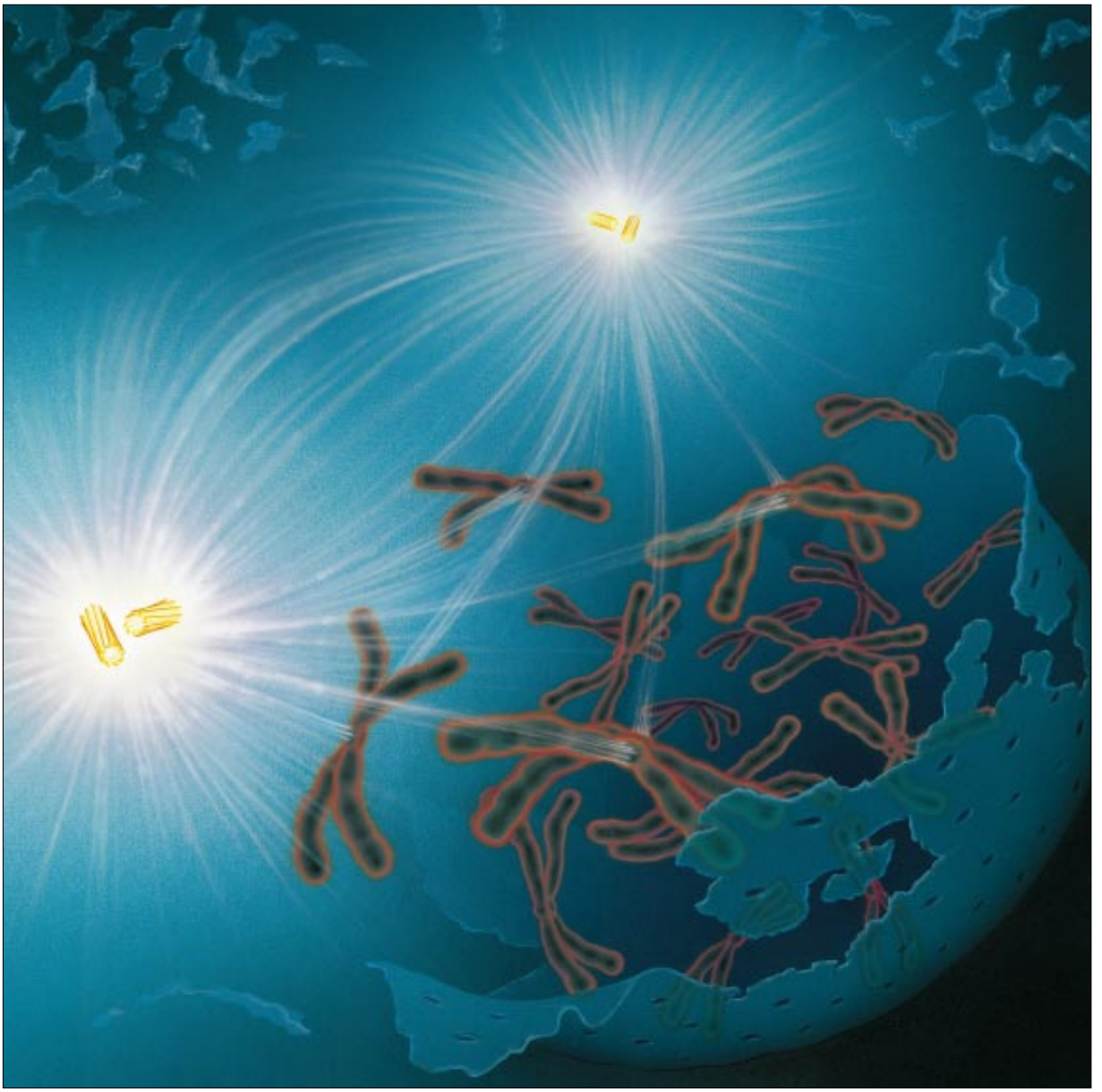
JUNE 1993
$3.95

SCIENTIFIC AMERICAN

Tuning in the radio signals of ancient galaxies.

Fossil heat: an archive of climatic change.

The dubious link between genes and behavior.



***Centrosomes** surrounded by starlike webs of protein filaments are the master architects of cell division.*

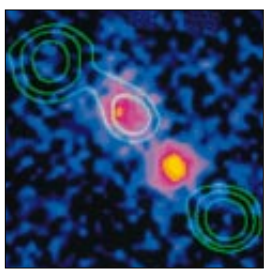
SCIENTIFIC AMERICAN

June 1993 Volume 268 Number 6

44 

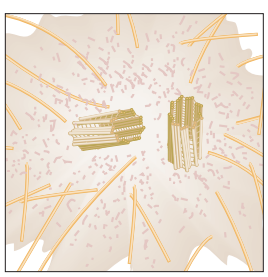
Underground Records of Changing Climate
Henry N. Pollack and David S. Chapman

Temperature readings taken over the past 150 years show that the climate grows warmer. But what was the trend before such records were kept? Ancient temperatures archived in continental crust may hold the answer. By correlating thermal gradients from boreholes with data about the composition of the primeval atmosphere, geophysicists are creating a more detailed picture of global climate.

54 

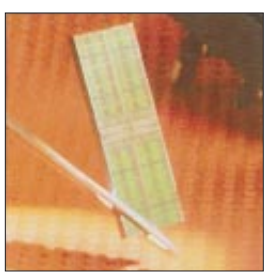
The Most Distant Radio Galaxies
George K. Miley and Kenneth C. Chambers

These blaring sources of radio waves glow with an intensity that is as much as a million times that of the Milky Way. By focusing on their powerful signals, astronomers have detected galaxies so remote that they are seen as they were when the cosmos was but one tenth its present age. Observations of these primitive objects offer clues to the formation of galaxies and the origin of the universe.

62 

The Centrosome
David M. Glover, Cayetano Gonzalez and Jordan W. Raff

The master architects of cells are organelles surrounded by asterlike blooms of fibers. By organizing the web of protein filaments that form the cellular skeleton, centrosomes govern shape, polarity and movement. During cell division, they set up the spindle that partitions the chromosomes into two daughter cells. Biologists are beginning to discover details of their structure and function.

70 

The Future of the Transistor
Robert W. Keyes

In the 45 years since the transistor was invented, the number of devices that can be packed onto a silicon chip has increased by eight orders of magnitude. And at every step of the way, critics have predicted that the physical limit to miniaturization lay just ahead. This author argues that there is still plenty of room for the trend to continue, possibly extending into the atomic realm.

100 

Monogamy and the Prairie Vole
C. Sue Carter and Lowell L. Getz

These unassuming little rodents form lifelong partnerships in which both male and female share pup rearing. Research suggests that the well-known hormones oxytocin and vasopressin play a major role in the development of this behavior. Such mechanisms may be at work in other species, including our own. The authors warn that physicians prescribing hormones should consider their behavioral effects.

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

108 

Autism
Uta Frith

For decades, parents of these tragic, isolated children have been haunted by the notion that traumatic experiences are the cause of the condition. But recent studies indicate that autism is a biological disorder. Understanding the handicap is a first step toward improving the limited lives of those afflicted with autism.

116 

The Great Well of China
Hans Ulrich Vogel

Add deep drilling to the list of technological triumphs of the ancient Chinese. One thousand years ago the inventors of gunpowder and paper sank a well 100 meters to obtain brine. Europeans did not match the engineering feat for 400 years. The crowning achievement 158 years ago was a 1,001-meter well in Sichuan.

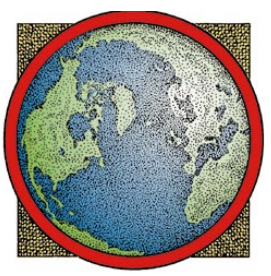
122 

TRENDS IN BEHAVIORAL GENETICS

Eugenics Revisited
John Horgan, senior writer

The century-old idea that genetics can explain, predict and even modify human behavior is back in vogue. With new molecular tools, researchers have linked such diverse phenomena as mental illness, alcoholism, homosexuality and even high intelligence to specific genes. But some of these findings have been retracted, and critics charge that the others are based on flimsy evidence.

DEPARTMENTS

18 

Science and the Citizen

The promise of an artificial pancreas.... Shades of green in Washington.... Quantum teleportation.... The legacy of the bubble boy.... How cells transport proteins.... Time lens.... PROFILE: United Nations population expert Nafis Sadik.

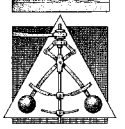
132 

Science and Business

Keeping pathogens out of the food supply.... Sporting micrographs.... Flying without halons.... Commercializing wavelets.... Photorefractive memories.... THE ANALYTICAL ECONOMIST: How reliable are economic indicators?

12 

Letters to the Editors
Racism or not?... Neither sleet, nor rain.... Reproducible wealth.

16 

50 and 100 Years Ago
1943: Can "judicious mating" eliminate nearsightedness?

142 

Mathematical Recreations
Packing problems in a sports-gear shipping room.

146 

Book Reviews
Fearing finality.... Materialistic chimps.... Planets and galaxies.

152 

Essay: *George E. Brown, Jr.*
Science must confront the new political imperatives.

THE COVER painting depicts cell division during the early stage called prophase. As the replicated chromosomes condense and the nuclear membrane begins to break down, the organelles called centrosomes migrate to opposite sides of the nucleus. The centrosomes are the centers of the starlike assemblages of microtubules. Each one contains a pair of structures called centrioles. Details of the structure and functions of centrosomes have only recently come to light (see "The Centrosome," by David M. Glover, Cayetano Gonzalez and Jordan W. Raff, page 62).

THE ILLUSTRATIONS

Cover painting by Gary Carlson

Page	Source
45	Dan Wagner
46	Roberto Osti (*top*), Jared Schneidman Design (*bottom*)
47	Jared Schneidman Design
48	Patrick Cone (*top*), Jared Schneidman Design (*bottom*)
49–50	Jared Schneidman Design
55	Alfred T. Kamajian (*top*), George K. Miley and Kenneth C. Chambers, Willem J. M. van Breugel, Lawrence Livermore National Laboratory, and F. Duccio Macchetto, Space Telescope Science Institute (*bottom*)
56	Alfred T. Kamajian (*top*), Alan Stockton, University of Hawaii (*bottom left*), National Radio Astronomy Observatory/Associated Universities, Inc. (*bottom right*)
57	Alfred T. Kamajian (*top*), Johnny Johnson (*bottom*)
58–59	Alfred T. Kamajian (*top*), George K. Miley et al. (*bottom*)
60–61	Johnny Johnson
62	Science Photo Library/ Photo Researchers, Inc.
63	Photo Researchers, Inc.
64–65	Dimitry Schidlovsky
66	David M. Glover
67	Lisa Frenz, University of Dundee
68	Dimitry Schidlovsky; David M. Glover (*photograph*)
70–71	IBM Corporation
72–74	Johnny Johnson
75–78	IBM Corporation
101	Lisa Davis, University of Illinois
102–103	Lisa Davis, University of Illinois (*photographs*)
104	Patricia J. Wynne and Michael Goodman
105–106	Patricia J. Wynne
108	Rodica Prato
109	Abraham Menashe
110	Rodica Prato
111	Jared Schneidman Design
112–114	Abraham Menashe
116–117	Enhancement by Jason Küffer (*computer*) and Tomo Narashima (*painting*)
118	Courtesy of Hans Ulrich Vogel (*top*), Johnny Johnson (*bottom*)
119–120	Michael Goodman
121	Zigong Salt History Museum, Sichuan, China
122–123	Jason Goltz
124	Bob Sacha
127	Nick Kelsh
128	E. Fuller Torrey, National Institute of Mental Health
130	American Philosophical Society
142–143	Johnny Johnson

SCIENTIFIC AMERICAN®

Established 1845

EDITOR: Jonathan Piel

BOARD OF EDITORS: Alan Hall, *Executive Editor;* Michelle Press, *Managing Editor;* John Rennie, Russell Ruthen, *Associate Editors;* Timothy M. Beardsley; W. Wayt Gibbs; Marguerite Holloway; John Horgan, *Senior Writer;* Philip Morrison, *Book Editor;* Corey S. Powell; Philip E. Ross; Ricki L. Rusting; Gary Stix; Paul Wallich; Philip M. Yam

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

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;* Tanya DeSilva, *Prepress;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems;* Leo J. Petruzzi, *Manufacturing & Makeup;* Carl Cherebin

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

ADVERTISING: Robert F. Gregory, *Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* William Buchanan, *Manager, Corporate Advertising;* Peter Fisch, Elizabeth Ryan. Michelle Larsen, *Director, New Business Development.* CHICAGO: 333 N. Michigan Avenue, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager.* WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Kate Dobson, *Advertising Manager;* Tonia Wendt. Lisa K. Carden, Lianne Bloomer, San Francisco. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

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

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; Vivienne Davidson, Linda Kaufman, Intermedia Ltd., Paris; Barth David Schwartz, *Director, Special Projects,* Amsterdam. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

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

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

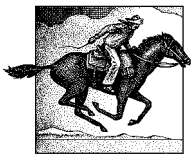
PRESIDENT AND CHIEF EXECUTIVE OFFICER: John J. Hanley

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

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

CHAIRMAN EMERITUS: Gerard Piel

PRINTED IN U.S.A.



LETTERS TO THE EDITORS

Racism in Science

The poignant truth that Howard M. Johnson describes in "The Life of a Black Scientist" ["Essay," SCIENTIFIC AMERICAN, January] was a factor that drove quite a few blacks from the U.S. Those of us who wanted to make something meaningful of our lives and saw that the prevailing climate was not conducive to intellectual growth had to make the journey elsewhere. It was either be run out of the country or out of your mind.

The profile of Rita Levi-Montalcini in the same issue offers some hope, however. Levi-Montalcini declares, "If I had not been discriminated against…I would never have received the Nobel Prize." I think also that she benefited from an engineer father, who perhaps set an intellectual standard. She may have had the luck to live in a place where there were those who realized the error of their countrymen and did all they could to provide a calmer working environment.

LAWRENCE A. ZUMO
Debrecen, Hungary

Johnson states that to succeed "as a black scientist in a white intellectual environment," one must possess an "insatiable appetite for discovery" and a "love of research" and be "ambitious" and "internally tough." But aren't those qualities required for anyone of any race to succeed?

Frankly, if Johnson experienced racial discrimination during his education and career, I couldn't find it in his essay. Whatever injustices he experienced seem to have been related more to class than to race. Although being poor gave him a slow start, no one denied him a scholarship or a job because he was black; quite the contrary, by his account. He belittles the serious problems of racism when he suggests that all problems a black man experiences in life can be attributed to the racist attitudes of whites.

JAMES M. DONOVAN
New Orleans, La.

Johnson replies:

I disagree that those qualities are shared by all. I am in contact on a regular basis with people who are academically successful but lack most of them. Further, I indicated that my primary and secondary schooling was segregated by law. What would satisfy Donovan as evidence of racial discrimination?

Violence and the Environment

Thomas F. Homer-Dixon, Jeffrey H. Boutwell and George W. Rathjens ["Environmental Change and Violent Conflict," SCIENTIFIC AMERICAN, February] dismiss one argument too readily. Although the stock of natural resource wealth is being degraded and depleted at rates unknown in history, we have also seen an unprecedented growth in other forms of wealth. No sensible person would argue that reproducible capital and knowledge can perfectly substitute for losses in natural resources and environmental wealth. It would be equally foolish, however, to argue against any substitution possibilities. Indeed, in many countries, losses in natural capital have been more than offset by gains in human and reproducible capital, although such favorable results cannot be guaranteed for all time.

In explaining conflict, it may be more useful to focus on the uneven distribution of the total wealth rather than on the distribution of particular forms of wealth. If this explanation is correct,

good social policy should support balanced efforts to conserve and enhance both natural and other forms of wealth, as well as how that wealth is distributed.

HENRY M. PESKIN
Silver Spring, Md.

The authors reply:

We suspect there are limits to the rate and extent of substitutions of knowledge and capital for renewable resources, especially in poor countries. First, the substitution task is extremely demanding: resources such as forests, good soils and abundant water simultaneously play many key roles in human-ecological systems. Second, by definition, poor countries have less knowledge and capital. Third, substitution depends on a society's ability to apply enough ingenuity at the right times and places, which depends in part on appropriate and stable social institutions, such as markets, research centers and governments. Those institutions are weak in many poor countries and vulnerable to the intense rivalries among interest groups normally engendered by scarcity. If substitutions therefore cannot be made smoothly, violence will probably increase as scarcities of renewable resources worsen.

Model Thinker

The answer to the question "How Should Chemists Think?" is very obvious: like Roald Hoffmann [SCIENTIFIC AMERICAN, February]. Without detracting from the exposition of scientific details, Hoffmann correlates science with philosophy, art, literature, history and myth and laces it with a delightful sense of humor. Such a synthesis was the aim of Renaissance Man. This is how we should all think, to make sense of the world and keep our sanity. Now I must read his poetry.

VIVIENNE HAYWARD
Stockport, England

Post Haste

"Zip Code Breakers," by Gary Stix ["Science and Business," SCIENTIFIC AMERICAN, February], discussed the difficulties of machines reading handwritten addresses and cited the figures $40 per 1,000 for hand sorting versus only $4 for machine sorting. Perhaps the U.S. Post Office is approaching the problem in reverse. Why not influence the writers of letters to provide machine-readable addresses?

I recommend that the postal service sell, at nominal cost, a hand-stamp numbering device for zip coding. There are various ways to persuade the public to use a small stamping machine; one would be to charge a penny less per letter.

PAUL H. BANNER
Chicago, Ill.

For many years, students facing multiple-choice tests have indicated their answers by filling in grids. The same technology is appropriate for mail sorting. A small grid could be marked with the zip code by the user. Envelopes with blank grids on them could be printed inexpensively, and the postal service could supply a pad of self-adhesive grids for users to mark and attach to unmarked envelopes.

RICHARD ROTHWELL
Sutton Valence School
Kent, England

Bar codes on peel-off labels or bar-coded envelopes could be sold like stamps. They would be generated by the local post office for a small fee. Alternatively, envelopes or labels could feature squares that would guide the position and spacing of handwritten characters. Using the techniques developed for reading forms, post office equipment could then read the addresses and add bar codes.

FRED FEND
Highland Park, Ill.

Because of the volume of mail, letters to the editor cannot be acknowledged. Letters selected for publication may be edited for length and clarity. Unsolicited manuscripts will not be returned unless accompanied by a stamped, self-addressed envelope.

50 AND 100 YEARS AGO

JUNE 1943

"The helium welding process renders possible and practical the fusion of inflammable metals, such as magnesium; the latter, owing to its extreme lightness, is being employed increasingly in airplane construction. To obtain the protective blanket of helium, the inventors of the process have designed a special electric torch having a hollow handle and nozzle through which the noninflammable gas can be passed. Helium has more than five times the specific heat of air and when in motion forestalls the amassing of heat around the weld. Thus the welding process is surrounded by relatively cool atmosphere, affording a better fusion and penetration with less distortion than that obtained in other welding processes."

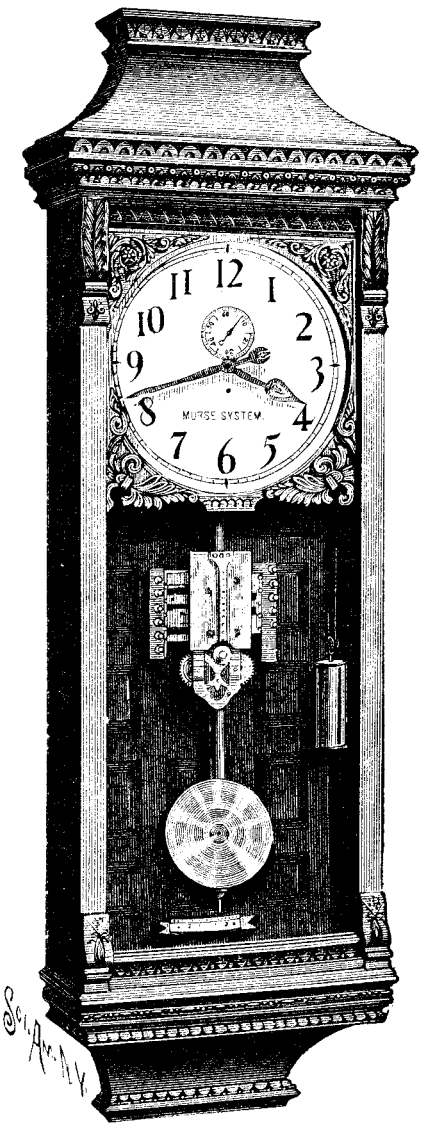
"It has long been realized that, if the stars have planets circulating around them, there is no hope at all of detecting them as we observe the planets of our own system, by reflected light. A planet twice the diameter of Jupiter and distant from the nearest star, Alpha Centauri, as far as Jupiter is from the Sun, would appear to us like a star of the 21st magnitude—that is, barely bright enough to be photographed by a 100-inch telescope, under the best conditions, if it stood alone on a dark sky. It would actually be within a few seconds of arc of its primary, whose light, a hundred million times brighter, would drown it out hopelessly. There is, however, another way in which a planet might reveal its presence. Two bodies circulate in orbits of the same shape but different sizes about their common center of gravity, keeping on opposite sides of it. (A small oscillation of Neptune, due to the attraction of its one known satellite, has actually been observed.)"

"Myopia is believed by science to be hereditary. In an address before the American Medical Association, Lawrence T. Post, M.D., St. Louis ophthalmologist, stated that 'there is little evidence to show that this is usually anything but a hereditary defect handed down just as other physical characteristics are. Continued stressing of the importance of judicious mating may result in its diminution and finally bring about its end. Even if it is impossible to bring about completely eugenic mating, it may at least be feasible to prevent the marriage of two people affected with extreme nearsightedness. Failure to do so is probably the principal reason for the very large incidence of this defect among the Germans today.' "

SCIENTIFIC AMERICAN

JUNE 1893

"An instance of rare presence of mind attended by success in the use of an antidote to poisoning occurred recently at Sag Harbor, N.Y. Flora Sterling, the five-year-old daughter of Dr. Sterling, while playing about the house found a bottle which had formerly contained citrate of magnesia and still bore the label. The child put it up to her lips and took a long swallow. With a scream she dropped the bottle and began to clutch her little throat in an agony of pain. Her father, who had heard her screams, found that what the little one had taken for citrate of magnesia was oxalic acid. Seeing that not a moment was to be lost, if he wished to save the child's life, the doctor looked about for an alkaline antidote. Seizing his penknife the doctor sprang to the whitewashed wall and scraped some of the lime into his hand. This he threw into the glass partly filled with water, and poured the mixture down the almost dying child's throat. The antidote took effect at once."



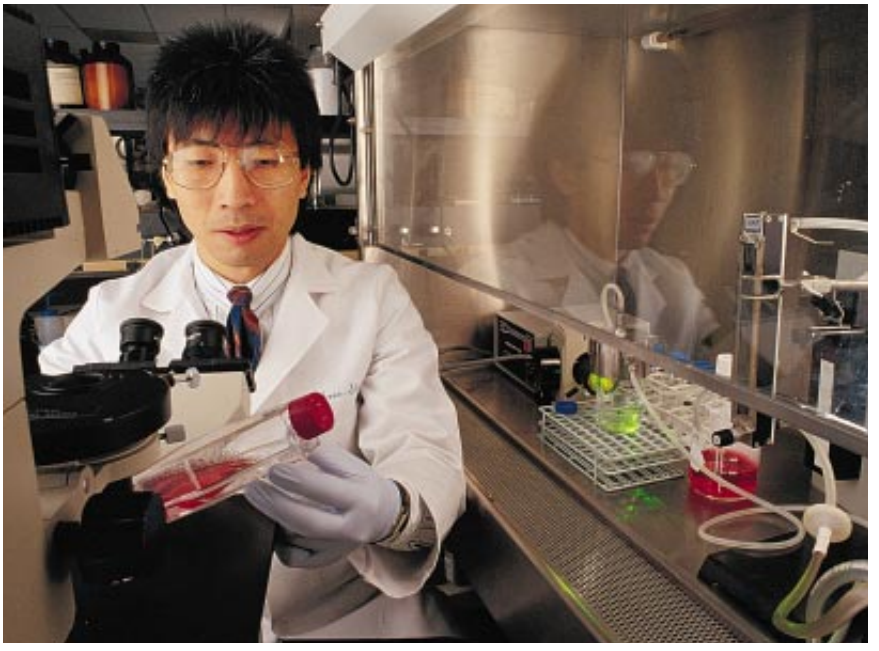
Self-winding master clock

"Professor Dewar communicated to the Royal Society on March 9 that he has succeeded in freezing air into a clear, transparent solid. The precise nature of this solid is at present doubtful, and can be settled only by further research. The doubt arises from the fact that Professor Dewar has not been able by his utmost efforts to solidify pure oxygen. Nitrogen, on the other hand, can be frozen with comparative ease. It thus becomes a question of whether the cold produced is sufficiently great to solidify oxygen, or whether its mixture with oxygen raises its freezing point, or whether it is not really frozen at all, but merely entangled among the particles of solid nitrogen, like the rose water in cold cream.—*London Times*."

"One of the most satisfactory of all the systems which have been devised for the regulation and maintenance of uniform time throughout the various rooms and buildings of a factory, or in different departments of any extended business, is that which has been perfected by the 'American Watchman's Time Detector Company,' New York. The system comprises a self-winding regulator, as shown in our illustration (*left*), to be placed in the main office or some central position, and any number of secondary clock dials placed in the various rooms and departments and electrically connected with the central regulator. The regulator is wound by electricity; that is, it is self-winding."

Living Cure

Insulin-secreting implants approach human testing



WARREN FAUBEL *Black Star*

TINY GEL CAPSULES containing human insulin-producing cells were produced by Patrick Soon-Shiong and his colleagues at Wadsworth Medical Center.

In the past year researchers have brought within reach a long-sought therapy for diabetes: an artificial pancreas. Such a device would secrete insulin in precise relation to the level of glucose in the blood, improving the management of the disease and the comfort of the patient. For years, no one could make the therapy work in animals larger than rodents, but now two groups have demonstrated its efficacy in diabetic dogs. Human clinical trials could begin as early as this summer.

The first encouraging results were published last summer by investigators at BioHybrid Technologies in Shrewsbury, Mass. That team announced in *Science* that they had weaned diabetic dogs from insulin injections for several months by implanting islets of Langerhans, warding off rejection with a semipermeable membrane. Now a group at the Islet Transplant Center, part of the Veterans Administration Wadsworth Medical Center in Los Angeles, will soon report in the *Proceedings of the National Academy of Sciences* that it may have beaten BioHybrid to the punch with a similar technique. "I have dogs going two years without additional transplants," says Patrick Soon-Shiong of the Wadsworth group.

For its part, BioHybrid has found a way to make islet therapy practical on a large scale. In the December 1992 issue of *Diabetes,* the company's investigators announced that they had successfully treated diabetic dogs with islets harvested from pigs. This species-to-species technique, called xenografting, is crucial because only 5,000 cadaver pancreases become available every year, too few to supply the two million diabetics in the U.S. who take insulin.

The Wadsworth researchers protect their islets with a gel membrane that unfortunately stimulates inflammation during the first few weeks. It therefore must be accompanied by low doses of cyclosporine, an immunosuppressive drug. Despite that requirement, the group was the first to win permission from the Food and Drug Administration to begin human trials; it plans to undertake preliminary trials in 20 human diabetics who have had kidney transplants and so already require cyclosporine. "We are scouring the West Coast" for cadavers, Soon-Shiong says.

The interest in grafts stems from their ability to do what even the cleverest human contrivance cannot do: respond rapidly to changes in the concentration of glucose in the blood. Even frequent home blood testing to fine-tune diet, exercise and dosages of insulin cannot fully normalize blood glucose. But increasing numbers of clinicians endorse this strict regimen as the best way to prevent vascular damage, blindness, kidney failure and stroke—complications that make diabetes the third-largest cause of death in the U.S., after heart disease and cancer.

Indeed, in June the National Institutes of Health expects to release the results of a nine-year study proving, once and for all, the value of near-normalization. Yet even that report may not induce many more diabetics to adopt the strict regimen, which demands great dedication. Many physicians remain suspicious of the attempt to normalize blood sugar, in part because it raises the risk of having insulin reactions—spells of low blood sugar that can lead to coma or death if not treated promptly. Islet grafts could solve all these problems.

But until recently, the immune system foiled such transplants, particularly in insulin-dependent diabetics. In this form of the disease, which usually strikes early in life, the immune system attacks beta cells, the islets' insulin makers. Efforts to preserve native beta cells with cyclosporine have failed, as have most attempts to use the drug to sustain transplants of unprotected islets. Even if cyclosporine worked perfectly, however, its side effects would outweigh the benefits in most patients.

If it is impractical to fight the immune system, then the only remaining option is to hide from it. William L. Chick, president of BioHybrid, conceived this strategy of immunoisolation more than a decade ago, when he was affiliated with the Joslin Diabetes Center in Boston. Amicon Corporation, now part of W. R. Grace, had developed an acrylic copolymer membrane whose pores block the passage of any molecule weighing more than about 50,000 daltons. That limit is large enough to allow insulin and all necessary nutrients to pass but small enough to exclude killer cells and most immunoglobulins.

Chick initially experimented with vascular shunts on the assumption that no other design could expose enough islets to enough blood to keep them all active.

Blood flows from an artery through a tube of semipermeable membrane and into a vein. Islets packed in agar surround the tube, and a plastic housing surrounds the islets. The early units could hold only enough islets to produce 15 to 20 units of insulin a day, half of what dogs and people normally require. Workers therefore had to put in two devices, cutting into four blood vessels. Still, the surgery nearly normalized the dogs for several months.

BioHybrid has also studied a less invasive containment strategy that uses strawlike chambers floating in the peritoneal cavity, where they exchange glucose and insulin with the blood vessels by way of the intervening fluids. Robert P. Lanza, a senior researcher at BioHybrid, who also holds an appointment at Harvard Medical School, says he and his colleagues have sustained dogs for many months by implanting hundreds of chambers at a time, all seeded with canine islets. "We'd like to try xenografts in large animals now," he says.

Paul E. Lacy of Washington University School of Medicine was the first to use diffusion chambers. In 1991 he normalized diabetic mice by putting the chambers under the skin—"the worst place possible" for viability, he says, although it is one of the best in terms of convenience. Today he is associated with CytoTherapeutics in Providence, R.I., which is financing his efforts to fashion the membranes into configurations calculated to house and nourish the 500,000 or more islets a human patient requires. To obtain that many cells, Lacy's associate David W. Scharp and Camillo Ricordi, now at the University of Pittsburgh Medical Center, developed a way of using enzymes to digest a pancreas into an islet-rich fluid.

The main obstacle is fibrosis: the body's attempt to wall off and destroy foreign substances. CytoTherapeutics and BioHybrid work to avoid fibrosis by making their membranes very smooth. Neocrin, a biotechnology firm backed by Baxter Healthcare, instead has tried to design a membrane that stimulates a tolerable form of fibrosis, one that leaves a space into which the capillaries can grow, nourishing the islets. Neocrin hopes to protect the islets from re-

jection by encasing them in a semipermeable gel.

Soon-Shiong pioneered such microencapsulation by using alginate, a gel derived from seaweed. To avoid islet starvation, a problem in the larger chambers, he put the cells in capsules just 600 microns wide, producing a high enough ratio of surface area to volume to facilitate the flow of nutrients. Moreover, such capsules are small enough to be injected into the peritoneal cavity by needle, a minimally invasive procedure. Unfortunately, the early capsules broke easily and often provoked fibrosis. Changes in the geometry of the gel capsules have solved some of the mechanical problems, Soon-Shiong says, and extended the life of the majority of capsules to about six months. "I'm aiming for a year between retransplantations," he adds.

Fibrosis, meanwhile, has yielded to a biochemical insight. "We found that alginate is composed of two types of polysaccharides based on mannuronic acid and guluronic acid: M-blocks and G-blocks," Soon-Shiong notes. "M-blocks stimulate interleukin-1 and tumor necrosis factors; G-blocks do not."

The workers were able to improve the biocompatibility of the capsules by increasing the ratio of G-blocks to M-blocks. To counteract the effects of residual M-blocks leaching from capsules injected into dogs, the group administered about a tenth of the normal dosage of cyclosporine. Six months later, when the dogs again required supplementary insulin injections, the cyclosporine therapy was stopped—but the implants continued to function. Some are still churning out insulin more than a year after injection. Soon-Shiong asserts that these results, together with unpublished data from more recent experiments, suggest that superpurified alginate capsules may require no drug therapy at all.

If the Phase I trials show the microcapsules to be safe and effective, Soon-Shiong intends to use porcine islets in subsequent trials. First, however, he must catch up with BioHybrid's pig-to-dog results. "Do you know where I can find a herd of pathogen-free pigs?" Soon-Shiong asks. He is not joking. *—Philip E. Ross*

A Bus for Scotty

Added to the list of weird phenomena in the quantum world is an effect that resembles teleportation. For non-Trekkers, that's the dissolution of a body or object at point A and its reconstitution at point B. An international team of investigators argues that it is possible to disembody the quantum state of a particle into classical and quantum parts and then, at another location, recombine those parts into an exact replica of the original quantum state. The convenience of this kind of transport, if fantasy for humans, seems to exist for quantum particles.

One of the architects of the scheme, Charles H. Bennett of the IBM Thomas J. Watson Research Center, reported the calculations at the March meeting of the American Physical Society. The idea makes use of the distinctions between information transmitted by classical methods and that conveyed by quantum means. Classical data, such as these words, can be observed and copied but cannot travel faster than the speed of light. Quantum news, in contrast, cannot be observed without disturbing the particle and destroying its quantum state, nor can it be copied reliably. Furthermore, quantum information under the right circumstances seems to travel faster than light.

Perhaps the most famous example of instantaneous communication is the Einstein-Podolsky-Rosen, or EPR, effect. Say an atom emits two photons simultaneously and in different directions. The photons are in some undefined quantum state until someone measures them. Discovering that the quantum state—as defined by polarization—of one photon is up fixes the polarization of the other photon into the opposite direction, or down. The effect, which is instantaneous no matter what the distance between the photons, emerges inescapably from quantum reality and has been demonstrated in the laboratory.

The EPR phenomenon is the linchpin of quantum teleportation. Alice takes one of the EPR photons and gives the other to Bob. Bob then moves to another location with the photon. Some time later Alice finds a quantum particle (another photon, for instance) whose state she wants to send to Bob. She measures the quantum state of this mystery particle with respect to that of her EPR photon. For example, Alice might find out that the polarizations of the mystery particle and her EPR photon are "perpendicular" to each other. Of course, her observation disrupts the quantum state of the system, effectively destroying the mystery particle and her EPR photon. Alice relays the relational information about her EPR photon and the mystery particle to Bob via such classical means as a telephone call or a holler.

Alice's measurement has a second, subtler effect: it forces the other EPR photon, which Bob is holding, into a definite quantum state. Bob combines the quantum information in his EPR photon with the classical message from Alice. In this way, he can transform his EPR photon into an exact replica of Alice's original, mystery particle. In essence, Bob brings back to life at another location the particle Alice killed. Bob is not simply copying information; Alice's mystery particle must be destroyed (by observing it) before Bob can resurrect it. "It is an unexpected consequence of elementary quantum mechanics," remarks Bennett, who did the work with William K. Wootters of Williams College and Asher Peres of the Technion–Israel Institute of Technology, among others.

Nothing practical is likely to emerge from quantum teleportation. Bennett explains that it is not the kind of tool for assisting communications schemes such as quantum cryptography, "but it is something that helps us understand the nature of quantum information." Indeed, no one yet knows how to test quantum teleportation in the laboratory. Bennett notes, however, that experimentalists are at least not completely discouraged. He imagines that quantum teleportation might be useful in physics experiments in which a particle is created in one place and must be measured somewhere else.

What of beaming up Scotty? "The unfortunate aspect of it," Bennett observes, "is that it makes everyone think of *Star Trek*." But the intricate and vast number of particles that make up living organisms is likely to keep transporter rooms firmly rooted in science fiction. There's always the bus.

—Philip Yam

Ecolocation

Where will the administration stand on the environment?

When President Bill Clinton and Vice President Al Gore won the election last November, environmentalists cheered. They saw Gore, the author of a best-selling book on the environment, as one of their own and a dependable ally. Chemical-based industry, traditionally at loggerheads with the green lobby, feared the worst. Both sides have been surprised in the first few months of the Clinton regime.

Certainly, the environment is assuming a larger profile. For the first time, it has an advocate on the staff of the National Security Council, in the person of Eileen B. Claussen, a former official at the Environmental Protection Agency. In another gesture that could also be of more than symbolic importance, a special commission will scrutinize the impact of the North American Free Trade Agreement. And green types were gratified that the president's proposed—and now failed—"economic stimulus package" included spending on water treatment plants. Many of the administration's appointments have also pleased the environmental lobby.

Nevertheless, the "green group," an informal coalition of environmental organizations, has been dismayed by several of the administration's political compromises. Chief among them has been Clinton's capitulation on increasing fees for mining and grazing on federal land and phasing out subsidies for logging in federal forests. When Senator Max Baucus of Montana, chairman of the Senate Environment Committee, indicated that his colleagues from west of the Mississippi might have misgivings about the promised land-use reforms, Clinton quickly agreed to take them out of his proposed fiscal 1994 budget. At risk, the president feared, was his economic program.

Although Clinton promised the measures will be introduced administratively and in legislation, many in the green group feel the fumble has lost him the political initiative. "I will predict that a 12.5 percent royalty on mining will not be included in a mining reform bill coming out of the Senate," says D. Reid Wilson, political director of the Sierra Club. And like-minded leaders worry that Clinton is softening his campaign pledge to freeze emissions of carbon dioxide—a probable cause of global warming—at 1990 levels by 2000.

Environmentalists were also startled by Clinton's decision to abolish the Council on Environmental Quality, which

has adjudicated environmental arguments between executive departments. When the green group, which consists of relatively conservative organizations such as the Izaak Walton League of America as well as more radical groups such as the Sierra Club and Friends of the Earth, registered its complaint with the president, its leaders were summoned to a meeting with a furious Vice President Gore. He "read the riot act and told them to get out of the way," according to Marchant Wentworth of the Izaak Walton League. Wilson, who worked for Gore during his tenure as senator and during his presidential bid in 1988, says he has rarely seen Gore more angry. The bill that would abolish the Council on Environmental Quality would also enact the president's plan to elevate the EPA to a cabinet-level department. Gore apparently felt snubbed by the rebellion. Wentworth sees the spat as political amateurishness.

Even so, Wilson says relations between the green group and the administration are better than they were under former president George Bush. "Now we sometimes agree to differ—before, we agreed not to talk to each other," he points out. Meanwhile the administration seems to be treading carefully to avoid making enemies in the world of commerce. Chemical manufacturers say they are encouraged by the professed willingness of the EPA's new administrator, Carol M. Browner, to institute "a new era in communication between the EPA and America's business community," as she put it in her Senate confirmation hearing. "We see hopeful signs that our relationship with the EPA will be less confrontational," says John F. McCarthy, a vice president of the National Agricultural Chemicals Association.

Robert J. Hirsch, chair of the committee on energy, environment and natural resources of the National League of Cities, echoes that opinion. Hirsch says his committee is accustomed to battles with the EPA over the cost of regulations. In March, however, negotiations between the EPA and the league seemed to have concluded satisfactorily with an agreement about levels of contamination by disinfectants.

The major battles that will reveal the true shade of green in the Clinton administration have yet to be joined, however. Those will be the solid waste act (known as RCRA, for Resource, Conservation and Recovery Act), the clean water act and the Superfund act, which are all up for reauthorization.

Industry is mobilizing, and Superfund is the principal target. Some $10 billion has been spent on the program since 1980, which was intended to rectify past abuses by cleaning up contaminated sites even if the guilty parties could not be found. Yet most of the monies the program has spent have gone into lawyers' and consultants' pockets. Only 47 Superfund sites have been fully cleaned up, while 1,275

Mr. Clinton, Put Down That Watering Can

President Bill Clinton would not be flattered. When he announced his $17-billion technology initiative this past March, many Europeans dismissed it as just another example of the "watering can" approach to nurturing new technology and fostering industrial competitiveness. Why the lukewarm response to the idea of showering money on critical technologies? The European Community's plunge into similar industrial policy under the rubrics of Esprit, Eureka and Race has not produced a tangible return.

So now the EC is nurturing a hot, new idea bearing the buzz name "megaproject." Whereas past EC research projects brought together researchers and manufacturers, megaprojects would tie research more tightly to market needs by not only enlisting potential users as collaborators but actually putting them in a leadership position. In the case of designing computer networks for health care, for example, hospitals and health authorities would define goals and direct a cluster of research projects charged with developing standards and technology. And rather than stop at the demonstration of feasibility, megaprojects would even go so far as to build factories.

Up to this point the Commission of the European Community in Brussels has talked only informally about setting up megaprojects, citing such applications as computer networks. The idea, however, has received widespread support from industry and research policy officials and is expected to form the centerpiece of the fourth phase of the commission's research programs, called Fourth Framework, beginning next year.

With Europe's flagship high-tech companies—most notably Groupe Bull, Siemens, N.V. Philips and Olivetti—losing money, a shift in thinking was a political necessity for the commission. The lingering recession has made it difficult to argue that past programs have had an effect on competitiveness. According to Nigel Horne, a special adviser at KPMG Peat Marwick and an adviser to the commission, "the time has come when we should expect more from research than progress on a broad technological front."

Much of the impetus behind these policy proposals has come from dissatisfaction with the results of previous research efforts. The Esprit program's original goal in 1985 was merely to foster research collaboration. Since then, critics of the program have succeeded in convincing the commission to sharpen project definitions and to require tangible "deliverables" every few years. Despite these efforts, however, Esprit has never proved its effectiveness in improving the crucial linkage among the research and development and marketing departments of large corporations. Similarly, the Race program, created in 1988 to develop transnational broadband communications networks, fell short of its goal of implementing the networks. "Race has done a good job in certain technology areas," says John Forrest, chief executive of National Transcommunications in Winchester, England, "but the vision has gotten nowhere."

Industry seized on these shortcomings as evidence that programs should be selected that have greater "market pull." A review of EC projects showed that some of the more peripheral ones targeted at specific industries such as health care and air-traffic control had the best record. The notion of combining the pragmatism of these efforts with the technological depth of Race and Esprit took hold.

The current state of policy limbo in Brussels is tempering optimism for the megaproject concept. The commission is only just finding its feet after a massive, 18-month reorganization. Uncertainty over the fate of the Maastricht Treaty, which will not be ratified before the summer, if at all, has put a hold on any formal proposals. As a result, details about how to structure and pay for the megaprojects are now the subject of vigorous behind-the-scenes lobbying. Nevertheless, the idea has enough impetus that megaprojects may soon become the new paradigm for European competitiveness. *—Fred Guterl, London*

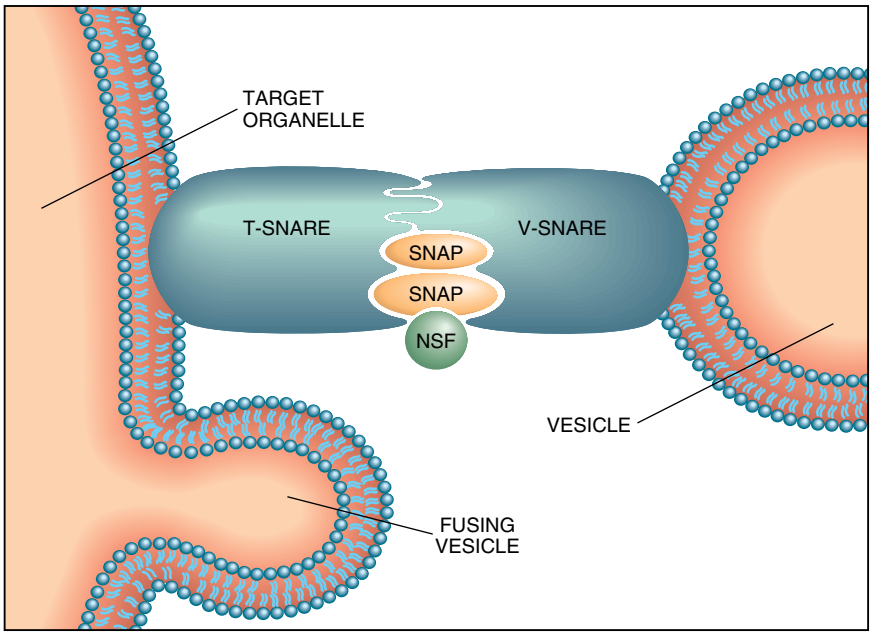
remain on the National Priorities List.

Industry was heartened by Clinton's comment at his pre-Inauguration economic summit in Little Rock that he was "appalled by the paralysis and the political divisions and the fact that the money's being blown" on Superfund. Frank Popoff, chairman of the Dow Chemical Company and of the board of the Chemical Manufacturers Association, has written to Clinton that Superfund's "harshly punitive nature" is what "warps the cleanup remedies…and has fostered the litigious climate." The "retroactive strict, joint and several liability" principle in the act means that anyone who has ever polluted a site can be held responsible for all cleanup costs.

But the legal aspects are not the only ones that will come under scrutiny. For some years, the EPA has been assessing the methodology it uses to set limits for toxic chemicals in fresh foods and the environment in general. Although all sides in the debate agree that animal tests will continue to be a principal basis for screening compounds for carcinogenicity, many scientists question the current standard technique for estimating those risks. Animals are now fed the maximum dose of a chemical that they can tolerate and then examined for malignancies. The EPA has circulated draft proposals for a scheme that would allow it to consider everything that is known about why a chemical is toxic as well as its observed carcinogenicity. The effect would be to reduce the number of chemicals listed as probable carcinogens.

The EPA is also collaborating informally with Senator Daniel Patrick Moynihan of New York, who is drafting legislation that would encourage the EPA to consider relative risks when making all types of regulations. Although Browner has not yet formally endorsed Moynihan's proposals, she has expressed doubts about the Delaney clause, a 1954 law that bans food additives that can cause cancer in laboratory animals, regardless of the size of the risk. "The thrust of the new thinking is that we should be able to distinguish big risks from small risks," says Donald G. Barnes, the current head of the EPA's science advisory board.

EPA officials have acknowledged that they must consider complexities such as the distribution of risk across different sectors of the population and the degree of voluntary control over exposures. But the continuing press to reform is another sign that the Clinton administration is seeking a broader consensus on rational policy, to end the standoff between the engine of economic recovery and the green lobby. —*Tim Beardsley*



LAURIE GRACE

VESICLE FUSION inside cells is mediated by specific combinations of SNARE, SNAP and NSF proteins, according to one new model.

SNAPs and SNAREs

Protein hooks help vesicles grab cell membranes

Many essential proteins in the cells of higher organisms are ferried from one organelle to the next inside small membrane packages. When they arrive at their target, these vesicles merge with the membrane they find there, an event called fusion. Growth, secretion and other vital processes all depend on this complex phenomenon. But details of this aspect of intracellular protein transport have been slow to emerge. Biologists still do not entirely understand how the vesicles recognize their destination or how they incorporate themselves into another membrane.

That situation has begun to change because of work by James E. Rothman and Thomas Söllner and their colleagues at the Memorial Sloan-Kettering Cancer Center in New York City. They have identified cellular proteins that seem to control fusion mechanisms in all eukaryotic (complex) cells, from yeast to humans. Moreover, the same proteins seem to be involved both in fusion events that occur spontaneously and in those that are regulated, such as the release of neurotransmitters from brain cells. "So we have a finding that unites several different fields," Rothman observes. "Seemingly different questions in cell biology and neurobiology are revealed to be the same question"—a neat feat of fusion in itself.

In the past few years the Sloan-Kettering researchers have determined that certain cytoplasmic proteins—*N*-ethylmaleimide-sensitive fusion (NSF) protein and soluble NSF attachment proteins (SNAPs)—are essential for membrane fusion inside mammalian cells. It quickly became apparent that the NSF and SNAP proteins were identical to those in yeast that served a similar purpose and had been identified by Randy W. Schekman of the University of California at Berkeley.

Those discoveries were good news in that they showed all eukaryotes used NSF and SNAPs. At the same time, Rothman notes, they created "kind of a paradox" because both NSF and SNAPs are very general components of the intracellular fusion machinery. "Yet there must be extraordinary specificity in these fusions," he adds. The fact that vesicles do not fuse randomly with the wrong membranes "suggests that there is some kind of targeting mechanism." Rothman and his colleagues therefore set out to look for more fusion-related molecules on cell membranes.

Working with extracts from neurons, they recently isolated four membrane proteins that act as the attachment points for SNAPs during fusion. Rothman says, "We call them SNAREs, both because it's short for SNAP receptors and because a snare is a trap for small game." The "game" here is microscopic: the SNAREs, SNAPs and NSF form a particle that presumably allows vesicles and their targets to fuse.

Rothman and his colleagues initially assumed that they were the first to isolate SNAREs. To their surprise, however, Sloan-Kettering chemist Paul Tempst helped show that all four SNAREs had previously been identified as components of the synapses between neurons. Although the functions of those proteins had not been known, neurobiologist Richard Scheller of Stanford University had found one of them on the vesicles that contain neurotransmitters and two others on the surface membrane of the neuron. (The position of the fourth has not been determined precisely.) Similar proteins had also been found in the organelles of yeast.

To Rothman and Söllner, the placement of the SNAREs suggested a model that linked the targeting and fusion mechanisms. The proteins may be of two types: v-SNAREs (those on the vesicles) and t-SNAREs (those on the target membranes). "The seductive proposal," Rothman says, "is that every vesicle carries a particular v-SNARE that pairs it with a t-SNARE found only on the appropriate target membrane." In the presence of NSF and SNAPs, interactions between the right v-SNAREs and t-SNAREs may stabilize the association of vesicles and their targets long enough for fusion to begin.

Because the same components of the fusion machinery appear throughout the eukaryote kingdom and in regulated and unregulated fusion processes, the same mechanism is almost certainly at work everywhere. "This is one area in the membrane field in which there have been very few insights until now," Rothman remarks. As he and his co-workers reported this past March in *Nature,* cells may regulate some types of vesicle fusion by modifying SNAREs or other parts of the fusion complex. Scheller has noticed that the t-SNARE referred to as syntaxin associates closely with calcium channels in neural membranes; calcium fluxes are known to trigger the fusion of neurotransmitter vesicles.

Schekman hails the new hypothesis as "very attractive." He readily acknowledges that the riddle of vesicle fusion is not yet solved. "It's getting close," he says, but researchers still have not determined which component of the fusion complex causes the membranes to merge with one another. It might be one of the identified molecules, but it could also be "a separate entity that is recruited only after the fusion complex has formed or after NSF and the SNAPs have left the scene. So there are plenty of open questions." Still, the discovery of SNAREs tightens the noose considerably. —*John Rennie*

Cosmic Diagnosis

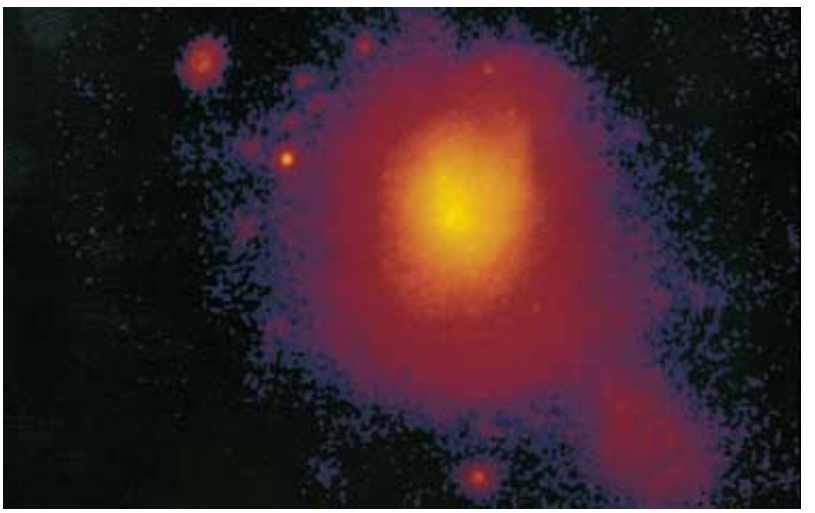
Like doctors, astronomers are finding that x-rays offer an invaluable means for examining otherwise hidden structures. Last year Trevor Ponman and his colleagues at the University of Birmingham in England announced that x-ray observations of hot gas in the Coma galaxy cluster show that the cluster's mass follows a surprisingly complicated, lumpy distribution. "It supports the notion that clusters have grown by the accumulation of blobs of galaxy groups and that the process is still happening now," Ponman explains. That discovery is especially significant because the Coma cluster, located 300 million light-years away in the constellation Coma Berenices, is the nearest and one of the best-studied rich clusters of galaxies.

