howell and I dug there alone (Arambourg died in 1969).

From the very first expedition, the stratigraphy of this site was eminently visible, a superb column more than 1,000 meters deep. The fauna contained in these beds appeared to change so markedly as it progressed from base to summit that the site was obviously capable, even at mere glance, of telling a story. When dating by potassium-argon and by paleomagnetism finally became available, so that a chronological grid could be placed on this sequence, the history became clear.

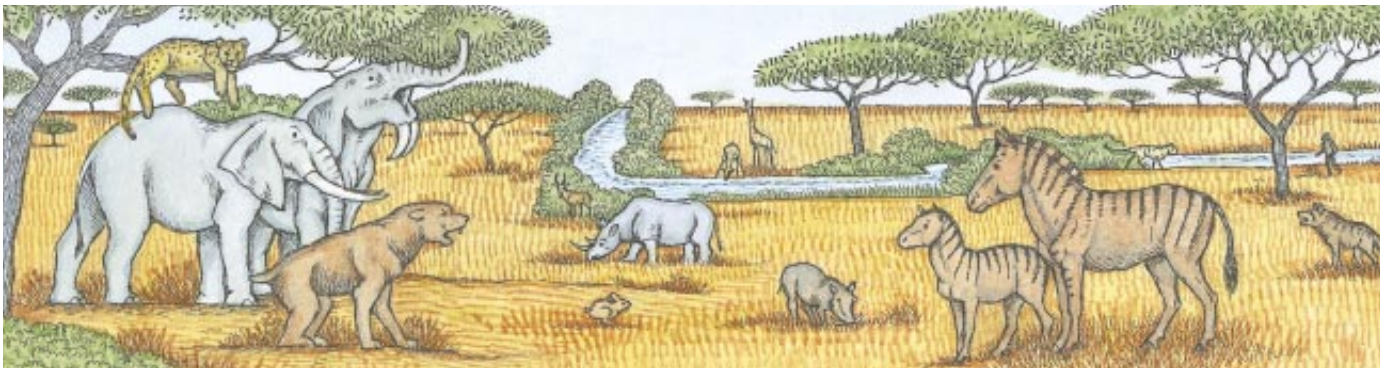
Starting four million years ago (the age of the oldest Omo level, the Mursi formation) and ending one million years ago (the age of the most recent level, the top of the Shungura formation), the climate had clearly changed from humid to distinctly less humid. As a consequence, the vegetation had evolved from plants adapted to humidity to those capable of thriving in a drier climate. The fauna had also changed from one suited to a brushwood assemblage to one characteristic of a grassy savanna. And the Hominidae, subject like the other vertebrates to these climate fluctuations, had changed from so-called gracile australopithecines to robust australopithecines and, ultimately, to humans.

In 1975 I informed the international paleontological community of this clear correlation between the evolution of the climate and the evolution of the hominines. I did so in a note to the *Proceedings of the Academy of Sciences* in Paris and in a communication to a congress in London at the Royal Geological Society. The reaction was very skeptical.

Of all the great eastern African paleontological sites, the strata of Omo were the only ones that could have permitted such observations. This site alone offered a continuous sedimentary column that ran from four million years ago to one million years ago. It is precisely between three and two million years ago, or to be very exact between 3.3 to 2.4 million years ago, that the whole earth cooled and that eastern Africa became dry. (Laetoli and Hadar were too old, Olduvai was too young and East Turkana presented a stratigraphic gap at that point, so they could not offer the same demonstration.) We know this fact through several other tests conducted in various regions of the world.

This climatic crisis appears clearly in the fauna and flora records of the Omo sequence. By indexing, both qualitatively and quantitatively, the animals and plants gathered in the various levels, we can interpret the differences that emerge from these species, with regard to changes in the environment.

MARY AND LOUIS LEAKEY examine the *Zinjanthropus* skull and upper jaw at Olduvai Gorge in Tanzania in 1959. Their discovery of a hominid fossil at this site led to a bone rush: paleontologists flooded in, and hundreds of thousands of fossils were excavated in subsequent decades.

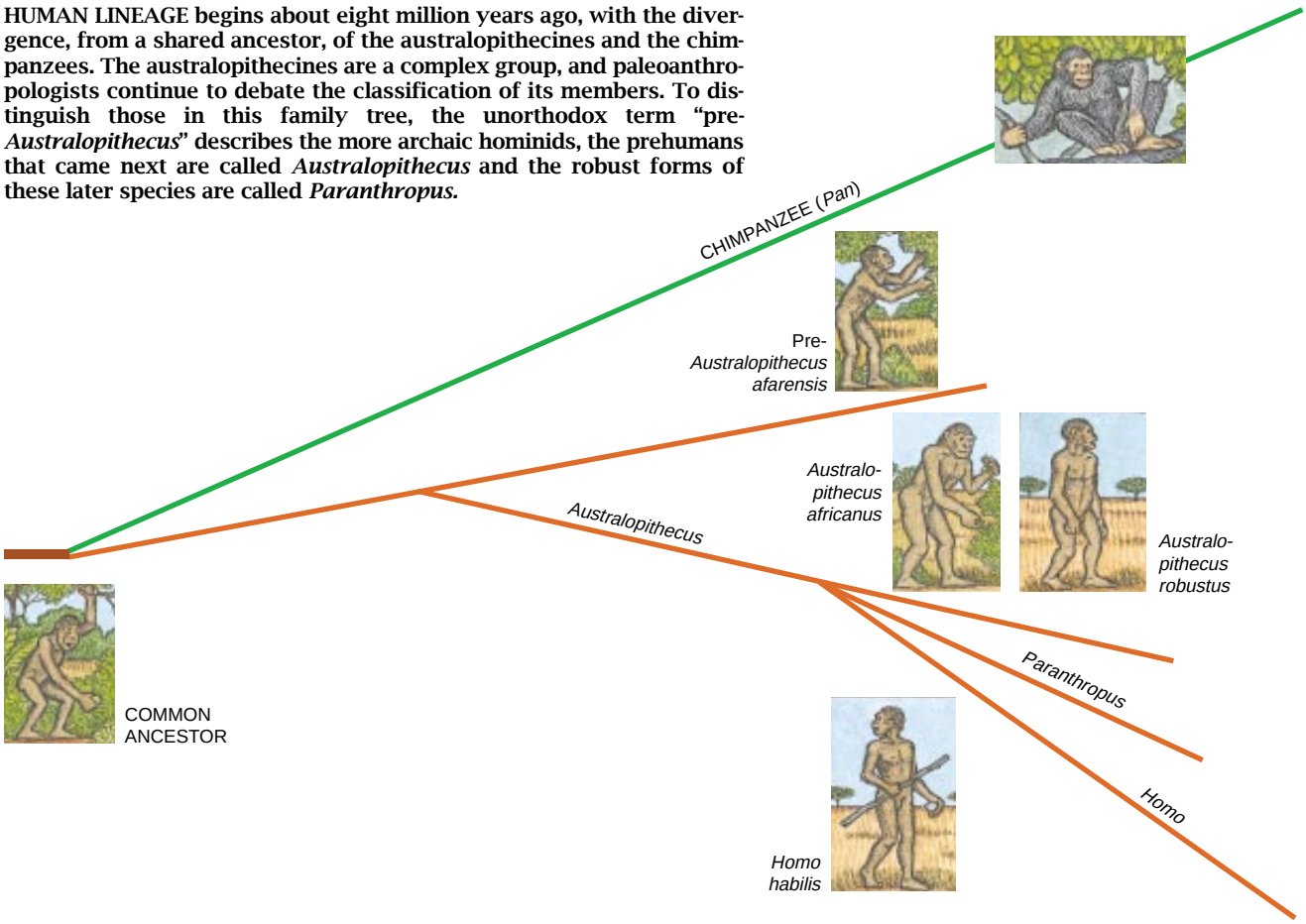


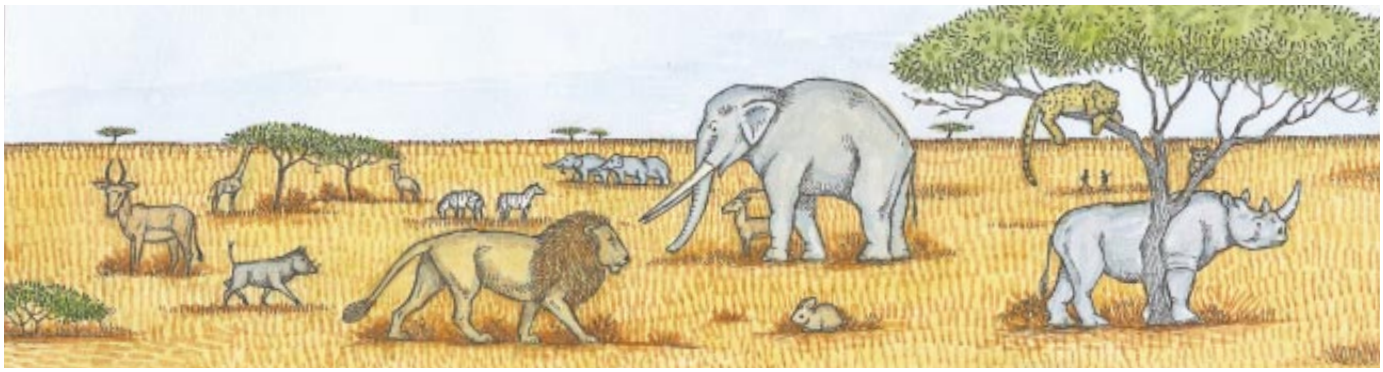
2.5 MILLION YEARS AGO — **UPPER SHUNGURIAN (UPPER PLIOCENE)** — 1.8 MILLION YEARS AGO

We know, for example, that the cheek teeth—that is, the premolars and molars—of herbivore vertebrates have a tendency to develop and become more complex when the diet becomes more grassy and less leafy. This change takes place because grass wears down the teeth more than leaves do. We know also that the locomotion of these same herbivores becomes more digitigrade in open habitats in which they are more vulnerable: one runs better on tiptoe than in boots. A certain number of anatomical features corresponding to very precise functions can also be good indicators: the tree-dwelling feet of some rodents or the feet of others that are adapted to digging. We use, with appropriate caution, of course, a method called actualist; in other words, we believe that the varieties of animals or plants we are considering acted then as they act today.

Many examples demonstrate this transition to a drier environment, and they are extraordinary in their agreement. As one moves from the older strata on

HUMAN LINEAGE begins about eight million years ago, with the divergence, from a shared ancestor, of the australopithecines and the chimpanzees. The australopithecines are a complex group, and paleoanthropologists continue to debate the classification of its members. To distinguish those in this family tree, the unorthodox term "pre-*Australopithecus*" describes the more archaic hominids, the prehumans that came next are called *Australopithecus* and the robust forms of these later species are called *Paranthropus*.





the bottom to the younger strata on the top, there is an increase in the hypsodonty—that is, in a tooth's height-to-width ratio—among Elephantidae (elephants close to the ones living in Asia today), Rhinocerotidae (specifically the white rhinoceros), *Hipparion* (ancestors of the horse), Hippopotamidae (precursors of the hippopotamus) and some pigs and antelopes. In other words, these groups exhibited the increasing complexity that we associate with a shift from a diet of leaves to a diet of grass. The Suidae, or precursors to swine, also show an increase in the number of cusps on their molars as they evolved.

On the lower strata are many antelopes—including Tragelaphinae and Reduncinae, which live among shrubs. All these creatures must have lived in an environment of wooded savanna close to water. On the top levels the true horse, *Equus,* appears, as do the high-toothed warthogs, *Phacochoerus* and *Stylochoerus.* We also see the development of the swift antelopes, *Megalotragus, Beatragus* and *Parmularius,* animals found on open grasslands.

On the bottom, three species of small *Galago,* or monkey, and the two Chiroptera, *Eidolon* and *Taphozous,* indicate a well-developed forest and a dense savanna. This conclusion is supported by the large number of Muridae rodents, such as *Mastomys,* as well as the rodents *Grammomys, Paraxerus, Thryonomys* and *Golunda*. At the top, the rodents *Aethomys, Thallomys, Coleura* and *Gerbillurus* in conjunction with *Jaculus* and *Heterocephalus,* the Chiroptera, and the *Lepus,* or hare, replace the previous inhabitants. All the later rodents inhabit dry savanna.

Pollen specimens on the bottom indicate 24 taxa of trees, whereas the top is characterized by 11. At the bottom, the ratio of pollens from trees to pollens from grasses equals 0.4. But at the top, it is less than 0.01. At the bottom, pollens from species that grow in humid conditions are abundant—they include *Celtis, Acalypha, Olea* and *Typha*. In the more recent strata, however, these pollens diminish considerably or even disappear from the record, whereas pollens from Myrica, a plant typical of dry climates, appear. The number of pollens transported by the wind, called allochtone pollens, dwindles from 21 percent at the bottom, where the forest edge is near the Omo River, to 2 percent at the top, where the Omo was low and the forest edge far away.

The story with the hominids is similar. They are clearly represented by *Australopithecus afarensis* on the lower strata. But the younger strata on the top reveal *A. aethiopicus, A. boisei* and *Homo habilis.* The oldest species of australopithecines, the graciles, are more ensconced in tree-filled habitats than are the more recent species, those called robust. As for humans, we are unquestionably a pure product of a certain aridity.

I called this climatic crisis "the (H)Omo event" using the simple play on words of Omo and *Homo,* because it permitted the emergence of humans—an event that affects us quite specifically—and because it was the Omo sequence that revealed it for the first time. Some years later the same data were reported from South Africa.

Thus, it appears strikingly clear that the history of the human family, like that of any other family of vertebrates, was born from one event, as it happens a tectonic one, and progressed under the pressure of another event, this one climatic.

These changes can be but quickly summarized here. Essentially, the first adaptation changed the structure of the brain but did not increase its volume, as suggested by the interpretation of endocasts, latex rubber casts of fossil skulls, done by Ralph L. Holloway of Columbia University. At the same time, the changes caused Hominidae to retain an upright stance as the most advantageous and to diversify the diet while keeping it essentially vegetarian. The second adaptation led in two directions: a strong physique and a narrow, specialized vegetarian diet for the large australopithecines and a large brain and a broad-ranging, opportunistic diet for humans.

Some hundreds of thousands of years later, it was the latter development that proved to be the more fruitful, and it is this one that prevailed. With a larger brain came a higher degree of reflection, a new curiosity. Accompanying the necessity of catching meat came greater mobility. For the first time in the history of the hominids, humanity spread out from its origin. And this mobility is the reason that in less than three million years, humanity has conquered this planet and begun the exploration of other worlds in the solar system.

FURTHER READING

EVOLUTION DES HOMINIDÉS ET DE LEUR ENVIRONNEMENT AU COURS DU PLIO-PLÉISTOCÈNE DANS LA BASSE VALLÉE DE L'OMO EN ETHIOPIE. Yves Coppens in *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences,* Vol. 281, Series D, pages 1693–1696; December 3, 1975.

EARLIEST MAN AND ENVIRONMENTS IN THE LAKE RUDOLF BASIN: STRATIGRAPHY, PALEOECOLOGY AND EVOLUTION. Edited by Yves Coppens, F. Clark Howell, Glynn Ll. Isaac and Richard E. F. Leakey. University of Chicago Press, 1976.

RECENT ADVANCES IN THE EVOLUTION OF PRIMATES. Edited by Carlos Chagas. Pontificia Academia Scientiarum, 1983.

L'ENVIRONNEMENT DES HOMINIDÉS AU PLIO-PLÉISTOCÈNE. Edited by Fondation Singer-Polignac. Masson, Paris, 1985.

