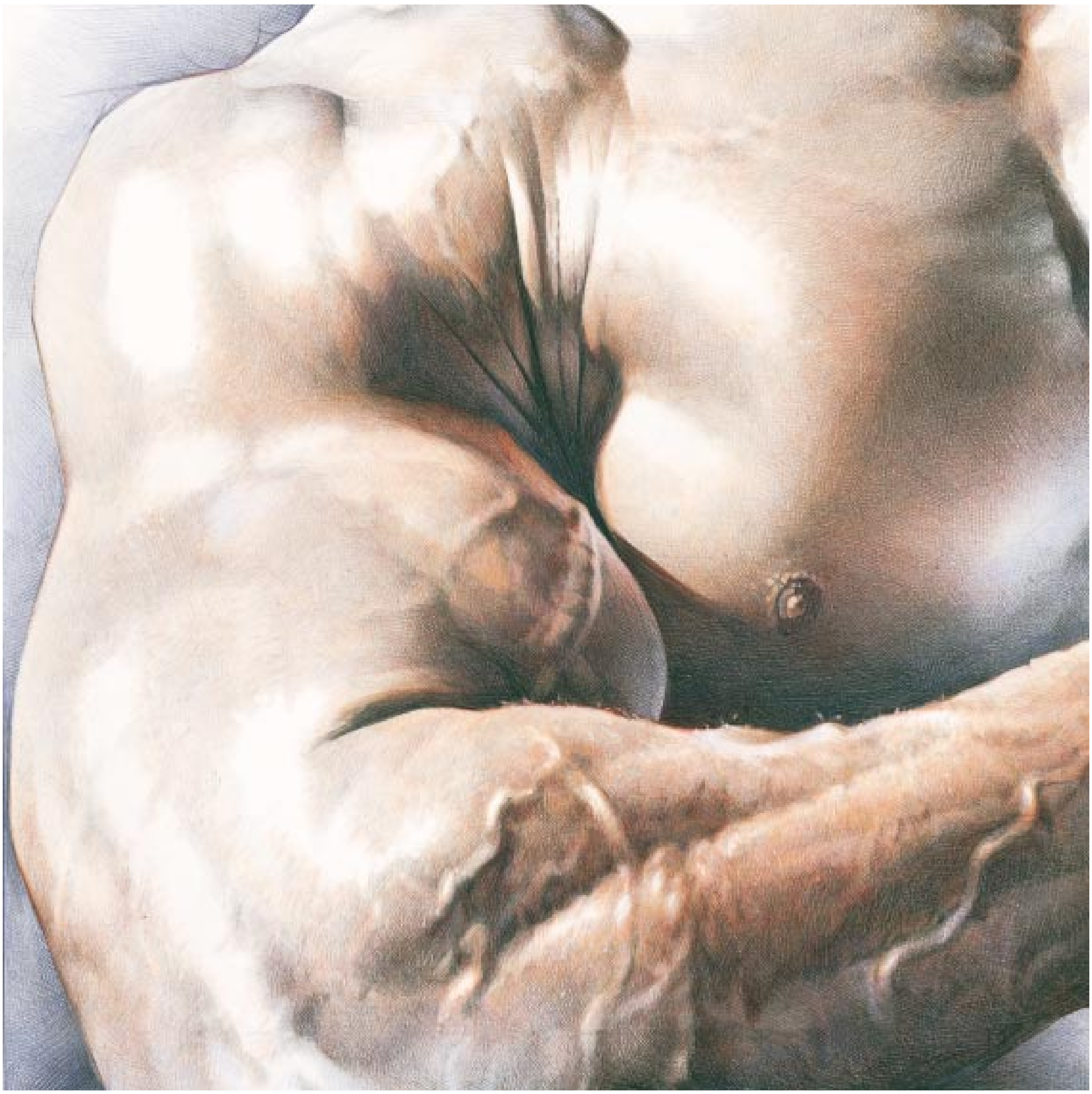
SCIENTIFIC AMERICAN

FEBRUARY 1995
$3.95

Bubbles turn sound into light.

Debunking The Bell Curve.

Microchips: How much smaller?



***Building strong muscles** is only one of the uses tried for anabolic steroids.*

SCIENTIFIC AMERICAN

February 1995 Volume 272 Number 2

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

DEPARTMENTS

THE COVER drawing shows the toned physique of a bodybuilder. To achieve such muscularity, athletes have reportedly taken androgenic-anabolic steroids since the 1940s, but the practice has been banned for the past 25 years. Physicians are now giving these drugs to growing numbers of aging men to improve their well-being. The trend may help bridge the gap between illicit and legitimate steroid use (see "The History of Synthetic Testosterone," by John M. Hoberman and Charles E. Yesalis, page 76). Drawing by C. Bruce Morser.

LETTERS TO THE EDITORS

Is Anybody Out There?

In "The Search for Extraterrestrial Life" [SCIENTIFIC AMERICAN, October 1994], Carl Sagan makes the point that alien life would almost certainly be based on carbon because no other element "comes close to carbon in the variety and intricacy of the compounds it can form." But certain polyhedral borane molecules (near-spherical compounds of boron and hydrogen) support a chemical diversity that approaches that of organic chemistry. Assuming life has to be based on molecules with a carbon framework because of carbon's "unique" chemistry may be a bit parochial.

WALTER H. KNOTH
Mendenhall, Pa.

Sagan mentions only one of the three major SETI efforts actively looking for radio signals from extraterrestrial civilizations. For the past 22 years, Ohio State University has had a SETI program under the direction of Robert Dixon, and for the past 18 years, our group at the University of California at Berkeley has been conducting search operations on some of the world's best radio telescopes. Our project, SERENDIP, is a piggyback system that operates alongside and simultaneous with other radio astronomy observations, so our costs are very low. Unfortunately, our project is currently unfunded.

STUART BOWYER
DAN WERTHIMER
CHUCK DONNELLY
JEFF COBB
University of California, Berkeley

Planning a Sustainable World

I was disappointed that "Sustaining Life on the Earth," by Robert W. Kates [SCIENTIFIC AMERICAN, October 1994], made no mention of how urban and regional planning could help reduce the threats of pollution and biota loss. For example, Sacramento County, California, recently experimented with subdivision design standards that encourage alternatives to traveling by car. These standards emphasize a close-grained mix of land uses (including transit stops), short city blocks, radial street patterns and the prohibition of culs-de-sac. In addition, urban planning increasingly calls for compact development, to reduce the loss of agricultural land and wilderness areas. Is it enough to place all our faith in technological fixes, or should we place greater importance on the "softer" solutions offered by contemporary planning approaches?

ROBERT YOUNG
Guelph, Ontario

Missing the Forest

In the October 1994 issue Stephen Jay Gould's article "The Evolution of Life on the Earth" is mistitled; replace "Life" with "Animals," and the contents are accurately described. The evolutionary history of plants is a fascinating one as revealed by the fossil record and the study of contemporary species. Without plants we would have neither Dr. Gould nor the paper on which *Scientific American* is printed.

ROBERT ORNDUFF
Berkeley, Calif.

Life on the Fringe

The world no doubt greeted your October 1994 single-topic issue, "Life in the Universe," with unrestrained enthusiasm. Here in remote Saskatchewan, we are even more eagerly awaiting its sequel: "Life Elsewhere."

CHRISTIAN STUHR
Swift Current, Saskatchewan

Evans's Emotions

Antonio R. Damasio's claim that "emotion is integral to the process of reasoning" ["Essay," SCIENTIFIC AMERICAN, October 1994] is perhaps less counterintuitive than he supposes. Nearly 130 years ago, in a commentary on W.E.H. Lecky's *History of the Rise and Influence of the Spirit of Rationalism in Europe,* Mary Ann Evans (better known as George Eliot) observed that "Mr. Lecky has the advantage of being strongly impressed with the great part played by the emotions in the formation of opinion." She then chides him for failing to distinguish properly "between the complexity of the conditions that produce prevalent states of mind, and the inability of particular minds to give distinct reasons for the preferences or persuasions produced by those states," that is, the inability of most of us to recognize the role emotion has played in our thinking. She notes that the connection must be "a result of definite processes, if we could only trace them." Evans would surely salute the rigorous method by which Damasio has begun to trace those processes.

THOMAS P. YUNCK
Pasadena, Calif.

A Horse Is a Horse

The anomalous taxonomy described by Madhusree Mukerjee in "What's in a Name?" [SCIENTIFIC AMERICAN, October 1994] has a modern-day counterpart. In *Regina* v. *Ojibway,* a Canadian court found that a horse carrying a down pillow in place of a saddle had legally become a bird. The Small Birds Act defined a bird as "a two-legged animal covered with feathers," and the court agreed that two legs were merely the minimum requirement. The case report was certainly meant as satire, but textbooks have reprinted *Regina* v. *Ojibway* without comment, and generations of law students have repeated it. What Will Rogers said of Congress might apply equally to judicial humor: "Every time they make a joke, it's a law."

ALAN WACHTEL
Palo Alto, Calif.

Letters selected for publication may be edited for length and clarity. Unsolicited manuscripts and correspondence will not be returned or acknowledged unless accompanied by a stamped, self-addressed envelope.

ERRATA

In the illustration of the four-stroke cycle on page 55 of "Improving Automotive Efficiency" [December 1994], the crankshaft was mistakenly labeled as the camshaft. In the same issue "The Annual Ig Nobel Prizes" on page 22 incorrectly describes Alfred Nobel as the inventor of TNT; actually he invented dynamite.

50 AND 100 YEARS AGO

FEBRUARY 1945

"'Industry has begun to appreciate the service that university laboratories can provide,' Dr. Harvey A. Neville of Lehigh University said recently. 'There is an increasing realization that certain types of research can be conducted more effectively in these laboratories, where the academic atmosphere, isolated from the production process, allows a fresh perspective.'"

"Keep an eye on lithium and its industrial applications in the near future. One fifth the weight of aluminum, this lightest of all metals is yielding to the probe of research. Today, lithium and its compounds are finding various uses in copper castings, tin bronzes and other alloys, as well as applications in the ceramic, glass and air-conditioning industries."

"Asparagus butts, a waste product of the canning industry, may find useful application. Three scientists found that juice pressed from these butts can be used as a culture medium to produce bacterial proteinase, an enzyme that digests proteins. Bacterial proteinase is used in the brewing industry and in the leather and textile industries."

"A photographic technique has been worked out that is so sensitive it could presumably take a picture of a ghost, if there actually were such things. This new process, utilizing an illuminating flashlight with an exposure of less than one millionth of a second, photographs things which are invisible, such as the finest details of air disturbances even to the extent of making an image of a heat wave rising from the palm of one's hand."

SCIENTIFIC AMERICAN

FEBRUARY 1895

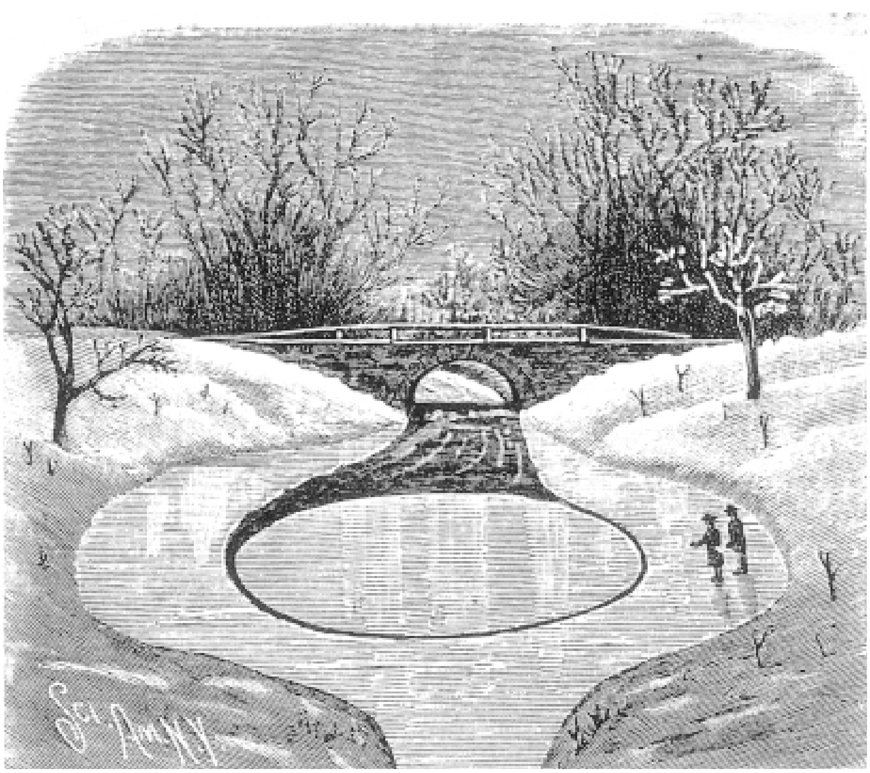
"Such a drop in temperature as was experienced over the greater portion of the United States from the Rocky Mountains to the Atlantic, and from the Canada border to the Gulf of Mexico, during the week ending February 9, has hardly had a parallel since the recording of weather changes has become a regular system. The temperature was below the freezing point for nearly three days throughout the entire United States, except a small area on the southern extremity of Florida and the California coast up to about Portland."

"If we look through all of chemistry, we will find that the one great desire of the chemists was the synthesis of carbon hydrogen for use as fuel. At last it seems as if the great synthesis has been accomplished. By exposing a mixture of lime and anthracite coal to an electric arc, a heavy semi-metallic mass is produced. If the material is immersed in water, the hydrocarbon gas acetylene is given off."

"Over the street doors of one of our most extensively patronized dry goods stores, arc lights are suspended for purposes of illumination. Throngs of ladies are constantly passing to and fro under these lights. The inflammable nature of women's apparel is such as to render it dangerous to stand or pass under arc lights. We noticed a narrow escape for a lady the other evening. Fire fell from the arc lamp and just grazed her dress as she passed under the lamp."

"Paper is now made of such inferior materials that it will soon rot, and very few of the books now published have much chance of a long life. The papermaker thus unwittingly assumes the function of the great literary censor of the age. However, his criticism is mainly destructive, and it is too severe. Without the power of selective appreciation, he condemns to destruction good and bad alike."

"At a place on the Mianus River, near Bedford, N.Y., known locally as the 'ten foot hole,' the stream widens out into a pool forty or fifty feet wide. In this pool there has formed a cake of ice about twenty-five feet in diameter and perfectly circular in shape. This cake, shown in the accompanying illustration, is slowly revolving and is surrounded for about two-thirds of its circumference by stationary ice. There is a space of about three inches between the revolving cake and the stationary ice, except at the up stream side of the cake, where the water is open and the current quite swift. Each revolution takes about six minutes."



A revolving ice cake

SCIENCE AND THE CITIZEN

Dangerous Sex

New signs of risk taking prompt rethinking about AIDS prevention

The messages have been loud and clear: anyone can get AIDS, so wear a condom, don't share needles, get tested. For more than a decade, the admonishments have helped in slowing the spread of the human immunodeficiency virus (HIV) in the U.S.

DONNA BINDER *Impact Visuals*

INFECTION STATUS, willingly tattooed on an AIDS activist, serves as a reminder that the crisis persists.

Yet several recent signs point to a disturbing fact. Small pockets of the population—most notably, teenagers and young adults—appear to be ignoring the warnings about risky sexual behavior. "It could really blossom up again if we don't do something," warns Thomas J. Coates, director of the University of San Francisco Center for AIDS Prevention Studies (CAPS).

Most prevention strategies take a shotgun approach. "Some of the public information campaigns emphasize that everybody is at risk," says Don C. Des Jarlais, an AIDS researcher at Beth Israel Medical Center in New York City. "And that's probably creating some backlash." The message may have reached a saturation point with some groups. A series of CAPS studies found that in urban areas, nearly a quarter of heterosexual adults between the ages of 18 and 25 reported that they had had more than one sex partner during the past year—a proportion that is almost three times greater than that of the general population in large cities. Forty percent of those with multiple partners never used a condom; moreover, condom usage declined with an increasing number of partners.

"Heterosexuals generally don't feel at risk," remarks M. Margaret Dolicini, a CAPS investigator. In the U.S., HIV infection outside the high-risk groups—gay and bisexual men, injecting drug users and partners of these users—is quite low. "A nonzero risk is nothing to be casual about," Dolicini says, especially in light of the fact that heterosexual transmission is the primary vector in many other countries.

The broadside attack may also be unrealistic. "We need to deal differently with AIDS prevention," remarks Derek Hodel of the Gay Men's Health Crisis in New York City. It is one thing to wear a condom at the outset of an epidemic, he remarks, "but when you consider using a condom for the rest of your life, it is a very different prospect." Indeed, a European investigation of monogamous couples in which only one partner was positive for HIV found that condom usage was highly variable: nearly half had unprotected intercourse. "The speculation is that [when] people become involved with a person, they become more committed," Dolicini says. "It becomes a trust issue, so they stop using condoms."

In addition, prevention information can be irrelevant. CAPS researcher Olga A. Grinstead studied the risk behavior of urban women, who become infected mainly through heterosexual contact. Women are told to be monogamous, she says, yet that message is meaningless if their partners are injecting drugs or are not monogamous themselves. "Programs have to be targeted to enhance a woman's empowerment, so that she can insist on condom use or else refuse sex," says Anke A. Ehrhardt of the HIV Center for Clinical and Behavioral Studies at the New York State Psychiatric Institute.

Part of that empowerment means giving women more choices. One possibility would be a viricide that could kill pathogens yet permit pregnancy. The main holdup seems to be the uncertainty of whether the virus is transmitted through the seminal fluid or by the sperm itself. "You will never get very far with a viricide until you close up this problem," predicts Zena A. Stein of the Columbia University School of Public Health. Unfortunately, "there are no labs concentrating on it."

Maintaining protective behavior may mean trying to render condoms as a normally accepted part of sexuality. "Among heterosexuals, protected behavior has not become the norm," Ehrhardt notes. "We know we can change behavior, but we need to make it normative with more consistent messages." It may be sufficient for individuals to practice safe sex until they move into a low-risk category, such as a truly monogamous partnership. "Behavior change won't be perfect forever," says James W. Curran, associate director of the Centers for Disease Control and Prevention. "Whether gay or straight, the idea is to keep people uninfected before they get into long-term relationships."

That advice is problematic for adolescents and young adults, who engage in risk taking and have traditionally been hard to persuade. According to statistical analysis done by Philip S. Rosenberg and his colleagues at the National Cancer Institute, the median age at the time of HIV infection has dropped from the mid-30s during the early years of the epidemic to about age 25 now.

Of young people, gay and bisexual men constitute the most vulnerable segment. A study by the San Francisco Department of Public Health showed that one third of these men engaged in unprotected intercourse; 70 percent did not know they were infected. According to George F. Lemp, the principal researcher, there are several reasons this group is throwing caution to the wind. "They are fairly isolated and alienated from the community of older gay men," he explains. "They haven't built the peer-support network" that has dramatically slowed the spread among older men—three quarters of whom reported they always practice safe sex.

Targeting disparate groups has not been a strong point of federal spending on the AIDS crisis. The CDC allocates about $200 million annually to change high-risk behavior. "Of the actual dollars being spent for HIV prevention efforts, more than half of them go into counseling and testing programs for people who are at low risk," says Des Jarlais, whose work focuses on drug users who inject. "We need to think about specific subgroups, not how to reduce the risk of the population as a whole." To remedy that problem, the CDC recently initiated a program that puts allocations for prevention planning in the hands of communities, which can then determine their priorities.

"Prevention has taken on a new urgency," Ehrhardt says. "A vaccine is further away than we hoped." The World Health Organization expects to begin large-scale trials of two experimental vaccines by 1996 in Thailand and Brazil. Yet experts think these compounds will be only partially effective—raising concerns that they will give recipients a false sense of security. Last June the U.S. opted to drop out of the clinical trials because of doubts about the vaccines.

Despite the current reexamination of prevention programs, some recommendations clearly are too politically hot to ever come to fruition. Needle-exchange programs succeeded in slowing the spread of HIV. But "you can't get [the federal government] to adopt a syringe strategy," Coates laments.

Further, attitudes toward sex education in the schools would have to change. "Young people think of sex as vaginal intercourse," Grinstead says. "What would be useful is to expand their repertoire of behavior." Teaching "noncoital sex" and other safe behaviors is not likely to happen, either—as recently fired Surgeon General Joycelyn Elders can attest (she mentioned that masturbation could be a part of sex-education courses). "The proposals may be controversial," Rosenberg says, "but you have to think the unthinkable with AIDS." *—Philip Yam*

Putting Alzheimer's to the Tests

Several new techniques may detect the disease

A physician examining an elderly person suffering from mild dementia has a difficult diagnosis to make. Alzheimer's disease would be immediately suspected, but in about a third of such patients the cause is actually something different. Because brain biopsies—the only clear means of identifying the neural changes caused by the disease—are rarely performed, definitive diagnosis must wait until after a patient dies. Finding the true problem is crucial, however, because some conditions that mimic the symptoms of Alzheimer's, such as a brain tumor, may be readily treatable.

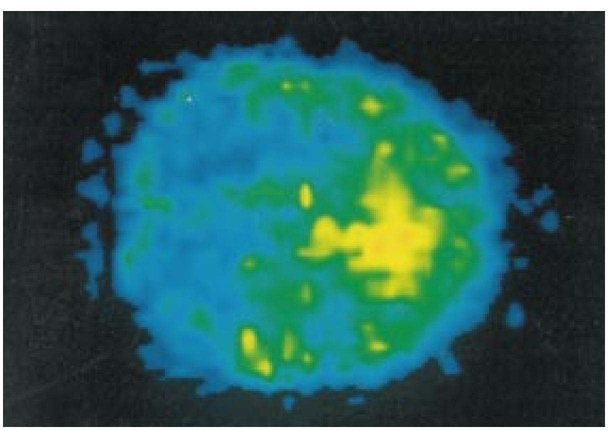
Last year saw an explosion of research on Alzheimer's, and several new techniques offer the hope of more certain identification. They also point to better ways of monitoring the disease's progression, which could speed the discovery of effective drugs.

One approach employs imaging to detect the neuritic plaques and neurofibrillary tangles that are characteristic of Alzheimer's. Daniel R. Weinberger and his colleagues at the National Institute of Mental Health found that a form of magnetic resonance imaging (MRI) called frequency-shifted burst imaging may offer a way of detecting the changes from outside the skull. The process was developed by Jozef H. Duyn and his colleagues at the National Institutes of Health.

Although other imaging techniques can also detect the disease, they cannot be used routinely. Positron emission tomography, for instance, takes several hours, uses highly specialized equipment and involves injecting the patient with a radioactive isotope. Some nonburst MRI techniques can detect changes in patients' brains, but they have not proved useful in diagnosis.

Burst MRI, unlike most other approaches, takes just a few minutes and can be done with an ordinary scanner. Uniquely, it allows the entire brain to be imaged with fair resolution in less than two seconds, without radioisotopes.

According to Weinberger, such scans from patients with probable Alzheimer's show areas where blood flow is lower than in healthy people. So far, however, only eight patients have been studied. "What we don't know yet is whether this information will help in cases where the clinical diagnosis is in doubt," Weinberger cautions. If the technique indeed proves useful, it would be relatively easy to adopt, he notes, because most patients who have suspected Alzheimer's are given conventional MRI scans anyway. Huntington Potter of Harvard Medical School points out that any type of scan can probably detect damage only after it has advanced to a significant degree. Potter, Leonard F. M. Scinto, also at Harvard, and others are working on an even simpler test that might detect the disease in its earliest stages. They put a few drops of a very dilute solution of tropicamide—a synthetic relative of atropine—into the eyes and monitor the response of the pupil 30 minutes later.



DANIEL R. WEINBERGER *National Institute of Mental Health*

ALZHEIMER BRAIN has impaired blood flow in certain areas **(dark regions at far left).**

Tropicamide is used in a 100-fold-greater concentration to dilate pupils for eye examinations, but in a study of 58 individuals published in *Science,* probable Alzheimer's patients showed pronounced dilation even from the dilute solution. The test agreed impressively with diagnoses made by neurologists. The compound blocks the action of the neurotransmitter acetylcholine. Potter's investigation was inspired by the observation that Down's syndrome patients, who often develop Alzheimer's-like brain damage, are sensitive to acetylcholine inhibitors. The sensitivity of Alzheimer's patients to tropicamide jibes with the observation that acetylcholine-producing neurons in the brain are among those destroyed by the disease. Genica, a company that just merged with Athena Neurosciences in San Francisco, intends to market a test based on the research.

Workers at Athena are pioneering yet another strategy. Building on recent advances in understanding the biochemistry of Alzheimer's, they have studied levels of proteins called tau and beta-amyloid in the cerebrospinal fluid of 120 patients and controls. According to John Groom, Athena's president, high levels of beta-amyloid are "strong negative predictors" of Alzheimer's, whereas high levels of tau indicate presence of the disease. Athena's full results have not yet been published, but the company is recognized as a leader, Potter says. Groom hopes that tests for tau and beta-amyloid might be able to monitor the progression of the disease and aid in diagnosis. Groom would also like to commercialize a test that would not detect Alzheimer's itself but merely provide an estimate of how likely an individual is to develop the condition. The test is based on the observation—made more than a year ago and extensively corroborated—that people who have a type of blood protein called apolipoprotein E—specifically, type 4—are more likely to acquire Alzheimer's than are people with other forms. That information, too, could help forestall the disease.

The scientific gains are encouraging, but time is short for patients—and for society. Today there are about three million individuals with Alzheimer's in the U.S.; that number is predicted to reach 14 million in 10 years. —*Tim Beardsley*

Global Warming Is Still a Hot Topic

Arrival of the seasons may show greenhouse effect

The question of whether the earth has succumbed to global warming has loudly been argued by scientists and politicians alike. Now the quiet voice of an electrical engineer has been added to the debate. Although the new appraisal is yet to be fully published, the analysis cries out for attention because it is novel in its approach and conclusion: not only has global warming arrived, the signal should have been obvious years ago.

The new message is reminiscent of when climate researcher James E. Hansen of the National Aeronautics and Space Administration Goddard Institute for Space Studies testified before Congress in the summer of 1988. Hansen said he was 99 percent sure that global warming was here—and few were in the mood to disagree. That summer was one of the hottest and driest in recent memory, and the temperature in Washington, D.C., was, if anything, leading the national trend in unpleasantness. So it came as no surprise that the testimony provoked a great deal of public interest and concern.

But the past few summers have been neither particularly hot nor dry, and some researchers suspect the scorching 1980s may have been the result of natural variability. The scientific community has simply not reached a consensus on whether greenhouse warming has yet been demonstrated.

So the latest contribution to the subject by AT&T Bell Laboratories engineer David J. Thomson is especially intriguing. Delivered without fanfare in San Francisco on a pleasantly cool December day, Thomson's presentation to the American Geophysical Union offered dramatic findings. He reported that "...changes in carbon dioxide resulting from human activities are causing large, and readily observable, changes both in the average temperature and in the seasonal cycle." Thomson reached such a concrete conclusion by taking a fresh look at the problem.

Recognizing the difficulty of constructing the history of global average temperature from a meager set of sampling locations, Thomson instead considered in detail particular sites with long historical records. But he did not examine average temperature. Rather he carefully tracked the annual cycle—that is, the timing of the seasons—using measurements from, among other places, central England between 1651 and 1991. Thomson recognized that one dramatic shift during this period was simply a result of the switch from the Julian to Gregorian calendars in 1752. When corrected for this artifact, most of the 340-year record indicates that the timing of the annual temperature cycle shifts gradually by a little over a day each century. At least that was the pattern until 50 or so years ago. Since about 1940 a pronounced anomaly in the timing of the seasons has appeared in Northern Hemisphere records.

To understand this variation, one has to appreciate the controls on the timing of the seasons. The seasonal cycle results from competition between direct solar heating (which peaks up north at the summer solstice, June 22) and transport of heat from elsewhere on the globe. Were transport perfectly efficient, peak heating would occur everywhere on January 3, when the earth is closest to the sun. Because transport is imperfect, different sites experience both radiative and transport modes of heating to different extents.

In the Northern Hemisphere the greatest radiative heating occurs many months before the highest transport heating; in the Southern Hemisphere, peak radiative and transport heating occur at nearly the same time. Moreover, a gradual shift in the timing of the seasons is expected as the earth's axis of rotation reorients.

But what a significant shift there has been. Thomson has seen a change in the balance between the two forms of heating superimposed on the natural trend of precession. The Northern Hemisphere is seemingly being forced away from the transport mode toward the radiative mode—just as might be expected from greenhouse amplification of solar warming. Conversely, the timing of the seasons in the Southern Hemisphere has hardly been affected. But, according to Thomson's reasoning, the Southern Hemisphere would not be expected to show large changes in the annual cycle, because the radiative and transport modes down under peak at nearly the same time.

Thomson's focus on the changes in seasonal timing allowed him to sidestep completely the nasty problem of compiling an accurate global average temperature from limited historical records. He has further managed to identify greenhouse warming and eliminate any natural increase in solar output as the cause of the past several decades of change.

It remains unclear how the new analysis will be received by climatologists. But Jeffrey J. Park of Yale University points out that Thomson, developer of a highly respected technique in spectral analysis, is one of the notable figures in signal-processing research, and it will be difficult for scientists to take the engineer's suggestions lightly. Thomson himself played down the statistical aspects of his recent appraisal of global warming, remarking that "this is not a very subtle analysis." If his disclaimer proves true, the lack of subtlety will make Thomson's detection of global greenhouse warming only that much more impressive. —*David Schneider*

Broken Dreamtime

Will the koala go the way of the dodo?

The bushfires that raged in the past year or so during one of the worst dry spells in recent Australian history destroyed scores of houses. They also consumed trees that are home to animals that have helped sell airplane tickets to tourists visiting this island continent. The blazes put an additional strain on diminishing koala habitat: the land where these creatures live in eastern Australia is increasingly being sought by real-estate developers.

Koalas have come to live cheek-by-snout with people moving into coastal areas populated with the animals' prized food. Koalas prefer to eat the leaves of less than a dozen of the 650 native varieties of eucalyptus trees. Undeniably, the past 100 years have not been good to this marsupial (koalas are bears only in their resemblance to the genus *Teddy*). Millions of pelts went to England around the turn of the century as a sought-after, cheap and durable fur.

Despite mounting threats, it is unclear just how endangered this age-old creature is. The koala—which plays a critical role in the Dreamtime, the Aboriginal myth of the creation of the world—has a reclusive nature, so it is difficult to perform an accurate census. Although estimates by the Australian Koala Foundation (AKF) suggest that its numbers have dropped from 400,000 in the mid-1980s to between 40,000 and 80,000 today, no one really knows how many koalas remain.

Notwithstanding concerns voiced by a few activist groups, the Australian government has not put the koala on its endangered list, which comprises 75 vertebrates and 223 vascular plants. The state government in New South Wales, with its abundance of vacation and retiree homes, has designated the koala "vulnerable," a notch below endangered. Yet park officials have had to move koalas from several islands off the coast in the state of Victoria because of the marsupial's overpopulation, says Jim Crennan of the Australian Nature Conservation Agency.

The federal government in Canberra has actually tried to generate interest in species it considers to be more threatened. But "the koala is a national icon," notes Crennan to explain why there is more popular attention devoted to it than to an endangered species such as the Northern Hairy-Nosed Wombat. The government also supported a campaign to substitute a chocolate Easter bunny with a chocolate bilby. (The rabbit-eared bilbies are threatened, whereas rabbits are considered a serious pest.)

MARTIN HARVEY *Wildlife Collection*

EUCALYPTUS HOMES of the koala are being chopped down to provide space for residential development.

The state governments do maintain some wildlife management programs for the koala; the federal government places tight restrictions on exports to foreign zoos, and a number of university research programs exist. But perhaps because of the cute-and-cuddly factor, a great deal of research and care for the koala occurs at the grassroots level. One notable example is the privately run Koala Hospital located in the New South Wales town of Port Macquarie, 300 miles to the north of the city of Sydney.

The hospital, founded 21 years ago by two shopkeepers and a local newspaperman, provided assistance to some 250 koalas last year, a figure that rose dramatically with the fires. The average number of patients seen by the hospital is usually closer to 170—most frequently the result of car accidents, dog attacks and bacterial diseases such as *Chlamydia.* The hospital survives largely on volunteer labor. (Twenty percent of the AUS$70,000 a year in expenditures needed to run the facility are from interest on royalties for a song about koalas, "Goodbye Blinky Bill," written by Australian singer John Williamson.)

Another approach to saving koalas is more conceptual. The AKF has a database that combines on-site surveys and satellite data into a Koala Habitat Atlas. It has begun to provide both a census and an assessment of how much koala living space has been lost. "It's not how many animals are left—it's how many trees are left and how many trees can be sustained," says Deborah Tabart, the AKF's executive director.

This information can be employed to divert builders away from stands of eucalyptus. It may also give the AKF or another group enough data to apply to the government to have the koala listed as an endangered species. Unfortunately, the atlas, the compilation of which began in 1990, can also be used by real-estate developers seeking untouched areas, Tabart says.

Koalas, which have low fecundity, are not particularly well adapted to survive the destruction of their arboreal homes or to live near people. Having abnormally small adrenal glands in relation to body weight, the animals do not cope well with stress, states Ken Phillips, a volunteer researcher at the Koala Hospital who is also a professor of psychology and telecommunications at New York University. The nocturnal creature is easily blinded by car lights, writes Phillips in a recent book, *Koalas: Australia's Ancient Ones.* And despite long teeth and claws, which could make them a worthy adversary, they are slow, lumbering and easily upset. Koalas do not fight back when a dog attacks.