Simon D. M. White of the Institute of Astronomy at the University of Cambridge and his collaborators have since amplified and expanded on Ponman's findings. Using data collected by the *Roentgen Satellite* (*ROSAT*), White's group has produced an x-ray image of the Coma cluster revealing unprecedented detail (*below*). White describes his work as "x-ray archaeology" because it enables him to reconstruct the process by which the Coma cluster came together. "It's fairly clear that you can see the remnants of previous subclumps," White says. The bright extensions of the cluster, most clearly seen at the bottom right, consist of hot gas surrounding giant galaxies that probably were once the dominant objects in their own, smaller clusters before being swallowed and merging into Coma.

The perceived structure of the Coma cluster fits well with leading ideas regarding the origin of cosmic structure, which hold that such vast clusters of galaxies form by capturing and absorbing smaller masses. Alternative cosmological models, in which clusters such as Coma originate all of a piece, look increasingly unappealing given the current data, White notes.

Not all is necessarily rosy for the theorists, however. X-ray observations of galaxy clusters enable astronomers to calculate the total mass of those clusters and to determine what fraction of that mass consists of ordinary matter ("baryonic matter" in the scientific argot); the remainder must be the mysterious dark matter. White finds that in the inner regions of the Coma cluster, 11 to 35 percent of the mass is ordinary matter. The favored cosmological models predict that the fraction of ordinary matter should be much lower, "by about a factor of five," he says. "In my opinion, that's a major discrepancy."

So where is all the dark matter hiding? A group led by John S. Mulchaey of the Space Telescope Science Institute in Baltimore used another set of *ROSAT* data to argue that it may be tucked away in clusters much smaller than Coma, a conclusion Ponman considers "a bit dodgy." Then again, White points out that the fault could lie in the x-ray data or in an improper understanding of how galaxy clusters coalesce, how dense the universe is or even how the universe began. In the intellectual realm, as in the physical, upheaval seems to be the rule. —*Corey S. Powell*



SIMON D. M. WHITE, ULRICH G. BRIEL and J. PATRICK HENRY

X-RAY-EMITTING GAS traces out mass in the Coma galaxy cluster.

David's Victory

Gene causing "bubble boy" illness is finally found

Severe combined immunodeficiency (SCID) is its name, but most people think of it as "bubble boy" disease. Born without an immune system, a Texas child known publicly only as David was mortally vulnerable to even the mildest infection. He spent all 12 years of his short life inside protective sterile rooms and a miniature space suit. David died nine years ago, leaving behind many mourners and cultures of his cells that have been nurtured by SCID researchers.

Now, working with DNA from those cells, a team of investigators led by Warren J. Leonard of the National Heart, Lung and Blood Institute has discovered a common genetic cause of SCID. The work holds the promise of better diagnostic tests and treatments for the rare disorder; it also seems likely to help piece together more general puzzles about the genesis of the cells of the immune system.

SCID occurs in about one out of every 100,000 live births. Medical researchers have long known that in about half of those cases, the genetic defect responsible for the disease lay somewhere on the X chromosome. That form of SCID occurs exclusively among boys, who have only one X chromosome. Girls, who have two X chromosomes, remain healthy but can eventually pass SCID on to their sons. Boys like David, who exhibit X-linked SCID, possess virtually none of the white blood cells called *T* lymphocytes that defend the body from disease.

The new work by Leonard and his collaborators reveals that X-linked SCID (X-SCID) is caused by an abnormality in the gene that makes the gamma-chain subunit of the receptor for the cytokine interleukin-2. This receptor protein, which is made of alpha, beta and gamma chains, sits on the surface of cells in the immune system. Its function is to bind with circulating molecules of interleukin-2, a chemical signal that cues lymphocytes to grow and divide during immune responses. Because their receptor is defective, cells in X-SCID patients cannot bind to interleukin-2. Moreover, for reasons that are not entirely known, the defect in the gamma chain apparently impairs the generation of their *T* cells.

GAMMA-LIAISON NETWORK

DAVID THE BUBBLE BOY had to live in a germ-free environment because of a rare genetic condition that left him without an immune system. Using DNA derived from his cells, researchers have now found the ultimate cause of his ailment.

"We were not a lab that was working to find the cause of X-SCID," Leonard points out. Rather he and his collaborators Masayuki Noguchi and Stephen Adelstein were engaged in basic research that paid a clinical dividend. Last summer Japanese researchers announced that they had cloned DNA that encoded the gamma chain. Leonard and the other members of his laboratory, who had long studied the interleukin-2 receptor and its subunits, were attempting to learn more about the gamma-chain gene. In collaboration with William S. Modi and O. Wesley McBride's group at the National Cancer Institute, they mapped it to a position on the X chromosome. To their pleasure, they realized that previous genetic studies had implicated roughly the same part of the chromosome in SCID.

They decided to test the hypothesis that defects in the gamma-chain gene were causing the immunodeficiency. With the further assistance of Howard M. Rosenblatt of the Baylor College of Medicine and Alexandra H. Filipovich of the University of Minnesota, the researchers looked at DNA derived from David and two other SCID patients. All three, they found, had mutations in the gamma-chain gene. "Each of them had a different mutation," Leonard summarizes, "but the bottom line was that each of the mutations resulted in a defective interleukin-2 receptor gamma chain."

Conceivably, better knowledge of the gene defect underlying X-SCID will someday improve treatment. Currently SCID patients can sometimes be restored to health with bone marrow transplants from compatible donors. Genetic therapies that could correct or compensate for the gamma-chain problem might also be possible, although Leonard notes that they will probably take years to develop. "The application I hope will be available much sooner is better diagnosis," he adds. In theory, if genetic analyses became sufficiently easy and inexpensive, physicians could identify the specific gamma-chain mutation in an X-SCID patient and then screen his female relatives to determine whether they are carriers of the trait. Those tests could be of value in family-planning decisions and prenatal diagnoses.

What the recent X-SCID discovery reveals about the development of the immune system may ultimately be at least as significant as its clinical applications. As Leonard and his co-workers discussed this past April in *Cell,* a few human patients are known to acquire SCID because of mutations in the gene for interleukin-2 itself. Whereas the people without a complete interleukin-2 receptor lack *T* cells, those without interleukin-2 seem to have a normal complement of *T* cells, albeit unresponsive ones. Those findings are perplexing: one might expect that both types of disruptions of the interleukin-2 response system would have the same effect.

One possible explanation, the researchers have speculated, is that the gamma chain may also be a component of other cytokine receptors. If so, the loss of a functional gamma chain may interfere broadly with intercellular signaling that is essential to the differentiation and maturation of *T* cells. No direct evidence yet shows that this is the case, Leonard emphasizes, but the model has precedents: for example, the receptor proteins for the interleukin-3 and interleukin-5 cytokines share the same beta-chain subunit.

Whether or not that theory proves correct, it seems certain that further studies of the gamma chain and SCID will deepen understanding of the mechanisms of immune system development. Few 12-year-olds have ever left so rich a legacy. —*John Rennie*

Time Warp

Resonating crystals squeeze light beams into pulses

Activate the time lens" sounds as if it should stand right next to "Reverse the polarity of the neutron flow" in a gallery of bad science fiction dialogue. Instead it describes an optical trick that Asif A. Godil and his colleagues have been performing regularly for about a year in a physics laboratory at Stanford University.

The Stanford time lens is a lithium niobate crystal, which can change its refractive index—and thus the speed of light waves traveling through it—in response to an electric field. A microwave cavity surrounds the crystal, setting up an oscillation that alternately delays and accelerates segments of a light beam traversing the lens. When a 30-picosecond pulse travels through the lens, the leading waves are held back and the trailing ones eased forward until they are less than two picoseconds apart. Previous pulse-squeezing techniques have relied on light pulses containing a range of wavelengths, but the time lens can operate as easily on monochromatic light.

Although initial tests of the time lens employed it to focus light at a single point in time, the device can also stretch out and thereby magnify short pulses, says David M. Bloom, a professor of electrical engineering who works with Godil. Events that take place too fast for sensors to capture could be stretched out and studied in detail.

Indeed, Michael T. Kauffman, also of Stanford, recently devised a variation of the time lens that eliminates the need for high-speed electronics to study short pulses. As the time lens speeds up or slows down the crests and troughs of a light pulse, it reduces or increases the wavelength of different parts of the pulse, converting time differences to wavelength (or frequency) differences that can be measured by spectrograph. Eventually, Bloom predicts, it may be possible to study chemical reactions and other processes that last just a few femtoseconds using only time lenses and the equivalent of a simple prism. —*Paul Wallich*

A Powerful Voice for Women

Like a jet stream, Nafis Sadik girds the globe, often sweeping through six countries in a month while she talks tomes, seemingly without pausing to breathe. Given Sadik's subject matter—sex, abortion, women's rights, exploitation of children, environment—she needs nothing less than the wind behind her.

Because she is executive director of the United Nations Population Fund, Sadik's energy and expertise are in demand now more than ever. After a decade or so of politically enforced quiet, concerns about population growth have resurfaced. Estimates of one billion more people on the planet by the end of the century, persistent poverty, mass migrations and environmental degradation catalyzed discussion at the 1992 Earth Summit in Rio de Janeiro. Publicity about efforts to stifle such talks also stoked the debate.

While some issues appear to have cooled after the exodus of delegates from Rio, family planning remains hot. In 1994 the International Conference on Population and Development in Cairo will extend the dialogue. In addition, President Bill Clinton has promised to restore funding that was cut off in the mid-1980s for family-planning programs—including those at the U.N. Just as significantly, the Clinton administration has made it clear that subjects such as a woman's right to abortion are no longer taboo.

Throughout the vagaries of public and political opinion, Sadik's voice and message have been unwavering. When she arrived at the U.N. in 1971, "population was not discussed so openly. U.N. organizations were uncomfortable if you talked about women's health and family planning. And it has taken a while to get over that embarrassment," she observes. "But I suppose they get comfortable with someone. I mean, they *know* I am going to talk about it, so they get used to hearing it."

Sadik, whose own life has both embraced and overturned tradition, seems well suited to tackle the subject of sexuality and women's rights: she is direct but diplomatic, and her occasional monotone suggests she could blunt the thorns of any prickly topic. Her office at U.N. headquarters in New York City reflects her work and travel. Statues, plates, paintings, photographs and bibelots from Africa, Japan, the Middle East and, in short, everywhere else can be found in all corners of the room. Yet there is also a composed, public aspect to it: reference and family-planning books are arranged on her desk so that the titled spines face visitors.

ROBERT PROCHNOW

NAFIS SADIK, executive director of the U.N. Population Fund, notes "all the preferences in our society are for men."

Sadik, who continues to wear traditional Pakistani attire and whose voice preserves the cadences of Urdu, was born in Jaunpur in 1929 to a conservative Islamic family. But her father, a finance minister and former vice president of the World Bank, did not share the common view that women must marry and raise children. "He was a visionary, and he believed in educating girls and boys, because, you know, in our part of the world girls are often not educated," Sadik notes. "And all the family members kept saying, 'Oh, you are going to send your daughter to work, how terrible. Why are you sending her to college?' "

After completing high school, she considered two professions: engineering and medicine. "But then I decided that the world was not ready to accept women engineers." So she entered Dow Medical College in Karachi and, because her most inspiring teacher was a obstetrician and gynecologist, went on to specialize in women's health. Her international disposition also took shape at that time. She did her internship at City Hospital in Baltimore.

Sadik returned to Pakistan and in 1952 married a businessman, Azhar Sadik, and practiced obstetrics and gynecology in the towns where her husband's work took them. The contact with women in small, rural communities made explicit to her the link between family planning and the status of women, a link that shaped her career. "The role of women is seen only as reproductive, even if they do many other things," she explains. "When I would tell a woman after her most recent child, 'Now you must have proper spacing between this child and the next,' she would say, 'Oh, I can't do that because of my husband,' or 'My family won't allow it.' Especially if she had had a daughter, there was pressure to have a son."

Her already emphatic voice gains urgency as she describes the situation of her patients. "They were really burdened. I mean this childbearing was just like they were machines for having children," she recounts. "Their life was like a continuing bondage, and it still hasn't changed all that much. Most of the women in the rural areas have that same cycle, and they teach the same values to their children. They teach their sons to order; they teach their daughters that they must serve even their brothers."

So Sadik began trying to provide family-planning services to the women she treated. "At that time, only condoms and diaphragms were available, and some of these women had infections, so the diaphragm was not suitable," she recalls. "To get condom usage, you had to get the husbands to agree. I had to call them in and say, 'You have to make sure that your wife doesn't get pregnant.' "

Unexpectedly, Sadik found that most couples did follow her advice. "It meant quite a lot of hard work, persuasion and coaxing," she says. But, in the end, "if one of the women became pregnant, her husband was quite embarrassed about it." The idea that men and women must work together remains central to Sadik and to her conception of family planning. "I have some disagreement with the idea that only women can control everything," she states. "I think there has to be a proper definition of roles and a collaboration and a cooperation between women and men."

As a result of her fieldwork, Sadik joined Pakistan's national family-planning service in 1964 and, ultimately, became director of the agency. In 1971 she came to the U.N. Population Fund, then in its third year. Despite the problems of relocation and of finding a job, her husband said it was his turn to follow her. "If it had been someone else, who had said no, I am sure I would have gone back. I wouldn't have stayed here," she muses. "He was very liberal in his attitudes and had no hang-ups about my working and doing whatever I wanted."

Despite the hurdles that she had overcome while seeking an education in Pakistan, Sadik describes being taken aback by the atmosphere at the U.N. "When I first joined, I thought the U.N. was not very forthcoming as far as women were concerned," she remembers. "I found that I had better respect in Pakistan." In order to be heard, Sadik says she had to repeat herself aggressively. An idea would be picked up if a man in a meeting presented it, even though "I might have already said the same thing, and it had been ignored."

In 1987 she was appointed head of the fund, becoming the first woman to be made director of a U.N. agency. This time, however, no extra assertiveness was required. "For many years, I was the only woman in the group, and I got special attention paid to what I said. After a year, other people would talk about population issues or women's issues, and then they would look at me to see if I had heard them," she laughs.

In the more than 20 years that Sadik has been at the U.N., the Population Fund's budget has grown from $3 million to $250 million (all contributions are voluntary). The number of countries with U.N.-supported family-planning programs has expanded from about three to 135. During the same period, global fertility rates have fallen from 6.1 to 3.4 children per woman. The agency continues to make family-planning services available and to support maternal and child health programs and education, as well as to collect data on fertility and population.

They teach sons to order; they teach daughters to serve even their brothers.

Although the role and the budget of the fund have expanded, the organization has experienced setbacks. When the agency was established in 1969, the U.S. was a major sponsor. But in 1984, at the second world conference on population in Mexico City, U.S. policy changed drastically. President Ronald Reagan (and, later, President George Bush) blocked money for any group that provided abortions or counseling about abortions. Immediately after, charging that the U.N. fund was involved in coercive family-planning programs in China, the U.S. dropped $10 million of its annual support.

The Clinton administration has promised to reverse this policy and to resume allocations. Sadik says she is pleased, of course, by the renewed U.S. support and by the more open attitude toward abortion. Indeed, abortion is one of the topics slated for discussion in Cairo. "Half of the [500,000] maternal deaths each year are the result of unsafe and illegal abortions," she points out. "In 1984 it was said that abortion was not to be used as a method of family planning. But that is not the issue here. Abortion should be safe, and the lack of services should not result in the deaths of women." She is prepared for a fight.

That Sadik can turn a controversy to advantage—or at least not be buffeted about by it—is quite clear. By now the story of population at the Earth Summit has been well chronicled. The topic was used as a bargaining tool and was absent from the initial discussions. Developing countries did not want to be blamed for overpopulation or to talk about controlling their growth rates; developed countries did not want to discuss their megaconsumption of resources. After population was finally introduced, Agenda 21—a document described as a blueprint for environmental policy and development in the next century—was altered to satisfy representatives from several Catholic countries, the Vatican and some women's groups. (The women's organizations objected to the suggestion of an association between environmental degradation and women.)

The changes in the text and the late appearance of the subject made for great drama. Government leaders and the media discussed the fact that the planet gains 250,000 people every day and that the population, currently 5.4 billion, is expected to double by 2050. Although Sadik said in Rio that some people attributed the blitz to her maneuvering, she demurred, saying the Vatican deserved all the credit.

Since Rio, Sadik has been planning the Cairo conference. The emphasis will be on population and economic growth. She asserts that the involvement of nongovernmental organizations, an important component of the Earth Summit, is vital. These special-interest groups are often considered closer to communities than are national or federal agencies. "Working with them is a better way to identify people in need," Sadik declares.

Sadik hopes the relation between environment and population growth, which was taken as a priori in Rio, can be made more explicit and that countries reluctant to define sustainable development can be forced to do so. "They are going to have to think about a standard of living that may include a minimal level of education, health and employment. But not necessarily that everyone is going to be rich and jet around the world," she cautions. "The developed countries have to think about how long they can keep using the world's resources out of proportion to their numbers."

But at heart, the focus for Sadik remains the same. "You have to address the root cause, which is the low status of women," she urges, the speed and momentum of her speech as forceful as they were an hour ago—and as they will be in another hour. "All the preferences in our society are for men. That has to be changed to make it equal." —*Marguerite Holloway*

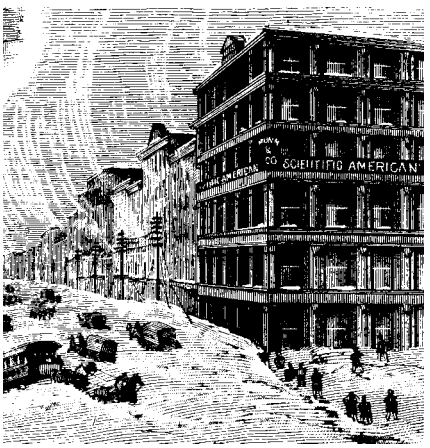
SCIENTIFIC AMERICAN

Underground Records of Changing Climate

Boreholes drilled into continental rock can recover fossil temperatures that reveal the climate of past eras. The results require careful interpretation

by Henry N. Pollack and David S. Chapman

Is the earth's climate growing warmer? Persuasive evidence exists to support the proposition. According to meteorologic records, the mean temperature of the atmosphere has increased by slightly more than half a degree in the past century. Preserved air samples and other data show that levels of gases that trap the earth's heat have also risen during this period. The proportion of carbon dioxide in the atmosphere has risen by more than 20 percent and that of methane has roughly doubled. This correlation suggests a possible cause for the apparent effect. The proposition seems reasonable that the greenhouse gases are responsible for the warming trend. Yet the case is not airtight. It is conceivable that the matching increases in temperature and greenhouse gases are a statistical coincidence and that the two variables have nothing to do with each other in the long run.



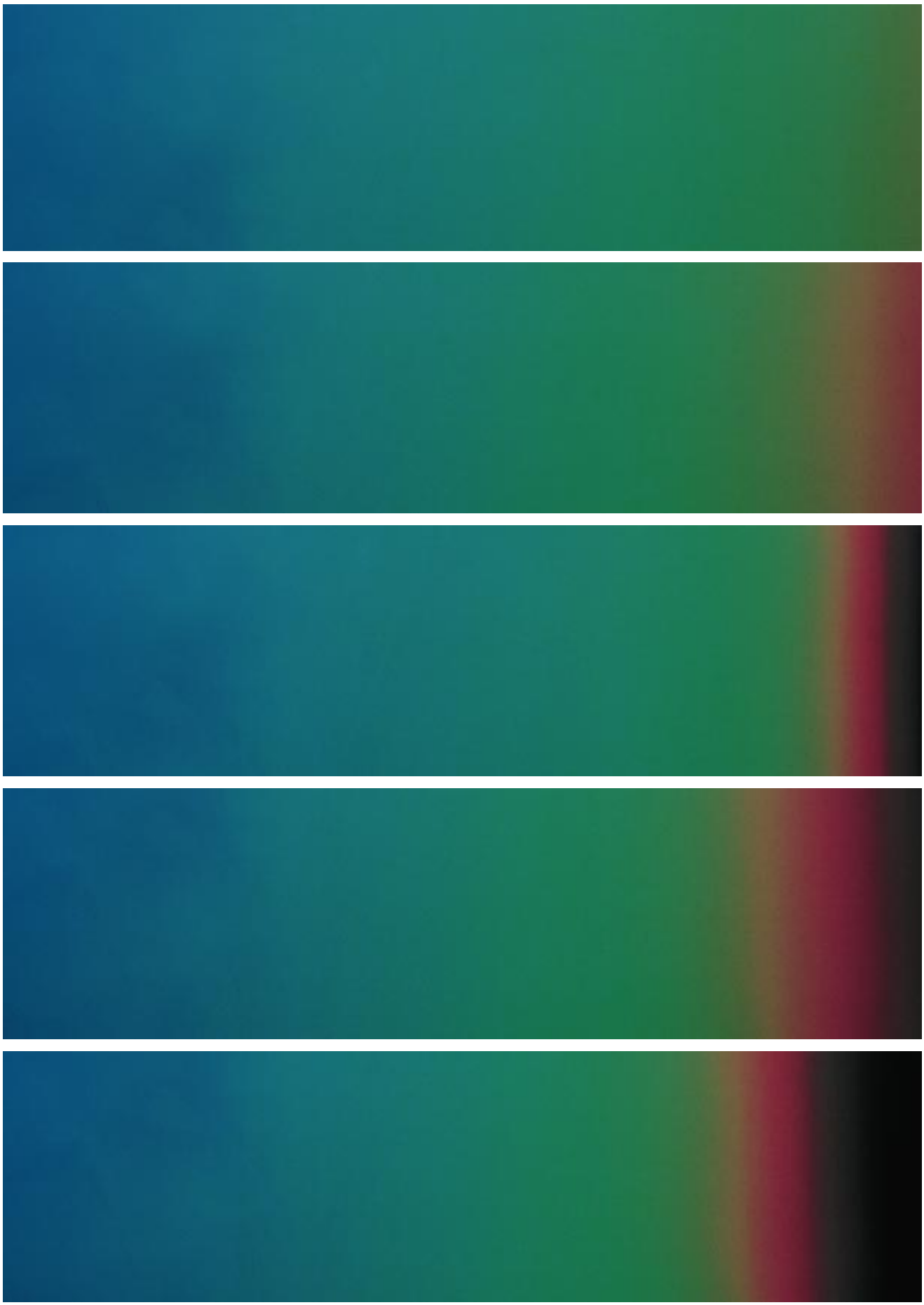
How can climatologists resolve the ambiguity? Half of the necessary data are clearly available: air bubbles trapped in the polar caps and glacial ice archive change in atmospheric composition across a span of millennia. The temperature record is more problematic: widespread meteorologic data reach back no more than 150 years. Effective coverage of the Southern Hemisphere began only in this century, and until the past few decades there were important gaps in the polar regions [see "Global Warming Trends," by Philip D. Jones and Tom M. L. Wigley; SCIENTIFIC AMERICAN, August 1990]. There is nonetheless an archive to be read if one knows where to look for it. Just as the annual layers of Arctic and Antarctic ice preserve tiny bubbles of primordial air, so the ground retains fossil temperatures whose history can be traced back to the climate of previous centuries.

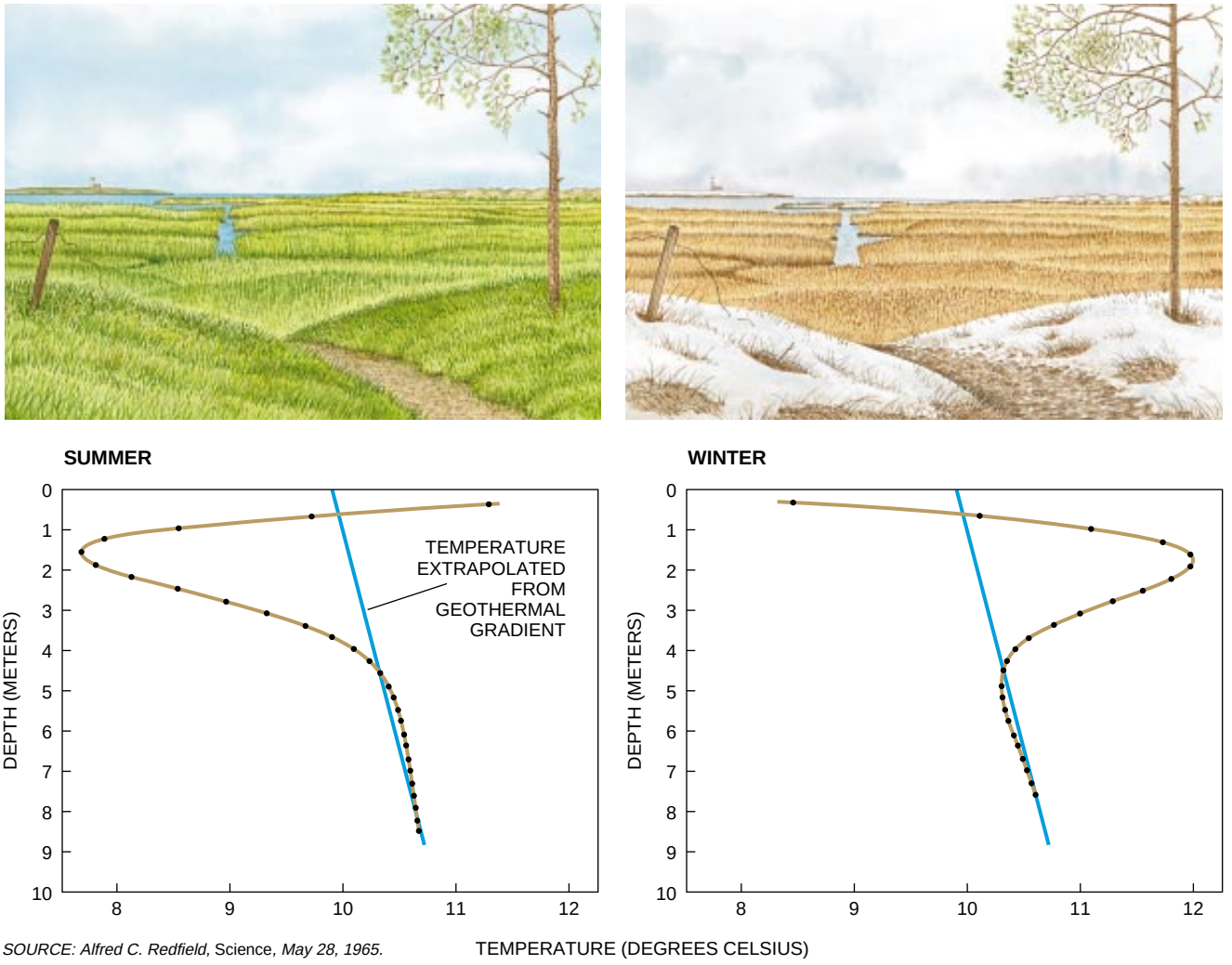
This archive exists in principle everywhere on the continents and can be tapped simply by drilling a borehole and lowering a sensitive thermometer to obtain a profile of temperature versus depth. Although many obstacles must be overcome before subsurface logs can yield an unambiguous reconstruction of past terrestrial surface temperatures, geothermal researchers are confident that they will be able to decipher the earth's buried text.

Geophysicists who have been system-

HENRY N. POLLACK and DAVID S. CHAPMAN have been collaborating on geothermal research for more than 20 years. The two met in Africa in 1970; Pollack was on sabbatical visiting the University of Zambia, where Chapman was a lecturer in physics. Pollack, a professor of geophysics at the University of Michigan, received his doctorate from the university in 1963. He also chairs the International Heat Flow Commission. Chapman followed his six years of teaching in Zambia by studying with Pollack at Michigan. He received his doctorate in 1976 and is now a professor of geology and geophysics at the University of Utah.

THERMAL GRADIENT in an aluminum sheet heated on one side and cooled on the other is made visible by temperature-sensitive liquid crystals (*top*). This gradient is conceptually similar to that normally observed within the earth's crust. If the right side is warmed slightly—in analogy to climatic warming or cooling, the resulting thermal disturbance propagates into the material (*succeeding images*). The authors have found similar anomalies in their measurements of subsurface temperature profiles and are using them to reconstruct past climate. (Engraving on this page is from a *Scientific American* report on the blizzard of 1888; this past spring's massive snowstorm came on the same date but caused somewhat less disruption.)





SOURCE: Alfred C. Redfield, Science, *May 28, 1965.*

TEMPERATURE PROFILES taken in the peat bog behind a salt marsh show how seasonal changes propagate downward, dying out as they go. At depths below 15 meters, yearly variations fade, and only longer-term climate changes are visible.

atically measuring subsurface temperatures for more than three decades have already begun reading this archive—albeit serendipitously. Their original intent was to determine the geothermal gradient (the rate at which temperature increases with depth) and measure the associated heat flux from the earth's crust [see "The Flow of Heat from the Earth's Interior," by the authors; SCIENTIFIC AMERICAN, August 1977]. Recently they have come to realize that the "noise" afflicting the top few hundred meters of their subsurface temperature data is actually the signature of external factors—such as climatic change—that modify the temperature in the uppermost part of the crust.

An early intimation that borehole readings contained useful information about climate came late in 1986. Arthur H. Lachenbruch and B. Vaughn Marshall of the U.S. Geological Survey found that the temperature profiles of a number of holes drilled in the Alaskan permafrost showed common patterns of near-surface perturbation. The patterns were consistent with the notion that the surface of the permafrost had warmed by two to four degrees Celsius during the 20th century. Although they were not the first to suggest that borehole temperature profiles contained information about changing surface conditions, Lachenbruch and Marshall made their discovery at a time when earth scientists were having their attention inexorably drawn to the possibility of global warming.

Little more than a year later, at a meeting of the American Geophysical Union, we remarked to each other that we, too, had seen many borehole temperature records that exhibited similar perturbations. Since then, we and several of our geothermal colleagues have begun exploring this subsurface resource to determine the regional variation of the earth's surface temperature over the past few centuries.

To understand how the earth retains the progression of temperatures at its surface, one must start with the theory of heat flow. Heat tends to travel through the rocks of the crust by conduction (moving groundwater can also carry heat, and so climate researchers must avoid regions where this effect is significant). When the surface of a conducting material experiences a temperature change, that alteration propagates into the interior as more energetic molecules jostle their neighbors and transfer heat to them. The effect can be demonstrated by playing a torch on the end of a metal rod: not only does the end become incandescent, but after a time adjacent sections of the rod begin to glow as well. Furthermore, if the hot end of the rod is then plunged into ice, a wave of cooling will follow the wave of heat down the length of the metal. In the same way, temperature fluctuations at the surface of the earth propagate downward into the rocks.

At shallow depths, subsurface temperature fluctuations lag surface tem-

perature variations by a few weeks or months—thus the old farming adage "Springtime drives the frost deeper." Although in spring the ground surface has already begun to warm from the winter months, the colder temperatures of the winter have gone underground. They can be found in the subsurface at depths of a few meters.

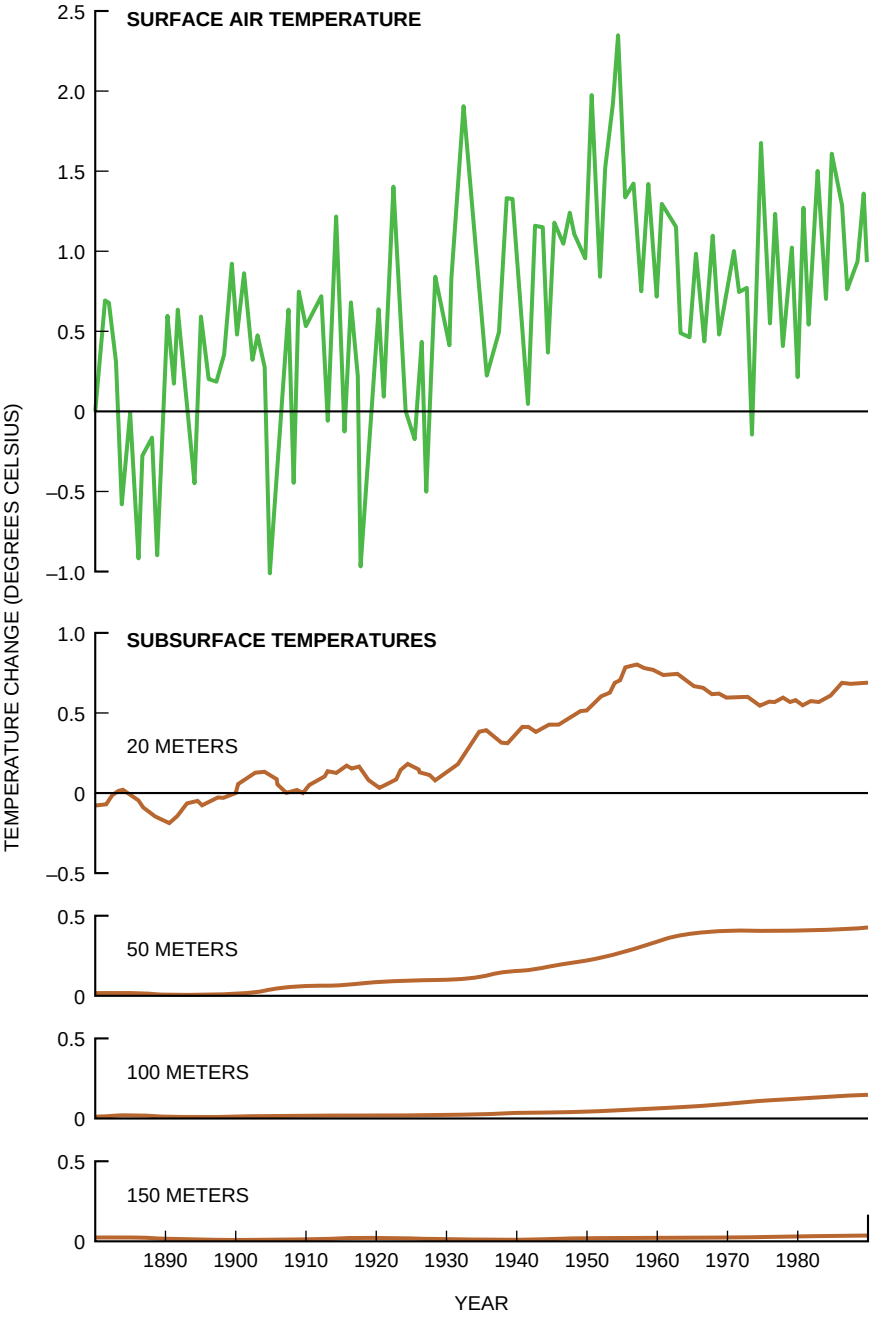
As surface temperature oscillations propagate downward, they become progressively smaller and die out. Shorter-period fluctuations, however, attenuate more rapidly than do longer ones. Only longer-term variations penetrate to great depths. The daily cycle of warm days and cool nights disturbs only the top meter of soil or rock, and the seasonal oscillation penetrates only about 15 meters before the signal is lost. A century-long cycle, in contrast, can be observed to depths of around 150 meters and a millennial one to about 500 meters. In this way, the earth selectively retains long-term trends and excludes short-period excursions from the archive, an excellent trait for recording climate.

Furthermore, subsurface records of climatic change are readily accessible. Because thermal signals travel slowly, in general all the changes in surface temperature that have occurred in the past millennium are imprinted in the uppermost 500 meters of the crust, a depth easily attainable by inexpensive drilling.

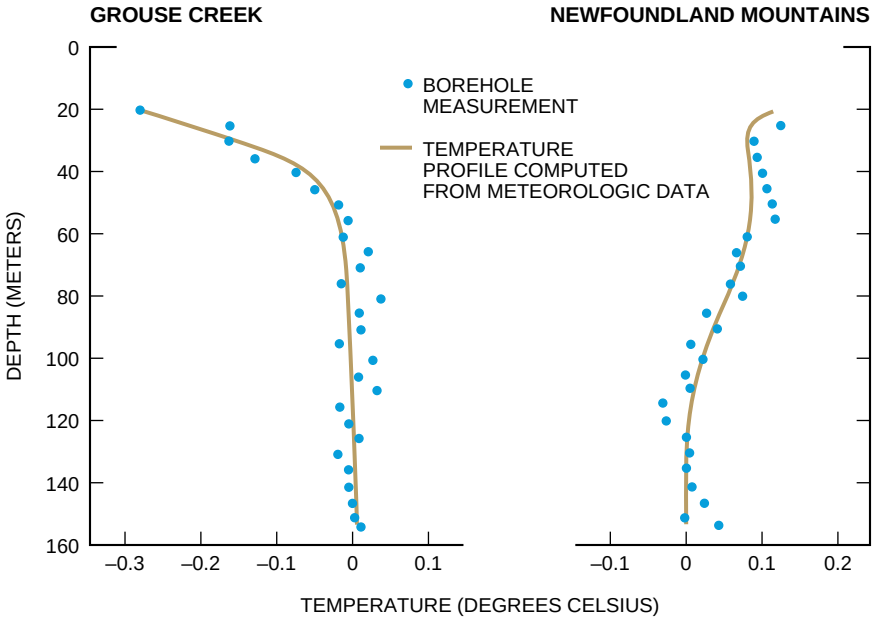
Once the mechanism by which propagating thermal disturbances leave traces of past climates is understood, the process can be reversed to recover that history from borehole temperature logs. The first step is to identify the thermal signature of the heat that is making its way upward through the crust so that it can be isolated from the climatic signal. In regions where the rock is all of one type, this deeper heat flow is characterized by temperatures that increase at a constant rate with depth. Such a constant gradient generally appears within a few hundred meters below the surface.

If the earth's climate were unchanging, this linear profile would extend all the way up to the surface. Consequently, by extrapolating the linear part of the temperature profile upward, geophysicists can tell what the temperature would have been at shallower depths before the onset of a surface temperature excursion. The difference between the surface value of the extrapolated geothermal gradient and the present-day surface temperature indicates the total amount of warming or cooling that has taken place. Moreover, the depth at which the measured profile departs from the undisturbed geothermal gradient is related to the time that climatic change began. The details of the profile between the surface and the undisturbed lower zone can be unraveled to yield information about the pace and variability of the changes. For example, a warming episode following an extended cool interval would be marked by anomalously high borehole temperatures near the surface and anomalously low ones further down.

When meteorologic, topographic and vegetative conditions are favorable, borehole temperatures track climatic change surprisingly well. In 1990 Timothy J. Chisholm, then a graduate student at the University of Utah, analyzed temperature profiles from six boreholes in the desert of western Utah. The holes, drilled in 1978 specifically for thermal measurements, were located in spots where thermal disturbances caused by



RECENT CLIMATIC HISTORY can be seen in both meteorologic records (*top graph*) and subsurface temperatures at varying depths (*lower graphs*). Graphs show annual surface temperatures in New England during the past century and the subterranean excursions that follow as the surface change propagates downward. The warming trend has only recently become visible 150 meters down, but temperatures there will continue to reflect the century's warming for many years regardless of what happens at the surface.



BOREHOLE MEASUREMENTS reveal a close match to subsurface temperatures synthesized from records at meteorologic stations at two sites in western Utah (a photograph of the Newfoundland Mountains is shown above). Subsurface temperatures in other regions may not correlate as well with air temperatures because snow cover and other factors insulate the ground from temperature extremes.

topography, streams, lakes, snowpack or human activity were minimal. Even more significant, they were geographically interspersed with seven meteorologic stations where air temperatures had been recorded since 1891.

Chisholm's results suggest that the area has been getting warmer. Five of the boreholes have temperature profiles consistent with an increase averaging 0.4 degree C during the past few decades, and one shows a cooling of 0.8 degree C. The subsurface record at each location correlates closely with the air temperatures at the nearest meteorologic stations. Indeed, the borehole whose temperatures bear the mark of recent cooling is closest to the only weather station in the region where average air temperatures have fallen during the past century. Chisholm also constructed temperature profiles based on the known flow of heat out of the earth and on the meteorologic data; these theoretical curves bear a remarkable resemblance to the actual curves of the nearest boreholes.

This close agreement is encouraging, but unfortunately it is also the exception rather than the rule. Borehole temperature profiles and meteorologic data usually do not agree in detail. Temperatures within the earth can faithfully document the thermal history of the solid surface, but meteorologists are generally more concerned about the temperature of the air. The thermal coupling of the atmosphere to the ground is not a simple process, and the temperature signal the ground receives is often already a filtered version of what the atmosphere is undergoing.

In regions that accumulate winter snow, the resulting surface blanket effectively insulates the earth from the coldest phases of the annual cycle. In central Canada the air temperature may plummet to –20 degrees C in midwinter, but the ground temperature hovers near freezing. The heat of summer, however, encounters no barrier and is transmitted into the subsurface. This winter shielding can lead to a difference of several degrees between mean annual ground and air temperatures; the effect is smaller where winters are not so severe.

At even higher latitudes, the top of the permanently frozen ground is separated from surface air by both snow and an active layer that thaws and freezes every year. Consequently, although permafrost provides an excellent medium in which to record surface temperature excursions, the complex pattern of heat transfer through these layers must be unraveled to reveal the effects of climatic change.

Temperate and tropical regions present yet a different set of confounding factors. Crops or shade trees may insulate the ground from summer heat while allowing it to cool in winter, and underground water flows can also perturb subsurface temperatures. Where humans have been at work, the picture becomes even more complicated. Deforestation and agricultural expansion exposes the ground to increased solar radiation. Draining or filling of marshlands eliminates the cooling effect of evaporation and causes surface warming. Urbanization also leads to warming because roads and buildings absorb solar energy and transmit it to the ground. Even the heat that leaks out from basements in winter affects the relation between subsurface and air temperatures. Many of these environmental modifications have become widespread during the past century and so may either magnify or mask the local archive of global warming stored in the earth.

In addition, some aspects of local topography, hydrology and patterns of vegetation can cause subsurface heating or cooling that could be mistaken for regional climatic change. The geothermal gradient generally increases below valleys and decreases below hills. Both effects diminish with depth below the irregular surface, but at shallow depths they produce temperature distortions that mimic a changing surface temperature. Meanwhile many lakes do not freeze completely in winter, and their warm bottoms influence nearby subsurface temperatures. Groundwater movements can likewise affect subsurface temperatures and leave a signature that in some circumstances looks remarkably like a response to surface temperature change.

Frustrating though these geologic thermal disturbances may be to someone seeking a straightforward correspondence between borehole logs and

climatic change, most of them can be modeled and their magnitudes estimated. In many cases, the borehole temperature profile can be corrected for these effects. Moreover, the geothermal archive is not limited to a single borehole—to see if a change is real or apparent, one can check whether boreholes spread across hundreds of kilometers of continental terrain have common perturbations in their temperature profiles. It is highly unlikely that all the boreholes would have identical topography, vegetation, geologic structure or hydrologic settings and disturbances. As a result, a common temperature pattern might safely be ascribed to climate.

Already several geothermal data sets from North America have been analyzed for evidence of surface temperature changes. Investigations in the Alaskan Arctic by Lachenbruch and his colleagues at the USGS provided dramatic evidence of warming. Temperature profiles from wells spread across 500 kilometers of northern Alaska show anomalous warming in the upper 100 to 150 meters of the permafrost and rock. The duration of the warming event appears to vary at different sites, but nearly everywhere it has a 20th-century onset.

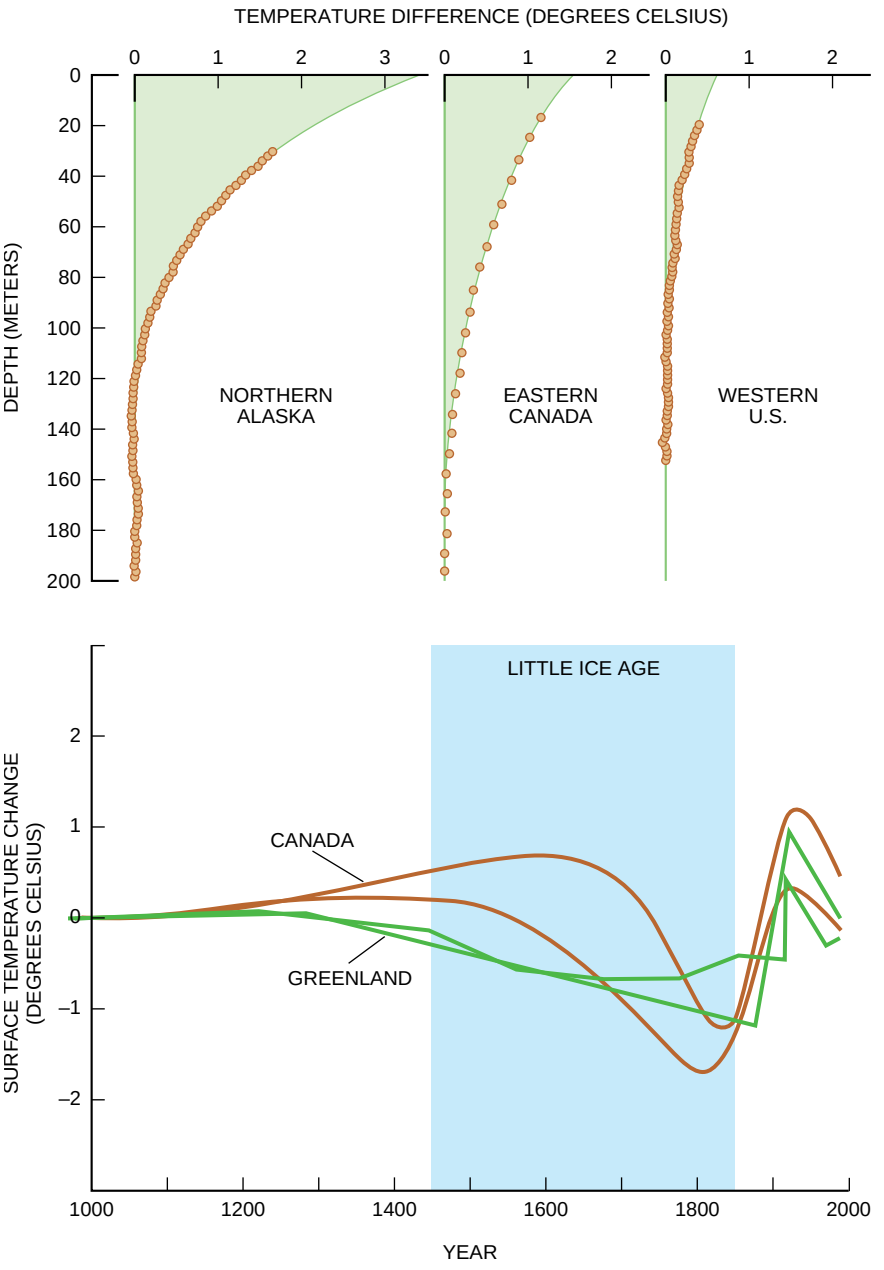
The additional heat required to produce the warming seen in the upper 100 meters of the earth in northern Alaska is small—only about 0.2 percent of the solar radiation received annually in this region. This imbalance is far too small to be measured directly, but it shows up clearly in the geothermal record. Furthermore, although the warming of between two and four degrees C is substantially greater than the global average warming of the 20th century, it is consistent with polar meteorologic records.

Boreholes distributed across Ontario, Quebec and the northern Great Plains document a less dramatic but equally clear warming. Separate investigations were made by Hugo Beltrami and Jean-Claude Mareschal of the University of Quebec at Montreal, by Kelin Wang, Trevor Lewis and Alan Jessop of the Geological Survey of Canada and by Paul Shen and Alan E. Beck of the University of Western Ontario. They have all delineated a warming that appears to be in part a recovery from an earlier one- or two-century cooling trend that bottomed out sometime between 1850 and 1900; their results show mean temperature elevations between one and two degrees C during the past 100 to 150 years. Further, William D. Gosnold of the University of North Dakota has inferred surface temperature increases of about two degrees C in North Dakota and Wyoming.

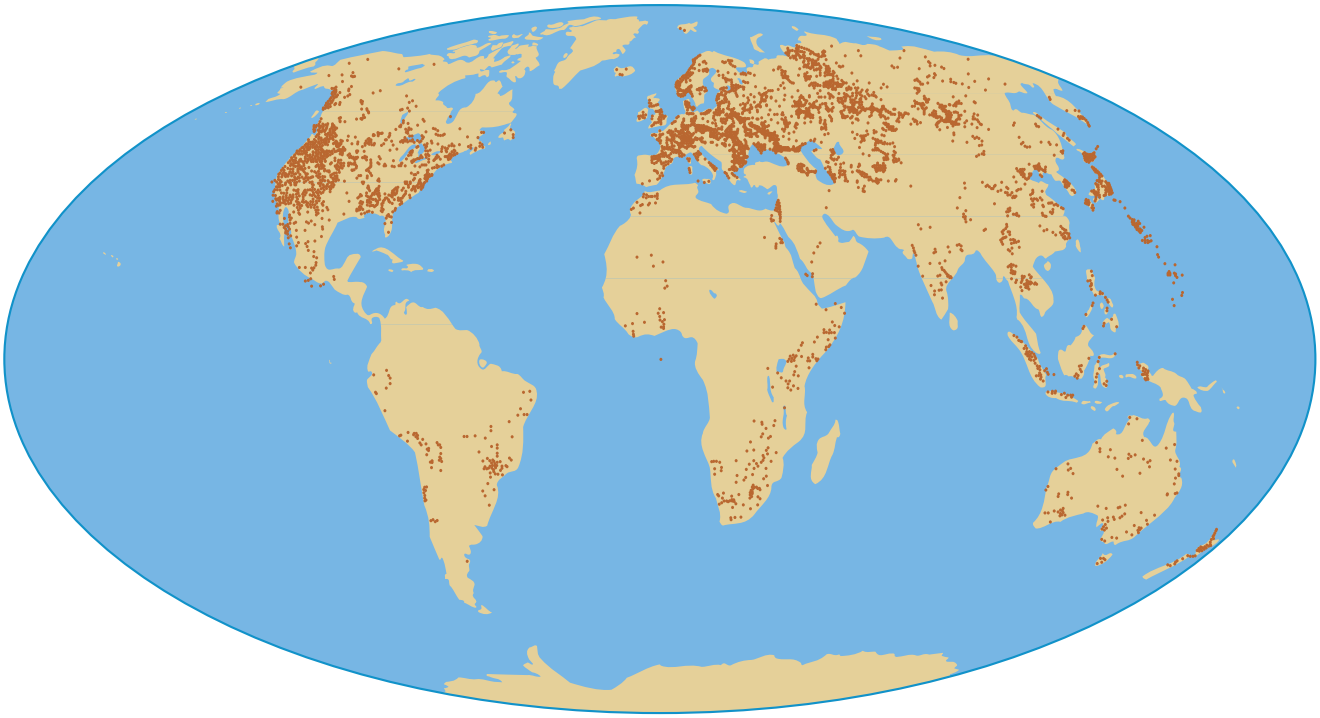
Data from southern South Dakota and Nebraska, however, indicate little change over the past 100 years, as does our own work in the desert of western Utah. This lack of a clear warming signal is consistent with the work of climate modelers, who have predicted that global warming should be most vigorous at high latitudes but minimal or even nonexistent in some temperate regions.

These preliminary results, mostly from North America, indicate that the broad outlines of the regional and temporal variation of the earth's surface temperature over at least the past century can be recovered from subsurface thermal data. More recent work suggests that the subterranean climatic archive can be read even further back in time and over much of the earth's surface.

Workers drilling at many sites in Europe, North America and Greenland have found the signature of several centuries of colder temperatures, starting at various times during the 1400s or



BOREHOLE TEMPERATURE ANOMALIES (*top*) show the difference between actual temperatures measured at three sites and those expected from the geothermal gradient. Warming appears to have begun about 100 years ago in eastern Canada and northern Alaska; climatic change in the western U.S. is more recent and less pronounced. Long-term climate histories reconstructed from boreholes in Greenland and Canada (*bottom*) indicate not only the current warming trend but also the Little Ice Age that began in the 1400s and ended in the 1800s.



WORLD DISTRIBUTION of borehole records available for analysis is uneven. Additional drilling and measurements in South America, Africa, Asia and Antarctica would significantly enhance the resulting picture of global climatic history.

1500s and ending in the 1800s. These data are consistent with contemporary accounts and other evidence of the Little Ice Age, during which glaciers advanced in many parts of the globe. The borehole data provide information about even earlier periods, but those epochs can be seen only "through a glass darkly." The reconstructed surface temperature histories show a progressive loss of detail and become more generalized. Such a loss, however, is more than compensated for by the increasingly robust estimate of the long-term mean temperature for each region.