Aging Airways

by Gary Stix, *staff writer*

The Sperry-Univac 8300s are as tall as Shaquille O'Neal and heavier than a Sumo wrestler. The ribbon cables that trail from their cabinets, the toggle switches and banks of lights make these metal hulks relics of another era, a time when the National Aeronautics and Space Administration was doing moon shots and the nation was preoccupied with antiwar protests, civil-rights marches and flower children. Although the computers were designed in the 1960s, the Federal Aviation Administration (FAA) has yet to pull the plug on them. Up to eight of the machines process radar data for display on the round green monitors at air-control centers that direct thousands of aircraft through the skies over the North American continent every day. Yet the computers share a paltry 256 kilobytes of total memory. If they were personal computers, they could not run Flight Simulator, a popular computer game for aviation aficionados that requires about eight times that memory capacity.

***DELAYED TAKEOFF** for deployment of new air-traffic equipment has meant that controllers at this Long Island control center must*

For 15 years, the FAA has been aware of the growing obsolescence of its air-traffic system—the world's largest—which handles about seven million flights a year. The agency is in the process of spending $36 billion for a top-to-bottom overhaul of the system, an effort scheduled to run through the year 2001. But costs may exceed budget by billions of dollars. And delays in the program threaten to deprive travelers and airlines of some of the benefits of the new system through the beginning of the next decade. "The new technology is so far down the road that it's not going to affect my career," says Christopher S. Boughn, a controller and union representative at a Long Island air-traffic center.

Even as the FAA wrestles with changing the hardware and software that controllers use to manage the nation's air traffic, the agency has been outpaced by users of the system. In the early 1980s airlines began to equip their craft with on-board processors that could calculate the fastest or the most fuel-efficient route. Such flexibility undermines the highwaylike structure of the existing system, which sends flights along rigidly fixed air corridors. As technologies continue to emerge, the effort to remake air-traffic control will become further complicated. The Navstar global positioning system

The FAA struggles to replace its winking, blinking air-traffic control equipment. But will the skies be

work with radar displays that have been tracking a growing traffic load in the New York metropolitan area for about 20 years.

(GPS) can determine the location of an aircraft, vehicle or ship to within 100 meters; additional instrumentation could better that number to only a few meters. These developments encourage a degree of independence for aircraft that emphasizes the antiquated technology under which the system operates.

The FAA's difficulties raise the question of whether decade-long planning cycles, ideal for building roads and bridges, can accommodate the rapid changes in computer and communications systems that are the foundations of air-traffic control. A panel appointed by the Clinton administration and Congress to study airline competitiveness last year concluded: "There is a fundamental inconsistency between the processes of government and the operation of a 24-hour-a-day, high-technology, capital-intensive air-traffic control system."

Heeding these concerns, administration officials recommended in January that air-traffic control be spun off from the FAA into an independent corporation, owned by the government, that would be exempt from the straitjacket of purchasing regulations, hiring rules and the federal budget process. "One could probably make an argument that we would have done a much better job of modernizing air-traffic control had air-traffic control been in a corporate environment," says FAA administrator David R. Hinson.

While the government, the airlines and electronics manufacturers grapple with the issues, the nation's nearly 18,000 air-traffic controllers orchestrate a flow of commercial air traffic that amounted to 478 billion passenger miles in 1992. The job begins before the aircraft leaves the gate. Before then, the airline has already filed with the FAA a flight plan, which includes time of departure, aircraft type and intended routes of flight. From a tower, one of almost 500 the FAA operates, a ground controller tells the pilot about runway for takeoff, radio frequency and other information. As the aircraft reaches the head of the runway, responsibility passes to another controller, who gives final permission for takeoff. Once aloft, the locus of air control changes.

In some airports the tower controller actually sends a paper flight-progress strip—a rectangular sheet of paper with information about a

given aircraft—down a tube to a departure controller who also uses a radarscope to track the aircraft in the vicinity of the airport. This room is one of about 200 terminal radar-approach control facilities, a name that is shortened to "TRACON" in airspeak. In some instances, TRACONs are in a separate building.

TRACON controllers are the traffic cops who direct airplanes during the initial leg of a journey, through what is known as the terminal airspace. The terminal airspace resembles a cylinder whose invisible walls rise 10,000 to 15,000 feet at a radius of 30 to 45 miles from the perimeter of the airport.

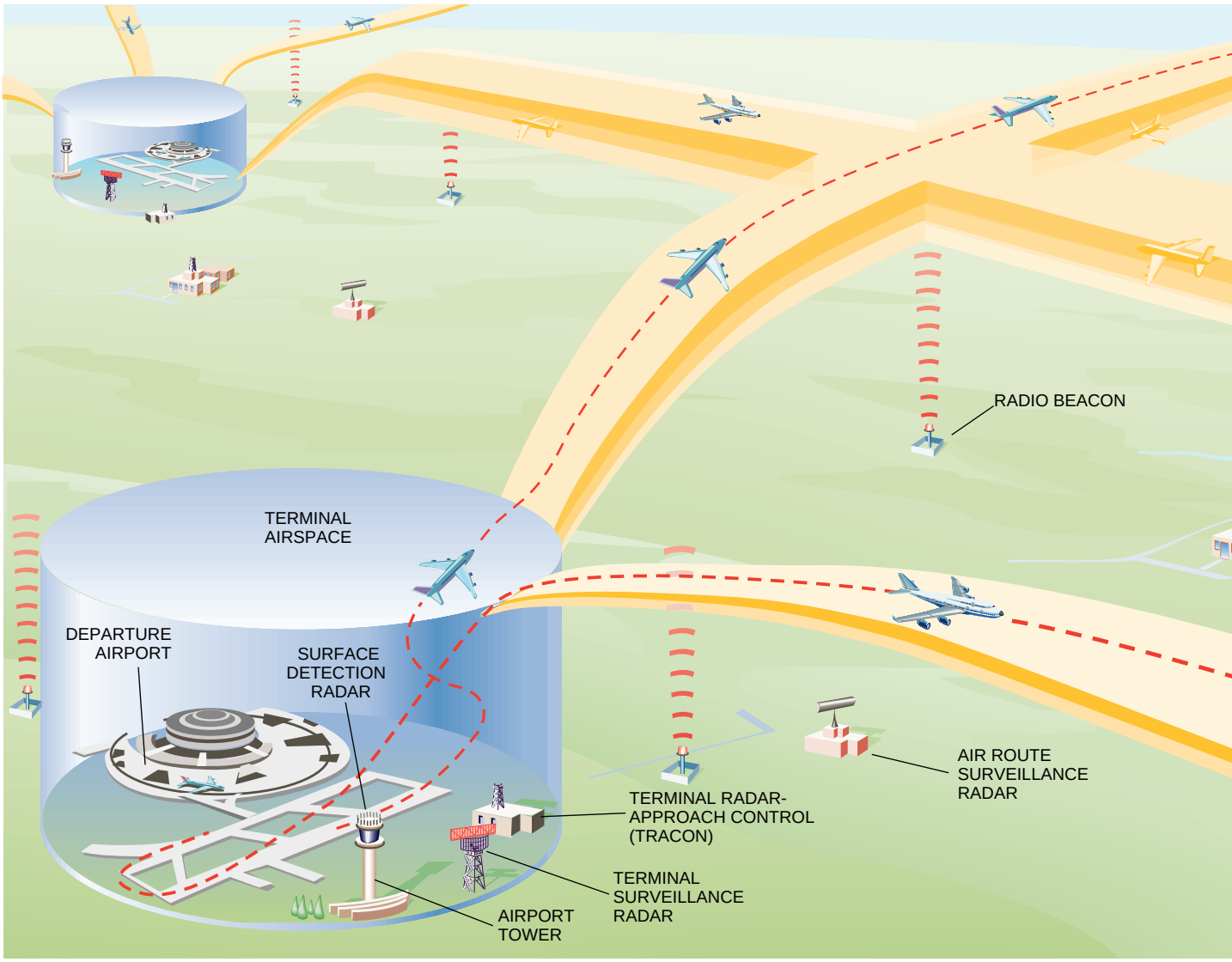
When the aircraft reaches the border of the terminal airspace, the TRACON hands off the airplane to a controller at one of the 22 nationwide air route traffic control centers, also known as en route centers. The centers guide airplanes along the network of airways. These skyways, which layer the nation from 18,000 to more than 40,000 feet, resemble an interstate highway system. They are used largely by commercial airlines and business jets. (There are lower airways, too, along which mostly commuter and private aircraft navigate.)

The roadways and intersections are defined by radio beacons—such as VHF omnidirectional range (VOR) transmitters—that emit signals in all directions. An aircraft's navigation system often takes its bearings from two or more VOR transmitters. During this segment of the journey, aircraft must maintain a horizontal separation of at least five miles and a vertical separation of 1,000 to 2,000 feet, depending on altitude. Control shifts from one en route center to the next as an airplane hurtles across the country until it reaches the airspace encircling a destination airport.

Air control relies on an intricate support network of communications, navigation and tracking technology. Different radar stations watch aircraft along each segment of the journey. These so-called primary radars determine the position of the aircraft and the course on which it is heading by bouncing a signal off its skin. Secondary radars that track the aircraft in flight consist of a ground-based transmitter that sends a signal to "interrogate" an instrument on the aircraft, a transponder that sends back altitude information and a numeric code that identifies the aircraft.

Computers, such as the Sperry mainframes in the TRACONs, process signals from these radars. The controller's screen displays the classic radar blip accompanied by a data block, a notation that shows speed, altitude and other data received from or computed from the radar inputs.

For all its quaintness, the system works extremely well, at least from the perspective of safety. Even if air travelers roll their eyes knowingly at one another when they hear "it will be just a few minutes until we're cleared so that we can pull back from the gate," they



know their chances of surviving a flight on a scheduled airliner are better than those they experience when riding in the family car. A major reason is the reliability of the air-traffic controllers and their winking, blinking aged hardware.

Major airlines have not had any passenger fatalities in about two years. Moreover, the number of near collisions has steadily decreased since the mid-1980s. The FAA has kept things running by instituting conservative spacing margins between aircraft and by retaining extensive backup systems. The burden of regulating the flow of air traffic eased during the early 1990s when the number of passengers dropped because of a languishing economy and the Persian Gulf War.

So, if it ain't broke—Twenty-year-old mainframes are as capable as souped-up engineering workstations of placing flight data on a radar display. A risk analyst, in fact, might rate their record as better than that of new, untested computers running software that gives controllers advice on how to direct aircraft. Yet the case that the FAA makes for rebuilding has some compelling features. In the early 1980s government officials, airline managers, pilots and controllers knew that these highways in the sky would eventually sprout potholes. Simple accounting extrapolations showed that the FAA's $200 million a year in capital spending would not support the explosion of air traffic that came about after airline deregulation in 1978. The emergence of the "hub and spoke" network—in which most flights are routed through a large regional city—increased traffic faster than controllers could be trained or equipment could be put in place.

Budget considerations and a crisis also argued for change. In 1982 FAA officials reasoned that automation of air-traffic control would allow the agency to cope with the personnel shortage that President Ronald Reagan created by firing 11,500 controllers.

Technological Artifacts

Equipment does not last forever—or does it? True, some navigation gear reportedly still shows the insignia of the Civil Aeronautics Administration, the FAA's predecessor before 1958. But keeping things running may not always be worth the trouble. Display terminals at the en route centers suffer 12,000 component failures every year, an average of one failure every five weeks for each display, according to the Department of Transportation. Computer capacity at TRACONs has been so taxed at times that the Government Accounting Office once warned that safety could be affected.

Maintenance is becoming a black art. Parts are scarce. Old equipment has to be cannibalized. Many of the technicians and support staff who were schooled in the subtleties of 1960s vintage computers are retiring and have not been replaced. "There are very few programmers still competent in Jovial," says Heinz Erzberger of the NASA Ames Research Center, referring to a virtually extinct programming language used on some of the air-traffic mainframes.

In 1982 the FAA started what was supposed to be an 11-year, $12.6-billion revamping of the entire air-traffic system. Since then, the agency has added about 150 projects that will bring overall costs to more than $36 billion through the year 2001. A General Accounting Office review noted that through the 1993 budget year only 54 of the total 238 projects within the capital program had been completed, a number that represents about 8 percent of the total estimated costs. "There is some reason to believe that this system, which was intended to bring 1960s technology into the 1980s, might not even be ready for the next century," concluded the federally appointed airline-industry study panel.

The centerpiece of the original renewal effort, according to agency officials, was—and is—the Advanced Automation System (AAS). It will replace the green glow of the round radar displays and rickety moon-shot-era computers that are often less powerful than the personal computers used by agency secretaries for word processing. Also on the capital spending list are solid-state digital radar units that would supplant tracking systems for the terminal airspace that still run on vacuum tubes. Improved secondary radar would give better tracking information and provide a two-way digital communications link between ground and air. An integrated voice communications system would replace the patchwork of telephone and radio links.

In some respects, the undertaking seemed relatively straightforward. The first stages of the AAS would replace the 25-year-old radarscopes with large-screen color workstations that would be connected by a local-area network to one another and to a mainframe computer. The network would process incoming radar and transponder signals for display on a controller's monitor.

Twelve years later much of the overall program has yet to be implemented. The agency did increase the capacity of computers in the radar rooms of the en route centers. But other large programs in the capital plan are three to five years

AIRPORT-TO-AIRPORT guidance of commercial airliners begins with tower controllers who give clearances for taxi and takeoff. Once an airplane has left the ground, responsibility for the aircraft passes to the terminal radar-approach control centers (TRACONs). The terminal airspace under TRACON control stretches in a radius of 30 to 45 miles from the airport. Air route traffic control centers, also called en route centers, conduct air traffic along multitiered airways in the sky, which, for commercial jets, usually extend from 18,000 feet upward. Aircraft must often adhere to this rigid route structure (*shown by the yellow tracks*) instead of flying along the fastest or most fuel-efficient path. Radar, including sensors to watch runway traffic, monitor each stage of flight.

behind schedule, and waiting for the AAS has itself proved costly. The FAA will have to spend at least $300 million to keep existing computer systems running.

The most glaring example of the problems that overtook the agency and its contractors can be found within the AAS effort. The network of workstations had to run 99.99999 percent of the time; it could be out of operation only for a minute and 20 seconds over a 20-year period. In addition, computers cannot simply be shut down overnight when bringing up a new air-traffic system: the cutover has been compared to doing brain surgery on a patient who is driving a car.

The AAS project was structured like a typical defense contract. It started in 1984 with a "bake-off" between Hughes Aircraft and IBM to come up with competing design prototypes. When the contract was awarded to IBM in 1988, Hughes filed a protest, which created several months of delay. The legal problems behind them, the FAA and IBM dropped into a serious pitfall, one likely to threaten the advancement of any software project: the propensity of the customer to change objectives and needs as work proceeds.

The FAA showered IBM with endless lists of system requirements. In the end, the documents towered to a height of 20 feet. "This was a textbook case of mistakes," says David Garlan, a professor of computer science at Carnegie Mellon University, who was a member of an outside team that evaluated the program. "They were pulling every trick in the book to meet the schedule and accommodate these requirements, and they got caught."

Major FAA Capital Projects

	SCHEDULE SLIPPAGE (YEARS)	ESTIMATES (MILLIONS OF DOLLARS)	CHANGE IN UNIT COST (PERCENT)
ADVANCED AUTOMATION SYSTEM Replaces computer hardware, software and workstations used by controllers in towers and radar facilities.	6–8	$6,500.00–7,300.00	––
AIR ROUTE SURVEILLANCE RADAR-4 Used for long-range tracking by air-traffic centers.	9	383.10	11
AIRPORT SURFACE DETECTION EQUIPMENT-3 Enables airports to monitor ground activity of aircraft even in rain and fog.	6	191.00	10
AIRPORT SURVEILLANCE RADAR-9 Replaces 96 radar units that use vacuum tubes with solid-state systems. Provides monitoring of aircraft within a 60-mile radius of airports.	4	839.90	(30)
AUTOMATED WEATHER OBSERVING SYSTEM Takes automatic readings on wind velocity, temperature, dew point, cloud height, visibility and precipitation and relays them with a synthesized voice to pilots.	3	229.90	35
MICROWAVE LANDING SYSTEM Provides an electronic guidance beam to aircraft on an approach. The FAA has delayed going to full production until the feasibility of using satellites for landing guidance has been determined.	12	740.80	––
MODE SELECT SECONDARY SURVEILLANCE RADAR Improves the tracking of signals from aircraft transponders, which provide altitude data and aircraft identification. Permits ground-to-air communications of air-traffic clearances and weather data.	7	473.20	40
TERMINAL DOPPLER WEATHER RADAR Detects wind shear wind shifts and precipitation around airports.	––	350.70	38
VOICE SWITCHING AND CONTROL SYSTEM Improves ground-to-air communications between controllers and pilots and voice and data linkages among air-traffic centers.	6	1,407.00	444

SOURCE: General Accounting Office's report "Air Traffic Control: Status of FAA's Modernization Program," April 1993, and House aviation subcommittee hearing report, March 3, 1992. Delay and cost figures are through the 1993 federal budget year, except for those for the Advanced Automation System and the Microwave Landing System, which are current.