Most Australians have never seen a live koala in the wild. If human incursions continue unabated into stands of eucalyptus, Phillips notes, they may never see one. Aboriginal mythology holds that koalas, when abused, have powers that can induce drought. The story seems to have a strange parallel in fact. Australia has experienced a severe drought, and in places such as Port Macquarie, the koala has definitely had its placid existence disrupted. *—Gary Stix*

Finessing Fermat, Again

The wily proof may finally be finished

When Andrew J. Wiles of Princeton University announced in December 1993 that his proof of Fermat's Last Theorem was incomplete, some mathematicians predicted that it could take years to finish. Only 10 short months later Wiles seemingly proved them wrong and Fermat right. He has now simplified his proof of Pierre de Fermat's proposal—which the French mathematician scribbled in a book margin in the late 1630s—that the equation $x^n + y^n = z^n$ has no integer solutions if the exponent is greater than 2. Most experts now say the new argument looks solid.

Four scholars deemed Wiles's second proof incontestable last October. He then sent E-mail messages to some 20 colleagues, telling them a surprise package was on its way. Each received two manuscripts via express mail: *Modular Elliptic Curves and Fermat's Last Theorem,* offering the revised proof, and *Ring Theoretic Properties of Certain Hecke Algebras,* which validates an assumption used in the proof. Wiles devised the work in the latter text with a former student, Richard L. Taylor of the University of Cambridge. Both papers have been submitted to the *Annals of Mathematics.* "People are quite confident that this proof works," reports Henri R. Darmon of McGill University. "All the concepts involved have been studied at length, and what he's added is small."

Indeed, the second proof uses essentially the same strategy as the first, relying on certain connections between Fermat's famed assertion and the theory of elliptic curves. These links were first noted a decade ago by Gerhard Frey of the University of Essen in Germany. He observed that any solutions contradicting Fermat's claim would generate a strange class of semistable elliptic curves. Further, these curves would violate certain conditions set forth in another famous supposition in number theory, the Taniyama-Shimura conjecture. In 1986 Kenneth Ribet of the University of California at Berkeley proved that if the Taniyama-Shimura conjecture were true—at least for semistable elliptic curves—then Fermat's Last Theorem would be true, too.

Wiles's original attempt at proving the theorem by way of the Taniyama-Shimura conjecture stumbled near the end. To complete the proof, Wiles tried to construct a so-called Euler system using a technique developed by Viktor A. Kolyvagin of the V. A. Steklov Institute of Mathematics in Russia. "He attempted what looked like the most logical way to proceed," explains Karl C. Rubin of Ohio State University, "but now it seems very difficult to do things that way."

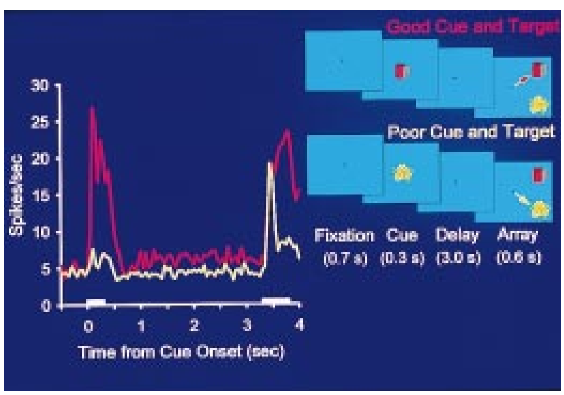
This time Wiles avoided Euler systems and the troublesome technique. "Instead of tackling the gap head-on, he has found an elegant and beautiful way around it," Darmon explains. The new ending invokes Hecke algebras—an approach Wiles toyed with four years ago and abandoned. Darmon urged Wiles to rethink this earlier tack while taking his seminar on the proof last spring.

For now, the suggestion seems to have

Commanding Attention

Studying consciousness is a tricky task, so researchers tease apart aspects of mental processing in the hope that the parts may yet illuminate the whole. One of those lines of inquiry recently produced attention-grabbing results—literally. At the annual meeting of the Society for Neuroscience last November, researchers presented new findings on how animals pay attention to visual cues, a process that is being studied as a surrogate for consciousness. It appears that remembered properties of objects can influence which neurons in the visual pathway show sustained activity. The outcome determines which objects' representations are relayed to higher brain centers.

The findings come from work in macaque monkeys. Robert Desimone and his associates at the National Institute of Mental Health studied the activity of neurons in the brains of these creatures. In one set of experiments the animals had been trained to respond to a symbol when it was flashed on a screen; an irrelevant, distracting symbol was displayed simultaneously. The scientists found that in at least two higher regions of the visual pathway, neurons that started to respond to the distracting symbol were quickly inhibited by their neighbors. When responding to the target, in contrast, neurons were not inhibited.



ROBERT DESIMONE *National Institute of Mental Health*

RESPONSE OF NEURONS in part of the visual pathway to two symbols **(red and yellow at right)** ***depends on which one is sought. If the animal seeks the "poor" stimulus, activity is fast inhibited*** **(yellow line at left).**

Desimone suggests that a form of competition is taking place. In his view, nerve cells extending from regions of the brain where memories are stored—probably the prefrontal cortex—bias the outcome as neurons in the visual pathway vie to become active. The bias operates in such a way that unfamiliar objects and remembered objects of great significance are more likely to win in the competition than are familiar, unimportant ones. "The memory system filters what should get into consciousness," Desimone states. Other experiments have shown that whether a cue appears in an expected or unexpected location also influences the competition.

The notion that competition occurs has been discussed for some time, but the latest experiments show it occurring throughout most of the visual pathway and in varied experimental settings. The details are still controversial, however. Some researchers disagree about where in the pathway the suppression of irrelevant stimuli takes place and how exactly it exerts its effects. But, for Desimone, the idea that competition could explain attention is gaining ground on older views that envisaged the process as a sequential search. For more information, keep watching this space. And be sure to pay attention. *—Tim Beardsley* paid off. "It will take some time to verify, but this proof looks very promising," Rubin says. Hundreds of mathematicians are now studying Wiles's work in search of errors. "This is probably the most scrutinized math manuscript in history," Darmon comments. It may prove to be one of the most consequential as well. The Taniyama-Shimura conjecture joins key concepts from calculus and algebra—a union that could breed novel insights in both fields. Proving "Fermat's Last Theorem is a symbolic victory, but it's the proof of the Taniyama-Shimura conjecture that counts," Darmon explains. After 350 years, Fermat's 15 minutes of celebrity may finally be up. *—Kristin Leutwyler*

It's Getting Easier to Find a Date

Geochronologists reconcile two timescales

In their quest to determine accurate ages for everything from superplume eruptions to hominid fossils, geologists have recently turned to the heavens. No, they are not praying for further funding. They are using astronomy to improve on their traditional geologic chronometer, the decay of radioactive elements. These researchers are forging the gears of a geologic clock from traces of the earth's orbital changes. And in the process, they are recalibrating history.

Scientists have long recognized that variations in the earth's orbit influenced ancient climates. This phenomenon occurs because shifts in the orientation of the rotation axis, in the angle of axial tilt and in the circularity of the orbit control the amount of sunlight falling at different latitudes. Such changes have, for example, ushered in and out a series of Pleistocene ice ages. These climate fluctuations are, in turn, imprinted in the sediments of the geologic record.

Because the timing of orbital oscillations can be precisely calculated, the age of strata bearing identifiable climate cycles can also be determined. And over the past few years ever more researchers have been doing just that. Their studies provide what is now a well-accepted astronomical calibration for the past five million years.

A pivotal analysis was reported in 1990. That year Nicholas J. Shackleton of the University of Cambridge, Andre L. Berger of the Catholic University of Louvain and William R. Peltier of the University of Toronto correlated changes in the ratio of oxygen isotopes in microscopic shells from Pacific Ocean sediments with astronomical oscillations. Oxygen isotopes serve to track climatic change because they indirectly reveal the amount of water that has evaporated from the ocean and been stored in polar ice. Soon afterward Frits J. Hilgen of Utrecht University published an astronomically based chronology for climate cycles found in exposures of Mediterranean sediments.

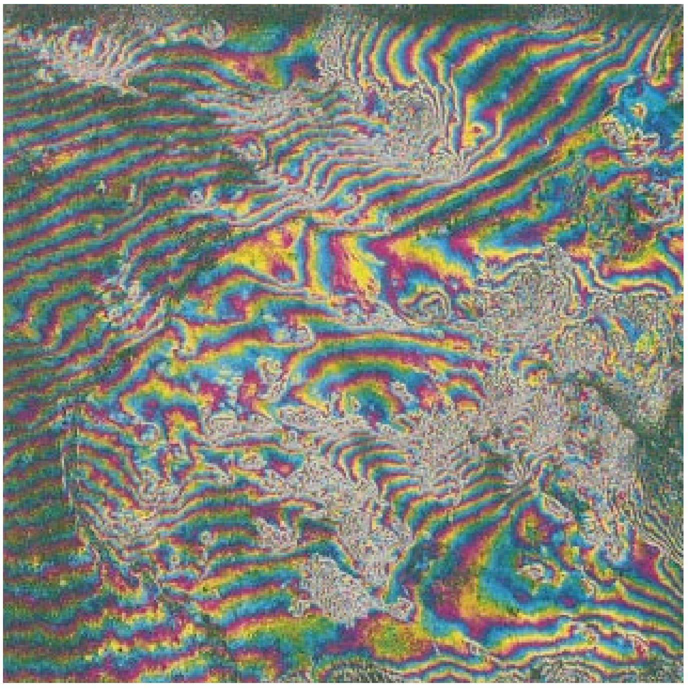
The new chronology was initially at odds with accepted ages based on the decay of radioactive elements. In this dating technique, researchers measure how much radioactive potassium within an igneous rock has decayed into argon. Because they know the half-life of the potassium, geochronologists can calculate how much time has passed for said amount of argon to have been produced. Age estimates for several reversals in the earth's magnetic field—which can serve as markers for both radiometric and astronomical dating—at first suggested that the two systems were off by about 7 percent.

The discrepancy prompted many scientists to review early potassium-argon radiometric work. "It took Shackleton's orbitally tuned timescale to force a reassessment of radiometric ages established during the 1960s and 1970s," remarks Carl C. Swisher III of the Berkeley Geochronology Center. He points out that the early work had been so important to geochronology that until recently most researchers hesitated to challenge it. Geochronologists are currently confident that the early radiometric dates were, in fact, too young. But Swisher also notes that some decade-old radiometric dates from East Africa done by Ian McDougall of Australian

Seeing How the Earth Moved

Running interference is not confined to the football field. Scientists at the National Center for Space Studies in Toulouse, France, and at the Jet Propulsion Laboratory in Pasadena, Calif., are using the interference principle to develop new maps of earthquakes. These radar "interferograms," as they are called, can reveal the extent of deformation of the earth's crust that took place—even if those changes were centimeters in size. The image reproduced at the right shows, among other things, ground motion in southern California after the Landers earthquake of June 1992, which measured in at magnitude 7.3 on the Richter scale.

By juxtaposing radar images obtained by the European Space Agency's ERS-1 satellite before the quake with images taken several months later, researchers created interference patterns similar to those made by oil spreading on water. The color banding in the picture corresponds to the relative phase in the two superimposed radar images, which, in turn, depends on the height of the local topography and on changes in topography caused by the quake. *—David Schneider*



National University agree with the new astronomical ages.

Last May, Paul R. Renne, along with Swisher and other colleagues at the Berkeley Geochronology Center and Berkeley's Institute of Human Origins, suggested that the astronomical timescale could be used to refine a fundamental laboratory standard for argon-argon radiometric dating. This variation on the potassium-argon technique provides only a relative measure of age: a standard mineral of known age must be used with each analysis to give an absolute calibration. Swisher now cautions that geochronologists should not go too far in completely accepting the astronomical timescale's recalibration of the radiometric clock. "What you really want is for the astronomical and radiometric dates to agree independently—if they don't, then you need to figure out why," he explains. Agreement between the two timescales is not quite perfect, but it appears that the discrepancy is becoming negligible.

Employing sediments to check the absolute age of a critical volcanic standard reverses the traditional roles for these kinds of rock: typically geologists use volcanic ash layers to date important sedimentary sections, not vice versa. But the inverted strategy seems to have worked for Renne and his colleagues. And running counter to the establishment must not have felt that strange to this research group—after all, they do come from Berkeley. *—David Schneider*

Nothing Personal, You're Just Not My Type

Most movie aliens cannot reproduce successfully

The world was safe all along. Back in the 1950s, moviemakers regularly served up the spectacle of creatures from other planets attempting to take over our bucolic little orb. Heroic earthlings fought the aliens with dynamite, napalm, atomic torpedoes and bad acting. But had the heroes been better acquainted with life-history strategies—the reproductive behaviors that determine patterns of population growth—they might not have bothered.

"In general," says May R. Berenbaum, professor of entomology at the University of Illinois, "none of [the aliens] exhibit the opportunistic sorts of reproductive traits or characteristics of organisms that successfully colonize." Her findings help to explain why earthlings should be afraid of at least some contemporary invaders—such as zebra mussels, bark beetles, medflies and, perhaps, the sluglike aliens featured in one of last year's movies, *The Puppet Masters.*

Berenbaum's first try at sharing her interest in cinematic biology was an abortive attempt to organize an insect film festival while she was a graduate student in entomology at Cornell University. "I thought it could be a way to attract a large audience to insect issues," she recalls. Shortly after joining the faculty at Illinois, however, she teamed up with Richard J. Leskosky, assistant director of the university's Unit for Cinema Studies, to get the bug film festival finally flying. The couple went on to produce several papers on insects in movies and cartoons, as well as a daughter—the aficionado of entomology and the film buff were married in 1988.

In 1991 Berenbaum was invited to lecture at the Midwest Population Biology Conference. "I thought it might be entertaining to look at population biology in the movies," she says. "And a recurrent biological theme in films is the idea of invading organisms." It seemed a testable hypothesis to see whether fictional invaders share the attributes that invading organisms in real biotic communities display.

So Berenbaum and Leskosky looked at the life histories of aliens in science-fiction movies released in the 1950s, a time when movies were lousy with invading organisms. (Film historians attribute the obsession to the recent memory of Nazi aggression and to cold war paranoia.) The two used *Keep Watching the Skies,* an exhaustive compilation of science-fiction flicks, as their database. Their lecture—a version of which was published in 1992 in the *Bulletin of the Ecological Society of America*—ended with films from 1957, the last year the book covered.

Of the 133 movies described in the text, 67 fulfilled Berenbaum's requirement for inclusion in the study: they depicted an extraterrestrial species. Analysis showed that invading is a dicey lifestyle choice. "We determined that, collectively, alien beings in science-fiction films do suffer from high mortali-

ALIEN INVADERS from two movies in the 1950s have different life-history strategies. The creature from* The Giant Claw *(1957) is a* K *type and has low fecundity—it lays one egg. The pods from* Invasion of the Body Snatchers *(1956) are smaller, more prolific* r *types and should have been successful in taking over the planet. Oh, well.

ty," Berenbaum and Leskosky wrote. Indeed, in only three of the movies do aliens survive to see the credits. They die at the hands of humans and through acts of God or the director—earthquakes, volcanoes and avalanches all come to the rescue of humans. But heroic deeds or natural disasters were probably less threatening to the long-term survival of the invading species than their own poor fecundity.

Opportunistic species, those good at colonizing new environments, exhibit so-called *r*-selection. "These species have a set of traits—small body size, rapid growth, huge brood sizes," Berenbaum explains. Those qualities lead to a high *r,* the intrinsic rate of increase, which can cause big problems in real life as well as in real bad movies. "European bark beetles, just reported in Illinois a year or two ago, almost shut down the entire Christmas tree industry," Berenbaum notes. "A National Academy of Sciences study showed that introduced species have caused about $90 billion worth of economic damage."

On the other hand, those species marked by slow development, reproduction later in life and large body size—traits of so-called *K*-selection—are good at competing in a stable environment but poorer at colonizing a new one. Thus, California farmers find themselves fearing *r* medflies far more than *K* elephants. The typical 1950s alien invader, however, is far closer biologically to an elephant than to an insect.

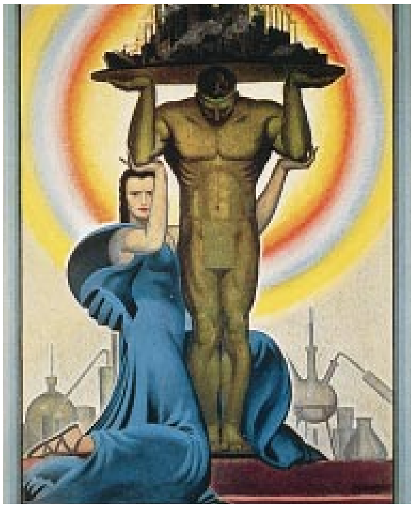
The aliens also suffer from overconfidence. Berenbaum and Leskosky found that 42 of the movies showcased either a lone invader or a pair. Only 21 films have the earth threatened by more than six intruders. The small initial invading force, combined with failure to go forth and multiply once they reach the planet, renders most movie aliens nothing more than short-term threats.

The few invaders who do try to reproduce once they land make efforts that are biologically questionable. For example, the attempts of the title character in *Devil Girl from Mars* (1955) to mate with humans is "an undertaking fraught with hazards associated with postzygotic reproductive isolating mechanisms," Berenbaum and Leskosky point out. (Strictly speaking, the humanoid Devil Girl was less interested in colonization than in the abduction of human males that she could import back to her home planet for breeding stock.)

Students of Stanislavsky would therefore do well to contemplate population biology in addition to "The Method" before accepting roles in science-fiction films starring *K*-type invaders. They would not act so scared. —*Steve Mirsky*

Out of the Lab and into the Fire

Two exhibits put science under the microscope



LEON H. SODERSTON *National Museum of American History*

TERRY MIURA *National Museum of American History*

TWO VIEWS of science are shown in a Smithsonian Institution exhibit: "Chemical Industry Upheld by Pure Science" **(left)** ***in 1937 and the recent "Frankentomato"*** **(right).**

Downstairs from the First Ladies' inaugural gowns and not too far from the television-set chairs of Edith and Archie Bunker in the National Museum of American History in Washington, D.C., sprawls the show "Science in American Life." The exhibit, which opened last April, as well as an upcoming one, "The Last Act: The Atomic Bomb and the End of World War II," which debuts in May at the National Air and Space Museum, has provoked heated debate about the way science and technology are portrayed. Behind this contentious argument lies a larger issue: whether scientists are no longer perceived by the public as revered truth-seekers but as flawed humans whose theories and technology simply reflect contemporary cultural concerns.

Some observers claim that the exhibits sacrifice scientific and historical accuracy to concentrate on social issues. The current show, for instance, looks at the environmental movement and discrimination against women and minorities within the scientific community. Two life-size talking mannequins re-create researchers arguing over who deserves credit for discovering saccharine. And the area devoted to the present day depicts both "spectacular advances in space exploration, electronics and medicine" and disasters such as Three Mile Island and the explosion of the space shuttle *Challenger.* Such events have, according to the exhibit's literature, encouraged people to question all authority, scientific or otherwise.

"There are a handful of places in 'Science in American Life' where the negative impact of science is not adequately balanced by good things," comments Ned D. Heindel of the American Chemical Society, which contributed $5.3 million to the project. Robert L. Park of the American Physical Society feels the museum presentation is severely skewed. "Ring the bell of evil, and viewers will automatically blame a scientist," he wrote in a recent editorial in the *Washington Post.*

Still under construction, "The Last Act" has elicited similar reactions. Veterans charged that the first version of the accompanying text unfairly portrayed the Japanese as victims and the Americans as aggressors. Further, in a letter to the director of the Smithsonian Institution, 48 historians objected to the revised text. By neglecting to mention that some individuals questioned the use of atomic bombs, the scholars state, the document now "utterly fails" to depict the event appropriately—despite a legal mandate to do so. (In response to veterans' concerns, the Senate passed a resolution declaring that the museum "has an obligation under the Federal law to portray history in the proper context of the times.")

Mike Fetters of the National Air and Space Museum contends that the show was intended to cover science as well as history. The Manhattan Project section "shows science and personalities," Fetters explains. And in the Hiroshima and Nagasaki area, planners "tried to show scientifically objective qualities [of dropping the two atomic bombs], such as radiation effects, as well as the human effects on the populations of the two cities," he elaborates.

By looking at the human side of science and technology, both exhibits reflect shifts in the history of science. Once the province of scientists with an interest in the past, the field has evolved into one in which practitioners may know more about society than about, say, chemistry. Over the past 30 years, museums have changed their focus from "the hardware of science to the social context," says Arthur P. Molella of the National Museum of American History. "Instead of just looking at how ideas evolved, science historians now look at scientists as human beings within their culture. In 'Science in American Life,' we wanted to show that science is very much a part of American history, as much as politics or business."

Viewing scientists this way is the basis for an approach known as social constructivism. According to this theory, institutional, political, economic or personal interests influence our theories about science. Science is not the revelation of an independent reality but a reflection of these underlying forces.

Many scientists worry that these historians have taken the discussion too far. Social constructivists conclude that "science is just another narrative and has no greater claim to authority than any other narrative," Park notes. "On that basis, a Native American folk legend of the origin of humans should be taken as seriously as the theory of evolution." In their book, *Higher Superstition: The Academic Left and Its Quarrels with Science,* published last year, Paul R. Gross of the University of Virginia and Norman Levitt of Rutgers University argue that attacks on science reflect political power struggles, not careful philosophical study. Because "most other aspects of capitalistic society seem to be working, and if the object of your attack is Western culture, then sooner or later you've got to attack science," Gross comments.

Gross points out that historians have been talking about cultural influences on science for many years but that scientists have not paid much attention until now. The sudden defensiveness may have some roots in the end of the cold war—a cultural phenomenon itself—which left many research programs bereft of an apparent mission. Furthermore, academic life is increasingly competitive as the government wrestles with budget cuts. "Many scientists now perceive their position as more precarious," says Peter L. Galison of Harvard University.

Galison feels, however, that "most historians of science don't see themselves as trying to attack science." They consider viewing science within society as a way to understand its complexities. The current context of science in our society as demonstrated by "Science in American Life" seems to be one of isolation from the public accompanied by considerable skepticism about scientific progress. Scientific achievements—both good and bad—have been set aside for preservation, just as the gowns upstairs have.

Indeed, the exhibit quotes S. B. Woo of the University of Delaware on this very separation: "Nowadays, science is getting so abstract, specialized, and complex that the public tends to regard it as hopelessly esoteric and irrelevant to their lives. We in the scientific community need to develop better strategies for 'selling' science to the public and convincing them of its value."

Regardless of the debate about context, there are some points about content that do not seem open for discussion among scientists. Susan Solomon of the National Oceanic and Atmospheric Administration is certain about her field. "As a physical scientist," she declares, "I have to believe there are physical truths." —*Sasha Nemecek*

No, Really, It Was *This* Big

Researchers at Fisheries and Oceans Canada in West Vancouver have engineered a fly-fisherman's fantasy. Robert H. Devlin and his colleagues altered the DNA of Pacific salmon to create fish that are, on average, more than 11 times bigger than their natural counterparts.

To spawn these gargantuan creatures, the group used the process that has stimulated similar growth in transgenic mice. The investigators microinjected growth promoter genes from two sources into the Pacific salmon eggs. The first source was a nonhomologous species—in this case, the mouse. The second, homologous source was sockeye salmon. The scientists then hatched the some 3,000 eggs and examined the offspring that survived to at least one year of age.

In almost all cases, salmon containing the mammalian gene were of normal size. The salmon containing the piscine growth promoter gene, however, generally showed dramatic increases in size and weight—in fact, one fish was 37 times larger than a standard Pacific salmon. Winter levels of the growth hormone produced by the gene were also 40 times higher than normal in the transgenic salmon.

"We're not necessarily trying to produce gigantic salmon here," Devlin explains. "What we'd like to do is increase their size so that aquaculture food production of the fish is more efficient and profitable." —*Steven Vames*

ROBERT H. DEVLIN *Fisheries and Oceans Canada*

Once Upon a Time There Was a Theory

Determining what drives economic growth or decline depends as much on storytelling as on data. For the past decade or so, a new crop of theorists, including Paul Romer of the University of California at Berkeley and Robert Lucas of the University of Chicago, has been pushing "endogenous" growth. These economists argue that development results entirely from economic factors: once upon a time the U.S. was poor; then its population grew and became urbanized, allowing business to exploit economies of scale. As a result, the country became rich. There are even mathematical models to prove it. Economists understand all the variables in this story—population, production costs and profits—and so it is called endogenous (inside the economics).

Economic historians such as Joel Mokyr of Northwestern University and Nathan Rosenberg of Stanford University, meanwhile, favor "exogenous" explanations based on outside factors, in particular technological change. Once upon a time we were all poor; then a wave of gadgets swept over England. As a result, we are all rich, or well on our way to it, if we will let people alone. This story does a better job of explaining, for instance, why China's per capita income grows by 10 percent a year: the Chinese, like the Koreans and Japanese before them, adopt the best methods invented thus far and quickly catch up with more advanced nations, regardless of endogenous factors in their economy.

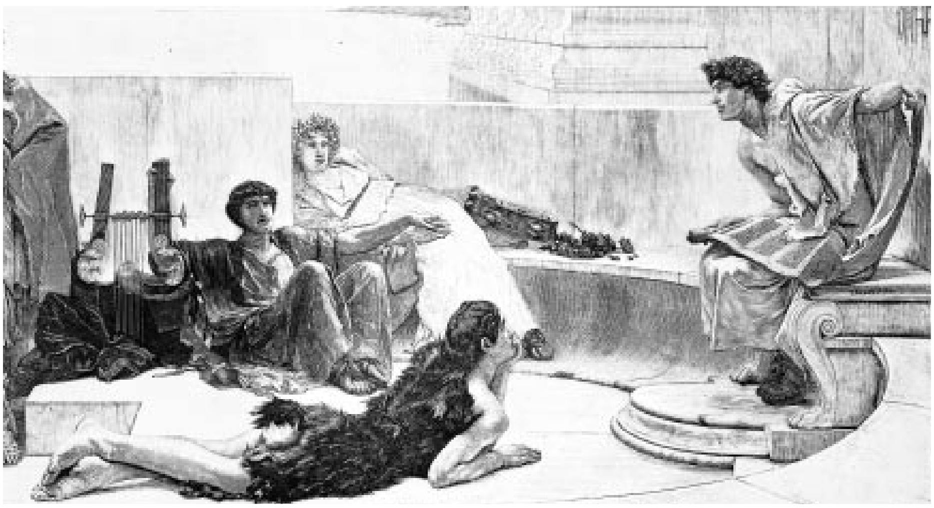
The exogenous version has its own problems, but one of the major reasons the endogenist economic theorists argue against it seems to be that it offends their narrative sense. They do not like to have to step outside of economics to talk about the nature and causes of the wealth of nations.

Are endogenists being unscientific in wanting to tell one kind of story rather than another? Is economics as a whole simply not a science because its practitioners rely on narrative? Nobel Prize–winning physicist Steven Weinberg wrote a paper in 1983 called "Beautiful Theories" to make the point that aesthetic principles are at the heart of good physics. Indeed, astrophysicist Subrahmanyan Chandrasekhar wrote an entire, beautiful book on the matter, *Truth and Beauty: Aesthetics and Motivations in Science.* The same issues of narrative aesthetics appear in paleontology. Classical Darwinian evolution proceeds like a film in dignified slow motion: punctuated equilibrium interleaves still photographs with bursts of silent movies.

The notion of "science" as divorced from storytelling arose largely during the past century. Before then the word—like its French, Tamil, Turkish and Japanese counterparts—meant "systematic inquiry." The German word for the humanities is *Geisteswissenschaft,* or "inquiry into the human spirit," as opposed to *Naturwissenschaft,* which singles out the external world. When Sigmund Freud's translators rendered *Geisteswissenschaft* as "mental science," they left many readers wondering why a science had so much to do with Oedipus and other literary tales.

Most sciences do storytelling and model building. At one end of the gamut sits Newtonian physics—the *Principia* (1687) is essentially geometric rather than narrative. Charles Darwin's biology in *The Origin of Species* (1859), in contrast, is almost entirely historical and devoid of mathematical models. Nevertheless, most scientists, and economists among them, hate to admit to something so childish-sounding as telling stories. They want to emulate Newton's elegance rather than Darwin's complexity. One suspects that the relative prestige of the two methods has more to do with age than anything else. If a proto-Darwin had published in 1687, and a neo-Newton in 1859, you can bet the prestige of storytelling versus timeless modeling would be reversed.

Even when economists rely on models, decisions about what to include or what conclusions to draw turn on some principle of storytelling. Particularly important is the sense of beginnings and endings. To an eclectic Keynesian, the story "oil prices went up in 1973, causing inflation" is full of meaning. But for a monetarist, it ends too soon: a rise in oil prices without some corresponding fall elsewhere is not an equilibrium.



BETTMANN ARCHIVE

STORYTELLING is essential to science and literature. Mathematical models may be in fashion now, but aesthetic principles guide scientists much as they did early readers of Homer.

Meanwhile Keynesians accuse the monetarist plotline of an ill-motivated beginning: focusing on money, the end result of production, ignores where it comes from and why.

So when forecasters debate the impact of Federal Reserve Chairman Alan Greenspan's latest hike in interest rates, they are not just contesting the coefficients for their equations. They are debating which narrative style best describes the economy. And in economics, as in other sciences, you cannot get away from the aesthetics of human stories. Or, as Damon Runyon put it: " 'I thank you, Herbie, just the same,' I say, 'but I must do without your tip,' and with this I start walking away. 'Now,' Herbie says, 'wait a minute. A story goes with it,' he says." Well, of course, this is a different matter entirely.

DONALD N. McCLOSKEY is professor of economics and history at the University of Iowa.

A Budgetary Storm Is Brewing

The new Congress may chop technology funds

Prospects for federal research and development are up in the air as Republicans looking for budget cuts take control on Capitol Hill. Although it is too early to say where the chips will fall, clear signs indicate that science and technology will not be spared in the housecleaning.

Before last November's elections, Republican staff of the House Budget Committee startled science watchers by publishing a series of draconian possible cuts. The list included abolishing the U.S. Geological Survey and the National Biological Survey as well as limiting the annual growth of the National Science Foundation to 1 percent less than the rate of inflation, which is now 2.7 percent. The Advanced Technology Program of the National Institute of Standards and Technology—which in recent years has become the centerpiece of the administration's technology development efforts—was also targeted.

Although the cost-cutting litany was rushed out in the heat of the campaign and is likely to be forgotten when committees come to decide on expenditures, it does indicate the spirit of Capitol Hill. Senate Republicans showed the same mood when they voted in December to abolish the congressional Office of Technology Assessment, a bipartisan research agency that has provided analyses of technical topics for lawmakers.

The battle over budgets might, however, avoid turning into a massacre. The game of musical chairs that followed the election ended with some supporters of science in charge of key committees. Representative Robert S. Walker of Pennsylvania, who has championed support for basic research, wound up at the helm of the House Committee on Science—formerly the Committee on Science, Space and Technology. The reconstituted committee has an expanded jurisdiction that includes energy research, and Walker said in December he "was prepared to continue to move in the direction" of creating a government Department of Science, Space, Energy and Technology out of existing agencies.

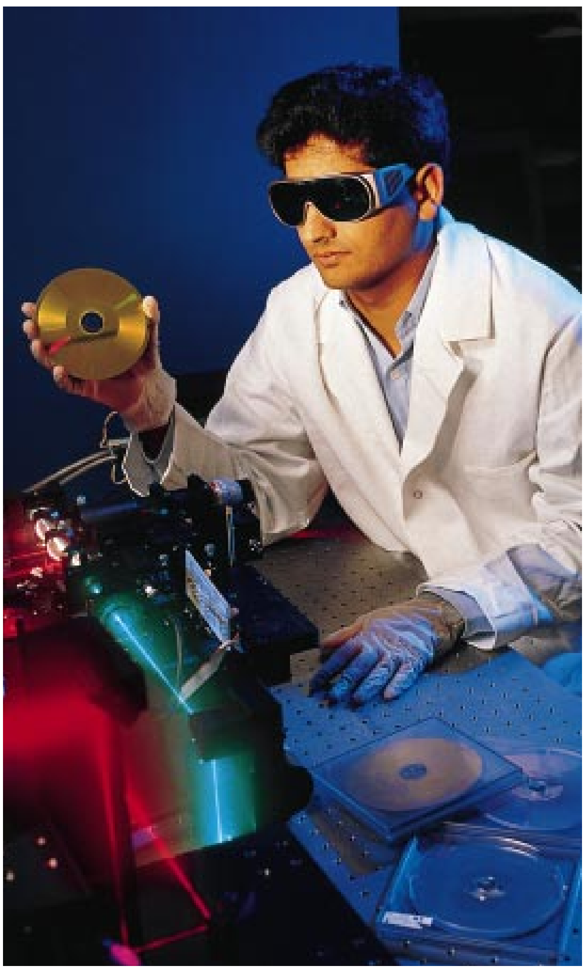
Walker declares himself a supporter of independent university-based research. He promises to continue the campaign of his predecessor, Representative George E. Brown, Jr., of California, to eliminate earmarked appropriations—better known as academic pork—for universities and even suggests that colleges seeking earmarks—which escape peer review—might be penalized.