Encouraged by results thus far, geophysicists have embarked on a concerted project to gather more subsurface climate data—first by looking into their own archives. In the fall of 1991 the International Heat Flow Commission, an association of geothermal researchers organized under the auspices of the International Association of Seismology and Physics of the Earth's Interior, established a new working group to consolidate existing data from the thousands of boreholes that have been drilled for research or for mineral exploration during the past three decades. The group will develop a unified data base of subsurface temperatures and other relevant information. This information, originally gathered to aid the understanding of global tectonic processes, will then serve as the basis of a worldwide analysis of historical temperature trends.

As might be expected, the record is not evenly distributed. Northern continents have been drilled and logged more thoroughly than southern ones. Significant gaps exist in such crucial regions as the Amazon basin, the Sahara Desert and Antarctica. If the best possible use is to be made of existing information, boreholes should be drilled in these regions to gather climate data.

Workers have begun to explore the possibility of revisiting existing boreholes to determine directly how subsurface temperatures have changed in the past few decades. Locating and reentering old boreholes in remote areas is often akin to the proverbial search for needles in haystacks, but it is not impossible. In collaboration with Edward R. Decker of the University of Maine, we have recently relocated and surveyed a set of boreholes in New England, drilled for geothermal research purposes in the 1960s, for example. We are analyzing the new data to determine the evolution of the subsurface temperature field during the 28-year interval between measurements.

The most important task for those who would recover global climate data from subsurface temperatures is integrating coverage from as many widely scattered sources as possible. As the meteorologic records have documented, there is significant regional variability in the 20th-century history of atmospheric temperatures: some areas evince warming that exceeds the global average, some show warming that falls short of the global mean and some have even cooled. No single region—except coincidentally—yields a signal that represents the global average.

Furthermore, a complete reconstruction of the recent history of the earth's climate will ultimately require more than just a knowledge of surface temperatures. Climate is a composite of temperature, precipitation, wind and many other variables. Information about some of these factors can be gleaned from many sources, including tree ring chronology and chemistry, coral growth patterns, ice core stratigraphy, lake and ocean sediments and historical, commercial and agricultural records. The challenge for climatologists is to weave these diverse regional observations into a global picture.

FURTHER READING

CHANGING CLIMATE: GEOTHERMAL EVIDENCE FROM PERMAFROST IN THE ALASKAN ARCTIC. Arthur H. Lachenbruch and B. Vaughn Marshall in *Science*, Vol. 234, pages 689–696; November 7, 1986.

CLIMATIC CHANGE INFERRED FROM UNDERGROUND TEMPERATURES. Special issue of *Global and Planetary Change*, edited by Trevor Lewis, Vol. 6, Nos. 2–4; December 1992.

The Most Distant Radio Galaxies

Astronomers have identified powerful radio-emitting galaxies that existed when the universe was only one tenth its present age. These objects offer a glimpse at the early evolution of giant galaxies

by George K. Miley and Kenneth C. Chambers

In 1946 a group of researchers at the Royal Radar Establishment in Malvern, England, discovered that intense radio emissions were emanating from a tiny region of the constellation Cygnus. Seven years later Walter Baade and Rudolph Minkowski of Hale Observatories trained the giant, 200-inch Mount Palomar telescope on the site of that radio source and found a peculiar double object they speculated might be two galaxies in collision. Subsequent work established that the source, known as Cygnus A, lies at a surprisingly great distance, 650 million light-years from the earth. That Cygnus A could be detectable at such a distance led Baade and Minkowski to conclude that it is a source of extraordinary power.

Since then, astronomers have found that Cygnus A is just one member of an entire class of active galaxies that radiate with as much as a million times the luminosity of the Milky Way. The relations between the disparate kinds of active galaxies and the nature of the mechanism that enables them to radiate so intensely have remained enduringly mysterious. Over the past two decades, however, observers and theorists have increasingly become convinced that loud radio emission is but one manifestation of the energetic processes taking place near an extremely massive collapsed object—a black hole having perhaps a billion times the mass of the sun.

By homing in on radio signals from such objects, we and several colleagues have located the most distant galaxies known. These objects are so remote that it has taken their radiation billions of years to reach the earth. Astronomers are seeing them as they were when the universe was only one tenth its present age of about 15 billion years. The most youthful active galaxies differ in several noteworthy ways from their older, more proximate relatives, and hence they offer clues about how massive galaxies form and evolve. These findings are even providing insight into the origin of the universe.

GEORGE K. MILEY and KENNETH C. CHAMBERS have spent the past seven years hunting for the most distant galaxies on the basis of their distinctive radio emissions. Miley is a professor of astronomy at Leiden University in the Netherlands. He obtained a Ph.D. in radio astronomy from the University of Manchester in 1968. During the 1970s, he used the large radio telescope at Westerbork, the Netherlands, to study steep-spectrum radio sources, laying the groundwork for the technique described in this article. From 1984 to 1988 Miley was at the Space Telescope Science Institute in Baltimore, where he headed the academic affairs branch; while there he initiated a search for distant radio galaxies, which became Chambers's Ph.D. thesis project. Chambers is now a professor of astronomy at the Institute for Astrophysics of the University of Hawaii at Honolulu.

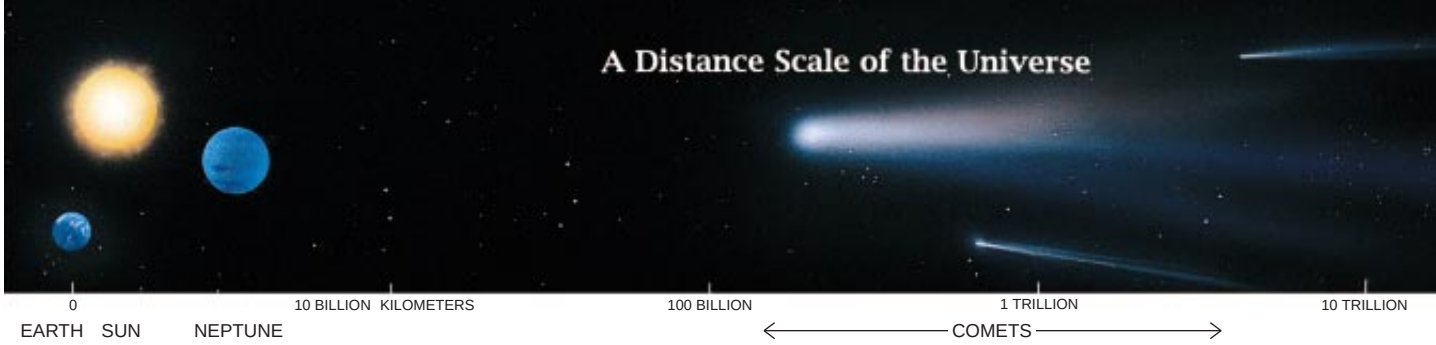
The diversity of radio-emitting galaxies became apparent as researchers sought out the visible counterparts to the radio sources listed in the 3C (third Cambridge) catalogue compiled in the late 1950s by Martin Ryle and his group at the University of Cambridge. Roughly 70 percent of the sources in the catalogue, including Cygnus A, are classified simply as radio galaxies. Most of the relatively nearby examples of these objects look more or less like normal giant elliptical galaxies. During the past few years, astronomers have observed extraordinarily distant radio galaxies; these objects have peculiar, irregular structures.

Scientists now know that radio galaxies are just one of a bewildering assortment of active galaxies that radiate with astounding power. The other main class of active galaxy consists of the quasistellar radio sources, or quasars, so named because of their starlike appearance. Unlike radio galaxies, quasars in no way resemble normal galaxies; furthermore, contrary to their name, 90 percent of quasars are quiet at radio wavelengths.

In 1963 Maarten Schmidt of the California Institute of Technology deduced from the spectra of the brightest quasars that they lie far beyond the stars of the Milky Way. Researchers have since established that quasars are the brilliant, tiny central regions of distant galaxies whose outer parts are difficult to detect because of the intense glare. The starlike appearance of quasars belies the fact that they are among the most luminous objects in the universe.

Some active galaxies display less extreme forms of activity. For example, Seyfert galaxies have bright centers that resemble tame quasars, but the body of the spiral galaxy surrounding the center is clearly evident. Indeed, astronomers are coming to recognize that the line between active galaxies and ostensibly normal ones is far blurrier than once believed. The central regions of many—perhaps most—massive galaxies

MOST DISTANT KNOWN GALAXY, which is called 4C 41.17, may lie more than 12 billion light-years from the earth. This false-color image was taken by the *Hubble Space Telescope*. The galaxy's irregular shape looks startlingly unlike the smooth, elliptical form of most relatively nearby radio-emitting galaxies. The contour lines map out the intense radio emanations from 4C 41.17. Astronomers are debating why the shape of the radio source around the more distant of these so-called radio galaxies roughly aligns with their visible appearance.



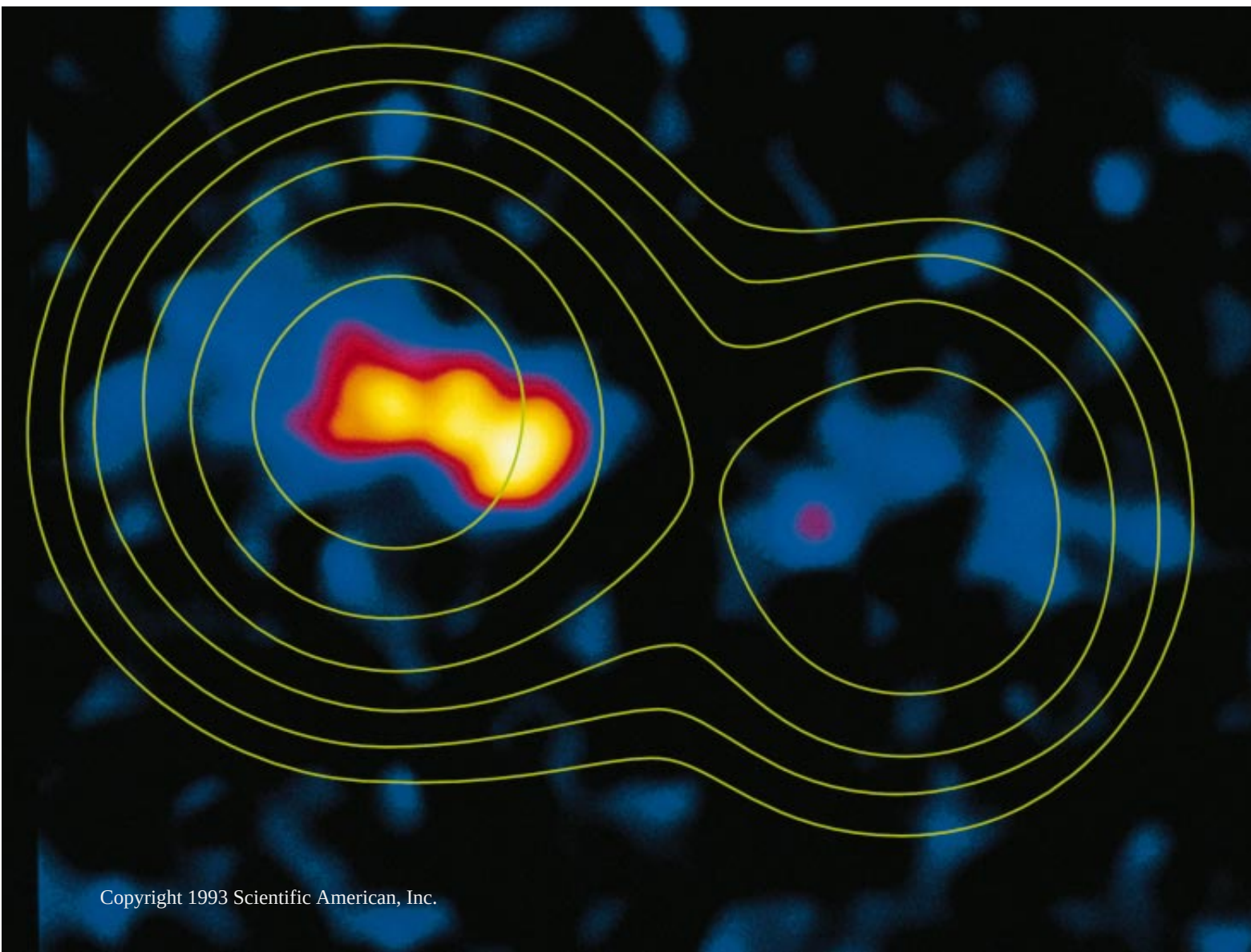
contain radio sources and some heightened concentration of light.

Using a newly developed technique known as radio interferometry, astronomers showed that many kinds of active galaxies share a common radio structure. Interferometry is accomplished by linking together two or more telescopes to create, in essence, a single, much more precise instrument. When so arranged, radio telescopes can provide much sharper images than those from even the largest optical telescopes.

Studies made during the past three decades using interferometry revealed that radio galaxies and quasars usually display two symmetric, radio-emitting lobes that straddle and dwarf their optical parent galaxies. The largest sources stretch more than 10 million light-years across, or more than 20 times the visible extent of the typical host galaxy and more than 100 times the diameter of the Milky Way. The nature of the radio emission implies that it is produced by electrons traveling at velocities approaching the speed of light as they pass through a magnetic field.

In 1971 Martin J. Rees of Cambridge suggested that hidden engines located within the nuclei of the parent galaxies generate the energy needed to power the giant radio lobes. Rees and Roger D. Blandford, now at Caltech, proposed that high-speed particles shooting along narrow channels could transport the energy. A few years later other investigators demonstrated that in many sources jetlike features do indeed seem to connect a radio-bright core in the galaxy's nucleus with knots of radio emission emanating from the outer extremities of the lobes. The jets are thought to mark the path of the subatomic particles racing from the nucleus.

The nature of the engine that powers the violent processes in radio galaxies and quasars is still a mystery, but most astronomers think a massive rotating black hole lies behind all the commotion. Einstein's theory of relativity pre-



dicts the existence of objects whose gravity is so strong that nothing, not even light, can escape from within them; observers are actively seeking unequivocal evidence of such objects.

Theorists commonly suppose that material spiraling toward a black hole becomes compressed and heated to a temperature of millions of degrees before it vanishes into the hole's interior. The superheated particles circling the hole are thought to be responsible for the various exotic phenomena that occur in and around the centers of active galaxies, such as the formation of radio jets. The jets are thought to consist of collimated beams of particles that are spewed out along the black hole's rotation axis, perhaps by a kind of electromagnetic dynamo process.

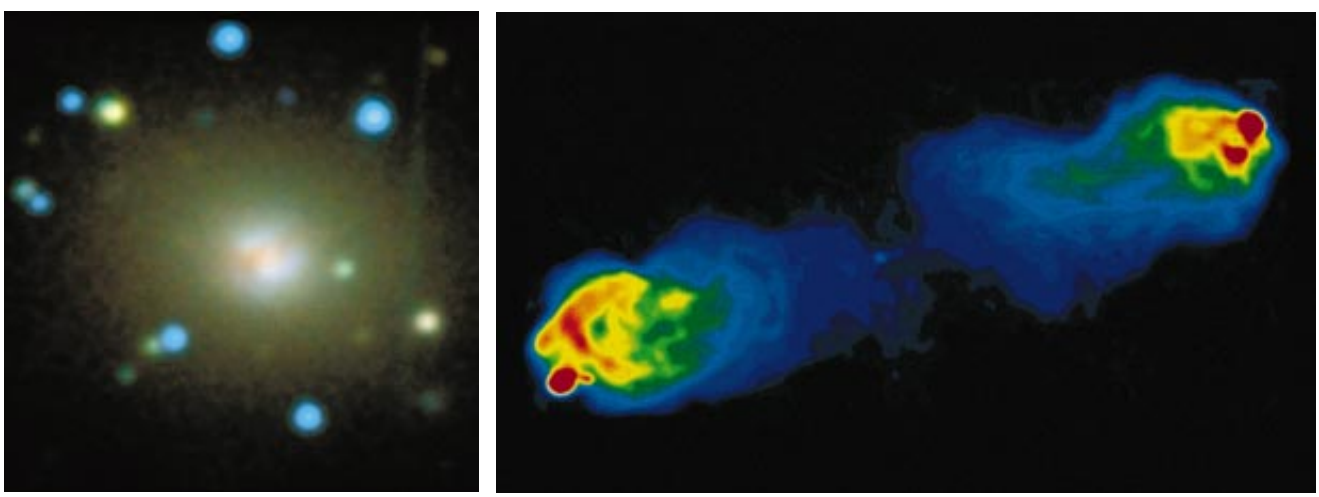
As a result of the advances in theory and observation, astronomers have begun to piece together a satisfying picture that unifies the different kinds of active galaxies. According to present thinking, one of the most significant factors determining the appearance of an active galaxy is the orientation of the radio jet—in particular, whether or not the jet is aimed toward the earth.

Several observations made during the past few years suggest that dust in the central region of an active galaxy can block all radiation except that emitted along the axis of the radio jet. Peter D. Barthel, while at Caltech, therefore proposed that all radio-emitting galaxies contain bright, embedded nuclei. If the radio source points earthward, the nucleus is visible, and the object is classified as a quasar. If the radio source is aligned in any other direction, the nucleus is more likely to be hidden from us, in which case the object is considered to be a radio galaxy.

Another likely influence on the observed properties of radio-emitting galaxies is the nature of the local environment surrounding the radio source. If the jets encounter regions of dense dust and gas, they will be unable to propagate outward, and the galaxy will not develop powerful radio-emitting lobes. Such environmental factors may explain why the most luminous radio sources form around giant elliptical galaxies, not around gas-rich spiral galaxies like the Milky Way.

The passage of time also must affect the behavior of an active galaxy. Variations in the degree of activity or the orientation of the central black hole would alter the luminosity and appearance of the radio source. Radio and optical emissions from the inner regions of active galaxies are known to fluctuate in intensity from year to year, so it is clear that conditions near the hole can change quite rapidly. Over millions of years, the black hole would gradually increase in mass and might deplete the nearby region of all material, snuffing out the activity.

As the number of known active galaxies has increased, astronomers have come to appreciate just how drastically the population of these objects has changed over the history of the universe. At great distances, where the universe is being seen as it was billions of years ago, quasars are far more abundant than they are nearby. Current surveys indicate that two billion years after the big bang, the density of bright quasars and other active galaxies in the universe was several hundred times greater than it is today [*see illustration on page 61*]. Many researchers have speculated that the era during which quasars were most common is somehow related to the formation of galaxies, but no direct link has yet been established. By examining the youngest, most distant radio galaxies, we and our colleagues hope to in-



CYGNUS A, one of the closest bright radio galaxies, is about 650 million light-years away. A new composite optical image made at several wavelengths (*left*) reveals a previously unseen darkish lane near the center of the galaxy, possibly the leftovers from a recent merger with a smaller galaxy. The huge filamentary, radio-emitting lobes (*right*) measure about 400,000 light-years across, several times the diameter of the visible part of the galaxy. The lobes are believed to be powered by twin jets of fast-moving particles ejected along the rotation axis of a black hole at the center of Cygnus A.

vestigate that relation and to uncover clues about the very early history of the universe.

Before we can discuss events that happened long ago in galaxies faraway, we must first introduce a few fundamental cosmological concepts. Measurements of distance in the universe depend on the fact that every atomic element emits and absorbs light at certain characteristic colors, or wavelengths, which show up as bright or dark lines in the spectra. In 1929 Edwin P. Hubble reported that the spectral lines associated with hydrogen, calcium and other elements show up redder (at longer wavelengths) in most galaxies than they do in the laboratory. This so-called redshift is caused by the overall expansion of the universe, which reddens, or stretches, the light. The farther away one looks, the greater the amount of expansion that has occurred and hence the greater the redshift.

The fractional shift in wavelength is usually denoted as *z*. Astronomers can measure the redshift of even a faint galaxy to within a fraction of a percent. If they knew the precise rate of the cosmic expansion and the true geometry of the universe, they could determine to a similar precision the distance to the galaxy and thereby infer its size and the amount of time that has passed since the light left that galaxy.

At present, however, the size and age of the universe are uncertain by a factor of two. Astronomers therefore find it more convenient to discuss how faraway an object is in terms of its redshift rather than its distance in light-years. Assuming that the universe is 15 billion years old and that its density matches that of the most popular cosmological models, a galaxy having a redshift of two is seen as it was 80 percent of the way back to the beginning of the universe, meaning it lies roughly 12 billion light-years from the earth; a galaxy at a redshift of four is seen as it was 90 percent of the way back.

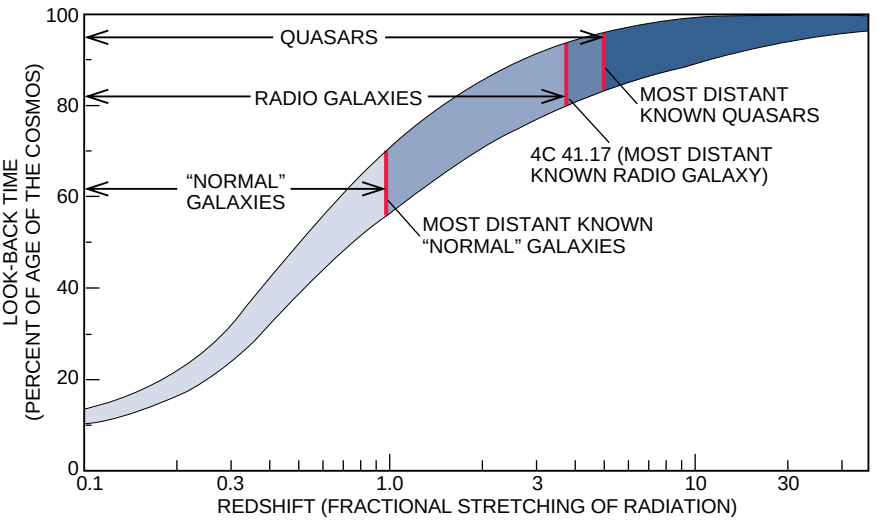
Studies of such distant radio galaxies provide a way to learn about events that occurred during the very first moments after the birth of the universe. Most cosmologists believe galaxies grew around small density fluctuations that arose less than 10^{-32} of a second after the big bang. According to current theory, most of the mass of the universe exists in the form of exotic particles known as cold dark matter. These particles interact with normal matter only through gravity, so they were able to collapse into clumps soon after the big bang, when normal matter was still too hot to do so. As the universe cooled, normal matter fell into the clumps of dark matter and ultimately formed galaxies.

In the simplest version of the cold dark matter model, galaxies would have coalesced so slowly that few of them could have formed at redshifts higher than two or three—that is, within the first couple of billion years after the big bang. Therefore, studies of galaxies dating from that time or earlier are crucially important in learning which cosmological models are promising and which ones need to be discarded.

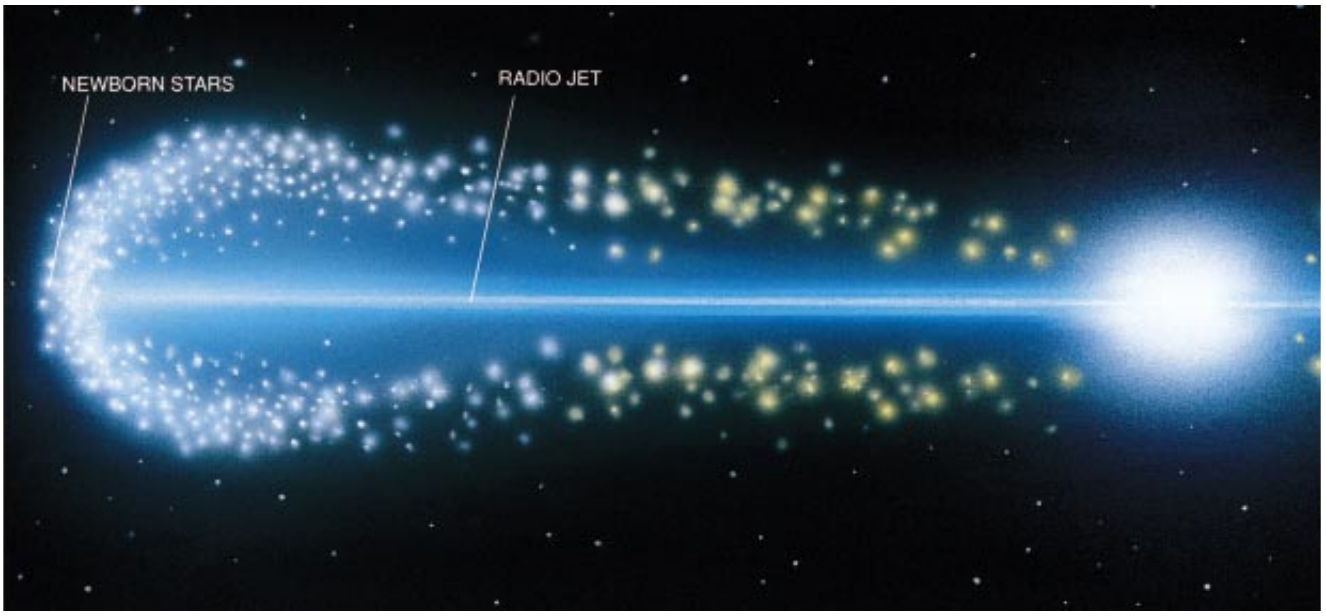
The powerful, easily detected radio emission produced by quasars and radio galaxies gives astronomers an effective way—at present, the only effective way—to locate galaxies at redshifts of two or higher. During the past decade, light detectors that incorporate charge-coupled devices, or CCDs, have revolutionized this line of inquiry by enabling astronomers to capture images of much fainter galaxies and to make spectroscopic measurement of their redshifts.

Throughout the 1970s, before CCDs became available, Hyron Spinrad and his colleagues at the University of California at Berkeley painstakingly examined the visible counterparts to the sources in the 3C catalogue. That effort revealed the first-known galaxies having redshifts greater than one. It culminated in the discovery of a radio galaxy at a redshift of 1.8, which for some time held the title of most distant known galaxy.

It is impractical to carry out time-consuming, high-sensitivity optical observations of the tens of thousands of fainter radio sources that are now known. During the past few years, several methods have been used to select the best, most distant candidates. Patrick J. McCarthy of the Carnegie Observatories in Pasadena, Calif., working with Spinrad and Willem J. M. van Breugel of Lawrence Livermore National Laboratory, has found



REDSHIFT AND DISTANCE are intimately related because of the expansion of the universe. The farther away an object is, the more its light has been stretched, or redshifted. In this graph, distance is expressed in terms of relative look-back time, the time that light took to travel from an object to the earth divided by the time that has elapsed since the big bang. Radio galaxies and quasars display bright emission lines that can be seen at redshifts of between four and five, when the universe was only about a tenth its present age.



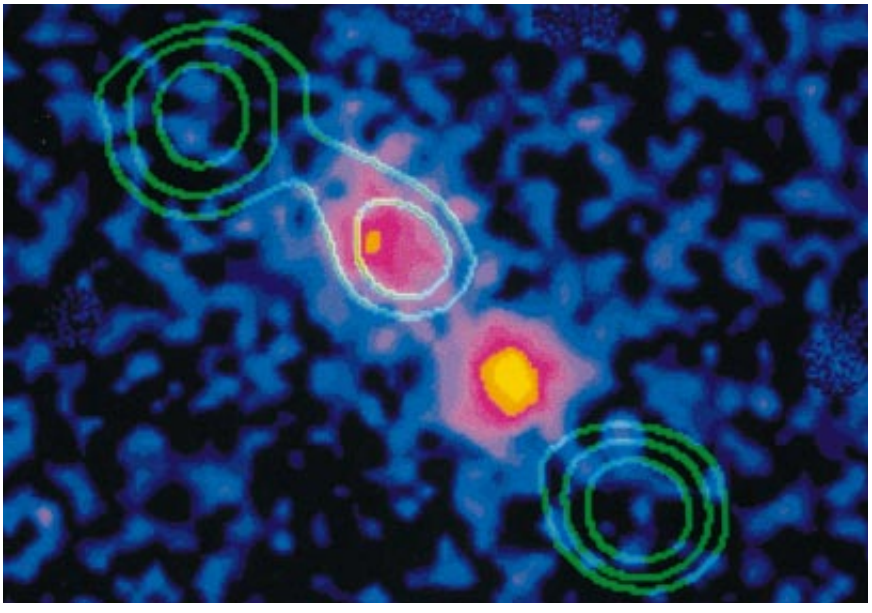
APPARENT ALIGNMENT of the shapes of the optical and radio components of radio galaxies may result from bursts of star formation. A two-sided jet of fast particles produced in a galaxy's central region propagates outward into the interstellar and intergalactic gas. As the front of the jet plows through the surrounding gas, it creates shocks that accelerate electrons to near-light speeds; these electrons generate radio waves as they race through the local magnetic field. The gas compressed by the shock cools to form clusters of stars that appear spread out along the direction of the radio source.

several galaxies lying at redshifts greater than two by concentrating on radio sources that have no bright optical counterpart. In a similar vein, Simon J. Lilly of the University of Toronto measured the optical colors of faint objects associated with the "1 Jansky" sample, a list of faint radio sources (several times fainter than those in the 3C catalogue) compiled using a radio telescope at Bologna, Italy. In 1988 Lilly, then at the University of Hawaii, reported discovery of a radio galaxy having a redshift of three.

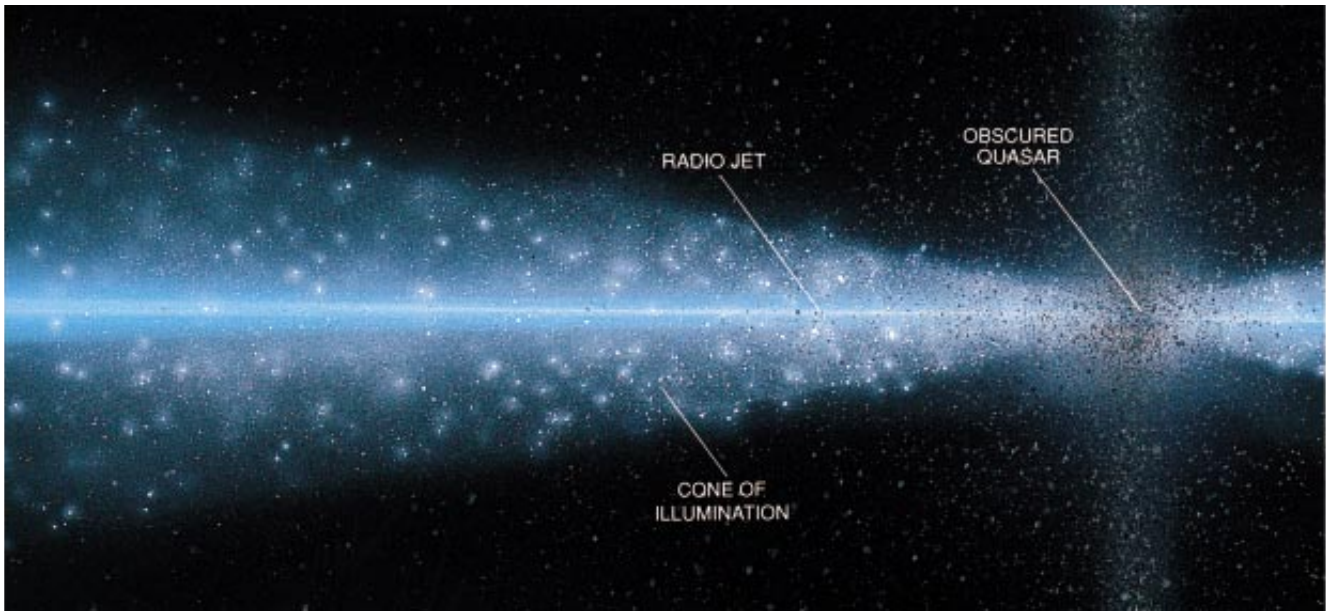
But nature has supplied another way to identify distant galaxies that has proved even more effective, one that is based solely on their properties at radio wavelengths. The spectral slope, or color, of the radio emission from an active galaxy correlates closely with its distance. The most remote objects have the steepest radio spectra—that is, their brightness falls off most rapidly from low frequencies to higher ones. Although the reason for the correlation between radio spectrum and distance is not yet fully understood, it serves as the empirical basis for a search method that works remarkably well.

We became the first researchers to take advantage of that correlation eight years ago, when we began to concentrate on the visible galaxies associated with radio sources that have particularly steep radio spectra. Among the first and most exciting objects we examined was 4C 41.17, which derives its name from the fourth Cambridge catalogue of radio sources. We identified the host galaxy connected with the radio source and determined its redshift to be 3.8. This source currently holds the title of most distant known galaxy.

Encouraged by that success, we continued our observations in collaboration with Huub Röttgering and Rob van Ojik of Leiden Observatory in the Netherlands and with several other colleagues. The work is progressing at an extremely rapid pace. Before we embarked on our project, not a single galaxy was known to have a redshift greater than two. So far our project has revealed about 35 galaxies at such high redshifts, bringing the total number known to more than 60. Our efforts are unveiling details of galactic evolution by showing how young, distant radio



REMOTE RADIO GALAXIES exhibit the alignment effect far more strongly than do more proximate ones. These two galaxies (*left* and *opposite page*), selected from a recent survey conducted by the authors and their collaborators, have



ALTERNATIVE MODEL proposes that the observed radio-optical alignment is caused by radiation scattered off electrons or dust particles. In this model, the radio galaxy contains a bright quasar that is mostly hidden from view by an obscuring shroud of dust. Light from the quasar can escape only along the axis of the radio jet. That cone of light illuminates material—either electrons or dust—and is scattered off this material and polarized by it. Because the light escapes along the direction of the jet, the observer sees the visible part of the galaxy oriented at the same angle as the radio source.

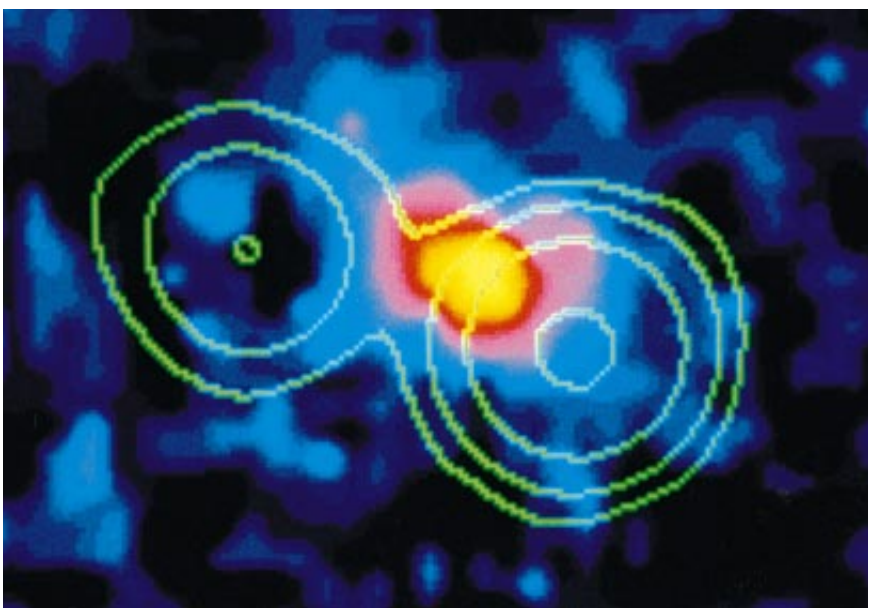
galaxies differ from more mature ones lying closer to the earth.

A remarkable property of giant elliptical galaxies (the ones that harbor the brightest radio sources) is the uniformity of their infrared luminosities. In 1984 Lilly and Malcolm S. Longair, both then at the University of Edinburgh, observed radio galaxies from the 3C catalogue in the infrared using the U.K. Infrared Telescope in Hawaii. They then constructed a graph of redshift versus infrared brightness out to a redshift of about 1.5. The resulting plot displayed a fairly neat, linear pattern, which seemed to imply that the intrinsic infrared luminosity of radio galaxies varies little over space or time.

When those observations were made, astronomers believed that the infrared emission in giant elliptical galaxies was produced by stars at least a few billion years old. The natural interpretation was that radio galaxies contain a sizable population of mature stars that all have fairly uniform properties. Lilly and Longair therefore hoped radio galaxies could serve as standard candles, objects whose absolute luminosity is known, so their apparent brightness can be used to measure accurately the size, age and geometry of the universe.

Improved observations have revealed that the situation is not that simple. Ten years ago most researchers thought radio emission was merely a useful tool for finding distant galaxies and that it did not influence the optical properties of the objects. That assumption, which was based on the fact that nearby giant elliptical galaxies look exactly the same whether or not they are strong radio sources, has turned out to be incorrect. New CCD images of the most distant radio galaxies show them to be lumpy and elongated, indicating that they are far from stable and uniform.

Those images have led to a tremendously exciting and unexpected finding. About six years ago, working independently, we and a group of researchers at Berkeley discovered that the axis of the radio emission from the galaxies lines up with the shapes seen when they are viewed in optical continuum radiation (that is, the entire band of visible



redshifts of about 2.5 (*left*) and 2.9 (*right*). The green lines show the contours of the radio emission; the false-color optical images were taken using the New Technology Telescope of the European Southern Observatory in Chile.

light, not just certain emission lines). That effect becomes noticeable in some radio galaxies having redshifts of about 0.5; at redshifts of one or more, most of these systems exhibit roughly aligned radio and optical morphologies. No comparable phenomenon is seen in nearby radio galaxies.

Much to cosmologists' regret, the observed radio-optical alignments undermine the use of radio galaxies as standard candles. The correlation between the prominence of the alignment and the galaxy's redshift strongly implies that the nature of the light from a radio galaxy depends strongly on the galaxy's distance. Furthermore, astronomers can no longer feel justified in assuming that sources selected because they are intense radio sources are, in every other respect, normal, representative galaxies. Only after the alignment phenomenon has fully been understood can we hope to disentangle optical distortions caused by the geometry of the universe from true changes caused by physical evolution.

Researchers have advanced two major hypotheses to explain the origin of the elongated visible morphology of these galaxies. One possibility is that the jet that powers the radio source sets off an enormous burst of star formation along its path; the other is that dust scatters radiation from a bright but obscured central energy source. If the former answer is correct, then the optical emission would be the light from the newborn stars. One of us (Chambers) and Stephane Charlot, then at the Space Telescope Science Institute in Baltimore, demonstrated that a burst of star formation that had been under way for a few hundred million years could indeed account for the optical and infrared colors of high-redshift radio galaxies. Such ages are also plausible for the radio sources.

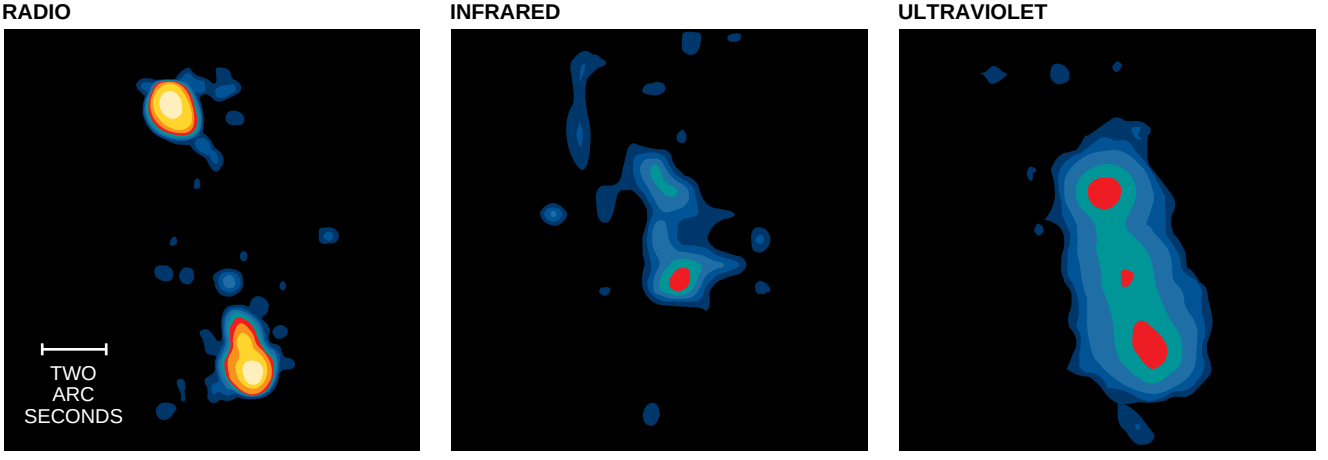
Theoretical work by Rees of Cambridge and by Mitchell C. Begelman of the University of Colorado and Denis F. Cioffi of the National Aeronautics and Space Administration also bolsters the idea of star formation. Their research suggests that shocks produced by the radio jet could compress clouds of gas surrounding the galaxy, allowing them to collapse into stars. David S. De Young of Kitt Peak National Observatory ran computer simulations of collisions between radio jets and clouds; his results confirm the plausibility of such a scenario.

Detailed studies of nearby radio sources offer additional observational support for the notion that jets might be able to trigger star formation. Although the radio source does not generally affect the optical appearance of nearby radio galaxies, observations now hint that jets can perturb the interstellar material within a galaxy. Timothy M. Heckman of Johns Hopkins University, van Breugel and one of us (Miley), using the Mayall Telescope at Kitt Peak, observed clouds of ionized gas lying along the radio sources in nearby radio galaxies. That finding suggests that the radio jets are interacting vigorously with the gas in these systems. One radio-emitting galaxy, the peculiar Minkowski's Object, shows what appears to be a newly formed dwarf galaxy located at a bend in the jet.

An alternative explanation for the radio-optical alignment was prompted by the observations of the optical polarization (the degree to which light waves are preferentially oriented in a particular direction) of 3C 368 and several other bright, aligned radio galaxies. A group led by Spirello di Serego Alighieri, Robert A. Fosbury and Clive N. Tadhunter of the European Southern Observatory and Peter J. Quinn of Mount Stromlo Observatory studied 3C 368 using a telescope in Chile; Michael Scarrott and C. D. Rolph of the University of Durham and Tadhunter conducted follow-up work at the William Herschel Telescope in the Canary Islands. The researchers found that the light from 3C 368 is highly polarized.

One of the easiest ways to polarize light is to scatter it. That fact led Tadhunter and his colleagues and, independently, Andrew C. Fabian of Cambridge to suggest that some of the light from the distant radio galaxies consists of scattered emission from a hidden quasar nestled in the galaxy's nucleus. Because the quasar's radiation is absorbed in every direction except along the radio axis, it cannot be seen directly. Like a searchlight passing through the fog, however, the quasar beam bounces off electrons or dust in its path, rendering it visible to terrestrial viewers. That scattered light would appear aligned along the radio jet.

Neither the starburst nor the scattering models can explain all the features of distant radio galaxies. The presence of polarization means that some light must be scattered. But electrons scatter all wavelengths of light equally well, so one would expect the scattered light to resemble the spectrum of a quasar, which it does not. Dust scatters blue light more efficiently than it does red and so could produce the strong color gradients that are observed. There is



DIFFERENT WAVELENGTHS capture subtly different aspects of the bright radio galaxy 3C 368; false colors denote intensity of radiation, from low (*dark blue*) to high (*light yellow*). Radio emission (*left*) is generated by high-speed electrons moving through the galaxy's magnetic field; a double-lobed structure is clearly evident. An infrared image (*center*) shows radiation emitted by stars and gas in the body of the galaxy. Ultraviolet rays having a wavelength of 3,727 angstroms come from oxygen atoms that have been disrupted and ionized (*right*). Note that all three images have similar orientations.

evidence that some radio galaxies display clear alignments between their radio and infrared emission, however, and a dust model has difficulty producing enough infrared scattering to account for this effect. In any case, dust consists of heavy elements that form only in the interior of stars, so it could be present only if some stars have already formed along the radio axis. A composite picture of distant radio galaxies that includes both star formation and scattering along the radio axis therefore seems most plausible.

Attempts to decipher the processes responsible for the radio-optical alignment effect are already leading to the development of more sophisticated theories about the early evolution of radio galaxies. Further advances will depend on finding additional radio galaxies at very high redshifts. We and a number of our colleagues are working hard to do just that.

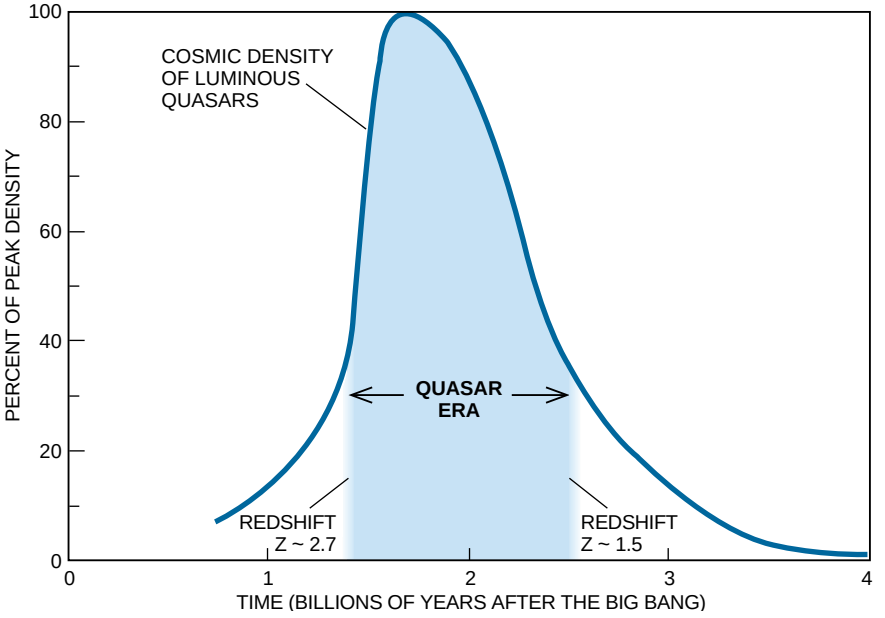
Some intriguing preliminary results have emerged from our follow-up studies of 4C 41.17. Last year, in collaboration with van Breugel and F. Duccio Macchetto of the Space Telescope Science Institute, we used the *Hubble Space Telescope* to observe 4C 41.17. The resulting image has an angular resolution roughly 10 times better than the previous ground-based photograph. This sharp picture shows the inner region of this galaxy to have an irregular, clumpy form. Those clumps may represent gas clouds illuminated by a central quasar, or they may be giant star clusters caught in the act of coalescing. Analysis of the spectrum of 4C 41.17 may determine which of these explanations is correct and hence could reveal the mechanism that produces the radio-optical alignments.

Optical and radio telescopes could detect objects like 4C 41.17 at redshifts of about six, if such remote, young galaxies exist. Finding and examining radio galaxies at ever greater redshifts will help settle many profound questions about how the universe was born and how galaxies formed. Sensitive images and spectra of regions surrounding those galaxies will enable astronomers to search for nearby companions or clusters of galaxies. Studies of the galaxies' colors will yield information about the stars they contain and hence about how soon after the big bang those stars began to shine. Related observations of shapes and motions within the galaxies may establish whether these objects are being seen while they are still in the process of collapsing from primordial clouds of hydrogen and helium gas.

Furthermore, spectral analysis of extremely distant radio galaxies enables astronomers to observe whatever may lie along the line of sight between those galaxies and the earth. Intervening gas clouds or other galaxies, for example, could produce detectable absorption lines in the radio galaxy's spectrum. Those lines can reveal information about the shape, composition and kinematics of the intervening objects, as well as their distance. Because radio galaxies, unlike quasars, are spatially extended, they can serve as valuable probes for investigating closely separated lines of sight a few hundredths of a degree apart.

Simply compiling a more complete tally of distant radio sources will clarify how many quasars and radio galaxies existed during the first couple of billion years after the big bang and how the density of those objects evolved over time. As with the extinction of dinosaurs on the earth, the stunning decline of the active galaxy population as the universe has aged signals a dramatic change in the celestial environment. Many researchers are now conducting a census of distant radio galaxies as a function of their redshift and absolute radio luminosity. Detailed comparisons of the population densities of quasars and galaxies at redshifts greater than two may elucidate the processes responsible for exterminating the species.

Over the next decade, an impressive arsenal of instruments will facilitate these investigations. At the end of this year, NASA plans to correct the optics of the *Hubble Space Telescope,* improving the sensitivity with which it can observe distant radio galaxies by about a factor of five. A new generation of large ground-based optical telescopes, such as the European Southern Observatory's Very Large Telescope and the two 10-meter Keck telescopes in Hawaii, will be able to map faint objects that have eluded detection thus far. These tools hold out the promise that within the next few years astronomers will greatly expand their understanding of a universe that has been 15 billion years in the making.



QUASARS AND RADIO GALAXIES were nearly 1,000 times more abundant two billion years after the big bang than they are now. The reason for the rapid rise and decline of the active galaxy population is a mystery, one that may hold important clues about the formation and early development of massive galaxies.

FURTHER READING

4C 41.17: A RADIO GALAXY AT A REDSHIFT OF 3.8. K. C. Chambers, G. K. Miley and W.J.M. van Breugel in *Astrophysical Journal,* Vol. 363, No. 1, pages 21–39; November 1, 1990.

HIGH REDSHIFT RADIO GALAXIES AND THE ALIGNMENT EFFECT. K. C. Chambers and G. K. Miley in *Edwin Hubble Centennial Symposium: Evolution of the Universe of Galaxies.* Edited by R. G. Kron. Astronomical Society of the Pacific, 1990.

DISTANT GALAXIES. G. K. Miley in *Proceedings of the ST-ECF/STSCI Workshop on Science with the Hubble Space Telescope: ESO Conference and Workshop Proceedings,* No. 44. Edited by P. Benevenuti and E. Schreier. European Southern Observatory, 1992.

HUBBLE SPACE TELESCOPE IMAGING OF DISTANT GALAXIES: 4C 41.17 AT $z = 3.8$. G. K. Miley, K. C. Chambers, W.J.M. van Breugel and F. Macchetto in *Astrophysical Journal,* Vol. 401, No. 2, Part 2, pages L69–L73; December 20, 1992.

The Centrosome

By directing the assembly of a cell's skeleton, this organelle controls division, motility and shape. The details of its structure and function are just beginning to emerge

by David M. Glover, Cayetano Gonzalez and Jordan W. Raff

The cytoplasm that fills a cell is not simply an amorphous medium in which organelles float. Rather it has a skeleton of protein fibers and accessory molecules that serves as the highly dynamic scaffolding for the many external and internal alterations that a cell undergoes during its life. The workings and organization of that cytoskeleton are only now beginning to be understood. Nevertheless, it is clear that an organelle known as the centrosome plays a pivotal role in those processes.

The centrosome is a somewhat shapeless body from which radiate microtubules, one of the major cytoskeletal elements. The microtubules, in turn, seem to influence the distribution of the smaller actin and intermediate filaments, which are the other major protein fibers that compose the cytoskeleton. Thus, the centrosome acts as a master architect of cytoskeletal design. Through its effects on that complex lattice of fibers, the centrosome governs the shape, polarity and movement of cells as well as the transport of materials within cells. During cell division, it assumes the critical function of setting up the mitotic spindle, the cellular apparatus that partitions the chromosomes into daughter cells.

Despite its importance, the centrosome has remained an enigma. Now, with the development of molecular biology, key components of the centrosome are at last being identified and characterized. We may be on the verge of answering questions about the division, differentiation and movement of cells that have puzzled biologists for more than 100 years.

Centrosomes were first described independently in 1887 by Theodor H. Boveri and Edouard Joseph Louis-Marie van Beneden, who were studying cell division in the eggs of the roundworm *Ascaris*. During mitosis, when a cell divides, the two sets of chromosomes produced by replication of the cell's DNA are equally partitioned between the two daughter cells. The mitotic spindle is essential to this process. During the part of mitosis called prophase, the replicated chromosomes condense; during metaphase, they line up along the equator of the spindle. Finally, during anaphase, one full set of chromosomes migrates toward each of the spindle's poles.

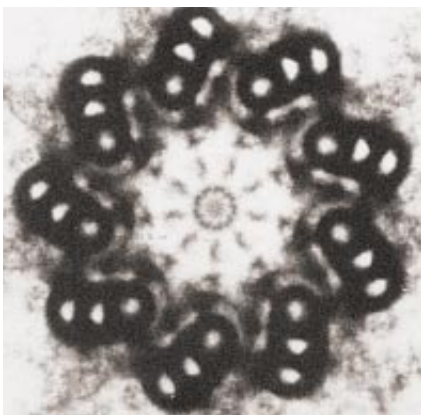
Boveri and van Beneden both noticed that the mitotic spindle appeared to extend from two dotlike objects, which they called polar corpuscles or centrosomes, one at each pole of the spindle. When cells were not dividing, a single centrosome could be observed next to the nucleus. At the onset of cell division, the centrosome divided into two and began to move apart. The twin centrosomes became the centers that organized the microtubular structures radiating from each pole of the spindle. The separation of the centrosomes, which occurs only once during a mitotic cycle, constitutes a mechanism that ensures that the replicated chromosomes are shared equally by the two daughter cells.