Mesmerized by the flood of detail, both contractor and customer failed to pay attention to how a more automated system would change the way an individual controller performs. The AAS, in fact, was supposed to reduce the daily grind. The workstations would, the FAA hoped, do away with the antiquated paper flight-progress strips, which are a record of altitude, destination and other information a controller tracks. The computers were supposed to store an electronic database that would help automate the time-consuming task of annotating flight strips by hand.

What IBM devised, according to one controller, was a perfect electronic facsimile of the paper flight strip. The database contained all the information that can be listed on a flight strip—about 140 separate units of information, or fields, enough to keep the controller tapping on a keyboard all day. "The controller should be paying more attention to watching aircraft than to inputting information," says Mitch Coleman, a controller at the Atlanta en route center who was brought in a year ago to help streamline the system. Controllers suggested a version of the electronic flight strips with a more manageable 35 fields. The union to which Coleman belongs, the National Air Traffic Controllers Association, has cited the lack of controllers' full-time involvement early on in AAS development as a reason for these snags.

IBM's 500-member software engineering team has also run into problems with what is known as continuous operations. In the networks planned for en route centers, sometimes more than 200 workstations must talk to one another and to the mainframe that crunches incoming radar and transponder data. Any machine must be able to back up another when a processor fails. If the mainframe itself goes down, each workstation has to be able to handle information independently. IBM's software group has had difficulty reconciling inconsistencies in data between the workstations and the mainframe when the central computer is reactivated.

Continuous operations has helped contribute to the extraordinary cost problems and delays that the project has experienced. In announcing a shakeup of the AAS program in early March, Hinson presented results of an internal FAA report that showed that the $2.3 billion spent so far on the system could reach between $6.5 billion and $7.3 billion by the date of completion.

The direst total cost estimate is some $3 billion more than the agency's projection in 1988, when the contract was awarded to IBM. That amount is more

ESTIMATED TIME OF ARRIVAL of these new controller workstations for monitoring aircraft from 22 en route control centers is still in doubt—they may not be up and running before the turn of the decade.

than triple the original budget projected for the AAS by the FAA in 1983.

The report also noted that, yet again, the delivery date for the first network of workstations may slip. The Seattle en route center, which was originally to begin using the workstations in October 1996, may now get the new equipment sometime between mid-1997 to early in 1999. When it started planning for the AAS in 1983, the FAA had hoped to have the first workstations up and running in late 1990.

The report's cost estimates fail to give the whole picture because they do not include more than $400 million for AAS research and development. And the FAA has scaled back on what it expects to buy with this contract. It has decided against upgrading aging hardware and software in 108 of the 258 airport towers that were originally slated for a makeover under the AAS contract. "We keep getting less program for more money," says Robert E. Levin, an assistant director with the General Accounting Office.

In making the announcement, Hinson said he would appoint new program management, while suspending work on one part of the contract. The agency is also looking at technologies that might substitute for some of the hardware and software now slated for deployment under the current contract.

The AAS has experienced another surprise that might slow things further. In March, IBM completed the sale of its Federal Systems Company, the one handling the AAS, to defense contractor Loral for $1.58 billion. (The FAA still must approve transfer of the contract to Loral, however.)

Even when AAS hardware and software are finally in place, each unit will be little more than a glorified personal computer. "Controllers are getting a new color monitor," says Timothy Hancock, a controller with the Atlanta en route center. The agency does indeed have more ambitious plans for the AAS. The hardware and software should eventually let controllers direct air traffic in less rigid ways.

For that reason, failure to get the AAS on the air has proved particularly frustrating for the airlines. Since the early 1980s they have been buying aircraft with on-board computers and navigational systems that can calculate a route with the shortest time of flight—say, from Kansas City to Los Angeles—and even determine the approach to the airport that will require the least fuel. By enabling aircraft to deviate from the track of a roadway in the sky, the onboard equipment could save the airlines more than $2 billion every year by reducing fuel use, shortening delays and improving operating efficiencies.

But because the AAS is not fully developed, the air-traffic system operates very much as it always has. Consequently, the airlines can use the flight-management computers to travel on direct routes only about 10 percent of the time. "We can't take advantage of a lot of the technology because of the antiquated airspace management system," says Raymond J. Hilton, director of air-traffic management for the Air Transport Association, a trade group that represents the airline industry.

Software Traffic Cops

Indeed, some researchers have developed, and begun to test, software that does not have to rely on the AAS hardware platform. Since the mid-1980s Erzberger of the NASA Ames Research Center has been designing software to steer aircraft through traffic jams that arise at the hub airports. The software, called the Center-TRACON Automation System (CTAS), can now run on a network of ordinary engineering workstations. The network can be linked to the 20-year-old displays still in use in some radar rooms. "I'm fortunate to have been outside the FAA and not under their research people," Erzberger says.

Erzberger started his career designing algorithms for the cockpit flight-management computers. The CTAS is designed to mimic the calculations of these airborne systems. A software module acts as an "adviser" to a TRACON or an en route center, suggesting, as much as 45 minutes ahead of time, the best sequencing for aircraft converging on an airport.

The CTAS traffic-management adviser schedules the flow of aircraft with an expert-system-like database containing information about winds and an airline's preferred operating procedures and by taking into account instructions issued to pilots by individual controllers. Eventually it may be able to accept a suggested route sent down from an airplane's flight-management computer. It can group similar aircraft together—a DC-10 and a 747, for example. The biggest airliners cast off a turbulent wake that can endanger smaller craft, although airplanes of comparable size or weight can be more closely spaced. The CTAS is being tested at the Dallas/Fort Worth and the Denver airports.

Later this year the en route center that feeds traffic to the Denver TRACON will experiment with another component of the CTAS—a step toward controlling individual aircraft with the aid of a computer. This "descent adviser" will actually suggest to the controller how an aircraft should execute the early stages of an approach to an airport. It does so by comparing a radar-derived estimate of the airplane's time of arrival at the terminal airspace with a time scheduled by the traffic-management adviser for an optimal flow of traffic.

If the airplane is not projected to arrive on schedule, the descent adviser recommends turns or changes in speed to ensure that aircraft conform to the plan devised by the traffic-management adviser. The descent adviser can also look 10 or 15 minutes ahead to check for routing conflicts. Another module assists with the final approach to the airport and suggests the best runway for each flight. If the computer errs, there is still enough time for the controller to correct a computer instruction that brings two airplanes into dangerous proximity.

Besides helping controllers to become better crossing guards, these tools are supposed to reduce human error. Erzberger claims that the CTAS can raise the skills of an average controller to that of an expert and make up for a good controller's lapses. "You have controllers who are truly virtuosos," Erzberger says. "They operate at maximum efficiency, and when fully engaged they're like people in the Olympics. But nobody can do that day after day."

Before this vision materializes, the FAA will have to convince those who staff radar rooms that it is not trying to consign their work to a program on a floppy disk. Unfortunately for the agency, history tends to support such fears. In 1982, shortly after the controllers' strike, the agency sold its modernization plans to Congress on the promise that automation would allow the air-traffic system to operate with fewer people. Furthermore, early plans for the AAS called for software that would relay an instruction to an aircraft without first consulting the controller, a now dormant proposal. And despite union claims of a continuing deficit of 2,500 controllers, FAA administrator Hinson believes the new technology should let the air-traffic system operate with more or less the same 18,000-member workforce. (The union says only 14,400 controllers are employed full-time.)

Air controllers are not alone in their concerns. Experts who study the ways humans use machines worry about committing to software an occupation that has been compared to a cross between a video game and three-dimensional chess. "Air-traffic control looks like a simple set of rules that you could capture in a limited set of algorithms; it's only when you get into it that you realize it's not that simple at all," says V. David Hopkin, a human-factors specialist affiliated with Embry-Riddle Aeronautical University. "The more complex things get," Hopkin notes, "the more likely it is you will build in a complex interaction of things that might not be discovered for years."

Controllers may also be skeptical because automation has often created more problems than it has solved. The history of the Oceanic Display and Planning System (ODAPS) serves as an example. Budgeted at $12 million and not yet complete at a cost of more than $50 million, the ODAPS was supposed to depict air traffic moving over the oceans while identifying flight routes that might bring aircraft too close to one another. The system was to be a first cut at replacing a tracking system using flight-progress strips that has not changed much since the 1950s. But the ODAPS has encountered so many problems since development began in the mid-1980s that the FAA plans to scrap it gradually and try another approach.

Meanwhile, at the en route center in Ronkonkoma, N.Y., the ODAPS serves as a feeble backup system to the present outdated manual tracking operation. Air-traffic control over the ocean is hindered by a lack of radar beyond 200 miles from the coast. A pilot identifies the aircraft's position by talking over what is often a wall of static to a high-frequency radio operator located in a building a mile away from the center. The operator sends a teletype message to a controller. That information is then hand-annotated on a flight strip. In periods of intense sunspot activity, communications may be lost for hours at a time. Pilots bound for Europe are sometimes reduced to sending a message to an aircraft 100 miles or so behind that may still be able to get a call through to an operator.

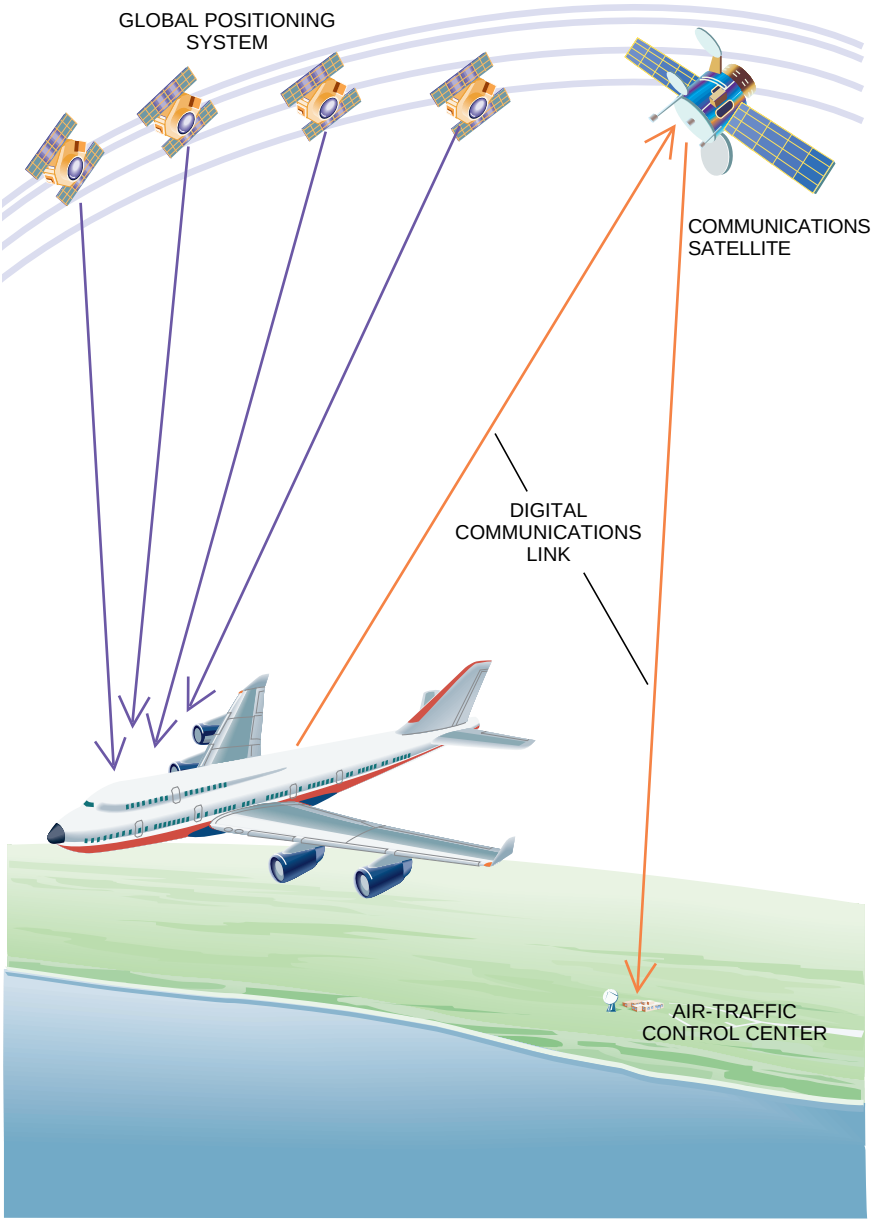
Back to the Present

Controllers use the information on flight strips—altitude, destination, flight path and type of aircraft—to visualize where each Boeing 747, DC-10 or Airbus is located in its procession across the Atlantic. A procedure used for helping to determine whether airplanes come too near is simple enough to please the most die-hard Luddite. First, plot the expected course of the airplanes by drawing lines with a grease pencil on the plastic that covers an aviator's map of the Atlantic. Then take a flight strip, curl its edge over at a 45-degree angle and place it on the map. If the angle of approach is less than 45 degrees, as measured by the paper marker, the separation rules for two airplanes converging from the same direction apply. If it is more, other regulations hold. "The technology we're using on the control room floor is a lot further behind than what we'd like to see," says Roger Kiely, a controller and union representative at the Ronkonkoma center. "Strips work fine," he adds, "but when you get 25 or 30 planes there's room for error."

Yet another wave of technological progress has begun to curl over the FAA's embattled technical staff. A combination of satellite navigation and digital communications has the potential

MANUAL LABOR is still the main method for keeping airplanes from coming too close over the oceans. Paper flight strips reveal the location of each airplane in the three-million-plus square miles of ocean for which the New York air route traffic control center is responsible.

SATELLITE TRACKING of airplanes, called automatic dependent surveillance, is carried out by first determining aircraft position. An airplane receiver gauges the travel time of signals beamed from four or more of 24 satellites orbiting the earth 11,000 miles aloft—known as the Navstar global positioning system. Aircraft location can then be relayed to an air-traffic control center by a communications satellite.



to become the infrastructure for a standardized worldwide air-traffic control system. The system could even extend to developing nations while enhancing the landing capabilities of smaller airports in the U.S. For two decades, the armed forces have been developing the technology, known as the global positioning system. Desert Storm witnessed its first deployment in battle. Truckers, merchant seamen, recreational boaters and backpackers have joined the military as users of satellite guidance. Now it is commercial aviation's turn. The GPS "is arguably the most important advance in the history of navigation," says Hinson of the FAA.

The GPS uses signals from 24 satellites orbiting 11,000 miles above the earth. By measuring the travel time from four or more of the 24 satellites, an aircraft can determine its horizontal position to within 100 meters and altitude to within 140 meters. Supplemented with a digital communications link, either through a communications satellite or a direct air-to-ground connection, the GPS can relay the exact location of the aircraft anywhere from Kansas City to Buenos Aires to a controller.

Combining satellite navigation and digital communications, one form of a technique termed automatic dependent surveillance (ADS) could supplement or even replace ground-based radar and radio-navigation beacons. Long-suffering ocean air controllers would not have to wait for a call to get already dated information from a radio operator to know where airplanes are over the Atlantic or Pacific.