Walker's priorities include using tax incentives to encourage the development of hydrogen-fueled automobiles and to spur the development of a commercial space sector. Nevertheless, he is not friendly to much of current federal support of technology development. He says he would "ultimately" favor eliminating the Advanced Technology Program, which he sees as "a form of national industrial policy." He also complains that "too much of the National Science Foundation's money is being diverted to applied research." Walker indicates he might support the continued existence of the Office of Technology Assessment, but "in a much restructured and downsized way." And he has doubts about the present level of research on global change, which "might be more in tune with politics than with its scientific measure."

Representative Jerry Lewis of California heads the appropriations subcommittee overseeing budgets of independent agencies and thus will have important authority over the budgets of the National Aeronautics and Space Administration and the National Science Foundation. Lewis is recognized as a voice for science, and like Walker he has supported the space station. But whether he and Walker will be able to hold the line for research budgets remains an open question.

The person in overall charge of spending on the House side, Representative Bob Livingston of Louisiana, has not distinguished himself on scientific matters. Livingston, who chairs the Appropriations Committee, was a keen proponent of the Strategic Defense Initiative, a program that could return from the grave. Republicans have pledged to deploy a cost-effective antiballistic missile defense system as soon as possible, although nobody has yet indicated where the funds for such an undertaking would come from.

Biomedical research enjoys strong bipartisan support in Congress, and so the National Institutes of Health is perhaps less likely than other research agencies to be battered badly. It has a champion in the Senate in the person of Mark O. Hatfield of Oregon, chairman of the Senate Appropriations Committee, who has proposed setting up a special fund to protect biomedical research from the vagaries of the budget process. Hatfield's influence may be valuable, but research pertaining to AIDS, sex and reproduction could still be vulnerable to conservative opposition. In the current radical climate, nothing should be taken for granted. —*Tim Beardsley*



ROBERT RATHE *Robert Rathe Photography*

PROJECTS of the Advanced Technology Program, such as work on this high-density optical disk, may be cut.

Agents and Other Animals

Good software help is hard to find

Great things are expected of agents, little pieces of software designed to roam around computer networks making themselves useful. Agents "give people the magical ability to project their desires into cyberspace," raves Marc Porat, who, together with veterans of the team that designed the Apple Macintosh, founded a company called General Magic in Mountain View, Calif., to put agent technology on the market. Meanwhile Microsoft, Apple and others are touting "wizards" and "intelligent agents" that, their advertisements promise, will make complex tasks a snap.

People should know better. Agents are rapidly catching a bad case of the worst kind of computer industry hype: misplaced expectations. It's not that agents won't deliver great things; they probably will. But those great things will almost certainly be different from the ones consumers now expect. Marketers promise Jeeves the perfect butler, but researchers are quite literally struggling to build Bonzo the wonder dog—"and, frankly, even a dog looks pretty ambitious," says Pattie Maes, a researcher at the Media Lab of the Massachusetts Institute of Technology.

The heart of the problem is that most people, quite naturally, expect software to perform the same kind of tasks that, say, travel agents or insurance agents do. Such knowledgeable servants achieve your goals without forcing you to learn the details of their work. In fact, the software agents being built promise nothing of the kind.

When Porat talks about agents, he is referring to a mobile program—that is, any computer program that can send itself across a network and do work on a remote machine. And when other companies create intelligent agents, they are relying on a subtle redefinition of the word "intelligent" made by the school of "nouvelle AI" led by Maes and Rodney A. Brooks, also at M.I.T. These researchers eschew the symbolic reasoning needed by humans to solve problems for the basic intelligence an insect needs to stay alive.

A student of Brooks's, for example, built a robot modeled on a sea slug. The slug wandered around the edges of tidal pools looking for plants; this robot wandered around the edges of desks, looking for soft-drink cans. Nowhere did it have the ability to "think" consciously about what a soft-drink can might be. Its sensors and actuators were simply wired so that it reacted to the shape of a can by moving its arm toward it.

Both Porat's mobile programs and software-only versions of Brooks's virtual critters can do useful work. But neither knows anything about human desire, let alone how to satisfy it. Jeeves might shimmer into the room bearing a restorative pitcher of martinis because he knows Wooster has broken off another engagement that morning. Bonzo, in contrast, trots in with paper and slippers because that is what he always does when his master comes home, so long as Wooster rewards him for it. Programs do not take the initiative; they just do as they are told.

This lack of empathy (for lack of a better word) does not make mobile programs or virtual critters any less useful; it just makes them different. As James E. White of General Magic points out, a mobile program can save a lot of work. Say you have bought 500 shares of Acme Investments at $20 and want to sell when it reaches $30. One way to do this is to get the machine that tracks prices to keep your computer constantly informed of the price—which is likely to result in thousands of disappointing messages. A mobile program, on the other hand, could do the job with two messages: one to send the program over to the quote machine and the other to say, "Sell now!"

The drawback to a mobile program, though, is passivity. Because it will do only what it is told, when it is told, it leaves its owner with the responsibility of issuing instructions at the proper time. Virtual critters go a step further. They initiate their own actions. The trick to making these creations function seems to be to separate acting and understanding.

"MY REAL NAME IS 'FOSSILIZED TREE SAP' BUT MY FRIENDS CALL ME 'AMBER'."

MICHAEL CRAWFORD

Maes and her students are designing a virtual dog, a virtual hamster and other, less animistic critters—all of which exist as disembodied software—that help to sort E-mail and schedule meetings. The hope is that these pets can be trained to do useful tasks, just as real dogs can.

Encouragingly, the kinds of instincts that make a good bird dog, or even a crumb-grabbing cockroach, may yet make a good information retriever. Maes's helper watches how its owner sorts E-mail and, when it senses a pattern, offers to complete the task on its own: picking up the letter and dropping it in the appropriate box. A click on the "OK" button is as good as a virtual dog biscuit and a scratch behind the virtual ears.

But critter-makers have some way to go in teaching their creations to interact more easily with humans—which brings us back to expectations. Critters that can learn to mimic their owner's actions are an improvement over mobile programs because people without programming skills can teach agents new tricks. They still cannot impart skills they do not know. (Imagine if you could not employ a travel agent unless you knew how to be one.) One remedy may be to buy ready-trained agents—although this solution begs the question about expectations.

The social conventions that tell people what to anticipate from dogs—sit, fetch and so on—do not exist for agents. It is one thing for a computer to be trained according to a single person's actions; it is another for one individual to use a computer accustomed to another's quirks.

An additional problem is that even without marketing hype to encourage them, consumers consistently overestimate the intelligence of computer programs—particularly those using language. (Given that the only other items we know that use language are other people, and they are pretty smart, this expectation is not so surprising.) During the 1970s, Joseph Weizenbaum of M.I.T. produced a program that crudely mimicked a Rogerian psychoanalyst by twisting statements into questions. Thus, "I'm <adjective>" became "Why are you <adjective>?" Much to Weizenbaum's surprise, people became deeply emotionally involved with his device—and deeply disappointed when they discovered it was a form of computerized party trick.

PATTIE MAES; ALIVE © 1994 M.I.T. Media Laboratory

VIRTUAL DOG is one of many software critters that could rove around sorting E-mail and keeping out electronic intruders.

To save agents from a similar fate, researchers propose to make them look more like what they are. One strand of Maes's research is to invent animated computer environments. Her virtual dog looks a bit like Walt Disney's Pluto as it lopes across the screen—which dispels any illusions of knowledge and wisdom more quickly than any explanation of how it works. Better, it is starting to learn to use expressions to communicate puzzlement, eagerness and other "emotions" that signal its intentions.

Eventually, our world may contain a bestiary of such critters—none of them Jeeves, but each as familiar as a dog or a horse. Virtual cockroaches might scurry across hard disks, disposing of old files; virtual bees might buzz across networks in search of rich sources of information; and virtual Dobermans might nab electronic intruders. In the meantime, those hoping to be neither confused nor disappointed by the new world might want to try a simple exercise. Every time you see the phrase "intelligent agent," substitute "trainable ant"—which is a better description of exactly what tomorrow's critters might be. Perhaps one depicted as a cartoon crab with scissors for one claw and a pot of glue for another. *—John Browning*

SCIENTIFIC AMERICAN CORRESPONDENCE

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

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

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

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

Photocopying rights are hereby granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue of *Scientific American* for the fee of $3.00 per copy of each article plus $0.50 per page. Such clearance does not extend to the photocopying of articles for promotion or other commercial purposes. Correspondence and payment should be addressed to Copyright Clearance Center, Inc., 222 Rosewood Drive, Danvers, MA 01923. Specify CCC Reference Number ISSN 0036-8733/95. $3.00 + 0.50.

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

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

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

The Chilling Wind of Copyright Law?

Legal changes may reshape Internet activity

If one of the committees trying to define the future of the Internet is right, pretty much everything anyone does in cyberspace may be illegal. The Working Group on Intellectual Property Rights of the White House Information Infrastructure Task Force is not even talking about hacking or software piracy or thefts of confidential information. The team is crafting a whole new definition of copyright law.

According to the group's draft report—issued last summer and the subject of recent public hearings—random browsing of World Wide Web pages, transmission of Usenet postings, reading of electronic mail or any of the other Internet activities may already violate the law. "It's really that bad," says Jessica Litman, a professor of copyright law at Wayne State University. She explains that the team's chairman, Bruce A. Lehman of the Department of Commerce, has made a peculiar reading of a part of the copyright act that applies to computer software and has extended it to all digital data.

The rule in question says that loading a program from a disk into working memory constitutes making a copy—even though the bits and bytes vanish once the computer is turned off. It thus follows, according to the draft, that transferring the text of a document across the Internet so that it can be displayed on a user's screen is also copying—and, unless specifically licensed by the owner, copyright infringement.

Although the principle of fair use may legitimize making such copies, the report suggested that exemptions might shrink as automated licensing schemes are put into place. Jane C. Ginsburg of Columbia University notes that a recent court decision rejected a fair-use defense on the grounds that the defendant, Texaco, could easily have obtained permission to copy. (At the same time, she points out, the court suggested that copyright owners might be compelled to grant permission.)

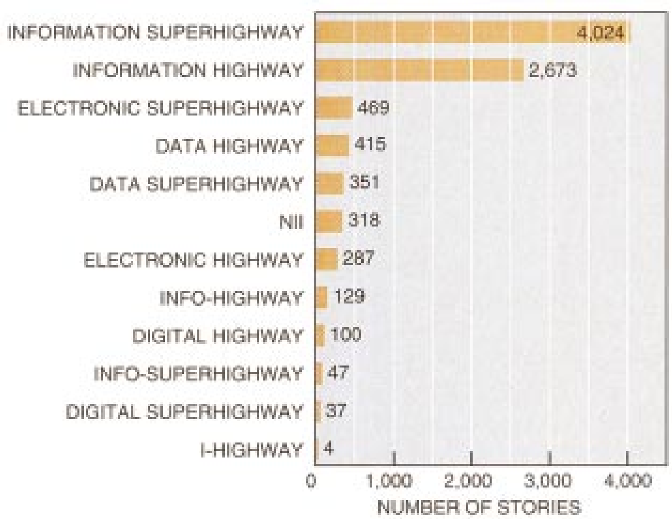
The working group has also advocated abolition of the "first sale" doctrine, which states that someone who buys a copyrighted work, such as a textbook, can freely sell, give or lend it to anyone else without paying additional royalties. As a result, says Pamela Samuelson, professor of law at the University of Pittsburgh, the electronic equivalent of flipping through magazines at a newsstand would be illegal, as would the analogue of wandering into a bookstore and skimming a novel before deciding whether to buy it. Interlibrary loans that make articles in obscure journals available to researchers would also be a thing of the past, remarks Prudence Adler of the Association of Research Libraries. In short, predicts L. Ray Patterson of the University of Georgia, the constitutional mandate that copyright should "promote the progress of science and the useful arts" would be a dead letter.

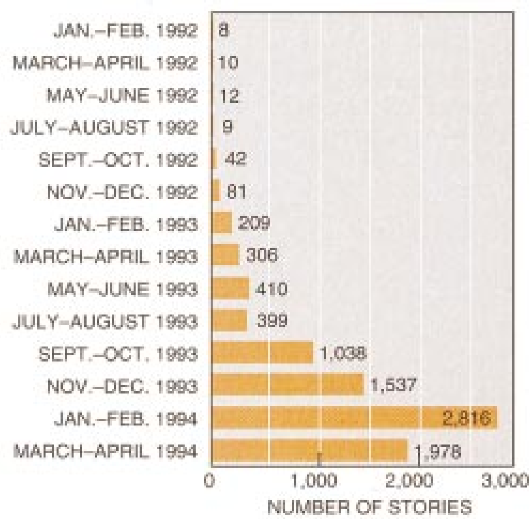
Not surprisingly, Terri A. Southwick of the Patent and Trademark Office takes a different view. Unless they believe their property will be protected, copyright holders will not trust their works to the Internet, and development of a global information highway will be stunted, she explains. If these concerns are not addressed, "the Internet will still thrive," Southwick says, "but it won't reach its full potential." Ginsburg reported that artists appear more or less evenly split between enthusiasm for cyberspace as a wonderful new medium and fear that they may lose all control of their works once recordings, texts or pictures are converted to digital form. (*Scientific American,* for example, currently sharply restricts redistribution of its text on networks.)

No one is willing to bet which viewpoint will prevail. The task force received more than 150 responses and held four days of public hearings. Six weeks after the filing deadline, its members had yet to finish reading responses. The final report, to be issued in March or April, will contain proposed legislation to clarify who can do what to whose data. But the congressional subcommittee that deals with the issue has been cut in half, and its roster is still unsettled. Outgoing staffer William Patry suggests that anything could happen, "including nothing at all." —*Paul Wallich*

How Do They Call It? Let Us Count the Ways

Since the phrase first appeared in 1992, the "information superhighway" has become a familiar part of the American lexicon. Its synonyms, according to a report by the Freedom Forum Media Studies Center at Columbia University, remain somewhat less popular in the newspapers, magazines and broadcasts that were reviewed (*left*). At the same time, however, the concept seems to have peaked before its prime, well before the highway is laid down (*right*).





LAURIE GRACE

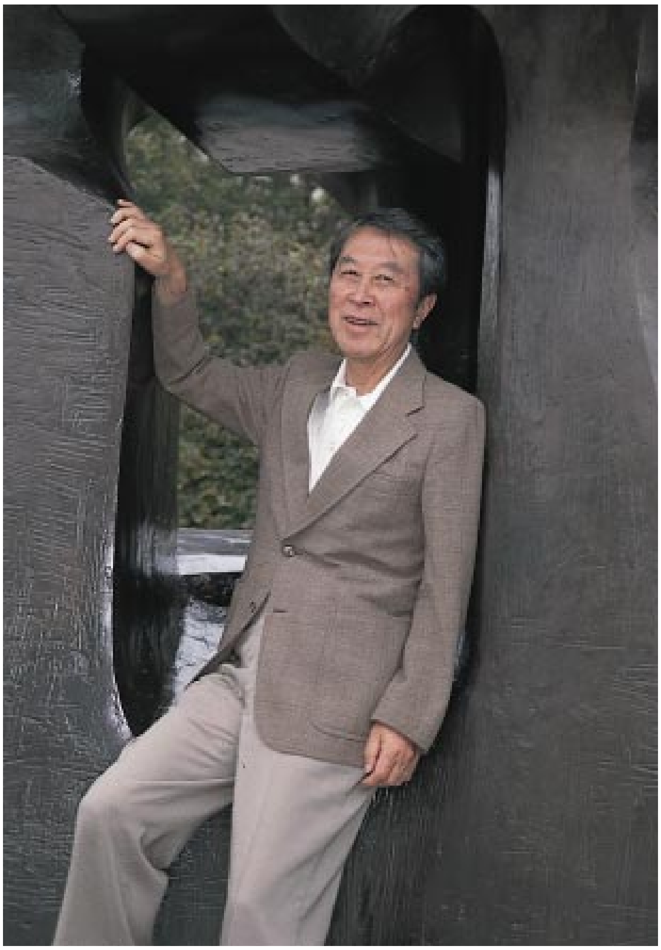
PROFILE: YOICHIRO NAMBU

Strings and Gluons—The Seer Saw Them All

I first saw Yoichiro Nambu almost 10 years ago, from the back row of a graduate seminar in physics at the University of Chicago. A small man in a neat suit, he was sketching long, snaking tubes on the blackboard. Sometimes he said they were vortex lines, found in superconductors; other times he called them strings, connecting quarks. Mystified, and yet fascinated by a bridge between such disparate realms, I later asked him to be my thesis adviser.

Face to face, Nambu was still hard to understand. I was clearly not the first to try. Bruno Zumino of the University of California at Berkeley once recounted his own attempts: “I had the idea that if I can find out what Nambu is thinking about now, I’ll be 10 years ahead in the game. So I talked to him for a long time. But by the time I figured out what he said, 10 years had passed.” Edward Witten of the Institute for Advanced Study in Princeton, N.J., explains: “People don’t understand him, because he is so farsighted.”

Nambu was the first to see that when a physical system such as a superconductor—or an ocean of quarks—defies the symmetry imposed by physical laws, a new particle is born. Along with Moo-Young Han, then a graduate student at Syracuse University, he proposed the existence of gluons, the objects that hold quarks together. He also realized that quarks act as if they are connected by strings, an idea that became the foundation of string theory. “Over the years,” remarks Murray Gell-Mann of the Santa Fe Institute, “you could rely on Yoichiro to provide deep and penetrating insights on very many questions.”



MADHUSREE MUKERJEE

PHYSICIST YOICHIRO NAMBU found that nature, in trying to repair broken symmetry, creates a new particle.

The roots of Nambu’s originality may lie in a singular childhood in prewar Japan. Born in Tokyo in 1921, he was two when the city was destroyed by an earthquake. (He still has a vague recollection of flames.) Kichiro Nambu, his father, had run away from home to attend university and there had met his bride, Kimiko. The earthquake forced him to return to his hometown of Fukui, near Kyoto, with his wife and young son.

The prodigal was forgiven (although his wife never was). Retaining traces of defiance, Kichiro Nambu became a schoolteacher and built his house on the outskirts of town—an act that was later to save him from Allied bombs. From Tokyo he had brought back an eclectic library. Browsing there, his growing boy learned of ideas that allowed him to flee, at least mentally, the excruciating regimen at the local school.

Fukui, in those days, prided itself on having the most militaristic school in Japan. The boys dressed in scratchy army uniforms and were taught to march, shoot and salute. “If you didn’t see a senior boy and so didn’t salute him, he would punch you out,” Nambu recalls. “You had to keep one eye on every person.” At 4:00 A.M. in midwinter, he would walk a mile to school to learn Samurai sword fighting, barefoot on bare floors in unheated halls. To the frail child, school proved as trying as, later, the real Imperial army.

Nor did the school neglect the mind. Heroic deeds—notably, that of a schoolteacher who died saving the emperor’s picture from a fire—embellished the curriculum. Nambu was protected from such teachings by his father’s antiestablishment diatribes. Yet they also prevented him from fitting in. “I had a longing to be like the other boys,” he smiles ruefully. As he grew, he came to realize that his father’s opinions were dangerous in an increasingly warlike Japan.

Thus, Nambu learned to keep his thoughts to himself. This trait served him well later, through years in the army—and perhaps even as a physicist. His originality might come from having to think everything through for himself, from being aware of, but ignoring, ideas in the world outside.

Moving on to a premier college in Tokyo in 1937, Nambu discovered a freer intellectual atmosphere and smart classmates who awed the country boy. Of his courses, physics caused him special trouble: “I couldn’t understand entropy and flunked thermodynamics.” Yet, possibly inspired by Hideki Yukawa, the pioneer who realized that particles transmit force, Nambu chose to aim for a master’s in physics at Tokyo University.

Among his new classmates, he found some underground radicals. Japan was fighting China. “We were told of the victories,” Nambu says, “but these communists somehow also knew about the massacres and defeats.” The academic program turned out to be short: the class graduated six months early so that its members could be drafted.

In the army Nambu dug trenches and

carried boats. “Physically it was hard,” he shrugs, “but inside I was free. As long as you said, ‘Yes, sir, yes, sir,’ they left you alone.” After a year he was assigned to help develop short-wavelength radar. The navy already had such radar, but the army had no confidence in that equipment. Nor was Nambu’s team especially successful: “To test our system, I set it up on a hilltop and hired a boat to take a metal rod out into the ocean. You could see it with your bare eyes—but not with our radar.”

He was then ordered to steal a secret navy document, a paper on field theory by Sin-Itiro Tomonaga, who was applying his discoveries on particle waves to radar waveguides. (Werner Heisenberg’s publications on field theory had arrived from Germany shortly before, after traveling by submarine for a year.) Obtaining these papers—simply by asking a professor—Nambu became acquainted with some of the newest ideas in physics.

Life was quite easy. The unit was housed in a golf club, and romance was budding between Nambu and his assistant, Chieko Hida. For the most part, the war seemed far away. Yet one night Nambu watched a fleet of B-29s fly over Osaka. For a change, they did not drop their bombs on the city but moved on to Fukui. Nambu lost his grandparents; his parents were spared.

After the war, Nambu and Hida married, whereupon he left for Tokyo to take up a long-promised research position. (Hida stayed on in Osaka to look after her mother.) Housing was scarce, and Nambu moved into his laboratory for three years. Gas and electricity were free, and he could bathe in the water basin intended for extinguishing air-raid fires. But his officemate, Ziro Koba, a diligent young man (he once shaved his head for missing a calculation), would come in early and often embarrassed Nambu, who was sleeping across both their desks.

“I was hungry all the time,” Nambu says. Finding food took up most of the week. For the rest, he thought about physics, calculating on rolls of cash-register paper. Koba, a student of Tomonaga, kept Nambu informed about the latter’s work. A group of solid-state physicists in a neighboring office also provided stimulating company.

All that these researchers knew of scientific developments in the West came from sporadic issues of *Time* magazine. Later, journals in a library set up by the Occupation forces helped to fill in the gaps. Yet much had to be reinvented by the Japanese physicists. Sometimes they got there first. After moving to Osaka City University in 1949, Nambu published a formula describing how two particles bind, now known as the Bethe-Salpeter equation. Along with others, he also predicted that strange particles should be created in pairs, a discovery usually attributed to Abraham Pais.

Describing Nambu’s early work, Laurie M. Brown of Northwestern University writes of its “exuberant sense of play.” As his student, I enjoyed Nambu’s sheer pleasure in ideas and his ready laugh (even if I did not always get the joke). In the belief that too much work is harmful, he urged me to attend baseball games and to read the exploits of V. I. Warshawski, the fictional Chicago sleuth.

In 1952 Nambu was invited to visit the Institute for Advanced Study. There he found many brilliant and aggressive young men. “Everyone seemed smarter than I. I could not accomplish what I wanted to and had a nervous breakdown,” Nambu wrote to me decades later, trying to bring comfort during my own travails as a postdoctoral fellow. In 1957, after having moved to Chicago, he proposed a new particle and met with ridicule. (“In a pig’s eye!” Richard Feynman shouted at the conference, Brown recalls. The omega was discovered the next year, in an accelerator.) Meanwhile Nambu had heard J. Robert Schrieffer describe the theory of superconductivity that he had just devised with John Bardeen and Leon N. Cooper. The talk disturbed Nambu: the superconducting fluid did not conserve the number of particles, violating an essential symmetry of nature. It took him two years to crack the puzzle.

Imagine a dog faced with two bowls of equally entic-

COURTESY OF YOICHIRO NAMBU

NAMBU just before his fifth birthday **(top)** ***and studying past midnight in his college dormitory*** **(bottom).**

ing food. Being identical, the bowls present a twofold symmetry. Yet the dog arbitrarily picks one bowl. Unable to accept that the symmetry is entirely lost, Nambu discovered that the dog, in effect, cannot make up her mind and constantly switches from one bowl to the other. By the laws of quantum physics, the oscillation comes to life as a new particle, a boson.

Nambu points out that others, such as Bardeen, Philip W. Anderson, then at Bell Laboratories, and Gerald Rickayzen, then at the University of Illinois, also saw that a superconductor should have such a particle. It was Nambu, however, who detailed the circumstances and significance of its birth and realized that the pion, as well, was born in like manner (by breaking the chiral, or left-right, symmetry of quarks). While he searched for more of its siblings in nature, Nambu circulated his findings in a preprint.

Jeffrey Goldstone, then a postdoctoral fellow at CERN, the European laboratory for particle physics, realized the import of this work and soon published a simpler derivation, noting that the result was general. Thereafter the new particle was dubbed the Goldstone boson. ("At the very least, it should be called the Nambu-Goldstone boson," Goldstone comments.) When Nambu finally published his calculations in 1960, his paper also showed how the initially massless particle mixes with a magnetic field in a superconductor to become heavy. Recognized by Anderson, Peter Higgs, then at the Institute for Advanced Study, and others as a general phenomenon, it later became the Higgs mechanism of the Standard Model of particle physics.

In the years that followed, Nambu studied the dynamics of quarks, suggesting they were held together by gluons carrying a color quantum number to and fro. "He did this in 1965, while the rest of us were floundering about," Gell-Mann says. (Nambu, however, believed the quarks to be observable and assigned them integer electrical charges, an error that Gell-Mann and others corrected.) In 1970, perusing a complex mathematical formula on particle interactions, Nambu saw that it described strings. In the 1980s his "string action" became the backbone of string theory.

"He has an amazing power of coming up with pictures," says Peter G. O. Freund, a colleague at Chicago. While working with Nambu, I noticed that he would look at a problem from several different, yet simultaneous, points of view. It was as if instead of one mind's eye he had at least two, giving him stereoscopic vision into physical systems. Where anyone else saw a flat expanse of meaningless dots, he could perceive vivid, three-dimensional forms leaping out.

Over time, Nambu became known as a seer, albeit a shy one. "I can think of no one who gives such good advice," Witten says. Pierre M. Ramond of the University of Florida observes that the directions of particle physics were often predicted by Nambu's papers—encrypted in the footnotes.

These days Nambu puzzles over how quarks acquire their diverse masses. He suggests they might come from historical accident, such as the quarks being born at different stages of the early universe. His thoughts have also turned to biology and to an old bane, entropy. Nambu calculates that virus-size particles, when placed in a cusp-shaped container, seem to violate gravity and entropy. Perhaps they conceal a clue as to how life-forms defy entropy and become ever more organized. Prophecy or quixotic fancy? Ten years from now, we might know. *—Madhusree Mukerjee*

SCIENTIFIC AMERICAN

Population, Poverty and the Local Environment

As forests and rivers recede, a child's labor can become more valuable to parents, spurring a vicious cycle that traps families in poverty

by Partha S. Dasgupta

CHRISTOPHER PILLITZ *Matrix*

As with politics, we all have widely differing opinions about population. Some would point to population growth as the cause of poverty and environmental degradation. Others would permute the elements of this causal chain, arguing, for example, that poverty is the cause rather than the consequence of increasing numbers. Yet even when studying the semiarid regions of sub-Saharan Africa and the Indian subcontinent, economists have typically not regarded poverty, population growth and the local environment as interconnected. Inquiry into each factor has in large measure gone along its own narrow route, with discussion of their interactions dominated by popular writings—which, although often illuminating, are in the main descriptive and not analytical.

Over the past several years, though, a few investigators have studied the relations between these ingredients more closely. Our approach fuses theoretical modeling with empirical findings drawn from a number of disciplines, such as anthropology, demography, ecology, economics, nutrition and political science. Focusing on the vast numbers of small, rural communities in the poorest regions of the world, the work has identified circumstances in which population growth, poverty and degradation of local resources often fuel one another. The collected research has shown that none of the three elements directly causes the other two; rather each influences, and is in turn influenced by, the others. This new perspective has significant implications for policies aimed at improving life for some of the world's most impoverished inhabitants.

In contrast with this new perspective, with its focus on local experience, popular tracts on the environment and population growth have usually taken a global view. They have emphasized the deleterious effects that a large population would have on our planet in the distant future. Although that slant has its uses, it has drawn attention away from the economic misery endemic today. Disaster is not something the poorest have to wait for: it is occurring even now. Besides, in developing countries, decisions on whether to have a child and on how to share education, food, work, health care and local resources are in large measure made within small entities such as households. So it makes sense to study the link between poverty, population growth and the environment from a myriad of local, even individual, viewpoints.

The household assumes various guises in different parts of the world. Some years ago Gary S. Becker of the University of Chicago was the first investigator to grapple with this difficulty. He used an idealized version of the concept to explore how choices made within a household would respond to changes in the outside world, such as employment opportunities and availability of credit, insurance, health care and education.

One problem with his method, as I saw it when I began my own work some five years ago, was that it studied households in isolation; it did not investigate the dynamics between interacting units. In addition to understanding the forces that encouraged couples to favor large families, I wanted to understand the ways in which a reasoned decision to have children, made by each household, could end up being detrimental to all households.

In studying how such choices are made, I found a second problem with the early approach: by assuming that decision making was shared equally by adults, investigators had taken an altogether too benign view of the process. Control over a family's choices is, after all, often held unequally. If I wanted to understand how decisions were made, I would have to know who was doing the deciding.

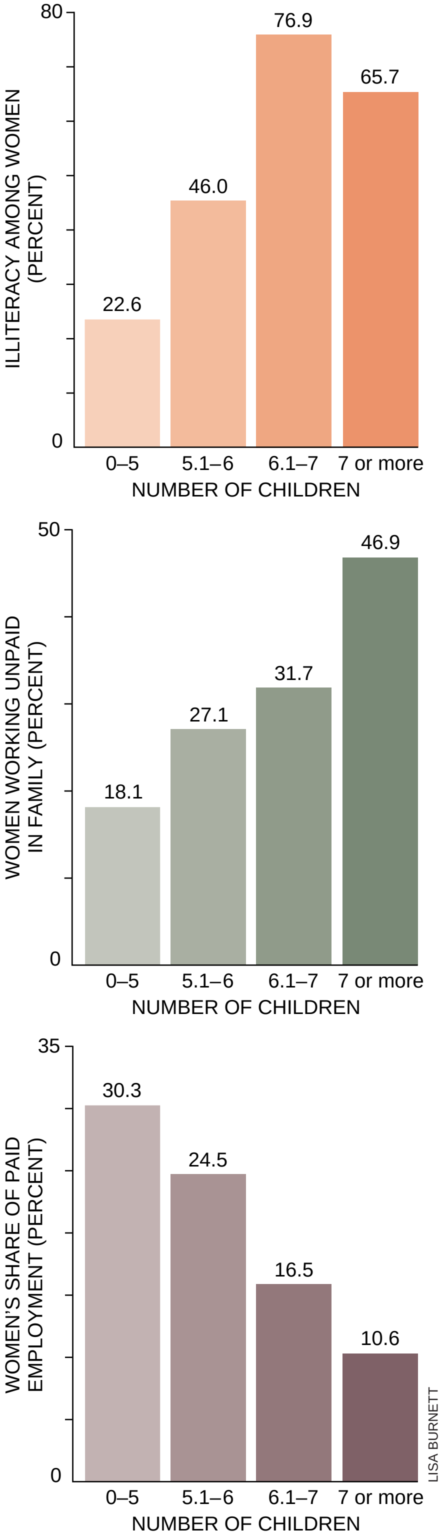
Power and Gender

Those who enjoy the greatest power within a family can often be identified by the way the household's resources are divided. Judith Bruce of the Population Council, Mayra Buvinic of the International Center for Research on Women, Lincoln C. Chen and Amartya Sen of Harvard University and others have observed that the sharing of resources within a household is often unequal even when differences in needs are taken into account. In poor households in the Indian subcontinent, for example, men and boys usually get more sustenance than do women and girls, and the elderly get less than the young.

Such inequities prevail over fertility choices as well. Here also men wield more influence, even though women typically bear the greater cost. To grasp how great the burden can be, consider the number of live babies a woman would normally have if she managed to survive through her childbearing years. This number, called the total fertility rate, is between six and eight in sub-Saharan Africa. Each successful birth there involves at least a year and a half of

FETCHING WATER in Rajasthan, in the west of India, takes up several hours a day for each household. As resources become increasingly sparse and distant, additional hands become more valuable for such daily tasks, creating a demand for families to have more children. The burgeoning population puts more pressure on the environment, spurring a need for even more offspring in a cycle of increasing poverty, population and environmental damage.

PARTHA S. DASGUPTA, who was educated in Varanasi, Delhi and Cambridge, is Frank Ramsey Professor of Economics at the University of Cambridge and Fellow of St. John's College. He is also chairman of the Beijer International Institute of Ecological Economics of the Royal Swedish Academy of Sciences in Stockholm and Fellow of the British Academy. Dasgupta's research has ranged over various aspects of environmental, resource and population economics, most recently poverty and malnutrition.



TOTAL FERTILITY RATE around the world (the average number of children a woman produces) generally increases with the percentage of women in a country who are illiterate (*top*) or work unpaid in the family (*middle*). Fertility decreases when a larger share of the paid employment belongs to women (*bottom*). Bringing in a cash income may empower a woman in making decisions within her family, allowing her to resist pressure to bear more children.

pregnancy and breast-feeding. So in a society where female life expectancy at birth is 50 years and the fertility rate is, say, seven, nearly half of a woman's adult life is spent either carrying a child in her womb or breast-feeding it. And this calculation does not allow for unsuccessful pregnancies.

Another indicator of the price that women pay is maternal mortality. In most poor countries, complications related to pregnancy constitute the largest single cause of death of women in their reproductive years. In some parts of sub-Saharan Africa as many as one woman dies for every 50 live births. (The rate in Scandinavia today is one per 20,000.) At a total fertility rate of seven or more, the chance that a woman entering her reproductive years will not live through them is about one in six. Producing children therefore involves playing a kind of Russian roulette.