By the beginning of the 20th century, researchers had described centrosomes in a wide variety of species. Under the light microscope, however, the appearance of the organelle differed greatly among cell types. It became known by a bewildering assortment of names: centriole, central corpuscle, division center, central body, mitotic center and centrosphere.

Adding to the confusion was the observation that higher plants and some lower eukaryotic cells did not seem to have centrosomes at all. That observation was interpreted as evidence that centrosomes were not essential parts of the mitotic apparatus. Even as late as the 1930s, some biologists thought that centrosomes were merely artifacts induced during the preparation of cells for microscopy or that their presence at the poles of spindles was a consequence rather than the cause of spindle formation.

Fortunately, the development of electron microscopy helped to clarify the structure of centrosomes and the microtubules linked to them. At the heart of the centrosome in an animal cell are two geometric arrangements of microtubules described as centrioles. Each centriole is a cylindrical bundle of nine rods; each rod consists of three micro-

DAVID M. GLOVER, CAYETANO GONZALEZ and JORDAN W. RAFF have contributed prominently to current understanding of the centrosome. Glover is professor of molecular genetics and director of the Cancer Research Campaign (CRC) Cell Cycle Genetics Group at the Medical Sciences Institute of the University of Dundee. A graduate of the University of Cambridge, he received his doctorate in biochemistry in 1972 from University College, London. Gonzalez, a postdoctoral research assistant in the CRC group at Dundee, earned his doctorate at the Universidad Autónoma de Madrid. Raff is a postdoctoral fellow in the department of biochemistry and biophysics of the University of California, San Francisco. He received his Ph.D. from Imperial College, London.



STRUCTURE OF CENTROSOME became evident only with the advent of electron microscopy, even though this organelle has been known for more than a

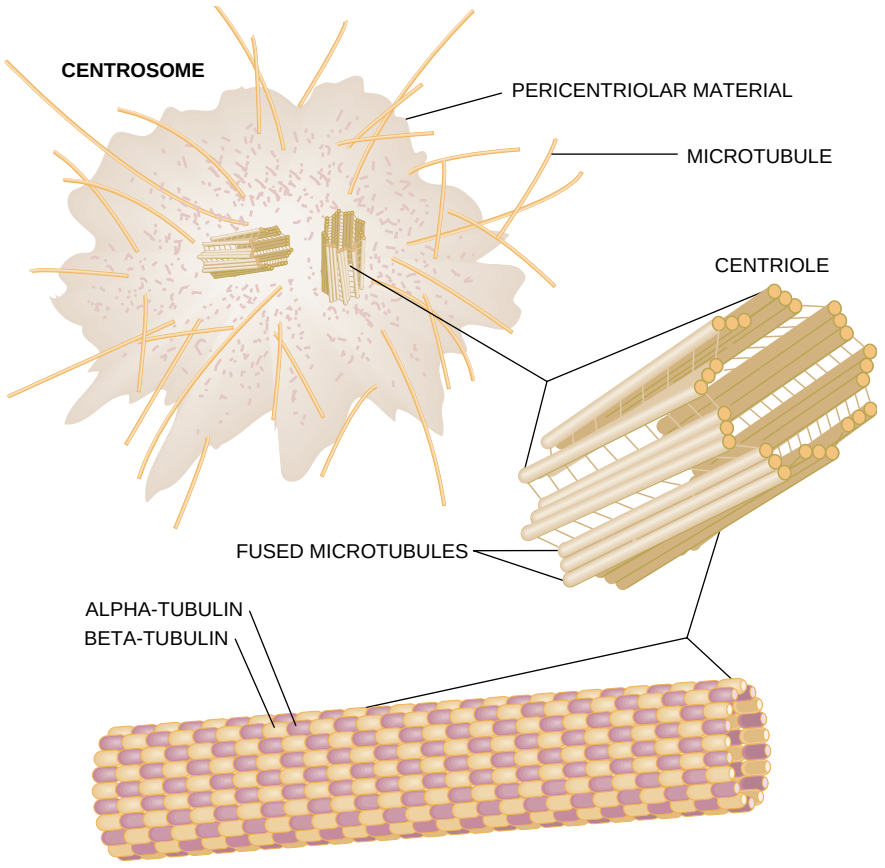
tubules about 5,000 angstroms (one 50,000th of a millimeter) long, fused along their length. Because of their appearance in cross section, centrioles are sometimes described as pinwheel structures. In animal cells, each centrosome has two centrioles that are at right angles to each other and are surrounded by a cloud of amorphous material.

Our current ignorance about the nature of the surrounding substance is reflected by the fact that it is referred to simply as pericentriolar material. Electron microscopy has shown that microtubules emanate from the pericentriolar material and not directly from the centrioles in the centrosome. It is therefore likely that this material functions as the true microtubule-organizing center, a term coined by Jeremy D. Pickett-Heaps of the University of Colorado at Boulder. In plant cells the microtubule-organizing center is more diffuse, and the centrioles are missing, which explains why plant cells were initially thought to lack centrosomes.

As it turns out, centrioles are not found exclusively in centrosomes. They also appear in the basal bodies underneath flagellae and cilia—the long, threadlike structures that project from the surface of some cells. Many unicellular eukaryotes swim by beating their flagellae or cilia; some specialized cells in the body, such as those lining the intestinal tract, use their cilia to move extracellular secretions. Because the basal bodies seem to coordinate the organization of the microtubules in the flagellae and cilia, they are fundamental to the movements of those organelles.

The discovery of centrioles in centrosomes and basal bodies strengthened a hypothesis, put forward independently by L. F. Henneguy and M. Lehossek in 1898, that these structures are inter-

century. It consists of two structures called centrioles set at right angles to each other and surrounded by a cloud of pericentriolar material (*right*). Seen in cross section, a centriole reveals a pinwheel structure made of structural elements called microtubules (*left*). The function of the centrosome is to organize the microtubules in the cytoplasm, particularly during cell division. Because microtubules influence other protein fibers, the centrosome acts as the architect of the cytoskeleton.



CENTRIOLES at the heart of the centrosomes in animal cells are tubelike structures. Each one is formed from nine rods; each rod consists of three microtubules fused along their length. A microtubule is a hollow fiber made of subunits that contain an alpha- and a beta-tubulin protein. The cytoplasmic microtubules organized by the centrosome seem to grow out of the pericentriolar material and not directly from the centrioles themselves.

convertible. Support for that idea can be found in a variety of organisms. For example, *Chlamydomonas* and some other single-cell algae have two flagellae, each with a basal body. Just before mitosis, the cells resorb their flagellae. The basal bodies then appear to migrate close to the nucleus, where they are incorporated into the centrosomes that organize the mitotic spindle.

Although the centrosome and the basal body both organize microtubules, they do so in dramatically different ways. The microtubules of the mitotic spindle grow out of the pericentriolar material, but the microtubules in a cilium grow directly out of the centriole in the basal body. More precisely, the formation of a cilium involves the emergence of a structure called the axoneme. The axoneme grows through the addition of tubulin to two of the three microtubules in each component rod of the pinwheel. It elongates until it extends throughout the length of the cilium. In addition, an extra pair of microtubules not found in the centriole develops along the axis. The structure of the axoneme mirrors the ninefold symmetry of the centriole.

The ordered, stable structure of the axoneme contrasts with the highly dynamic arrangement of microtubules organized by the centrosome, which have properties that change during the cell cycle. When a cell is in interphase—the period between mitoses—an extensive array of microtubules emanates from the centrosomes and stretches throughout most of the cytoplasm. As the cell enters mitosis, this interphase array breaks down, and the released tubulin subunits reassemble themselves into a completely different structure, the mitotic spindle.

How does that reorganization occur? Microtubules extending from centrosomes are dynamic structures: they continually grow and shrink because of a phenomenon called dynamic instability. One end of each tubule, the plus end, grows rapidly because tubulin subunits, each of which consists of an alpha- and a beta-tubulin protein, readily attach there. The minus end of each microtubule grows much more slowly and would in fact depolymerize if it were not stabilized by binding to the centrosome. The inherent instability of microtubules permits them to redistribute themselves very quickly, which is crucial for cells trying to change their shape, migrate or divide.

As the cell enters mitosis, cytoplasmic microtubules become particularly unstable, and the interphase array depolymerizes. The centrosomes nucleate, or promote the growth of, many extremely dynamic microtubules. Those fibers alternately grow out in random directions and swiftly shrink back. If the end of a microtubule comes in contact with a kinetochore—a specialized region on each chromosome—the microtubule attaches to it and stops growing or shrinking. In effect, the centrosomes send out microtubule "feelers" that search for chromosomes.

Eventually, microtubules from both centrosomes bind to the kinetochores on all the replicated chromosomes. During metaphase, the pairs of replicated chromosomes align on the equator of the spindle. Subsequently, they split apart and move toward opposite poles. The forces that move the chromosomes depend on force-generating molecules—microtubule motors—that are present at several sites on the mitotic apparatus and that have been the subject of intense study in recent years.

The observations of the cytoskeleton demonstrate that centrosomes do organize microtubules but do little to explain how they do so. The recent discovery of a protein called gamma-tubulin may represent a breakthrough in attempts to answer that question.

Gamma-tubulin was first identified in 1989 in the fungus *Aspergillus nidulans* by Berl R. Oakley of Ohio State University and his colleagues, who were seeking proteins that could interact with beta-tubulin. Because they had already isolated a mutation of the beta-tubulin gene that affected the behavior of microtubules, they set out to find a second mutation that could "rescue" the mutant cell from that defect. They reasoned that such a second mutation would probably occur in genes for proteins that interacted with beta-tubulin and could compensate for its abnormality. One of the genes that the researchers found encoded a protein that was closely related to both alpha- and beta-tubulin and was given the name gamma-tubulin. To the surprise of Oakley and others, gamma-tubulin was not a component of the microtubules themselves but rather of the spindle pole body—the fungal equivalent of a centrosome.

Those findings immediately suggested that gamma-tubulin might act as a nucleation point for microtubule growth. Gamma-tubulin is found in the pericentriolar material, and it does appear to be essential for microtubule nucleation. The general hypothesis is

further supported by the observation that gamma-tubulin is highly conserved in all eukaryotes, which suggests that it plays an important role in all microtubule-organizing centers. Gamma-tubulin may therefore turn out to be the molecular handle that researchers have long sought for analyzing how the centrosome organizes microtubules.

The mechanism of centrosome replication has also puzzled biologists since the discovery of the organelle. Boveri and van Beneden originally regarded the centrosome as a permanent and autonomous cell organelle that, like the nucleus, arose through the replication and division of a previously existing body of the same kind. Electron microscopic studies of centrosome duplication in cultured cells supported that idea: as a cell prepares to divide, the pair of centrioles within a centrosome splits apart, and each centriole nucleates a second one at right angles to itself. The new centriole initially consists of just nine single microtubules arranged in a cylinder, but they soon transform into triplet microtubules. The two pairs of centrioles migrate to opposite sides of the nucleus, each taking some of the pericentriolar material with them. The cell now has two centrosomes. Because this model of centrosome replication requires the presence of a preexisting centrosome, it is sometimes described as template-driven.

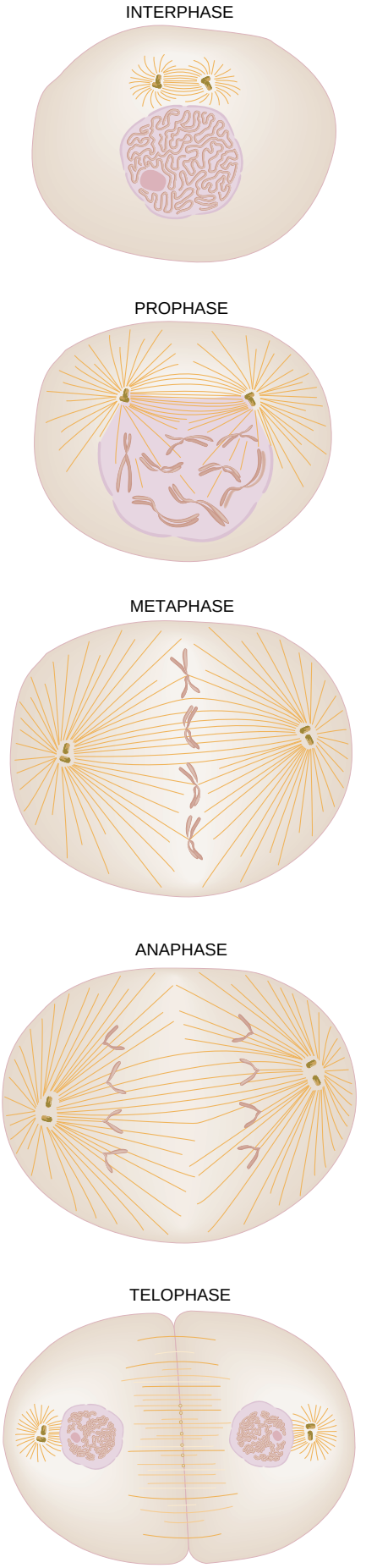
On the other hand, there are many well-documented cases in which centrosomes appear to arise spontaneously. For example, depending on its growth conditions, the protozoan *Naegleria* assumes either a flagellated form or an amoeboid form lacking flagellae. When researchers have taken a *Naegleria* amoeba, sliced it into thin sections and inspected each section under the microscope, they have never been able to find centrioles in the cytoplasm. Yet when the amoeboid form changes and becomes flagellated, typical basal bodies develop. The centrioles in those basal bodies seem to have appeared without a precursor.

The apparently spontaneous generation of centrosomes is not necessarily inconsistent with the possibility that their replication depends on the existence of a template. In *Naegleria,* for example, the replicative element of the centrosome might simply pass through a phase in the life cycle of the organism in which it no longer looks like a centrosome.

Some investigators have proposed that if centrosomes do replicate from some kind of template, they may contain genetic information in the form of DNA or RNA. Such nucleic acid molecules, which have the power of self-duplication, might endow the centrosome with replicative properties. Many laboratories have engaged in a search for nucleic acids in centrosomes, but their results have been conflicting and ambiguous. A review article written in 1971 cited seven reports that favored the theory that nucleic acids are present in centrosomes—and eight against. Today the question of whether centrosomes have a nucleic acid component is as hotly disputed as ever.

Notwithstanding the lack of progress on that front, investigators have learned some intriguing truths about the regulation of centrosome division. It is important for a cell that its centrosome divides once—and only once—per cell cycle. If a centrosome does not replicate, a bipolar mitotic spindle cannot form, and so the cell cannot divide. If the centrosome replicates more than once, multipolar spindles may form, and the chromosomes will not be segregated correctly. Studies of the cell cycle have uncovered some surprising facts about the controls over centrosome division and its connection to other events in the cell cycle.

Biologists customarily divide the cell cycle into mitosis, the act of cell division, and interphase, the period between mitoses. Interphase is itself subdivided into three parts: G1, a gap phase following mitosis; S phase, during which the chromosomal DNA is duplicated; and G2, a second gap phase preceding mitosis. Genetic studies of yeast suggest that each step of the cycle must be completed before the next can begin. If S phase is blocked by the mutation of a gene required for DNA replication, the cycle arrests at that point, and cells never enter mitosis.

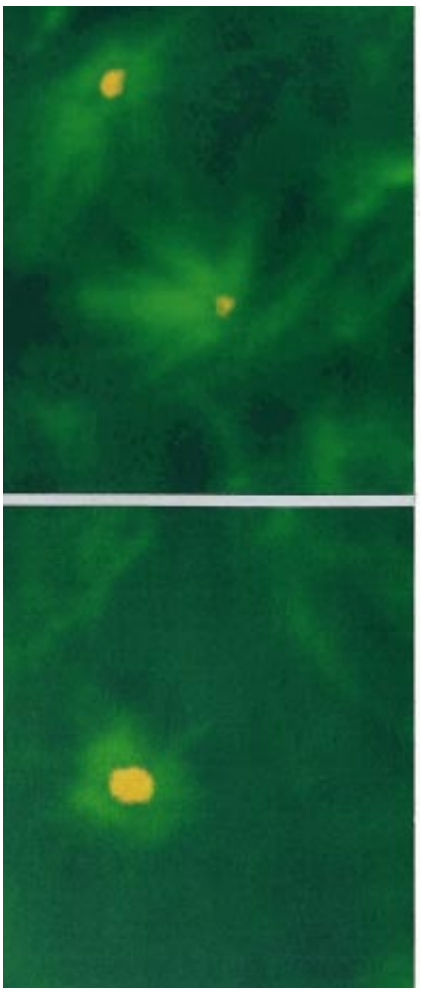


MITOSIS, or cell division, is orchestrated by the centrosomes. During interphase, a cell's centrosomes associate with the nucleus and are surrounded by starlike arrangements of microtubules called asters. During prophase, the centrosomes migrate toward opposite sides of the nucleus, the chromosomes begin to condense into distinct structures and the nuclear membrane breaks down. The microtubules extending from the centrosomes form a bipolar mitotic spindle. The chromosomes line up along the middle of the spindle during metaphase and then migrate toward each pole during anaphase. During telophase, the nuclear membranes re-form, and a deepening furrow in the outer plasma membrane pinches the cell in two.

Likewise, if the chromosomes are not aligned on the metaphase plate, anaphase cannot proceed and so on.

It is as though a set of checkpoint controls ensures that critical steps are performed correctly before allowing subsequent ones to proceed. Cells in the tissues of multicellular organisms also appear to cycle in such a regulated manner. In these cells it has been shown that blocking DNA synthesis prevents centrosome replication, which suggests that the two events may be linked.

Yet that connection cannot be generalized. The early embryos of many organisms undergo extremely rapid division cycles that seem to lack checkpoint regulation. The embryonic cells alternate rapidly between S phases and mitoses, without intervening gaps. Attempts to disrupt their cell cycles with either mutations or drugs often succeed in inhibiting only one replicative process. Other aspects can continue to cycle for some time.



MERRY-GO-ROUND MUTANTS of *Drosophila* fruit flies exhibit a centrosome that does not split. In a wild-type fly (*top*), the microtubules (*green*) form a mitotic spindle with the centrosomes (*yellow*) at the poles. In the mutant fly (*bottom*), a monopolar spindle forms. The chromosomes become arranged into a circle surrounding the centrosome.

For that reason, centrosome replication has been studied intensively in the rapid mitotic cycles of the embryos of many organisms. Many of the crucial experiments were carried out in the 1960s by Daniel Mazia, then at the University of California at Berkeley, and his colleagues, who uncoupled centrosomal and nuclear replication in the fertilized eggs of sand dollars and sea urchins. They proved that centrosome replication can occur in the absence of nuclear replication, although that dissociation does not normally occur in most cells. Our work with embryos of the fruit fly *Drosophila melanogaster* also illustrates the extent to which centrosomal and nuclear replication can be uncoupled. In those embryos the mitotic cycles are among the shortest known, lasting only 10 minutes. Even the most rapidly proliferating mammalian cells divide only once every 12 hours or so.

Drosophila development begins with 13 rounds of rapid nuclear division cycles, during which the replicating nuclei are not partitioned into separate cells by membranes. Instead all the nuclei reside in a common cytoplasm and membrane, or syncytium. After the first seven rounds of synchronous division, most of the nuclei migrate from the interior of the embryo toward the surface. The synchronous mitoses continue until about 6,000 nuclei are arranged in a monolayer just below the surface of the syncytial embryo. At that point, individual membranes grow around each nucleus and create a multicellular embryo. Thereafter, the cells adopt a regulated cycle of division.

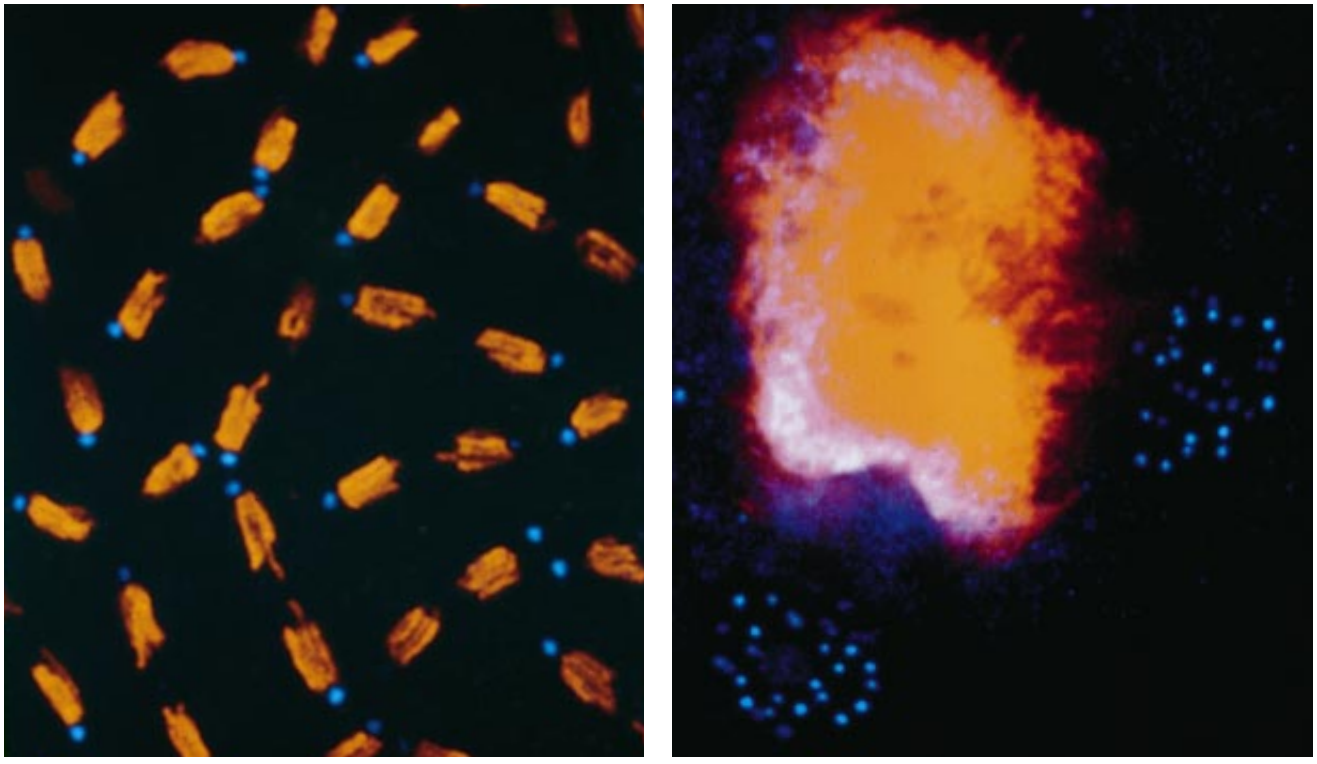
The early cycles of nuclear division are too short to permit the expression of the embryo's own genes—the embryonic DNA is either undergoing replication or segregation into daughter nuclei. Not until the cell cycle lengthens does expression of the embryo's own genes begin. All the protein requirements for the first 13 mitoses must therefore be supplied by the mother to her egg. That maternal dowry needs to be sufficient to make at least 6,000 nuclei and the associated mitotic apparatuses, including centrosomes and other components. Mutations in the maternal genes that are producing those proteins can cause defects in the embryonic mitoses. In the giant nucleus mutation of *Drosophila,* for example, the nuclear DNA and the centrosomes do go through repeated rounds of replication. For unknown reasons, the centrosomes dissociate from the nucleus. These embryos develop with a small number of nuclei that grow to gigantic proportions.

A related effect is seen in ordinary *Drosophila* embryos that have been injected with aphidicolin, a drug that inhibits an enzyme essential to DNA replication. In those embryos the centrosomes dissociate from nuclei and replicate independently. Other mitotic events also continue to cycle in aphidicolin-treated embryos, including the breakdown and reformation of the nuclear envelope and the condensation and decondensation of the chromosomes.

Mitosis in the early embryo seems to consist of several cyclic processes happening in parallel. Because they are only loosely coordinated with one another, they can run independently. We now know that the dissociation of centrosomes from nuclei and their apparently autonomous replication is a common feature of several mutations that affect the early mitoses in *Drosophila* embryos. At later developmental stages, those mutations have very different effects: usually they halt the cell cycle at some point, and often they prevent centrosome replication. The cell cycles of later *Drosophila* development seem to require the correct progress through a series of regulated checkpoints, much like those in yeast.

The accumulating evidence suggests that by nucleating microtubules, centrosomes indirectly influence the organization of other cytoskeletal elements, particularly the actin filaments. Perhaps the most striking manifestation of that effect is seen after mitosis, when the cytoplasm of the two daughter cells pinches apart. Several experimental treatments artificially activate the development of frog eggs in the absence of sperm. Yet because an embryo's functional centrosome is usually derived from the fertilizing sperm, eggs activated in those ways do not have a centrosome and cannot divide. Nevertheless, the egg proceeds through many aspects of the cell cycle as though it were attempting to divide. If purified centrosomes are injected into an activated egg, the egg can divide.

The centrosomes have that effect because they organize the cytoskeletal components essential to cell division. Actin and myosin filaments form a contractile ring between the two poles of the mitotic spindle. This ring, which is attached to the plasma membrane that surrounds the cell, gradually contracts, squeezing the cell in two. The mechanism responsible for generating the force of this contraction is similar to



REPLICATION of nuclear DNA and centrosomes is not always linked. In these micrographs, fluorescent dyes have stained DNA orange and the centrosomes blue. In a wild-type *Drosophila* fruit fly embryo (*left*) during anaphase, each set of chromosomes is pulled toward a centrosome. In giant nucleus mutants (*right*), the centrosomes dissociate from the nucleus. Both the centrosomes and the DNA continue to replicate but do so independently. Mitotic spindles do not form.

that in muscle contraction, which also involves interactions between actin and myosin filaments.

Remarkably, the contractile ring always assembles precisely halfway between the poles and perpendicular to the spindle. The two centrosomes must somehow organize actin and myosin filaments, but how they guide the assembly of the contractile ring at the midpoint between them is a mystery.

Centrosomes have the striking ability to direct the arrangement of actin filaments in the *Drosophila* embryo. As the rapidly dividing nuclei migrate to the embryo's surface, the cytoskeleton dramatically reorganizes itself in such a way that the actin at the surface of the embryo assembles into baskets of filaments over each nucleus. At the posterior pole of the embryo, much larger actin caps form over the nuclei, which pinch off from the surface to form the first set of cells. Later in development the progeny of those pole cells become the germ cells (the precursors of the eggs or of the sperm).

The nuclei at the surface of other regions of the embryo undergo four further rounds of division before membranes grow around them to form individual cells during cycle 14. That stage is reached about an hour after the pole cells form. If cytoplasm from the posterior pole of an embryo is injected into any region of a different embryo, the injected embryo will initiate pole cell formation as nuclei migrate to the site of the injection. This experiment demonstrates the presence of information in the posterior cytoplasm that initiates pole cell formation when nuclei arrive at the posterior cortex.

Recently we made the surprising discovery that centrosomes alone can trigger the formation of pole cells. We found that if aphidicolin is injected into a *Drosophila* embryo early enough, it inhibits nuclear division and as a consequence prevents nuclei from migrating to the surface. The centrosomes, however, continue to divide and to migrate.

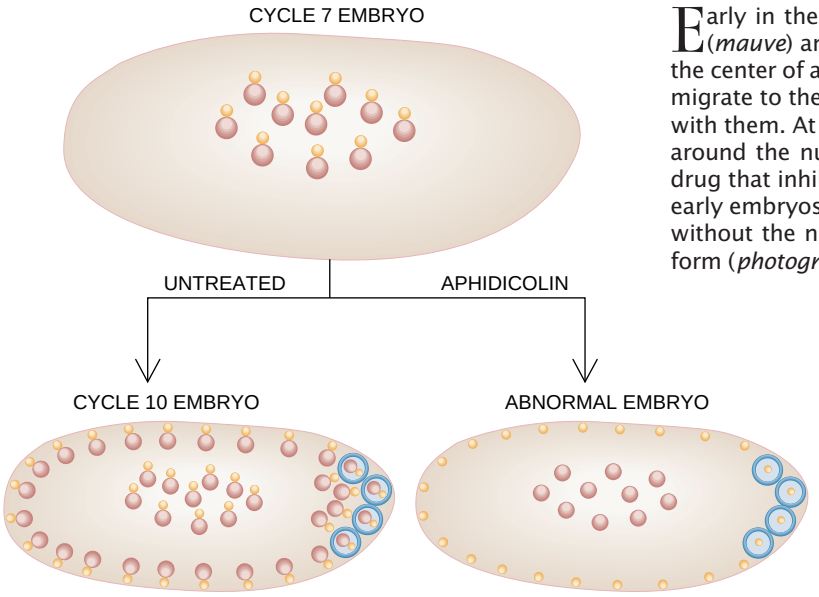
It is as though each centrosome were a locomotive that normally pulls a nucleus to the surface of the embryo along a railroad of microtubules. In aphidicolin-treated embryos the inhibition of DNA replication uncouples the nucleus from the centrosome, so only the centrosome migrates to the surface. Those centrosomes that reach the posterior pole still initiate the formation of pole cells—but each pole cell lacks a nucleus. The centrosomes at the surface of other regions of the embryo can spur the formation of actin caps despite the absence of nuclei, but they do not induce cell formation.

Thus, the centrosomes, in addition to being capable of organizing actin filaments, are able to respond to information in the posterior cytoplasm and to initiate pole cell formation. The nature of the interaction between the centrosomes and the posterior cytoplasm is unknown. Presumably, components of the posterior cytoplasm alter the properties of the cytoplasmic microtubules, with the result that the microtubules direct cell formation rather than just actin cap formation.

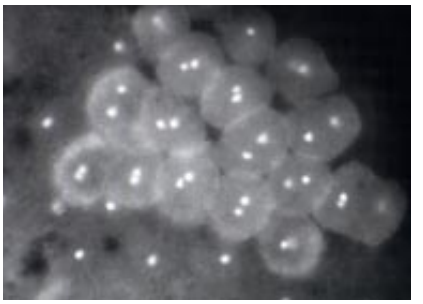
As we have described, centrosomes can influence other components of the cytoskeletal network by organizing microtubules. Those microtubules also have other vital functions: they organize the intermediate filaments; they help to determine cell polarity; they direct the intracellular transport of molecules; and they position other organelles, such as the Golgi apparatus and the endoplasmic reticulum, within a cell. In this way, the centrosomes and the microtubules they nucleate control many aspects of cellular organization.

Remarkably little is known about how the centrosome works at the molecular level. Although it will be important to characterize the individual macromolecules that make up the centrosome, it is

Centrosomes and Pole Cell Development



Early in the development of *Drosophila,* the nuclei (*mauve*) and the centrosomes (*yellow*) replicate at the center of a common cytoplasm. Later, centrosomes migrate to the surface of the embryo and bring nuclei with them. At the posterior end, pole cells (*blue*) form around the nuclei and centrosomes. If aphidicolin, a drug that inhibits DNA replication, is injected into the early embryos, the centrosomes replicate and migrate without the nuclei. They also still cause pole cells to form (*photograph*).



unlikely that studying those molecules in isolation will answer all the questions. To that end, researchers will need to reconstitute the function of the centrosome outside the cell, in a cell-free system in which individual components can be manipulated experimentally.

Several laboratories have already begun to develop such systems. Cell-free extracts of frog eggs hold particular promise. Recently Eric Karsenti of the European Molecular Biology Laboratory in Heidelberg and his colleagues added purified centrosomes to extracts from immature frog eggs, which are in an interphaselike state. The centrosomes nucleate arrays of long microtubules that have properties similar to those of cytoplasmic microtubules in interphase cells. If centrosomes are added to extracts from mature frog eggs, which are in a mitotic state, the centrosomes create arrays of much shorter, less stable microtubules, similar to those in the mitotic spindle. The added centrosomes therefore respond to cytoplasmic signals that change during the cell cycle.

One of these signals has been identified as a protein kinase, an enzyme that adds phosphate groups to specific cellular proteins and thereby alters their behavior. When added to frog egg extracts containing centrosomes and interphaselike arrays of microtubules, that protein kinase converts the microtubules into the more dynamic mitosis-like arrays. The same protein kinase seems to be required in all eukaryotic cells for many other aspects of the entry into mitosis.

Eventually, many components of the centrosome will be isolated, and it will be possible to reconstruct a functional centrosome in the test tube. Several centrosome-associated proteins have been identified using antibodies: our own laboratory at the University of Dundee used them to find and clone the gene for a protein in *Drosophila* that associates with the centrosome during mitosis. Other laboratories are using the inherent affinity of microtubules for the molecules with which they interact: purified microtubules can serve as a fishing line to hook molecular components of centrosomes.

This biochemical approach complements a genetic approach. Genetic analysis has proved to be a very powerful tool for identifying the components of many biological functions. Through treatment of the chromosomes of an organism with mutagenic drugs or ionizing radiation, it is possible to produce individuals in which specific genes have been functionally destroyed. By studying the resulting changes in those mutant individuals, researchers can infer the role of the normal form of the gene's protein product.

Such genetic studies with yeasts and fruit flies are already yielding mutations that alter centrosome behavior. The mutation merry-go-round is one that changes the behavior of centrosomes during mitosis. Instead of forming a normal bipolar spindle, the mutant cells produce a spindle that pulls all the chromosomes toward a single centrosome (hence the inspiration for the name of the mutation). The function of the merry-go-round gene is not yet understood, but it serves to demonstrate how single mutations can disrupt the functions of complex structures.

Given a sufficiently detailed description of such genetic interactions, investigators will eventually be able to deduce the functions of all the components of a centrosome. Genetic manipulations, in concert with biochemical studies, should make the goal of understanding centrosome function at the molecular level attainable. When that day arrives, cell biologists should finally be able to solve many of the puzzles that have troubled them for so long.

FURTHER READING

THE DISSOCIATION OF NUCLEAR AND CENTROSOMAL DIVISION IN *GNU,* A MUTATION CAUSING GIANT NUCLEI IN *DROSOPHILA*. M. Freeman, C. Nüsslein-Volhard and D. M. Glover in *Cell,* Vol. 46, No. 3, pages 457–468; August 1, 1986.

CENTROSOMES, AND NOT NUCLEI, INITIATE POLE CELL FORMATION IN *DROSOPHILA* EMBRYOS. J. W. Raff and D. M. Glover in *Cell,* Vol. 57, No. 4, pages 611–619; May 19, 1989.

MOLECULAR BIOLOGY OF THE CELL. Bruce Alberts et al. Garland Press, 1989.

WHAT CONTROLS THE CELL CYCLE. Andrew W. Murray and Marc W. Kirschner in *Scientific American,* Vol. 264, No. 3, pages 56–63; March 1991.

THE CENTROSOME. Edited by V. I. Kalnins. Academic Press, 1992.

The Future of the Transistor

As it has grown smaller and cheaper, engineers have scoffed at theoretical barriers to its progress—so far

by Robert W. Keyes

I am writing this article on a computer that contains some 10 million transistors, an astounding number of manufactured items for one person to own. Yet they cost less than the hard disk, the keyboard, the display and the cabinet. Ten million staples, in contrast, would cost about as much as the entire computer. Transistors have become this cheap because during the past 40 years engineers have learned to etch ever more of them on a single wafer of silicon. The cost of a given manufacturing step can thus be spread over a growing number of units.

How much longer can this trend continue? Scholars and industry experts have declared many times in the past that some physical limit exists beyond which miniaturization could not go. An equal number of times they have been confounded by the facts. No such limit can be discerned in the quantity of transistors that can be fabricated on silicon, which has proceeded through eight orders of magnitude in the 46 years since the transistor was invented [*see box on page 75*].

I do not have a definitive answer to the question of limits. I do, however, have some thoughts on how the future of solid-state electronics will develop and what science is needed to support continuing progress.

Several kinds of physical limitations might emerge as the size of the transistor continues to shrink. The task of connecting minute elements to one another might, for example, become impossible. Declining circuit size also means that researchers must cope with ever stronger electric fields, which can affect the movement of electrons in several ways. In the not too distant future the transistor may span only hundreds of angstroms. At that point, the presence or absence of individual atoms, as well as their behavior, will become significant. Diminishing size leads to increasing density of transistors on a chip, which raises the amount of waste heat that is thrown off. Today's chips shed about 10 times as much heat as does a cooking surface of comparable size, a flux that can rise at least 10-fold without exceeding the cooling capacity of known designs. As the size of circuit elements drops below the wavelength of usable forms of radiation, existing manufacturing methods may reach their limits.

To see how such problems might arise and how they can be addressed, it is useful to review the operation of the field-effect transistor, the workhorse of modern data processing. Digital computers operate by manipulating statements made in a binary code, which consists of ones and zeroes. A field-effect transistor is operated so that, like a relay, it is switched only "on" or "off." The device therefore represents exactly one binary unit of information: a bit. In a large-scale system, input signals control transistors, establishing connections that produce signals on output wires. The wires carry the signals to other switches that produce outputs, which are again sent on to another stage. The connections within the computer and the way in which the input signals determine an output together represent a logical statement. A series of such statements, in turn, determines a word in a document or an entry in a spreadsheet.

ROBERT W. KEYES is a research staff member at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y. His interests have focused on semiconductor physics and devices and on the physics of information-processing systems, subjects on which he has written and lectured widely; he has also received eight issued patents. A native of Chicago, Keyes studied at the University of Chicago, where he earned a doctorate in physics. He is an active participant in the programs of the National Research Council, the American Physical Society and the Institute of Electrical and Electronic Engineers.

MINIATURIZATION has made transistors cheaper than staples by spreading man-

The field-effect transistor contains a channel that interacts with three electrodes: a source, which supplies electrons to the channel; a drain, which receives them at the other side; and a gate, which influences the conductivity of the channel [*see illustration on next page*]. Each part contains different impurity atoms, or dopants, which modify the electrical properties of the silicon.

The gate switches the transistor on when a positive voltage applied to it attracts electrons to the interface between the semiconductor and the gate insulator. These electrons then establish a connection between the source and drain electrodes that allows current to be passed between them. At this point, the transitor is "on." The connection persists for as long as the positive charge remains on the gate. An incoming signal is applied to the gate and thus determines whether the connection between source and drain is established. If a connection results, the output is connected to the ground potential, one of the standard digital voltages. If no connection results, the output is connected through the resistor to the positive power supply, the other standard digital voltage.

Circuits of transistors must be oblivious to the operations of neighboring arrays. Existing concepts of insulation, impedance and other basic electrical properties of semiconductors and their connections should work well enough, for designers' purposes, in the next generation of devices. It is only when conducting areas approach to within about 100 angstroms of one another that

ufacturing costs over millions of devices on each of the hundreds of chips on a single wafer. This worker holds a nearly completed wafer. Its components will be connected by the condensation of metal in a vacuum chamber (*foreground*).

quantum effects, such as electron tunneling, threaten to create problems. In laboratory settings, researchers are already at the brink of this limit, at about 30 angstroms; in commercial devices, perhaps a decade remains before that limit is reached.

Another challenge is the strengthening of the electric field that inevitably accompanies miniaturization. This tendency constrains the design of semiconductor devices by setting up a basic conflict. Fields must continually get stronger as electron pathways shrink, yet voltages must remain above the minimum needed to overwhelm the thermal energy of electrons. In silicon at normal operating temperatures, the thermal voltage is 0.026 electron volt. Therefore, whenever a semiconductor is switched so as to prevent the passage of electrons, its electrical barrier must be changed by a factor several times as large. One can minimize the thermal problem by chilling the chip (an expensive proposition).

Even cooling cannot end the problem of the electric field. Signals must still have the minimum voltage that is characteristic of a semiconductor junction. In silicon this electrical barrier ranges between half a volt and a volt, depending on the degree of doping. That small voltage, applied over a very short distance, suffices to create an immensely strong electric field. As electrons move through such a field, they may gain so much energy that they stimulate the creation of electron-hole pairs, which are themselves accelerated. The resulting chain reaction can cause an avalanche of rising current, thereby disrupting the circuit. An avalanche begins when fields exceed about 500,000 volts per centimeter, compared with about 400,000 volts in today's chips.

Workers resort to a variety of tricks to mitigate the effects of strong electric fields. They have designed field-effect transistors, for example, in which the field can be moved to a place where it does not disrupt other electronic functions. This strategem is just one of many, all of which entail trade-offs with other desired characteristics, such as simplicity of design, ease of manufacture, reliability and long working life.

Miniaturization also increases the heat given off by each square centimeter of silicon. The reason is purely geometric: electrical pathways, and their associated energy losses, shrink in one dimension, whereas chip area shrinks in two. That relation means that as circuits get smaller, unit heat generation falls, albeit more slowly than does the number of units per square centimeter.

Devices already pour out as much as 30 watts per square centimeter, a radiance that one would expect of a material heated to about 1,200 degrees Celsius. Of course, the chips cannot be allowed to reach such temperatures, and so cooling systems remove heat as fast as it is produced. We have not even come close to exhausting the existing cooling technologies, which can siphon off energy many times faster than is now required.

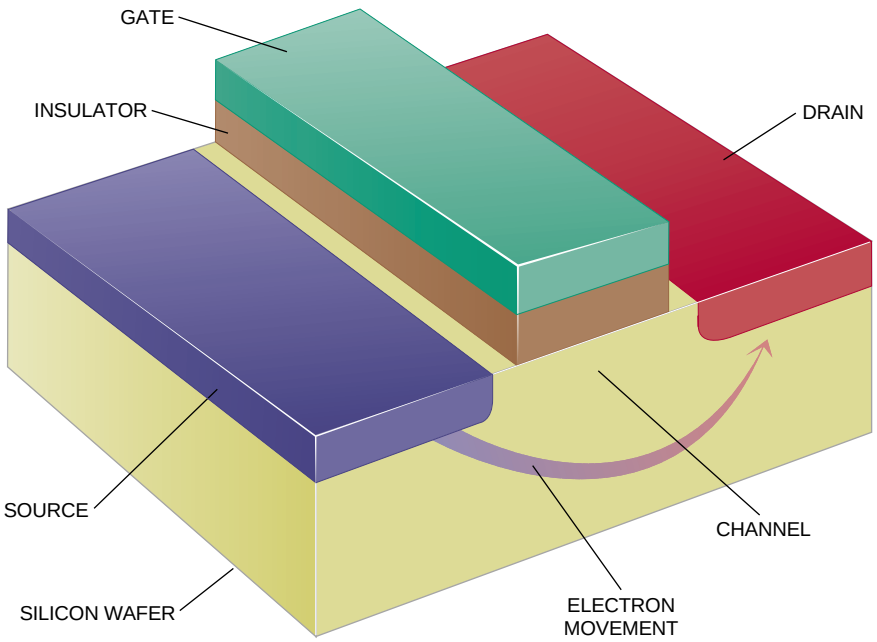
The exigencies of manufacturing impose constraints on the performance of electronic devices that might not be apparent from a purely theoretical discussion. Low-cost manufacturing leads, perforce, to small differences among the devices that are made on each wafer, as well as among those fabricated on different wafers. This variability cannot be banished—it is inherent in the way solid-state devices are made.

A semiconducting material, such as silicon, is made into a transistor in an integrated process involving many steps. Templates, called masks, are applied to the silicon in order to expose desired areas. Next, various operations involving chemical diffusion, radiation, doping, sputtering or the deposition of metal act on these areas, sometimes by constructing device features, other times by erecting scaffolding to be used in succeeding steps and then torn down. Meanwhile other devices—resistors, capacitors and conductors—are being built to connect the transistors.

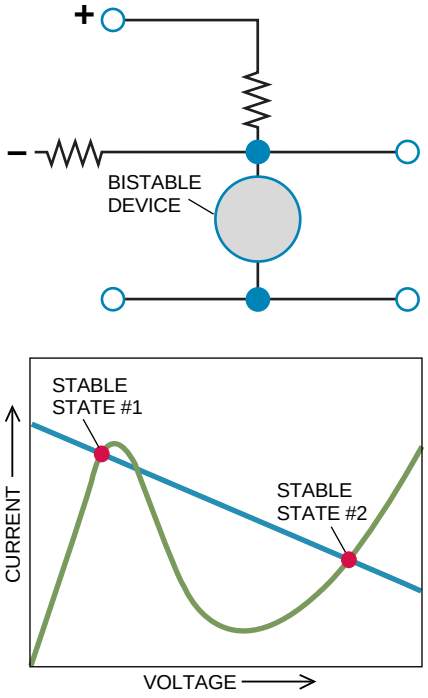
Variations intrude at every step. For example, perfect focusing of the source of radiation over a large wafer is hard to achieve. The temperature of the wafer may vary slightly from one place to another during processing steps, causing a difference in the rate of chemical reactions. The mixing of gases in a reaction chamber may not be perfect. For many reasons, the properties of devices on a given wafer and between those on different wafers are not identical. Indeed, some devices on a wafer may be no good at all; the proportion of such irremediable errors places a practical limit on the size of an integrated circuit.

A certain amount of fuzziness is inherent in optical exposures. The light used in photolithography is diffracted as it passes through the holes in the template. Such diffraction can be minimized by resorting to shorter wavelengths.

When photolithographic fabrication was invented in the early 1970s, white light was used. Workers later switched to monochromatic laser light, moving up the spectrum until, in the mid-1980s, they reached the ultraviolet wavelengths. Now the most advanced commercial chips are etched by deep ultraviolet light, a difficult operation because it is hard to devise lasers with output in that range. The next generation of devices may require x-rays. In-



FIELD-EFFECT TRANSISTOR, the workhorse of data processing, is built as a sandwich of variously doped silicon layers. It contains a channel, a source, a drain and an insulated gate. When a positive voltage is applied to the gate, electrons move near the insulation, establishing a connection underneath it that allows current to pass from source to drain, switching the transistor on.



BISTABLE CIRCUIT does the transistor's job by exploiting nonlinear effects. A device such as a tunnel diode is placed at the junction of two main electrodes and a minor one (*top*). If the minor electrode injects some extra current, the circuit will move from one stable state to the other (*bottom*). Such devices are impractical because they cannot tolerate much variation in signal strength.

deed, each generation of circuitry requires manufacturing equipment of unprecedented expense.

Other problems also add to the cost of making a chip. The mechanical controls that position wafers must become more precise. The "clean rooms" and chambers must become ever cleaner to ward off the ever smaller motes that can destroy a circuit. Quality-control procedures must become more elaborate as the number of possible defects on a chip increases.

Miniaturization may at first glance appear to involve manipulating just the width and breadth of a device, but depth matters as well. Sometimes the third dimension can be a valuable resource, as when engineers sink capacitors edgewise into a chip to conserve space on the surface. At other times, the third dimension can constrain design. Chip designers must worry about the aspect ratio, that is, the relation of depth to surface area. Their device resembles a sandwich (sometimes even a Dagwood sandwich). If the layers are multiplied excessively, etching will tend to undercut the structure before it can be completed. Such constraints compel workers to achieve tighter control over the thickness of films and over the penetration of impurities into the semiconductor. At present levels of control, about 20 layering operations—called mask steps—are feasible.

The formulas that are used to design large devices cannot be used for the tiny transistors now being made in laboratories. Designers need to account for exotic new phenomena that appear in such extremely small devices. Because the effects cannot be accurately treated by purely analytic methods, the designers have recourse to computer models that simulate the motion of electrons in a device.

A computer follows a single electron through a device, keeping track of its position as time is increased in small steps. Physical theory and experimental information is used to calculate the probability of the various events that are possible. The computer uses a table for the probabilities, stored in its memory, and a random number generator to simulate the occurrence of these events. For example, an electron is accelerated by an electric field, and the direction of its motion might be changed by a collision with an impurity. Adding the results of thousands of electrons modeled in this fashion gives a picture of the response of the device.

Consider the seemingly trivial question of how to represent the motion of an electron within an electric field. When path lengths were comparatively long, an electron quickly accelerated to the point at which collisions robbed it of energy as fast as the field supplied new energy. The particle therefore spent most of its time at a constant velocity, which can be modeled by a simple, linear equation. But when path lengths became shorter, the electron no longer had time to reach a stable velocity. The particles now accelerate all the time, and the equations must account for that complication.

If such difficulties can arise in modeling a well-understood phenomenon, what lies ahead as designers probe the murky physics of the ultrasmall? Simulations can be no better than the models that physicists make of events that happen in small spaces during short periods. To refine these models, researchers need to carry out experiments on femtosecond time scales.

More knowledge of solid-state physics is required, because as chips grow more complex they require more fabrication steps, and each step can influence the next. For instance, when doping atoms are introduced into a crystal, they tend to attract, repel or otherwise affect the motion of other dopants. Such effects of dopants on other dopants are not well understood; further experiments and theoretical investigations are therefore needed. Chemical reactions that take place on the surface of a silicon crystal demand a supply of silicon atoms, a kind of fluid flow within the solid lattice; does such motion carry other constituents along with it? These questions did not exercise designers of earlier generations of chips, because the transistors were then large enough to swamp such ultramicroscopic tendencies.

The prospect of roadblocks aside, the transistor has only itself to blame for speculation about alternative technologies. Its extraordinary success in the 1950s stimulated an explosive development of solid-state physics. In the course of the work, investigators discovered many other phenomena, which in turn suggested a host of ideas for electronic devices. Several of these lines of research produced respectable bodies of new engineering knowledge but none that led to anything capable of even finding a niche in information processing.

Some workers have argued that the transistor owes its preeminence to being first off the block. Because of that head start, semiconductors have been the center of research, a position that guarantees them a margin of technological superiority that no rival can match. Yet I believe the transistor has intrinsic virtues that, by themselves, could probably preserve its dominant role for years to come.

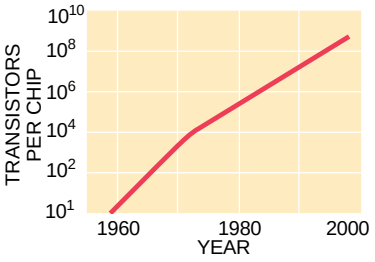
I participated, as a minor player, in some of the efforts to build alternative switches, the repeated failures of which made me wonder what was missing. Of course, quite a few new fabrication methods had to be developed to implement a novel device concept. But even though these could be mastered, it was difficult to get a large collection of components to work together.

What gave the transistor its initial, sudden success? One difference stood out: the transistor, like the vacuum tube before it, has large gain. That is, it is capable of vastly amplifying signals of the kind processed in existing circuits, so that a small variation in input can produce a large variation in output. Gain makes it possible to preserve the integrity of a signal as it passes through many switches.

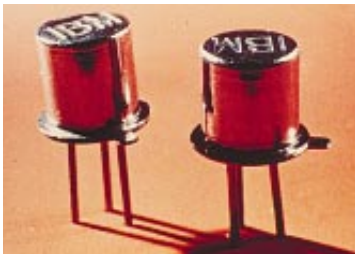
Rivals to the transistor may have been equally easy to miniaturize, but they exhibited far less gain. Take, for instance, bistable devices [*see illustration on this page*], which perform log-

The Shrinking Transistor

Miniaturization is manifest in this comparison between an electromechanical switch, circa 1957, and a recent chip containing 16 million bits of memory (*right*). Progress appears in these snapshots (*counterclockwise*): Bell Laboratories' first transistor (1948); canned transistors (1958); salt-size transistors (1964); 2,000-bit chip (1973); boards with 185,000 circuits and 2.3 megabits of memory apiece (1985); and a 64-megabit memory chip (1992).