If controllers knew exactly where every craft was at every moment, conservative 60-mile lateral distances that have to be maintained between airplanes over the oceans because of the absence of radar could be reduced. Traffic flow on the increasingly congested transoceanic flight corridors could be accelerated. The precision of satellite positioning might also allow reduced aircraft spacings over the continental U.S.

The FAA has evolved into a GPS advocate. In September it demonstrated GPS guidance for a poor-weather landing approach at Washington's National Airport. Implementation of the GPS may proceed more smoothly than has that of the AAS. Most of the satellites needed are already aloft, and private companies have brought to market the receivers for the satellite signals.

The FAA has begun to open the door to making the GPS the main navigation system used by pilots in commercial airliners. The agency has announced that it will permit airlines to use the GPS to supplement existing aircraft navigation equipment. The GPS may now be used for making some low-visibility landing approaches. Within a year or two, the FAA is considering letting aircraft cross the oceans without any navigation instruments other than a GPS receiver. This technology will eventually prove a boon at thousands of small airports that lack the ground-based instrumentation to guide pilots in poor weather conditions. Continental Airlines has already been using the GPS for reduced-visibility and nighttime landings on commuter flights into Steamboat Springs and Aspen in Colorado.

The FAA's stance toward the GPS hints that the agency may have learned some lessons from its mistakes. Chastened by the problems of implementing the AAS, the staff has embraced research programs jointly managed with industry. The scope of GPS programs under the FAA is less ambitious than was that of the AAS. "Before, things were orchestrated over the need to take great leaps in capability," says Martin Pozesky, associate administrator for system engineering and development

at the FAA. "Today we take smaller leaps, more incremental steps."

In the international air-travel sector the challenge for the FAA and other worldwide aviation authorities is less technical than institutional. Agreeing on standards and equipment will probably delay implementing a global air-traffic control system, perhaps for a decade or more. The U.S. has offered use of the GPS to other nations without charge, and the newly formed states that replaced the Soviet Union have done the same with their own network of satellites, the global orbiting navigation satellite system (GLONASS). Together the GPS and GLONASS could become a basis for an international navigation network, or a global navigation satellite system—the term of art used by international standards groups.

Europe, though, is ambivalent about heavy reliance on satellites owned by individual nations. It worries, too, about depending on U.S. satellites originally launched for military navigation. Members of a study panel for the European Union have emphasized the need for agreements that would ensure availability of the system for civilian users. "There's a fear that the U.S. might flip the switch on GPS," says Kevin Dopart, a senior analyst at the congressional Office of Technology Assessment.

Still, plans for an independently launched civilian constellation of satellites, at an expense of several billion dollars, remain largely speculative. What may also slow agreement are existing European plans to upgrade land-based tracking and navigation.

Unifying European air-traffic control requires rationalizing the kludge of 31 radar data processing systems, from 18 different manufacturers, that are controlled by 22 different kinds of operating system software. "The U.S. has the worst air-traffic control and technology management, but it's still better than all the others," Dopart says.

Meanwhile, back home, the FAA has still been subjected to intense scrutiny as an institution. The airlines, among other interested parties, believe the agency as structured is not nimble enough, despite the response it has brought to the GPS.

The FAA wholeheartedly embraced satellites only after a decade in which ground-based radar and navigation systems remained the focus of most of its research and development efforts. As little as two years ago, the agency had forwarded a proposal for the 1993 federal budget year calling for only $7 million for development of the GPS for low-visibility landings and other uses, despite an agency planning document's recommendation to allocate more than twice that amount of funding.

But the GPS, augmented with a ground- or satellite-based correction signal, could provide accuracies of seven meters or less for low-visibility approaches. If proved, this technique, called differential GPS, might even serve as a guide for an airplane's automatic landing equipment when airports are fogged in. This precision landing technology could make obsolete the more than $200 million in development costs the FAA spent over more than 10 years for a landing system that employs microwave transmitters next to a runway to help an airplane land.

Because of continuing doubts about the FAA's ability to manage technology programs, the airline-industry study panel recommended last summer that air-traffic control be removed from the agency. A similar spin-off proposal came from the National Performance Review, a group, led by Vice President Al Gore, that studies ways to streamline government bureaucracy.

All Aboard Airtrak

With a nod to both recommendations, the Secretary of Transportation Federico Peña proposed in January to vest a government-owned corporation with the air-traffic control operation. Proponents argue that such a corporation, Airtrak one might call it, would be exempt from the vicissitudes of federal budget-making and the thicket of procurement rules. Rather than haggle with Congress, the corporation could raise capital funds on the bond market. The discipline of trying to maintain a decent bond rating would ensure efficient management.

Last winter the administration was considering whether a corporation's operating expenses would be funded by passenger taxes, fees paid by aircraft owners for air-traffic services, dollars from the general treasury or a combination of these options. Private and business airplane owners opposed the plan because they fear airline domination of a corporate air-traffic system would inflict prohibitive fees on them.

The FAA, which now devotes two thirds of its $9-billion budget to air-traffic services, would probably remain responsible for safety regulations for the air-traffic network run by the corporation. In fact, some advocates of corporate air-traffic control point out that divesting these responsibilities would help the FAA avoid an inherent conflict of interest. It is difficult to maintain the necessary arms-length stance of the regulator when the agency also owns and manages the air-traffic control system.

A justification for Airtrak can be found as part of a worldwide trend toward privatizing air-traffic control. In places as far apart as Germany, South Africa, Switzerland and New Zealand, such corporations have been founded. Canada is also considering such a change for its air-traffic system. And there have been successes.

After New Zealand switched to a government-owned corporation, total annual operating costs dropped from $120 million (in New Zealand dollars) in 1988 to less than $80 million in 1993. Airways Corporation of New Zealand ran a total surplus of $50 million in its first four years of operation. "In fact, we believe we are probably one of the only air-traffic management systems in the world that earns more than it spends," Peter Proulx, the corporation's executive officer, told a panel of transportation specialists at a meeting in Washington, D.C., in early January. In Germany, which privatized its air-traffic system a year ago, pay incentives for air-traffic controllers have contributed to the drop in air delays.

Strong opposition in Congress clouds the prognosis for the proposal. Members may fear the loss of the opportunity to earmark funds for a home district. But the issue of public safety also cannot be ignored. James L. Oberstar, chair of a House subcommittee on aviation, which oversees the FAA, predicts that financially strapped airlines might pressure the board of directors of a corporation to reduce spacing between aircraft to increase capacity.

"To be blunt about it, when airline profit margins start to influence air-traffic control practices, the safety margin may be eroded," Oberstar said in a speech delivered in mid-January. Critics of Airtrak recall less stellar examples of monopolistic enterprises owned by the government—the U.S. Postal Service, for one. Oberstar does favor immediate efforts to bolster FAA management. The average tenure for an FAA administrator has been less than two years since 1987. Setting a fixed term of five years, Oberstar believes, would improve the agency's ability to manage long-term capital projects.

The ultimate answer remains opaque. The public policymaker must ensure the dual goals of efficiency and safety, as he or she makes decisions that involve buying and managing the high-technology systems on which people's lives depend. On time and in one piece. The institution that can meet both of these demands may not yet have been invented.

What's Bred in the Bone

Is a patent dispute slowing umbilical blood therapy?

MAX AGUILERA-HELLWEG

***UMBILICAL CORD BLOOD** is stored and frozen in an experimental bank established by Pablo Rubinstein of the New York Blood Center. Cord blood has already been used successfully to reconstitute the bone marrow of critically ill patients.*

The gospel of free enterprise holds that the opportunity to make a profit constitutes the most powerful stimulus for economic and social progress. Yet every rule has its exceptions. A case in point may be the controversy surrounding a promising technique for reconstituting the bone marrow of individuals whose ability to create blood cells has been destroyed by disease, radiation or chemotherapy. Every year an estimated 10,000 such patients die because a genetically compatible bone marrow donor cannot be found. The growing list of illnesses treated with bone marrow transplantation (mostly various forms of leukemia, lymphoma and anemia) amplifies the problem. Yet there is, some researchers believe, a bountiful, unexploited source of the crucial stem cells needed to establish functioning bone marrow. Preliminary research shows that umbilical cord blood can reconstitute a patient's ability to form a complete panoply of blood cells, including a functioning immune system. Cord blood has also been used as a vehicle for gene therapy.

The first transplant using cord blood was carried out in Paris and reported in 1989. Since then, 40 more cord blood transplants have been performed (all in children), according to John E. Wagner, Jr., of the University of Minnesota, a pioneer in the field. Most tissue has come from siblings, but two donors were unrelated to the recipients. Both of the unrelated donor transplants engrafted despite a mismatch of tissue type. Other recipients have done well even though there were multiple mismatches with the donor.

Hal E. Broxmeyer of Indiana University, who collaborated with Elaine Gluckman of the Hôpital Saint-Louis in Paris on the first umbilical cord blood transplant, comments that "many of us have a gut feeling that you're going to be able to cross some immunologic barriers with cord blood transplantation." If that feeling is correct, cord blood could be of particular help to African-Americans. They have a more difficult time finding compatible marrow donors than do Americans of European descent because there are few African-American donors.

Some hurdles still exist, says Paul R. McCurdy, head of the bone marrow transplantation branch of the National Heart, Lung and Blood Institute. One is the unanswered question of whether the amount of blood collected from an umbilical cord contains enough stem cells to regrow an adult's blood-forming system.

Another is of an economic nature. Research into transplantation of cord blood, according to several experts in the field, is being impeded by two U.S. patents that apparently assign all rights to the therapeutic use of the critical cells in frozen cord blood to one company, Biocyte in Stamford, Conn. "People are somewhat afraid to start collecting cord blood on a large scale for fear of the patent issue," remarks Wagner, an adviser to Biocyte.

Biocyte is riding the wave of enthusiasm for cord blood by offering parents the opportunity of banking their offspring's cord blood for 10 years. The blood is, however, reserved for the exclusive use of the donor. Biocyte's service costs $1,500, plus a $75 annual maintenance fee.

Biocyte's two patents, not its service, have provoked the controversy. The company's fundamental patent, issued in 1991, covers any use of cryopreserved fetal or neonatal stem cells for reconstituting the blood system or the immune system. That is an "unusually broad" scope, notes Pablo Rubinstein of the New York Blood Center, who was one of Biocyte's founders but is no longer affiliated with the firm. "There is no precedent for this," he adds. Rubinstein has established an experimental bank of cord blood and has stored 1,800 frozen samples. His bank was the source of the blood for the two unrelated transplants done so far.

Cryo-Cell International in Baldwin, N.Y., has already raised a challenge to Biocyte's patents. The company, which manufactures automated freeze-storage systems for medical specimens, has asked that the patent office reexamine Biocyte's fundamental patent, a request

that was approved last year. Daniel D. Richard, the president of Cryo-Cell, asserts that Biocyte's patents are unlikely to prevent his firm from selling equipment to store frozen cells from umbilical cord blood even if the patents survive reexamination. And Rubinstein, whose own experimental cord blood bank stands in an uneasy relation to Biocyte's patents, says he would put his money on Cryo-Cell in the dispute. Perhaps someday the patients, rather than the patents, will get everyone's attention. *—Tim Beardsley*

Devilish Details

Businesses and arms controllers square off on export restrictions

Export controls on dual-use technologies—equipment and materials that have both military and civilian applications—create tension between manufacturers keen to export to as many countries as possible and proponents of arms control. Arms-control advocates, who fear the spread of weapons of mass destruction, argue that export controls are the main reason that Saddam Hussein has been unable to build a nuclear bomb and that North Korea may yet be prevented from doing so. Industrial managers argue that unilateral controls do little to stop adversaries from obtaining strategic goods, because the products are often readily available from other countries.

Efforts to settle such questions reveal that the devil lurks in the details, and he is likely to get a thorough workout in the coming months. Congress will soon be taking up the Clinton administration's bill to establish an export-control regime to replace Cocom, the Coordinating Committee for Multilateral Export Controls, which lapsed at the end of March.

Last fall, an administration committee published a document on national export strategy that seemed to promise a radical overhaul. The U.S., the document said, would eliminate ineffective restrictions, create a more efficient control process and enhance the force of remaining controls by pursuing new multilateral approaches. In effect, the U.S. would try to establish Cocom-like groups whose members would reflect the changed geopolitical climate. Participants in such multilateral compacts—of which several examples already exist, geared to specific technologies—would be able to trade strategically sensitive products with greater freedom.

Industry cheered. Exporters have complained long and bitterly about the bureaucratic labyrinth they must negotiate in order to secure an export license for a sensitive item. Manufacturers also resent unilateral U.S. regulations that have prevented them from selling tractors and civilian aircraft to Iran, for example, while competitors overseas have taken the orders instead. Computer makers have fumed at pirated processors selling briskly in proscribed markets while the original product languishes under Cocom restrictions.

When the U.S. issued its legislative proposals late in the winter, however, the cheers turned to groans. The ad-

Cold Calculation

Catalytic converters have helped reduce automobile pollutants at the tailpipe by as much as 90 percent since they were first installed in American cars more than two decades ago. Yet the devices still do not work well enough for manufacturers to meet the low-emissions standards scheduled to come into effect in California—and possibly other states—in 1997. The law has created an opportunity, and Corning has responded. The company recently announced the introduction of an electrically heated catalytic converter that it says sharply reduces pollutants emitted by a car during its first few minutes after starting, before conventional converters begin working.

In catalytic converters, palladium, platinum or other rare metals catalyze reactions that remove hydrocarbons, sulfur compounds and other harmful substances from an automobile's exhaust. The trouble is that the catalysts must be hot to work. According to federal tests, cars generate some 80 percent of their total emissions in the brief period before the converter heats up.

FRANK J. BORKOWSKI *Foto Forum*

ELECTRICALLY HEATED catalytic converter developed by Corning cuts down on cold-start emissions.

"As we tried to squeeze emissions down further, it quickly became obvious that we needed to concentrate on the first minute or so of operation," says Douglas D. Teague of Chrysler Corporation. "Even if you get the emissions from a warm engine down to zero, you would still not meet the standards set by California."

Corning's converter consists of a honeycombed chamber lined with platinum, palladium and rhodium catalysts and a steel-alloy heating element that draws power from either the battery or the alternator. The converter reaches a temperature of 400 degrees Celsius—enough for the catalysts to take effect—in as little as five seconds, remarks Robert W. DeMallie, a Corning spokesperson.

At least two other companies are developing similar electrically heated devices: a division of W. R. Grace based in Hiram, Ohio, and NGK, a Japanese firm. Chrysler and other manufacturers are now testing the Corning converter in fleets of cars.

Other strategies have been considered. One involves heating the converter before the engine turns over; the disadvantage is that the driver would have to wait at least several seconds before the car starts. Another scheme in which gasoline is ignited at the mouth of the converter raises "safety issues," DeMallie comments.

Corning's converter may add no more than $200 to the cost of a car. Why didn't the automobile industry develop heated converters earlier? Because it didn't have to, DeMallie replies. "If it weren't for this [low-emission] legislation," he says, "this wouldn't be occurring."—*John Horgan*

ministration's bill does declare a preference for multilateral controls in place of unilateral U.S. actions. But the bill also explicitly provides for the continuation of unilateral controls. Furthermore, it gives the State Department leadership in resolving disputes between other agencies. As a result, exporters fear that controls inspired by foreign policy, as opposed to military ones, will flourish. "I think the State Department is driving this train," comments Howard Lewis III of the National Association of Manufacturers. Lewis runs a newly formed industry group, the Coalition for Export Control Reform, which has been lobbying aggressively to defeat the administration's proposals in Congress.

Lewis's coalition strongly favors multilateral controls because they would have to be agreeable to other countries and because they would permit exports to all countries in the control group. Although the coalition grudgingly admits the need for some unilateral controls, it wants "discipline" in their application. What industry would ideally like, Lewis says, is a list of which countries cannot receive which goods. He would also like the government to justify items on the list by specifying what military advantage the goods would confer.