Given such a high cost of procreation, one expects that women, given a choice, would opt for fewer children. But are birth rates in fact highest in societies where women have the least power within the family? Data on the status of women from 79 so-called Third World countries display an unmistakable pattern: high fertility, high rates of illiteracy, low share of paid employment and a high percentage working at home for no pay—they all hang together. From the statistics alone it is difficult to discern which of these factors are causing, and which are merely correlated with, high fertility. But the findings are consistent with the possibility that lack of paid employment and education limits a woman's ability to make decisions and therefore promotes population growth.

There is also good reason to think that lack of income-generating employment reduces women's power more directly than does lack of education. Such an insight has implications for policy. It is all well and good, for example, to urge governments in poor countries to invest in literacy programs. But the results could be disappointing. Many factors militate against poor households' taking advantage of subsidized education. If children are needed to work inside and outside the home, then keeping them in school (even a cheap one) is costly. In patrilineal societies, educated girls can also be perceived as less pliable and harder to marry off. Indeed, the benefits of subsidies to even primary education are reaped disproportionately by families that are better off.

In contrast, policies aimed at increasing women's productivity at home and improving their earnings in the marketplace would directly empower them, especially within the family. Greater earning power for women would also raise for men the implicit costs of procreation (which keeps women from bringing in cash income). This is not to deny the value of public investment in primary and secondary education in developing countries. It is only to say we should be wary of claims that such investment is a panacea for the population problem.

The importance of gender inequality to overpopulation in poor nations is fortunately gaining international recognition. Indeed, the United Nations Conference on Population and Development held in Cairo in September 1994 emphasized women's reproductive rights and the means by which they could be protected and promoted. But there is more to the population problem than gender inequalities. Even when both parents participate in the decision to have a child, there are several pathways through which the choice becomes harmful to the community. These routes have been uncovered by inquiring into the various motives for procreation.

Little Hands Help…

One motive, common to humankind, relates to children as ends in themselves. It ranges from the desire to have children because they are playful and enjoyable, to the desire to obey the dictates of tradition and religion. One such injunction emanates from the cult of the ancestor, which, taking religion to be the act of reproducing the lineage, requires women to bear many children [see "High Fertility in Sub-Saharan Africa," by John C. Caldwell and Pat Caldwell; SCIENTIFIC AMERICAN, May 1990].

Such traditions are often perpetuated by imitative behavior. Procreation in closely knit communities is not only a private matter; it is also a social activity, influenced by the cultural milieu. Often there are norms encouraging high fertility rates that no household desires unilaterally to break. (These norms may well have outlasted any rationale they had in the past.) Consequently, so long as all others aim at large families, no household on its own will wish to deviate. Thus, a society can get stuck at a self-sustaining mode of behavior that is characterized by high fertility and low educational attainment.

This does not mean that society will live with it forever. As always, people differ in the extent to which they adhere to tradition. Inevitably some, for one reason or another, will experiment, take risks and refrain from joining the crowd. They are the nonconformists, and they help to lead the way. An increase in female literacy could well trigger such a process.

Still other motives for procreation involve viewing children as productive assets. In a rural economy where avenues for saving are highly restricted, parents value children as a source of security in their old age. Mead Cain, previously at the Population Council, studied this aspect extensively. Less discussed, at least until recently, is another kind of motivation, explored by John C. Caldwell of the Australian National University, Marc L. Nerlove of the University of Maryland and Anke S. Meyer of the World Bank and by Karl-Göran Mäler of the Beijer International Institute of Ecological Economics in Stockholm and me. It stems from children's being valuable to their parents not only for future income but also as a source of current income.

Third World countries are, for the most part, subsistence economies. The rural folk eke out a living by using products gleaned directly from plants and animals. Much labor is needed even for simple tasks. In addition, poor rural households do not have access to modern sources of domestic energy or tap water. In semiarid and arid regions the water supply may not even be nearby. Nor is fuelwood at hand when the forests recede. In addition to cultivating crops, caring for livestock, cooking food and producing simple marketable products, members of a household may have to spend as much as five to six hours a day fetching water and collecting fodder and wood.

Children, then, are needed as workers even when their parents are in their prime. Small households are simply not viable; each one needs many hands. In parts of India, children between 10 and 15 years have been observed to work as much as one and a half times the number of hours that adult males do. By the age of six, children in rural India tend domestic animals and care for younger siblings, fetch water and collect firewood, dung and fodder. It may well be that the usefulness of each extra hand increases with declining availability of resources, as measured by, say, the distance to sources of fuel and water.

MIKE GOLDWATER *Matrix*

COLLECTING FIREWOOD is one way a brother and sister in Eritrea contribute needed labor to a family. Households throughout much of the Third World count on youngsters for a variety of tasks, such as herding cows and goats, looking after younger siblings, carrying water and searching for fuel and fodder. Older children often work as much as one and a half times the number of hours as men. Many are sold into "bonded labor," where they work to repay parents' debts.

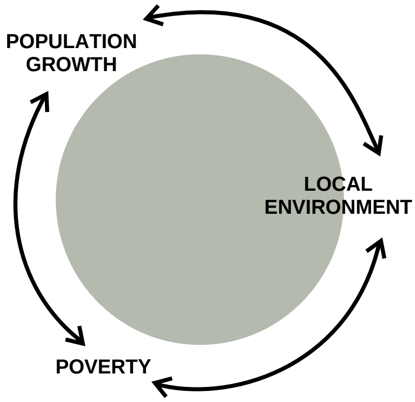
...But at a Hidden Cost

The need for many hands can lead to a destructive situation, especially when parents do not have to pay the full price of rearing their children but share those costs with the community. In recent years, mores that once regulated the use of local resources have changed. Since time immemorial, rural assets such as village ponds and water holes, threshing grounds, grazing fields, and local forests have been owned communally. This form of control enabled households in semiarid regions to pool their risks. Elinor Ostrom of Indiana University and others have shown that communities have protected such local commons against overexploitation by invoking norms, imposing fines for deviant behavior and so forth.

But the very process of economic development can erode traditional methods of control. Increased urbanization and mobility can do so as well. Social rules are also endangered by civil strife and by the takeover of resources by landowners or the state. As norms degrade, parents pass some of the costs of children on to the community by overexploiting the commons. If access to shared resources continues, parents produce too many children, which leads to greater crowding and susceptibility to disease as well as to more pressure on environmental resources. But no household, on its own, takes into account the harm it inflicts on others when bringing forth another child.

Parental costs of procreation are also lower when relatives provide a helping hand. Although the price of carrying a child is paid by the mother, the cost of rearing the child is often shared among the kinship. Caroline H. Bledsoe of Northwestern University and others have observed that in much of sub-Saharan Africa fosterage is commonplace, affording a form of insurance protection in semiarid regions. In parts of West Africa about a third of the children have been found to be living with their kin at any given time. Nephews and nieces have the same rights of accommodation and support as do biological offspring. In recent work I have shown that this arrangement encourages couples to have too many offspring if the parents' share of the benefits from having children exceeds their share of the costs.

In addition, where conjugal bonds are weak, as they are in sub-Saharan Africa, fathers often do not bear the costs of siring a child. Historical demographers, such as E. A. Wrigley of the University of Cambridge, have noted a significant difference between western Europe in the 18th century and modern preindustrial societies. In the former, marriage normally meant establishing a new household. This requirement led to late marriages; it also meant that parents bore the cost of rearing their children. Indeed, fertility rates in France dropped before mortality rates registered a decline, before modern family-planning techniques became available and before women became literate.



	TOTAL
SUB-SAHARAN AFRICA	6 TO 8
INDIA	4
CHINA	2.3
JAPAN AND WESTERN INDUSTRIAL DEMOCRACIES	1.5 TO 1.9

POVERTY, population growth and environmental degradation interact in a cyclic pattern (*top*). The chart (*bottom*) shows that fertility is higher in countries that are poorer.

The perception of both the low costs and high benefits of procreation induces households to produce too many children. In certain circumstances a disastrous process can begin. As the community's resources are depleted, more hands are needed to gather fuel and water for daily use. More children are then produced, further damaging the local environment and in turn providing the household with an incentive to enlarge. When this happens, fertility and environmental degradation reinforce each other in an escalating spiral. By the time some countervailing set of factors—whether public policy or diminished benefits from having additional children—stops the spiral, millions of lives may have suffered through worsening poverty.

Recent findings by the World Bank on sub-Saharan Africa have revealed positive correlations among poverty, fertility and deterioration of the local environment. Such data cannot reveal causal connections, but they do support the idea of a positive-feedback process such as I have described. Over time, the effect of this spiral can be large, as manifested by battles for resources [see "Environmental Change and Violent Conflict," by T. F. Homer-Dixon, J. H. Boutwell and G. W. Rathjens; SCIENTIFIC AMERICAN, February 1993].

The victims hit hardest among those who survive are society's outcasts—the migrants and the dispossessed, some of whom in the course of time become the emaciated beggars seen on the streets of large towns and cities in underdeveloped countries. Historical studies by Robert W. Fogel of the University of Chicago and theoretical explorations by Debraj Ray of Boston University and me, when taken together, show that the spiral I have outlined here is one way in which destitutes are created. Emaciated beggars are not lazy; they have to husband their precarious hold on energy. Having suffered from malnutrition, they cease to be marketable.

Families with greater access to resources are, however, in a position to limit their size and propel themselves into still higher income levels. It is my

Green Net National Production

Some economists believe population growth is conducive to economic growth. They cite statistics showing that, except in sub-Saharan Africa, food production and gross income per head have generally grown since the end of World War II. Even in poor regions, infant survival rate, literacy and life expectancy have improved, despite the population's having grown much faster than in the past.

One weakness of this argument is that it is based on statistics that ignore the depletion of the environmental resource base, on which all production ultimately depends. This base includes soil and its cover, freshwater, breathable air, fisheries and forests. No doubt it is tempting to infer from past trends that human ingenuity can be relied on to overcome the stresses that growing populations impose on the natural environment.

Yet that is not likely to be the case. Societies already use an enormous 40 percent of the net energy created by terrestrial photosynthesis. Geoffrey M. Heal of Columbia University, John M. Hartwick of Queens University and Karl-Göran Mäler of the Beijer International Institute of Ecological Economics in Stockholm and I have shown how to include environmental degradation in estimating the net national product, or NNP. NNP is obtained by deducting from gross national product the value of, for example, coal extracted or timber logged.

This "green NNP" captures not only present production but also the possibility of future production brought about by resources we bequeath. Viewed through NNP, the future appears far less rosy. Indeed, I know of no ecologist who thinks a population of 11 billion (projected for the year 2050) can support itself at a material standard of living of, say, today's representative American.

impression that among the urban middle classes in northern India, the transition to a lower fertility rate has already been achieved. India provides an example of how the vicious cycle I have described can enable extreme poverty to persist amid a growth in well-being in the rest of society. The Matthew effect—"Unto every one that hath shall be given, and he shall have abundance: but from him that hath not shall be taken away even that which he hath"—works relentlessly in impoverished countries.

Breaking Free

This analysis suggests that the way to reduce fertility is to break the destructive spiral. Parental demand for children rather than an unmet need for contraceptives in large measure explains reproductive behavior in developing countries. We should therefore try to identify policies that will change the options available to men and women so that couples choose to limit the number of offspring they produce.

In this regard, civil liberties, as opposed to coercion, play a particular role. Some years ago my colleague Martin R. Weale and I showed through statistical analysis that even in poor countries political and civil liberties go together with improvements in other aspects of life, such as income per person, life expectancy at birth and infant survival rate. Thus, there are now reasons for thinking that such liberties are not only desirable in themselves but also empower people to flourish economically. Recently Adam Przeworski of the University of Chicago demonstrated that fertility, as well, is lower in countries where citizens enjoy more civil and political freedom. (An exception is China, which represents only one country out of many in this analysis.)

LOUIS PSIHOYOS *Matrix*

DESTITUTES sleep in the Indian city of Bombay, having migrated from villages where spiraling poverty, population and environmental decay have made life impossible. In time, some of these dispossessed become the emaciated beggars and laborers common to urban areas in the Third World.

The most potent solution in semiarid regions of sub-Saharan Africa and the Indian subcontinent is to deploy a number of policies simultaneously. Family-planning services, especially when allied with health services, and measures that empower women are certainly helpful. As societal norms break down and traditional support systems falter, those women who choose to change their behavior become financially and socially more vulnerable. So a literacy and employment drive for women is essential to smooth the transition to having fewer children.

But improving social coordination and directly increasing the economic security of the poor are also essential. Providing cheap fuel and potable water will reduce the usefulness of extra hands. When a child becomes perceived as expensive, we may finally have a hope of dislodging the rapacious hold of high fertility rates.

Each of the prescriptions suggested by our new perspective on the links between population, poverty and environmental degradation is desirable by itself, not just when we have those problems in mind. It seems to me that this consonance of means and ends is a most agreeable fact in what is otherwise a depressing field of study.

FURTHER READING

POPULATION, NATURAL RESOURCES, AND DEVELOPMENT. Special issue of *Ambio,* Vol. 21, No. 1; February 1992.

AN INQUIRY INTO WELL-BEING AND DESTITUTION. Partha Dasgupta. Oxford University Press, 1993.

POPULATION: THE COMPLEX REALITY. Population Summit Report of the World's Scientific Academies, Royal Society, London. North American Press, 1994.

POPULATION, ECONOMIC DEVELOPMENT, AND THE ENVIRONMENT. Edited by Kerstin Lindahl Kiessling and Hans Landberg. Oxford University Press, 1994.

WORLD DEVELOPMENT REPORT. World Bank, annual publication.

POVERTY, INSTITUTIONS□ AND THE ENVIRONMENTAL RESOURCE BASE. Partha Dasgupta and Karl-Göran Mäler in *Handbook of Development Economics,* Vol. 3. Edited by T. N. Srinivasan et al. North Holland Publishing, Amsterdam (in press).

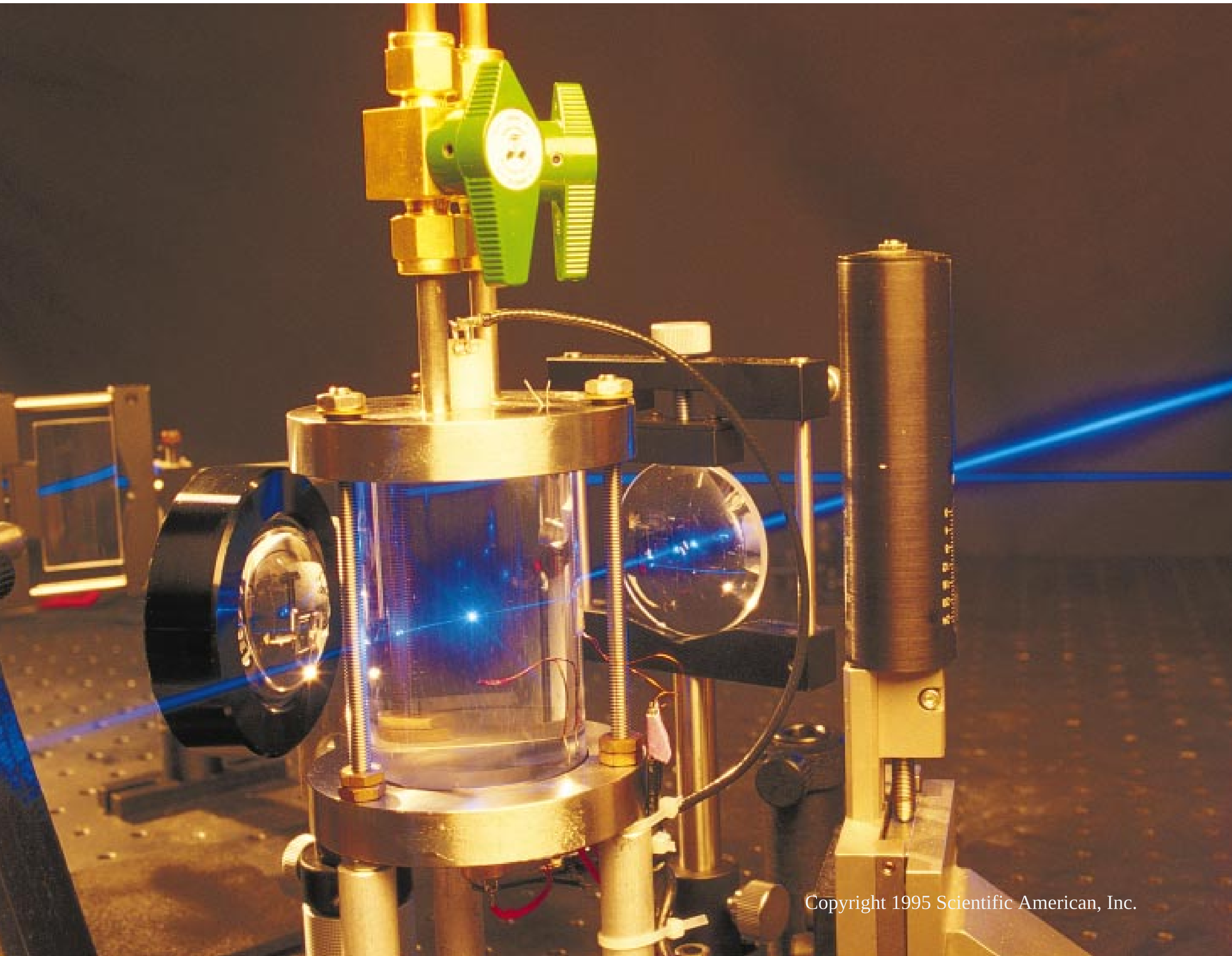
Sonoluminescence: Sound into Light

A bubble of air can focus acoustic energy a trillionfold to produce picosecond flashes of light. The mechanism eludes complete explanation

by Seth J. Putterman

Imagine you are riding a roller coaster. First, you chug up a long incline slowly. When you get to the top, your car free-falls, speeding up until it reaches the bottom of the drop, where the deceleration crams you into your seat. That sensation is what you would feel if you were riding a pulsating bubble of air trapped in water—except that the drop would reach supersonic speeds and at the bottom you would be crushed into your seat with a force equal to 1,000 billion times your weight.

Obviously, more than your stomach would react to such a ride. As for the bubble, it responds to the extraordinary force by creating a flash of light only a tiny fraction of a second long. The light



is mostly ultraviolet, which indicates that when the bubble's free fall stops, its interior becomes much hotter than the surface of the sun. A sound wave can make the bubble repeat this wild ride more than 30,000 times a second, so that the flashes burst out with clocklike regularity.

In sonoluminescence—as the process of converting sound into light is called—the bubble is concentrating the energy of the acoustic vibrations by a factor of one trillion. That is, the sound wave that drives the bubble is centimeters long, but the light is emitted from a region of atomic dimensions.

A detailed accounting of this inexpensive yet unusual illumination source remains elusive. The flashes are so brief that to measure the properties of light, we must use photodetectors that respond more quickly than those employed by high-energy physicists. (In fact, sonoluminescence is the only means of generating picosecond flashes of light that does not require expensive lasers.) The physical process by which sonoluminescence achieves such a huge focusing of energy may serve as a useful model for researchers seeking to develop controlled nuclear fusion. Current attempts to fathom the mysteries of sonoluminescence in my laboratory at the University of California at Los Angeles and in other institutions are generating new paradoxes faster than the existing questions can be answered.

SETH J. PUTTERMAN received his Ph.D. at the Rockefeller University in 1970 before joining the faculty of the University of California, Los Angeles. His research interests include turbulence, superfluidity and the quantum mechanics of single atoms. He writes that he is indebted to his student Ritva Löfstedt for valuable assistance in formulating the ideas in this article and thanks the U.S. Department of Energy for supporting the sonoluminescence research.

Skeptical Inquirer

I was actually quite incredulous of sonoluminescence when I first heard about it in the mid-1980s from my scholarly colleague Thomas Erber of the Illinois Institute of Technology. One day at the U.C.L.A. coffeehouse, he taunted me about my long-standing interest in fluid mechanics, focusing on the Navier-Stokes equations, which describe the flow of fluids. He asked, "If you think the Navier-Stokes equations are so great, then please explain to me how sound can be made into light." Based on my intuition, I replied that I did not believe sonoluminescence was possible. But he insisted that this effect had been documented some time ago. So along with Ritva Löfstedt, who was then a U.C.L.A. undergraduate, I went back through the old papers to see if sonoluminescence was for real.

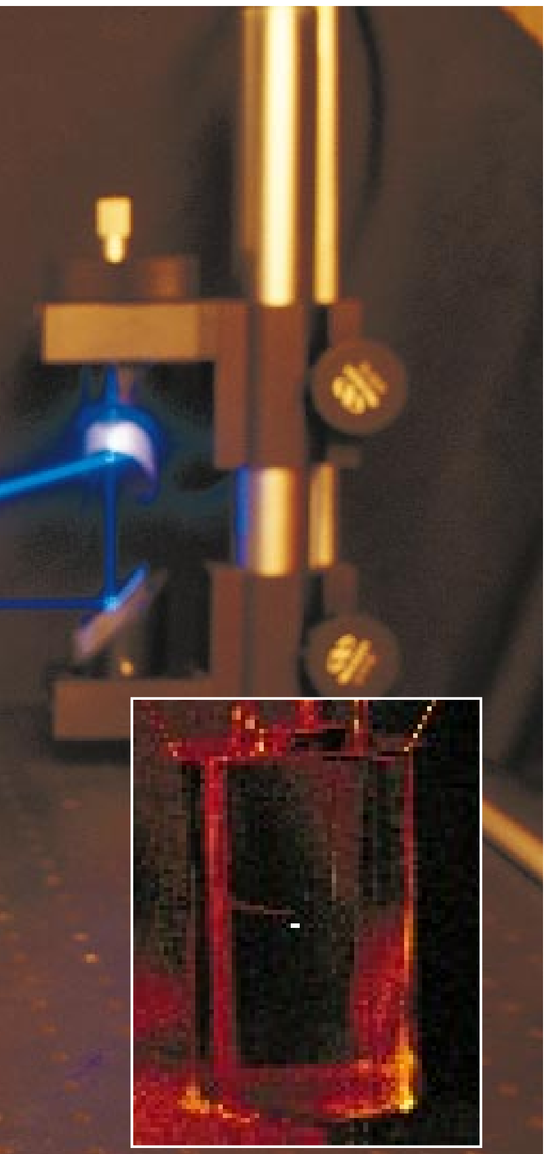
In the 1920s and 1930s, we learned, chemists working with loudspeakers developed for sonar systems during World War I came across an interesting phenomenon: a strong sound field could catalyze reactions that take place in an aqueous solution. A German scientist, Reinhard Mecke of the University of Heidelberg, then commented to his coworkers that the amount of energy needed for a chemical reaction is the same as that needed to excite the emission of light from an atom. So he suggested a search for such a signal. Soon afterward, in 1934, H. Frenzel and H. Schultes of the University of Cologne discovered sonoluminescence in a bath of water excited by acoustic waves.

Perhaps it was the common observation that one can generate a spark of light by touching a doorknob after walking on a carpet. Whatever their inspiration, Frenzel and Schultes explained the light emission in terms of *Reibungselektrizität,* or "frictional electricity." In their experiment the sound wave initiated the process of cavitation—the growth and collapse of bubbles in water. They pictured the bubbles' motion through the liquid as analogous to that of shoes shuffling on a carpet. The abrasion causes electrical charges to separate in the originally neutral media. A spark releases the built-up charge. Then they concluded their paper by saying they had more important matters to attend to.

Other researchers, seeking clues to the mechanism, proceeded to carry out spectral measurements of this new light source. These studies were inconclusive because of the transient nature of the phenomenon. The strong sound fields they used created clouds of bubbles that grew, collapsed and gave off light in an unpredictable and unsynchronized manner.

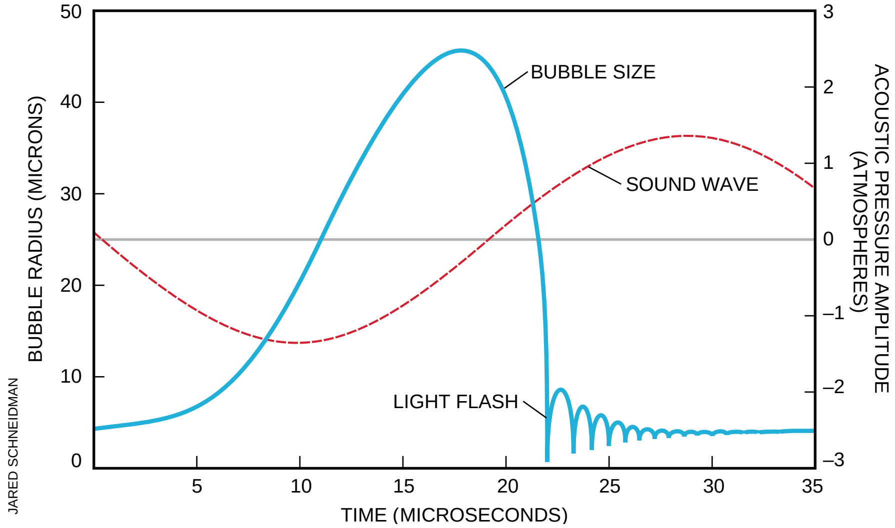
At U.C.L.A., Bradley P. Barber, a graduate student, and I became enthusiastic about characterizing and understanding the mechanism responsible for sonoluminescence. I learned that other investigators had just succeeded in trapping a single, light-emitting bubble in water that was partially degassed. They were D. Felipe Gaitan, now at the Naval Postgraduate School, and Lawrence A. Crum, now at the University of Washington. It seems that my enthusiasm for their advance far exceeded theirs. They had dismantled the experiment and abandoned this avenue of research. But they did show us how to adjust our apparatus to find single-bubble sonoluminescence.

So with a boiling flask from the chemistry laboratory, an oscilloscope from the undergraduate lab, my home stereo and a photomultiplier tube (light sensor) purchased with my credit card, we were off and running [see "The Ama-



ED KASHI; SETH J. PUTTERMAN AND ROBERT A. HILLER (*INSET*)

MAKING LIGHT OF SOUND is accomplished by a bubble of air trapped in a cylindrical flask of degassed water. Sound from speakers above and below the flask trap the bubble. A flash of light 50 picoseconds long emerges during the compression part of the acoustic wave. A laser measures the bubble size as it pulses in time with the sound. The light emission itself is rather faint (*inset*).



ROLLER-COASTER RIDE of a pulsating bubble lags slightly behind the expansion and compression of sound waves. The bubble expands to its maximum radius just after the acoustic pressure becomes negative. During compression, the bubble rapidly shrinks to less than one micron in radius and emits a flash of light. The bubble continues to swell and contract briefly before settling down.

teur Scientist," page 96]. For some of our initial work, we injected an air bubble into water with a syringe. Over the years we have refined our setup. Our current apparatus consists of a piezoelectric transducer on the top of a cylindrical flask filled with water. The transducer is a ceramic material that turns an oscillating voltage into a mechanical vibration and thereby sets up sound waves—alternating fields of compression and expansion—in the water. Submerged in the water is a small piece of toaster wire. When current flows through it, the wire heats up, boiling the water nearby. As a result, a bubble filled with water vapor forms. Before the vapor recondenses, air dissolved in the water flows into the pocket to create an air bubble.

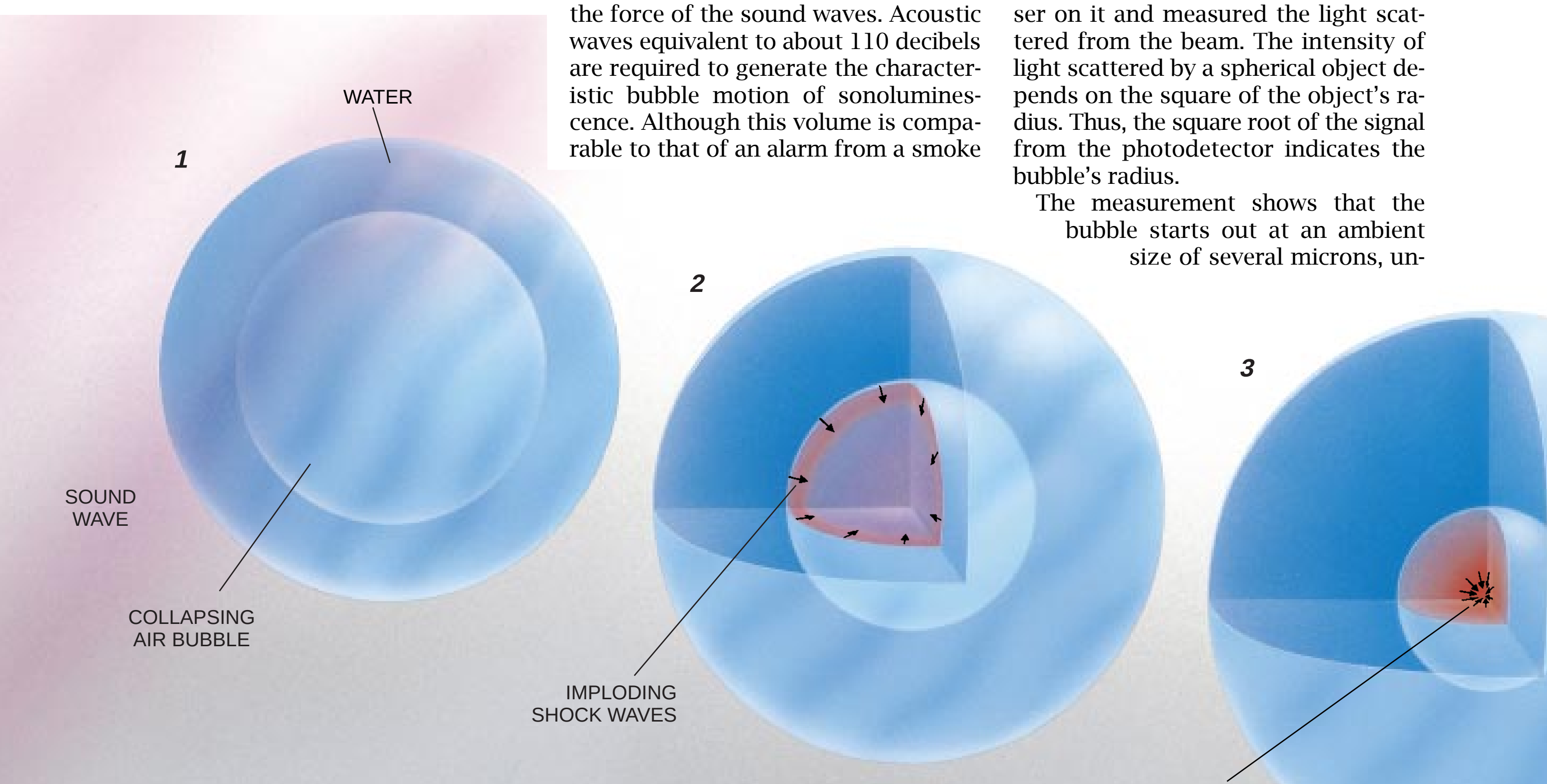
This bubble is then trapped at the center of the cylindrical flask, where the buoyancy force that would make the bubble rise to the top is balanced by the force of the sound waves. Acoustic waves equivalent to about 110 decibels are required to generate the characteristic bubble motion of sonoluminescence. Although this volume is comparable to that of an alarm from a smoke detector a few centimeters away, the frequency of the sound lies just beyond the range of human hearing.

Probing the Bouncing Bubble

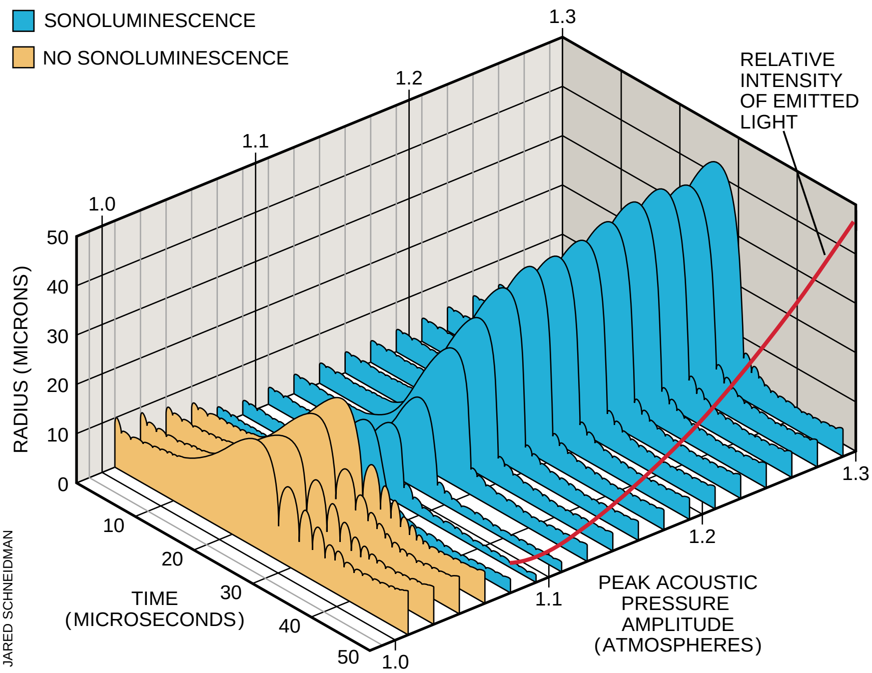
As physicists attempting to characterize sonoluminescence, our first goal was to identify the time scales involved in the process—specifically, the duration of the flash. We were amazed to find that such a measurement would require the use of the fastest known light sensors. Our analysis yielded an upper bound of about 50 picoseconds. We also found that the flashes came out with an incredible regularity. The timing between consecutive flashes, typically about 35 microseconds, varies by no more than 40 picoseconds.