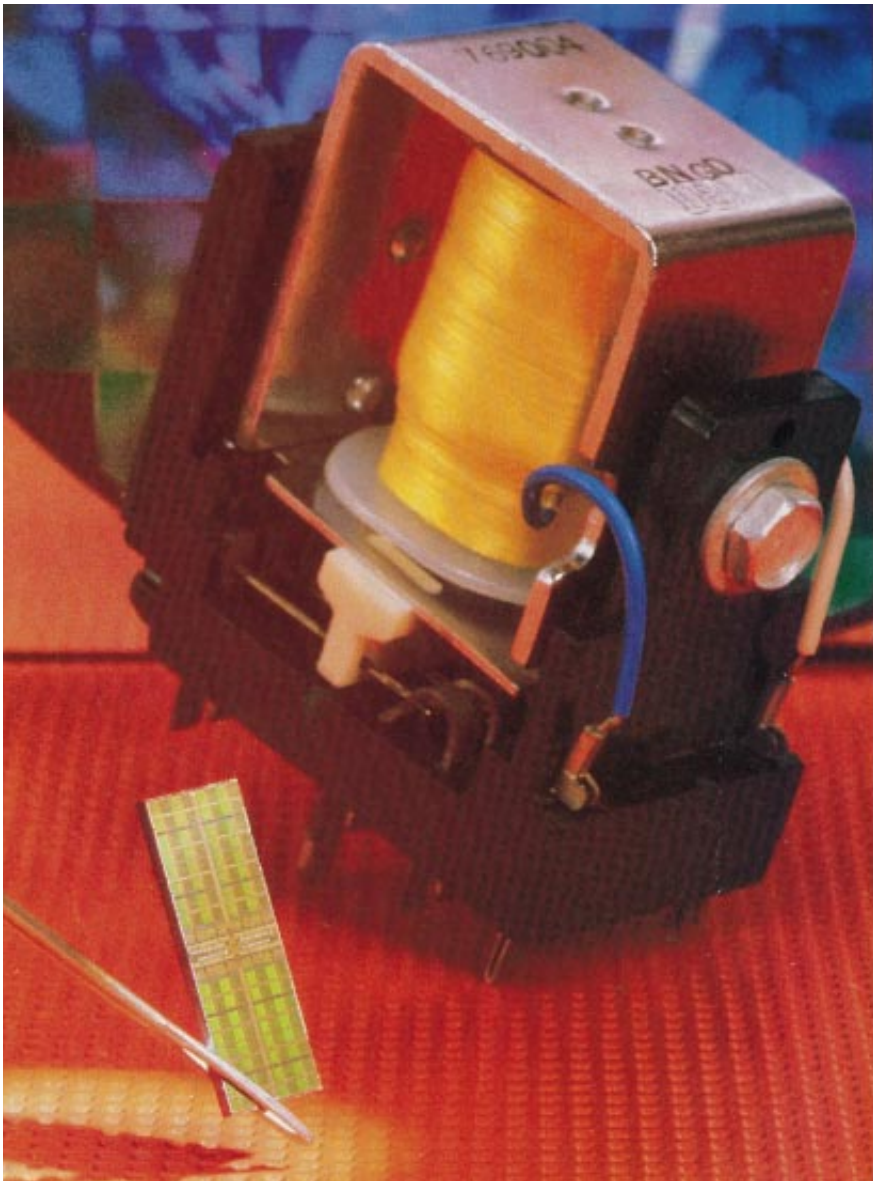


The first transistor

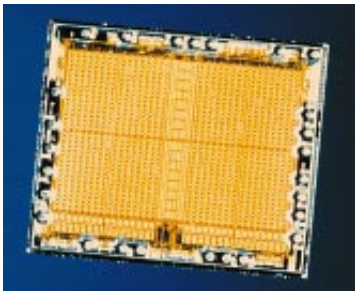


Early commercial transistors

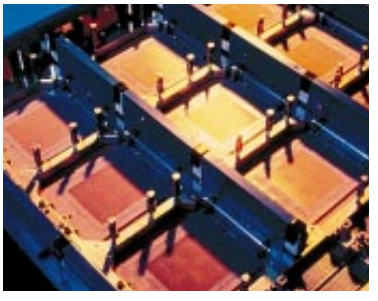
Salt-size transistors



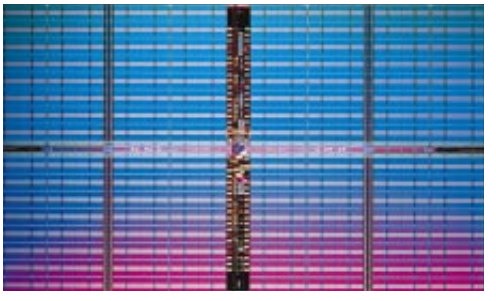
Switches, now and then



Early integrated circuit



Circuit assembly



Contemporary memory chip

ic functions by moving between two stable states that are separated by an unstable transition. Researchers have produced such a transition by designing circuits having a range of values in which current declines as voltage increases. Any slight disturbance, such as that obtained by injecting extra current through the device, will switch the circuit between its two stable states. Because this slight input can bring about large changes in the current and voltages, there is a sense in which gain is achieved.

Yet the gain is far less useful than that provided by a transistor because it operates within narrow tolerances. A bistable switch thus performs deceptively well in the laboratory, where one can fine-tune the circuit so it stays near the crossover point. A collection of such switches, however, does not lend itself to such painstaking adjustments. Because not all the circuits will work, no complex device can be based on their operation. Negative resistance therefore plays no role in practical data processing.

IMMENSE AND DENSE: this active-matrix liquid-crystal panel shows that today's electronic structures can achieve great complexity over large areas. Each liquid-crystal pixel is controlled by its own transistor, providing extraordinary resolution.

The same difficulty has plagued the development of nonlinear optical devices, in which the intensity of optical beams replaces the currents and voltages of electrical circuits. Here, too, the operation depends on fine-tuning the system so that a small input will upset a delicate balance. (Such switches have occasionally been termed "optical transistors," a label that misconstrues the principles of transistor action.) Optical switches face a problem even more fundamental. Light, unlike electricity, hardly interacts with light, yet the interaction of signals is essential for logic functions. Optical signals must therefore be converted into electrical ones in a semiconductor. The voltage thus produced changes the optical response of another material, thereby modulating a beam of light.

Another proposed switch, sometimes called a quantum interference device, depends on the interference of waves. In the most familiar case, that of electromagnetic radiation, or light, one wave is divided into two components. The components begin oscillating in phase, that is, their peaks and troughs vibrate in tandem. If the components follow routes of different lengths before reuniting, the phase relation between their waveforms will be changed. Consequently, the peaks and troughs either cancel or reinforce one another, producing a pattern of bright and dark fringes. The displacement of the fringes measures the relative phase of the system.

Electrons also possess a wave nature and can be made to interfere. If the two components of a wave move at equal speeds over similar paths to a rendezvous, they will reconstitute the original wave; if they move at different speeds, they will interfere. One can manipulate the velocity of one wave by applying a tiny electric field to its pathway. The correct field strength will cause the waves to cancel so that no current can flow through the device.

At first sight, this action duplicates a field-effect transistor, which uses an electric field to control a current through a semiconductor. In an interference device, however, conditions must be just right: if the applied voltage is too high or too low, there will be some current. This sensitivity means that an interference device will not restore the binary nature of a degraded input signal but will instead add its own measure of noise. Data passing from one such device to another will quickly degenerate into nothingness.

The lack of real rivals means that the future of digital electronics must be sought in the transistor. The search begins anew with each voyage into a smaller scale or a different material. The latest reexamination was occasioned by the introduction of new semiconductor materials, such as gallium arsenide and related compounds, several of which may even be incorporated into a single device. These combinations may be used to produce what are called heterojunctions, in which crystalline lattices of different energy gaps meet. Lattices may mesh imperfectly, creating atomic-scale defects, or they may stretch to fit one another, in the process creating an elastic strain. Either defects or strain can produce electrical side effects.

These combinations complicate the physics but at the same time provide a variable that may be useful in surmounting the many design problems that miniaturization creates. For instance, the dopants that supply electrons to a semiconductor also slow the electrons. To reduce this slowing effect, one can alternate layers of two semiconductors in which electrons have differing energies. The dopants are placed in the high-energy semiconductor, but the electrons they donate immediately fall into the lower-energy layers, far from the impurities.

What, one may ask, would one want with a technology that can etch a million transistors into a grain of sand or put a supercomputer in a shirt pocket? The answer goes beyond computational power to the things such power can buy in the emerging information economy. It has only recently been taken for granted that anyone with a personal computer and a modem can search 1,000 newspapers for references to anything that comes to mind, from kiwifruit to quantum physics. Will it soon be possible for every person to carry a copy of the Library of Congress, to model the weather, to weigh alternative business strategies or to checkmate Gary Kasparov?

FURTHER READING

FIELD-EFFECT TRANSISTORS IN INTEGRATED CIRCUITS. J. T. Wallmark and L. G. Carlstedt. John Wiley & Sons, 1974.

THE PHYSICS OF VLSI SYSTEMS. R. W. Keyes. Addison-Wesley Publishing, 1987.

CAN WE SWITCH BY CONTROL OF QUANTUM MECHANICAL TRANSMISSION? Rolf Landauer in *Physics Today,* Vol. 42, No. 10, pages 119–121; October 1989.

LIMITATIONS, INNOVATIONS, AND CHALLENGES OF CIRCUITS AND DEVICES INTO A HALF MICROMETER AND BEYOND. M. Nagata in *IEEE Journal of Solid-State Circuits,* Vol. 27, No. 4, pages 465–472; April 1992.

THE FUTURE OF SOLID-STATE ELECTRONICS. R. W. Keyes in *Physics Today,* Vol. 45, No. 8, pages 42–48; August 1992.

Monogamy and the Prairie Vole

Studies of the prairie vole—a secretive, mouselike animal—have revealed hormones that may be responsible for monogamous behavior

by C. Sue Carter and Lowell L. Getz

Observation of the mating and pup-rearing habits of nondescript, brown rodents that live under weeds and grasses might not seem an obvious way to improve knowledge of monogamy. After all, most humans can attest to the complexity of male-female relationships. Yet studies of the prairie vole (*Microtus ochrogaster*), a common pest throughout the midwestern U.S., have led us on a fascinating scientific journey from our starting point in ecology to the exploration of the neuroendocrinology of social bonds. Unlike most rodents, prairie voles form long-lasting pair bonds, and both parents share in raising their young. Our studies have provided a new understanding of the importance of two hormones, oxytocin and vasopressin, which are well known for their respective roles in reproduction and body water regulation. Work with voles now suggests that these hormones are involved in the development of monogamy.

The chief criterion that defines monogamy is a lifelong association between a male and a female. Within this broad definition lie several characteristics that are easily observed. Males and females of monogamous species tend to be about the same in size and appearance. Mated pairs will defend the nest and territory from intruders, and both parents care for the young. Monogamous mammals may form complex social groups that include an extended family and offspring of various ages. Incest is avoided within these families; adult young usually do not reproduce as long as they live with related family members. Finally, we should point out that although common in birds, monogamy is rare in mammals. In an exhaustive survey, Devra G. Kleiman of the National Zoological Park in Washington, D.C., found that only about 3 percent of mammals are monogamous.

Sexual exclusivity, however, is not a feature of monogamy. Studies of the prairie vole as well as those of other mammals and birds have indicated that absolute sexual monogamy is not necessarily associated with social monogamy. In fact, DNA fingerprinting tests have shown that offspring of female prairie voles are not always fathered by the cohabiting males. In some cases, a litter may have mixed paternity.

Because prairie voles incorporate the defining features of monogamy, they make excellent subjects for the exploration of the biological foundations of monogamy, at least as it exists among nonhumans. Prairie voles are also small, weighing only a few ounces, and are easily reared in the laboratory. But of particular importance for understanding the biology of monogamy is the fact that not all voles are monogamous. The meadow vole (*M. pennsylvanicus*) and the montane vole (*M. montanus*) show no indications of monogamy. Voles of these species are rarely retrapped with the same partner and do not establish stable families, and males of these species do not usually care for their young. Therefore, comparisons of prairie voles with their nonmonogamous relatives can yield insights into the causes of monogamy.

One of the first surprises that came from studies of prairie voles was the observation that social cues regulate the reproductive physiology of this species. Even to enter estrus (sexual heat), a female prairie vole must sniff a male. Indeed, Milo E. Richmond, now at Cornell University, found that female prairie voles do not have the ovarian cycles that are typical of nonmonogamous mammals. In monogamous voles, a female must have a male partner to induce estrus.

Furthermore, not just any male can bring a female into heat. Fathers and brothers do not seem capable of eliciting sniffing. This may be an adaptive mechanism designed to prevent incest. In fact, both males and females will essentially remain prepubescent as long as they stay with their families.

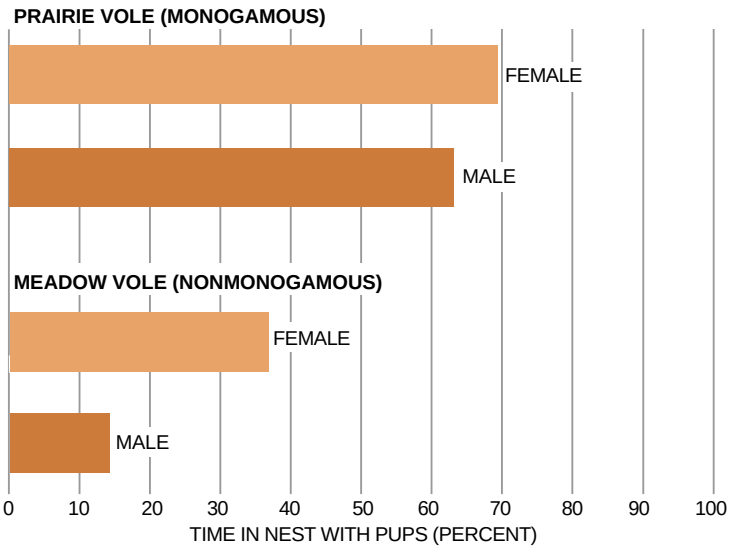
By sniffing an appropriate male, the female picks up a chemical signal called a pheromone. Pheromones in turn trigger the hormonal events needed to activate the ovaries and to induce heat. A small chemical sense organ, known as the vomeronasal organ, helps to mediate the effects of pheromones. John J. Lepri, now at the University of North Carolina at Greensboro, and Charles J. Wysocki of the Monell Chemical Senses Center in Philadelphia found that removal of the vomeronasal organ in the female prevented the start of heat. A similar effect occurs when the olfactory bulb is detached. By removing the bulb, Jessie R. Williams, Brian Kirkpatrick and Burton Slotnick, working in our University of Maryland laboratory, disrupted the sexual and social behaviors of the prairie voles.

In our laboratory, Dean E. Dluzen found that once a female is exposed

C. SUE CARTER and LOWELL L. GETZ have collaborated on their respective studies of behavioral endocrinology and fieldwork to investigate the biology of monogamy. Both are fellows of the American Association for the Advancement of Science. Carter received her Ph.D. from the University of Arkansas at Fayetteville. Before taking on her current position as professor of zoology at the University of Maryland, she was professor of psychology and ecology, ethology and evolution at the University of Illinois. Getz heads the department of ecology, ethology and evolution at the University of Illinois. He received his Ph.D. from the University of Michigan. The work described here reflects a 15-year collaboration and arose from observations made during approximately 35 years of fieldwork by Getz.

PRAIRIE VOLES (*Microtus ochrogaster*) engage in prolonged periods of mating, often past the time that is needed to ensure pregnancy. These extended bouts may help to facilitate the formation of monogamous social bonds.

to male odors, levels of norepinephrine, a neurotransmitter, and luteinizing hormone–releasing hormone (LHRH) change within minutes. These biochemical events occurred within the area of the olfactory bulb that receives input from the vomeronasal system. The stimulation of the olfactory system and the secretion of LHRH cause the pituitary gland to release a surge of luteinizing hormone into the bloodstream. In conjunction with other endocrine changes, the release of luteinizing hormone begins a cascade of chemical and neural events that stimulates the ovary to secrete gonadal steroids. Two of the most important steroids secreted are estradiol, a potent kind of estrogen, and progesterone.



PARENTAL CARE demonstrated by prairie voles far exceeds that shown by nonmonogamous meadow voles. The difference is most apparent with male prairie voles, which are with the pups four times as often as male meadow voles are.

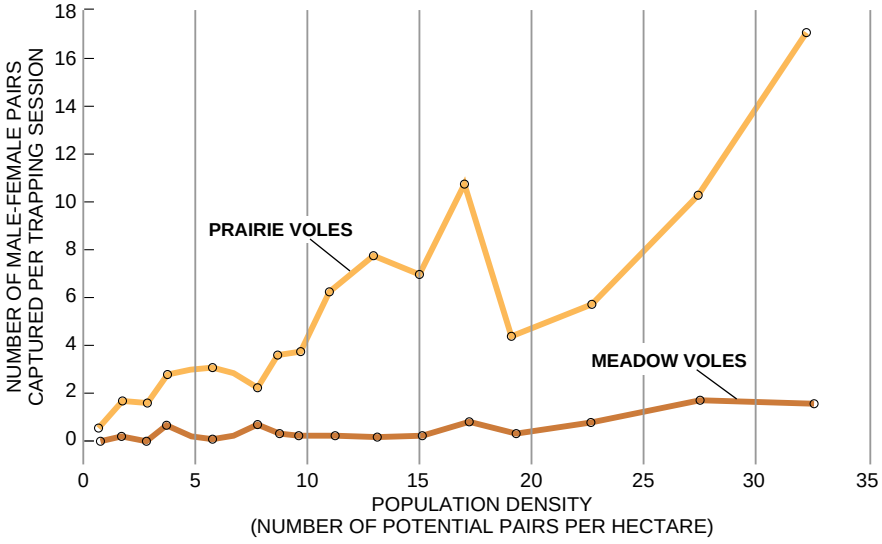
Might estrogen and progesterone also be involved in monogamous behavior? In collaboration with Janice M. Bahr of the University of Illinois, we searched for patterns of gonadal steroid production that varied between estrus and nonestrus female prairie voles and compared the results with data from nonmonogamous species. Estradiol, a hormone known to be essential in inducing estrus in rodents, was elevated only in female prairie voles in heat. It declined after mating. This pattern is similar to that displayed by polygamous rodents. Analysis of the patterns of progesterone levels, however, presented an unexpected finding. In the nonmonogamous rats and montane voles, progesterone is released in the bloodstream shortly after mating begins. This rise in progesterone probably helps to regulate the duration of sexual activity by bringing these rodents into and out of heat. In contrast, we found that in prairie voles progesterone levels in the blood did not increase until many hours after coitus began.

The delayed secretion of progesterone explains an observation made in previous studies: that female prairie voles in their first estrus mate for prolonged periods. In our laboratory, Diane M. Witt observed that when the female was in natural estrus, males and females continued to engage in bouts of mating for about 30 to 40 hours. This extended mating period contrasts sharply to that seen in nonmonogamous species. Mating in meadow and montane voles persists for a few hours, and Syrian hamsters become nonreceptive after about 45 minutes of mating.

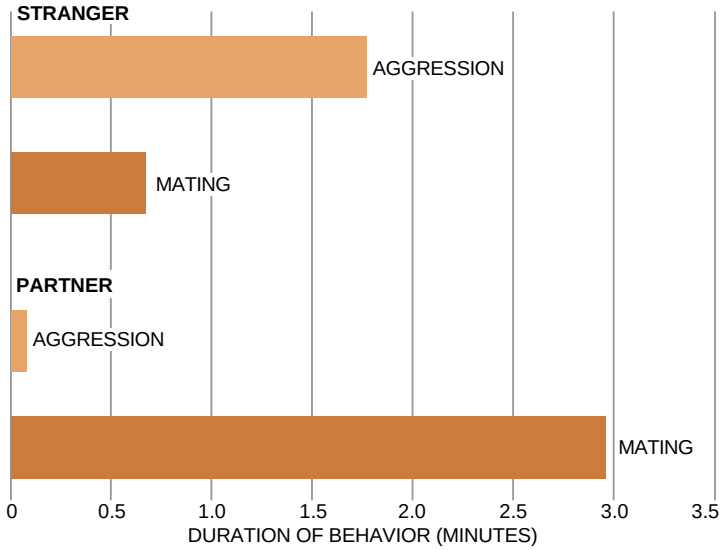
It is possible that the lengthy sexual interactions of prairie voles help the sperm enter the uterus and reach the egg. Studies of rats by Norman T. Adler of the University of Pennsylvania have shown that complex patterns of sexual behavior can influence the release of hormones and alter the ability of sperm to enter the female's reproductive tract and fertilize an egg.

Yet improving the chances of fertilization is probably not the sole reason for these extended bouts of mating. Once mating begins, females ovulate within about 12 hours, and successful pregnancy can occur shortly thereafter. Thus, prairie voles in their first heat continue to copulate for hours after they have met the requirements for pregnancy.

We suspect that, like humans and some other primates, prairie voles may copulate to facilitate the formation of monogamous social bonds. Protracted mating would be particularly crucial for prairie voles that are interacting for the first time, because they need to establish their lifelong monogamous bond. Indeed, some evidence for this idea comes from observations of females that have previously mated and become



MALE-FEMALE PAIRS of prairie voles are caught far more frequently than are such pairs of meadow voles. Furthermore, the same pairs are often captured repeatedly. Such studies provided the first clue of monogamy in the prairie vole.



AGGRESSION by female prairie voles is revealed in 10-minute tests comparing hostility with mating preference. Females spent more time attacking strangers rather than mating with them. In contrast, they show little aggression toward their partners.

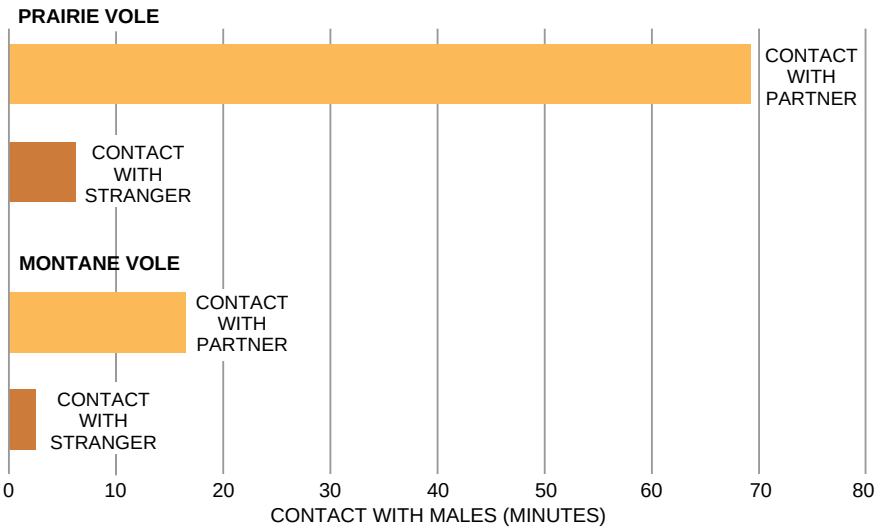
pregnant. Witt found that these experienced females engaged in brief copulations, sometimes limited to a few minutes. Having established a social bond, experienced males and females may not need to mate for long periods.

Social interaction that follows mating may be one of the mechanisms that reinforces monogamy in a species. Such interplay in nonmonogamous species often is restricted to a brief interval when the female is in heat. For example, Michael H. Ferkin, now at Cornell, observed that male and female meadow voles did not remain in physical contact after mating. In the Syrian hamster, which is an especially solitary animal, one of us (Carter) found that a female that has mated becomes extremely aggressive toward the male. In fact, the female may try to kill her sexual partner if he does not leave after coitus.

In contrast, mated monogamous mammals remain highly social toward their mates, even during nonreproductive periods. Leah Gavish, in our laboratory, demonstrated that prairie voles often touch and remain near their sexual partner. But this friendliness does not extend to strangers. After mating, both males and females became exceptionally aggressive toward unfamiliar members of their own sex. In nature, this behavior translates into territoriality or mate defense. In the laboratory we have used this model to examine the physiological processes responsible for pair bonding.

Specifically, we hypothesized that hormonal events induced by copulation might account for the dramatic behavioral changes that occurred after mating. Working in our laboratory, Kerry O'Banion took a first step toward examining this idea. O'Banion studied how females choose male partners before and after mating. In his experiments, familiar and unfamiliar males were tethered at opposite ends of an arena. O'Banion gave a female 10 minutes to choose. For the most part, females chose to mate with familiar and unfamiliar males equally. But if they had lived with a male, females showed a tendency to engage in nonsexual physical contact with the familiar male, not the stranger. These results illustrate the importance of social contact as an index of partner choice. They also confirm the DNA tests revealing that in nature female voles do not show absolute sexual monogamy.

More recently Williams examined female preferences in tests that lasted for at least three hours. She placed female prairie voles in a relatively large maze that contained three chambers. The females could elect to spend time alone or with males tethered in the two other chambers. The animals were monitored on videotape for their social and sexual preferences. After exploring both the stranger and the partner for about



SOCIAL PREFERENCE for sexual partners is demonstrated by female prairie voles in three-hour tests. They prefer contact with their partners over contact with male strangers. Female montane voles actually spend more time alone.

30 minutes, females usually chose the familiar male.

In similar studies, Williams discovered that a female in her first heat developed a preference for a male if she was allowed to live with him for at least 24 hours. If the pair copulated, however, cohabitation produced clear social preferences in as few as six hours. These studies demonstrate that some aspect of the sexual interaction hastens the onset of a partner choice. We believe that hormones or neurochemicals released during mating or cohabitation may explain the experimental results.

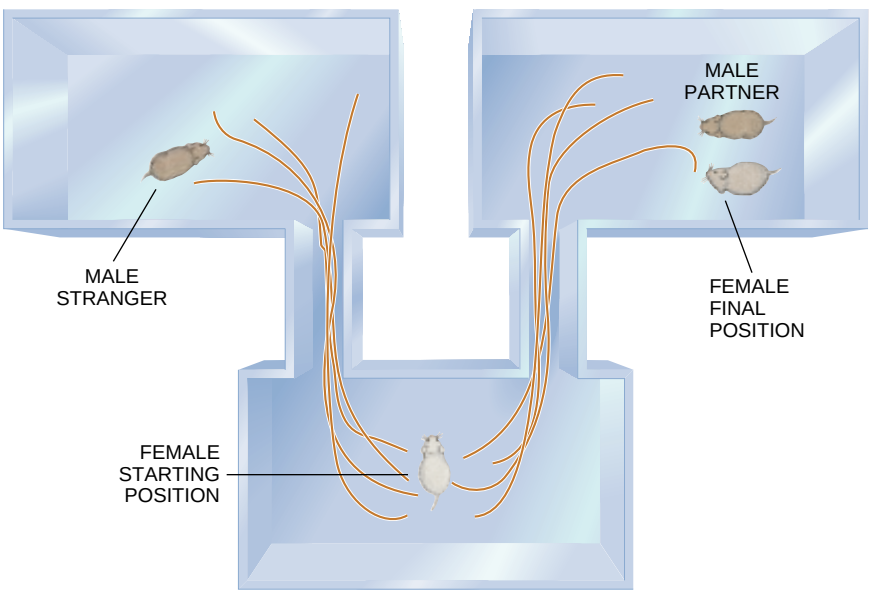
One clue to the identity of the hormones came from work by Peter H. Klopfer of Duke University. He recognized that social bonds between mothers and their offspring were associated with the release of oxytocin and hypothesized that the compound might be the hormone of "mother love." Niles Newton of Northwestern University extended these observations to speculate that maternal and sexual bonds could be influenced by the secretion of the hormone. Oxytocin is produced primarily as a result of breast or genital stimulation, such as that which occurs during mating, birth and lactation. More recently E. Barry Keverne and Keith M. Kendrick of the University of Cambridge have shown that in sheep either vaginal stimulation or oxytocin treatments can speed the formation of mother-infant bonds. Kerstin Uvnäs-Moberg of the Karolinska Institute in Stockholm has demonstrated that even simple touch can release oxytocin.

Based on these studies, we hypothesized that in prairie voles stimulation experienced during mating, or perhaps more slowly by touch and cohabitation, might release oxytocin. Oxytocin would, in turn, hasten the formation of social bonds between males and females.

Several recent findings support this supposition. Witt injected oxytocin into the central nervous system of females. As a result, the females became more sociable and less likely to fight with males, as compared with females that did not receive the oxytocin or females that received the hormone administered into the peripheral circulation. The positive social effects of oxytocin in the brain have now been documented in other species. Witt found improved social behavior in rats, and James Winslow and Thomas R. Insel of the National Institute of Mental Health (NIMH) reported similar results in squirrel monkeys.

Williams examined the role of oxytocin more directly. She repeated her preference tests on prairie voles whose cerebral ventricles were infused with oxytocin. She found that females formed rapid preferences for males if they were exposed to oxytocin over a six-hour period. But when combined with a drug that blocks the oxytocin receptors, oxytocin no longer exerted the social effect. These results suggest that oxytocin's action within the brain may be one of the physiological events that lead to the formation of monogamous pairs.

Because the receptors for a hormone can regulate the behavioral effects of that hormone, we also looked at the patterns of oxytocin receptors in the prairie vole. These receptors are scattered throughout the mammalian central nervous system. Witt found that the distribution of oxytocin receptors in prairie voles differed from the pattern in rats. The differences were especially striking within the limbic system, the area of the brain involved in sexual and social behavior. Insel and his NIMH colleague Larry E. Shapiro subsequently showed that the distribution of oxytocin receptors in prairie voles and in pine voles, another monogamous species, differs from that in the polygamous montane and meadow voles. That the patterns of oxytocin receptors correlate with monogamy further substantiates the idea that oxytocin has an essential role in social organization.

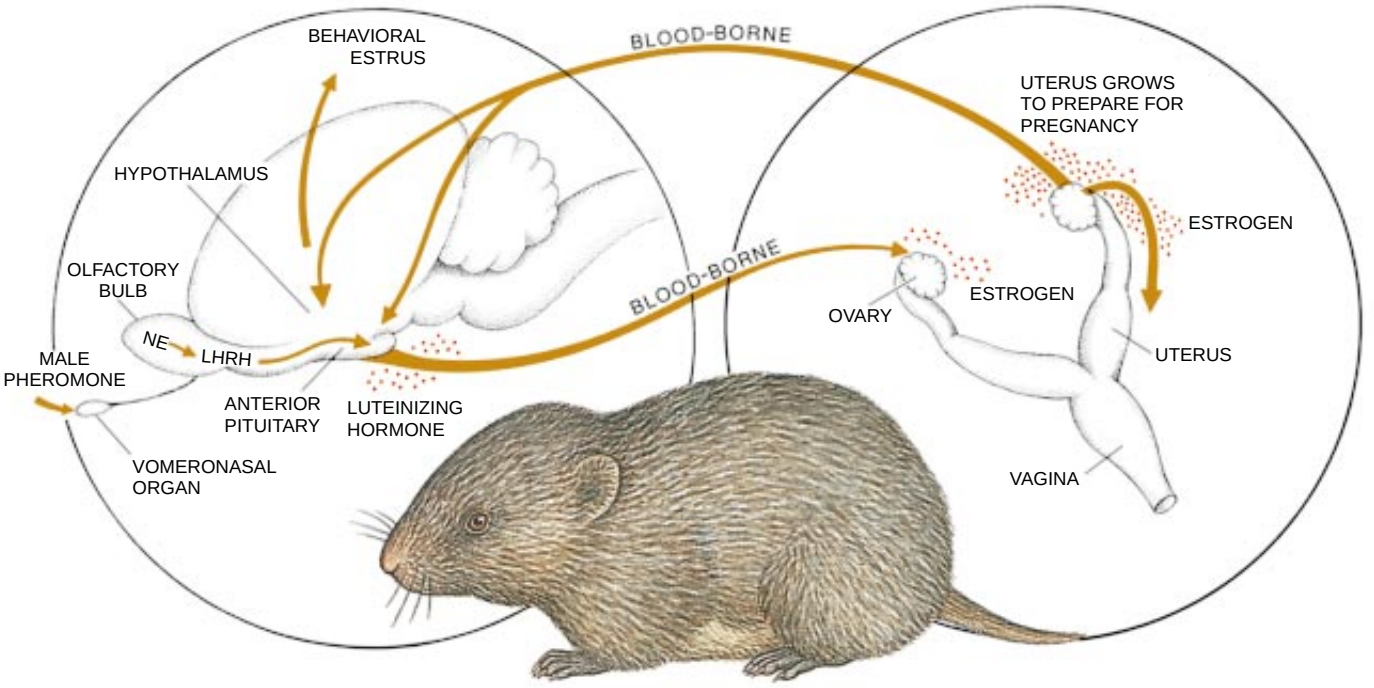


CHOICE TEST given to female prairie voles in the laboratory reveals a social preference for the mated males. Initially, females enter the cages of both strangers and their partners (*represented by brown lines*) and will mate with both. Within about 30 minutes, however, females tend to remain near the familiar male.

The pair bonding in monogamy also leads mated pairs to guard one another or the shared territory. Reproductively naive prairie voles rarely fight, but mated prairie voles can be extremely vicious toward strangers. Because studies have implicated testosterone, a major reproductive hormone manufactured by the testes, in aggression in other animals, we initially hypothesized that testosterone might also be responsible for the postmating hostility in prairie voles. But in our laboratory Nicholas Hastings found that neither castration nor testosterone injections had an effect on male aggression after mating.

If testosterone does not regulate aggression or mate guarding, then what does? Many pieces of evidence suggested that vasopressin, a hormone best known for its role in regulating the human body's water content, might play a role in mate guarding. First, Craig F. Ferris of the University of Massachusetts Medical Center in Worcester and Elliott H. Albers of Georgia State University had implicated vasopressin in territoriality and aggression in hamsters. Second, vasopressin shares a molecular structure similar to that of oxytocin; the molecules differ from one another in only two of their nine amino acids. In addition, both hormones may be released under similar circumstances, such as during sexual behavior and other social actions. The cellular and behavioral functions of vasopressin and oxytocin, however, tend to be antagonistic. Therefore, we reasoned that if oxytocin encourages social contact, perhaps vasopressin causes the antisocial or mate-guarding behavior shown by male prairie voles after they have mated.

Winslow, Hastings and Insel tested this hypothesis in a collaborative study. In one experiment, males were injected before mating with a drug that blocks



HORMONAL CASCADE that triggers estrus in the female prairie vole begins when she sniffs a male. The vomeronasal organ picks up pheromones, stimulating the olfactory bulb. Norepinephrine (NE) and luteinizing hormone–releasing hormone (LHRH) are secreted and start the production of luteinizing hormone. Luteinizing hormone reaches the ovaries via the bloodstream and stimulates them to produce estrogen. Estrogen is then carried to the hypothalamus, where it induces estrus.

vasopressin receptors. The injections eliminated the increase in attacks directed toward strangers that usually follows mating. The effect was not a general inhibition of aggression. The antivasopressin drug did not stem attack behavior when given to males that had completed their mating. In a separate experiment, Winslow and Insel infused vasopressin into a male while a female was present. Such males then displayed increased hostility toward male intruders.

Vasopressin may also play a role in male parental care. Recent findings by Maryam Bamshad, Melinda A. Novak and Geert J. De Vries of the University of Massachusetts show the behavior in male prairie voles correlates with characteristic changes in vasopressin levels.

The experiments on the effects of oxytocin and vasopressin on prairie vole behavior suggest that these two compounds have a much broader behavioral significance than was previously thought. Rather than just being a homeostatic compound, vasopressin may have a more general role as a neuropeptide involved in eliciting parental care and defensive behavior with respect to self and family. Oxytocin, which has a well-established role in reproduction, might block the more primitive, antisocial actions induced by vasopressin, thus permitting social behaviors to emerge. Finally, monogamy may be a refined expression of sociality in which interactions between oxytocin and vasopressin are particularly apparent.

Although these studies have provided strong clues to some of the neuroendocrine mechanisms underlying monogamous behavior, a major puzzle persists. Besides prairie voles, monogamy occurs in such diverse species as wild dogs, tamarins and marmosets. Why, in a physiological sense, should such taxonomically different mammals show the unique features of monogamy?

One solution to this mystery may be found in the adrenal system and its effects on the developing embryo. The adrenal system produces steroids called glucocorticoids. Individuals release these hormones, particularly corticosterone and cortisol, in response to stress. Yet the endocrine systems of adult prairie voles and marmosets secrete unusually copious amounts of glucocorticoids, even when the animals are not under stress. Our work with prairie voles leads us to hypothesize that the interactions between the adrenal and gonadal hormones during early life might account for some of the monogamous patterns that emerge later.

This supposition is based in part on research pioneered by Ingeborg L. Ward of Villanova University. Ward documented the developmental effects of interactions between adrenal and gonadal steroids in rats. Exposure to stress during the perinatal period—the period of sexual differentiation in mammals—influences subsequent reproductive development. For example, male rats that have been stressed early in life tend to show a more feminine pattern of development as adults. Even the genital anatomy is somewhat demasculinized. Apparently, high levels of stress during the perinatal period inhibit the normal secretion or the action of masculinizing hormones called androgens. Craig H. Kinsley, now at the University of Richmond, and Robert S. Bridges of Tufts University demonstrated that perinatal disturbance also increases the probability that male rats will show, as adults, parental care. Thus, in rats stress appears to alter reproductive functions in a direction that is considered normal in monogamous mammals.

We believe that adrenal activity in prairie voles might account in part for their monogamy. Shapiro and Insel found that shortly after birth prairie voles have an unusually reactive adrenal system. Simply removing the mother for a few minutes elevates the levels of glucocorticoids in pups. In contrast, the nonmonogamous montane voles and rats require greater disturbance before adrenal activity increases.

In prairie voles the reactivity of the adrenal glands, during late pregnancy or the early postnatal period, might con-

MONOGAMOUS CHARACTERISTIC		HORMONE
	FEMALE BONDING TO MALE	OXYTOCIN (RELEASED BY MATING OR CONTACT)
	AGGRESSION BETWEEN MALES AFTER MATING	VASOPRESSIN (HIGH)
	SIMILAR SIZE AND APPEARANCE OF MALE AND FEMALE	CORTICOSTERONE (HIGH) TESTOSTERONE (LOW)
	MALE AND FEMALE PARENTAL CARE	CORTICOSTERONE (HIGH) VASOPRESSIN (HIGH)
	SOCIAL REGULATION OF REPRODUCTION	CORTICOSTERONE (HIGH)

FEATURES OF MONOGAMY in the prairie vole correlate with specific hormones and their amount in the body. The high and low levels of hormone are relative to those found in nonmonogamous but closely related species.

tribute to the appearance in the adult of some of the definitive characteristics of monogamy, including reduced sexual dimorphism and increased male parental care. Preliminary experiments conducted in our laboratory by Luci Roberts offer some evidence. In those analyses, postpartum exposure to higher than normal levels of androgen reduced the tendency of adult male prairie voles to care for pups. Research in progress is examining the importance of interactions between the adrenal and gonadal systems in the development of other components of monogamy.

Although we can now identify some of the physiological underpinnings of monogamy in the prairie vole, the ultimate (or evolutionary) cause of this behavior and its adaptive significance remain unclear. It is widely assumed that, from the perspective of the offspring, having two parents is better than having one. Trapping studies in prairie voles, however, have not provided support for this assumption. Based on field data from more than 700 prairie vole families, single mothers are as successful as mother-father pairs in rearing litters to maturity.

That sexual exclusivity is not a dominant feature of monogamy in prairie voles also raises an evolutionary question. Parental care in mammals, and especially in prairie voles, represents a significant investment of time and energy. It is usually assumed that such commitment from the male is a benefit of monogamy because males can increase their own reproductive success by caring for their offspring. Perhaps this strategy represents a probabilistic function. Monogamous males are increasing their fitness in general while accepting the burden of rearing some pups that are not their own.

The theoretical implications of this finding is uncertain. The prairie voles we have studied in Illinois live in an environment that provides abundant supplies of food, water and other essential resources. We believe that monogamy in prairie voles evolved when food was not plentiful. Under such conditions, monogamy might offer additional benefits not evident in the habitats we have examined. We are currently comparing Illinois prairie voles with those that live in a much harsher environment in Kansas.

Our studies also frequently elicit questions concerning the applicability of our findings to human behavior. Monogamy in Old World primates and in humans probably takes on a different form from that described here for voles. Yet there are some parallels. Clearly, human monogamous partners do not always stay sexually exclusive. Animals, including humans, may be more particular in the selection of a social companion than in their choice of a sexual partner.

In addition, our research highlights the general significance of positive social behavior and bonds, which are at least as meaningful to humans as to prairie voles. Disruption of these bonds in humans, such as that which occurs in the loss of a child, parent or spouse, can have disastrous consequences for mental and physical health. Yet we understand little about the behavioral physiology of the formation of such relationships in humans. In fact, the notion that social bonds have a "biology" has not been generally understood.

It is tempting to speculate that oxytocin and vasopressin could also affect human behavior. But the role of these hormones, even in animals, is difficult to document, and many obvious questions remain unexplored. Most human research is limited to correlations between changes in hormonal blood levels and behavior.

The paucity of human data raises concerns about the medical effect of these hormones, which are often administered as treatments. For example, oxytocin is used to induce uterine contractions in childbirth, and vasopressin is prescribed to treat bed-wetting in children. Actions that indirectly affect hormonal levels, such as delivering a child by cesarean section or bottle-feeding, may also affect the amount of oxytocin received by either a mother or her infant. Because oxytocin and vasopressin were at first believed to act outside the brain, on tissues such as the uterus, breast and kidney, little attention has been given to the behavioral consequences of these treatments or actions. We now know that oxytocin and vasopressin are active in the central nervous system. Findings from animal research, such as those described here for prairie voles, should encourage the biomedical community to look more carefully at the potential effects of these powerful hormones on behavior.

FURTHER READING

THE COMPARATIVE PSYCHOLOGY OF MONOGAMY. D. A. Dewsbury in *Nebraska Symposium on Motivation, 1987,* Vol. 35, edited by D. W. Leger, pages 1–50; 1988.

SOCIAL ORGANIZATIONS AND MATING SYSTEM OF THE PRAIRIE VOLE, *MICROTUS OCHROGASTER*. L. L. Getz, B. McGuire, J. Hofmann, T. Pizzuto and B. Frase in *Social Systems and Population Cycles in Voles.* Edited by R. H. Tamarin, R. S. Ostfeld, S. R. Pugh and G. Bujalska. Birkhauser Boston, 1990.

OXYTOCIN AND SOCIAL BONDING. C. S. Carter, J. R. Williams, D. M. Witt and T. R. Insel in *Annals of the New York Academy of Sciences,* Vol. 652, pages 204–211; 1992.

Autism

Autistic individuals suffer from a biological defect. Although they cannot be cured, much can be done to make life more hospitable for them

by Uta Frith

The image often invoked to describe autism is that of a beautiful child imprisoned in a glass shell. For decades, many parents have clung to this view, hoping that one day a means might be found to break the invisible barrier. Cures have been proclaimed, but not one of them has been backed by evidence. The shell remains intact. Perhaps the time has come for the whole image to be shattered. Then at last we might be able to catch a glimpse of what the minds of autistic individuals are truly like.

Psychological and physiological research has shown that autistic people are not living in rich inner worlds but instead are victims of a biological defect that makes their minds very different from those of normal individuals. Happily, however, autistic people are not beyond the reach of emotional contact.

Thus, we can make the world more hospitable for autistic individuals just as we can, say, for the blind. To do so, we need to understand what autism is like—a most challenging task. We can imagine being blind, but autism seems unfathomable. For centuries, we have known that blindness is often a peripheral defect at the sensory-motor level of the nervous system, but only recently has autism been appreciated as a central defect at the highest level of cognitive processing. Autism, like blindness, persists throughout life, and it responds to special efforts in compensatory education. It can give rise to triumphant feats of coping but can also lead to disastrous secondary consequences—anxiety, panic and depression. Much can be done to prevent problems. Understanding the nature of the handicap must be the first step in any such effort.

Autism existed long before it was described and named by Leo Kanner of the Johns Hopkins Children's Psychiatric Clinic. Kanner published his landmark paper in 1943 after he had observed 11 children who seemed to him to form a recognizable group. All had in common four traits: a preference for aloneness, an insistence on sameness, a liking for elaborate routines and some abilities that seemed remarkable compared with the deficits.

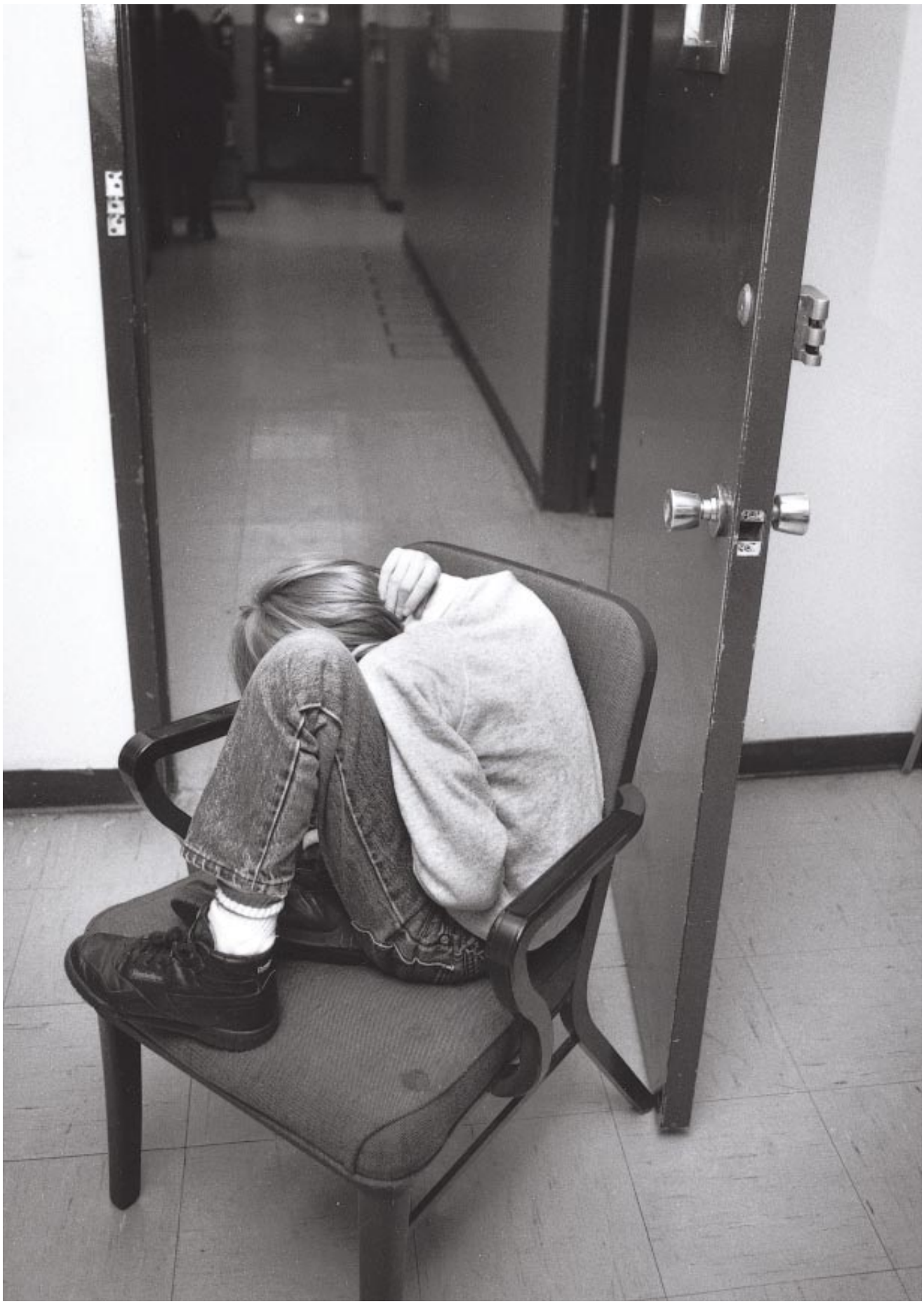
Concurrently, though quite independently, Hans Asperger of the University Pediatric Clinic in Vienna prepared his doctoral thesis on the same type of child. He also used the term "autism" to refer to the core features of the disorder. Both men borrowed the label from adult psychiatry, where it had been used to refer to the progressive loss of contact with the outside world experienced by schizophrenics. Autistic children seemed to suffer such a lack of contact with the world around them from a very early age.

Kanner's first case, Donald, has long served as a prototype for diagnosis. It had been evident early in life that the boy was different from other children. At two years of age, he could hum and sing tunes accurately from memory. Soon he learned to count to 100 and to recite both the alphabet and the 25 questions and answers of the Presbyterian catechism. Yet he had a mania for making toys and other objects spin. Instead of playing like other toddlers, he arranged beads and other things in groups of different colors or threw them on the floor, delighting in the sounds they made. Words for him had a literal, inflexible meaning.

Donald was first seen by Kanner at age five. Kanner observed that the boy paid no attention to people around him. When someone interfered with his solitary activities, he was never angry with the interfering person but impatiently removed the hand that was in his way. His mother was the only person with whom he had any significant contact, and that seemed attributable mainly to the great effort she made to share activities with him. By the time Donald was about eight years old, his conversation consisted largely of repetitive questions. His relation to people remained limited to his immediate wants and needs, and his attempts at contact stopped as soon as he was told or given what he had asked for.

CHARACTERISTIC ALONENESS of autistic children has evoked the image of a child in a glass shell (*above*). The drawing suggests, falsely, that a normal child would emerge if the shell could be broken. Aloneness is also exhibited by a boy (*right*) at the Association in Manhattan for Autistic Children, Inc., the scene of all the accompanying photographs except the last one.

UTA FRITH is a senior scientist in the Cognitive Development Unit of the Medical Research Council in London. Born in Germany, she took a degree in psychology in 1964 at the University of the Saarland in Saarbrücken, where she also studied the history of art. Four years later she obtained her Ph.D. in psychology at the University of London. Besides autism, her interests include reading development and dyslexia. She has edited a book in the field of reading development, *Cognitive Processes in Spelling,* and is the author of *Autism: Explaining the Enigma.*



Autistic Behavior

The traits most characteristic of autistic people are aloneness, an insistence on sameness and a liking for elaborate routines. At the same time, some autistic individuals can perform complicated tasks, provided that the activity does not require them to judge what some other person might be thinking. These traits lead to characteristic forms of behavior, a number of which are portrayed here.

Displays indifference

Indicates needs by using an adult's hand

Parrots words

Laughs and giggles inappropriately

Does not make eye contact

Does not pretend in playing

Joins in only if an adult insists and assists

Does not play with other children

Prefers sameness

Is one-sided in interactions

Talks incessantly about one topic

Behaves in bizarre ways

Handles or spins objects

Yet some do certain things well if the task does not involve social understanding.

Some of the other children Kanner described were mute, and he found that even those who spoke did not really communicate but used language in a very odd way. For example, Paul, who was five, would parrot speech verbatim. He would say "You want candy" when he meant "I want candy." He was in the habit of repeating, almost every day, "Don't throw the dog off the balcony," an utterance his mother traced to an earlier incident with a toy dog.

Twenty years after he had first seen them, Kanner reassessed the members of his original group of children. Some of them seemed to have adapted socially much better than others, although their failure to communicate and to form relationships remained, as did their pedantry and single-mindedness. Two prerequisites for better adjustment, though no guarantees of it, were the presence of speech before age five and relatively high intellectual ability. The brightest autistic individuals had, in their teens, become uneasily aware of their peculiarities and had made conscious efforts to conform. Nevertheless, even the best adapted were rarely able to be self-reliant or to form friendships. The one circumstance that seemed to be helpful in all the cases was an extremely structured environment.

As soon as the work of the pioneers became known, every major clinic began to identify autistic children. It was

found that such children, in addition to their social impairments, have substantial intellectual handicaps. Although many of them perform relatively well on certain tests, such as copying mosaic patterns with blocks, even the most able tend to do badly on test questions that can be answered only by the application of common sense.

Autism is rare. According to the strict criteria applied by Kanner, it appears in four of every 10,000 births. With the somewhat wider criteria used in current diagnostic practice, the incidence is much higher: one or two in 1,000 births, about the same as Down's syndrome. Two to four times as many boys as girls are affected.

For many years, autism was thought to be a purely psychological disorder without an organic basis. At first, no obvious neurological problems were found. The autistic children did not necessarily have low intellectual ability, and they often looked physically normal. For these reasons, psychogenic theories were proposed and taken seriously for many years. They focused on the idea that a child could become autistic because of some existentially threatening experience. A lack of maternal bonding or a disastrous experience of rejection, so the theory went, might drive an infant to withdraw into an inner world of fantasy that the outside world never penetrates.