By comparison, the government's bill represents something far closer to the status quo: "tinkering at the margins," according to Lewis. In addition to its other concerns, the coalition is apprehensive that licenses will still be subject to lengthy bureaucratic turf wars. The lobby also complains that reexports from one country to another will continue to need licenses, even when such shipments are made from countries that cooperate with the U.S.

William J. Perry, the secretary of defense, may have taken an early defeat in the skirmishing. Perry has frequently spoken out against unilateral export controls and in favor of the kind of multilateral regimes advocated by the Coalition for Export Control Reform. Perry was a member of a panel of the National Academy of Sciences that reported in 1991 on "six myths about the use of export controls for nonproliferation." The panel noted that the growing diffusion of high technology to countries throughout the world "will render inoperative any go-it-alone approach." Stemming the tide of proliferation "will demand a heretofore unknown level of multilateral cooperation and coordination," the panel said. Despite Perry's strong distaste for unilateral controls, it evidently failed to win enough support to dominate the administration's plans.

Some arms-control advocates are relieved that Perry was not more influential. "I have some serious concerns" about the administration's proposals, says Jon B. Wolfsthal of the Arms Control Association. Wolfsthal, taking a view exactly counter to that of Lewis, worries that the White House is putting too many restrictions on the president's ability to limit exports. "The U.S. has always taken a leadership position," Wolfsthal states. This bill "appears to embody a presumption to approve exports that may contribute to the development of weapons of mass destruction."

Gary Milhollin, a professor at the University of Wisconsin, who runs the Wisconsin Project on Nuclear Arms Control in Washington, D.C., warns that if the manufacturers' "unprecedented campaign" to weaken export administration legislation succeeds, "developing countries will find it easier to build atomic bombs and long-range missiles under Clinton than they did under either presidents Bush or Reagan." Milhollin argues that export controls are more important after the cold war than before—"the nuclear and missile race is still on between India and Pakistan and still on in the Middle East." And he maintains that the U.S. must control critical exports even if they are also available from foreign sources.

The stakes could hardly be higher. The nascent economic recovery in the U.S. will need growth in high-tech export markets if it is to maintain its momentum. Yet the continuing disintegration of the cold war world order is creating a fractallike replication of opportunities for nuclear, chemical and biological confrontation [see "When Treaties Fail," page 14]. Several competing bills, some mirroring the administration's cautious moves and others that would entail more drastic change, are starting to percolate through Congress. The process of reinventing export controls is only just beginning, and events may well outpace it. *—Tim Beardsley*

Office Romance

A Japanese builder spins dreams of a workplace utopia

Now on the drawing board of Hazama Corporation, one of Japan's largest construction firms, are plans for a building that will make the average salaryman, standing nose to nose with his counterpart on the 6:15 A.M. train from Saitama to Tokyo, even more eager than ever to get to the office. It might find favor with commuters on the 7:23 A.M. from Princeton Junction as well.

From the moment that an employee arrives at work, the building's distributed systems will recognize him, open doors for him, serve him refreshments and make sure that both temperature and humidity in the workplace are just the way he likes them. Callers will never be told, "She's away from her desk." Instead incoming calls will be routed to wherever the worker is in the building. If she is moving between two points, the nearest telephone will ring. Back at the worker's desk, a single instrument will serve as her telephone, fax, copier, word processor and data manager. And all this technology will be as friendly as it is powerful—designed for all employees, not just the technically proficient.

This is the TRON Hyper Intelligent Building, still in the theoretical stage but nearing the point at which its transformation into reality will begin. The energetic visionary behind the project is Ken Sakamura, associate professor of computer science at Tokyo University. Sakamura shared his concept with SCIENTIFIC AMERICAN at the inaugural meeting of the Association for Computing Machinery, held in early February at United Nations University in Tokyo.

Central to the idea of the Hyper Intelligent Building is what Sakamura calls a "handy terminal," a special identification card that includes a miniature microphone, speaker and liquid-crystal display. When the cardholder approaches an entry door, sensors will identify the individual, open the door if the person has access privileges for the area, turn on the lights and adjust local temperature to the individual's preset preferences. Any calls for the person will be directly routed to that location.

The first TRON Hyper Intelligent Building will be erected within three to four years. Others, Sakamura says, will quickly follow once the ground is broken. The concept provides a number of related benefits. Decentralized climate control makes workers more comfortable, but unlike most luxuries, it provides an increase in comfort at a lower cost because energy will no longer be wasted heating or cooling unoccupied areas.

The building will be the first application to an entire office environment of the TRON project's central idea of the Highly Functionally Distributed System. Computer intelligence is incorporated into a wide variety of everyday products, objects and systems. These then

become "intelligent objects," which collectively form organic systems designed to achieve shared values and objectives such as comfort, efficiency, safety, convenience or productivity. In a Hyper Intelligent Building, energy conservation will be a key objective.

Intelligent buildings are not new. Decentralized climate control, computer-readable identification cards and voice-recognition systems have all been used in buildings before. What is new about the approach is that all the systems used throughout the structure will be based on TRON specifications that have been developed over the past 10 years and that are completely open and available for use without licensing fees. TRON-specification floor-control processors installed on each floor of the building will be interconnected by a 400-megabits-per-second optical-fiber local-area network that, in effect, makes up the central nervous system of the structure.

Another planned innovation is the use of various office robots that will relieve workers of heavy lifting and other onerous tasks. Robots not only will file and retrieve materials but will track their use. Infrequently used files will be stored farther away from the workplace. After an extended period of disuse, the system will suggest off-site storage. The result, Sakamura says, will be work areas that are less cluttered and more efficient.

What does the future hold for the TRON Hyper Intelligent Building? Sakamura acknowledges that the first one will primarily be a showpiece that the world will come to inspect. He is convinced, however, that once this building has been erected, many others will follow. *—Robert Patton, Tokyo*

As Advertised

A Japanese commercial radiates charm; glowing claims for a beer

A new character has joined the ranks of Joe Camel, Beavis and Butt-head, Bart Simpson and other 'toons you love to hate. He is Mr. Pluto, a wide-eyed little fellow who wears a shiny, green helmet labeled Pu, the chemical tag for plutonium. Mr. Pluto stars in an 11-minute video developed last fall by the Japanese Power Reactor and Nuclear Fuel Corporation (PNC). The company hoped Mr. Pluto would calm the public's concerns over the plutonium fueling their prototype fast-breeder reactor, Monju, which was slated to reach criticality in April.

PNC's animated spokesman cheerfully assures his audience that "unlike radium, there is no proven case of plutonium in the body causing cancer." Moreover, he insists that terrorists could not build a working nuclear weapon from the reactor-grade plutonium used and produced by Monju. Mr. Pluto explains that even if someone drank water from a reservoir contaminated with plutonium, nearly all of the element would pass harmlessly through the body. Indeed, Mr. Pluto is shown holding hands with a trusting young buddy who drinks several mugs of plutonium-laced water. "There are many people out there who think of plutonium as a big, bad monster," Mr. Pluto frowns. "Do you have an image of *me* as something frightening?"

MAINICHI SHIMBUN *PNC*

MR. PLUTO reassures his friend, "Even if [plutonium] is drunk and enters the stomach and intestines, almost all of it is excreted, leaving the body."

U.S. Energy Secretary Hazel R. O'Leary does. In a letter dated February 7, she asked PNC president Takao Ishiwatari to consider withdrawing some 200 copies of the videotape that have been distributed near the Monju site. "As you know, plutonium is potentially dangerous, even in very small amounts, whether it is ingested or inhaled," she wrote. "Plutonium from reactors can also be used to make nuclear explosives."

The company answers critics by saying that the video contains no scientific errors. "They start with accurate facts," agrees Steven Dolley, research director at the Nuclear Control Institute (NCI), a Washington-based nonproliferation organization that sponsored a U.S. screening of the tape with Greenpeace in March. "But they jump to scientifically inaccurate and dangerous conclusions."

NCI president Paul Leventhal has dared Ishiwatari to take what Dolley calls their version of the Pepsi challenge. "Are you prepared to appear on national television and personally consume a gram of plutonium and run the risk of cancer from microgram quantities becoming lodged in your bone marrow?" Leventhal asked in a letter dated February 23. "If you are not, then you should withdraw the videotape showing a cartoon character drinking plutonium."

If Ishiwatari accepts, he does so aware that no antidote exists. Perhaps he should try strontium. After the Chernobyl accident, the Bulgarian defense department concocted Lulin Special Light Lager, a brew they claim contains a secret ingredient, called kantotonic, that removes strontium from the body. "That actually makes some sense," Dolley says, explaining that if one is exposed to radiation and has no internal injuries, the best thing to do first is drink a lot of fluids. Kantotonic remains in question, but drinking to your health may yet take on a whole new meaning. *—Kristin Leutwyler*

Lollipop, Lollipop

A candied sedative with a kick arouses opposition from doctors

Fentanyl is a powerful morphine substitute whose potency exceeds the real stuff 100-fold. The drug has won a place as an intravenous anesthetic for patients undergoing open-heart surgery. It is also popular on the street and with drug-abusing physicians. So what is it doing in the form of a lollipop/candied lozenge, called Fentanyl Oralet, for pediatric patients?

The answer, according to William C. Moeller, the president of Anesta, the firm that developed the pediatric preparation, is straightforward. Fentanyl Oralet is the only product that has been tested under Food and Drug Administration regulations for safety and effectiveness as a means of relieving both pain and intense anxiety in children facing surgery, bone marrow biopsies or other major procedures. Anesta, which has sold marketing rights for Oralet to Abbott Laboratories, has gained approval for the drug from the FDA.

But some clinicians and health officials believe this is definitely candy that should be taken from babies. Several physicians have asked the FDA to withdraw approval for Oralet or have told the agency they will not use it. They think a sucker made from fentanyl (Abbott characterizes it as a "lozenge on a plastic holder") is the wrong thing to give a child about to enter an operating room. The National Institutes of Health, in fact, informed FDA commissioner David A. Kessler in early March that it would not use Oralet.

So far the FDA is holding its ground. In the absence of approved sedatives for pediatric patients, FDA officials point out, doctors have concocted their own recipes from drugs that had previously been approved only for adults. "I have a file four inches thick with papers in which doctors talk about their favorite way to sedate children," shudders Curtis Wright, the acting director of the FDA's pilot drug evaluation staff. The agency has documented 133 instances in which children died, were hospitalized or were disabled by use of painkillers intended for adults.

Before Oralet reaches hospitals, the FDA has taken the unusual step of requiring Abbott to submit a marketing and educational plan for agency approval (which has not yet been granted). The drugmaker would have to limit distribution to larger hospitals with sufficient staff schooled in advanced resuscitation. Hospital personnel will have to undergo a training program in the use of the drug. A medical worker must observe any child ingesting the drug, and someone must be present who has the skills to restore a child's breathing, if needed.

The caution stems from proven risk. Fentanyl can slow or even stop respiration. The FDA has issued a warning about another fentanyl product, a patch applied to the skin for chronic pain that caused a few deaths when it was misused by health workers.

Oralet's dosage levels—200 to 400 micrograms—are markedly lower than those for the fentanyl transdermal patch, and the drug is ingested for shorter periods. The patch's maximum dosage reaches 10,000 micrograms. But Sidney M. Wolfe, who represents the Public Citizen Health Research Group and who opposes Oralet's going to market, claims that both forms of the drug reach levels in the bloodstream that produce a dangerous risk of breathing impairments.

Wolfe also worries that doctors may start prescribing Oralet for uses other than those approved by the FDA, as they have with the patch. A 17-year-old Florida boy died after an oral surgeon prescribed the patch for pain relief after extraction of teeth.

Oralet: lozenge or lollipop?

"Taking a potent narcotic and putting it in candy is not the right message to be sending," says Allen J. Hinkle, director of the anesthesia residency program at Dartmouth-Hitchcock Medical Center, who has urged the FDA to rescind its approval of the drug. At Dartmouth, Hinkle has instituted a drug-free program to lessen children's preoperative fear. A child and a parent will visit an operating room to get to know the hospital a week or so before the operation.

Hinkle acknowledged that for some children, particularly those younger than six years, anxiousness may linger, notwithstanding operating rooms made familiar and friendly. But, he adds, fentanyl, which has relatively weak anti-anxiety properties, may still not be the drug of choice. "I would encourage the FDA to keep looking for pediatric sedatives," Hinkle says. "Unfortunately, the first one out of the gate is the wrong one."

—Gary Stix

The Price of Clean

Can California afford the EPA's 15-year plan to clear its air?

It was the best of times; it was the worst of places. Smog in southern California is down 75 percent since the darkest days of the mid-1950s, thanks to emission regulations that are the strictest in the world. Yet the air remains the nation's dirtiest: the level of pollutants wafting around Los Angeles exceeded federal health standards on 143 days in 1992. The Clean Air Act gives the Los Angeles basin until 2010 to get rid of its smog but left it to California to figure out how.

The state fumbled, so environmentalists turned to the Environmental Protection Agency to impose a plan. When the EPA, which hates to prescribe to industry or state governments what they must do to meet its regulations, refused, the environmentalists sued. They won, and on February 15 the EPA released a 2,700-page list of proposed rules that would restrict emissions from virtually every kind of engine, most consumer products and even some farm animals within the Los Angeles air basin.

The breadth of targets reflects the depth of cuts that the EPA has determined are needed to meet public health standards in southern California: hydrocarbons must be reduced by 90 percent, nitrogen oxides by 70 percent and carbon monoxide by 45 percent. Such cuts would be easier, points out Bill Sessa of the California Air Resources Board, were it not for a rapidly growing population and the fact that all the easy measures were taken long ago.

The EPA plan helps most, Sessa says, by imposing exhaust limits on polluters that state agencies lack either the authority or the political wherewithal to regulate. Airlines operating in the L.A. area may have to meet declining emission targets or pay fines. Ships could be charged fees to enter South Coast ports unless they use cleaner engines. And heavy trucks would have to meet tighter exhaust standards or reduce the number of stops they make in the region.

The proposal cast its far-reaching net over immobile sources as well. Gas stations would have to buy more advanced vapor-recovery equipment. Forty different kinds of coatings, from house paint to concrete-curing compounds, would have to be made with fewer volatile organic chemicals. Any dairy farm housing more than 400 cattle would have to run half its manure through a methane-recovery system.

None of the proposed rules will necessarily make it into law—indeed, the EPA would prefer to replace the plan with the state's own proposal, which is due in mid-November. "The quality of the decisions that come from the states is better than the decisions we can make," says EPA administrator Carol M. Browner. In the meantime, the agency will hold public hearings through July. The cost that the proposed regulations could impose on California's economy, which is still listing from defense cuts and natural disasters, will probably be discussed intensely.

The EPA estimates the toll to be $4 billion to $6 billion each year from 1995, when the rules would begin to take effect, through 2010. Jeffrey A. Alson, a researcher with the EPA's Office of Mobile Sources in Ann Arbor, Mich., characterizes the proposal as "heavier on the technology and standards and lighter on changes in behavior." For example, the L.A. plan avoids such provocations as "no-drive days"—a measure to keep 20 percent of the vehicles off the road on a given workday—that were included in a scheme for Sacramento.

Tom Soto, president of the Coalition for Clean Air, which brought one of the original lawsuits, points out that the emphasis on technology might yield benefits as well as costs, by spurring innovation in cleaner fuels, cars and consumer products. The EPA endorsed the California board's plan to force automakers to build cars that run on electricity or natural gas by 1998. The EPA's rules would then require some commercial fleet owners to buy them.