To determine the radius of the sonoluminescing bubble, Barber shone a laser on it and measured the light scattered from the beam. The intensity of light scattered by a spherical object depends on the square of the object's radius. Thus, the square root of the signal from the photodetector indicates the bubble's radius.

The measurement shows that the bubble starts out at an ambient size of several microns, un-



COLLAPSING BUBBLE may heat up in stages. The first instance comes during the increase in pressure as a sound wave causes the bubble to shrink (*1*). The second occurs as the walls collapse at supersonic speeds, which in turn launches shock waves toward the bubble center (*2*). The shock further heats the interior. Even more heat is generated as the shock waves focus (*3*) and then explode out (*4*).



JARED SCHNEIDMAN

TRANSITION TO SONOLUMINESCENCE happens when the sound level reaches a critical state. The average radius of a bubble generally increases with a rise in acoustic amplitude. At the level at which sonoluminescence begins, however, the radius suddenly shrinks. The mechanism behind this transition is not understood.

til the expansion part of the sound field acts on it. Then the pressure drops, putting the fluid under tension and causing the bubble to swell to about 50 microns. The expansion continues until the sound field switches from rarefaction to compression.

At this point of maximum expansion, a near-vacuum has formed inside the bubble. That is because the volume of the bubble has greatly increased, but the number of molecules inside it has not changed. Atmospheric pressure, however, still acts on the outside of the bubble. The pressure difference between the inside and the outside leads to a catastrophic collapse. The bubble decreases from its 50-micron maximum radius to a size of about 0.5 micron. At that point, the surface of the bubble stops its inward rush as though it had suddenly slammed into a wall. It cannot become any smaller because of the repulsive force between the gas atoms and molecules. (We say at this point the size of the bubble is determined by the van der Waals forces of the hard core of its contents.) The light flash comes out as the bubble decelerates through its minimum radius [*see top illustration on opposite page*]. After the light emission, the bubble elastically bounces in size a few times and then sits dead in the water waiting for the next helping of sound.

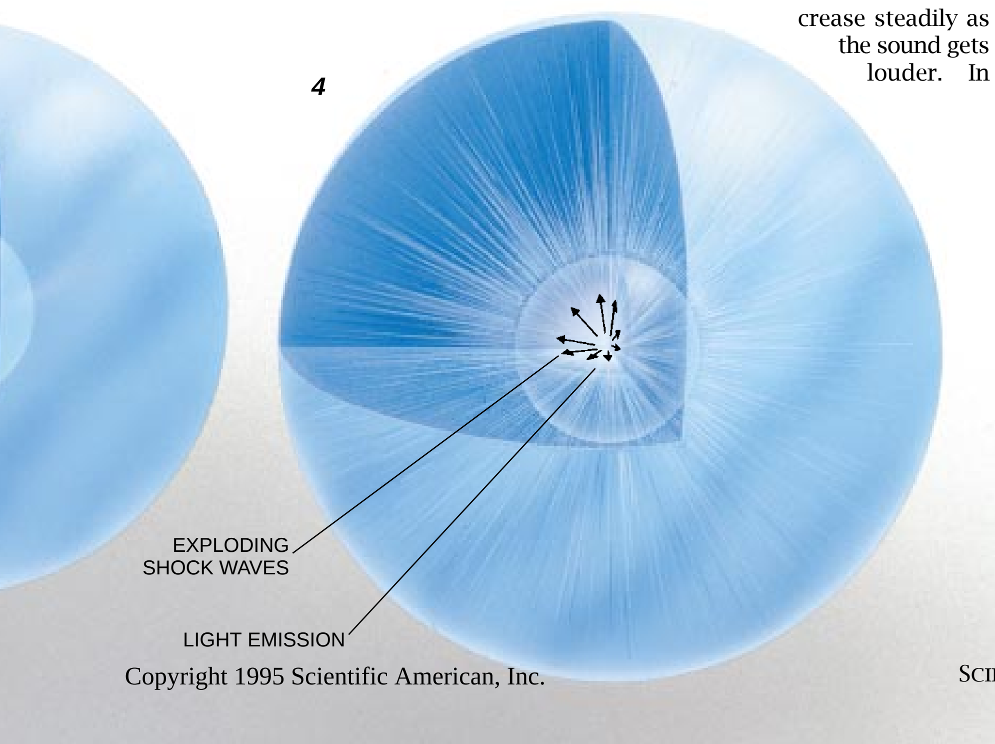
Although experiments can measure the size of the bubble, no theory can explain how those particular radii come about. The size of a bubble depends on the amount of gas trapped inside. Löfstedt, now one of my graduate students, and I are studying the mechanism whereby the gas dissolved in the surrounding water diffuses into the bubble. When the bubble is large, the pressure inside it is low; therefore, gas flows into it from the surrounding fluid. When the bubble is small, the reverse occurs. The balance between inflow and outflow of air molecules determines the average bubble size.

The radius of a bubble driven by a weak sound field seems to follow the predictions of this model. But applying the same reasoning to a high-amplitude, light-emitting bubble leads to a contradiction with the data [*see illustration above*]. The average radius of the bubble should be seen to increase steadily as the sound gets louder. In practice, however, this relation has an unusual discontinuity just as sonoluminescence sets in: the average radius suddenly decreases for a moment. Beyond that point, it rises with sound amplitude again. Some new (and as yet unknown) mass-flow mechanism must determine the sonoluminescent state.

Torrid Interior

To the unaided eye, the faint blue glow of a sonoluminescing bubble looks like a star in the night sky. In 1991 my graduate student Robert A. Hiller determined how much of this radiated light lies in the visible part of the spectrum. He found that there is literally more to the spectrum than meets the eye. The results showed that the bubble emits more purple light than red and more ultraviolet than purple. We could not follow the spectrum beyond photon energies of six electron volts, corresponding to an ultraviolet light wavelength of 0.2 micron, because above those energies light cannot propagate through water. (For the same reason, we had to construct our flask from quartz rather than plain glass, which blocks ultraviolet light.) An energy of six electron volts corresponds to a temperature of 72,000 kelvins, so the interior of the bubble must be scorching indeed.

That a collapsing bubble of gas becomes very hot can be explained in terms of an everyday experience for res-



TOMO NARASHIMA

idents of southern California and the Alps. These people suffer through particularly torrid weather when the wind blows from higher elevations to lower ones. In southern California, a "Santa Ana" condition occurs when air from the high desert heats up by 15 degrees Celsius as it blows into the Los Angeles basin. The sudden temperature increase results from the work performed by the atmosphere on the desert air mass as the air drops in altitude by 5,000 feet on its way to the ocean. At lower altitudes, the barometric pressure is higher. If the pressure difference compresses the air before it has time to exchange its heat with the ocean or other cooler bodies, the air becomes adiabatically heated—that is, its temperature rises without the addition of any heat energy.

The hot spot realized in a collapsing bubble is astronomical even when compared to a sizzling day in California. The volume of a sonoluminescent bubble drops by a factor of one million as its radius decreases 100-fold. In the 1950s B. E. Noltingk and E. A. Neppiras of Mullard Electronic Research Laboratory in Surrey, England, calculated that the resulting adiabatic compression of the bubble interior leads to a temperature of up to 10,000 kelvins and pressures greater than 10,000 atmospheres. (The bubble surface does not vaporize, perhaps because the high rate of pressurization and heating takes place well inside the bubble.)

Had the revered English physicist Lord Rayleigh lived in southern California, his experience with the weather might have led him to predict sonoluminescence as part of the bubble research that he carried out in 1917. The Royal Navy hired him to help understand the causes of the degeneration of ship propellers. Rayleigh determined that the small bubbles of air created as the propeller sliced through the water were the culprits. The bubbles would corrode the propeller as they collapsed onto it with a force greater than 10,000 atmospheres. But in describing the motion of the bubbles, he assumed the collapse of a bubble obeyed Boyle's law: in other words, he thought the temperature inside it remained constant. Had he realized that the collapse is so rapid that it is adiabatic, he surely would have predicted high temperatures and the associated light emission.

Exactly how would the high temperature produce light? According to researchers who study sonoluminescence and sonochemistry, the energy from the collapse is powerful enough to break apart molecules within bubbles. The dissociated molecules emit light as they recombine. This effect, referred to as chemiluminescence, was first presented by Virginia F. Griffing of Catholic University in 1952. It accompanies transient cavitation and has been used to initiate unusual chemical processes. An example is the fabrication of amorphous iron by Kenneth S. Suslick of the University of Illinois [see "The Chemical Effects of Ultrasound," by Kenneth S. Suslick; SCIENTIFIC AMERICAN, February 1989].

Although adiabatic heating of a collapsing bubble provides an impressive mechanism for energy concentration, it

Shock Waves in Bubbles

In the past, researchers who studied sonoluminescence and sonochemistry associated the transient clouds of cavitating bubbles with hot spots that formed within each bubble. In this traditional model the energy focused by the collapse of the bubbles creates dissociated molecules that emit light as they recombine.

Yet the prevailing wisdom about transient cavitation cannot explain the strongly ultraviolet spectrum emitted by a single bubble synchronized to the sound field. Our measurements indicate that the bubble's interior attains a temperature substantially higher than 10,000 kelvins. This value can be reached if the collapse of a single synchronized bubble is so fast and symmetrical that it launches a spherical shock wave into its interior. As the imploding shock wave of radius R_s focuses, its amplitude and speed increase. For this case, the solution to the equations of hydrodynamics takes the form

$$R_s = At^b$$

where A is a constant, time t is measured from the moment of focusing when $R_s = 0$, and b is 0.7 for air.

A Mach number is associated with every shock wave. This number represents the ratio of the shock velocity to the ambient speed of sound. The temperature behind a shock front is higher than that in front of it; the ratio of those temperatures is proportional to the square of the Mach number.

For an imploding air bubble, the Mach number approaches infinity as the shock front moves closer to the focal point, which means that a tremendous amount of heating takes place. Furthermore, when the shock hits the center and explodes outward, the molecules that were behind the shock are suddenly in front of it again. The hot molecules are hit a second time, and their temperatures go up by another factor of the square of the Mach number.

Temperatures that can be reached by this mathematical model are therefore unimaginably high. In reality, they are limited by the stability of the shock front. In the shock-wave model, sonoluminescence hinges on the shock front remaining spherical down to a radius of about 0.1 micron.

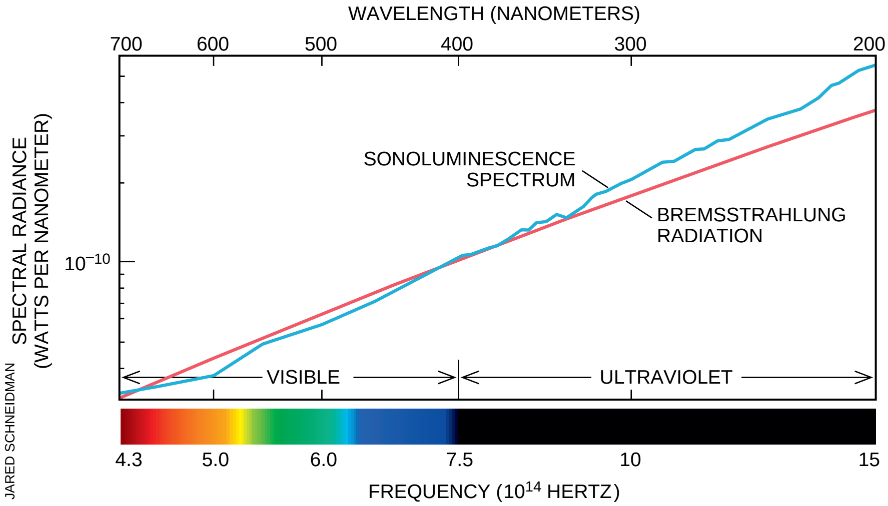
Spherical shock fronts played an important role in the design of nuclear weapons. British physicist Geoffrey I. Taylor used photographs and the corresponding expression for exploding shocks to obtain an unauthorized calibration of early hydrogen bomb tests. In the future, our understanding of how a sonoluminescing sound field generates such a beautifully spherical collapse could assist researchers at such institutions as Lawrence Livermore National Laboratory and the University of Rochester in designing improved versions of inertial confinement fusion. In this fusion process, huge lasers induce the implosion of a small pellet containing a mixture of the hydrogen isotopes deuterium and tritium. The spherical implosion is the key to reaching temperatures and densities sufficient to realize the fusion of these hydrogen nuclei to yield helium and neutrons.

There is a highly speculative chance that the comparison between inertial confinement fusion and sonoluminescence may indeed reveal a deeper similarity. If the sonoluminescent shock remains stable down to the incredibly small radius of 10 nanometers, then this tiny region would also reach temperatures appropriate to fusion. It is not hard to imagine many effects that would stand in the way of such an outcome—instability of the shock wave, thermal diffusion and radiation damping, to name a few. Given the shortcomings of current models, we bet that this issue cannot be decided by computer simulation. Only future experiments can tell if the interior of the bubble gets as hot as the interior of the sun.

cannot be the only or complete answer. Such heating alone would not be able to generate the largely ultraviolet spectrum we observed. Therefore, an additional stage of energy amplification must take place. Barber and I deduced a plausible mechanism. We realized that the supersonic speeds of the collapsing bubble could launch shock waves into the bubble's interior. Although the bubble's motion is arrested by the forces of the gas molecules against one another, the imploding shock wave could continue inward and further concentrate the energy of the collapse.

Our U.C.L.A. colleagues Paul H. Roberts and Cheng-Chin Wu also realized the potential importance of shock waves in sonoluminescence. They calculated the extent of the concentration. Building on a solution first developed in the 1940s by the German mathematician Karl G. Guderley of the Institute for Gas Dynamics in Braunschweig, they showed that the bubble's collapse could launch a shock wave into the bubble that becomes stronger as the shock implodes. The high temperature and pressure associated with this shock front become even more amplified when the converged shock subsequently explodes outward [*see box on opposite page*].

Typically, shock fronts are susceptible to instabilities that corrugate their surfaces, which thereby limit the extent of the implosion. If the inward-moving shock front launched by the bubble remains intact to a radius of 0.1 micron from the center of the bubble, the temperature near it would be 100,000 kelvins. This heat is about that required for the strongly ultraviolet spectrum we observed. If the shock front survives down to 20 nanometers, the temperature would reach one million kelvins, hot enough to make soft (relatively long wavelength) x-rays. Such photons do not propagate through water, so we do not know whether they are there. The possibility of getting weak x-rays from sound might seem far-fetched, and I am skeptical of such an outcome. Then again, I was quite doubtful of sonoluminescence in the first place.



SPECTRUM of sonoluminescence shows that most of the emitted light is ultraviolet. As pointed out by Paul H. Roberts and Cheng-Chin Wu of the University of California at Los Angeles, the signal compares closely with bremsstrahlung radiation—that is, light emitted by a plasma at 100,000 kelvins.

Noble Addition

Although the mechanism of sonoluminescence from a single bubble is difficult to explain, the phenomenon is easy to produce and modify. Despite being a robust phenomenon, it is highly sensitive to controllable experimental parameters, such as the intensity of the sound and the temperature of the water. For instance, the amount of light emitted with each flash increases by a factor of 200 as the temperature drops from 35 to 0 degrees Celsius. At 0 degrees, the bubble gives off about 10 million photons per flash.

The sensitivity to temperature suggested that we could learn more about sonoluminescence by changing other quantities. We attempted to find single-bubble sonoluminescence in liquids other than water, but without success. Because we could not change the driving fluid, we tried changing the gas in the bubble. This alteration entailed degassing the water by exposing it to a vacuum, a process that removes the dissolved air. Then we dissolved other gases of our choice into the water. Obviously, this procedure had to be carried out in an airtight system. Hiller, who built the apparatus, first used it to make pure nitrogen bubbles. He anticipated that their properties would be similar to air, which is 80 percent nitrogen.

To our surprise, pure nitrogen bubbles made hardly any light. We therefore expected that oxygen would prove quite amazing. But again, we found that a pure oxygen bubble was very dim. Similarly, an 80–20 mixture of nitrogen and oxygen was a weak emitter. So was gas from a liquid-air canister! We anxiously searched for our stupid mistake.

In fact, the measurements were good. Air is 1 percent argon, and argon is removed from commercial liquid air. Adding argon back boosted the light intensity. Helium and xenon also worked, although each noble gas produces a unique spectrum. A small gas impurity of about 1 percent seems to be the key to sonoluminescence. We do not yet know why that is the optimal amount.

In view of our experimental results, what do we understand about sonoluminescence? First and foremost, we are dealing with a "virtuoso" sound field, one that positions a gas bubble at just the right location to act on it symmetrically and with maximum force. The theory of an adiabatic compression followed by an imploding shock wave provides an appealing picture that is helping to guide research.

Still, this working model must be viewed as tentative, because it fails to explain so many unknowns. These mysteries include the size of the bubble, the role of inert gases and the mechanism of light emission. Most important, theory and experiment have failed to determine the limit of energy focusing that can be achieved. Surely the mechanism is nature's most nonlinear system, yet it can be controlled and made free of chaos. The joy of this problem is that the effect is so robust but so sensitive that whenever we change a parameter we find wonderful new physics.

FURTHER READING

OBSERVATION OF SYNCHRONOUS PICOSECOND SONOLUMINESCENCE. Bradley P. Barber and Seth J. Putterman in *Nature*, Vol. 352, No. 6333, pages 318–323; July 25, 1991.

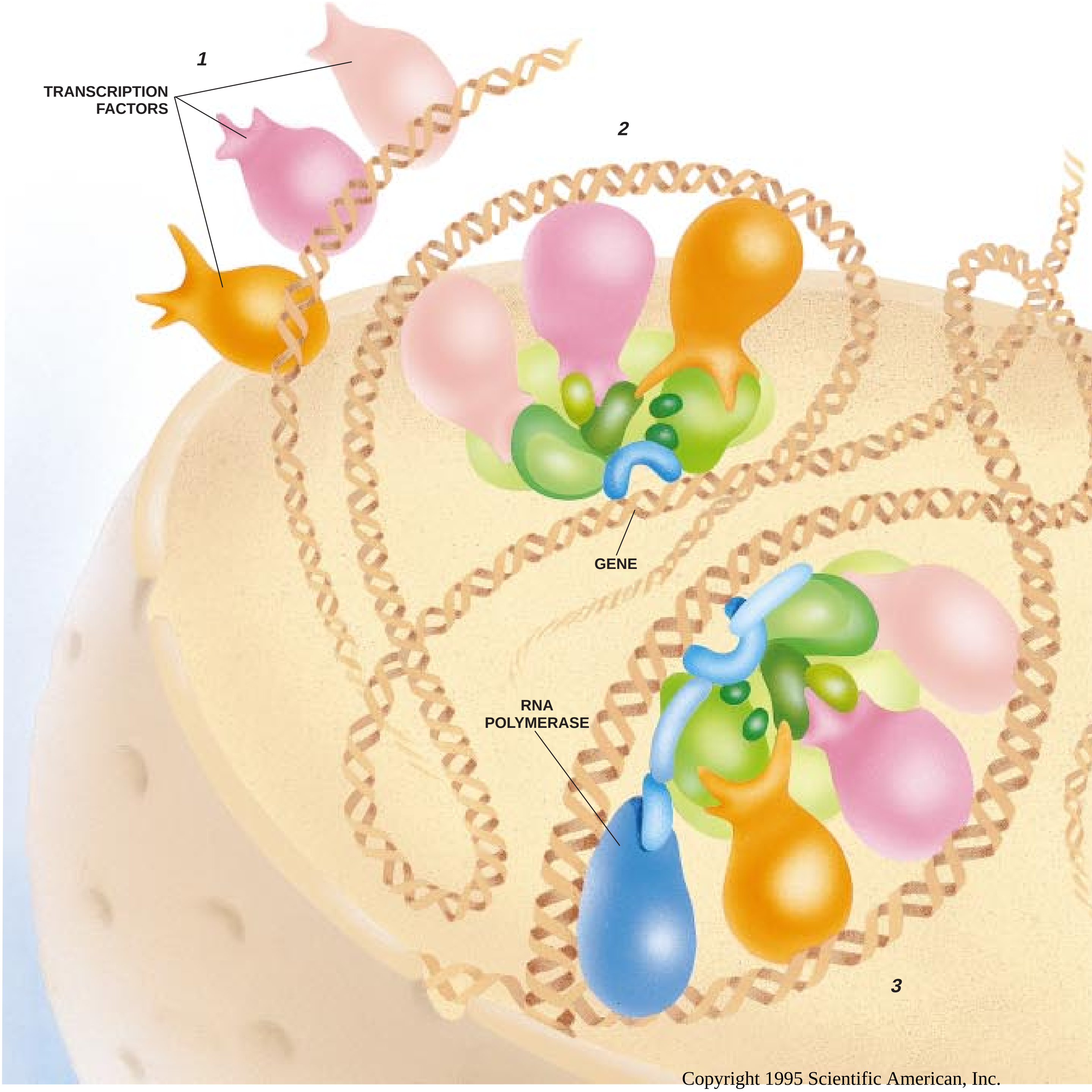
SENSITIVITY OF SONOLUMINESCENCE TO EXPERIMENTAL PARAMETERS. Bradley P. Barber, C. C. Wu, Ritva Löfstedt, Paul H. Roberts and Seth J. Putterman in *Physical Review Letters*, Vol. 72, No. 9, pages 1380–1383; February 28, 1994.

EFFECT OF NOBLE GAS DOPING IN SINGLE-BUBBLE SONOLUMINESCENCE. Robert Hiller, Keith Weninger, Seth J. Putterman and Bradley P. Barber in *Science*, Vol. 266, pages 248–250; October 14, 1994.

Molecular Machines That Control Genes

The activities of our genes are tightly regulated by elaborate complexes of proteins that assemble on DNA. Perturbations in the normal operation of these assemblies can lead to disease

by Robert Tjian

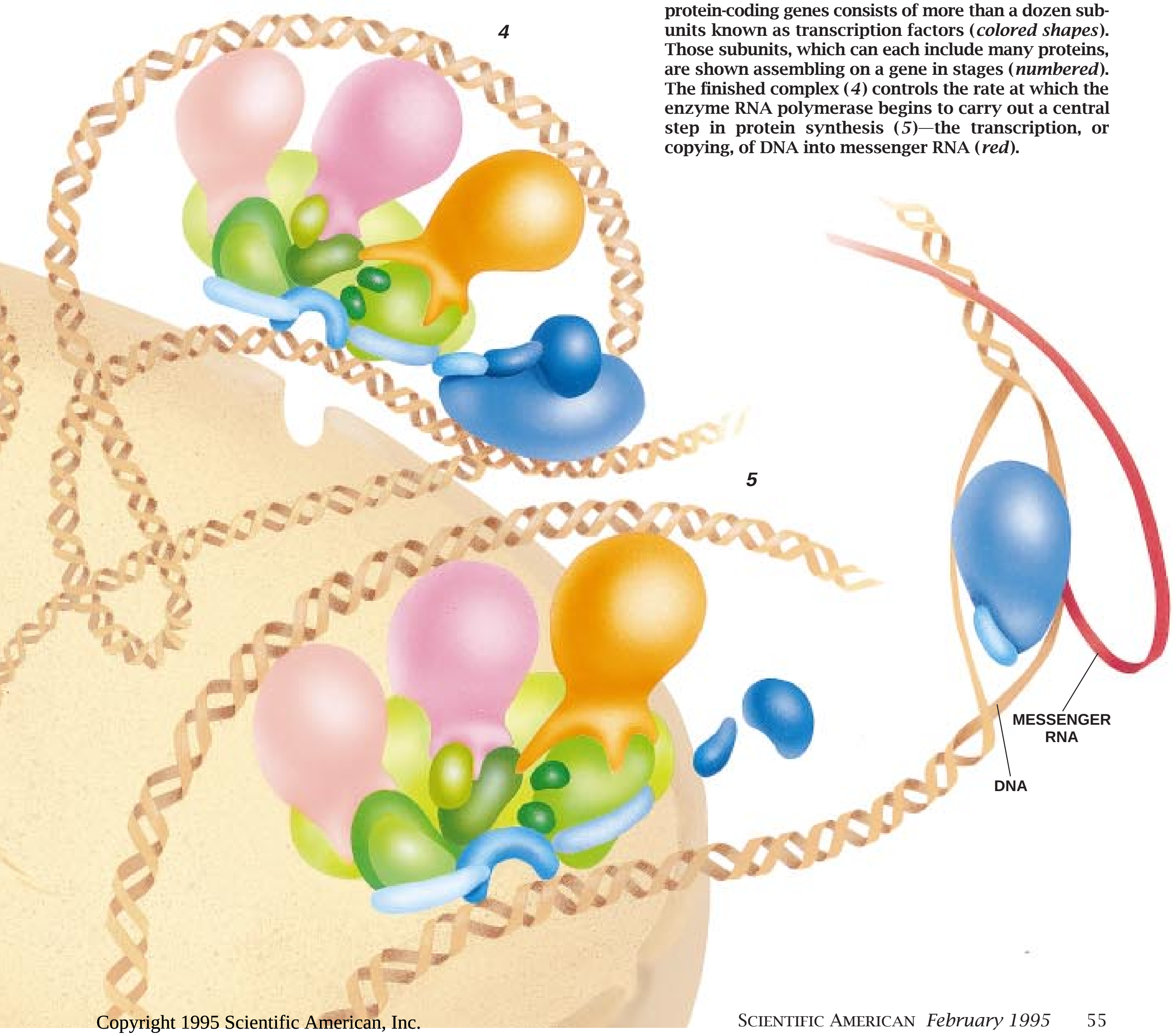


Asthma, cancer, heart disease, immune disorders and viral infections are seemingly disparate conditions. Yet they turn out to share a surprising feature. All arise to a great extent from overproduction or underproduction of one or more proteins, the molecules that carry out most reactions in the body. This realization has recently lent new urgency to research aimed at understanding, and ultimately manipulating, the fascinating biochemical machinery that regulates an essential step in protein synthesis: the transcription of genes. For a protein to be generated, the gene that specifies its composition must be transcribed, or copied, from DNA into strands of messenger RNA, which later serve as the templates from which the protein is manufactured.

Even before therapy became a goal, transcription had long captivated scientists for another reason: knowledge of how this process is regulated promises to clarify some central mysteries of life. Each cell in the body contains the same genome, the complement of some 150,000 genes that form the blueprint for a human being. How is it that the original cell of an organism—the fertilized egg—gives rise to a myriad of cell types, each using somewhat different subsets of those genes to produce different mixtures of proteins? And how do the cells of a fully formed body maintain themselves, increasing and decreasing the amounts of proteins they manufacture in response to their own needs and those of the larger organism?

To answer these questions and design

ROBERT TJIAN, who was born in Hong Kong, is an investigator with the Howard Hughes Medical Institute and professor of molecular and cell biology at the University of California, Berkeley. He earned his Ph.D. in biochemistry and molecular biology from Harvard University in 1976 and conducted research at the Cold Spring Harbor Laboratory in New York State before joining the faculty of Berkeley in 1979.



MOLECULAR MACHINERY that regulates the activity of protein-coding genes consists of more than a dozen subunits known as transcription factors (*colored shapes*). Those subunits, which can each include many proteins, are shown assembling on a gene in stages (*numbered*). The finished complex (*4*) controls the rate at which the enzyme RNA polymerase begins to carry out a central step in protein synthesis (*5*)—the transcription, or copying, of DNA into messenger RNA (*red*).

drugs able to modulate transcription, investigators need to know something about the makeup of the apparatus that controls reading of the genetic code in human cells. After some 25 years of exploration, the overall structure of that apparatus is becoming clear. Work in my laboratory at the University of California at Berkeley and at other institutions has revealed that one part of the apparatus—the engine driving transcription of most, if not all, human genes—consists of some 50 distinct proteins. These proteins must assemble into a tight complex on DNA before a special enzyme, RNA polymerase, can begin to copy DNA into messenger RNA. The putative constituents have now been combined in the test tube to yield a fully operational transcription engine. Still other proteins essentially plug into receptive sockets on the engine and, in so doing, "program" it, telling it which genes should be transcribed and how quickly. Critical details of these interactions are emerging as well.

Clues from Bacteria

When my colleagues and I at Berkeley began focusing on human genes in the late 1970s, little was known about the transcription machinery in our cells. But studies begun early in that decade had provided a fairly clear picture of transcription in prokaryotes—bacteria and other primitive single-celled organisms that lack a defined nucleus. That work eventually lent insight into human and other eukaryotic (nucleated) cells and helped to define features of transcription that hold true for virtually all organisms.

The bacterial research showed that genes are essentially divided into two functionally distinct regions. The coding region specifies the sequence of amino acids that must be linked together to make a particular protein. This sequence is spelled out by the nucleotides (the building blocks of DNA) in one strand of the DNA double helix; the nucleotides are distinguished from one another by the nitrogen-rich base they carry—adenine (*A*), thymine (*T*), cytosine (*C*) or guanine (*G*). The other region of a gene has regulatory duties. It controls the rate at which RNA polymerase transcribes the coding region of a gene into messenger RNA.

In bacteria, as in most prokaryotes, the regulatory region, called the promoter, resides within a stretch of nucleotides located a short distance—often as few as 10 nucleotides—in front of (upstream from) the start of the coding region. For transcription to proceed accurately and efficiently, RNA polymerase must attach to the promoter. Once it is so positioned, it slides over to the start of the coding region and rides along the DNA, like a train on a track, constructing an RNA replica of the coding sequence. Except in very long genes, the number of RNA molecules made at any moment depends mainly on the rate at which molecules of RNA polymerase attach to the promoter and initiate transcription.

Interestingly, RNA polymerase is a rather promiscuous molecule, unable to distinguish between the promoter and other DNA sequences. To direct the enzyme to promoters of specific genes, bacteria produce a variety of proteins, known as sigma factors, that bind to RNA polymerase. The resulting complexes are able to recognize and attach to selected nucleotide sequences in promoters. In this way, sigma factors program RNA polymerase to bypass all nonpromoter sequences and to linger only at designated promoters.

Considering the importance of sigma factors to the differential activation of genes in bacteria, my colleagues and I began our inquiry into the human transcription apparatus by searching for sigmalike molecules in human cells. But we had underestimated the complexity of the machinery that had evolved to retrieve genetic information from our elaborate genome. It soon became apparent that human sigma factors might not exist or might not take the same form as they do in bacteria.

Surprising Complexity

If there were no simple sigma factors in eukaryotes, how did such cells ensure that RNA polymerase transcribed the right genes at the right time and at the right rate? We began to see glimmerings of an answer once the unusual design of eukaryotic genes was delineated.

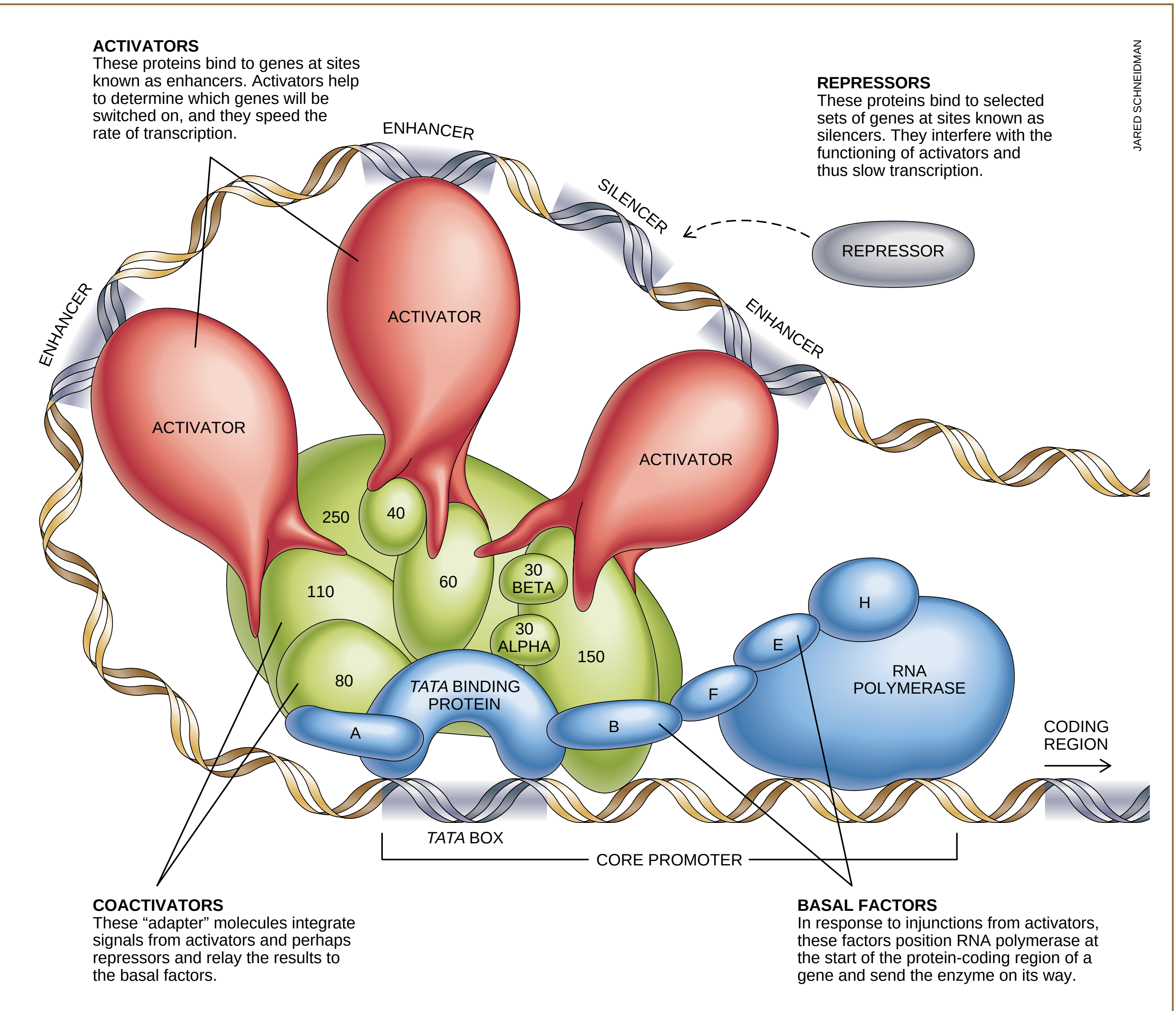
By 1983 investigators had established that three kinds of genetic elements, consisting of discrete sequences of nucleotides, control the ability of RNA polymerase to initiate transcription in all eukaryotes—from the single-celled yeast to complex multicellular organisms. One of these elements, generally located close to the coding region, had been found to function much like a bacterial promoter. Called a core promoter, it is the site from which the polymerase begins its journey along the coding region. Many genes in a cell have similar core promoters.