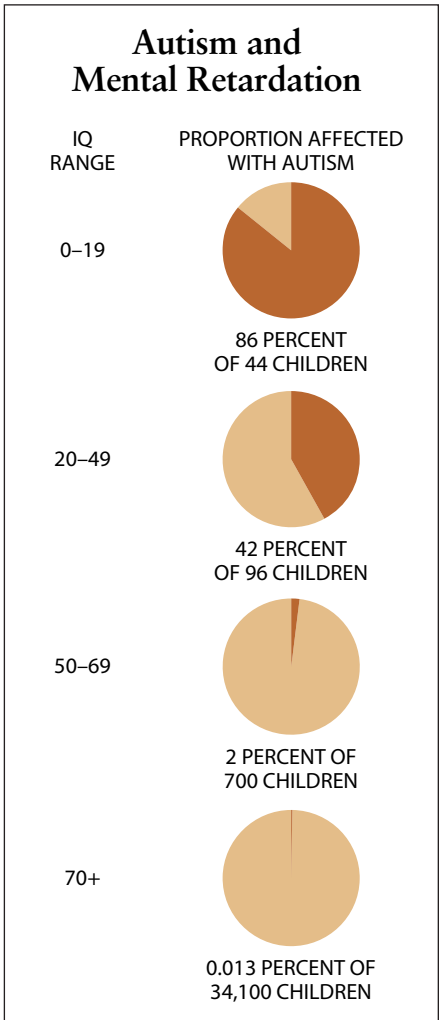
These theories are unsupported by any empirical evidence. They are unlikely to be supported because there are many instances of extreme rejection and deprivation in childhood, none of which have resulted in autism. Unfortunately, therapies vaguely based on such notions are still putting pressure on parents to accept a burden of guilt for the supposedly avoidable and reversible breakdown of interpersonal interactions. In contrast, well-structured behavior modification programs have often helped families in the management of autistic children, especially children with severe behavior problems. Such programs do not claim to reinstate normal development.

The insupportability of the psychogenic explanation of autism led a number of workers to search for a biological cause. Their efforts implicate a defective structure in the brain, but that structure has not yet been identified. The defect is believed to affect the thinking of autistic people, making them unable to evaluate their own thoughts or to perceive clearly what might be going on in someone else's mind.

Autism appears to be closely associated with several other clinical and medical conditions. They include maternal rubella and chromosomal abnormality, as well as early injury to the brain and infantile seizures. Most impressive, perhaps, are studies showing that autism can have a genetic basis. Both identical twins are much more likely to be autistic than are both fraternal twins. Moreover, the likelihood that autism will occur twice in the same family is 50 to 100 times greater than would be expected by chance alone.

Structural abnormalities in the brains of autistic individuals have turned up in anatomic studies and brain-imaging procedures. Both epidemiological and neuropsychological studies have demonstrated that autism is strongly correlated with mental retardation, which is itself clearly linked to physiological abnormality. This fact fits well with the idea that autism results from a distinct brain abnormality that is often part of more extensive damage. If the abnormality is pervasive, the mental retardation will be more severe, and the likelihood of damage to the critical brain system will increase. Conversely, it is possible for the critical system alone to be damaged. In such cases, autism is not accompanied by mental retardation.

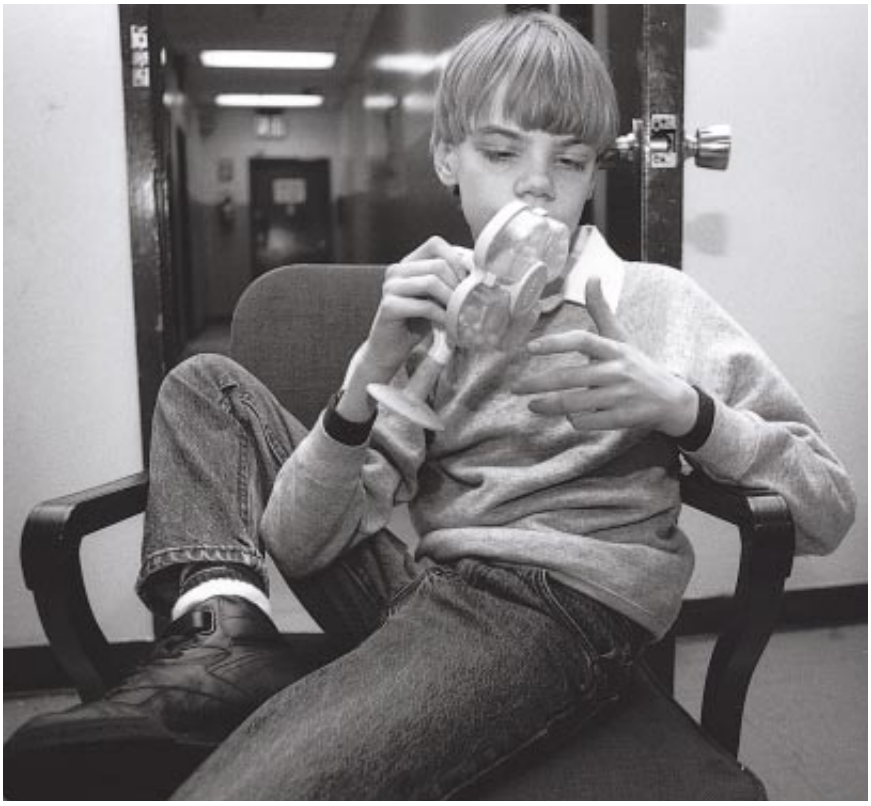
Neuropsychological testing has also contributed evidence for the existence of a fairly circumscribed brain abnormality. Autistic individuals who are otherwise able show specific and extensive deficits on certain tests that involve planning, initiative and spontaneous generation of new ideas. The same deficits appear in patients who have frontal lobe lesions. Therefore, it seems plausible that whatever the defective brain structure is, the frontal lobes are implicated.



CLOSE LINK between autism and mental retardation is reflected in this chart. The percentage of children showing the social impairments typical of autism is highest at low levels of intelligence as measured by tests in which an intelligence quotient (IQ) below 70 is subnormal. For example, 86 percent of 44 children in the lowest IQ range showed the social impairments of autism. The data are drawn from a population of about 35,000 children aged under 15 years.

Population studies carried out by Lorna Wing and her colleagues at the Medical Research Council's Social Psychiatry Unit in London reveal that the different symptoms of autism do not occur together simply by coincidence. Three core features in particular—impairments in communication, imagination and socialization—form a distinct triad. The impairment in communication includes such diverse phenomena as muteness and delay in learning to talk, as well as problems in comprehending or using nonverbal body language. Other autistic individuals speak fluently but are overliteral in their understanding of language. The impairment in imagination appears in young autistic children as repetitive play with objects and in some autistic adults as an obsessive interest in facts. The impairment in socialization includes ineptness and inappropriate behavior in a wide range of reciprocal social interactions, such as the ability to make and keep friends. Nevertheless, many autistic individuals prefer to have company and are eager to please.

The question is why these impairments, and only these, occur together. The challenge to psychological theorists was clear: to search for a single cognitive component that would explain the deficits yet still allow for the abilities that autistic people display in certain aspects of interpersonal interactions. My colleagues at the Medical Research Council's Cognitive Development Unit in London and I think we



UNUSUAL BEHAVIOR is often displayed by autistic individuals. Autistic children, for example, tend to fixate on making toys and other objects spin (*above*) and to play repetitively. Another trait is to ignore the efforts of others to communicate or to draw the autistic person into an activity (*below*).

have identified just such a component. It is a cognitive mechanism of a highly complex and abstract nature that could be described in computational terms. As a shorthand, one can refer to this component by one of its main functions, namely the ability to think about thoughts or to imagine another individual's state of mind. We propose that this component is damaged in autism. Furthermore, we suggest that this mental component is innate and has a unique brain substrate. If it were possible to pinpoint that substrate—whether it is an anatomic structure, a physiological system or a chemical pathway—one might be able to identify the biological origin of autism.

The power of this component in normal development becomes obvious very early. From the end of the first year onward, infants begin to participate in what has been called shared attention. For example, a normal child will point to something for no reason other than to share his interest in it with someone else. Autistic children do not show shared attention. Indeed, the absence of this behavior may well be one of the earliest signs of autism. When an autistic child points at an object, it is only because he wants it.

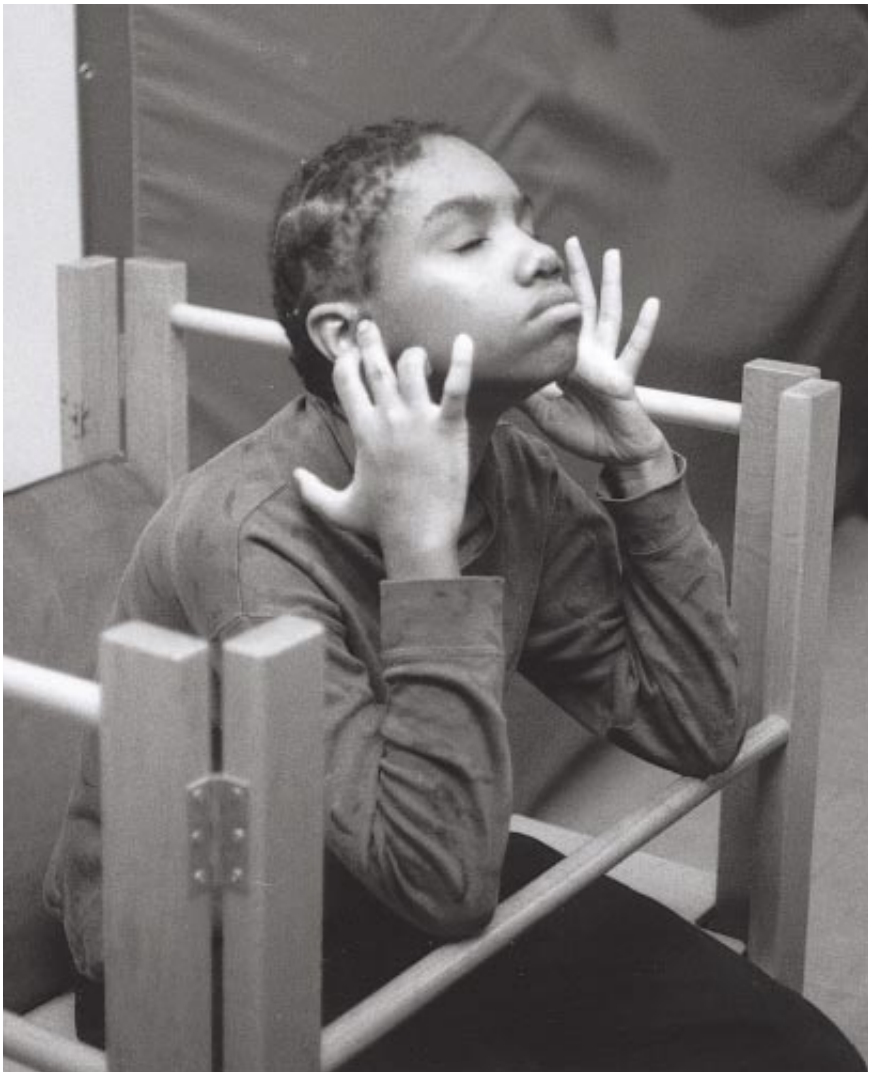
In the second year of life, a particularly dramatic manifestation of the critical component can be seen in normal children: the emergence of pretense, or the ability to engage in fantasy and pretend play. Autistic children cannot understand pretense and do not pretend when they are playing. The difference can be seen in such a typical nursery game as "feeding" a teddy bear or a doll with an empty spoon. The normal child goes through the appropriate motions of feeding and accompanies the action with appropriate slurping noises. The autistic child merely twiddles or flicks the spoon repetitively. It is precisely the absence of early and simple communicative behaviors, such as shared attention and make-believe play, that often creates the first nagging doubts in the minds of the parents about the development of their child. They rightly feel that they cannot engage the child in the emotional to-and-fro of ordinary life.

My colleague Alan M. Leslie devised a theoretical model of the cognitive mechanisms underlying the key abilities of shared attention and pretense. He postulates an innate mechanism whose function is to form and use what we might call second-order representations. The world around us consists not only of visible bodies and events, captured by

first-order representations, but also of invisible minds and mental events, which require second-order representation. Both types of representation have to be kept in mind and kept separate from each other.

Second-order representations serve to make sense of otherwise contradictory or incongruous information. Suppose a normal child, Beth, sees her mother holding a banana in such a way as to be pretending that it is a telephone. Beth has in mind facts about bananas and facts about telephones—first-order representations. Nevertheless, Beth is not the least bit confused and will not start eating telephones or talking to bananas. Confusion is avoided because Beth computes from the concept of pretending (a second-order representation) that her mother is engaging simultaneously in an imaginary activity and a real one.

As Leslie describes the mental process, pretending should be understood as computing a three-term relation between an actual situation, an imaginary situation and an agent who does the pretending. The imaginary situation is then not treated as the real situation. Believing can be understood in the same way as pretending. This insight enabled us to predict that autistic children, despite an adequate mental age (above four years or so), would not be able to understand that someone can have a mistaken belief about the world.



SELF-ABSORPTION displayed by this autistic girl is a common feature of the disorder. In the motion picture *Rain Man*, self-absorption was the key trait of the central character, an autistic adult, portrayed by actor Dustin Hoffman.

Together with our colleague Simon Baron-Cohen, we tested this prediction by adapting an experiment originally devised by two Austrian developmental psychologists, Heinz Wimmer and Josef Perner. The test has become known as the Sally-Anne task. Sally and Anne are playing together. Sally has a marble that she puts in a basket before leaving the room. While she is out, Anne moves the marble to a box. When Sally returns, wanting to retrieve the marble, she of course looks in the basket. If this scenario is presented as, say, a puppet show to normal children who are four years of age or more, they understand that Sally will look in the basket even though they know the marble is not there. In other words, they can represent Sally's erroneous belief as well as the true state of things. Yet in our test, 16 of 20 autistic children with a mean mental age of nine failed the task—answering that Sally would look in the box—in spite of being able to answer correctly a variety of other questions relating to the facts of the episode. They could not conceptualize the possibility that Sally believed something that was not true.

Many comparable experiments have been carried out in other laboratories, which have largely confirmed our prediction: autistic children are specifically impaired in their understanding of mental states. They appear to lack the innate component underlying this ability. This component, when it works normally, has the most far-reaching consequences for higher-order conscious processes. It underpins the special feature of the human mind, the ability to reflect on itself. Thus, the triad of impairments in autism—in communication, imagination and socialization—is explained by the failure of a single cognitive mechanism. In everyday life, even very able autistic individuals find it hard to keep in mind simultaneously a reality and the fact that someone else may hold a misconception of that reality.

The automatic ability of normal people to judge mental states enables us to be, in a sense, mind readers. With sufficient experience we can form and use a theory of mind that allows us to speculate about psychological motives for our behavior and to manipulate other people's opinions, beliefs and attitudes. Autistic individuals lack the automatic ability to represent beliefs, and therefore they also lack a theory of mind. They cannot understand how behavior is caused by mental states or how beliefs and attitudes can be manipulated. Hence, they find it difficult to understand deception. The psychological undercurrents of real life as well as of literature—in short, all that gives spice to social relations—for them remain a closed book. "People talk to each other with their eyes," said one observant autistic youth. "What is it that they are saying?"

Lacking a mechanism for a theory of mind, autistic children develop quite

SOCIAL CONTACT is sometimes achieved by autistic people, particularly in familiar and well-structured situations. The autistic teenage boy in this photograph enjoys music and is responding to it in a family setting as his father plays the guitar.

differently from normal ones. Most children acquire more and more sophisticated social and communicative skills as they develop other cognitive abilities. For example, children learn to be aware that there are faked and genuine expressions of feeling. Similarly, they become adept at that essential aspect of human communication, reading between the lines. They learn how to produce and understand humor and irony. In sum, our ability to engage in imaginative ideas, to interpret feelings and to understand intentions beyond the literal content of speech are all accomplishments that depend ultimately on an innate cognitive mechanism. Autistic children find it difficult or impossible to achieve any of these things. We believe this is because the mechanism is faulty.

This cognitive explanation of autism is specific. As a result, it enables us to distinguish the types of situations in which the autistic person will and will not have problems. It does not preclude the existence of special assets and abilities that are independent of the innate mechanism my colleagues and I see as defective. Thus it is that autistic individuals can achieve social skills that do not involve an exchange between two minds. They can learn many useful social routines, even to the extent of sometimes camouflaging their problems. The cognitive deficit we hypothesize is also specific enough not to preclude high achievement by autistic people in such diverse activities as musical performance, artistic drawing, mathematics and memorization of facts.

It remains to be seen how best to explain the coexistence of excellent and abysmal performance by autistic people on abilities that are normally expected to go together. It is still uncertain whether there may be additional damage in emotions that prevents some autistic children from being interested in social stimuli. We have as yet little idea what to make of the single-minded, often obsessive, pursuit of certain activities. With the autistic person, it is as if a powerful integrating force—the effort to seek meaning—were missing.

The old image of the child in the glass shell is misleading in more ways than one. It is incorrect to think that inside the glass shell is a normal individual waiting to emerge, nor is it true that autism is a disorder of childhood only. The motion picture *Rain Man* came at the right time to suggest a new image to a receptive public. Here we see Raymond, a middle-aged man who is unworldly, egocentric in the extreme and all too amenable to manipulation by others. He is incapable of understanding his brother's double-dealing pursuits, transparently obvious though they are to the cinema audience. Through various experiences it becomes possible for the brother to learn from Raymond and to forge an emotional bond with him. This is not a farfetched story. We can learn a great deal about ourselves through the phenomenon of autism.

Yet the illness should not be romanticized. We must see autism as a devastating handicap without a cure. The autistic child has a mind that is unlikely to develop self-consciousness. But we can now begin to identify the particular types of social behavior and emotional responsiveness of which autistic individuals are capable. Autistic people can learn to express their needs and to anticipate the behavior of others when it is regulated by external, observable factors rather than by mental states. They can form emotional attachments to others. They often strive to please and earnestly wish to be instructed in the rules of person-to-person contact. There is no doubt that within the stark limitations a degree of satisfying sociability can be achieved.

Autistic aloneness does not have to mean loneliness. The chilling aloofness experienced by many parents is not a permanent feature of their growing autistic child. In fact, it often gives way to a preference for company. Just as it is possible to engineer the environment toward a blind person's needs or toward people with other special needs, so the environment can be adapted to an autistic person's needs.

On the other hand, one must be realistic about the degree of adaptation that can be made by the limited person. We can hope for some measure of compensation and a modest ability to cope with adversity. We cannot expect autistic individuals to grow out of the unreflecting mind they did not choose to be born with. Autistic people in turn can look for us to be more sympathetic to their plight as we better understand how their minds are different from our own.

FURTHER READING

AUTISM: EXPLAINING THE ENIGMA. Uta Frith. Blackwell Publishers, 1989.

THE COGNITIVE BASIS OF A BIOLOGICAL DISORDER: AUTISM. Uta Frith, John Morton and Alan M. Leslie in *Trends in Neurosciences,* Vol. 14, No. 10, pages 433-438; October 1991.

AUTISM AND ASPERGER SYNDROME. Edited by Uta Frith. Cambridge University Press, 1992.

UNDERSTANDING OTHER MINDS: PERSPECTIVES FROM AUTISM. Edited by Simon Baron-Cohen, Helen Tager-Flusberg and Donald J. Cohen. Oxford University Press, 1993.

The Great Well of China

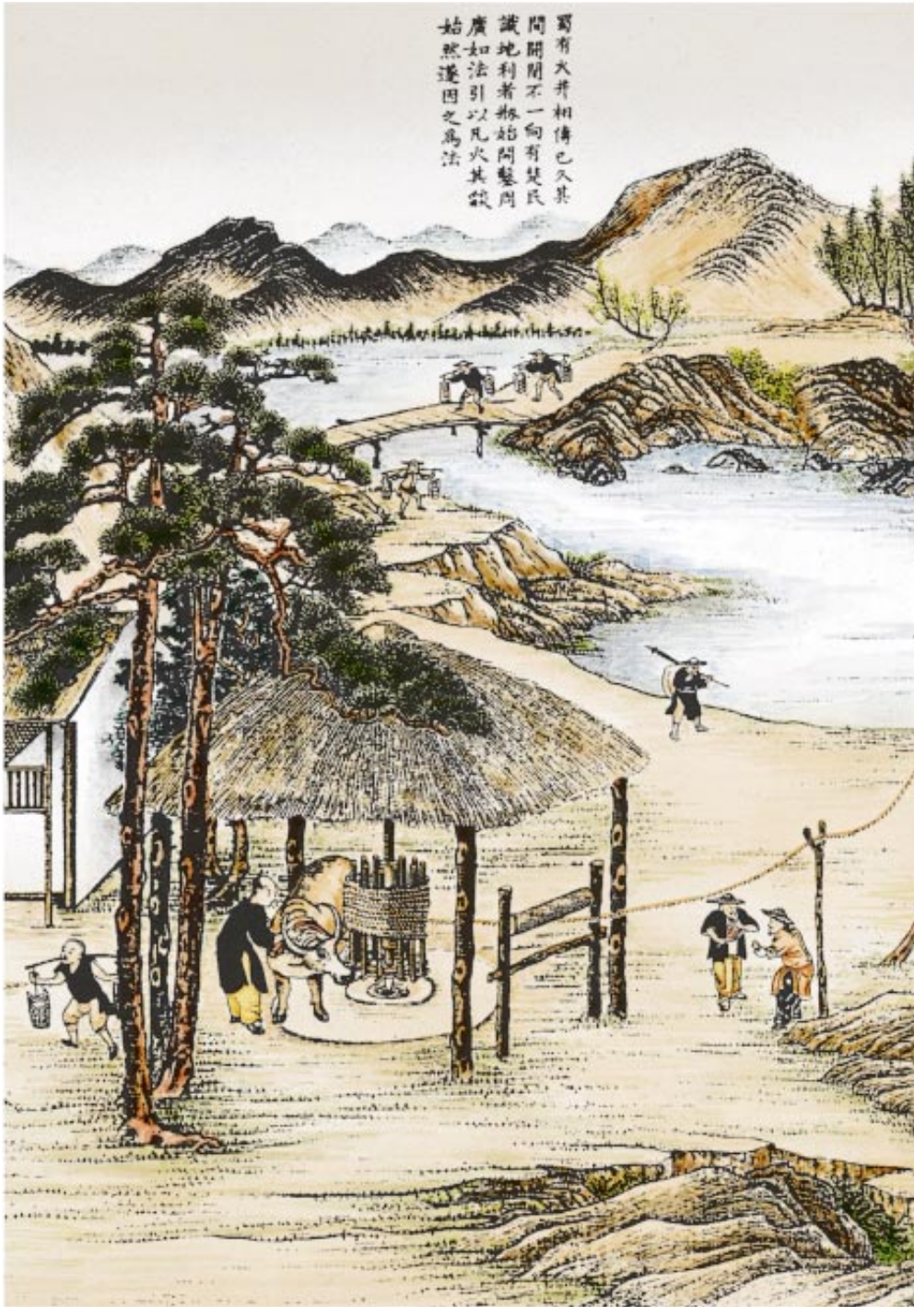
More than 150 years ago the Chinese drilled one kilometer into the earth to extract brine for making salt. The well was the culmination of an 800-year-old technology

by Hans Ulrich Vogel

The landlocked Chinese province of Sichuan harbors one of the technological wonders of world: the Xinhai well. It is remarkable not just for its age—158 years—and its depth—1,001 meters; the Xinhai well represents the crowning achievement of an industry that invented deep drilling four centuries before Europeans developed the technology.

Almost 1,000 years ago the Sichuanese made boreholes 100 meters deep to reach sources of brine. Elaborate systems were built for hoisting brine from the ground and refining it into precious salt. The Sichuanese even struck natural gas and found its first commercial use: fueling fires to evaporate the brine. Today deep drilling is praised as one of ancient China's greatest innovations—comparable to their invention of paper, printing, gunpowder and compasses. Indeed, the Chinese salt industry of the 19th century is, in some ways, the precursor of modern petroleum production.

In 1983, when I was first inspired to study salt production in China, I found little reliable information about the de-



WELLS DRILLED 100 meters or more into the earth yielded brine and natural gas for the ancient Chinese in the province of Sichuan. Brine, which was boiled down to make salt, was vital to the Sichuanese economy for more than 2,000 years. The illustration is an enhanced reproduction of a scroll painted in the 1750s. To lift brine from the well, workers used a bamboo tube suspended from a rope. The brine was heated over a fire fueled by natural gas.

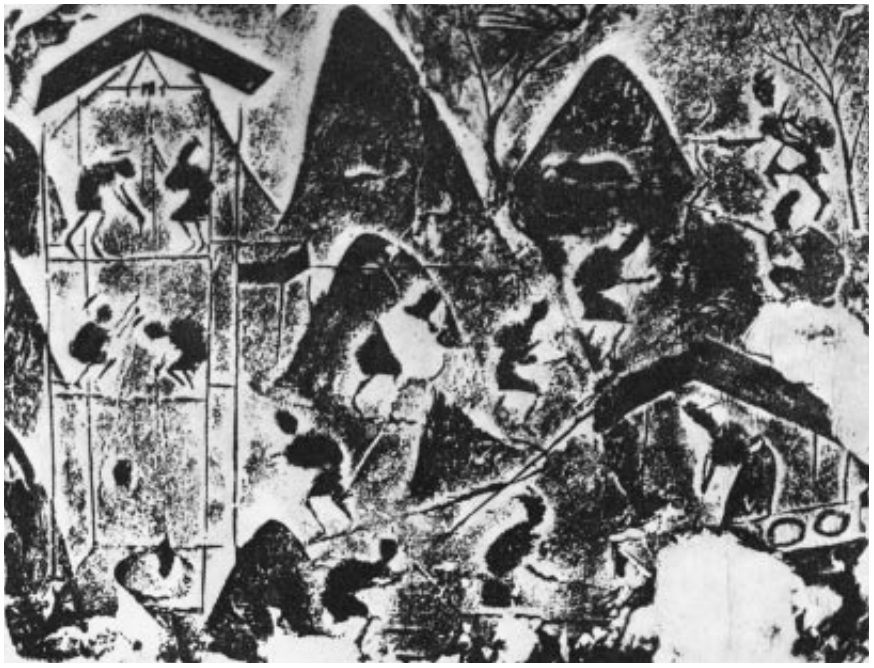
HANS ULRICH VOGEL is associate professor of Chinese studies at the University of Heidelberg. After earning his Ph.D. in Chinese history in 1983 from the University of Zurich, he began research on the history of the salt industry in Sichuan Province.

velopment of deep drilling. Most of the available sources were 20th-century descriptions of borehole wells, and some were based on Sichuanese lore. In recent years, Chinese scholars have uncovered a large number of writings and artifacts that provide clues about the production and trade of salt in China. As a result, publications dedicated to the subject have sprung up all over China. My colleagues and I can now challenge some of the tenacious legends and provide an accurate history of the spectacular Sichuan salt wells.

The salt industry was pivotal in the economic and social history of China. More than 2,100 years ago the rulers of China first endeavored to monopolize salt production and trade throughout the empire. By the 14th century the Yuan dynasty raised as much as 80 percent of its revenue through taxes on the salt industry; the Ming, who ruled China from 1368 to 1644, relied on salt taxes for approximately 40 percent of their income, and even the Qing Empire, which endured until 1911, depended on salt for a tenth of its wealth.

The major source of salt for China was brine from the sea, not from wells. Production of sea salt contributed about 80 percent of the total output for most of the past millennium. Yet wells were essential to the economic welfare of the Sichuanese, because for these inland Chinese, making well salt was cheaper than importing sea salt.

Before deep drilling was invented, the Chinese dug shallow wells to tap sources of brine. A relief image carved on a tombstone from the second century [*see top illustration on next page*] is definitive evidence that the Chinese had developed techniques

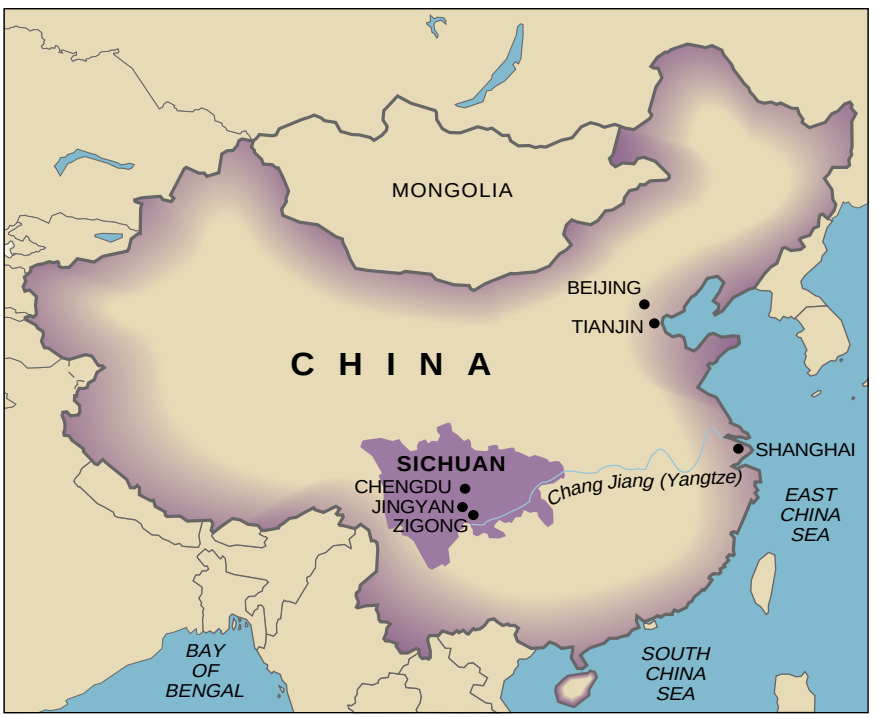


SALT PRODUCTION in second-century China is depicted in this rubbing of a tombstone. Workers hoisted brine from a shaft well (*left*) using a rope, animal-skin bags, a pulley and a scaffold. The brine was boiled over a hearth (*right*) to produce salt.

for hoisting brine from shaft wells. The image shows laborers hoisting brine from a hole in the ground.

Unfortunately, many of the details about how shaft wells were used in ancient times are not known. For a good description, historians must turn to the writings of later centuries. In 1177, for example, the supervisor of Sichuan, Hu Yuanzhi, gave a particularly vivid account. According to this official, workers began their search for brine by inspecting the terrain of mountain valleys and then digging exploratory holes to a depth of between 60 and 80 zhang (one zhang equals about 3.6 meters). If they were lucky enough to find a salty spring, they made an effort to reinforce the sides of the hole by lining it with stones. After the well was constructed, workers lowered buffalo-hide bags attached to a rope to the bottom, where the bags filled with brine. If 40 or more men labored continuously to pull the bags from the well, they could drain most wells almost completely in about 12 hours. In the final step of the process, the brine was poured into flat pans and boiled over a fire of wood and grasses until only the salt remained.

Shaft wells were often very profitable, but after the development of deep drilling and the emergence of borehole wells, such wells slowly lost their economic importance. Historical sources suggest that deep drilling was invented during the 11th century in the district of Jingyan, which is about 130 kilometers south of the Sichuan capital of Chengdu. One of the first to mention the use of deep drilling was Wen Tong—a prefect who oversaw the Jingyan region. Around 1071 he wrote: "[Two decades ago] the local populace began constructing wells by drilling into the ground and installing bamboo pipes. These borehole wells are used to extract [brine] which is then boiled and refined into salt. [The people of Jingyan] have become wholly skilled in this method, and the number of practitioners has increased tremendously."



JINGYAN in Sichuan Province is the site at which deep drilling was invented 900 years ago. In the 18th century the center of the industry shifted to Zigong.

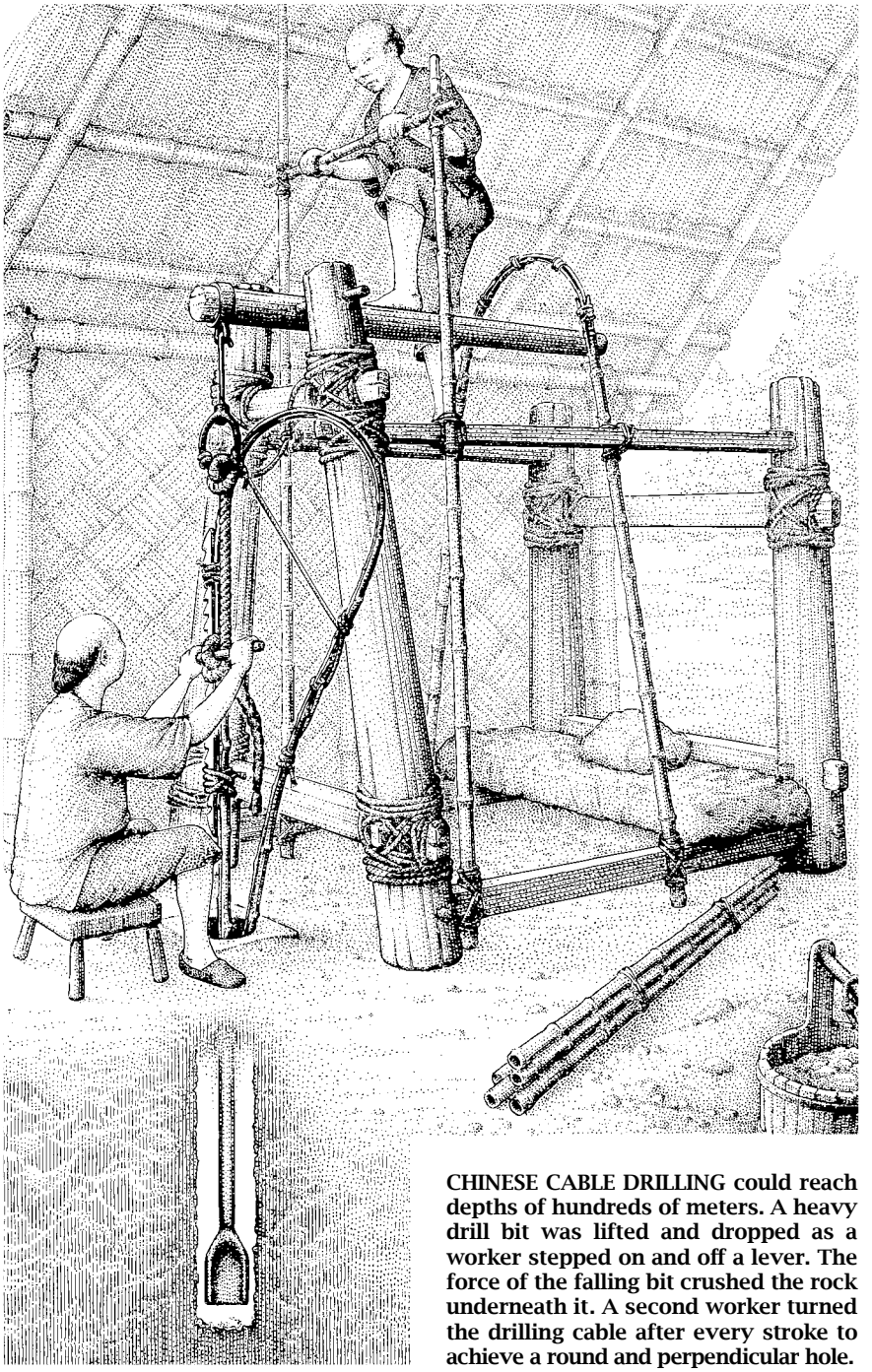
The wells were owned by several powerful families in Jingyan and were operated by migrant laborers. By making white salt cheaply, the salters of Jingyan could compete with the state-run shaft wells, which produced expensive, adulterated salt. Wen Tong seems to have been worried that this new, unregulated industry might undermine the state monopoly. He also expressed concern about establishing law and order among the growing population of migrant workers. It is perhaps for these reasons that he reported that Jingyan was "an important and complicated official post sorely in need of proper government." During Wen Tong's lifetime, the state felt it necessary to intervene in the entrepreneurial activities of the Sichuanese, particularly in the tea and salt industries. For a decade or so, the state prohibited the construction of borehole wells in most parts of Sichuan. But when the advocates of the interventionist policies lost power in 1089, the emperor's administration relaxed the ban.

Within a few decades the Sichuanese began drilling for brine all across the province. In 1132 officials registered more than 4,900 brine wells in Sichuan,

most of them borehole wells. This represented a tremendous growth in the entire salt industry; from the 10th to the 12th century, the number of brine wells had increased eightfold. The reason the industry grew so rapidly is clear. As the 11th-century scholar Su Shi writes, "People never miss an opportunity to make a profit."

The ancient art of constructing borehole wells is known to historians through the writings of a few scholars and officials, notably Su Shi and Hu Yuanzhi. Su Shi gave some technical information about the methods used during the middle of the 11th century. After the hole was drilled, workers hollowed out thick bamboo stalks, connected them end to end and inserted them into the hole. The bamboo tubes served as the walls of the well and kept fresh groundwater from contaminating the brine at the bottom of the well. By Su Shi's account, the wells were as wide as a drinking bowl and could reach depths of 120 meters or more. To hoist brine out of the well, the laborers used a hollow bamboo stalk with a valve installed at one end. It is unclear from Su Shi's description, however, how the valve worked and how brine was captured in the bamboo. He does state unambiguously that the amount of brine brought up by each tube was several dou (one dou equals about 6.6 liters).

In 1177 Hu Yuanzhi described some of the difficulties encountered in the operation of the wells. The bamboo tubes lining the well wore down and sometimes would become blocked by stones and earth. Workers employed a variety of instruments to remove the obstructions, but sometimes they spent weeks or months trying to repair wells without success.



CHINESE CABLE DRILLING could reach depths of hundreds of meters. A heavy drill bit was lifted and dropped as a worker stepped on and off a lever. The force of the falling bit crushed the rock underneath it. A second worker turned the drilling cable after every stroke to achieve a round and perpendicular hole.

Historical accounts from the 11th and the 12th centuries do not give much detail about deep-drilling techniques. Indeed, it is very difficult to figure out from these incomplete descriptions how wells were constructed and how brine was hoisted. To obtain a better understanding of the techniques developed in northern Sichuan during the 11th century, I examined documents from later periods, searched for archaeological evidence and conducted fieldwork. Using this information, I have attempted to reconstruct the early procedures for deep drilling, hoisting and repair.

To select a suitable drilling site, the Sichuanese could do little more than inspect topological features. Naturally, they often chose sites close to successful wells. The drilling apparatus consisted of an iron drill, weighing about 100 kilograms. The drill was suspended from a bamboo cable that was attached to the short end of a lever a few meters in length. A laborer would lift the drill by standing on the long end of the lever and would then jump off, causing the drill to drop and smash into the ground. After digging a wide hole to a depth of about 50 meters, workers typically used a smaller drill to produce a deeper, narrow hole.

To remove excess dirt and water from the hole, the Sichuanese used bamboo in an innovative way. A bamboo stalk consists of several sections, each of which is hollow in the middle and solid at the ends. If an opening is carved into the top of each section, a series of small buckets is created. Workers lowered such stalks to the bottom of the hole and allowed water and earth to flow into the small buckets. They then removed the stalk from the hole and poured out the material. If the earth was dry, they would pour water into the hole until they created a slurry. Bamboo pipes were installed in the upper wide section of the well but not in the lower narrow part. The pipes, which were 10 to 13 centimeters in diame-

ter, were connected together and were made watertight.

Hoisting brine out of the well required a bamboo tube with a one-way leather valve at the lower end. The top of the tube was tied to a bamboo cable, which in turn was wound around a large drum. To start, the tube was lowered to the bottom of the well. The leather valve at the end of the tube then opened as the apparatus sunk into brine and would close if the tube was lifted upward. By moving the tube up and down rapidly within the brine, an operator could force the liquid up inside the hollow interior and fill it. The full tube was lifted by winding the cable around the drum. (In later periods a similar tube was used during the drilling process to remove slurry.) After the tube was lifted out of the well, it was held in the vertical position by one of a variety of structures. The simplest of these was a ring tied to the top of a pole; the upper end of the tube was inserted through the ring, which clamped down on it. A worker placed a container under the bottom of the tube and then opened the valve with a hook, allowing the brine to empty out.

The design of Sichuan borehole wells remained essentially the same for nearly 500 years, possibly because the government's strict control over the salt industry stifled innovation. During the late 16th century, however, the Sichuanese came up with some innovations that allowed them to drill as deep as 300 meters.

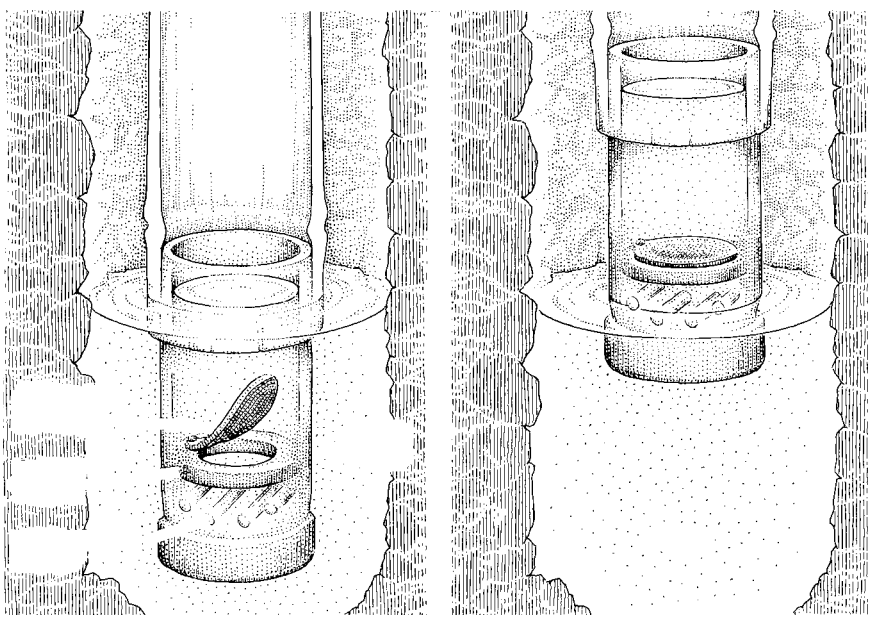
The most detailed depiction of that period is *An Illustrated Account of Well Salt Production (Yanjing tushuo),* created by Ma Ji in the 1580s. The text describes the salt wells of the Shehong District in Northern Sichuan. Unfortunately, the illustrations have been lost. From Ma Ji and other sources, historians have learned about four major developments in borehole well technology from the 16th century onward. First, a stone-lined hole in the top soil was constructed before the actual drilling process was begun; the structure probably prevented the soft earth from clogging the borehole during drilling. Second, in some cases, the workers hitched buffaloes to the hoisting drum to power the apparatus. Third, the Sichuanese discovered ways to utilize natural gas that was escaping from the brine wells. They developed remarkable types of gas storage equipment and built gas stoves for the purpose of evaporating brine. Natural gas was not used widely, however, until much later.

Fourth and most interesting, the Sichuanese developed a device known as a "jar," which was first used in the 16th century to help fish out drills that got stuck in boreholes. The top of the jar was connected to a heavy rod that was itself attached to the end of the drilling cable; the bottom of the jar was coupled to a fishing tool. The jar allowed some play between the rod and the fishing tool so that one could move upward or downward independent of the motion of the other. To remove a lost drill bit, laborers lowered the fishing tool until it came to rest on top of the drill. Then the drilling cable was pulled up just enough so that the heavy rod was lifted, but not the fishing tool. Next the drilling cable was released so that the heavy rod smashed down on top of the fishing tool, "jarring," or hammering, it downward into the drill bit. In this way, the fishing tool became firmly attached to the drill. Finally, by winding up the drilling cable, workers could, with luck, lift out the drill and everything else.

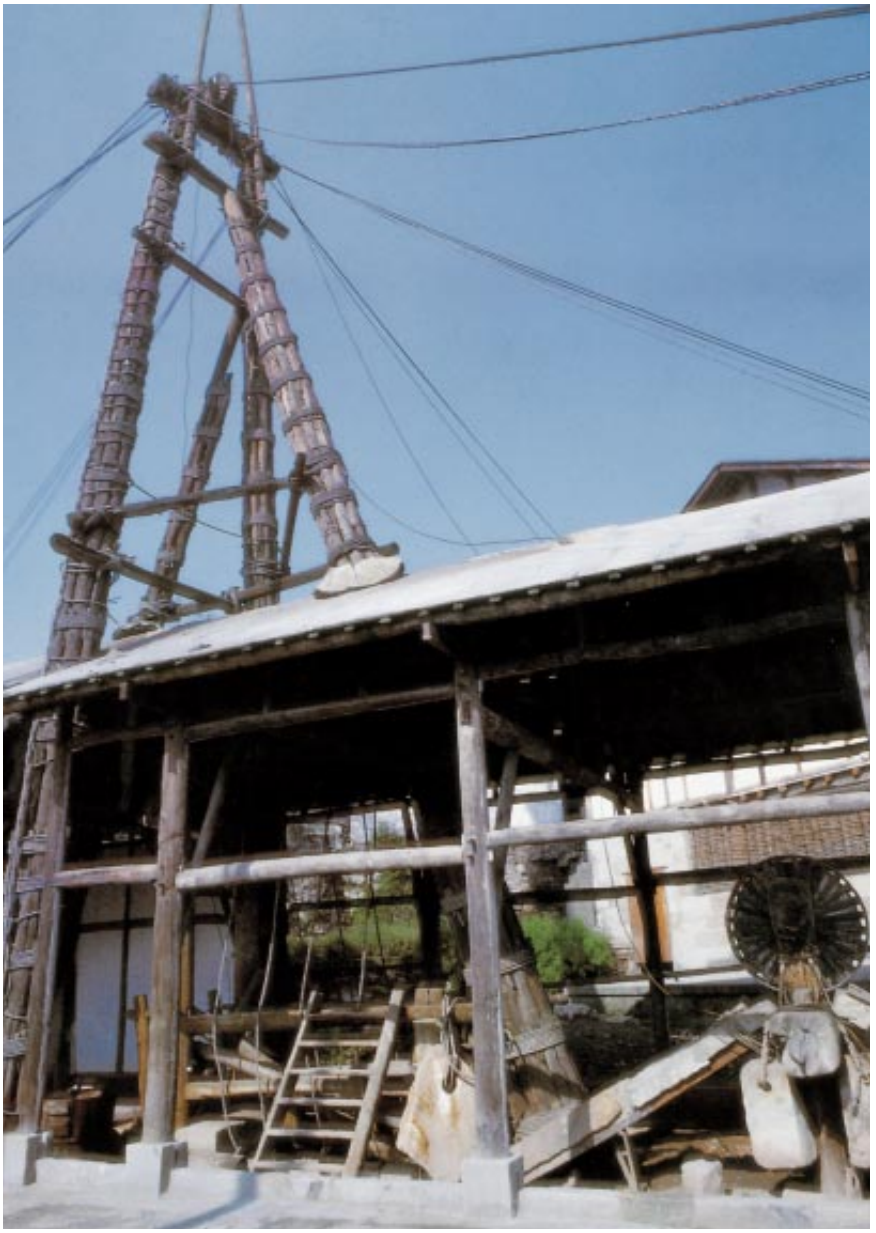
The Sichuan salt industry was disrupted in the 1640s and 1650s as the Qing armies waged several military campaigns in an attempt to conquer China. They succeeded and ruled for more than 260 years. During that time, the salt industry flourished. In the middle of the 18th century the northern Sichuanese in the districts of Shehong and Pengxi produced most of the well salt in China, but by the end of the century they had diminished their supply of fuel and their sources of brine. As production and revenues dropped, the Sichuan government collected fewer taxes and hence chose to reform the laws that regulated the salt industry. In the end, they decreased the overall tax burden and encouraged the construction of new wells.

Because of these events, the center of salt production shifted southeastward to Ziliujing, the site of the modern city of Zigong. The salt industry in Ziliujing prospered and captured a substantial portion of the Chinese market. In particular, the region enjoyed a surge in business during the Taiping Uprising of 1850–1864. Rebels blocked the trade of sea salt to the Yangtze River valley. The blockade left the well owners in the position of being the largest suppliers of salt in the region.

The Sichuanese of Ziliujing refined the art of borehole drilling to create an industry that was unmatched in China or anywhere in the world. Using essentially the same principles of drilling, hoisting and maintenance that were developed in the 11th century, they pushed the technology to its limits. By constructing many varieties of drilling and fishing tools and by increasing the size of the drilling and hoisting machinery, they managed to tap rich deposits of brine and natural gas in deep strata. Whereas in the 16th century the Sichuanese drilled as deep as 300 meters, their descendants achieved a depth of about 500 meters, in the 18th century, and nearly 800 meters, in the early 19th. The famous Xinhai well, which reached the astounding depth of one kilometer, was constructed in 1835. In contrast, during the 1820s European



BAMBOO TUBES were used to hoist brine from deep wells. A valve that had been installed at the bottom of the tube would open as the tube was pushed into the brine and would close again as the tube was removed.



XINHAI WELL, which reaches a depth of one kilometer, was by far the deepest in the world when it was drilled in 1835.

workers drilled down 370 meters, and in 1842 the German engineer C. G. Kind built the deepest borehole well outside of China; it measured only 535 meters.

The deep borehole wells of Ziliujing had two major advantages over the shallower wells in the area. First, whereas the salinity of the brine was rarely 10 percent at depths less than 300 meters, it rose to 18 percent at depths greater than 800 meters. Second, a natural-gas well of the mid-18th century typically yielded only enough gas for one or two burners, but the deep, prosperous wells of the 19th century could supply several hundred burners. As the Sichuanese of Ziliujing continued to increase the depth and number of wells, they increased the production of salt and natural gas considerably. Before the 1720s the annual output of the Sichuan salt industry was about 35,000 tons, yet by 1900 yearly production had grown by a factor of nine.

Although the salt industry in Sichuan was in many respects one of the most sophisticated enterprises in China, it never succeeded in dominating the Chinese salt industry. In the middle of the 19th century the Sichuanese of Ziliujing were spending 3,000 kilograms of monetary silver to set up each salt well, and their salt therefore had difficulty competing with that produced by less expensive techniques. The wells were economical only in the remote province of Sichuan, where drilling for salt was cheaper than importing it. Thus, Sichuan deep-drilling techniques never spread to other provinces.

Although the Europeans apparently developed methods for deep drilling independently of the Chinese, they eventually discovered and adopted the Sichuan technique, which was known as Chinese rope drilling. The earliest European description of Sichuan deep drilling appears in Martino Martini's *Novus Atlas Sinensis,* published in 1655. Martini's report was inaccurate in many respects, however, and it was for the most part ignored.

European engineers did not take a great interest in Sichuan deep drilling until the 1820s, when they learned of the boreholes of Ziliujing through the letters of a French missionary, L. J. M. Imbert. The missionary gave a rather detailed report of how the Sichuanese extracted brine and natural gas by drilling to depths of 800 meters and more—a feat many Europeans doubted anyone could accomplish.

Nevertheless, Europeans made many attempts to emulate the Chinese cable drilling and, in the 1840s, managed to reach depths of 100 meters. The main reason for their limited success was that they used rope instead of the stronger and more elastic bamboo cable. European and American engineers eventually improved on the Sichuan methods by replacing the ropes with iron cables, by incorporating jars into the drilling line and by powering the apparatus with steam engines. Despite these advances, Europeans favored percussion drilling with solid rods; Americans, on the other hand, found the cable method very useful, particularly in Pennsylvania and California.

Historians are still filling in many of the details about deep drilling and the salt industry in ancient China. Ironically, in 1984 when Joseph Needham asked me to revise a chapter on the salt industry in China for his monumental work *Science and Civilisation in China,* I thought it would be a simple and short project. As I uncovered a vast amount of information, the chapter turned into a book. Nine years later I am just putting the final touches on the manuscript.

FURTHER READING

SALT AND CIVILIZATION. Samuel A. M. Adshead. Macmillan, 1992.

SALT PRODUCTION TECHNIQUES IN ANCIENT CHINA: THE AOBO TU. Yoshida Tora. Translated and revised by Hans Ulrich Vogel. E. J. Brill, 1993.

TRENDS IN BEHAVIORAL GENETICS

EUGENICS REVISITED

by John Horgan, *senior writer*

Scientists are linking genes to a host of complex human disorders and traits, but just how valid—and useful—are these findings?

"How to Tell If Your Child's a Serial Killer!" That was the sound bite with which the television show *Donahue* sought to entice listeners February 25. On the program, a psychiatrist from the Rochester, N.Y., area noted that some men are born with not one Y chromosome but two. Double-Y men, the psychiatrist said, are "at special risk for antisocial, violent behavior." In fact, the psychiatrist had recently studied such a man. Although he had grown up in a "Norman Rockwell" setting, as an adult he had strangled at least 11 women and two children.

"It is not hysterical or overstating it," Phil Donahue told his horrified audience, "to say that we are moving toward the time when, quite literally, just as we can anticipate…genetic predispositions toward various physical diseases, we will also be able to pinpoint mental disorders which include aggression, antisocial behavior and the possibility of very serious criminal activity later on."

Eugenics is back in fashion. The message that genetics can explain, predict and even modify human behavior for the betterment of society is promulgated not just on sensationalistic talk shows but by our most prominent scientists. James D. Watson, co-discoverer of the double-helix structure of DNA and former head of the Human Genome Project, the massive effort to map our entire genetic endowment, said recently, "We used to think that our fate was in our stars. Now we know, in large part, that our fate is in our genes."