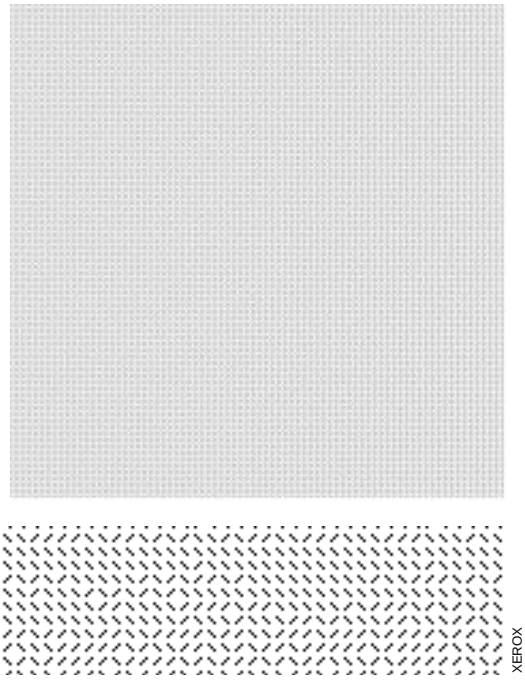
Soto also notes that Californians currently pay a price for their smog, in the form of health costs, reduced crop yields and lowered industrial productivity that total $9.4 billion a year, according to a 1989 study by the South Coast Air Quality Management District. "So there is still a net $5 billion from implementing this," he says. Or rather there could be, a generation or so from now, when the skies over the Los Angeles basin are an uncanny shade of blue. *—W. Wayt Gibbs and Gary Stix*

Gray Matter

There was a time when seers predicted that computers would banish paper from the modern workplace. Yet office machines still spend much of their time converting digital data into marks on a page. The conversion is often an annoyingly one-way trip—the vagaries of scanning effectively trap information in its printed form. David L. Hecht, a principal scientist at the Xerox Palo Alto Research Center, has now opened another lane between the electronic and paper worlds: software that can embed digital data in graphics on the page itself, using marks so small that they appear to human eyes as a gray patch. Hecht calls the scheme a "self-clocking glyph code."

The idea resembles a pointillist's vision of bar codes. A minuscule slash to the left represents binary 0; to the right, binary 1. These are the glyphs. Leave a bit of white space between the marks and arrange them within a latticelike frame of glyphs whose bits are instructions for reading the pattern. The patches can then grow to any shape or size. All the legerdemain that programmers wield on bits in computers can be set in ink as well, from data compression to encryption and error correction.

Therein lies the technique's real potential. Glyphs created by a standard laser printer can pack a square inch with only about 350 bytes of uncompressed data—fewer characters than are contained in this paragraph. But compression can double or treble that amount, and error-correction codes can be added to protect against smudging or other damage. "If you use a pattern with a fine enough grain," Hecht muses, "you could re-create a document after shredding it."



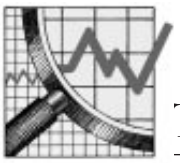
XEROX

***DIGITAL VERSION** of this story is stored on the page* **(top)** *using a technique developed by Xerox. Individual bits printed as tiny slashes* **(bottom)** *combine to form gray patches that hold more data than bar codes. Special scanning software can retrieve data even if they have been obscured by doodling.*

Xerox first attempted to commercialize a glyph code in PaperWorks, fax software released in 1992. The product garnered critical acclaim. But it has not conquered the personal productivity market at which it was aimed. PaperWorks's lackluster sales could feed worries that glyphs might simply be another innovation that got away from Xerox—like the desktop laser printer, graphical user interface and Ethernet network.

But Xerox is already talking to other software companies about licensing its glyph algorithms, says Brent G. Gnall, who is directing the business development of glyph codes. The company is also working on ways to embed glyph-packed graphics into documents produced by the most popular business software packages. "On the receiving end," says Beau Sheil, the director of advanced development, "we will provide the scanning software so that you can take a piece of paper and read it like a floppy disk."

Hecht reels off a list of possible uses. A gray background below a spreadsheet, he suggests, might encode the cell formulas. Printouts of charts could retain numerical data and color keys in shades of gray. Glyphs on invoices, shipping dockets and claim forms could alleviate some of the boring and error-prone data entry required of clerks. More far-reaching applications may come out of Xerox's agreement to help Microsoft produce its At Work software to make fax machines, copiers, printers and computers all communicate more intelligently. If so, glyph codes might one day play an important role in the grand unification of office electronics and of paper. *—W. Wayt Gibbs*



How Green Is My Label?

Buying groceries or appliances may seem an unlikely way to help the environment, but economists cite marketplace choice as the most efficient way to get manufacturers to abandon ecologically unsound practices. As sustainable products lead to increased profits and unsustainable ones to losses, companies should start to turn green. This wisdom is reflected in the rapid growth of eco-labeling programs: about 30 have arisen around the world since 1978, when Germany placed the Blue Angel seal on selected products.

Green labeling differs from other kinds of consumer information because it targets altruism rather than self-interest, says economist Thomas H. Tietenberg of Colby College. Although some labels note whether products contain harmful pollutants, most are more concerned with other criteria: whether items are produced with recycled materials; how much energy they consume during their manufacture and use; or what by-products their creation entails. Studies by *Advertising Age* and the Roper Organization have found that a majority of consumers say they would preferentially buy an eco-labeled product.

But does eco-labeling actually promote such behavior? Although classic economic dogma holds that markets are most efficient when buyers and sellers have access to all possible information, economist W. Kip Viscusi of Duke University thinks that less may be more on product labels. "There are overload issues," he says, citing studies that show that most people can process only about five separate pieces of information attached to a can of soup or a carton of garbage bags. Inundating consumers with details may lead to less thoughtful decisions, he contends.

The two economists and many of their colleagues agree that the crux of the matter is credibility: ecological data on products must be accurate and relevant, and consumers must be willing to rely on it. Gaining such trust is not easy. Consumers are aware that, as Viscusi says, "it's in everybody's interest to deceive them and to create the impression that the [products] are much better than they really are."

In the U.S., two organizations are vying for seals of approval from consumers: Scientific Certification Systems (SCS), a for-profit company in Oakland, Calif., and Green Seal, a nonprofit group in Washington, D.C. Although many organizations label products ranging from building materials to bananas, these two evaluate consumer products such as light bulbs and detergent.

SCS sells what it calls an environmental report card. Manufacturers pay the organization a fee (between $10,000 and $100,000) to evaluate the life cycle of a product—that is, the overall impact of its manufacture, distribution, use and disposal. The item then bears a label that resembles a table of nutritional contents, telling how many grams of particulates it releases, how many gallons of water it consumes and so on.

SCS representatives explain that they have chosen this approach because it educates the consumer and does not reduce labeling to a single seal of approval. Yet nutrition labels are based on well-known standards (such as U.S. government-established vitamin allowances), whereas the standards that SCS uses to devise the report cards are arcane. Furthermore, Richard A. Denison of the Environmental Defense Fund asserts that SCS's report cards are being interpreted by consumers as a seal of approval. "With a traditional seal of approval, you have to meet standards," he cautions. "But not in this case. You are not getting approval; you are just getting a rating."

Eco-labeling could make consumers see red instead of green.

In contrast, Green Seal takes an approval-only approach. The organization chooses a product category, such as fluorescent bulbs, in which its researchers anticipate that consumer choice could shape the market. They then set forth guidelines for manufacturers (reviewed by industry, environmental groups and the public) that are intended to ensure that products manufactured according to the standards cause the least environmental damage.

Products that meet the criteria for their industry are awarded a green seal. The cost of the seal—between $3,000 and $11,000—is covered by the company applying for green labeling. Underwriters Laboratories certifies the product by examining factories and testing the goods. Green Seal is modeled after the Canadian government's EcoLogo, which was first used in 1988 and now adorns more than 700 products.

These competing strategies have led many experts to suggest that perhaps someone ought to certify the certifiers. Currently the only regulation of green claims comes from the Federal Trade Commission, which evaluates false advertising. So far the commission has brought about 15 environmental cases. In one instance, a manufacturer attempted to pass off low-wattage, low-light bulbs as energy savers.

Some specialists view the FTC's work as insufficient. "It's not so much a question of telling consumers which labels are fraudulent as telling them which labels are uninformative," Tietenberg explains. He calls for environmental organizations to publicize cases where companies adopt ostensibly green standards that do not help the environment.

"The reason that this is so important is that consumers buy a product because they think it will benefit the environment," says the EDF's Denison. "If it is an insignificant benefit, then that is deception." He advocates federal intervention to deter misleading claims.

The government may have begun to assert itself. President Bill Clinton signed an executive order last year requiring federal agencies to use more recycled goods. The Environmental Protection Agency is developing guidelines for determining just what products will meet this requirement. The EPA has also asked Viscusi and his colleague Wesley Magat to study ecological labeling; he says that when the report is finished in December, it will recommend national standards.

Meanwhile at least some companies are trying to clean up their acts before the labelers arrive. William D. Browning of the Rocky Mountain Institute reports, for example, that when the American Institute of Architects compiled its environmental resource guide, the very act of asking questions about manufacturing processes caused some rug makers to unroll the green carpet.

—Marguerite Holloway and Paul Wallich

How Many Guards in the Gallery?

It was the monthly meeting of the directors of the Parrot Corporation. The president, Boss Parrot, nodded in satisfaction as the chairman moved to the final item on the agenda, the Warthog Collection. This was to be housed in a high-tech art gallery—Parrot liked to refer to it as a state-of-the-art gallery—that would be devoted to the complete works of the incomparable Sandy Warthog. Parrot, an early admirer of Warthog's slick devialist style, owned every picture the artist had ever produced.

"Security will be an important item," the head of finance noted.

"Naturally the Warthog Collection will be fitted with all the latest electronic surveillance gear," said Harry Sams, chief of security. "An individual microcamera will face every painting—"

"No, I want something special for the Warthog Collection," interrupted Parrot.

"Uh—*how* special?"

"*Real* special, Harry. I want guards. Human guards."

"Boss, let's not get too ambitious, huh? Do you realize the cost in positive security vetting alone? Not to mention medical insurance, severance pay, coffee-break tokens. Real people cost real money, Boss. No, what I suggest is the latest sniffer robots from Notsobitchi and maybe a coupla—"

"I want guards, Harry."

"Boss. Guards. Uh—how many?"

"Enough to make sure that every square inch of the building can be seen by at least one guard. I want each guard stationed on a swivel chair, so that in effect each one has all-around vision. Hire as many as the job needs—but not a single guard extra, you understand? Human guards cost real money, you know."

***STATE-OF-THE-ART GALLERY** houses the Sandy Warthog collection.*

"Yes, sir, Boss Parrot, sir. Uh—have the building plans been finalized?"

"Not yet."

"That makes it kinda awkward to work out the required number."

"The architect has decided that the building will all be on one level and that there will be 24 straight walls—one for each of Warthog's definitive style periods. What I'll do is this: since you're operating with inadequate information, I'll authorize you to hire precisely enough guards to keep any room with 24 straight walls under surveillance, assuming each guard stays in one fixed position and has all-around vision. But I don't want you hiring anymore than you really need. If it turns out that the Warthog Collection needs fewer guards than the theoretical maximum, I'll accept responsibility. But if *no* design of the gallery needs as many guards as you hire—or if you don't hire *enough*—then I hope you can find someone else to hire *you,* Harry."

When you're a high-powered executive with your job on the line, you don't fool around. You spend a lot of company dollars to hire the best outside consultants that money can buy and tell them to solve the problem. Harry found them in the phone book: an outfit called Moonlight Mathematicians. Alf Moon and Dee Light.

"That's a tough problem, Mr. Sams," said Moon.

"Real tough," said Light, who already knew the answer. But you don't make a profit by telling the customer things like that.

"Thing is," Moon went on, "there's an awful lot of ways to arrange 24 straight lines." The two of them had a habit of speaking alternately.

"And the number of guards you need depends on which arrangement you choose."

"Fortunately, there are some good techniques around. They belong to what we in the trade call Art Gallery Theorems."

"Yeah, right. Like, if a room needs just one guard, then it must be what we call star-shaped."

"Which means that there is some interior or boundary point such that any point in the room can be linked to it by a straight line that stays inside the room."

"Now, imagine you've got a number of guards inside a room, placed any old way. Then each of them sits in a star-shaped region—namely, that part of the room that he or she can see. So what you're asking is—"

"What is the minimal number of star-shaped regions into which any room with 24 walls can be decomposed," said Moon, sketching quickly on his notepad [*see top left illustration on opposite page*].

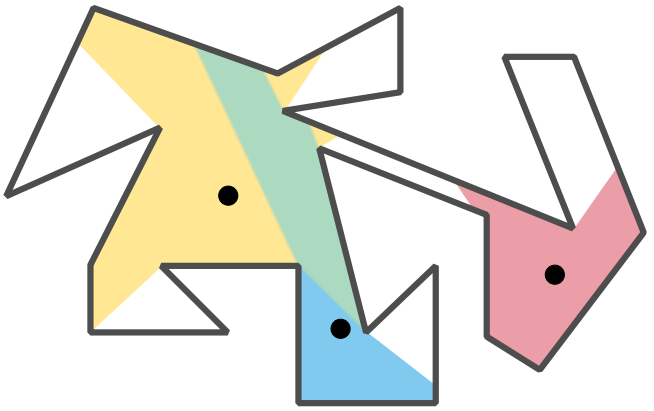
"You don't say," said Harry.

"So in this particular room, with this particular placement of guards," continued Moon, "you can see that the three guards I've drawn can survey all of the room except for six separate regions."

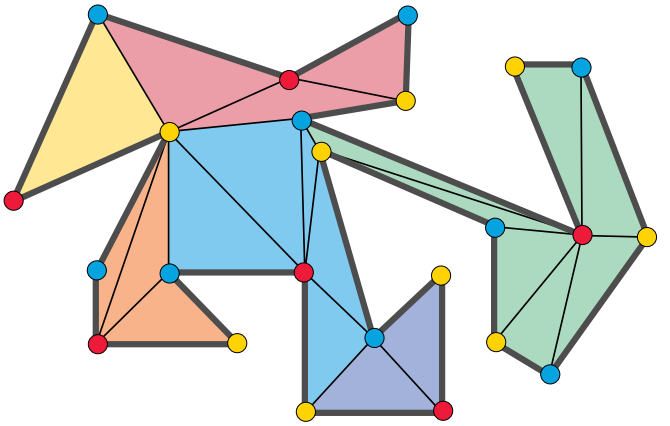
"Which are convex," mentioned Light.

"In particular, that means they're star-shaped."

"So if you put an extra guard in each of them, making a total of nine, then every part of the room would be in view."



THREE GUARDS (black dots) *are placed in the star-shaped regions that they can watch over. Note that the blue and yellow regions overlap* (shown in green). *Several parts of the gallery are not visible to any of the guards.*



COLOR THE VERTICES of a room so that each triangle contains each color once. Place guards at the vertices with the color that occurs the fewest times. For a room with n *walls, this method guarantees a solution with* $[n/3]$ *or fewer guards.*

"Great!" said Harry. "So I need nine guards?"

Moon and Light shook their heads knowingly. "Not at all, Mr. Sams. First, because some arrangement with fewer guards might be enough for this particular room—"

"And second, because some other room might need more."

"Let me summarize," said Sams. "Either the answer is nine, or it might be less, or it might be more."

"You got it."

"How much are we paying you guys?"

"Just getting the feel of the problem, Mr. Sams," said Moon. "Obviously, we gotta be systematic here."

"'If in doubt, chop bits out,'" said Light. "We need to chop the room up into simpler pieces and then try to reassemble them into star-shaped regions in as efficient a manner as possible."

"Right. And how do you propose to do that, Light?"

"'To get things straight, triangulate.' I think triangles are the natural pieces to use, Moon, don't you agree?"

"Yes, but there are lots of ways to chop a room into triangles."

"Sure. I propose we do so without introducing any extra vertices." She turned to Sams. "That's a technical term in the trade for the points at the corners." Sams nodded.

"You mean, using only the vertices that the room already has?" Moon asked. "Can you always do that?"

"Sure. I can prove it if you want."

"No, go on, it's getting interesting."