Walter Schaffner of the University of Zurich and Steven Lanier McKnight of the Carnegie Institution of Washington, among others, had additionally identified an unusual set of regulatory elements called enhancers, which facilitate transcription. These sequences can be located thousands of nucleotides upstream or downstream from the core promoter—that is, incredibly far from it. And subsequent studies had uncovered the existence of silencers, which help to inhibit transcription and, again, can be located a long distance from the core promoter.

In a somewhat imperfect analogy, if the core promoter were the ignition switch of a car engine, enhancers would act as the accelerator, and silencers as the brakes. Eukaryotic genes can include several enhancers and silencers, and two genes may contain some identical enhancer or silencer elements, but no two genes are precisely alike in the combination of enhancers and silencers they

Anatomy of the Transcription Apparatus

The molecular apparatus controlling transcription in human cells consists of four kinds of components. Basal factors (*blue shapes at bottom*), generally named by single letters, are essential for transcription but cannot by themselves increase or decrease its rate. That task falls to regulatory molecules known as activators (*red*) and repressors (*gray*); these can vary from gene to gene. Activators, and possibly repressors, communicate with the basal factors through coactivators (*green*)—proteins that are linked in a tight complex to the *TATA* binding protein (TBP), the first of the basal factors to land on a regulatory region of genes known as the core promoter. Coactivators are named according to their molecular weights (in kilodaltons).



carry. This arrangement enables cells to control transcription of every gene individually.

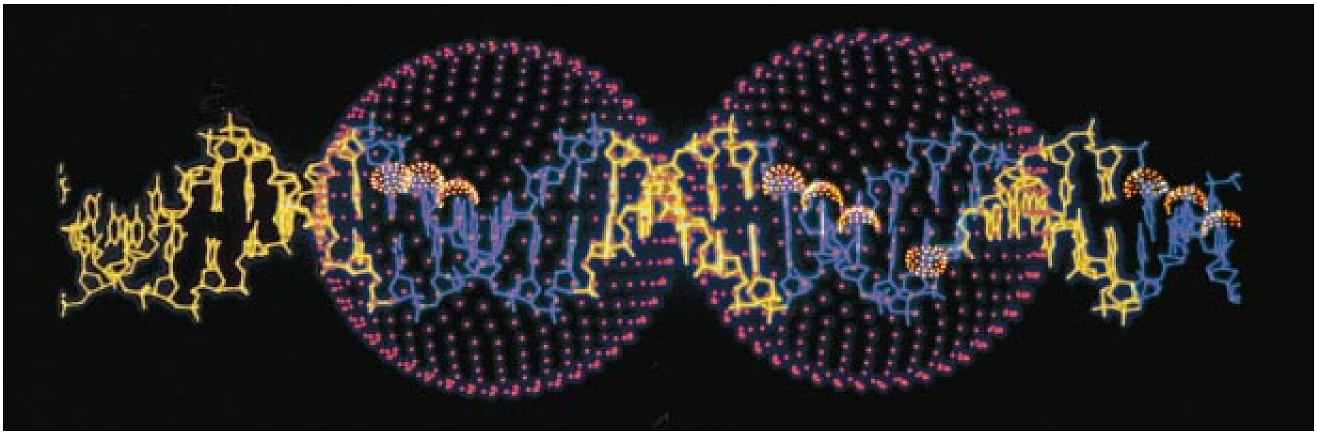
Discovery of these elements led to two related—and, at the time, highly surprising—conclusions. It was evident that enhancers and silencers could not control the activity of RNA polymerase by themselves. Presumably they served as docking sites for a large family of proteins. The proteins that bound to enhancers and silencers—now called activators and repressors—then carried stimulatory or repressive messages directly or indirectly to RNA polymerase (that is, pressed on the accelerator or on the brakes). It also seemed likely that the rate at which a gene was transcribed would be dictated by the combined activity of all the proteins—or transcription factors—bound to its various regulatory elements.

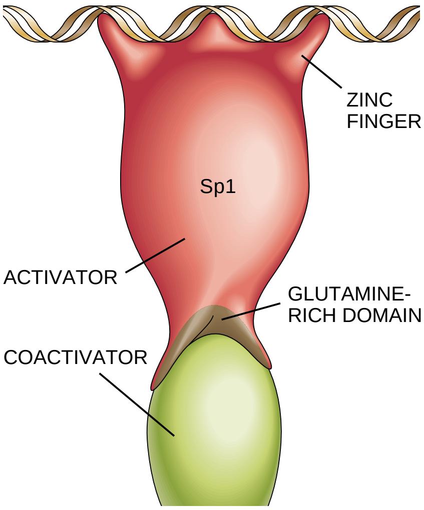
A Human Factor Is Discovered

Nevertheless, we were hard-pressed to explain how proteins that bound to DNA sequences far from the core promoter of a gene could influence transcription of that gene. As is true of other laboratories, we began attacking this puzzle by trying to isolate human transcription factors, none of which had yet been found (with the exception of RNA polymerase itself). We assumed that once we had pure copies of the factors we would be able to gain more insight into exactly how they function.

Because many proteins that bind to DNA play no role in reading genes, we could not find transcription factors efficiently by screening nuclear proteins solely according to their ability to associate with DNA. My group therefore adopted a more discriminating strategy, looking for proteins that in a test-tube reaction both combined with DNA and stimulated transcription.

In 1982 William S. Dynan, a postdoctoral fellow in my laboratory, determined that some protein in a mixture of nuclear proteins fit all the requirements of a transcription factor. It bound to a regulatory element common to a select set of genes—an enhancer sequence known as the *GC* box (because of its abundance of *G* and *C* nucleotides). More important, when added to a preparation of nuclear proteins that included RNA polymerase, the substance mar-





TWO MOLECULES of the activator protein Sp1 (*represented above as large dotted spheres*) have each attached to enhancer sequences called *GC* boxes (*purple regions*) by means of protrusions known as zinc fingers; the points of contact with DNA are highlighted by orange hemispheres. After Sp1 grabs on to DNA, it uses a region rich in the amino acid glutamine (*brown in diagram*) to convey transcription-stimulating signals to a specific coactivator (*green shape*).

kedly increased the transcription only of genes carrying the *GC* box. Thus, we had identified the first human transcription factor able to recognize a specific regulatory sequence. We called it specificity protein 1 (Sp1).

We immediately set out to purify the molecule. One daunting aspect of this work was the fact that transcription factors tend to appear only in minuscule quantities in cells. Typically, less than a thousandth of a percent of the total protein content of a human cell consists of any particular factor. In 1985 James T. Kadonaga in my laboratory found a way to overcome this substantial technical barrier—and in the process introduced a powerful new tool that has since been used to purify countless transcription factors and other scarce DNA binding proteins.

Because Sp1 selectively recognized the *GC* box, Kadonaga synthesized DNA molecules composed entirely of that box and chemically anchored them to solid beads. Then he passed a complex mixture of human nuclear proteins over the DNA, predicting that only Sp1 would stick to it. True to plan, when he separated the bound proteins from the synthetic DNA, he had pure Sp1.

From studies carried out by Mark Ptashne and his colleagues at Harvard University, we knew that bacterial transcription regulators are modular proteins, in which separate regions perform distinct tasks. Once we learned the sequence of amino acids in Sp1, we therefore looked for evidence of distinct modules and noted at least two interesting ones.

One end of the molecule contained a region that obviously folded up into three "zinc fingers." Zinc-finger structures, in which parts of a protein fold around a zinc atom, are now known to act as the "hooks" that attach many activator proteins to DNA. But at the time Sp1 was only the second protein found to use them. Aaron Klug and his colleagues at the Medical Research Council in England had discovered zinc fingers, in a frog transcription factor, just a short time before [see "Zinc Fingers," by Daniela Rhodes and Aaron Klug; SCIENTIFIC AMERICAN, February 1993].

The other end of Sp1 contained a domain consisting of two discrete segments filled with a preponderance of the amino acid glutamine. We strongly suspected that this region played an important role during transcription because of a striking finding. In test-tube experiments, mutant Sp1 molecules lacking the domain could bind to DNA perfectly well, but they failed to stimulate gene transcription. This outcome indicated that Sp1 did not affect transcription solely by combining with DNA; it worked by using its glutamine-rich segment—now known as an activation domain—to interact with some other part of the transcription machinery. The question was, which part?

In 1988 when we began searching for the target of Sp1, we had some idea of where it lay. Our guess was based on an emerging understanding of the so-called basal transcription complex, one part of which seemed to be a likely target.

Closing in on a Target

In the mid-1980s Robert G. Roeder and his colleagues at the Rockefeller University had shown that RNA polymerase cannot transcribe eukaryotic genes unless several other transcription factors—now called basal factors—also collect on the core promoter. And over the course of the 1980s, Roeder's laboratory and others had identified at least six of those essential factors, called A, B, D, E, F and H.

In a test tube, this assembly of factors enabled RNA polymerase to transcribe a bound gene at a basal—low and invariant—rate, but it could not by itself modulate that rate. It was as if someone had constructed and switched on the engine of a car but had lost use of the steering wheel, the accelerator and the brakes. For instance, when my group mixed the components of the complex (including RNA polymerase) with a gene containing a *GC* box, we obtained a low, unchanging level of transcription. We saw a marked increase in transcription only when we incorporated Sp1 into the mixture.

By the late 1980s it was apparent that human cells harbor at least two separate classes of transcription factors. Basal factors are required for initiation of transcription in all genes; other proteins—activators and repressors—dictate the rate at which the basal complex initiates transcription. Different genes are controlled by distinct combinations of activators and repressors. We now suspect that in the body the basal complex arises spontaneously only rarely; most of the time, cells depend on activators to initiate its construction.

These various discoveries suggested that the glutamine-rich domain of Sp1 enhanced transcription by contacting a basal factor. More specifically, we suspected that Sp1 latched on to factor D, and facilitated its attachment to the promoter. We focused on this subunit because Phillip A. Sharp and Stephen Buratowski of the Massachusetts Institute of Technology had shown that it can land on the core promoter before all other basal factors and can facilitate assembly of the complete basal engine. In fact, factor D is the only basal component able to recognize DNA. It binds selectively to a sequence called the *TATA* box, found in the core promoters of many eukaryotic genes.

To pursue our hypothesis, we needed to know more about the composition of factor D, which we assumed was a solitary protein. Other investigators also wanted to know its makeup, and so the race was on to attain pure copies. Isolation from human cells proved more challenging than anyone anticipated. Consequently, many groups eventually tried their luck with yeast cells. Finally, in 1989, several laboratories independently succeeded in isolating a yeast protein that displayed the expected properties of factor D. The protein, named TBP (for *TATA* binding protein), recognized and bound selectively to the *TATA* box and led to a low level of transcription when it was joined at the core promoter by RNA polymerase and other constituents of the basal machinery.

Believing that the TBP protein was factor D itself, we undertook to test this idea in additional studies. Once we did that, we intended to determine exactly which regions of TBP were contacted by Sp1 and other regulators. Little did we know that we were about to be completely thwarted—and to make a critical discovery.

Unexpected Trouble

When B. Franklin Pugh in our laboratory replaced the impure preparations of factor D previously used in our test-tube reactions with purified molecules of TBP, he had no trouble replicating the earlier finding that such substitution in no way disrupted basal transcription. To our surprise and consternation, though, he found that Sp1 was no longer able to influence the basal machinery. We had to conclude that factor D and TBP were not, in fact, equivalent and that factor D actually consisted of TBP plus other subunits. (It is now known that many transcription factors consist of more than one protein.) Apparently, those subunits were not needed for operation of the basal machinery, but they were essential to regulation of that machinery by activators.

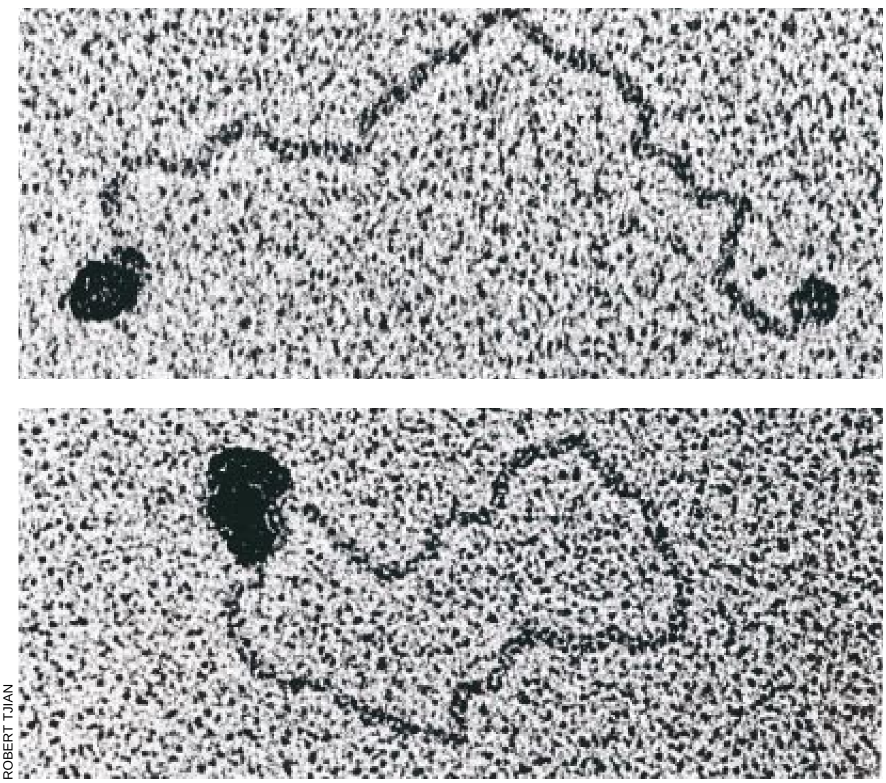
In other words, these additional components were not themselves activators, for they did not bind to specific sequences in DNA. Nor were they basal factors, because low, unregulated levels of transcription could be achieved without them. They seemed to constitute a third class of transcription factor, which we called coactivators. We further proposed that coactivators, not TBP, were the targets for the protein binding domains of activators. We envisioned that activators would bind to selected coactivators to speed up the rate at which the basal complex set molecules of RNA polymerase in motion.

We were attracted to this scenario because we had difficulty imagining how a single protein, TBP, would have enough binding sites to accommodate all the activators made by human cells. But if the coactivators that were tightly linked to TBP bore multiple binding domains, the coactivators could collectively provide the docking sites needed to relay messages from hundreds or thousands of activators to the transcription engine.

It was Pugh who originally proposed that coactivators might function as such adapter molecules. His data soon convinced me he was probably correct, but not everyone in our laboratory agreed. Indeed, our weekly meetings in early 1990 were often punctuated by heated discussions. Not surprisingly, when the coactivator concept was presented to other workers in the field, they, too, expressed considerable skepticism. This reaction to an unexpected and complicating result was probably justified at that stage, because our data were only suggestive, not conclusive. We had not yet isolated a single coactivator.

Coactivators: The Missing Links

To satisfy ourselves and the scientific community that we were correct, we had to devise an experimental procedure that would unambiguously establish whether coactivators existed and operated as the relays we envisioned. For approximately two years after Pugh formulated the coactivator hypothesis, we struggled to purify an intact and functional complex containing TBP and all the other associated constituents of factor D. I must admit to some dark mo-



ROBERT TJIAN

ACTIVATORS CAN COMMUNICATE with one another, not only with DNA and coactivators, as is demonstrated in an experiment involving Sp1. Copies of the protein (*dark dots*) each recognized and bound to *GC* boxes at the tips of a string of DNA. After binding, they joined together—causing DNA to loop out in the process.

ments when it seemed the rather unpopular coactivator hypothesis might be based on some error in our studies.

The breakthrough finally came in 1991, when Brian D. Dynlacht, Timothy Hoey, Naoko Tanese and Robert Weinzierl—graduate students and postdoctoral fellows in our laboratory—found an ingenious way to isolate pure copies of factor D. Subsequent biochemical analyses revealed that, aside from TBP, the complete unit included eight previously unknown proteins. Because we did not yet have proof that these proteins could function as coactivators, we referred to them more generically as TBP-associated factors, or TAFs.

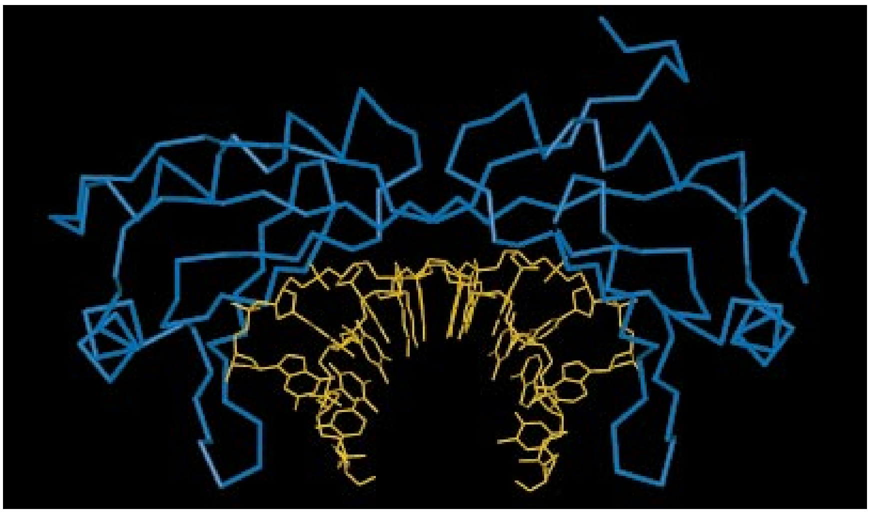
We became convinced that TAFs do indeed convey molecular signals from activators to the basal transcription apparatus after we separated the bound proteins from TBP and completed several more experiments. For instance, we were able to show that mixing of the activator Sp1 with basal factors and RNA polymerase enhanced production of messenger RNA from a gene containing a *GC* box only when TAFs were added as well. Later, Jin-Long Chen, a graduate student, combined purified TBP and the eight isolated TAFs in a test tube along with a human gene and the rest of the basal transcription machinery. The various proteins assembled on the gene and proved able to respond to several different types of activator proteins. These activators, we later showed, produced their effects by coupling directly with selected TAFs. Together the coactivators in factor D do indeed constitute a kind of central processing unit that integrates the regulatory signals issued by DNA-bound activators.

A Universal Theme

The complexes formed by activators, coactivators and the basal machinery appear to be human equivalents of sigma factors; they, too, draw RNA polymerase to specific genes at specific rates. In a way, the complexes can be viewed as sigma factors that have been elaborated into many subunits. Gratifyingly, recent evidence from our group and others suggests we have uncovered a universal mode of gene regulation in eukaryotes. Those studies confirm that coactivators also exist in yeast and that factor D consists of multiple subunits in fungi as well as in humans.

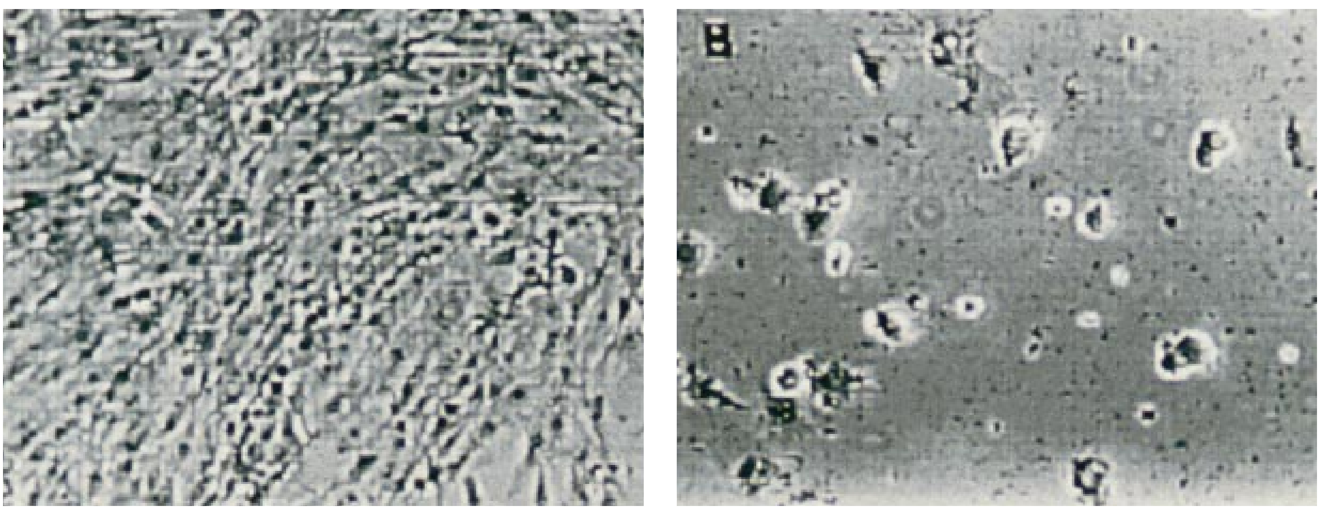
As satisfying as these results are, they do not fully explain how binding of activators to enhancers and to coactivators influences the rate at which RNA polymerase transcribes genes in living cells. It may be that linkage of activators to enhancers causes DNA to bend in a way that brings the enhancers closer to one another and to the core promoter. This arrangement may help activators (alone or in concert with one another) to dock with coactivators and position factor D on the promoter. This step, in turn, would facilitate assembly of the complete basal complex. Formation of this complex may distort the underlying DNA in a way that enables RNA polymerase to begin its journey along the coding region.

Researchers know less about the functioning of repressors. Nevertheless, many of us think repressors may also bind to coactivators at times. This bind-



STEPHEN BURLEY *Rockefeller University*; PAUL B. SIGLER *Yale University*

***TATA* BINDING PROTEIN (*blue*) is a remarkably symmetrical, saddle-shaped molecule. Its underside rides on DNA (*yellow*) and seems to bend it. This bending may somehow facilitate assembly of the complex that initiates transcription. Coactivators, which are not depicted, are thought to bind tightly to the upper surface.**



EDITH WANG *University of California, Berkeley*

CELLS GROWING IN CULTURE (*left*) stopped proliferating and began to die (*right*) when they were blocked from making a coactivator. They failed because they could no longer produce proteins critical to survival. This finding is one of many indicating that coactivators are essential to the transcription of most, if not all, genes.

ing could inhibit transcription by preventing activators from attaching to their usual sites on coactivators. Other times repressors might bypass the basal machinery, blocking transcription by preventing activators from connecting with enhancers.

Although there are gaps in our knowledge, we can now begin to sketch out an explanation as to why different cells make different mixtures of proteins during embryonic development and in mature organisms. A gene will be transcribed at a measurable rate only if the various activators it needs are present and can successfully overcome the inhibitory effects of repressors. Cells vary in the proteins they make because they contain distinct batteries of activators and repressors. Of course, this scenario begs the question of how cells decide which transcription factors to produce in the first place, but progress is being made on that front as well.

BETTINA CIRONE *Photo Researchers, Inc.*

ONE OF THE WORLD'S TALLEST WOMEN reportedly reached her height of seven feet, seven and a quarter inches because her pituitary gland produced an excess of growth hormone. Investigators hope eventually to treat this and many other disorders by increasing or decreasing the transcription of selected genes.

Therapies of Tomorrow

How might investigators use our newly acquired knowledge of gene regulation to develop drugs for combating life-threatening diseases involving excessive or inadequate transcription of a gene? In theory, blocking selected activators from attaching to enhancers or coactivators should depress unwanted transcription, and stabilizing the transcription machinery on a gene should counteract undesirably weak transcription.

Blockade could be achieved by fitting a molecular "plug" into an activator, thereby preventing its interaction with a coactivator, or by enticing an activator to attach to a decoy that resembles a coactivator. Stabilization of a complex might be achieved by deploying molecules that would strengthen the interaction between activators and DNA or between activators and coactivators. Such approaches are remote today, but it is exciting to consider a sampling of the applications that might eventually be possible.

Take, for example, the human immunodeficiency virus (HIV), which causes AIDS. To reproduce itself in human cells, HIV needs the viral transcription factor *TAT* to enhance transcription of HIV genes. If *TAT* could be inhibited by some agent that recognized *TAT* but ignored human transcription factors, replication of the virus might be halted without affecting production of proteins needed by the patient.

Conversely, treatment of some disorders—for instance, hypercholesterolemia—might involve enhancing the transcription of selected genes. Hypercholesterolemia increases a person's risk for heart disease. Cholesterol accumulates to destructive levels in the blood when low-density lipoprotein (LDL), otherwise known as the bad cholesterol, is not removed efficiently. In theory, the disease could be corrected by turning up transcription of the gene for the LDL receptor in liver cells. This receptor helps to clear LDL from the blood. This idea may soon be testable, because studies by Michael S. Brown and Joseph L. Goldstein of the University of Texas Health Science Center at Dallas are teasing apart the specific molecular constituents of the apparatus that regulates transcription of the receptor gene.

Until recently, no one put much effort into screening small molecules, natural products or other compounds for their ability to modulate transcription. Even so, a number of drugs already on the market have been found by chance to work by altering the activity of transcription factors. One of these, RU 486 (the French "abortion" pill), represses the function of particular steroid receptors, a class of activators that direct embryonic development. Similarly, the immunosuppressants cyclosporine and FK506 suppress transcription of a gene whose protein product is needed by certain cells of the immune system. These drugs act indirectly, however. They activate an enzyme that impedes the functioning of a transcription factor for the gene.

As time goes by, the precise combination of transcription factors that regulate individual genes is sure to be identified. And drug developers will probably use this information to devise sophisticated compounds for fighting cancer, heart disease, immune disorders, viral infections, Alzheimer's disease and perhaps even the aging process. How well these agents will succeed is anybody's guess, but it is likely that therapies of the future will benefit in one way or another from basic research into transcription—research that began not out of a wish to design drugs but rather out of a simple desire to get to the heart of the molecular machinery that controls the activity of our genes.

FURTHER READING

TRANSCRIPTIONAL SELECTIVITY OF VIRAL GENES IN MAMMALIAN CELLS. Steven McKnight and Robert Tjian in *Cell,* Vol. 46, No. 6, pages 795–805; September 12, 1986.

TRANSCRIPTIONAL REGULATION IN MAMMALIAN CELLS BY SEQUENCE-SPECIFIC DNA BINDING PROTEINS. Pamela J. Mitchell and Robert Tjian in *Science,* Vol. 245, pages 371–378; July 28, 1989.

EUKARYOTIC COACTIVATORS ASSOCIATED WITH THE TATA BOX BINDING PROTEIN. G. Gill and R. Tjian in *Current Opinion in Genetics and Development,* Vol. 2, No. 2, pages 236–242; April 1992.

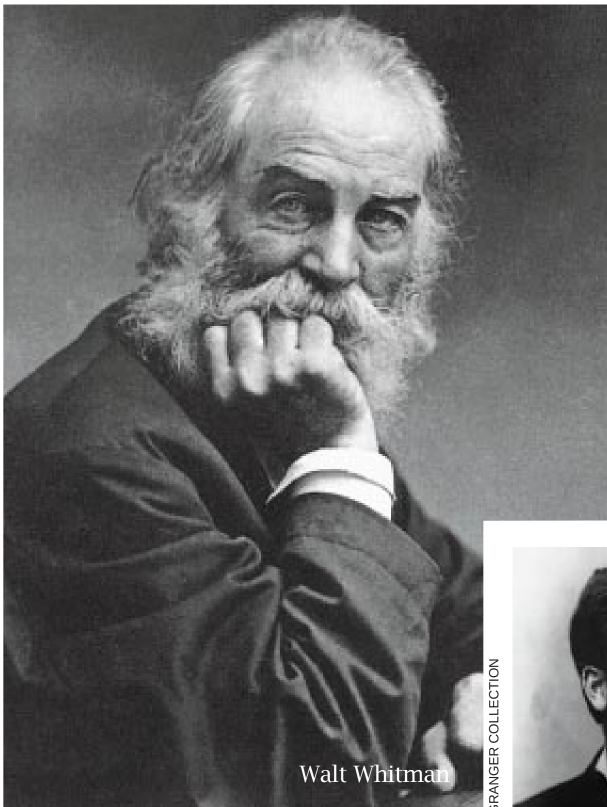
TRANSCRIPTIONAL ACTIVATION: A COMPLEX PUZZLE WITH FEW EASY PIECES. R. Tjian and T. Maniatis in *Cell,* Vol. 77, No. 1, pages 5–8; April 8, 1994.