Daniel E. Koshland, Jr., a biologist at the University of California at Berkeley and editor of *Science,* the most influential peer-reviewed journal in the U.S., has declared in an editorial that the nature/nurture debate is "basically over," since scientists have shown that genes influence many aspects of human behavior. He has also contended that genetic research may help eliminate society's most intractable problems, including drug abuse, homelessness and, yes, violent crime.

Some studies cited to back this claim are remarkably similar to those conducted over a century ago by scientists such as Francis Galton, known as the father of eugenics. Just as the British polymath studied identical twins in order to show that "nature prevails enormously over nurture," so do modern researchers. But the primary reason behind the revival of eugenics is the astonishing successes of biologists in mapping and manipulating the human genome. Over the past decade, investigators have identified genes underlying such

NATURE OR NURTURE? Identical twins Paula and Nina Sossen were brought up together and share the same genes. Yet Paula (right)*, a graduate student at the University of Michigan, is heterosexual; Nina, a graphic designer in Madison, Wis., is a lesbian.*

"EERIE" PARALLELS between identical twins raised apart—such as Jerry Levey (*left*) and Mark Newman, who both became firefighters—are said to support genetic models of human behavior. Yet skeptics say the significance of such coincidences has been exaggerated.

crippling diseases as cystic fibrosis, muscular dystrophy and, this past spring, Huntington's disease. Given these advances, researchers say, it is only a matter of time before they can lay bare the genetic foundation of much more complex traits and disorders.

The political base for eugenics has also become considerably broader in recent years. Spokespersons for the mentally ill believe demonstrating the genetic basis of disorders such as schizophrenia and manic depression—and even alcoholism and drug addiction—will lead not only to better diagnoses and treatments but also to more compassion toward sufferers and their families. Some homosexuals believe society will become more tolerant toward them if it can be shown that sexual orientation is an innate, biological condition and not a matter of choice.

But critics contend that no good can come of bad science. Far from moving inexorably closer to its goals, they point out, the field of behavioral genetics is mired in the same problems that have always plagued it. Behavioral traits are extraordinarily difficult to define, and practically every claim of a genetic basis can also be explained as an environmental effect. "This has been a huge enterprise, and for the most part the work has been done shoddily. Even careful people get sucked into misinterpreting data," says Jonathan Beckwith, a geneticist at Harvard University. He adds, "There are social consequences to this."

The skeptics also accuse the media of having created an unrealistically optimistic view of the field. Richard C. Lewontin, a biologist at Harvard and a prominent critic of behavioral genetics, contends that the media generally give much more prominent coverage to dramatic reports—such as the discovery of an "alcoholism gene"—than to contradictory results or retractions. "Skepticism doesn't make the news," Lewontin says. "It only makes the news when you find a gene." The result is that spurious findings often become accepted by the public and even by so-called experts.

The claim that men with an extra Y chromosome are predisposed toward violence is a case in point. It stems from a survey in the 1960s that found more extra-Y men in prison than in the general population. Some researchers hypothesized that since the Y chromosome confers male attributes, men with an extra Y become hyperaggressive "supermales." Follow-up studies indicated that while extra-Y men tend to be taller than other men and score slightly lower on intelligence tests, they are otherwise normal. The National Academy of Sciences concluded in a report published this year that there is no evidence to support the link between the extra Y chromosome and violent behavior.

Minnesota Twins

No research in behavioral genetics has been more eagerly embraced by the press than the identical-twin studies done at the University of Minnesota. Thomas J. Bouchard, Jr., a psychologist, initiated them in the late 1970s, and since then they have been featured in the *Washington Post, Newsweek,* the *New York Times* and other publications worldwide as well as on television. *Science* has favorably described the Minnesota team's work in several news stories and in 1990 published a major article by the group.

The workers have studied more than 50 pairs of identical twins who were separated shortly after birth and raised in different households. The assumption is that any differences between identical twins, who share all each other's genes, are caused by the environment; similarities are attributed to their shared genes. The group estimates the relative contribution of genes to a given trait in a term called "heritability." A trait that stems entirely from genes, such as eye color, is defined as 100 percent heritable. Height is 90 percent heritable; that is, 90 percent of the variation in height is accounted for by genetic variation, and the other 10 percent is accounted for by diet and other environmental factors.

The Minnesota group has reported finding a strong genetic contribution to practically all the traits it has examined. Whereas most previous studies have estimated the heritability of intelligence (as defined by performance on intelligence tests) as roughly 50 percent, Bouchard and his colleagues arrived at a figure of 70 percent. They have also found a genetic component underlying such culturally defined traits as religiosity, political orientation (conservative versus liberal), job satisfaction, leisure-time interests and proneness to divorce. In fact, the group concluded in *Science,* "On multiple measures of personality and temperament…monozy-

gotic twins reared apart are about as similar as are monozygotic twins reared together." (Identical twins are called monozygotic because they stem from a single fertilized egg, or zygote.)

The researchers have buttressed their statistical findings with anecdotes about "eerie," "bewitching" and "remarkable" parallels between reunited twins. One case involved Oskar, who was raised as a Nazi in Czechoslovakia, and Jack, who was raised as a Jew in Trinidad. Both were reportedly wearing shirts with epaulets when they were reunited by the Minnesota group in 1979. They also both flushed the toilet before as well as after using it and enjoyed deliberately sneezing to startle people in elevators.

Some other celebrated cases involved two British women who wore seven rings and named their firstborn sons Richard Andrew and Andrew Richard; two men who both had been named Jim, named their pet dogs Toy, married women named Linda, divorced them and remarried women named Betty; and two men who had become firefighters and drank Budweiser beer.

Other twin researchers say the significance of these coincidences has been greatly exaggerated. Richard J. Rose of Indiana University, who is collaborating on a study of 16,000 pairs of twins in Finland, points out that "if you bring together strangers who were born on the same day in the same country and ask them to find similarities between them, you may find a lot of seemingly astounding coincidences."

Rose's collaborator, Jaakko Kaprio of the University of Helsinki, notes that the Minnesota twin studies may also be biased by their selection method. Whereas he and Rose gather data by combing birth registries and sending questionnaires to those identified as twins, the Minnesota group relies heavily on media coverage to recruit new twins. The twins then come to Minnesota for a week of study—and, often, further publicity. Twins who are "interested in publicity and willing to support it," Kaprio says, may be atypical. This self-selection effect, he adds, may explain why the Bouchard group's estimates of heritability tend to be higher than those of other studies.

One of the most outspoken critics of the Minnesota twin studies—and indeed all twin studies indicating high heritability of behavioral traits—is Leon J. Kamin, a psychologist at Northeastern University. In the 1970s Kamin helped to expose inconsistencies and possible fraud in studies of separated identical twins conducted by the British psychologist Cyril Burt during the previous two decades. Burt's conclusion that intelligence was mostly inherited had inspired various observers, notably Arthur R. Jensen, a psychologist at the University of California at Berkeley, to argue that socioeconomic stratification in the U.S. is largely a genetic phenomenon.

In his investigations of other twin studies, Kamin has shown that identical twins supposedly raised apart are often raised by members of their family or by unrelated families in the same neighborhood; some twins had extensive contact with each other while growing up. Kamin suspects the same may be true of some Minnesota twins. He notes, for example, that some news accounts suggested Oskar and Jack (the Nazi and the Jew) and the two British women wearing seven rings were reunited for the first time when they arrived in Minnesota to be studied by Bouchard. Actually, both pairs of twins had met previously. Kamin has repeatedly asked the Minnesota group for detailed case histories of its twins to determine whether it has underestimated contact and similarities in upbringing. "They've never responded," he says.

Kamin proposes that the Minnesota twins have particularly strong motives to downplay previous contacts and to exaggerate their similarities. They might want to please researchers, to attract more attention from the media or even to make money. In fact, some twins acquired agents and were paid for appearances on television. Jack and Oskar recently sold their life story to a film producer in Los Angeles (who says Robert Duvall is interested in the roles).

Even the Minnesota researchers caution against overinterpretation of their work. They agree with their critics that high heritability should not be equated with inevitability, since the environment can still drastically affect the expression of a gene. For example, the genetic disease phenylketonuria, which causes profound retardation, has a heritability of 100 percent. Yet eliminating the amino acid phenylalanine from the diet of affected persons prevents retardation from occurring.

Such warnings tend to be minimized in media coverage, however. Writers often make the same inference that Koshland did in an editorial in *Science:* "Better schools, a better environment, bet-

Behavioral Genetics: A Lack-of-Progress Report

CRIME: Family, twin and adoption studies have suggested a heritability of 0 to more than 50 percent for predisposition to crime. (Heritability represents the degree to which a trait stems from genetic factors.) In the 1960s researchers reported an association between an extra Y chromosome and violent crime in males. Follow-up studies found that association to be spurious.

MANIC DEPRESSION: Twin and family studies indicate heritability of 60 to 80 percent for susceptibility to manic depression. In 1987 two groups reported locating different genes linked to manic depression, one in Amish families and the other in Israeli families. Both reports have been retracted.

SCHIZOPHRENIA: Twin studies show heritability of 40 to 90 percent. In 1988 a group reported finding a gene linked to schizophrenia in British and Icelandic families. Other studies documented no linkage, and the initial claim has now been retracted.

ALCOHOLISM: Twin and adoption studies suggest heritability ranging from 0 to 60 percent. In 1990 a group claimed to link a gene—one that produces a receptor for the neurotransmitter dopamine—with alcoholism. A recent review of the evidence concluded it does not support a link.

INTELLIGENCE: Twin and adoption studies show a heritability of performance on intelligence tests of 20 to 80 percent. One group recently unveiled preliminary evidence for genetic markers for high intelligence (an IQ of 130 or higher). The study is unpublished.

HOMOSEXUALITY: In 1991 a researcher cited anatomic differences between the brains of heterosexual and homosexual males. Two recent twin studies have found a heritability of roughly 50 percent for predisposition to male or female homosexuality. These reports have been disputed. Another group claims to have preliminary evidence of genes linked to male homosexuality. The data have not been published.

ter counseling and better rehabilitation will help some individuals but not all." The prime minister of Singapore apparently reached the same conclusion. A decade ago he cited popular accounts of the Minnesota research in defending policies that encouraged middle-class Singaporeans to bear children and discouraged childbearing by the poor.

Smart Genes

Twin studies, of course, do not indicate which specific genes contribute to a trait. Early in the 1980s scientists began developing powerful ways to unearth that information. The techniques stem from the fact that certain stretches of human DNA, called polymorphisms, vary in a predictable way. If a polymorphism is consistently inherited together with a given trait—blue eyes, for example—then geneticists assume it either lies near a gene for that trait or actually is the gene. A polymorphism that merely lies near a gene is known as a marker.

In so-called linkage studies, investigators search for polymorphisms co-inherited with a trait in families unusually prone to the trait. In 1983 researchers used this method to find a marker linked to Huntington's disease, a crippling neurological disorder that usually strikes carriers in middle age and kills them within 10 years. Since then, the same technique has pinpointed genes for cystic fibrosis, muscular dystrophy and other diseases. In association studies, researchers compare the relative frequency of polymorphisms in two unrelated populations, one with the trait and one lacking it.

Workers are already using both methods to search for polymorphisms associated with intelligence, defined as the ability to score well on standardized intelligence tests. In 1991 Shelley D. Smith of the Boys Town National Institute for Communication Disorders in Children, in Omaha, and David W. Fulker of the University of Colorado identified polymorphisms associated with dyslexia in a linkage study of 19 families exhibiting high incidence of the reading disorder.

Two years ago Robert Plomin, a psychologist at Pennsylvania State University who has long been active in behavioral genetics, received a $600,000 grant from the National Institute of Child Health and Human Development to search for genes linked to high intelligence. Plomin is using the association method, which he says is more suited than the linkage technique to identifying genes whose contribution to a trait is relatively small. Plomin is studying a group of 64 schoolchildren 12 to 13 years old who fall into three groups: those who score approximately 130, 100 and 80 on intelligence tests.

Plomin has examined some 25 polymorphisms in each of these three groups, trying to determine whether any occur with greater frequency in the "bright" children. The polymorphisms have been linked to genes thought to have neurological effects. He has uncovered several markers that seem to occur more often in the highest-scoring children. He is now seeking to replicate his results in another group of 60 children; half score above 142 on intelligence tests, and half score less than 74 (yet have no obvious organic deficiencies). Plomin presented his preliminary findings at a meeting, titled "Origins and Development of High Ability," held in London in January.

At the same meeting, however, other workers offered evidence that intelligence tests are actually poor predictors of success in business, the arts or even advanced academic programs. Indeed, even Plomin seems ambivalent about the value of his research. He suggests that someday genetic information on the cognitive abilities of children might help teachers design lessons that are more suited to students' innate strengths and weaknesses.

But he also calls his approach "a fishing expedition," given that a large number of genes may contribute to intelligence. He thinks the heritability of intelligence is not 70 percent, as the Minnesota twin researchers have claimed, but 50 percent, which is the average finding of other studies, and at best he can only find a gene that accounts for a tiny part of variance in intelligence. "If you wanted to select on the basis of this, it would be of no use whatsoever," he remarks. These cautions did not prevent the *Sunday Telegraph,* a London newspaper, from announcing that Plo-

The Huntington's Disease Saga: A Cautionary Tale

The identification of the gene for Huntington's disease, which was announced in March, was hailed as one of the great success stories of modern genetics. Yet it provides some rather sobering lessons for researchers seeking genes linked to more complex human disorders and traits.

The story begins in the late 1970s, when workers developed novel techniques for identifying polymorphisms, sections of the human genome that come in two or more forms. Investigators realized that by finding polymorphisms linked—always and exclusively—to diseases, they could determine which chromosome the gene resides in. Researchers decided to test the polymorphism technique on Huntington's disease, a devastating neurological disorder that affects roughly one in 10,000 people. Scientists had known for more than a century that Huntington's was caused by a mutant, dominant gene. If one parent has the disease, his or her offspring have a 50 percent chance of inheriting it.

One of the leaders of the Huntington's effort was Nancy Wexler, a neuropsychologist at Columbia University whose mother had died of the disease and who therefore has a 50 percent chance of developing it herself. She and other researchers focused on a poor Venezuelan village whose inhabitants had an unusually high incidence of the disease. In 1983, through what has now become a legendary stroke of good fortune, they found a linkage with one of the first polymorphisms they tested. The linkage indicated that the gene for Huntington's disease was somewhere on chromosome 4.

The finding led quickly to a test for determining whether offspring of carriers—either in utero or already born—have inherited the gene itself. The test requires an analysis of blood samples from several members of a family known to carry the disease. Wexler herself has declined to say whether she has taken the test.

Researchers assumed that they would quickly identify the actual gene in chromosome 4 that causes Huntington's disease. Yet it took 10 years for six teams of workers from 10 institutions to find the gene. It is a so-called expanding gene, which for unknown reasons gains base pairs (the chemical "rungs" binding two strands of DNA) every time it is transmitted. The greater the expansion of the gene, researchers say, the earlier the onset of the disease. The search was complicated by the fact that workers had no physical clues about the course of the disease to guide them. Indeed, Wexler and others emphasize that they still have no idea how the gene actually causes the disease; treatments or cures may be years or decades away.

The most immediate impact of the new discovery will be the development of a better test for Huntington's, one that requires blood only from the person at risk

NANCY WEXLER helped to find the gene responsible for Huntington's disease by studying a population in Venezuela that has been ravaged by the disorder.

and not other family members. By measuring the length of the mutant gene, the test might also predict more accurately when carriers will show symptoms.

As difficult as it was to pinpoint the gene for Huntington's, it will be almost infinitely harder to discover genes for behavioral disorders, says Evan S. Balaban, a biologist at Harvard University. Unlike Huntington's disease, he notes, disorders such as schizophrenia and alcoholism cannot be unambiguously diagnosed. Furthermore, they stem not from a single dominant gene but from many genes acting in concert with environmental effects. If researchers do find a statistical association between certain genes and a trait, Balaban says, that knowledge may never be translated into useful therapies or tests. "What does it mean to have a 10 percent increased risk of alcoholism?" he asks.

min had found "evidence that geniuses are born not made."

Evan S. Balaban, a biologist at Harvard, thinks Plomin's fishing expedition is doomed to fail. He grants that there may well be a significant genetic component to intelligence (while insisting that studies by Bouchard and others have not demonstrated one). But he doubts whether investigators will ever uncover any specific genes related to high intelligence or "genius." "It is very rare to find genes that have a specific effect," he says. "For evolutionary reasons, this just doesn't happen very often."

The history of the search for markers associated with mental illness supports Balaban's view. Over the past few decades, studies of twins, families and adoptees have convinced most investigators that schizophrenia and manic depression are not caused by psychosocial factors—such as the notorious "schizophrenogenic mother" postulated by some Freudian psychiatrists—but by biological and genetic factors. After observing the dramatic success of linkage studies in the early 1980s, researchers immediately began using the technique to isolate polymorphic markers for mental illness. The potential value of such research was enormous, given that schizophrenia and manic depression each affect roughly one percent of the global population.

They seemed to have achieved their first great success in 1987. A group led by Janice A. Egeland of the University of Miami School of Medicine claimed it had linked a genetic marker on chromosome 11 to manic depression in an Amish population. That same year another team, led by Miron Baron of Columbia University, linked a marker on the X chromosome to manic depression in three Israeli families.

The media hailed these announcements as major breakthroughs. Far less attention was paid to the retractions that followed. A more extensive analysis of the Amish in 1989 by a group from the National Institute of Mental Health turned up no link between chromosome 11 and manic depression. This year Baron's team retracted its claim of linkage with the X chromosome after doing a new study of its Israeli families with more sophisticated markers and more extensive diagnoses.

Schizophrenic Results

Studies of schizophrenia have followed a remarkably similar course. In 1988 a group headed by Hugh M. D. Gurling of the University College, London, Medical School announced in *Nature* that it had found linkage in Icelandic and British families between genetic markers on chromosome 5 and schizophrenia. In the same issue, however, researchers led by Kenneth K. Kidd of Yale University reported seeing no such linkage in a Swedish family. Although Gurling defended his result as legitimate for several years, additional research has convinced him that it was probably a false positive. "The new families showed no linkage at all," he says.

These disappointments have highlighted the problems involved in using linkage to study mental illness. Neil Risch, a geneticist at Yale, points out that linkage analysis is ideal for studying diseases, such as Huntington's, that have distinct symptoms and are caused by a single dominant gene. Some researchers had hoped that at least certain subtypes of schizophrenia or manic depression might be single-gene disorders. Single-gene mutations are thought to cause variants of breast cancer and of Alzheimer's disease that run in families and are manifested much earlier than usual. But such diseases are rare, Risch says, because natural selection quickly winnows them out of the population, and no evidence exists for distinct subtypes of manic depression or schizophrenia.

Indeed, all the available evidence suggests that schizophrenia and manic depression are caused by at least several genes—each of which may exert only a tiny influence—acting in concert with environmental influences. Finding such genes with linkage analysis may not be impossible, Risch says, but it will be considerably more difficult than identifying genes that have a one-to-one correspondence to a trait. The difficulty is compounded by the fact that the diagnosis of mental illness is often subjective—all the more so when researchers are relying on family records or recollections.

Some experts now question whether genes play a significant role in mental illness. "Personally, I think we have overestimated the genetic component of schizophrenia," says E. Fuller Torrey, a psychiatrist at St. Elizabeth's Hospital in Washington, D.C. He argues that the evidence supporting genetic models can be explained by other biological factors, such as a virus that strikes in utero. The pattern of incidence of schizophrenia in families often resembles that of other viral diseases, such as polio. "Genes may just create a susceptibility to the virus," Torrey explains.

The Drink Link

Even Kidd, the Yale geneticist who has devoted his career to searching for genes linked to mental illness, acknowledges that "in a rigorous, technical, scientific sense, there is very little proof that schizophrenia, manic depression" and other psychiatric disorders have a genetic origin. "Virtually all the evidence supports a genetic explanation, but there are always other explanations, even if they are convoluted."

The evidence for a genetic basis for alcoholism is even more tentative than that for manic depression and schizophrenia. Although some studies discern a genetic component, especially in males, others have reached the opposite conclusion. Gurling, the University College investigator, found a decade ago that identical twins were slightly *more* likely to be discordant for alcoholism than fraternal twins. The drinking habits of some identical twins were strikingly different. "In some cases, one drank a few bottles a day, and the other didn't drink at all," Gurling says.

Nevertheless, in 1990 a group led by Kenneth Blum of the University of Texas Health Science Center at San Antonio announced it had discovered a genetic marker for alcoholism in an association study comparing 35 alcoholics with a control group of 35 nonalcoholics. A page-one story in the *New York Times* portrayed the research as a potential watershed in the diagnosis and treatment of alcoholism without mentioning the considerable skepticism aroused among other researchers.

The Blum group claimed that its marker, called the A1 allele, was associated with a gene, called the D2 gene, that codes for a receptor for the neurotransmitter dopamine. Skeptics noted that the A1 allele was actually some 10,000 base pairs from the dopamine-receptor gene and was not linked to any detectable variation in its expression.

Since the initial announcement by Blum, three papers, including an additional one by Blum's group, have presented more evidence of an association between the A1 allele and alcoholism. Six groups have found no such evidence (and received virtually no mention in the popular media).

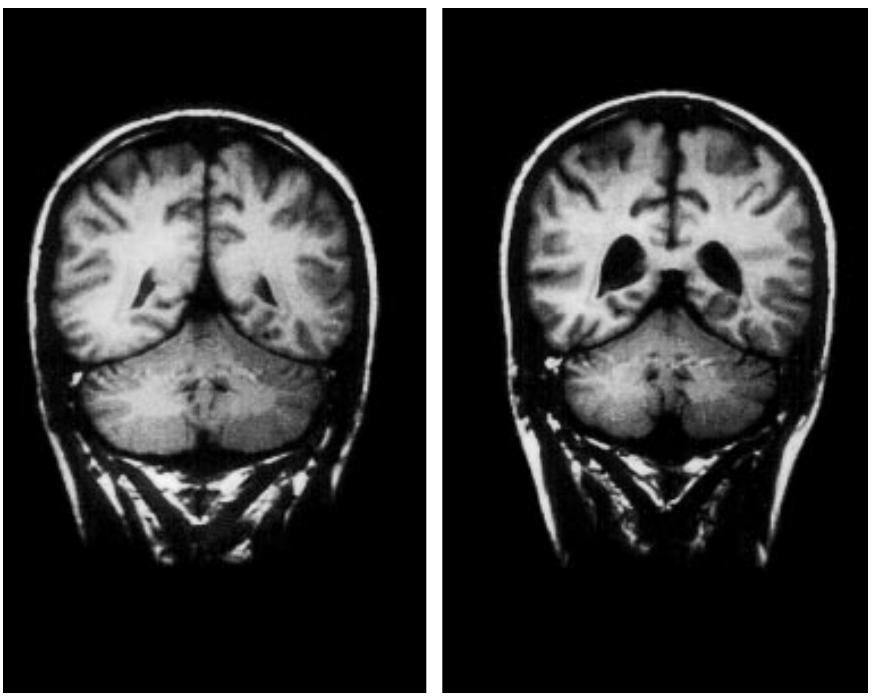
In April, Risch and Joel Gelernter of Yale and David Goldman of the National Institute on Alcohol Abuse and Alcoholism analyzed all these studies on the A1 allele in a paper in the *Journal of the American Medical Association*. They noted that if Blum's two studies are cast aside, the balance of the results shows no association between the D2 receptor and alcoholism, either in the disorder's milder or most severe forms. "We therefore conclude that no physiologically significant association" between the A1 allele and alcoholism has been proved, the group stated. "It's a dead issue," Risch says.

Gelernter and his colleagues point out that association studies are prone to spurious results if not properly controlled. They suggest that the positive findings of Blum and his colleagues may have derived from a failure to control for ethnic variation. The limited surveys done so far have shown that the incidence of the A1 allele varies wildly in different ethnic groups, ranging from 10 percent in certain Jewish groups to about 50 percent in Japanese.

Blum insists that the ethnic data, far from undermining his case, support it, since those groups with the highest prevalence of the A1 allele also exhibit the highest rates of "addictive behavior." He contends that the only reason the Japanese do not display higher rates of alcoholism is that many also carry a gene that prevents them from metabolizing alcohol. "They're pretty compulsive," explains Blum, who recently obtained a patent for a genetic test for alcoholism.

These arguments have been rejected even by Irving I. Gottesman of the University of Virginia, who is a strong defender of genetic models of human behavior. He considers the papers cited by Blum to support his case to be ambiguous and even contradictory. Some see an association only with alcoholism that leads to medical complications or even death; others discern no association with alcoholism but only with "polysubstance abuse," including cigarette smoking. "I think it is by and large garbage," Gottesman says of the alleged A1-alcoholism link.

By far the most controversial area of behavioral genetics is research on crime. Last fall complaints by civil-rights leaders and others led the National Institutes of Health to withdraw its funding from a meeting entitled "Genetic Factors in Crime: Findings, Uses and Implications." The conference brochure had noted the "apparent failure of environmental approaches to crime" and suggested that genetic research might yield methods for identifying and treating po-



BRAIN OF SCHIZOPHRENIC (*right*) appears different from the brain of his identical twin in these magnetic resonance images. Such findings suggest that factors that are biological but not genetic—such as viruses—may play a significant role in mental illness.

tential criminals—and particularly those prone to violence—at an early age.

Critics contend that such investigations inevitably suggest that blacks are predisposed to crime, given that blacks in the U.S. are six times more likely than whites to be arrested for a violent crime. In fact, some prominent scientists, notably Richard J. Herrnstein, a psychologist at Harvard, have made this assertion. Others reject this view but insist biological research on attributes linked to violent crime, such as aggression, may still have some value. "People who are unwilling to address genetic and biochemical factors are just putting their heads in the sand," says Goldman, the alcoholism expert. "It is not fair to say that just because there have been geneticists who have had a very narrow view of this in the past, we shouldn't explore this now."

In fact, investigations of the biology of violent crime continue, albeit quietly. Workers at City of Hope Hospital in Duarte, Calif., claim to have found an association between the A1 allele—the alleged alcoholism marker—and "criminal aggression." Last year a group led by Markus J. P. Kruesi of the University of Illinois at Chicago presented evidence of an association between low levels of the neurotransmitter serotonin and disruptive-behavior disorders in children. Kruesi concedes there is no way to determine whether the serotonin levels are genetically influenced. In fact, the serotonin levels might be an effect—a reaction to an environmental trauma—rather than a cause. "This might be a scar marker," he says.

One reason such research persists is that studies of families, twins and adoptees have suggested a genetic component to crime. Glenn D. Walters, a psychologist at the Federal Correctional Institution in Schuylkill, Pa., recently reviewed 38 of these studies, conducted from the 1930s to the present, in the journal *Criminology*. His meta-analysis turned up a small genetic effect, "but nothing to get excited about." He observes that "a lot of the research has not been very good" and that the more recent, better-designed studies tended to turn up less evidence. "I don't think we will find any biological markers for crime," he says. "We should put our resources elsewhere."

Gay Genes

The ostensible purpose of investigations of mental illness, alcoholism and even crime is to reduce their incidence. Scientists studying homosexuality have a different goal: simply to test whether homosexuality is innate, as many homosexuals have long professed. That claim was advanced by a report in *Science* in 1991 by Simon LeVay of the Salk Institute for Biological Studies in San Diego. LeVay has acknowledged both that he is gay and that he believes evidence of biological differences between homosexuals and heterosexuals will encourage tolerance toward gays.

LeVay, who recently left the Salk Institute to found the Institute of Gay and Lesbian Education, focused on a tiny neural structure in the hypothalamus, a region of the brain known to control sexual response. He measured this structure, called the interstitial nu-

"Better Breeding"

Fairly or not, modern genetics research is still haunted by the history of eugenics. "It offers a lot of cautionary lessons," says Daniel J. Kevles, a historian at the California Institute of Technology, who wrote the 1985 book *In the Name of Eugenics*. The British scientist Francis Galton, cousin to Charles Darwin, first proposed that human society could be improved "through better breeding" in 1865 in an article entitled "Hereditary Talent and Character." He coined the term "eugenics," from the Greek for "good birth," in 1883.

Galton's proposal had broad appeal. The American sexual libertarian John Humphrey Noyes bent eugenics into an ingenious argument for polygamy. "While the good man will be limited by his conscience to what the law allows," Noyes said, "the bad man, free from moral check, will distribute his seed beyond the legal limit."

A more serious advocate was the biologist Charles B. Davenport, founder of Cold Spring Harbor Laboratory and of the Eugenics Record Office, which gathered information on thousands of American families for genetic research. After demonstrating the heritability of eye, skin and hair color, Davenport went on to "prove" the heritability of traits such as "pauperism," criminality and "feeble-mindedness." In one monograph, published in 1919, he asserted that the ability to be a naval officer is an inherited trait, composed of subtraits for thalassophilia, or love of the sea, and hyperkineticism, or wanderlust. Noting the paucity of female naval officers, Davenport concluded that the trait is unique to males.

Beginning in the 1920s the American Eugenics Society, founded by Davenport and others, sponsored "Fitter Families Contests" at state fairs around the U.S. Just as cows and sheep were appraised by judges at the fairs, so were human entrants (such as the family shown above at the 1925 Texas State Fair). Less amusingly, eugenicists helped to persuade more than 20 U.S. states to authorize sterilization of men and women in prisons and mental hospitals, and they urged the federal government to restrict the immigration of "undesirable" races.

No nation, of course, practiced eugenics as enthusiastically as Nazi Germany, whose program culminated in "euthanasia" ("good death") of the mentally and physically disabled as well as Jews, Gypsies, Catholics and others. As revelations of these atrocities spread after World War II, popular support for eugenics programs waned in the U.S. and elsewhere.

cleus, in autopsies of the brains of 19 homosexual males, 16 heterosexual males and six heterosexual women. LeVay found that the interstitial nucleus was almost twice as large in the heterosexual males as in the homosexual males or in the women. He postulated that the interstitial nucleus "is large in individuals oriented toward women"—whether male or female.

Of course, LeVay's finding only addresses anatomic differences, not necessarily genetic ones. Various other researchers have tried to establish that homosexuality is not just biological in its origin—caused, perhaps, by hormonal influences in utero—but also genetic. Some have sought evidence in experiments with rats and other animals. A group headed by Angela Pattatucci of the National Cancer Institute is studying a strain of male fruit flies—which wags have dubbed either "fruity" or "fruitless"—that court other males.

In December 1991 J. Michael Bailey of Northwestern University and Richard C. Pillard of Boston University announced they had uncovered evidence of a genetic basis for male homosexuality in humans. They studied 161 gay men, each of whom had at least one identical or fraternal twin or adopted brother. The researchers determined that 52 percent of the identical twins were both homosexual, as compared with 22 percent of the fraternal twins and 11 percent of the adopted brothers.

Bailey and Pillard derived similar results in a study of lesbians published this year in the *Archives of General Psychiatry*. They compared 147 gay women with identical or fraternal twins or adopted sisters: 48 percent of the identical twins were both gay, versus 16 percent of the fraternal twins (who share only half each other's genes) and 6 percent of the adopted sisters. "Both male and female sexual orientation appeared to be influenced by genetic factors," Bailey and Pillard concluded.

This conclusion has disturbed some of Bailey and Pillard's own subjects. "I have major questions about the validity of some of the assumptions they are making," says Nina Sossen, a gay woman living in Madison, Wis., whose identical twin is heterosexual. Her doubts are shared by William Byne, a psychiatrist at Columbia University. He notes that in their study of male homosexuality Bailey and Pillard found more concordance between unrelated, adopted brothers than related (but non-twin) brothers. The high concordance of the male and female identical twins, moreover, may stem from the fact that such twins are often dressed alike and treated alike—indeed, they are often mistaken for each other—by family members as well as by others.

"The increased concordance for homosexuality among the identical twins could be entirely accounted for by the increased similarity of their developmental experiences," Byne says. "In my opinion, the major finding of that study is that 48 percent of identical twins who were reared together were discordant for sexual orientation."

Byne also criticizes LeVay's conclusion that homosexuality must be biological—although not necessarily genetic—because the brains of male homosexuals resemble the brains of women. That assumption, Byne points out, rests on still another assumption, that there are significant anatomic differences between heterosexual male and female brains. But to date, there have been no replicable studies showing such sexual dimorphism.

Byne notes that he has been suspected of having an antigay motive. Two reviewers of an article he recently wrote criticizing homosexuality research accused him of having a "right-wing agenda," he says. He has also been contacted by conservative groups hoping he will speak out against the admittance of homosexuals to the military. He emphasizes that he supports gay rights and thinks homosexuality, whatever its cause, is not a "choice." He adds that genetic models of behavior are just as likely to foment bigotry as to quell it.

"Hierarchy of Worthlessness"

Despite the skepticism of Byne and others, at least one group, led by Dean Hamer of the National Cancer Institute, is searching not merely for anatomic or biochemical differences in homosexuals but for genetic markers. Hamer has done a linkage study of numerous small families, each of which has at least two gay brothers. He says his study has turned up some tentative findings, and he plans to submit his results soon. Hamer's colleague Pattatucci is planning a similar study of lesbians.

What purpose will be served by pinpointing genes linked to homosexuality? In an information sheet for prospective participants in his study, Hamer expresses the hope that his research may "improve understanding between people with different sexual orientations." He adds, "This study is not aimed at developing methods to alter either heterosexual or homosexual orientation, and the results of the study will not allow sexual orientation to be determined by a blood test or amniocentesis."

Yet even Pillard, who is gay and applauds Hamer's work, admits to some concern over the potential uses of a genetic marker for homosexuality. He notes that some parents might choose to abort embryos carrying such a marker. Male and female homosexuals might then retaliate, he says, by conceiving children and aborting fetuses that lacked such a gene.

Balaban, the Harvard biologist, thinks the possible dangers of such research—assuming it is successful—outweigh any benefits. Indeed, he sees behavioral genetics as a "hierarchy of worthlessness," with twin studies at the bottom and linkage studies of mental illness at the top. The best researchers can hope for is to find, say, a gene associated with a slightly elevated risk of schizophrenia. Such information is more likely to lead to discrimination by insurance companies and employers than to therapeutic benefits, Balaban warns.

His colleague Lewontin agrees. In the 1970s, he recalls, insurance companies began requiring black customers to take tests for sickle cell anemia, a genetic disease that primarily affects blacks. Those who refused to take the test or who tested positive were denied coverage. "I feel that this research is a substitute for what is really hard—finding out how to change social conditions," Lewontin remarks. "I think it's the wrong direction for research, given that we have a finite amount of resources."

Paul R. Billings, a geneticist at the California Pacific Medical Center, shares some of these concerns. He agrees that twin studies seem to be inherently ambiguous, and he urges researchers seeking markers for homosexuality to consider what a conservative government—led by Patrick Buchanan, for example—might allow to be done with such information. But he believes some aspects of behavioral genetics, particularly searches for genes underlying mental illness, are worth pursuing.

In an article published in the British journal *Social Science and Medicine* last year, Billings and two other scientists offered some constructive criticism for the field. Researchers engaged in association and linkage studies should establish "strict criteria as to what would constitute meaningful data." Both scientists and the press should emphasize the limitations of such studies, "especially when the mechanism of how a gene acts on a behavior is not known." Billings and his colleagues strive to end their article on a positive note. "Despite the shortcomings of other studies," they say, "there is relatively good evidence for a site on the X chromosome which is associated with [manic depression] in some families." This finding was retracted earlier this year.

Red-Banner Burger

Toward food inspection that assures safety

A nation's identity is tied as much to its food as to its flag. So a note of irony sounded through the tragedy precipitated this past winter by the consumption of that icon of youthful Americana, the hamburger. Hundreds of people who had eaten at a fast-food chain in Washington and other states contracted a gastrointestinal illness caused by a bacterium, *Escherichia coli* 0157:H7—two died.

The outbreak, one of the first crises for the Clinton administration, underlined the serious deficiencies in the current system of food inspection. And it demonstrated how harmful microbes, not residues of hazardous chemicals and pesticides, pose the greatest threat to the nation's food supply. "If you rank the public's safety concerns and those of the experts, they're just backwards from each other," says Robert L. Buchanan, a research leader for the U.S. Department of Agriculture's Agricultural Research Service.

E. coli 0157:H7 is a variant of the normal gut bacteria that was awarded the status of human pathogen by the Centers for Disease Control only in 1982, when it caused outbreaks of gastrointestinal illness in Michigan and Oregon. The affliction ranges in severity. The bacterium may produce no symptoms at one end of the spectrum. Or it may cause bloody diarrhea with complications resulting in kidney failure and death. It can also be transmitted through human contact. Two other individuals who died in the recent outbreak did not eat any hamburger but are believed to have contracted the disease from secondary infections.

The virulent "H7" joins an expanding list of bugs that trigger illness. The National Academy of Sciences's Institute of Medicine reported last year that the number of known pathogens carried in food tripled during the past 20 years, in part because of improved methods of identification. Eliminating microbes on the farm or in a meat plant requires seeking out not just *E. coli* but *Listeria, Salmonella, Campylobacter* and a slew of other microorganisms.

That is exactly what the USDA says it wants to do. Less than two months after the outbreak, its Food Safety and Inspection Service issued a plan that set out the ambitious goal of instituting a "science-based" system of meat inspection. This Pathogen Reduction Program pledges efforts to help eradicate microbial contamination from "farm to table." It would go beyond the present procedure of inspecting each and every carcass for disease at the slaughterhouse by look, feel and smell—a system set up in 1906 after Upton Sinclair detailed abuses in meat-packing plants in Chicago. Despite far-reaching goals, the plan is short on specifics. The USDA has yet to delineate the role of its 7,500 meat and poultry inspectors.

Neither this plan nor other USDA documents make clear whether the department favors retraining inspectors as food technologists who, instead of inspecting every carcass, would focus on microbial testing and use statistical methods to target meat-production operations that pose the highest risk. Previous efforts to streamline and modernize inspectors' jobs led to charges by some inspectors that the department was trying to undercut their authority and institute de facto deregulation.

Although the inspectors' jobs remain to be clarified, the new plan does endorse a risk-analysis technique that was used by Pillsbury during the 1960s to help the National Aeronautics and Space Administration ensure that astronauts' food was kept safe. It tries to zero in on the points in meat production where pathogens can enter.

Updating the system—finding the critical checkpoints—will be tedious because no one is certain how microorganisms propagate along the production chain and how many are needed to cause human illness. In the case of *E. coli*, researchers are still trying to identify the animals that harbor the organism and the reasons they may be colonized by the bacterium, which does

LARRY LEFEVER *Grant Heilman*

***TRADITIONAL MEAT INSPECTION** requires that a carcass be examined by look, feel and smell. The USDA now has a program to reduce invisible pathogens in meat.*

not appear to affect the animals' health. "We know almost nothing about it on the farm," says Dale D. Hancock, a veterinary epidemiologist at Washington State University, who tracked the bug for the USDA in livestock, detecting it in only 25 of 7,000 calves in 28 states.

A better understanding of *E. coli* 0157:H7 and other harmful microbes may bring about measures that can be adopted on the farm. They include changes in the way livestock are raised, the novel step of vaccinating animals for a human disease and feeding them with bacteria that will keep the malign *E. coli* from establishing itself in the intestinal tract, a practice known as competitive exclusion.

But the research needed to implement such a farm-based food safety program has not thrived. Six universities that formed the Food Animal Production Medicine Consortium have amassed only $800,000 in funding over four years. "It is minuscule," concedes Bennie I. Osburn, a professor of animal pathology at the University of California at Davis and a consortium director.

Scientists, regulators and industrial food scientists want diagnostic tests for the slaughterhouse or food processor that would, in a matter of seconds, confirm the presence of the 20 or so most wanted microbes at levels as low as one bacterium per 25 grams of meat. (It now takes up to a week to culture enough of *E. coli* 0157:H7 bacteria to get meaningful test results.) Naturally, such tests should also cost just a few pennies. That dream will remain in the imagination of food scientists for some time to come.

A number of ideas are under consideration. In March, Organon Teknika, a unit of Akzo, the Netherlands-based conglomerate, released an immunoassay that uses a monoclonal antibody to target *E. coli* 0157:H7. The test still takes a day. For nearly instantaneous results, several companies are working on applying techniques that use enzymes to amplify nucleic acids rapidly from *E. coli* 0157:H7 and other microbes—the polymerase chain reaction is one. Today these methods have trouble distinguishing between dead and live cells, which can produce false positive results.

Even with fast, inexpensive tests, microbial evaluation will probably be no more than a statistical check. To dampen any heightened expectations, the Food Safety and Inspection Service has warned that examining 20 percent of beef for the most dangerous pathogens could cost $58 billion.

Enter what is euphemistically known as the meat pasteurizer: irradiation. A Florida company, bearing the perhaps unfortunate name of Vindicator, has been training a beam of gamma rays on fruits and vegetables. The same facility is preparing to zap chickens, a process that received final regulatory approval last year.

The *E. coli* incident may mean that red meat is next in line. The Pathogen Reduction Program calls for assessing the effectiveness of irradiation, which could result in a petition to the Food and Drug Administration to endorse the process. The USDA's Food Safety Research Unit in Philadelphia published a study in the April issue of the *Journal of Applied Environmental Microbiology* showing that irradiation could effectively control *E. coli* 0157:H7 in meats.

At midyear, the American Meat Institute expects first results of its study on production-sized lots of irradiated meat. James Marsden, vice president of scientific and technical affairs for the industry group, said he attended a meeting earlier this year at which Agriculture Secretary Mike Espy told a group of meat executives that irradiation would be mandated unless other means of controlling microbial contamination could be found. No single method of retarding pathogens—

Is That Shirt Ascorbic Acid?

As any three-year-old will show you, wearing food is more fun than eating it. That was not quite what Michael W. Davidson, a biophysicist from the National High Magnetic Field Laboratory at Florida State University, had in mind in the mid-1980s. Back then he was using photomicrography to see how DNA crystals organize themselves in such places as sperm heads and bacteria.

Now Davidson finds his colorful and unusual images—of medicines, hormones, superconductors and pesticides, to name a few—the focus of a mini-industry. Vitamins and other images from Davidson's laboratory will be smeared on T-shirts, neckties and other body wear and will be available this summer—without the laundry bill. The scientific images have already begun to work their way onto other mass-market items, among them greeting cards, gift wrapping and even a New Age album cover.

The graphics are the result of improved film emulsions and camera technology coupled to various kinds of microscope techniques. The primary technique Davidson relies on is optically polarized light micrography, in which, he explains, "the polarized light goes through, and the crystals bend the light." The degree of refraction and hence the resulting colors depend on the molecular orientation and the thickness of the crystal. The method is ideal for many biological molecules, including the AIDS drug zidovudine (AZT) (*top left, opposite page*).

For samples too thick to transmit light, Davidson uses the reflected differential interference contrast method. This technique calls for the apparatus to split the illumination into two components. One beam bounces off the sample and

PHIL SEARS

MICROSCOPIST-TURNED-ARTIST Michael W. Davidson of Florida State University models a T-shirt splashed with vitamin K.

for example, washing meat with organic rinses such as acetic acid—is as effective as irradiation, Marsden says.

Consumer acceptance of irradiation is still an imponderable. Food safety groups believe more assurances are needed on such matters as vitamin loss, chemicals formed by the irradiation of food and the types of regulatory procedures that should be put in place to protect workers from radioisotopes.

Espy, a former congressman, will have to act as arbiter on many of these issues. He took an active part in responding to the *E. coli* crisis, which occurred about the time he became head of the department. The administration wants to hire 160 new inspectors to fill more than a quarter of the 550 vacancies. Espy also met with a group of whistle-blowing inspectors, who detailed complaints about the whittling away of their authority.

For the moment, the nearly 100-year-old inspection system will remain in place, and all-American hamburger lovers may have to settle for culinary advice endorsed by those as far apart on the spectrum as the American Meat Institute and Jeremy Rifkin, the consumer advocate and biotechnology critic. "The days of rare hamburgers are over," Rifkin railed in February, as he announced a suit being filed against the USDA to get a stern warning label inside packages of ground meat.

The USDA, for its part, is considering a rule to mandate handling and cooking instructions on the package. But its seal of approval may remain, despite the possible presence of microbial contamination. Let the diner beware. Having it your way may mean the choice of a hamburger well done or just plain burned. *—Gary Stix*

then recombines with the other. The beams interfere to produce, with the appropriate filters and prisms, a colorful, three-dimensional effect, such as that of the high-temperature superconductor yttrium-barium–copper oxide silk-screened onto an alumina substrate (*top right*).

Davidson can now use his imaging technique to create surreal artwork he calls microscapes. For instance, his "Nuclear Sunrise" photograph (*bottom*) is made from multiple exposures of different compounds. The desert foreground is ascorbic acid, the morning sky is stretched polyethylene and the stars (mostly obscured by the polyethylene) are small spheres of polybenzyl-L-glutamate. The rising sun is an image of the microscope diaphragm, slightly out of focus. Two of his microscapes have won Davidson grand prizes in photomicroscopy competitions sponsored by Polaroid, and New Age musician Richard Burmer is using one for the cover of an upcoming album.

Artistic recognition rarely translates into untold wealth, but it helps with the laboratory overhead. "About a year ago we decided to make some money on things that aren't ordinarily funded," says Davidson, who estimates that he sent out portfolios to more than 300 companies. His shotgun approach paid off. Amber Lotus, a paper products company in Oakland, Calif., already markets calendars, gift wrap and greeting cards with Davidson's motifs. In late July, Danskin, an apparel firm based in New York City, expects to introduce T-shirts and exercise wear featuring images of vitamins. Davidson has signed licensing agreements with several other firms as well. Florida State gets the proceeds, so "the people in the laboratory are really happy," Davidson notes. "We never expected it to get to this point." *—Philip Yam*

MICHAEL W. DAVIDSON

Extinguished

A champion firefighter goes down for the count

courtesy of AVIATION WEEK & SPACE TECHNOLOGY

FIRE EXTINGUISHER that didn't. The military tests replacements for halon—widely used to put out engine fires but banned as of January 1994—on its gunnery range.

At one end of a horseshoe-shaped bunker at Wright-Patterson Air Force Base in Dayton, Ohio, seven B-52 bomber engines churn the air into a 500-knot wind. Downstream, ducts guide the exhaust into a three-story tower, where it washes over the fuselage of an F-15, making the airplane behave as if it were flying. In fact, the fighter jet is a sitting duck for engineers firing incendiary rounds from Soviet-built MiG cannons at its fuel tank. The researchers hope the aircraft's fire-control system will detect the explosion and quench it within three milliseconds. More often than not, they dodge great balls of fire.

After all, not many chemicals can accomplish such a feat of firefighting. Fewer still are nontoxic, lightweight, highly compressible, noncorrosive, nonconductive, chemically stable, and leave no residue to foul jet engines. During the 1940s, when U.S. Army researchers found that the halogenated chemicals known as halons fit this bill, they understandably stopped looking for others.

Now the search is on again. Unfortunately, the very ingredients that make halons such effective firefighting agents—chlorine and bromine—also catalyze the destruction of stratospheric ozone. Last November at the fourth meeting of the Montreal Protocol members in Copenhagen, environmental ministers of 87 countries agreed to halt halon production by the end of 1993; Du Pont, the largest halon producer, has already dropped its line. As of next January, the 284 million pounds of halon currently filling an estimated 10 million fire extinguishers in airplanes, tanks, submarines and computer rooms will have to last forever.

Faced with an imminent deadline, the Department of Defense put the air force's Wright Laboratory in charge of finding and testing replacements. The military, now one year into a three-year, $16-million research program with the National Institute of Standards and Technology, does not expect to have prototype replacement systems before 1998. The two laboratories are screening a dozen chemicals, from which they hope to cull three for full-scale testing on Wright's gunnery range. But they are not likely to complete the task before the ban takes effect.

J. Michael Bennett, head of the halon replacement program, doubts that a simple, drop-in substitute will be found. What is more, he concedes that because most of Wright's 12 candidates are particularly bad greenhouse gases, likely to spend millennia in the upper atmosphere trapping solar radiation, none may be a long-term solution.

So Bennett's team is also looking at less traditional means of fire suppression. They have found that sodium bicarbonate (baking soda) puts out some engine fires quickly and safely when it is ground finely and mixed with a desiccant. And the air force is pursuing development contracts for two new types of extinguisher. One system detonates an explosive in water to produce a supersonic fog that quenches a fire by cooling it. In another—"the most promising replacement yet," Bennett admits—a solid propellant (the same one used to inflate air bags in automobiles) floods the fire with nitrogen or carbon dioxide, suffocating it.

Perhaps the best way to preserve the dwindling supply of halon is to reduce false alarms. Wright Lab is evaluating new detectors, produced by Donmar in Newport Beach, Calif., that employ robot-vision technology to distinguish a fire from a flash. Current photodetector-based systems discharge at the first sign of light, regardless of its source.

No one is yet willing to claim they have a solution, however, and the situation will likely worsen before it improves. To discourage hoarding, the U.S. increased the tax on halon to more than $40 per pound. The price of recycled halon, though tax-exempt, is also expected to rise sharply. Even so, the airlines and the military see a seller's market ahead. They loathe the idea of retrofitting planes to use a replacement—ever—but acknowledge they have no idea how much recycled halon will be available.

Plenty, says the Environmental Protection Agency. "Every airline owns in its computer reservation facilities more halon than it could possibly use over the next several decades," maintains Stephen O. Andersen of the EPA's stratospheric ozone division. Indeed, industry reports estimate that three out of every four halon systems were installed in computer rooms. As relatively cheap and expendable desktop computer networks have replaced expensive mainframes, the need for such systems has all but disappeared.

The military and the aviation industry have recently created banks of recycled halon in the hope of draining enough from these rich veins to tide them over until a replacement is found. But they do have a loophole when supplies dry up. The Montreal Protocol allows continued production of halon for "essential uses," such as in tanks, submarines and control rooms in nuclear power plants, when no alternatives—including recycled halon—exist.

But the path to an essential-use exemption is an option of last resort, deliberately fraught with bureaucratic pitfalls. In the meantime, halon users will likely line up at the halon banks, pray for the government to deliver a breakthrough replacement and wring their hands next spring while the Montreal Protocol signatories reconsider the wisdom of permitting the unrestricted use of recycled halon. —*W. Wayt Gibbs*

Making Wavelets

New math resurrects Brahms and compacts computer data

I am Doctor Brahms, Johannes Brahms." The voice, raspy but intelligible, comes from the speakers of Ronald R. Coifman's desktop computer. It is soon followed by an excruciating cacophony of hiss, pops, static and tones distorted to the limits of perception. "That was indeed Johannes Brahms playing his First Hungarian Dance," says Coifman, a professor of mathematics at Yale University. But the performance, recorded in 1889 on a wax cylinder donated by Thomas Edison (and since lost) and then rerecorded from a static-ridden radio broadcast of a 78-RPM record made from the cylinder, is utterly unlistenable.

Or was, at least, until Coifman used a powerful and increasingly important mathematical tool known as adapted waveform analysis to strip out selectively the random noise in the recording while preserving the highly structured music beneath. With a click of his mouse, Coifman plays the cleansed version. Although not all the notes are correct in time or pitch—the original wax cylinder had melted a bit by the time it was copied—the noise is nearly eliminated. Yale musicologist Jonathan Berger and graduate student Charles Nichols are now dissecting the recording to determine whether the syncopation in it is caused by distortion or whether Brahms had a particularly jazzy style at the keyboard.

Adapted waveform analysis has been catching on as a deluxe model of so-called wavelet analysis, itself a recent refinement of Fourier analysis. Whereas Fourier's technique uses a panoply of infinitely long sine and cosine waves to describe a sound or an image, wavelet analysis takes a single pulse of a particular shape—a wavelet—as its building block. Adapted waveform analysis draws from an entire library of pulse shapes. "It's just like a musical score," Coifman explains, "except that the notes are mathematical coefficients and the instruments are fundamental waveforms. The goal is to simplify the notation."

While adapted waveform analysis is gaining currency in the academic community, wavelets have already entered the commercial realm. Coifman founded Fast Mathematical Algorithms & Hardware, which is working with Martin Marietta to improve the ability of military radar systems to distinguish between, say, an ambulance and a tank. M. Victor Wickerhauser of Washington

SCIENTIFIC AMERICAN

CORRESPONDENCE

Reprints are available; to order, write Reprint Department, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 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, N.Y. 10017-1111, or fax (212) 355-0408.

Editorial correspondence should be addressed to The Editors, SCIENTIFIC AMERICAN, 415 Madison Avenue, New York, N.Y. 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, N.Y. 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.

University is developing techniques that use wavelets to pinpoint tumors in cloudy mammograms. Various groups are pursuing other medical applications for wavelet analysis, such as fetal heart monitoring, ultrasound image enhancement and electrocardiogram analysis.