"Next, I'm going to color every vertex of the room, using only three colors—red, yellow and blue, say—so that each triangle has exactly one vertex with each color. Just color any particular triangle, using a different color at each vertex, and then kinda work outward. The colors on every adjacent triangle are completely determined because exactly one vertex of that triangle isn't already colored, so you just use whatever color is different. Keep going systematically, and you color all the vertices."

"Right. Now look what happens when we do that to your diagram. You end up with six red vertices, nine blue ones and nine yellow ones. If you put a guard at each of the red vertices, they can see the whole of the room. Because each guard can see every triangle that he or she is a vertex of—"

"And every triangle contains a red vertex. Of course. That's neat. We need only six guards for this room, not the nine we first found. And we put them at the *corners*."

"Vertices." Light leaned forward with a quizzical look on her face. "Now, Moon, my friend, I hope you see how to generalize that argument?"

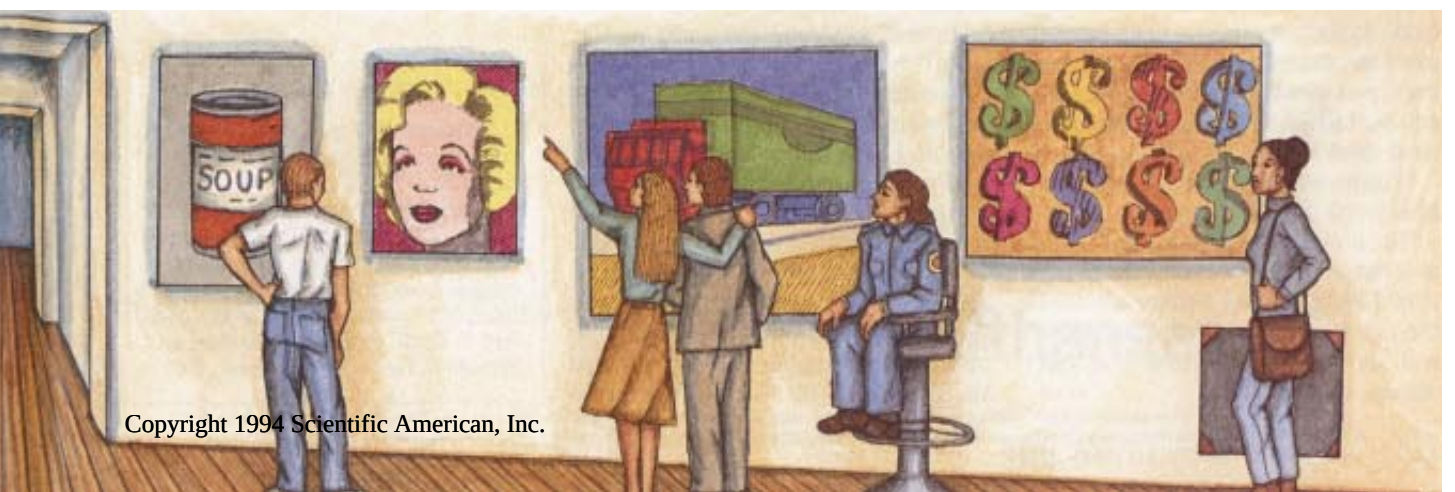
Moon stomped up and down the room, arms waving excitedly. "Yeah, yeah. If a room's got 24 walls, then it's got 24 corners. If we carve it up into triangles and color using your rules, then we choose whichever color occurs the smallest number of times and put the guards at the corresponding vertices."

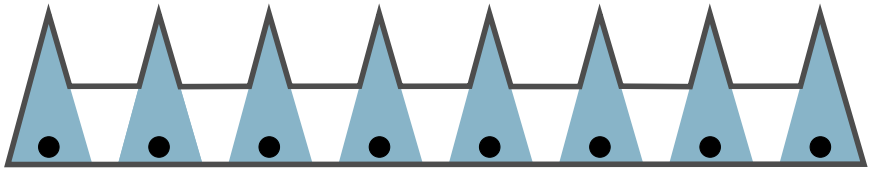
"Quite. Now, the total number of red, blue and yellow vertices is 24, so however the colors are assigned, at least one of them must occur eight or fewer times."

"You mean, if three numbers add up to 24, then they can't all be nine or more? Yes, I see that."

Light grinned. "So I've proved that *any* room with 24 walls can be guarded by at most eight guards."

"Hold it," said Sams, who had been listening to the byplay. "You just said





ROOM THAT NEEDS $[n/3]$ *guards is shown here with* $n = 24$.

this particular room needs only six, not eight."

"Correct, Mr. Sams. But we've got to find the smallest number of guards that will be enough to maintain security in *any* room with 24 walls. And what we've proved is that it's at most eight."

"So all we need," said Light, "is to find a room with 24 walls that needs eight."

"Which is easy," said Moon. "This'll do it." He drew a quick sketch [*see illustration above*]. "The guards have to be placed so that they can see into the eight triangular alcoves, and that gives eight regions that don't overlap, so each must contain a different guard. On the other hand, my placement of guards shows you can manage with eight."

"You haven't put the guards at the corners," protested Harry.

"We *can* put them there, and it helps to do so for the general proof," said Moon patiently. "But we don't *have* to, and in this case it's simpler not to."

"Okay," said Sams. "Eight guards. That's great. I'll tell Boss Parrot that the problem is solv—" The phone rang, and Sams picked it up. He nodded several times, grunted and placed the phone carefully back on the desk. Then he picked up a vase of flowers and hurled it across the room. He jumped up and down amid the wreckage, shouting incoherently.

"Trouble, Mr. Sams?" asked Light, with what she thought was tact.

"Damn fool architect has changed the specs," yelled Sams. "Now there are going to be 173 walls, so that the paintings can be grouped by thematic content as well as style period."

"No problem, Mr. Sams. Because we found a general method, it still applies. If you color the vertices of the room using Light's rules, you just have to place the guards at whichever color occurs least often. Now, if three numbers add up to 173, then one of them must be at most $173/3 = 57\ 2/3$."

"You're saying I need $57\ 2/3$ guards? Boss won't like that."

"No, it means that some color occurs at most 57 times, so you can manage with 57 guards. And a shape similar to the one I drew [*see illustration above*] will show that in general you can't manage with fewer."

"What we've proved," said Light, "is that a room with n vertices needs at most $[n/3]$ guards, where the brackets mean 'integer part.' That answer was found by the Czech mathematician Václav Chvátal in 1973. The proof I gave is a simplification found by Steve Fisk of Bowdoin College in 1978, and—"

The phone rang again. Sams answered it. "Would you believe it? After all that, the architect's gone back to a 24-sided room. Says there's too much dispute about which of Warthog's themes is which. Okay, you guys bill the company for your time, and send me a complete report with an executive summary by Thursday. I've got an appointment with Boss Parrot."

"We're real admirers of Boss Parrot," said Moon.

"Could we come, too?" pleaded Light. "We'd love to meet the great man. Probably never get another chance."

Sams thought for a moment and finally agreed. Wouldn't do any harm to have them along for moral support, anyhow. Just in case any technical queries came up.

"Eight guards, you say? Smart work, Sams." Parrot positively beamed with delight. "As it happens, the architect has just made the final decision on the gallery layout. You can show me where the guards have to be stationed." He spread the plans on the desk [*see illustration at right*].

Sams glanced at the plan, did a double take and looked hopelessly toward Moon and Light. Moon gave Light a nudge and whispered, "Whoops, *holes.* Nobody said anything about holes. Eight guards won't be enough for that room. I can tell just by looking."

"No. Hmm. There's a conjecture that if there are h holes and n walls—"

"Is that including the walls of the holes?"

"Mmm? Yeah, of course—then the maximum number of guards needed is $[(n+h)/3]$. The plans have $n = 24$ and $h = 3$, so nine guards ought to do—"

"What are you two whispering about?" interjected Parrot suspiciously.

"Beg your pardon, sir. We were—er—admiring the architect's design." Light gave Sams a conspiratorial wink and pulled a pen from her pocket. "Now, as Mr. Sams said, with 24 walls you can always manage with eight guards. Plus a superintendent, of course, that goes without saying: someone has to be in charge, make sure the job's done properly. Isn't that right, Mr. Sams?"

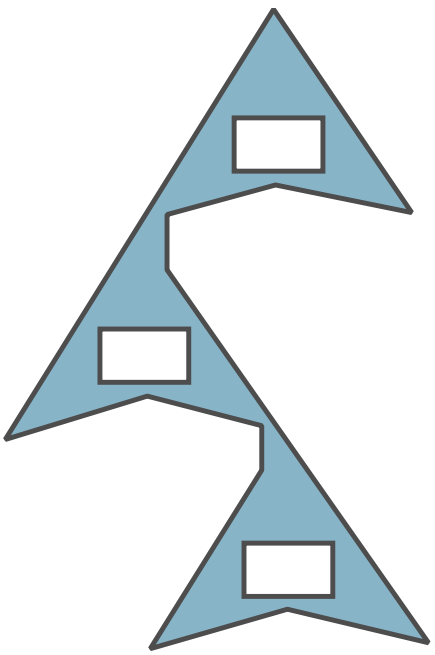
Sams, his brow beaded with sweat, pulled himself together. "Yes, Ms. Light, naturally I was intending a superintendent as well."

"So you'll need nine guards in total, sir, just as Mr. Sams was telling you."

"It will all be in the executive summary, Boss," said Sams, recovering his poise.

He gave Moon and Light a sideways glance and was rewarded with almost imperceptible nods. Nine guards. Wasn't that the figure they'd come up with right at the start? Anyway, the problem now was where to put the nine guards. The Moonlight Mathematicians would know, of course.

Somehow Sams had a feeling they'd just raised their rates.



WHAT HAPPENS when there are holes? An unsolved problem for readers to think about. This gallery needs nine guards: Where should they go?

FURTHER READING

MATHEMATICAL GEMS II. Ross Honsberger. Mathematical Association of America, 1976.

OLD AND NEW UNSOLVED PROBLEMS IN PLANE GEOMETRY AND NUMBER THEORY. Victor Klee and Stan Wagon. Mathematical Association of America, 1991.

UNSOLVED PROBLEMS IN GEOMETRY. Hallard T. Croft, Kenneth J. Falconer and Richard K. Guy. Springer-Verlag, 1991.

BOOK REVIEWS by Philip Morrison

A Volcano Is Born

PARÍCUTIN: THE VOLCANO BORN IN A MEXICAN CORNFIELD, edited by James F. Luhr and Tom Simkin, with Margaret Cuasay. Geoscience Press (12629 N. Tatum Blvd., Suite 201, Phoenix, AZ 85032), in association with the Smithsonian Institution, 1993 ($50; paperbound, $25).

As 1943 came in under seasonally blue skies, 10,000 Tarascan maize farmers on their beautiful pine- and oak-clad highlands knew steadily rising fear. They had heard subterranean noises and felt oscillatory tremors, occasionally at first, then daily. On the scary 19th of February the alert in San Juan Parangaricutiro, the largest village around, home to 2,000 people, could count 300 tremors. "The people could not…have confidence to remain in their houses to sleep," one local diarist noted.

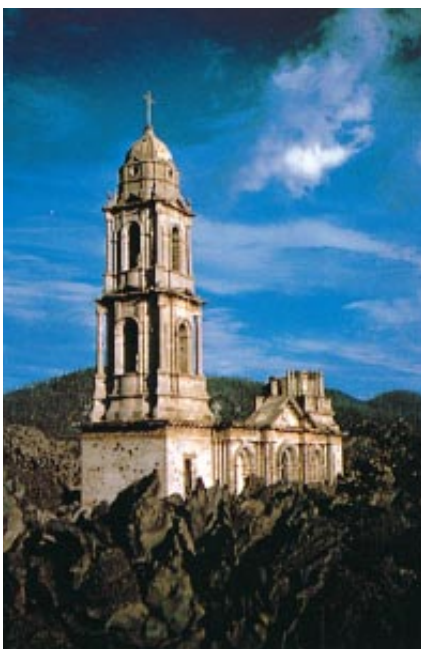
About 4 P.M. the next day, a Saturday, Dionisio Pulido, a Tarascan landowner from Parícutin, a smaller village nearby, went out to his field to tend his sheep. He took along his wife, their small son and a youthful plowman with a yoke of oxen. Pulido moved off to burn a small pile of branches, passing by an old pit, a few feet deep and 15 feet wide, that had always been there on one gentle knoll of his land. He saw a long fissure in the ground, a foot or two deep, extending across the pit into the far field. "Here is something new and strange, thought I." Ready to ignite the branches, he "felt a thunder" and saw the ground in the pit swell up a couple of meters high. Dust rose the length of the fissure, until a loud hiss brought smoke and perhaps fire. "I then became greatly frightened…mounted my mare and galloped" a mile or more to his village, where he found his frightened family and friends already at home, ready to confirm his account.

A photograph taken by a 14-year-old boy at about 5 P.M. from three miles north shows the initial wispy smoke rising. That evening some young men returned warily to the field, to see glowing stones as big as marbles and oranges thrown high, then to watch from a distance a rain of fire, huge clouds of vapor, strange blue flames. The village priest heard them out, consulted an old book on Vesuvius and concluded that a volcano had been born. On Sunday the San Juan officials met, listened well and sought help "by telephone and telegraph" from the leaders of the Michoacán state and the nation. All present agreed, "after ample deliberation," on a name for this unexpected babe: the Volcano of Parícutin (accented on the second syllable, pa-RIC-u-teen).

During Sunday came the most alarming tremors of all. Large bombs flew high to fall into the fields; that night, incandescent lava flowed out "like dough," forming what everyone could recognize as rocky *malpais.* (*Malpais,* meaning "badland," is a common Spanish term for the rugged surfaces of young lava flows with little soil or vegetation.) The huge infant stood 50 yards high in the field. On Monday aerial photographs showed the new cone and its vents; by week's end the cinder cone had grown into a notable hill 550 feet high, Pulido's fields forever covered.

By 1952 the volcano's heated life was over. No full counterpart is known to this little volcano so well watched from birth to death on dry land. There proceeded three slower pageants, the complex return of the injured flora and fauna, the resettlement of the disturbed people and the parade of geologists who still flock there as delighted inquirers. This bargain anniversary book is a well-made tale by two Smithsonian volcanologists with a gift for narrative. It is a collage of maps, air photographs in stereo pairs, commemorative postage stamps and newspaper clippings, a dozen color plates, one showing a colorful procession in resettled San Juan Nuevo, a detailed chronology, biological and sociological comparisons village by village, three dozen excerpted geological papers between 1946 and 1993, and end papers that amount to a lively animation.

One map reports the cone profile, week by week at first, then season by season, until finally the last layer stores a decade of impoverished deposition. The flat top of the cone stands today 1,400 feet above the 7,800-foot level of the maize roots. The lava had reached out its longest arm first, to entomb most of old San Juan, including the big pilgrimage church whose masonry towers alone protrude today out of a dozen or two feet of blocky *malpais.* Lava flowed thereafter to all points of the compass, no arm longer than that first rush, aided as it was by a valley. The pilgrims come more eagerly than ever to the new town to adore the image of El Señor de los Milagros, the very old crucifix that beckoned the faithful long ago. It was ceremonially removed for reset-



LAVA from the newly formed volcano at Parícutin, Mexico, buried the large pilgrimage church in the nearby village of San Juan Parangaricutiro.

tlement in 1944 once the lava finally reached the old church walls. The early ashfall buried poor Parícutin to the rooftops that first June, and the villagers left their homes. A year later the lava took all. The entire ashfall, out to a minimum thickness of one foot, covered an area of about 100 square miles. In time, the district's farmers learned to plow deep, mixing soil and ash, able to win a good crop of maize once they added fertilizer.

The geologists admire Parícutin as a type specimen. Its continuity of evolution suggests that it tapped one single chamber of fluid magma, often through one orifice only a meter or two across. The impelling gas was mainly superheated steam, moving at 300 meters per second, driving chunks of red-hot basalt, still plastic, high into the air. The rush of particulate ash led to lightning, in early times a zigzag flash every minute. The typical energy flow, mainly as heat stored in the outpouring of semifluid rock, was about that of the largest city power plants, up to 10 gigawatts, set at work for three or four years. The flow went on for nine years, steadily tapering off.