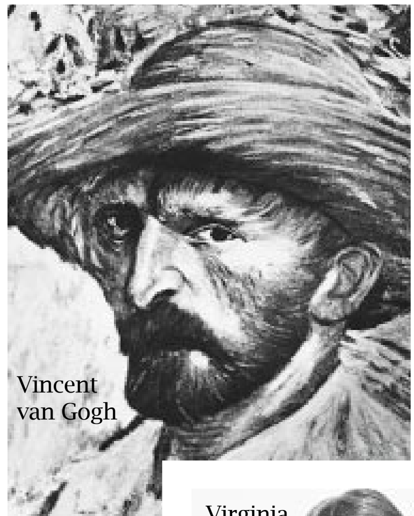
Manic-Depressive Illness and Creativity

Does some fine madness plague great artists? Several studies now show that creativity and mood disorders are linked

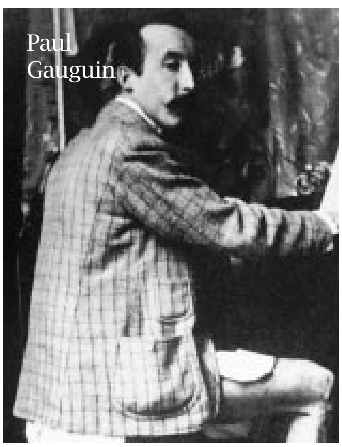
by Kay Redfield Jamison

AP/WIDE WORLD PHOTOS





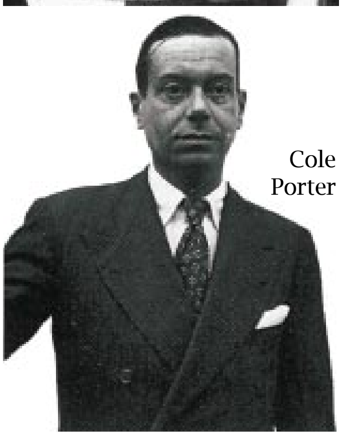
AP/ WIDE WORLD PHOTOS



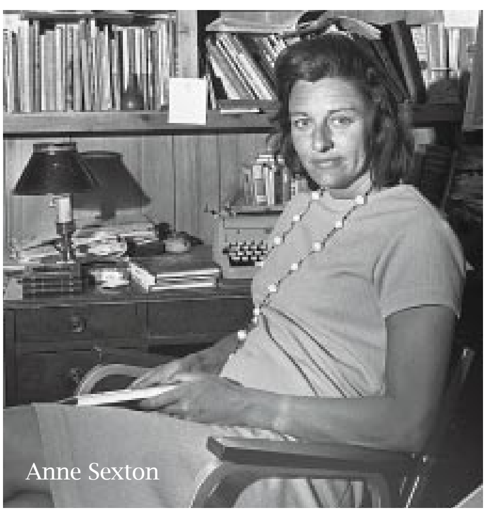
AP/ WIDE WORLD PHOTOS

GRANGER COLLECTION

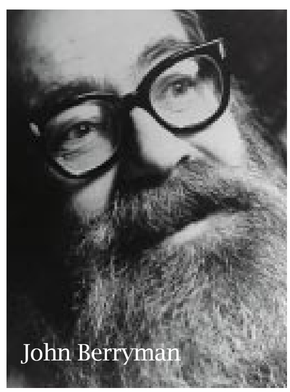
BETTMANN ARCHIVE



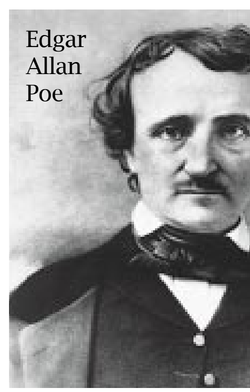
UPI/BETTMANN



UPI/BETTMANN



AP/ WIDE WORLD PHOTOS

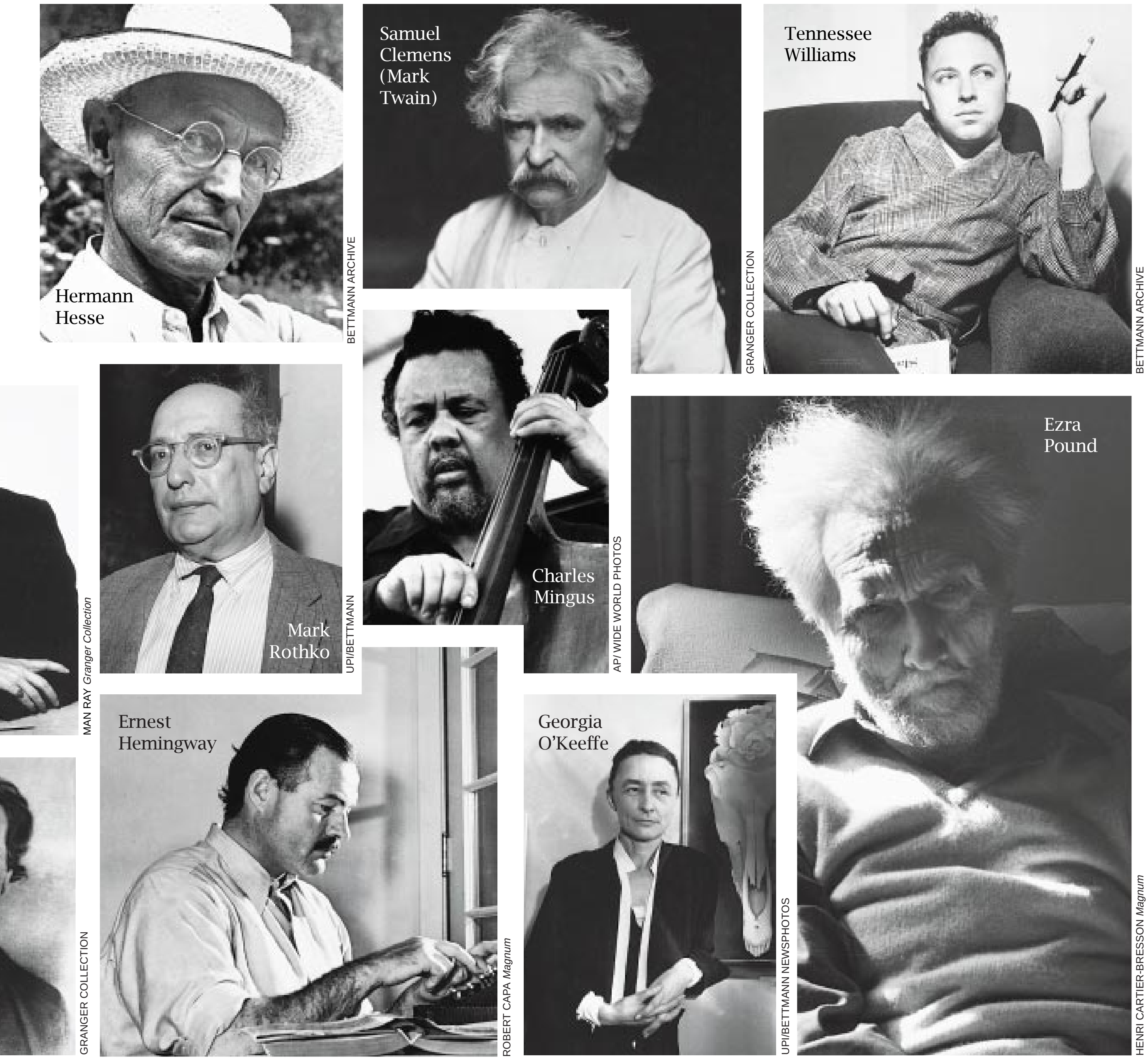


Men have called me mad," wrote Edgar Allan Poe, "but the question is not yet settled, whether madness is or is not the loftiest intelligence—whether much that is glorious—whether all that is profound—does not spring from disease of thought—from moods of mind exalted at the expense of the general intellect."

Many people have long shared Poe's suspicion that genius and insanity are entwined. Indeed, history holds countless examples of "that fine madness." Scores of influential 18th- and 19th-century poets, notably William Blake, Lord Byron and Alfred, Lord Tennyson, wrote about the extreme mood swings they endured. Modern American poets John Berryman, Randall Jarrell, Robert Lowell, Sylvia Plath, Theodore Roethke, Delmore Schwartz and Anne Sexton were all hospitalized for either mania or depression during their lives. And many painters and composers, among them Vincent van Gogh, Georgia O'Keeffe, Charles Mingus and Robert Schumann, have been similarly afflicted.

Judging by current diagnostic criteria, it seems that most of these artists—and many others besides—suffered from one of the major mood disorders, namely, manic-depressive illness or major depression. Both are fairly common, very treatable and yet frequently lethal diseases. Major depression induces intense melancholic spells, whereas manic-depression, a strongly genetic disease,

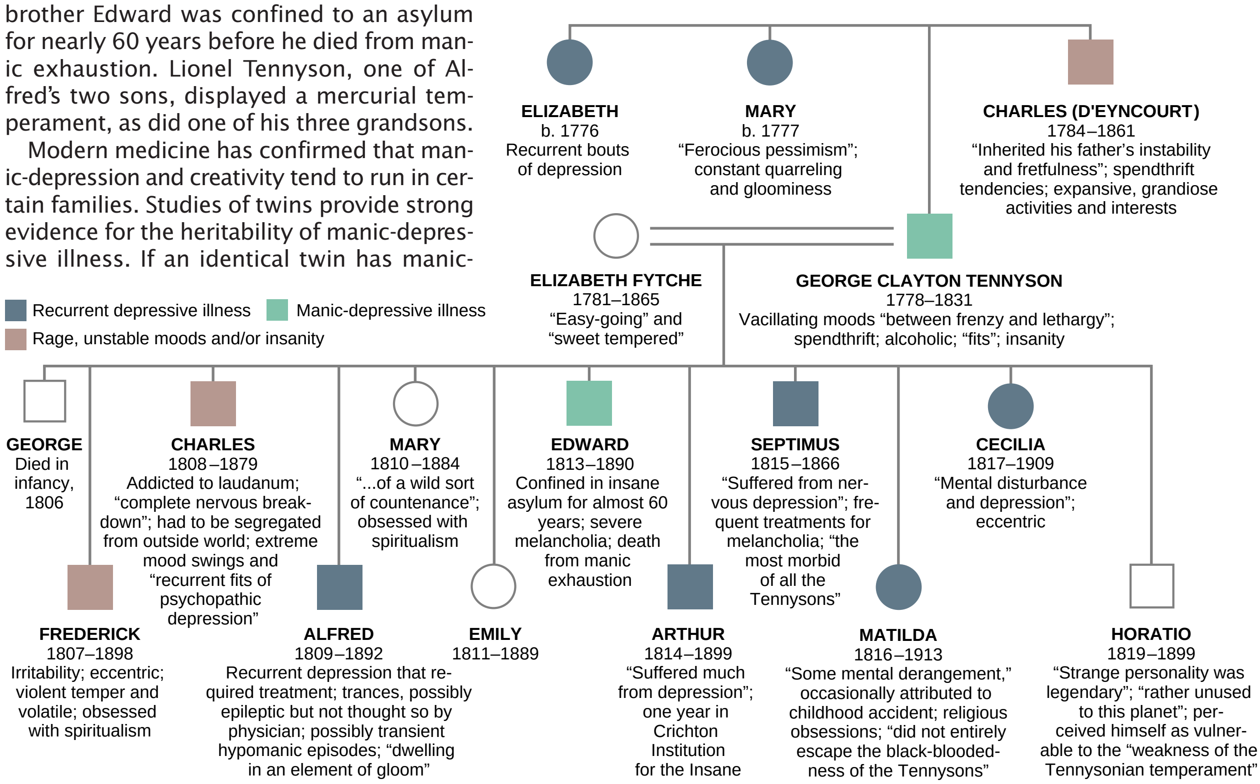
WRITERS, ARTISTS AND COMPOSERS shown in this montage all most likely suffered from manic-depressive illness or major depressive illness, according to their letters and journals, medical records and accounts by their families and friends. Recent studies indicate that the temperaments and cognitive styles associated with mood disorders can in fact enhance creativity in some individuals.



The Tainted Blood of the Tennysons

Alfred, Lord Tennyson, who experienced recurrent, debilitating depressions and probable hypomanic spells, often expressed fear that he might inherit the madness, or "taint of blood," in his family. His father, grandfather, two of his great-grandfathers as well as five of his seven brothers suffered from insanity, melancholia, uncontrollable rage or what is today known as manic-depressive illness. His brother Edward was confined to an asylum for nearly 60 years before he died from manic exhaustion. Lionel Tennyson, one of Alfred's two sons, displayed a mercurial temperament, as did one of his three grandsons.

Modern medicine has confirmed that manic-depression and creativity tend to run in certain families. Studies of twins provide strong evidence for the heritability of manic-depressive illness. If an identical twin has manic-depressive illness, the other twin has a 70 to 100 percent chance of also having the disease; if the other twin is fraternal, the chances are considerably lower (approximately 20 percent). A review of identical twins reared apart from birth—in which at least one of the twins had been diagnosed as manic-depressive—found that two thirds or more of the sets were concordant for the illness.



SOURCE: Adapted from Touched with Fire: Manic-Depressive Illness and the Artistic Temperament; *based on biographies, autobiographical writings and letters.*

LISA BURNETT

pitches patients repeatedly from depressed to hyperactive and euphoric, or intensely irritable, states. In its milder form, termed cyclothymia, manic-depression causes pronounced but not totally debilitating changes in mood, behavior, sleep, thought patterns and energy levels. Advanced cases are marked by dramatic, cyclic shifts.

Could such disruptive diseases convey certain creative advantages? Many people find that proposition counterintuitive. Most manic-depressives do not possess extraordinary imagination, and most accomplished artists do not suffer from recurring mood swings. To assume, then, that such diseases usually promote artistic talent wrongly reinforces simplistic notions of the "mad genius." Worse yet, such a generalization trivializes a very serious medical condition and, to some degree, discredits individuality in the arts as well. It would be wrong to label anyone who is unusually accomplished, energetic, intense, moody or eccentric as manic-depressive.

All the same, recent studies indicate that a high number of established artists—far more than could be expected by chance—meet the diagnostic criteria for manic-depression or major depression given in the fourth edition of the *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*). In fact, it seems that these diseases can sometimes enhance or otherwise contribute to creativity in some people.

Diagnosing Mood Disorders

By virtue of their prevalence alone, it is clear that mood disorders do not necessarily breed genius. Indeed, 1 percent of the general population suffer from manic-depression, also called bipolar disorder, and 5 percent from a major depression, or unipolar disorder, during their lifetime. Depression affects twice as many women as men and most often, but not always, strikes later in life. Bipolar disorder afflicts equal numbers of women and men, and more than a third of all cases surface before age 20. Some 60 to 80 percent of all adolescents and adults who commit suicide have a history of bipolar or unipolar illness. Before the late 1970s, when the drug lithium first became widely available, one person in five with manic-depression committed suicide.

Major depression in both unipolar and bipolar disorders manifests itself through apathy, lethargy, hopelessness, sleep disturbances, slowed physical movements and thinking, impaired memory and concentration, and a loss of pleasure in typically enjoyable events. The diagnostic criteria also include suicidal thinking, self-blame and inappropriate guilt. To distinguish clinical depression from normal periods of unhappiness, the common guidelines further require that these symptoms persist for a minimum of two to four weeks and also that they significantly interfere with a person's everyday functioning.

During episodes of mania or hypomania (mild mania), bipolar patients experience symptoms that are in many ways the opposite of those associated with depression. Their mood and self-esteem are elevated. They sleep less and have abundant energy; their productivity increases. Manics frequently become paranoid and irritable. Moreover, their speech is often rapid, excitable and intrusive, and their thoughts move quickly and fluidly from one topic to another. They usually hold tremendous conviction about the correctness and importance of their own ideas as well. This grandiosity can contribute to poor judgment and impulsive behavior.

Hypomanics and manics generally have chaotic personal and professional relationships. They may spend large sums of money, drive recklessly or pursue questionable business ventures or sexual liaisons. In some cases, manics suffer from violent agitation and delusional thoughts as well as visual and auditory hallucinations.

KAY REDFIELD JAMISON is professor of psychiatry at the Johns Hopkins University School of Medicine. She wrote *Touched with Fire: Manic-Depressive Illness and the Artistic Temperament* and co-authored the medical text *Manic-Depressive Illness.* Jamison is a member of the National Advisory Council for Human Genome Research and clinical director of the Dana Consortium on the Genetic Basis of Manic-Depressive Illness. She has also written and produced a series of public television specials about manic-depressive illness and the arts.

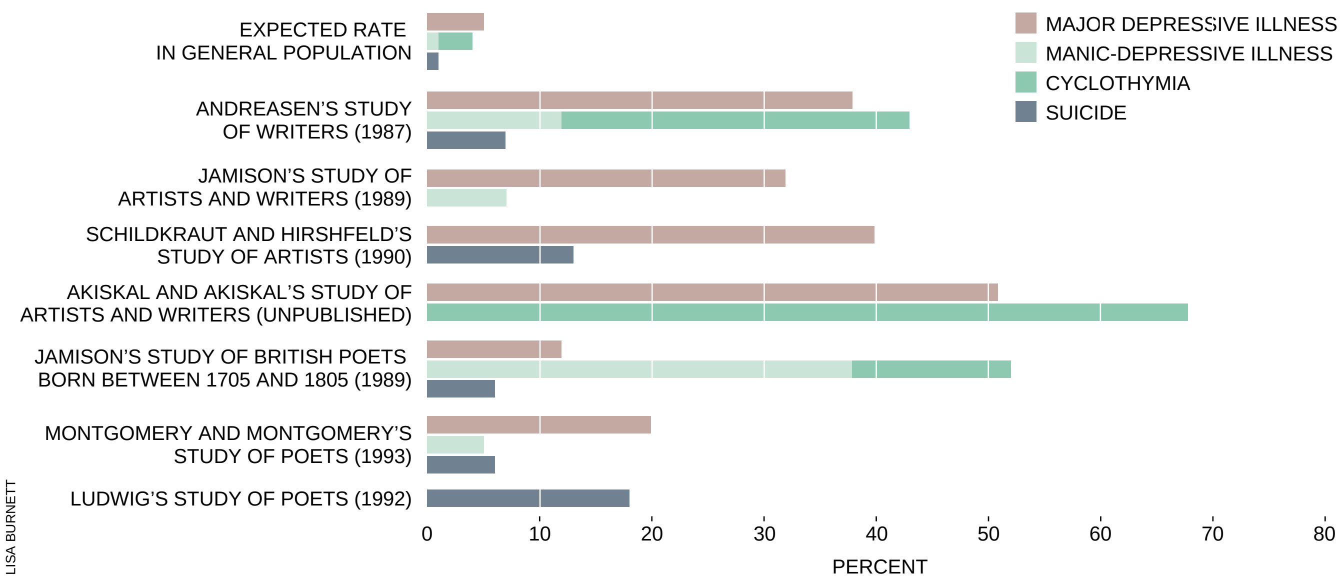
Rates of Mood Disorders

For years, scientists have documented some kind of connection between mania, depression and creative output. In the late 19th and early 20th centuries, researchers turned to accounts of mood disorders written by prominent artists, their physicians and friends. Although largely anecdotal, this work strongly suggested that renowned writers, artists and composers—and their first-degree relatives—were far more likely to experience mood disorders and to commit suicide than was the general population. During the past 20 years, more systematic studies of artistic populations have confirmed these findings [*see illustration below*]. Diagnostic and psychological analyses of living writers and artists can give quite meaningful estimates of the rates and types of psychopathology they experience.

In the 1970s Nancy C. Andreasen of the University of Iowa completed the first of these rigorous studies, which made use of structured interviews, matched control groups and strict diagnostic criteria. She examined 30 creative writers and found an extraordinarily high occurrence of mood disorders and alcoholism among them. Eighty percent had experienced at least one episode of major depression, hypomania or mania; 43 percent reported a history of hypomania or mania. Also, the relatives of these writers, compared with the relatives of the control subjects, generally performed more creative work and more often had a mood disorder.

A few years later, while on sabbatical in England from the University of California at Los Angeles, I began a study of 47 distinguished British writers and visual artists. To select the group as best I could for creativity, I purposefully chose painters and sculptors who were Royal Academicians or Associates of the Royal Academy. All the playwrights had won the New York Drama Critics Award or the Evening Standard Drama (London Critics) Award, or both. Half of the poets were already represented in the *Oxford Book of Twentieth Century English Verse.* I found that 38 percent of these artists and writers had in fact been previously treated for a mood disorder; three fourths of those treated had required medication or hospitalization, or both. And half of the poets—the largest fraction from any one group—had needed such extensive care.

Hagop S. Akiskal of the University of California at San Diego, also affiliated with the University of Tennessee at Memphis, and his wife, Kareen Akiskal, subsequently interviewed 20 award-winning European writers, poets, painters and sculptors. Some two thirds of their subjects exhibited recurrent cyclothymic or hypomanic tendencies, and half had at one time suffered from a major depression. In collaboration with David H. Evans of the University of Memphis, the Akiskals noted the same trends among living blues musicians. More recently Stuart A. Montgomery and his wife, Deirdre B. Montgomery, of St.



INCREASED RATES of suicide, depression and manic-depression among artists have been established by many separate studies. These investigations show that artists experience up to 18 times the rate of suicide seen in the general population, eight to 10 times the rate of depression and 10 to 20 times the rate of manic-depression and cyclothymia.

Mary's Hospital in London examined 50 modern British poets. One fourth met current diagnostic criteria for depression or manic-depression; suicide was six times more frequent in this community than in the general population.

Ruth L. Richards and her colleagues at Harvard University set up a system for assessing the degree of original thinking required to perform certain creative tasks. Then, rather than screening for mood disorders among those already deemed highly inventive, they attempted to rate creativity in a sample of manic-depressive patients. Based on their scale, they found that compared with individuals having no personal or family history of psychiatric disorders, manic-depressive and cyclothymic patients (as well as their unaffected relatives) showed greater creativity.

Biographical studies of earlier generations of artists and writers also show consistently high rates of suicide, depression and manic-depression—up to 18 times the rate of suicide seen in the general population, eight to 10 times that of depression and 10 to 20 times that of manic-depressive illness and its milder variants. Joseph J. Schildkraut and his co-workers at Harvard concluded that approximately half of the 15 20th-century abstract-expressionist artists they studied suffered from depressive or manic-depressive illness; the suicide rate in this group was at least 13 times the current U.S. national rate.

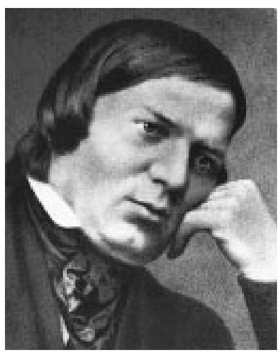
In 1992 Arnold M. Ludwig of the University of Kentucky published an extensive biographical survey of 1,005 famous 20th-century artists, writers and other professionals, some of whom had been in treatment for a mood disorder. He discovered that the artists and writers experienced two to three times the rate of psychosis, suicide attempts, mood disorders and substance abuse than did comparably successful people in business, science and public life. The poets in this sample had most often been manic or psychotic and hospitalized; they also proved to be some 18 times more likely to commit suicide than is the general public. In a comprehensive biographical study of 36 major British poets born between 1705 and 1805, I found similarly elevated rates of psychosis and severe psychopathology. These poets were 30 times more likely to have had manic-depressive illness than were their contemporaries, at least 20 times more likely to have been committed to an asylum and some five times more likely to have taken their own life.

Cycles of Creative Accomplishment

These corroborative studies have confirmed that highly creative individuals experience major mood disorders more often than do other groups in the general population. But what does this mean for their work? How does a psychiatric illness actually contribute to creative achievement? First, the common features of hypomania seem highly conducive to original thinking; the diagnostic criteria for this phase of the disorder include "sharpened and unusually creative thinking and increased productivity." And accumulating evidence suggests that the cognitive styles associated with hypomania (namely, expansive thought and grandiose moods) can lead to increased fluency and frequency of thoughts.

Studying the speech of hypomanic patients has revealed that they tend to rhyme and use other sound associations, such as alliteration, far more often than do unaffected individuals. They also use idiosyncratic words nearly three times as often as do control subjects. Moreover, in specific drills, they can list synonyms or form other word associations much more rapidly than is considered normal. It seems, then, that both the quantity and quality of thoughts build during hypomania. This speed increase may range from a very mild quickening to complete psychotic incoherence. It is not yet clear what causes this qualitative change in mental processing. Nevertheless, this altered cognitive state may well facilitate the formation of unique ideas and associations.

Manic-depressive illness and creative accomplishment share certain noncognitive features: the ability to function well on a few hours of sleep, the focus needed to work intensively, bold and restless attitudes, and an ability to experience a profound depth and variety of emotions. The less dramatic daily aspects of manic-depression might also provide creative advantage to some individuals. The manic-depressive temperament is, in a biological sense, an alert, sensitive system that reacts strongly and swiftly. It responds to the world



ROBERT SCHUMANN'S MUSICAL WORKS, charted by year and opus number (*above*), show a striking relation between his mood states and his productivity. He composed the most when hypomanic and the least when depressed. Both of Schumann's parents were clinically depressed, and two other first-degree relatives committed suicide. Schumann himself attempted suicide twice and died in an insane asylum. One of his sons spent more than 30 years in a mental institution.

The Case of Vincent van Gogh

Many clinicians have reviewed the medical and psychiatric problems of the painter Vincent van Gogh posthumously, diagnosing him with a range of disorders, including epilepsy, schizophrenia, digitalis and absinthe poisoning, manic-depressive psychosis, acute intermittent porphyria and Ménière's disease. Richard Jed Wyatt of the National Institute of Mental Health and the author have argued in detail that van Gogh's symptoms, the natural course of his illness and his family psychiatric history strongly indicate manic-depressive illness. The extent of the artist's purported absinthe use and convulsive behavior remains unclear; in any event, his psychiatric symptoms long predate any possible history of seizures. It is possible that he suffered from both epilepsy and manic-depressive illness.

Irises, ***1889***

METROPOLITAN MUSEUM OF ART, GIFT OF ADELE R. LEVY, 1958

with a wide range of emotional, perceptual, intellectual, behavioral and energy changes. In a sense, depression is a view of the world through a dark glass, and mania is that seen through a kaleidoscope—often brilliant but fractured.

Where depression questions, ruminates and hesitates, mania answers with vigor and certainty. The constant transitions in and out of constricted and then expansive thoughts, subdued and then violent responses, grim and then ebullient moods, withdrawn and then outgoing stances, cold and then fiery states—and the rapidity and fluidity of moves through such contrasting experiences—can be painful and confusing. Ideally, though, such chaos in those able to transcend it or shape it to their will can provide a familiarity with transitions that is probably useful in artistic endeavors. This vantage readily accepts ambiguities and the counteracting forces in nature.

Extreme changes in mood exaggerate the normal tendency to have conflicting selves; the undulating, rhythmic and transitional moods and cognitive changes so characteristic of manic-depressive illness can blend or harness seemingly contradictory moods, observations and perceptions. Ultimately, these fluxes and yokings may reflect truth in humanity and nature more accurately than could a more fixed viewpoint. The "consistent attitude toward life," may not, as Byron scholar Jerome J. McGann of the University of Virginia points out, be as insightful as an ability to live with, and portray, constant change.

The ethical and societal implications of the association between mood disorders and creativity are important but poorly understood. Some treatment strategies pay insufficient heed to the benefits manic-depressive illness can bestow on some individuals. Certainly most manic-depressives seek relief from the disease, and lithium and anticonvulsant drugs are very effective therapies for manias and depressions. Nevertheless, these drugs can dampen a person's general intellect and limit his or her emotional and perceptual range. For this reason, many manic-depressive patients stop taking these medications.

Left untreated, however, manic-depressive illness often worsens over time—and no one is creative when severely depressed, psychotic or dead. The attacks of both mania and depression tend to grow more frequent and more severe. Without regular treatment the disease eventually becomes less responsive to medication. In addition, bipolar and unipolar patients frequently abuse mood-altering substances, such as alcohol and illicit drugs, both of which can cause secondary medical and emotional burdens for manic-depressive and depressed patients.

The Goal of Treatment

The real task of imaginative, compassionate and effective treatment, therefore, is to give patients more meaningful choices than they are now afforded. Useful intervention must control the extremes of depression and psychosis without sacrificing crucial human emotions and experiences. Given time and increasingly sophisticated research, psychiatrists will likely gain a better understanding of the complex biological basis for mood disorders. Eventually, the development of new drugs should make it possible to treat manic-depressive individuals so that those aspects of temperament and cognition that are essential to the creative process remain intact.

The development of more specific and less problematic therapies should be swift once scientists find the gene, or genes, responsible for the disease. Prenatal tests and other diagnostic measures may then become available; these possibilities raise a host of complicated ethical issues. It would be irresponsible to romanticize such a painful, destructive and all too often deadly disease. Hence, 3 to 5 percent of the Human Genome Project's total budget (which is conservatively estimated at $3 billion) has been set aside for studies of the social, ethical and legal implications of genetic research. It is hoped that these investigations will examine the troubling issues surrounding manic-depression and major depression at length. To help those who have manic-depressive illness, or who are at risk for it, must be a major public health priority.

FURTHER READING

TENNYSON: THE UNQUIET HEART. R. B. Martin. Oxford University Press, 1980.

CREATIVITY AND MENTAL ILLNESS: PREVALENCE RATES IN WRITERS AND THEIR FIRST-DEGREE RELATIVES. Nancy C. Andreasen in *American Journal of Psychiatry,* Vol. 144, No. 10, pages 1288-1292; October 1987.

MANIC DEPRESSIVE ILLNESS. Frederick K. Goodwin and Kay R. Jamison. Oxford University Press, 1990.

CREATIVE ACHIEVEMENT AND PSYCHOPATHOLOGY: COMPARISON AMONG PROFESSIONS. Arnold M. Ludwig in *American Journal of Psychotherapy,* Vol. 46, No. 3, pages 330–356; July 1992.

TOUCHED WITH FIRE: MANIC-DEPRESSIVE ILLNESS AND THE ARTISTIC TEMPERAMENT. Kay R. Jamison. Free Press/Macmillan, 1993.

Masers in the Sky

Interstellar gas clouds produce intense, coherent microwaves. This radiation offers a glimpse of the size, content and distance of objects that may otherwise be invisible

by Moshe Elitzur

Three decades ago radio astronomers began to detect signals unlike any they had ever expected to see. In 1963 a team led by Alan H. Barrett of the Massachusetts Institute of Technology discovered radio signals from clouds of excited molecules in interstellar space. Many astrophysicists had believed molecular clouds could not form in the regions between stars, but the group's findings were anomalous for another reason as well. The first molecule they detected, the hydroxyl radical (OH), emitted radiation whose patterns appeared to contradict the rules of statistical physics. Emission lines that should have been strong were weak, and ones that should have been weak were intense.

In 1965 a team led by Harold F. Weaver of the University of California at Berkeley detected radiation whose properties were so unusual that they dubbed the emitting substance "mysterium," for lack of an obvious explanation. The microwaves they found were extraordinarily bright, every emission line covered a tiny range of wavelengths (as opposed to the broader peaks that researchers had seen in other sources), and the radiation was almost all polarized in the same sense, even though light and radio waves emitted by astronomical sources generally consist of photons each randomly polarized in a different sense.

It did not take too long, though, to realize that "mysterium" radiation was not a sign of a new molecule but rather a maser emission from interstellar hydroxyl. (Masers, which had been invented only in 1953, were the precursors of the now ubiquitous laser; the acronym stands for microwave amplification by stimulated emission of radiation.) The narrow line width and polarization were hallmarks of masers; only an amplified interstellar source could achieve such brightness. Albert Einstein had sketched the physical mechanism for stimulated emission in 1917, but building devices that could employ the phenomenon for amplification was very difficult; in 1964 Charles H. Townes shared the Nobel Prize in Physics for inventing masers.

Since those early observations, radio astronomers have detected maser emissions from many different interstellar molecules, including water vapor, silicon monoxide, methanol and even the hydrogen atom (by far the most abundant substance in the universe). Masers appear in a variety of astronomical sources, ranging from comets to galaxies.

The radiation of astronomical masers—when viewed with the unprecedented resolution of modern radio telescopes—provides information about their source that is otherwise unavailable. Masers form only under certain conditions, and so astronomers can infer detailed profiles of pressure, temperature and gas velocity. Furthermore, because masers are very small and very bright, they serve as markers of small-scale structure that would otherwise be impossible to observe.

MOSHE ELITZUR has studied astronomical masers and the gas clouds that give rise to them for more than 15 years. He is professor of astronomy at the University of Kentucky, whose faculty he joined in 1980. Elitzur received his Ph.D. from the Weizmann Institute of Science in Israel in 1971 and his B.Sc. from the Hebrew University of Jerusalem in 1966.

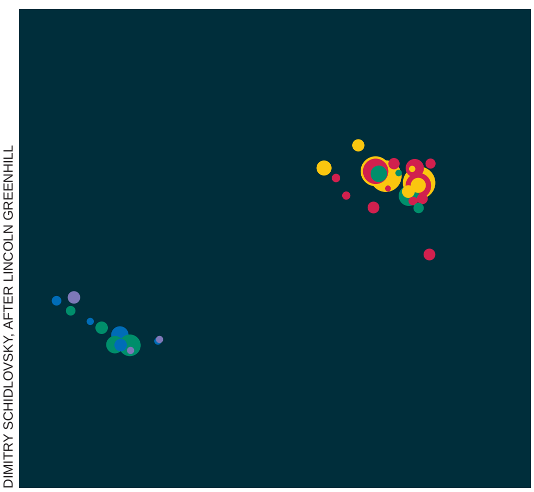
Amplification Is Easier in Space

By a fortuitous coincidence, conditions suitable for masers occur in circumstellar clouds during both early and late stages of the life of a star. These two phases are among the most challenging in stellar evolution, and extremely bright maser emission provides invaluable information on both.

How can interstellar gas clouds do spontaneously what physicists in well-equipped laboratories took decades to accomplish? Masers and lasers arise from a condition known as population inversion, in which the number of atoms or molecules in a higher-energy state exceeds that in a lower-energy state—the reverse of the normal state of affairs. As a result, the response of the inverted population to incoming photons also reverses conventional behavior. When atoms or molecules encounter photons of the appropriate wavelength, they generally move from a lower-energy state to a higher one. In inverted populations, however, most of the atoms or molecules are already in the higher-energy state, and so they respond instead by emitting a photon [*see box on page 71*]. Each emission then triggers the release of additional photons, and the incoming light is amplified rather than being absorbed.