But the big payoff for wavelets, if it comes, will likely be in data compression. Howard L. Resnikoff, a former Harvard University mathematics professor and co-founder of Thinking Machines, formed Aware in Cambridge, Mass., to capitalize on the ability of wavelets to condense information efficiently by representing it in terms of common patterns. Aware has put its wavelet compression techniques on a chip, and its sound and image compression software has been well received. "Our wavelet-based audio compression will be offered on all new Silicon Graphics workstations," Resnikoff beams, and he points to a recent government study that gives Aware's image compression top marks for quality over its competitors.

Wavelet compression seems particularly well suited for transmitted data because it spreads information around: images, for example, are divided into levels of detail rather than into rows and columns. Thus, like a hologram, each part of an image contains information about other parts. This makes communication over noisy channels more reliable, Resnikoff points out. "It also allows progressive transmission, gradually bringing an image into focus as more information arrives," adds Ingrid Daubechies of AT&T Bell Laboratories, one of the vanguard in wavelet mathematics. "This is useful when you are scanning through a library of images."

Resnikoff thinks that property may be lucrative as well as useful. Now that the Federal Communications Commission has allocated a portion of the spectrum for digital radio, Aware is positioning itself to provide the compression that must be used to make digital broadcasts economical. The company is also trying to capture the attention of the telephone companies that were recently awarded "video dial tone" rights by the government. Aware believes wavelet compression may be just what they need to force enough data through a pair of twisted copper wires to deliver movies to homes. Wavelets may yet make waves. *—W. Wayt Gibbs*

Micromachine Line

Industrial and academic researchers have wondered for years whether mechanical structures no bigger than healthy-sized bacteria would ever become more than laboratory playthings. They may now get a chance to find out. Beginning in late March an advanced semiconductor design and fabrication facility based in North Carolina began delivering its first batch of motors, resonators, grippers and mechanical drives to 15 separate clients ranging from a defense contractor to a high school class.

MCNC, formerly the Microelectronics Center of North Carolina, was set up 13 years ago with funds from the state as a nonprofit organization to develop advanced semiconductor technology. Last year the Defense Advanced Research Projects Agency, which was recently stripped of the word "Defense" in its title, decided to support the center in its micromechanics endeavor. MCNC is now open for business for a biannual production run.

Every May and December, MCNC will accept design specifications from each of its clients for one or more centimeter-size chips. The center makes the devices using the lithography and etching techniques borrowed from chip manufacturers. A thin wafer of silicon is "diced" into 50 chips, each of which contains dozens of minute components. Micromechanical elements have been considered as mechanical pumps that can deliver tiny doses of drugs, as sensing devices that can trigger an automobile air bag or as positioning elements for disk drives.

The center's micromechanics program is intended to encourage the kind of design creativity from small entrepreneurs that would be lost without access to expensive chip-manufacturing equipment and expertise. Larger companies can also use the facility to test prototypes. Besides manufacturing, the program will help clients design micromachines as well as combine them with electronic devices; it will also develop standard "libraries" of micromechanical components.

"For $500, you can play," says Karen W. Markus, an MCNC program manager. MCNC may become the first mail-order house for motors and grippers with dimensions measured in millionths of a meter. *—Gary Stix*

"Plastics, Benjamin…"

Photorefractive polymers make a big gain in data storage

Lugging around the *Encyclopaedia Britannica* might soon become as easy as carrying loose change. That is one promise of a new class of photorefractive polymers developed by investigators from the IBM Almaden Research Center. They recently announced that vast amounts of information can be stored in the form of two-dimensional holograms in inexpensive plastics.

The photorefractive effect was discovered in certain inorganic crystals more than 25 years ago. Unlike glass and other ordinary materials used for, say, lenses, the index of refraction of these substances changes when they are illuminated. That property suggested several applications, such as storing data, recovering distorted images and recognizing patterns. Yet despite two decades of work there are no widespread, practical uses. "The crystals are a pain in the neck," says Jack Feinberg, a physicist at the University of Southern California. "They are too expensive, too hard to grow, too variable."

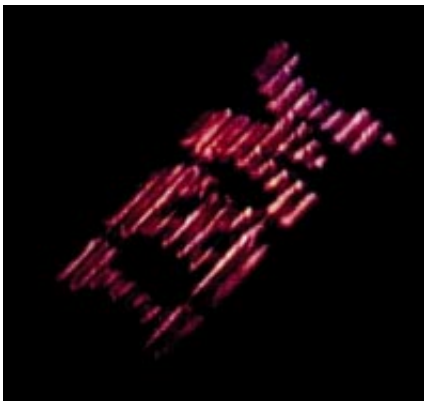
The problems of crystals led a few research groups to tackle an alternative—polymers. Polymers have several advantages: they can be made into many useful shapes, such as rods and thin films; they have electrical properties that are much easier to tailor; and they are much less expensive. The big break came in 1991, when W. E. Moerner and his IBM colleagues made the first photorefractive polymer and demonstrated that it could store data as holograms—one of the long-standing promises of photorefractivity. Since then, several researchers, including Jay S. Schildkraut of Eastman Kodak and Nasser Peyghambarian and his co-workers at the University of Arizona, have reported advances in creating other polymers that have photorefractive properties.

The photorefractive effect begins when two laser beams cross paths in the appropriate substance. The interference produces the classic pattern of alternating dark and light areas. Electrons from donor atoms in the dark areas travel to the light areas, where they are trapped by acceptor atoms. The migration of electrons produces a strong internal electric field that alters the structure and hence the optical properties of the crystal.

The compounds can store information because the changes in the material follow the spatially varying light-and-dark pattern established by the interfering beams. A weak, uniform beam sent through the substance reads the data. The process is similar to that used to form ordinary holograms, in which two laser beams produce an interference pattern that is recorded permanently on film and is read later by illumination.

But holograms in photorefractive substances have an important edge. Because the optical properties result from electrical rather than chemical changes, they are erasable. A strong enough, uniform beam will move the electrons back to their original locations and ready the material for another round of writing.

The early holograms produced in photorefractive polymers were not especially bright. But at the American Physical Society meeting held in Seattle this past March, Moerner announced that the team has created a new photorefractive polymer based on a compound called poly(*N*-vinylcarbazole). Properly doped, this material yields holograms that are 100 times brighter than the earlier ones—and that rival those formed in photorefractive crystals.



HOLOGRAM in a photorefractive polymer stores a pattern of bits laid out in the shape of a familiar corporate logo.

The polymers are also the first to show "net gain," which results from a property of many photorefractive materials referred to as beam coupling: when two laser beams cross, one beam can steal light intensity from the other. The robber beam—as well as any information imposed on it—becomes amplified. Net gain results when the coupling gain exceeds the unavoidable losses that occur when light passes through the material. "That was the difficult and more important breakthrough," Moerner says. "That's when the material starts getting useful."

Made thick, the IBM polymers can store several holograms in the same volume; one simply changes the angle of the laser beam. Moerner and his colleague Scott M. Silence were able to record five complex holograms in a disk one millimeter wide and 125 microns thick. A practical holographic storage system, extrapolated from this achievement, would be able to record several billion bits in about 225 cubic millimeters. In hard-copy units, that is the entire *Encyclopaedia Britannica* in a penny. The density is not quite so high as some other optical data storage techniques, such as those that rely on fine-needle tips to read and write information. But what makes the holographic storage exciting are its access time and the speed with which it could process data.

Indeed, random-access time for the polymer memories could be incredibly swift—from 100 to 1,000 times faster than the retrieval time from magnetic media. Unlike disk drives, which must rotate to the disk to read the information, holographic storage would rely only on the changing trajectories of moving photons. Because photons are massless, the devices would not be slowed by inertial delay, as are magnetic drives. Current beam deflectors take only a microsecond to divert light. Because the holograms are two-dimensional arrays of data, "all the bits come out at the same time," Moerner explains.

Don't throw away your hard drives just yet; there are several hurdles facing workers. One potential limitation of all hologram schemes is the effect of the read beam—each read causes the hologram to fade slightly. To combat fading, Moerner points out that the computer system would have to spend some time refreshing the data. Storage time is also an open question. No one really knows how long a polymer can retain information.

The polymers have a long way to go to demonstrate truly wide-ranging photorefractive properties. "Crystals can still do things polymers can't," Feinberg notes. "It's the difference between a Volkswagen and an airplane." The main problem is size. The IBM polymers were about an order of magnitude too thin to generate the total gain of the much larger photorefractive crystals. Furthermore, polymers need an external electric field, called a poling field, in order to function.

Still, "photorefractive polymers are rather new," Peyghambarian remarks. "There are going to be more groups working in this area," and that may lead to applications soon. Indeed, Moerner believes within five years something should come of photorefractive polymers. Perhaps the future really is in plastics. —*Philip Yam*

THE ANALYTICAL ECONOMIST

Are Economy Watchers Chasing a Mirage?

It was the best of times, it was the worst of times. Earlier this year the nation's economic pulse takers announced that the U.S. gross national product had chalked up yet another quarter of solid growth. Yet at the same time, more of its citizens received food stamps than ever before. A naive observer might well have been confused.

Do widely quoted indicators such as GNP, unemployment rate and consumer price index actually say something about the average person's financial well-being? If not, on what statistics should policymakers be basing their decisions? Some economists—Orley C. Ashenfelter of Princeton University, for one—consider the question "ancient" and of little import; others, namely Amartya Sen of Harvard University, think the U.S. should take its eyes off money-based numbers and focus instead on such measures as mortality, illiteracy, crime and homelessness.

No one really disputes the notion that gross economic statistics are inaccurate and potentially misleading, says Christina D. Romer of the University of California at Berkeley. But as long as they continue to be inaccurate in the same way, she explains, that is almost as good as being right.

For example, people have complained for decades that the unemployment rate does not include those discouraged workers who have given up looking for jobs, thus understating the number actually out of work. Nevertheless, current figures can still usefully be compared with past data "unless you have a story where this recession has more discouraged workers than usual," she says. (As it turns out, there may be just such a story. Romer notes that the present expansion is remarkable for being a largely "jobless" recovery, in which companies increase output by raising productivity rather than hiring more workers.)

Over the course of a few months or even a year, maintains Victor R. Fuchs of Stanford University, the standard economic aggregates are "all perfectly good." The numbers of discouraged workers—or, rather, the buying patterns that underpin the consumer price index or the industrial weighting factors that help to determine the official GNP—do not change that quickly. But over the course of a generation or so, he asserts, changes in the way the nation's business is structured may make past and present data incompatible.

One prime example of such incompatibility is the data used to measure overall U.S. productivity. Statisticians adjust them for inflation by normalizing the price of various inputs (raw materials, energy, labor and so on) to a base year, in this case 1987. Although some kind of adjustment is necessary, points out Robert J. Gordon of Northwestern University, the normalization process badly distorts the numbers because while the price of most things has gone up, the price of some, such as computers and electronic equipment, has plummeted. Manufacturing productivity, he contends, has been understated by about one percentage point for each year before 1987 (when computers cost far more) and overstated by the same amount for years since then (when computers cost significantly less).

As long as statistics are wrong consistently, that's almost as good as being right.

In addition to distorting indicia, however, the changing structure of the economy can make certain kinds of statistics fundamentally misleading. "From World War II through the 1970s," comments Greg J. Duncan of the University of Michigan, "average income did justice to both the bottom and the top of the distribution." When average income rose, it was a sign that both the poor and the rich were doing better; when it fell, the rich and the poor alike were doing worse.

But in the 1980s, Duncan observes, inequality increased, and so income statistics reflected the experience of a dwindling proportion of citizens. The fact that median income stagnated gave little sign that the top fifth of the population enjoyed modest but significant increases in income while the bottom fifth suffered a sharp drop. "The median misses everything," he says.

Paying more attention to the distribution of income would be easier if the numbers were more readily available, Duncan argues. But whereas "standard" indicators such as GNP, consumer price index and unemployment come out on a regular, timely basis, figures on income distribution or the number of people receiving public assistance are calculated less often and issued by different agencies after delays that range from months to nearly a year.

Gordon suggests that the U.S. might consider consolidating all its numbers operations in a central statistics office, as do Canada, France and some other nations. In a recent survey, he notes, economists ranked the quality of Canadian information gathering highest in the industrialized world and that of the U.S. seventh.

The dissatisfaction with the quality and scope of currently available numbers, however, does not fully address the question of whether any set of purely monetary figures will be adequate for judging economic health. Fuchs cites his boyhood in a middle-class neighborhood in the Bronx during the 1930s: "A few years ago I took my children and grandchildren back to where I had lived, and it was horrible—stores boarded up, buildings defaced." Yet in terms of inflation-adjusted income, today's Bronx residents are better off. It is factors such as health, vulnerability to crime and quality of education that make the difference, Fuchs asserts.

Indeed, in recent years the World Bank, set up to help poorer nations become more like the industrialized world, has mostly abandoned its steadfast reliance on financial growth as the goal of development. The World Bank "now sees its objective as poverty reduction," states Martin Ravallion, a principal economist in the bank's policy research department. "We have explicitly rejected GNP per capita" as a measure of progress, he says.

What measures is the World Bank using? According to Ravallion, development planners are looking at "what people can actually do and be"—whether they are properly fed, clothed and housed, whether they can read and have access to medical care—rather than how much money the national accounts declare they have. It is perhaps paradoxical that economists are now saying that "money isn't everything," but the results of their new focus may be instructive.

—Paul Wallich

A Bundling Fool Beats the Wrap

Silas Golding was finishing his first day of work as a tennis-ball packer for the Sportsware Shipping Shop. "Billy Jo, could you pass me another seven cans?" he asked. "And while you're at it, how about some shrink-wrap film?"

"Just let me wrap these basketballs," Billy Jo Rottweiler said. After cutting a piece of film, she handed him seven tennis-ball cans and the roll of film.

Silas covered his workbench with a long sheet of film, placed the cans on top, side by side, and then folded the sheet over, wrapping the cans as carefully as a parent diapers a baby. Then he put the entire package in a machine, which shrunk the film around the cans. The plastic fitted snugly around the edges of the bundle, but some spaces remained in the middle. Silas casually picked up the package, and its contents spilled onto the floor. The crash startled everyone in the shop.

"Golding! You've done that a hundred times today," the supervisor yelled.

"Sorry, sir."

"Would you like to work here tomorrow, you uncoordinated sloth?"

"Yes, sir."

"Then I suggest you find a way to wrap the cans properly. Why not try binding them up in a more compact bundle?"

"But Billy Jo packs her basketballs in a line—like a string of sausages," Silas protested. "And she stretches the wrapping tightly just as I do, so that there aren't any dents or anything. That is, the packs are always convex in shape."

"Yeah, but my packs don't fall apart," Billy Jo interjected.

"So, why doesn't it work with cylindrical cans?" the supervisor asked.

"Well, sir," Billy Jo continued, "I can see that if the total volume of the pack—including any air spaces—is as small as possible, then the pack won't collapse. Because if it did, you could shrink it even further, and you can't, so to speak. So we've got to think about packs whose volume—including everything inside the wrapper, air and all—is as small as possible but is still convex."

"I can see that the surface area of the bundle is a lot bigger when the cans are arranged in a line," the supervisor said. "Of course, that's determined by how long the strip of wrapping film must be. But I'm not so sure about the volume."

"Suppose," said Billy Jo, "the cans have a radius of length 1, to keep the calculation simple. If you look down on the top, then in essence you're just packing a lot of circles together and wrapping an elastic band round them. The volume of your pack is proportional to the area inside the band; the surface area of your wrapping is proportional to the length of the band, assuming you don't cover the ends when you wrap the cans up."

"Right."

"So we can find the area inside the band and the length of the band. It's a lot easier to think in two dimensions. Now, if you put seven circles in a line and wrap a band round to get a tight fit with convex edges, the total perimeter is $2\pi + 24 = 30.283$."

"Why?"

"Each straight side has length 12—six diameters of the circles—and then there are two semicircles, each of length π."

"Okay."

"But if you arrange six circles in a hexagon and place one in the center, the perimeter becomes $12 + 2\pi = 18.283$. It's a similar argument: there are six straight sides whose lengths are equal to a diameter and six arcs, each making up one sixth of a circle."

"I can see that, Rottweiler."

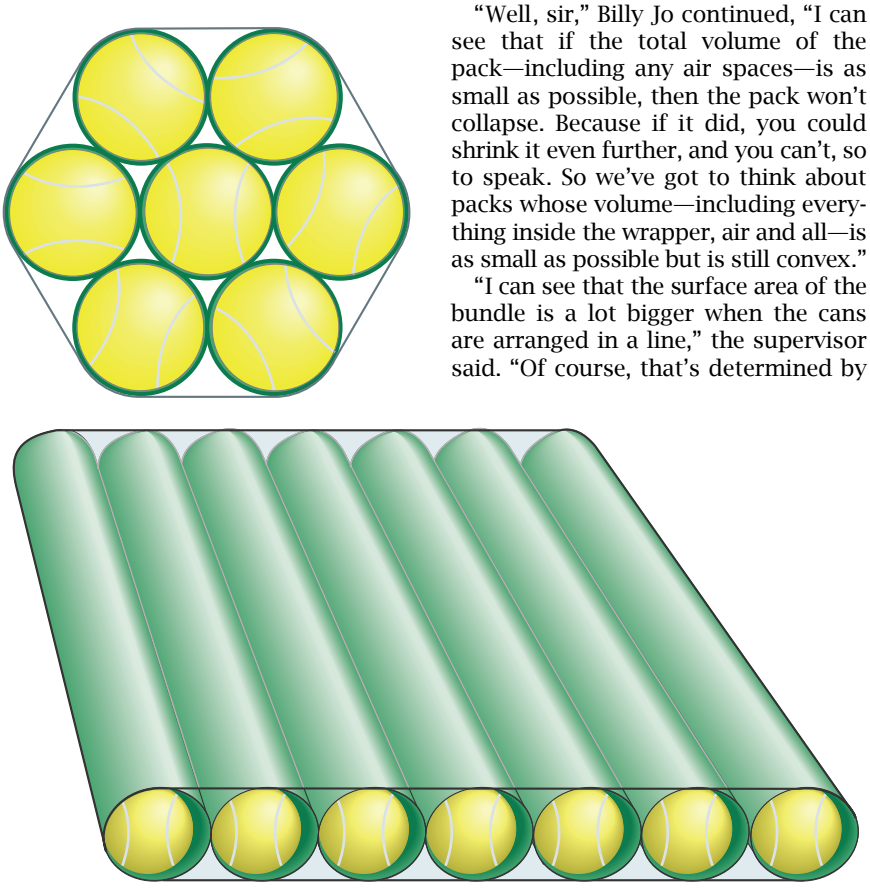
"So the perimeter for the hexagonal arrangement is only a bit more than half the length of the sausage. That's why Silas's packs keep collapsing: if the cans are moved a little bit, the wrapping gets loose."

"But all you've done is calculate the perimeter, Rottweiler. What about the volumes?"

"I should be able to use the same kind of method, sir. If you arrange the cans in a sausage, the total area inside the bundle can be split up as six squares with sides of length 2 plus two semicircles with radii of length 1; so that's $24 + \pi = 27.141$. On the other hand, if you arrange the cans in a hexagonal group, it's kind of tricky to work out the volume. There are lots of funny bits with curved edges."

"Isn't there a formula for things like that, Rottweiler?"

"Not in any textbooks I know of. I'll have to figure it out." She grabbed a pencil and some scrap paper. "Hmm.



SEVEN CYLINDERS can be wrapped together in a bundle or a row. Which of these packages has the smaller volume, and which is less likely to fall apart?

Seems to me I've got to work out the areas of two kinds of shape—apart from circles. There are 'center holes,' with three curved sides, and 'edge holes,' with two curved sides and one straight [*see illustration at right*]. A central hole is really an equilateral triangle minus three 60-degree sectors of a circle. The equilateral triangle has sides of length 2; therefore, its height is $\sqrt{3}$, and its area is half the base times the height, which is $(1/2) \times 2 \times \sqrt{3} = \sqrt{3}$. The three sectors together form a semicircle of area $\pi/2$. So a central hole has an area $\sqrt{3} - \pi/2 = 0.161$. Similarly, two edge holes plus two semicircles make a 2 by 2 square with area 4. Consequently, one edge hole plus one semicircle has an area half that, namely, 2. So an edge hole has an area $2 - \pi/2 = 0.429$.

"Finally, a hexagonal bundle of seven circles has an area made up from seven circles, six central holes and six edge holes. That comes to

$$7\pi + 6(\sqrt{3} - \pi/2) + 6(2 - \pi/2) = 25.533.$$

Which is smaller than 27.141. See? That's why Silas's sausages collapse. He should make hexagons instead."

"Fine," Silas said. "But I don't see why it doesn't work the same way with spheres. Surely if you pack your basketballs into a tight group, the total volume, wrapping and all, will be smaller than if they're arranged in a line."

"Not necessarily. It's a complicated problem, and it's genuinely three-dimensional. I don't see how to arrange seven spheres in a tight group without there being an awful lot of wasted space inside the wrapper. For my arrangement—that is, with all the balls in a line—the area of the wrapper is 28π, and the volume inside it is $40\pi/3$. But it's not so easy to work out the math for other arrangements. Still, my packs don't fall apart. So that's experimental evidence I'm right."

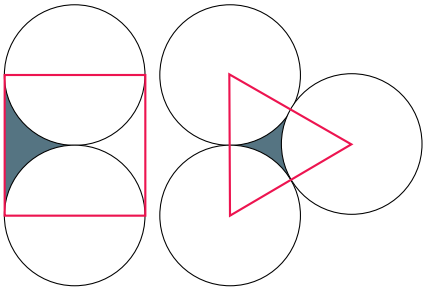
"That is all very well," the supervisor declared, "but you often have to tie bundles with a lot more than seven cans in them. What shape will those have to be to minimize the volume inside the packaging, Golding?"

"Sort of—like the hexagon, sir, but bigger," Silas said.

"Like the hexagon, sir, but bigger," the supervisor whined, mimicking Silas. "You call that a good answer, Golding?"

"No, sir."

"And as for you, Rottweiler, you don't seem to have a handle on the spheres problem. You can calculate the volume only for the sausage shape, and then you assume it has the smallest volume."



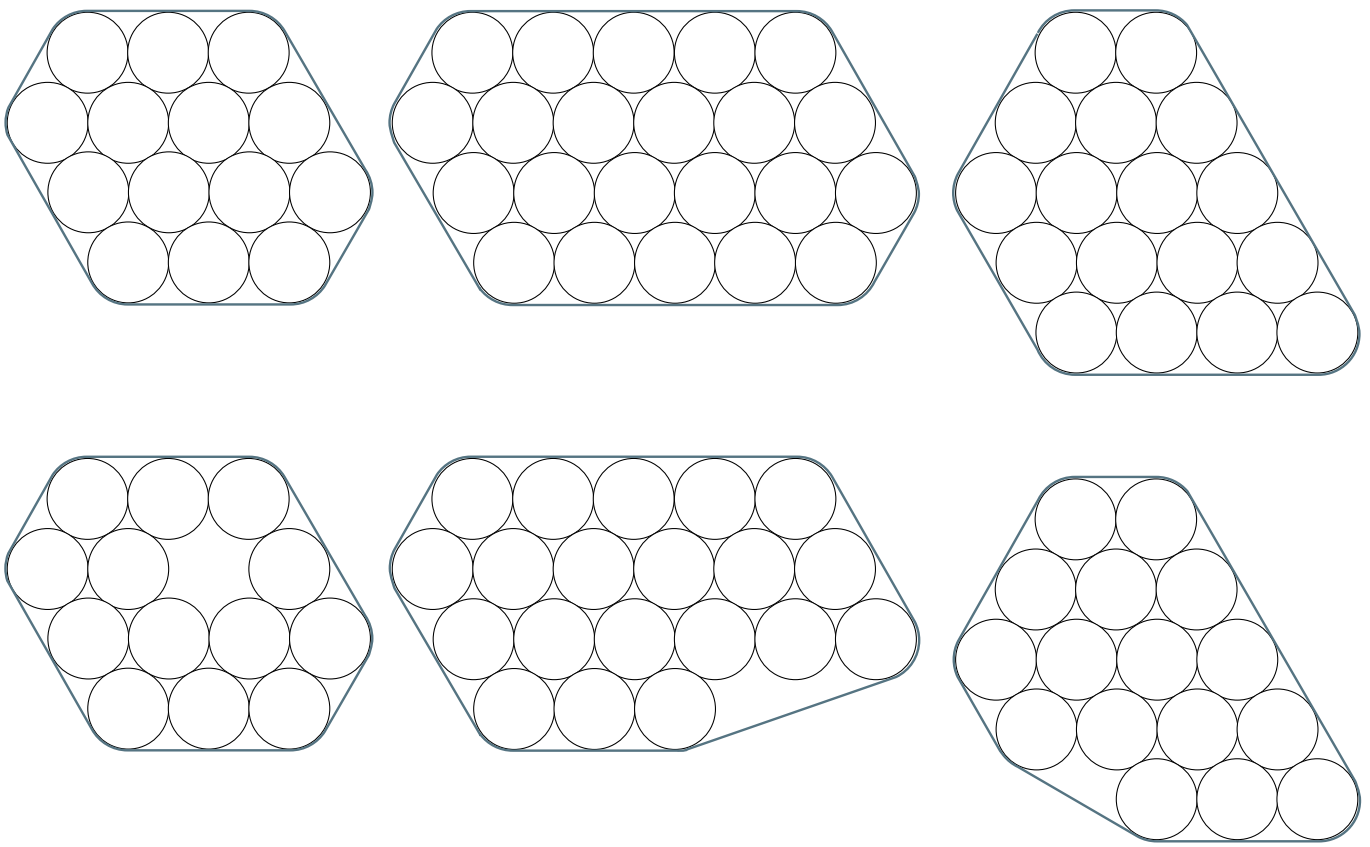
***THREE CIRCLES** packed one next to the other leave a space that is equal to the area of the triangle minus half the area of one of the circles. On the other hand, the space between two circles and an edge is half the area of the square minus half the area of one of the circles.*

She hung her head apologetically.

"I want you both to spend some time in the company library, digging up any useful information you can. I'm sure somebody must have studied the problem before. No point in reinventing the wheel, is there?"

"Yes, sir," Billy Jo and Silas chimed in unison. "We mean, no, sir."

The next morning the supervisor found Billy Jo and Silas asleep in the



***GROEMER PACKINGS** (top row) enclose the circles in the smallest possible area. The bundles in the bottom row are not Groemer packings, and in each of those cases, the circles can be packed into a smaller region.*

library, their heads resting on open books. "Good morning. Find anything, you two?"

"Yes, sir. Quite a lot, actually, sir," Billy Jo said, waking abruptly.

"Well, then."

"Sir, the basic problem is to find arrangements of n-dimensional spheres in n-dimensional space that minimize the n-dimensional 'volume' of their convex hull. Er, the convex hull is the smallest convex surface that contains all the spheres. In two dimensions, we're packing circles in the plane and trying to minimize the area of the convex hull; that's like packing Silas's cylindrical cans. In three dimensions, we're packing spheres, like my basketballs, and trying to minimize the volume."

"Go on."

"In two dimensions, the best arrangements are known to be 'as hexagonal as possible.'"

"I thought you'd found a precise answer, Rottweiler."

"Uh, well, sir, it's a bit tricky. I have to explain about Groemer packings. They're honeycomblike packings that form hexagons—not necessarily regular hexagons, but the sides have to be parallel in pairs. Oh, and some 'sides' may not occur at all—I mean, you could have pentagons and triangles and things like that [*see bottom illustration on preceding page*]. They're named after Helmut Groemer of Oregon State University, who proved a fundamental theorem about them in 1960. It relates the area inside the convex hull of a circle-packing to the number of circles inside it and its perimeter. Namely, for n circles and perimeter p, the total area of the convex hull is at least

$$2\sqrt{3}(n-1) + p(1 - \sqrt{3}/2) + \pi(\sqrt{3} - 1).$$

Moreover, the area is equal to this expression if and only if the packing is a Groemer packing."

"And?"

"Starting from this estimate, Gerd Wegner of Dortmund University pretty much solved the entire problem in 1986. He showed that the convex hull of a set of n equal circles is minimal when they are arranged in a Groemer packing whose sides are as close to being equal as possible—at least provided n is less than or equal to 120. The theorem is also true if n is a 'hexagonal number' of the form $3k^2 + 3k + 1$, when there is a Groemer packing forming a perfect regular hexagon."

"But there are still some unsolved cases?"

"Yes, but then the gap between the best-known results and the best that possibly could be achieved is very small."

The supervisor shook his head. "Not as precise as I had hoped, Rottweiler."

"No, sir," Golding said, leaping to her assistance. "But we never have to wrap more than 120 cans in one pack."

"True, but irrelevant to the theoretical problem," the supervisor replied.

"The strange thing is that in three dimensions, the answer is quite different—at least, for a smallish number of balls."

"A smallish number?" the supervisor asked quizzically.

"Anything up to and including 56, sir," Billy Jo added hastily. "For that many spheres, the arrangement whose convex hull has the smallest volume is a sausage—all the balls in a straight line. After that, though, the minimal arrangements get much more compact and rotund."

"Rotund, you say?"

"Less intuitive still is what happens in spaces of four or more dimensions," Rottweiler explained, ignoring her boss's banter.

"I don't think we make four-dimensional sports gear," the supervisor quipped.

"In any case, sir, as you just said, it's the theoretical point that matters."

"True. But what do you mean by four-dimensional space?"

"Well, two-dimensional space can be defined using two numbers as coordinates, (x,y), and three-dimensional space is every set of three numbers (x,y,z). Clearly, four-dimensional space is every set of four numbers (w,x,y,z), and n-dimensional space is every set of n numbers $(x_1,\ldots,x_n)$. That's it, really."

"And a sphere in n dimensions?"

"That's the set of points that lie within a distance of, say, one unit from a chosen point, the center. And of course you have to define the n-dimensional analogue of volume, but that's not too hard."

"I see."

"Anyway, for four dimensions, the arrangement of four-dimensional spheres whose convex hull has the least volume is a sausage for any number of balls up to about 50,000. It's not a sausage for 100,000 balls. So the best packing involves very long, thin strings of balls until you get an awful lot of them."

"How many exactly?"

"Somewhere between 50,000 and 100,000, sir. Nobody knows the precise value at which sausages cease to be the best. But the really fascinating change comes at five dimensions. At least, that's the conjecture. You might imagine that in five dimensions sausages are best for, say, up to 50 billion balls, but then something more rotund has a convex hull of smaller volume; for six dimensions, the same kind of thing holds up to 29 zillion balls, and so on. But in 1975 Laszlo Fejes Tóth formulated what is now called the sausage conjecture. It states that for five or more dimensions, the arrangement of balls whose convex hull has minimal volume is always a sausage, however large the number of balls may be."

"Good heavens," the supervisor exclaimed, genuinely surprised.

"What's so special about five dimensions?" Silas asked.

"Do bear in mind it's only a conjecture," Billy Jo pointed out. "But the heuristic evidence is quite strong. The basic idea is that as the number of space dimensions increases, it gets harder and harder to fill space efficiently with spheres. The gaps are multidimensional and become quite big. If you make the wrapping poke out along too many dimensions of the space, you leave a lot of gaps, which makes the volume quite large. Whereas if you confine the spheres to a single direction, in a sausage, you cut down the wasted volume so much that it all works. Tóth estimated the dimension at which the numbers for this argument work out, and the breaking point seems to be dimension five.

"It could all be rubbish, of course—nobody knows for certain. But it sure would be interesting to find out."

The supervisor tried hard not to look impressed. "Very well, you two," he said. "Good work. But now—back to your workbenches. You've got a lot of packs to wrap."

Silas spent the next hour wrapping 56 basketballs in a single sheet of plastic film. He was so proud of his accomplishment that he called over the supervisor to take a look.

"Pretty amazing," the supervisor cheered.

Silas carefully lifted the package over his head in triumph, but the package bent and broke from its own weight. Fifty-six basketballs bounced around the room.

FURTHER READING

RESEARCH PROBLEMS. L. Fejes Tóth in *Periodica Mathematica Hungarica*, Vol. 6, pages 197–199; 1975.

ÜBER ENDLICHE KREISPACKUNGEN IN DER EBENE. Gerd Wegner in *Studia Scientiarum Mathematicarum Hungarica*, Vol. 21, pages 1–28; 1986.

UNSOLVED PROBLEMS IN GEOMETRY. Hallard T. Croft, Kenneth J. Falconer and Richard K. Guy. Springer-Verlag, 1991.

BOOK REVIEWS by Philip Morrison

A Brief for Science

DREAMS OF A FINAL THEORY, by Steven Weinberg. Pantheon Books, 1992 ($25).

This confident account of a grand intellectual aspiration of our times is the work of a celebrated theorist who is as accomplished as any—articulate, cultivated—and more candid than most. He writes explicitly for the general reader as "a working physicist," who here uses not so much as one equal sign. His gaze centers on large issues from philosophy and the history of ideas, both the topical and the enduring. In a way, he has filed a sophisticated if sometimes impatient brief for science against philosophy, against sociology and—not without a certain tristesse—against religion as well.

"The word 'why' is notoriously slippery," he concedes. Yet the book opens with a few chapters around the nature and growing importance of whys within physics. To explain to a child that an apple falls to earth "because of gravity" is to say no more than that the event is common enough to elicit a learned word. But the account now given by gravitational physics is rich beyond mere certification. We see why a toy balloon also *rises* because of gravity, why grasshoppers and horses alike jump a few feet high, why the tides ebb and flow under the moon, why asteroids are chunky but Jupiter round and (perhaps) why our cosmos must be filled by matter yet unseen.

It is that unbroken, convergent, lengthy chain of connection that singles out the explanatory principles of physics. Reductionism, the claim that there is a hierarchy of sciences, deserves its two cheers. Some sciences hold priority, though not mainly because they generate wholes from smaller parts. Einstein's space-time did not arise from the study of parts but from new principles of profound universality. Yet three cheers for the reductionist stance would be too many, for even within physics we know of principles, such as the laws of thermodynamics, that govern change whatever the inner nature of the system. But their use is enriched by microscopic understanding and follows from certain grand symmetries of the world of particles and fields. Of course, all regularities, say, of macroeconomics, are not easily implied by particle physics, but they cannot contradict the bounds physics sets on human behavior, as the subtleties of DNA cannot contradict the quantum nature of molecular structures.

Chaos and chemistry alike are not in practice reducible to physics alone; emergent insights may be novel, but what they cannot be is autonomous. The aim of science is to connect. "The reason we give the impression that we think that particle physics is more fundamental than other branches…is because it is." Its ubiquity brings it closer to the "point of convergence of all our arrows of explanation." Let no congressperson rejoice; this is no basis for allocating research funds. Fundamental physics does not claim a first right to the public purse, but neither can its character be justly ignored.

Quantum theory, the subtle and diverse feedback loops that join experiment to theory, and even the aesthetic criterion for theories are examined closely, in a manner that will help thoughtful nonmathematical readers. There is a fascinating chronicle of Weinberg's own experience with an experimental contradiction to what we now know as the standard electroweak theory. It seemed at first that the theory could be made "just a little bit uglier and get it to agree." They tried hard; nothing worked. The naturalness, the formal inevitability, of the theory, was firmly built in. In the event, the conflicting experiments were repeated better elsewhere, and the disagreement went away. By now a tide of diverse experimental results has shown just what was forecast by the natural theory. The relation between theory and experiment is more art than uniformity; wishful fit is always a danger, but so is a premature contradiction.

The metaphysics of Descartes made good sense; how could forces be transmitted without some fluid medium? Newton's rash violation of that good sense was quantitative, hence greatly fruitful. By now it has been properly embedded within a far richer model of space and time than the 17th-century philosophers could dream up. Positivism won young Einstein, but soon enough he came to hold that "every theory…contains unobservable quantities." Philosophy is help and hindrance alike, though at least the philosophers mean well by science.

Author Steven Weinberg

A more hostile attack has lately come from an influential school of sociologists who assert that the results of physics are constructed by mere social convention. Of course, science is a social process, negotiated, historical and contingent. But something important works really well, or it is laid aside. Quarks or their equivalents are there within, not constructed any more than were the moons of Jupiter or the forces of gravity. Mountaineers, too, volubly negotiate equipment, route and schedule. Some teams succeed; some fail. Yet would anyone "give a book about mountain climbing the title *Constructing Everest*?"

The form of a final theory of forces and matter and change can be outlined from shadows we now see on the cave wall. "A fair fraction of today's young theoretical physicists are working on string theory." It is subtle and abstract beyond any theory of the past. Strings are tiny, one-dimensional rips in smooth space, enormously rich in energy, vibratory but without internal substructure. The promise comes from ability to parallel the broad picture of our more workaday views by intrinsic vibrations and, above all, from the hints of representing weak gravitation among the other forces—provided a choice is made that fixes both the weakness of gravity and the unprecedented mass-energy of the string.

One deep symmetry is essential to admit quantum behavior, and it can possess a large number of less general symmetries, suited to describe all the unchanging quantities (charge, energy and so on) that we know and many more. Enthusiasm was higher a few years ago when there seemed to be few such formulations; now we face thousands of candidate theories. Alike they understandably share freedom from the problem of today's particle theories, the infinities that arise from requiring that physical particles have no extension in space, for strings are not points.

How could any theory be final? "In my view, our best hope…is to show that the final theory…is logically isolated." That is, its requirements are so rigid that any significant change in form generates absurdities. Weinberg does not write idly; he himself tried cleverly to modify quantum mechanics to avoid certain much criticized features, to find that even small changes destroyed the tested inner consistency. The experience has marked him. If indeed we gain a final theory of all particles, so that we know why there are quarks and leptons, why there are three families of them, why all forces are of common mathematical form, why change conserves some properties and not others, science will certainly not be over. It is much broader than that. The rational will have made a real gain, but something old and mighty will have faded. Scientists of the future will envy us a little, that quest still before us.

Down in the rolling lands of Ellis County, Tex., we are building among the cottonwoods an ambitious particle accelerator. It has a good chance, not at all of leading us to those remote strings, but of testing the most likely extensions of theory beyond our patently incomplete but otherwise unflawed "standard model." The billion dollars a year is debated again and again, never on grounds of finding the uncertain Higgs boson but for economic advantages to one or another. Federal dollars, as we read daily, are more freely spent on aims far less noble. Yet we are upon hard times, and the Superconducting Super Collider remains uncertain beyond this year. Watching the drill pierce the chalk of Ellis County, Weinberg writes: "Perhaps I was under the spell of…Victorian optimism, but I could not believe that…in our time the search for the final laws of nature would be abandoned."

A rare chapter on religion is courageous, even generous. It ends with the eloquent rhetoric of the Venerable Bede, who compared our short life on the earth to that of the sparrow that flies by chance from wintry night into the King's bright banquet hall for fleeting comfort, then soon out again. "It is an almost irresistible temptation to believe with Bede…that there must be something for us outside the banqueting hall. The honor of resisting this temptation is only a thin substitute for the consolations of religion, but it is not entirely without satisfactions of its own."

This splendid book is as good reading about physics and physicists as this reviewer can name. Although its arguments could hardly prevail at every point of the long ramparts they defend, they are everywhere clear, honest and brilliantly instructive.

Our Wild Cousins

CHIMPANZEE MATERIAL CULTURE: IMPLICATIONS FOR HUMAN EVOLUTION, by W. C. McGrew. Cambridge University Press, 1992 ($79.95; paperbound, $27.95).

Among all living species the nearest of kin to humans is the chimpanzee. When we seek the origins of our humanity, those wild cousins, who match us up to all but a small fraction of the entire genetic message, have much to teach; their observable ways offer a touchstone for the shifting, shaky theories of the protohuman past. The author has been drawn by the chimps for 20 years, four of those years spent in the field in Africa. McGrew is an engaging, candid naturalist, no prehistorian but rather "imprinted…in the natural and not the social sciences," analytic, manifestly "committed to data" as the test of theory. While he freely cites the remarkable achievements of captive apes, those domesticated creatures remain auxiliary to what happens "in the real world."

It was in 1960 that Jane Goodall first found tool use and toolmaking in a natural population of nonhuman primates. Goodall's big pool of insightful observations from the Gombe forest in Tanzania is now 30 years deep. Yet only 10 other field studies of chimpanzees have persisted at least once around the seasons. There is a major ecological point here. We humans live over a wide spectrum of climates. Some "classic" hunter-gatherers such as the Kalahari San live in arid lands, places with 15 or 20 inches of seasonal rainfall a year, whereas African pygmies hunt and gather in rain forest with 60 splashy inches. The savanna dwellers of Tanzania's Olduvai Gorge 2.2 million years back enjoyed some 30 inches of seasonal annual rain.

It is something of a surprise to learn that chimpanzees, too, manage well in a wide variety of climates, from the rain forest of the Guinea coast with 10 feet of rain a year, to the Zaire woodlands, where moderate rainfall matches that in Olduvai long ago. The familiar argument that apes stayed arboreal in the dense forests while our forebears began to walk the open lands of the savanna gets no support from these chimpanzee facts. Neither chimpanzee nor human foragers have remained in the climate of their origin; desert and rain forest are probably rather novel settings for us both.

The most important goal of chimpanzee tool use is food; their most elaborate tool kit is one for taking honey and larvae from the wild beehive, a full meal worth some stings. Chimps in the Gambia rob a hive in a hollow tree with a set of four special tools, all made on the spot of sticks or twigs: a stout chisel to start an opening in the thick, outer seal, a finer chisel to enlarge it, a sharp bodkin to puncture the nest wall within and a long flexible dipstick to draw out the sweet reward. Such oxidizable, organic artifacts will not leave much of a mark after a million years.

One picture shows a wild chimpanzee wielding hammer and anvil made of stone, both of them found objects, to crack open hard-shelled oil palm nuts. The cracking kit used by humans nearby looks very much the same. Other chimpanzee populations live among plentiful oil palm crops yet never eat that rich food. Is that a sign of culture differences, as it would be interpreted among human populations?

Chimpanzees make and employ tools of plant material or unworked stone for much hard work, crushing roots, smashing bone. But "only humans use guided missiles…set untended…snare traps…use poisons…[or] dogs as hunting companions." No chim-

panzee tool makes use of a sharp edge, nor is one much needed, for the prey the chimps take are too small to demand real butchery. That inventory is securely hominid. Containers would be of the greatest value to these ape foragers of small food items, such as nuts or larvae. Chimpanzees in captivity readily use containers; a chimp depicted here is ably filling an empty coconut shell with water from a leaky tap. Yet no container has been seen with chimpanzees in the wild. Maybe the chimpanzees in the forests of West Africa, where nut cracking is essential to the diet, do use containers. "In the right… circumstances," chimps should be capable of making Oldowan-like tools and more, even to variations of style.

Does Promethean fire part us sharply from the apes? In captivity, both chimpanzees and orangutans can tend and handle fire. (Some of those chimps were introduced to the bad habit of smoking.) We need to watch the chimps of the savannas while bushfires sweep across their range. We need to test captive chimpanzees for spontaneous uses of fire. On cold nights would they sleep near fires? Would they roast tough pods, stoke a fire when the audiotaped roar of the lion was heard, fan dying embers or light one fire from another?

This book is as fascinating for its methodology and its questions as for the remarkable observations it gathers. Overall it strengthens, in spite of itself, the popular bias of our times: hominid culture grew not from routine tool use but out of new software abilities based on new neural hardware. It is the combinatorial richness of blossoming language that will be hardest to sift out of the ash layers of the Rift.

The Solar System and Beyond

THE PLANETS: A GUIDED TOUR OF OUR SOLAR SYSTEM THROUGH THE EYES OF AMERICA'S SPACE PROBES, by Nigel Henbest. Viking, 1992 ($35). **THE ANDROMEDA GALAXY,** by Paul Hodge. Kluwer Academic Publishers, 1992 ($79).

Our planet is only one among a short list, our galaxy one among a myriad of peers. We can grasp our own state better by understanding our neighbors, too. Both these books carry out the Copernican task at two wholly different levels of exposition and on sharply differing scales of physical size.

The Planets, by a British science writer of proven talent, is a deft work. Its big pages offer us a brief, entirely nontechnical text, absent all graphs (and of course without equations). What most catches the eye is plain in the subtitle: 100 colorful photographs (50 more in black-and-white), radioed pixel by pixel from many distant "hardy unmanned craft" over the years, *Mariners* to *Giotto.*

Some of the images are classics, such as the view of the rock-paved surface of Venus or of the sulfurous yellow-orange globe Io. Many images are fresh and wonderful: a Soviet shot of tiny, rocky Phobos dark against the very edge of its red planet, the underside of Saturn's rings, the subcontinent of India overspread by the wide, dark stain of the Deccan basalt. Captions and textual explanations are as simply put as they are telling. The present sterility of Mars, for instance, is made understandable; the lack of oxygen and hence of an ozone shield against solar ultraviolet has made its mineral surface into a fierce oxidant, corrosive to all organic compounds.

Our planetary neighborhood is laughably provincial on galaxian scale. Every autumn a dark northern sky discloses even to unaided eyes the candle flame that is our nearest neighbor among all spiral galaxies, the galaxy in the constellation Andromeda. Wide recognition that this object is akin to the Milky Way—a second huge, rotating pool of tens of billions of stars, with all their precursors, congeners, rarities, patterns, even wastes—is a 20th-century feat, owing much to Mount Wilson Observatory, especially to Edwin Hubble and Walter Baade.

Paul Hodge has himself studied this galaxy, directly in the sky and in the reports of other watchers, for more than 30 years. His book is an up-to-date guide to the big place. Such a guide has to treat the entire furniture of a galaxy. To that end, the author has collected and compared maps, graphs, tables of data and images over the entire spectrum to assemble a case study of extrasolar astronomy. He recounts not only what we know but how we know it and how well. Such a monograph demands a good deal from its readers by its very variety, even though it spares us the dense mathematical analysis. In some 15 chapters—a few others, full of interest, are mainly historical—we see the details of a great star city two million light-years away.

No optical telescope can see through the layer of dust in which we live to the center of the Milky Way, 100 times nearer to us than the Andromeda galaxy. Our neighbor is almost in the clear; we trade the loss of detail for new perspective and clear view. The spindle shape of Andromeda makes plain that its spinning and near-circular disk is

not seen face-on but tilted pretty close to our line of sight. We can hardly remedy this chance defect of our cosmic location.

In 1885 a famous new star shone out nearly as bright as the entire galaxy long before its new category, that of supernova, was recognized. Now we know 100 or so remnants of past supernova there, to compare with those here at home. No second supernova has appeared in Andromeda, nor have we picked up any pulsing relic so far-away. Counterparts abound there to all kinds of variable stars familiar in the Milky Way, to hundreds of novas and to Cepheids and eclipsing binaries. There are many open star clusters brighter than our Pleiades, and hundreds of globular clusters, star clouds like Orion, a thousand ionized diffuse nebulae, dust clouds and lanes, planetary nebulae, x-ray sources, radio emission from giant molecular clouds, swirling neutral hydrogen and cosmic-ray electrons circulating in weak interstellar magnetic fields.

An extraordinary pair of maps compares the spiral arms in an outer part of Andromeda with a similar pattern in the Milky Way. In Andromeda the arms are much more widely spaced; they have a smaller pitch angle, and between the arms there is less stellar background. We still hope to trace out the spiral arms in full, an unfinished aim both here and there; we see a warp in the disk, but little sign of a central bar. The galaxy extends in visible light wider than our own by around 50 percent but gives birth to new stars at a tenth the rate of our galaxy, except that within one dense ring of stars the fertility is comparable. The tiny, star-packed nucleus hides a condensed, quite massive object within, but one even less active than the similar, enigmatic center of our own galaxy.

Strangest of all is what we do *not* see. The internal mass judged from star motions at the extreme borders of Andromeda is perhaps twice what the Milky Way shows. But the simple energetics of the two big galaxies moving toward each other implies that the pair, if gravitationally bound over that distance of 15 diameters, must share a total mass tens of times greater than all the mass we detect from the motions of visible stars. Massive, unseen halos of dark matter are present. That was clearly grasped almost 20 years ago. Unseen matter dominates cosmic gravity not only faraway but here in our little group of galaxies as well. One riddle more: Have other stargazers there ever peered out in wonder at their neighbor, the Milky Way?

ESSAY by George E. Brown, Jr.

Can Scientists "Make Change Their Friend"?

For 50 years, federal science policy has been rooted in cold war perceptions born of a world in which the U.S. had only one true military rival and no true economic rivals. This policy has created the most active, innovative and expensive research system in history, the envy of our allies and adversaries alike. Now it is clear to many in the science policy community that public support for research and development must respond to a changed perception of reality—a "paradigm shift"—that requires us to reconsider the role of science in our society.

As the military imperative fades, we begin to see that we share a planet on which we confront overpopulation, environmental degradation, economic stagnation and injustice, as well as ethnic and religious division. But this unity in adversity has yet to lead to a comparable unity in vision about how to cope with the global challenges of a post–cold war world.

Those who have tried to define the nature of this paradigm shift (I count myself in this group) have quickly learned that scientists are profoundly unsettled by the prospect of change in the system that has nurtured them since World War II. As policymakers try to draft a blueprint for science that conforms with new global realities, many in the scientific community view this effort as a thinly veiled attempt to substitute the judgment of politicians for that of scientists in selecting the research agenda of the future. Those scientists seem genuinely unaware that political priorities have always dictated the general allocation of research funding: our priorities have been in place for so long that many assume them to be immutable.

I am deeply concerned that a serious misperception of current changes may prevent scientists from participating effectively in the fluid and evolving process of refocusing our national scientific and technical resources. My thesis is that scientists must be part of the reshaping that is taking place; they must accept President Bill Clinton's challenge to "make change our friend." But scientists cannot be involved unless they understand both the historical factors that have shaped research funding in the past and the political reality that underlies federally funded research.

Public support of science is a relatively recent phenomenon, which became significant in the mid-19th century when the federal government provided for land-grant institutions of higher education in every state and for the support of research in agriculture. In the years leading up to World War I, the government began to support research in such other practical areas as standards of weights and measures, geodesy and surveying, and aviation.

World War II and the cold war represented a paradigm shift that increased the amount and changed the nature of federal support for science. Budgets for research grew, as did the size of the system. The great bulk of the increase went into development of weaponry and control of the space environment. The success of science in enhancing our national security led to a mystical faith that continued support for science would help solve a range of other problems and lead to a heaven here on earth. Now that World War II and the cold war are over, national security has many new dimensions. And the American people have a feeling that heaven is not yet at hand. In fact, many feel that they were happier in the good old days and that as a nation we are losing out to our international competitors.

The science policy that satisfied the needs of the cold war must now be transformed to respond to a different world. There is an emerging consensus that national security needs to be redefined to encompass a strong and growing economy; a healthier environment; a reduction of the global tensions caused by economic disparities and ethnic, cultural and religious conflicts; and adaptive institutional structures (bureaucracies) able to respond to human needs promptly and efficiently.

I anticipate that our changing view of national security will lead to a variety of alterations in the federally funded research system. Clearly, the balance between civilian and military R&D expenditure must go from the current 40:60 ratio to at least about 50:50, as it was through most of the 1960s and 1970s. This balance would shift $15 billion or more from military to civilian R&D over the next few years, if the total expenditure remains at its current level. Much of the reduction on the military side will come in the area of advanced development of major weapons systems, while basic and applied military research will probably remain about the same.

On the civilian side, support for university investigators from the National Science Foundation, the National Institutes of Health and other civilian granting agencies may increase. There will continue to be a trend toward funding interdisciplinary teams working on large, complex questions. Likewise, a greater emphasis will be placed on the contributions of the social and behavioral sciences to solving these problems. The role of the federal laboratories may shift substantially: they will be encouraged to assist both private industry and universities with research tasks. They may also be called on to help improve the quality of science education.

Most of the research-funding agencies are reexamining their focus and priorities in the light of the paradigm I am describing. Increasingly, policymakers, such as members of Congress—who are charged by the public and the Constitution to secure the public welfare and safety—will not act to fund programs they perceive as unrelated to the evolving security goals of the nation. Nor will they accept on faith alone the statements of scientists that their research is intrinsically worthwhile and necessary to confront the challenges I have listed.

The scientific community must help reconcile these political standards with its own standards of excellence and intellectual autonomy. Such scientific traditions cannot be perceived as conflicting with, or superseding, the test of political and economic necessity, nor can they be justified through appeals to myth and privilege. Generous public support for a vibrant research system will be achieved only through the open and realistic articulation of the potential for science to contribute to a new national security in the 21st century.

GEORGE E. BROWN, JR., congressman from California, is chairman of the House Committee on Science, Space and Technology.