The rock type, like the flow rate, changed very smoothly. It began with a basalt, low in silica, high in steam and hence explosive; it ended in a silica-rich melt without steam, slowly moving over the ground to solidify as it went in big blocks below its fast-cooling crust. There is reason to believe the change in composition marks the flow of eroded country rock that in the end dominated the initial stream of hot magma from a few kilometers below the surface. About one cubic kilometer of rock was thrown out in total, most of it airborne; 1 percent of the mass was the working substance, steam. Five or 10 percent of the heat energy of this engine appeared as kinetic energy of cone construction, high-flung fiery bombs, white jets of steam and tall plumes black with ash.

Such volcanoes are not single, but gregarious like us. Similar old cones, rolling benches, smoothed slopes, useless badlands, are everywhere in this green and misty countryside. A couple of thousand cinder cones dot the Tarascan lands over an area as big as Vermont and New Hampshire. The geologists have plotted them, examined their remains, bigger than any dinosaur bones, and found their ages. For a few million years, each was built over a local magma pool, a new one oozing up every millennium or so. Not far away, another cone was watched from birth, recorded by the officials 250 years ago. The cinder cones tend to align in long strands as the whole region drifts northeastward with the continental plate; others will appear in their time, small-scale models of Pele's ocean islands, though much undernourished through thick subfloors below.

The wonder of Parícutin drew amazing witnesses, even in those months at the end of the battle of Stalingrad, hingepin of World War II. Dionisio himself poses, taut, rugged, in grave dignity, before the high smoking hill that was once his field. That first Sunday night the leading geologist of Mexico, age 76, "father of the Mexican oil industry," had arrived to explain volcanoes to the people, building on their familiarity with long-frozen *malpais.* Soon after him followed the celebrated Dr. Atl, white-bearded poet, painter, volcanologist, revolutionary, sponsor of the three great Mexican muralists, shown in the book at his easel below new Parícutin. Atl was no mere popularizer. Long before others, in 1950, he wrote in prescient italics: "*This abundance of volcanic structures is due to innumerable fractures caused by continental drift.*" Once upon a time Baron von Humboldt was less daring; on his visit to that same landscape before 1804 he stuck to a theory then in fashion: the old cones were mere thin coatings of cindery stuff on top of a swollen blister of rock. He lacked the spectacle of a Parícutin to demonstrate cone building on a plowed field.

A Mind We Ought to Know

ALEXANDER A. FRIEDMANN: THE MAN WHO MADE THE UNIVERSE EXPAND, by Eduard A. Tropp, Victor Ya. Frenkel and Artur D. Chernin. Translated by Alexander Dron and Michael Burov. Cambridge University Press, 1993 ($49.95).

Albert Einstein at the height of his powers wrote 50 papers from wartime Berlin. His first paper on cosmology, published in 1917 when he was still younger than 40 years, opened with what seems the simplest of cosmic models: uniform matter enduring in unbounded space-time. Yet neither Newton's gravitation nor his own curved space-time can consistently describe this concept. Avoid the infinite! Try a closed space, a finite sphere of uniform matter. It was workable, though only after a simple, if rather arbitrary, modification of his equations. The solution needed a new, weak, repulsive force to counter gravitational collapse; all is then balanced and unchanging. That work, badly flawed, nonetheless launched modern cosmology.

Five years later an intense theorist in

Petrograd sent a carefully presented calculation to the "most authoritative physics journal of the time." Its title was (translated from the German) *On the Curvature of Space.* The paper was explicit. Einstein's argument had been falsely restrictive; he had overlooked the possibility of cosmic change in his own theory. A symmetric cosmos remains so, whether finite or infinite, static, expanding, contracting or even oscillating. The general reader will recognize the broad consensus of our days—an evolutionary universe. What most cosmologists now study is rightly called after Friedmann. He knew that no one could then choose which possibility is in fact realized; even today we have insufficient grounds for truly decisive choice.

The exemplary little comedy unfolded. Einstein was by now a very bright star indeed, soon to be a Nobelist for his early work on photons, unrelated to space-time. He responded in print to Friedmann's piece by a few rather impatient lines, opening "The results… appear to me suspicious." Petrograd would have the last word all the same. Friedmann and his friends explained to Einstein by letter and visit just where the star had gone wrong. Pretty soon the readers of *Zeitschrift* would read under the name of the generous new laureate: "My criticism…was based on an error in calculation…. Mr. Friedmann's results are correct and shed new light."

Friedmann, though unknown in Berlin, was no youthful upstart. He was in his early 30s, a maturing young leader of physics in tumultuous postrevolutionary Red Petrograd. He lectured and wrote for the public, the engineers, the theorists, at no less than five institutions, from the naval academy to the university. His interests were remarkably wide; he held a combat decoration for his volunteer service as an observer-pilot, dropping bombs from a Farman biplane. His taste and experience with instrumentation and numerical methods supported his analytical powers, aimed closer at understanding the weather than designing aircraft.

It was no surprise that the cosmologist would in 1925 become director of the Main Geophysical Laboratory. It was more of a surprise that the young director soon flew with a famous balloon pilot on an ascent into the heights in order to test his view that turbulence might well be studied by elaborate measurements in the free air. The flight was heroic, if life-threatening. Friedmann and the pilot rose to the stratosphere; their oxygen reserves ran short at 24,000 feet. They got lost during those

half-conscious hours; luckily, they landed safely, deep in the impoverished neighborhood beyond Novgorod.

It is an irony that poor Alexander Alexandrovitch was fatally infected by typhoid fever only a few days later and died the following month, September 1925. His cosmology remains. It is tied to what we believe at present by two strong threads. The first is his young Petrograd student George A. Gamow, who first brought particle physics into the geometer's space-time at the end of World War II in the U.S. The second is a powerful school of physical cosmology that flourished in Moscow and Leningrad, in the work of Sakharov and Zeldovich and Linde and many another.

The turbulence of the story is echoed in the biography itself, published in Russian in still turbulent 1988. The photographs present a cultivated man of refined features, clearly castable in New York as the young intellectual in any Chekhov play. We rather expect what we learn: he was a revolutionary youth and a keen high school mathematician in 1905, and he would usually attend the symphony with orchestral score in hand. His father was a ballet dancer and composer; the parents' marriage dissolved when the boy was nine. The original relativist is foreseeable, the bold pilot perhaps less so.

The three authors speak with distinct voices, sometimes in unreadable detail, sometimes broadly vague. The book is not at all easy to read, but its wealth and its generous citation of Friedmann's popular works on space and time help to fill in a place, a time and a mind we ought to know. No other such source is found in English.

A Dictionary of Plant Species

FLORA OF NORTH AMERICA: NORTH OF MEXICO, edited by Flora of North America Editorial Committee. Fourteen volumes. Vol. 1: **INTRODUCTION;** Vol. 2: **PTERIDOPHYTES AND GYMNOSPERMS.** Oxford University Press, 1993 (each volume, $75).

What you get here is a lot more than what you see. Perhaps more apt than the iceberg metaphor would be a flourishing tree, its deep-buried roots as widespread as its green branches. This authoritative, many-headed work is centered at the Missouri Botanical Garden in St. Louis, where a much enriched digital counterpart, the TROPICOS database, will reside on-line for manipulation and updating. This hard-copy version will be completed by a dozen more big volumes, scheduled to appear about one a year, the final one an overall index and source list.

Volume 1 is a set of some 15 essays by a score of botanists, intended to summarize the current state of knowledge about plants of our continent north of the subtropics, including Greenland. A history of this collaborative project itself, now well under way for a dozen years after an earlier false start, is supported by a fine chronicle of flora. Although the modern pioneers were the energetic 17th-century herbalists of England and France, just about every traditional society held a treasure of its own floristic wisdom. But the ideal of a comprehensive account, not a merely regional one, its terms universal and known to all, was very much the work of the Enlightenment, in the person of Carl Linnaeus. His artificial classifying principles, though not the main form and purpose of his work, were transcended in the 1840s in the synthesis by John Torrey and Asa Grey, the first *Flora of North America*. A long list of tireless collectors, the freeing Darwinian insights, a wide university-based systematics and a maturing biology have since brought the old taxonomy to a new place. New species no longer drive our taxonomy; now it is new understanding. "By the end of [World War II]…walking across a meadow enjoying the flowers was often replaced by…a well-resolved chromatograph."

Other essays review the entire setting in space and time: climate, relief, soils, the fossil record, the present continental zones of vegetation, relationships with humankind. The basis of the classification system here adopted is analyzed. The history of most visible North American plants begins in the Upper Cretaceous, some 70 million years ago, when the groups we see today, in particular the flowering plants, became dominant. The present vegetation shows in detail a much smaller depth in time, two million years and less, the glaciers a major influence. Especially is this the case because the largest plant regime in our continent is the boreal forest, the young conifer-rich taiga of the north that greens nearly 30 percent of the land area described.

The work draws authority from real plant specimens, the myriad pressed parts and meticulously labeled sheets with which collectors of living plants have long endowed our herbaria. Some 60 million specimens, of course with many repeats and variants, are filed in the U.S. and Canada; a few individual herbaria can boast of millions of specimens each. From these samples and sheets and the lengthy published rec-

ord, this *Flora,* together with its plant drawings, distribution maps, keys, formal descriptions of plant characters, is being formed and documented. Today's authors discriminate by the use of field records, genetics, microscopy and biochemical comparisons of alkaloids and phenolics, enzymes and now DNA sequences, especially from chloroplasts.

For this reader, the key introductory essay is that by G. Ledyard Stebbins on the species concept and its history. Animal species are much neater; they do present reproductive isolates. Plants are not so simple: reproduction can be vegetative or sexual in one form. An individual may endure for a few years or over centuries, through steady cloning. Gene exchange is more complicated than any simple answer, yes or no. Examples make a strong case that nature "forces botanists to adopt a pluralistic species concept," although one that is biological and not merely subjective.

These books at the root are lists of identifiable species. They fulfill the simple democratic slogan: one species, one note. In the second volume, for example, the giant sequoia, the most voluminous of living organisms, is only given half a page of text, a postage-stamp map and a small line drawing shared with the Coast redwood that shows a twig and a cone or two for each. The same amount of space is offered to a little climbing fern. The work is as visually austere as it is verbally concise; its line drawings in ink are small, significant close-ups that are rarely supplemented with a microphotograph.

Volume 2, the first to appear of the dozen main volumes, is the work of no fewer than 55 contributors from the U.S. and Canada, plus their skilled illustrators. It details 430 species, the ferns and fern allies, the conifers and their kin. (Extrapolation would predict only 5,000 species in the whole work, which on various weak grounds seems low by a factor of three or even four.) Will most volumes to come include many shorter accounts? Note that even these ambitious books admit only those extant land or aquatic plants, not marine, that develop through an embryonic stage and then grow spontaneously (not under cultivation) within the wide but sharp geographic boundary.

Those who seek showy, colorful portraits or dramatic descriptions will find none. This is a dictionary of plant species, the working vocabulary of plant biodiversity, as essential to its knowing, productive users as any big dictionary and just about as dry to a casual glance. But its purpose is to become the arbiter of our plants for conservation and study.

ESSAY by Victor F. Weisskopf

Endangered Support of Basic Science

Fundamental science, based on the urge to know more about nature and ourselves, is in danger today. One need look no further for evidence than the offices of the National Science Foundation. The foundation's original mission was to support basic science. Now even this leading funder of research for its own sake directs an increasing proportion of its resources toward applied goals. Investigators who try to win approval for pure research projects by emphasizing the work's practical implications and administrators who are sympathetic to such proposals have received a stern warning from Senator Barbara Mikulski of Maryland, who chairs the Senate committee responsible for the funding of much basic and applied research.

The senator has said, "The foundation should make clear how it specifically defines each area so as not to shroud curiosity-driven activities under the rubric of strategic activities." She has recently defined strategic research to mean investments in science that are focused on important national goals such as climatic change, advanced manufacturing and high-performance computing.

Perhaps such a shift in emphasis was to be expected. For the past 100 years, the theory of relativity and the theory of quantum mechanics have pointed investigators toward a series of brilliant basic discoveries and insights. Today science, at least in physics, seems to be approaching a peak of sophisticated perception. From both physics and biology, innovative technologies constantly take shape. At the same time, an absence of new, big theoretical ideas has become apparent. Social and economic factors have also contributed to the shift in emphasis. As concern about economic competitiveness and the environment began to grow, public officials turned to science for solutions.

The tendency to support science so that it might produce technological fixes is especially dangerous for those parts of basic science that are quite far removed from practical applications. Astronomy, cosmology, particle physics and part of nuclear physics exemplify such areas. They deal with phenomena that are distant from us in time and space. Processes such as element formation in stars; the big bang and what happened in the first three minutes; the discovery of quarks, mesons and short-lived heavy electrons; and the study of highly energetic nuclear reactions tell us much about the innermost structure of matter and about the whence and whither of our universe. They have little to do with the processes in our terrestrial environment. These studies can be referred to as cosmic sciences.

In contrast, atomic physics, condensed-matter physics, that part of nuclear physics concerned with reactors and radioactivity, as well as chemistry, biology, and geology are studies of the properties of matter here on the earth. In this realm, quarks and other elementary particles do not come into play. We may call these studies terrestrial sciences, and for these fields, applications are obvious and easily predictable.

This analysis leads to a division of physics into "nonapplicable" and "applicable" fields. The cosmic sciences fall into the first category, although one never can tell. Some development lying in the future, unforeseen, may transmute them into applicable disciplines. The nonapplicable fields have already produced some spin-offs: for example, the ultrasensitive and discriminating detectors developed in particle physics have turned out to be most useful in medicine, biology and environmental research. Basic research has also proved to be the best training ground for future applied scientists.

But such considerations are precisely the wrong grounds for supporting inquiry whose goal is simply to understand. Science is like a tree whose roots correspond to basic research. If the roots are cut, the tree degenerates. Philosopher Michael Polanyi said, "[Basic science] was devised precisely for the purpose of elucidating the nature of things under more carefully controlled conditions and by more rigorous criteria than are present in the situations created by practical problems.... Such sensibility cannot be switched on at will for purposes alien to its inherent passion."

Basic science, cosmic as well as terrestrial, also creates important educational, ethical and political values. It fosters a critical, antidogmatic spirit, a readiness to say, "I was wrong," and an idealistic inclination to do work where there is little financial gain. Basic science establishes a bond between humans and nature; it does not recognize industrial, national, racial and ideological boundaries.

Today we cannot afford the kind of lavish funding that basic science enjoyed during the decades following World War II. But the support must not be reduced to such an extent that important fields of fundamental research wither away so that young people cannot pursue them. Particle physics is an example. The Superconducting Super Collider has been canceled, and the construction of a somewhat smaller European particle collider is being held up by lack of funds (and perhaps by lack of competition from across the Atlantic). If these projects are not realized, or are postponed by many years, young people will not enter the field, and our search for the fundamental structure of matter will be seriously endangered. Only international collaboration, neglected in the past, could perhaps save this extremely worthwhile inquiry from collapse. The decline of this branch of work would not be good for science despite its limited practical applicability.

In his 1995 budget President Bill Clinton proposes to spend less on research and development as a proportion of gross domestic product than the nation has spent in any year since 1958, according to the House Committee on Science, Space and Technology. This trend does not bode well for the future.

Polanyi was right. Science cannot flourish unless it is pursued for the sake of pure knowledge and insight. That kind of inquiry cannot be turned on and off as if it were an electric light. On the other hand, we as a species will not survive unless science is applied intensely and wisely for the betterment of the human condition. Human existence depends on compassion and knowledge. Knowledge without compassion is inhuman; compassion without knowledge is ineffective.

VICTOR F. WEISSKOPF, emeritus professor of physics at the Massachusetts Institute of Technology, was a group leader in the Manhattan Project and director general of CERN, the European laboratory for particle physics.