Laboratory masers and lasers commonly create population inversions by shifting atoms or molecules among three energy states: the stable bottom, which they would usually occupy; a short-lived upper state; and a longer-lived middle, to which the upper state preferentially decays. A source of energy pumps atoms or molecules into the upper state, from which they rapidly

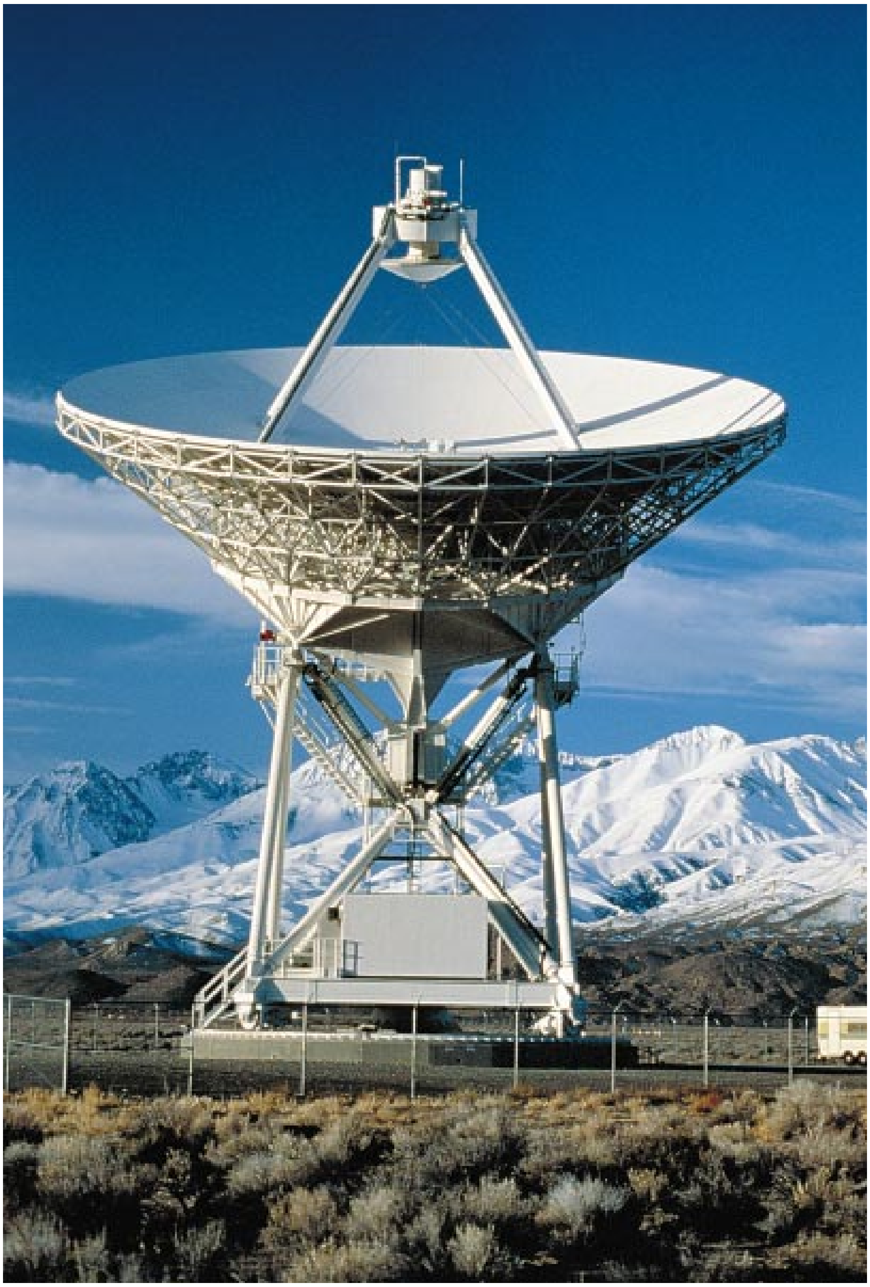
DAVID MALIN *Anglo-Australian Observatory*



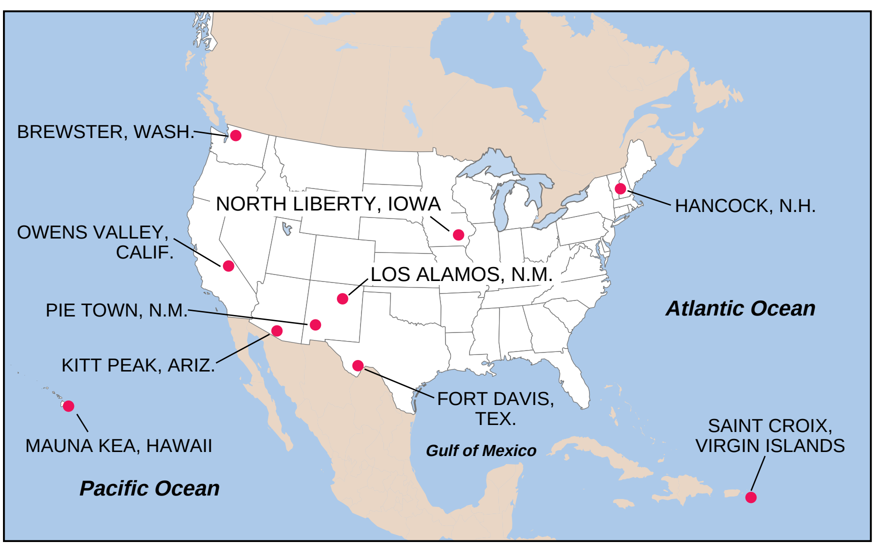
DIMITRY SCHIDLOVSKY, AFTER LINCOLN GREENHILL

SPIRAL GALAXY M33 is the site of at least six maser regions. (The colors of the circles designating the masers indi-

cate their velocity after correcting for the motion of M33: red is receding; blue is approaching.) Radio astronomers have measured the motion of these masers across the sky to determine the distance to M33. Images of masers are much more detailed than are their optical counterparts; each region contains many sources whose motions can be plotted (*inset, left*).



ROGER RESSMEYER *Starlight*



JOHNNY JOHNSON

RADIO TELESCOPE (*left*) in Owens Valley, Calif., is one of 10 that make up the Very Long Baseline Array. This array, distributed across the U.S. from Hawaii to the Virgin Islands (*map, bottom*), can act in effect as a single radio telescope, producing images of unprecedented resolution.

enter the middle one. The population of the longer-lived middle state increases until it becomes greater than that of the bottom one: a population inversion. Atoms or molecules can drop from the middle state to the bottom one by emitting a photon, but without stimulated emission they do so only rarely.

Population inversions can also appear in more complex systems, with four or more energy levels through which atoms or molecules circulate. Indeed, virtually all interstellar masers involve such complex configurations. Inversions are very difficult to set up on the earth because longer-lived middle states are hard to come by. In a gas at room temperature and pressure, collisions enforce the Boltzmann distribution of particles among energy states: the population of higher-energy states decreases exponentially. Any atom or molecule in an excited state will transfer its internal energy to a lower-energy neighbor when the two collide, and thus any inversion will be smoothed out quickly.

Deviations from this equilibrium, including population inversions, are possible only when the density is lower than a critical minimum. The density of ordinary air is roughly 2×10^{19} particles per cubic centimeter, but a maser operating on the wavelengths favored by interstellar hydroxyl can work only with about 100,000 particles per cubic centimeter, some 100 trillion times fewer. Although comparable to the best vacuum that can be created in a laboratory, this density is rather high for interstellar space and found only in those regions that stand out as clouds between the stars.

Masers Provide Unique Information

Amplified radiation gives observers a view of astronomical objects that differs strongly from the picture produced by the unamplified variety. Its two most important characteristics are high brightness and small size. Astronomers measure the intensity of a radiation source in terms of brightness temperature—how hot an object in thermal equilibrium (a so-called black body) would have to be to emit the same amount of radiation at a given wavelength. The brightness temperature of the sun's surface is approximately 5,800 kelvins, but that of some masers exceeds 10^{15} kelvins.

Such spectacular brightness must result from substantial amplification. Although any volume with an inverted population amplifies the radiation propagating through it, amplification increases sharply with the number of particles that a photon encounters during its passage. The density of maser regions must be low to keep the collision rates down and to allow population inversion; consequently, the extent of these regions must be enormous (on terrestrial scales) to provide enough molecules for amplification. A typical interstellar maser spans more than 150 million kilometers, roughly the radius of the earth's orbit around the sun.

Large though such a volume may be compared with a laboratory maser, it is minuscule compared with interstellar distances or the size of features that astronomers can observe by other means. The giant molecular clouds that house star-forming regions are typically between 10 and 100 parsecs across (a parsec is 3.26 light-years, or about 30 trillion kilometers). Their cores measure perhaps a few parsecs across, and the star-forming regions themselves, about a parsec. Masers might originate from a volume no more than a trillion kilometers across—the smallest distance that can be resolved by other instruments. Clusters of masers span perhaps 100 billion kilometers, and single maser spots are 1,000 times smaller still.

One reason the individual maser spots occupy such a small fraction of the clouds is that only a small fraction of the molecules in a region can participate in coherent maser emission. Because most material in interstellar clouds partakes in rapid, turbulent motions, many of the molecules cannot interact effectively with the radiation emitted at a given location. When a moving molecule emits a photon, the Doppler effect shifts the photon's wavelength (this effect also accounts for the "red shifts" of distant, rapidly receding galaxies). If a photon's wavelength changes by more than a tiny amount, corresponding to a relative motion of perhaps 1,000 meters per second, it will be unable to stimulate other molecules to radiate, and no maser amplification will occur. Instead maser amplification takes place only along paths where by chance the molecules all have sufficiently similar velocities. Most strong astronomical maser sources appear as collections of many small, bright spots, like a cloud of intensely bright fireflies. Each one radiates at a frequency corresponding to a different, well-defined Doppler shift that marks the spot's velocity relative to the earth.

Some gas clouds, however, move in a more ordered fashion—perhaps the most common example is the stellar wind blowing out from the surface of a red giant star. In this late phase of stellar evolution, stars shed gas like great leaky balloons as they puff up to radii more than 1,000 times that of our sun. (The sun itself will become a red giant in another four billion years or so.) These stellar winds are rich in molecules that emit maser radiation. Each maser molecule has a different set of energy levels; each radiates from the part of the stellar wind where conditions trigger its particular population inversion. Hydroxyl molecules emit maser radiation from a shell about 100 billion miles away from the star, about 25 times the distance from the sun to Pluto. Water vapor masers shine from a region roughly 10 billion miles from the star, and silicon monoxide masers from just above the edge of the stellar atmosphere. By tuning to the wavelength of each maser in turn, astronomers can probe red giant winds in considerably more detail than they can that of many other distant structures.

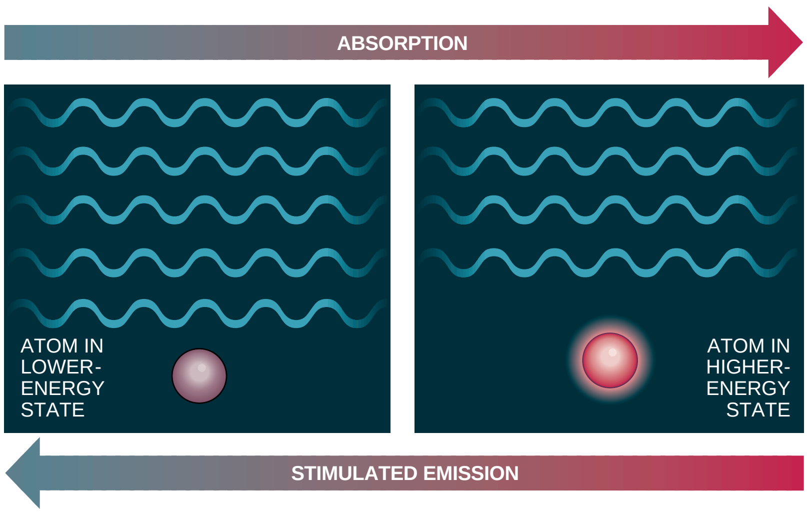
The plot of intensity versus frequency for a hydroxyl maser around a red giant has a distinct double-peak profile. The separation between the "blue" and "red" peaks typically corresponds to a velocity difference of about 10 miles per second. (Because the radiation consists of microwaves, it is not really blue or red, but the terminology is rooted in the Doppler shifts of visible light. Blue is at the high-frequency end of the spectrum and denotes an approaching source; red is at the low-frequency end

The Physics of Stimulated Emission

Stimulated emission, the foundation of maser operation, results from an important property of virtually all known physical laws: invariance under time reversal. If a microscopic process, such as the collision of two molecules, is recorded by a video camera, it is impossible to tell whether a playback is being run in the forward or the reverse direction. (Macroscopic events do not display this invariance because of the statistical behavior of systems containing many particles.)

In particular, consider an atom that can inhabit one of two distinct energy states [*see illustration below*]. There is no way to distinguish between a sequence of events in which the atom absorbs a photon from the surrounding radiation field and jumps into its excited state and the time-reversed interaction in which it adds a photon to the field and jumps to the lower state. Both events obey conservation of energy; both are in principle equally likely.

Both absorption and stimulated emission require an almost perfect match between the photon frequency and the velocity and energy levels of the interacting atom. As a result, the photon that a stimulated atom gives off is a twin of the one that caused the emission. Within a gas cloud whose particles are moving at nearly the same velocity, one photon with a particular frequency and direction can thus become two, two can become four and so on. The shower of photons that makes up an astronomical maser therefore maintains close coherence of frequency and direction.



DIMITRY SCHIDLOVSKY

and denotes a receding one.) This signature is so distinctive that it has been used to identify red giants at various locations in the galaxy even when the optical emission from the star itself is obscured by intervening dust.

The profile arises naturally from the radial motion of the stellar wind. Hydroxyl molecules in different sectors of the shell move in different directions and thus have large relative velocities. They cannot interact radiatively. Conversely, molecules on a given radial line move in the same direction at similar speeds; they are almost at rest with respect to one another. As a result, photons emitted by a molecule will affect only other molecules on the same line, and maser amplification is possible only for radiation propagating inward and outward along a line through the center of the shell.

The hydroxyl maser emission pattern resembles a hedgehog, with rays sticking out like spines from the gas cloud. At any given location, an observer can detect the emission only from the two regions on the shell along the line of sight to the central star. The signal from the front is blue-shifted, and that from the back is red-shifted. Each region is a small, caplike section of a sphere.

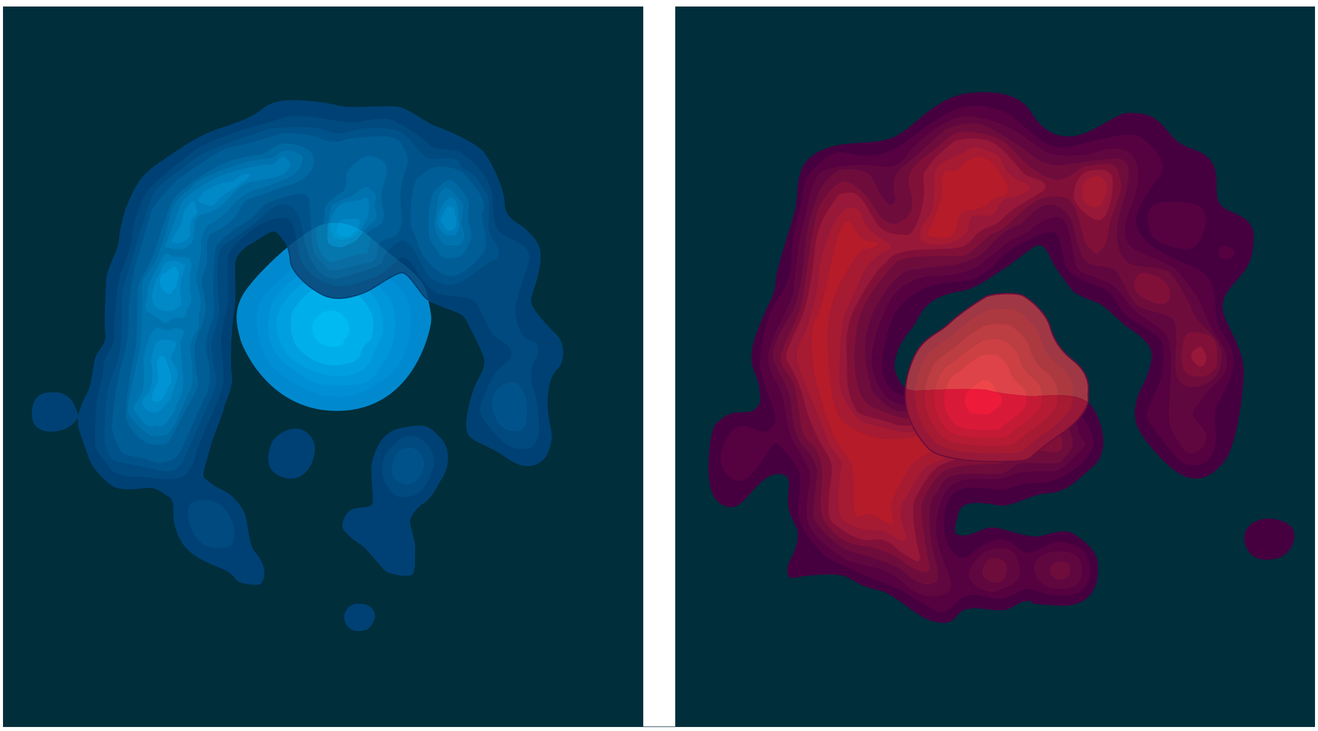
A team led by Roy S. Booth of Jodrell Bank in England produced an elegant, direct confirmation of this explanation by observing a star known as OH127.8. The researchers mapped both the peaks of maser emission and the weaker, less strongly Doppler-shifted "shoulders" between the peaks. Their maps show clearly that the radiation in the two intensity peaks comes from compact, well-defined caps on the line of sight. The weaker radiation from the shoulders covers a larger circular region, as expected from an expanding shell. The incomplete and clumpy appearance of the shell reflects the maser region's deviations from pure spherical symmetry. Such irregularities are to be expected in a turbulent stellar wind.

An Astronomical Yardstick

In addition to broadcasting details of the stellar wind, red giant masers also provide information about the distance to their associated stars. Distance determination is one of the most difficult problems in astronomy—indeed, the cosmic distance scale is still uncertain. Astronomers can measure objects' angular size with exquisite precision, but only seldom can they also determine the associated linear size. In the rare cases when both are measured, the distance to the source is directly determined—it is simply the ratio of linear to angular size.

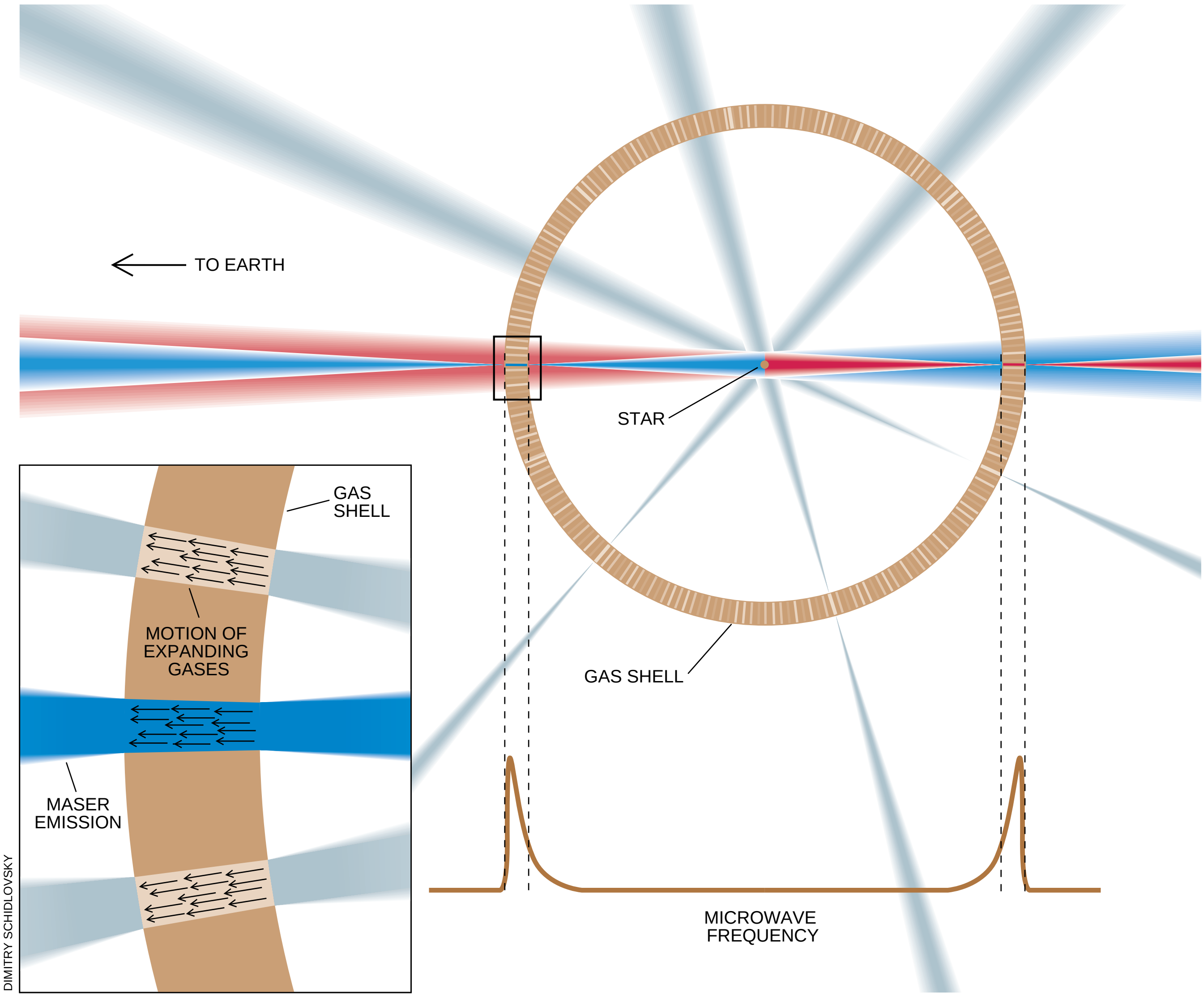
Hydroxyl maser shells around red giants furnish one of these rare opportunities. Many red giant stars vary regularly in brightness, with periods about a year long. Stellar radiation is the pump for the hydroxyl population inversion, and so the maser intensity should follow the star's temporal pattern. The red- and blue-shifted lines show a matching variation, but the red-shifted one is delayed by several weeks, the time required for its signal to cross from the back of the shell to the front. The precise length of the delay marks the shell's diameter. By simultaneously measuring the shell's angular diameter and the delay between red- and blue-shifted signals, astronomers can determine the distance to the star. Jaap Herman and Harm J. Habing of Leiden Observatory in the Netherlands have used maser observations to determine distances to several red giants in our galaxy. They point out that distance measurements to other galaxies may also be possible.

Radio telescopes are far better suited for making such measurements than are optical ones because they can resolve much smaller features. The angular resolution of optical telescopes is limited by atmospheric distortions and reaches a plateau around one arc-second—larger telescopes simply gather more light. In contrast, every increment in radio telescope size produces finer resolution.



DIMITRY SCHIDLOVSKY

RADIO EMISSIONS appear from an expanding gas cloud around a red giant star, shown here in separate views for the parts of the cloud between the star and the earth (*left*) and for those on the far side of the star (*right*). The Doppler effect causes photons from gases approaching the earth to be blue-shifted and those from receding gases to be red-shifted. Maser radiation is visible only from two caplike regions directly along the line of sight to the star.

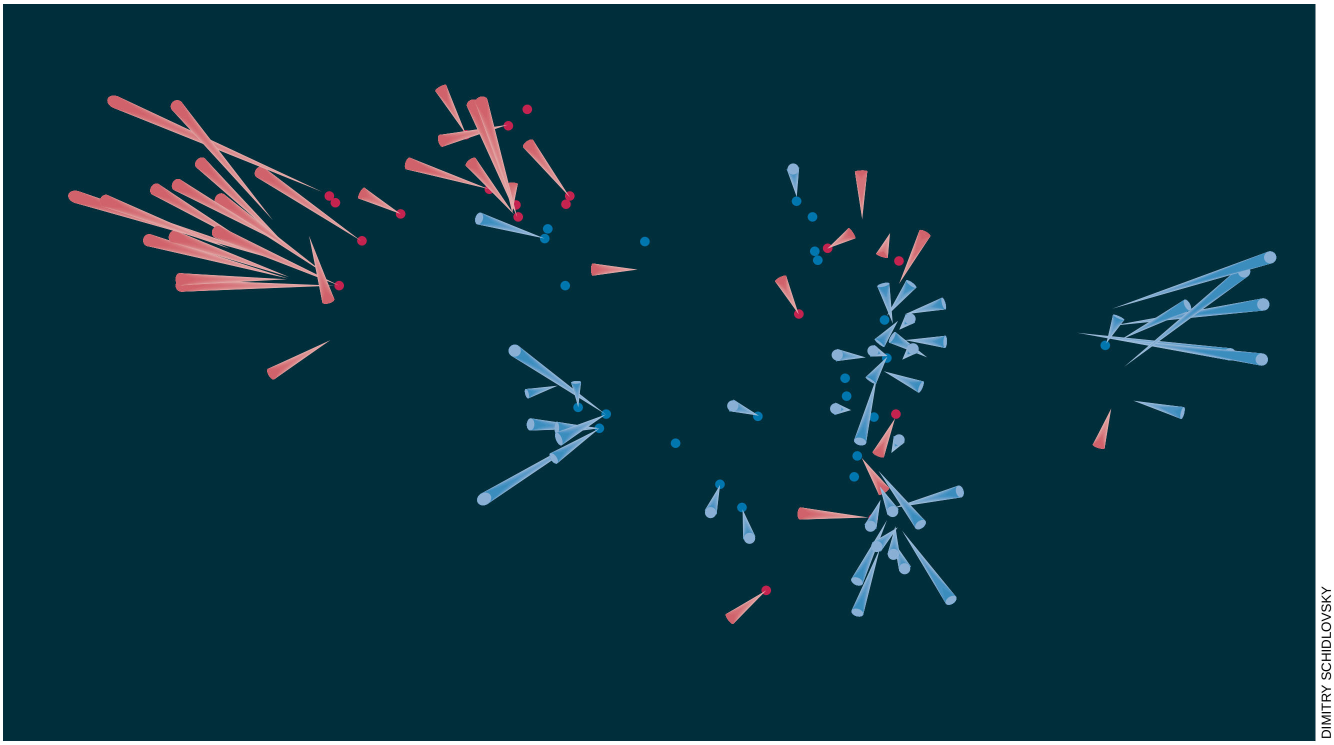


HYDROXYL MASER SHELL around a red giant is depicted in a schematic side view. Because amplification takes place only within a region of gas that is all moving at roughly the same velocity (*inset*), most maser emissions occur along radial lines (only some appear here). Thus, the only maser radiation seen from the earth comes from two small parts of the shell. Maser emissions from different parts of the shell can be distinguished by their Doppler shifts; the characteristic double peak (*bottom*) is a signature of red giants that can be detected even if the star itself is invisible to optical telescopes.

Radio astronomers have developed methods for combining signals from radio telescopes scattered around the globe, creating an effective aperture 8,000 miles across. When this technique, known as very long baseline interferometry (VLBI), is applied to sources of radio emission lines, only masers have sufficient intensity and the narrow line width necessary for correlating the signals from multiple receivers. VLBI has elevated astronomical angular resolution measurements to new heights: maser images can be made accurate to 300 microarc-seconds (were the human eye to have this resolving power, you could see these words from about 3,000 miles away). Moreover, radio telescopes can resolve the separation between neighboring maser spots even more closely—researchers can distinguish sources as close to one another as 10 microarc-seconds. Radio transmitters on the earth and the moon would register as separate spots from roughly 1,000 light-years away.

Such extraordinary precision makes feasible the determination of distances to maser sources other than red giants. Young stars, still in the process of formation inside dust clouds, may be invisible from the earth, but the masers they power shine clearly in the radio sky. Measuring the distance to these stellar nurseries is more complicated because astronomers cannot assume that the maser sources are symmetrically arranged within an expanding spherical shell. For any single object, there is no way of knowing the relation between its velocity along the line of sight and the transverse velocity that gives rise to its motion across the sky. Instead astronomers must study a cluster of maser sources; they then can construct equations that relate the velocities of all the members of the cluster. Once the relative motions have been found, the angular velocities and Doppler shifts can be correlated to determine the most probable distance to the cluster.

Working with high-resolution VLBI, astronomers can detect the apparent motions of masers in months instead of the decades that it takes to track optical sources. A group centered at the Harvard-Smithsonian Center for Astrophysics, led by James M. Moran and Mark Reid, has pioneered observations of water vapor maser motions. The an-



STAR-FORMING REGION (*above*) contains many masers, all emanating from the gas cloud around a very young star. Each maser seen in the map is color-coded according to whether it is approaching the earth or receding from it; the length of the cones indicates the direction and velocity of each maser's motion. By correlating the masers' velocities along the line of sight with their angular motion, astronomers can determine the distance to the region.

gular velocities they found have been likened to the growth of an astronaut's toenails on the moon or the motion of snails on Jupiter. Carl R. Gwinn of the University of California at Santa Barbara recently produced a map of the most luminous maser in our galaxy, a star-forming region called W49(N). The map charts the positions of each water vapor maser feature, along with its most probable future position and its speed of recession or approach. The spread of speeds and directions makes it apparent that the maser cluster is expanding from a common center; the complete three-dimensional structure of the velocity field can be modeled accordingly. Gwinn has estimated that the cluster resides at a distance of 10.4 kiloparsecs, with an uncertainty of 1.3 kiloparsecs.

The group has also employed this method to model water vapor maser motions in Sagittarius B2(N), a star-forming region very near the center of our galaxy. The maser spots were spread over an area roughly 0.3 light-year across, about eight kiloparsecs (25,000 light-years) from the earth. This number is the only direct measurement of the distance to the galactic center and has been adopted by the International Astronomical Union.

In 1993 members of the Harvard-Smithsonian group reported the first measurements of water vapor maser motions in another galaxy, M33. In the near future they expect to be able to calculate the galaxy's distance, removing significant sources of error in the cosmic distance scale. More recently they mapped maser sources in M106, a more distant spiral galaxy. Analysis of the rotation of the maser sources around their common center has yielded a tentative figure of about 5.4 megaparsecs (17.6 million light-years).

A Distant Vision

Unlike other scientists, we astronomers can neither manipulate nor touch most of the objects we study. The only information we gather on celestial objects, except for those in the solar system, is from the radiation they emit. Maser radiation, when it can be detected, carries unique information about interstellar structures, and it permits observations of small-scale astrophysical phenomena that cannot be seen in any other way.

Continuous progress in interferometric techniques allows researchers to probe increasingly smaller details. Recently U.S. radio astronomers inaugurated the Very Long Baseline Array, an array of 10 identical telescopes distributed across American territory from Hawaii to the Virgin Islands. Dedicated to interferometric observations, the array can operate as a single instrument with superb resolving capabilities.

Radio astronomers in Japan and Russia are currently engaged in plans for even more ambitious VLBI networks, including a telescope orbiting in space. The completion of the Very Long Baseline Array and the prospects for space-based interferometry ensure that the resolution of these "interstellar microscopes" will continue to improve. Maser studies (just like the sources they look at) will continue to expand, providing irreplaceable insights into the inner workings of interstellar clouds, evolving stars and distant galaxies.

FURTHER READING

COMPACT MASER SOURCES. R. J. Cohen in *Reports on Progress in Physics,* Vol. 52, No. 7, pages 881–943; July 1989.

ASTRONOMICAL MASERS. M. J. Reid and J. M. Moran in *Galactic and Extragalactic Radio Astronomy.* Edited by G. L. Verschuur and K. I. Kellerman. Springer-Verlag, 1991.

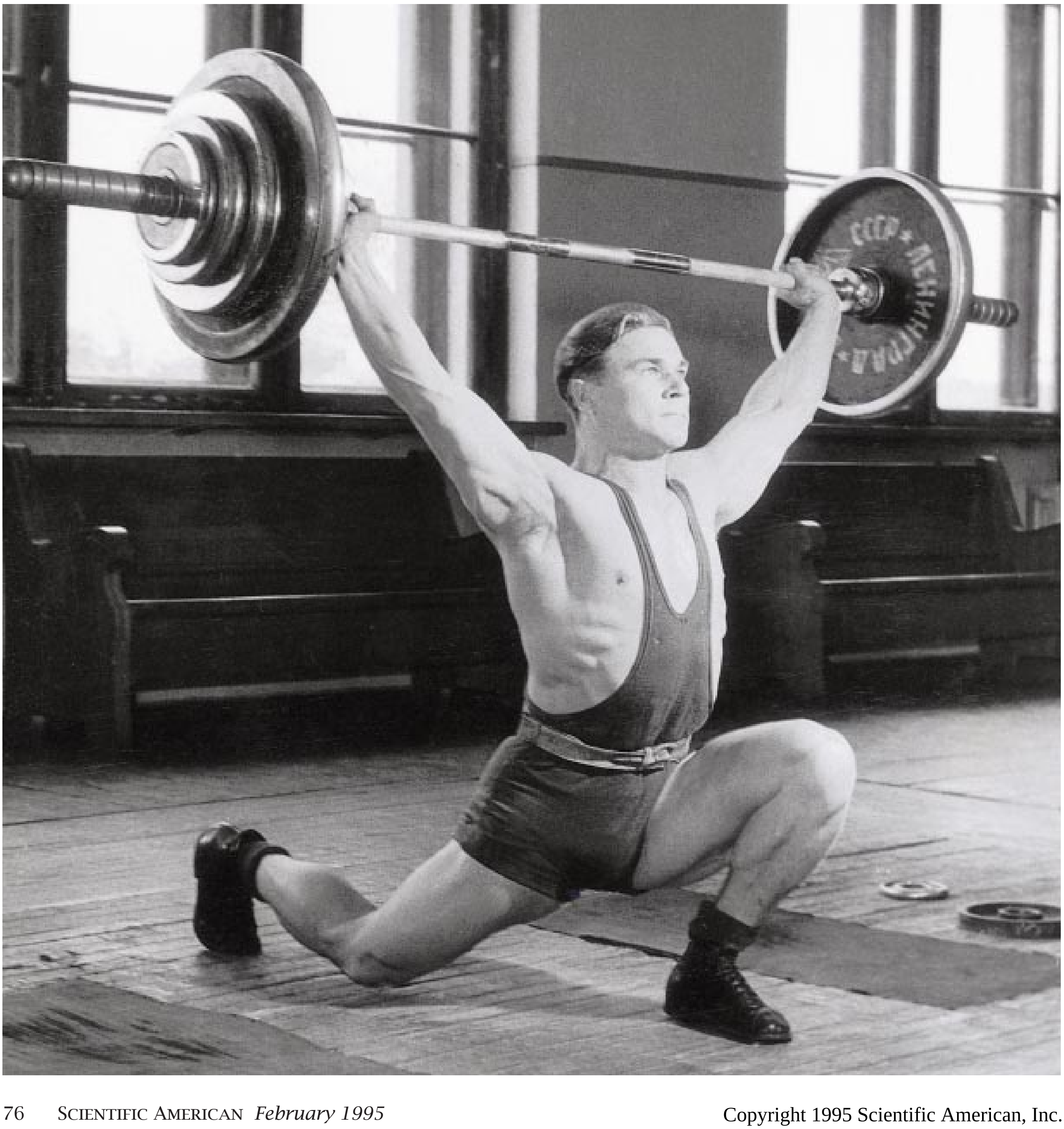
ASTRONOMICAL MASERS. M. Elitzur. Kluwer Academic Publishers, 1992.

ASTROPHYSICAL MASERS. Edited by A. W. Clegg and G. E. Nedoluha. Springer-Verlag, 1993.

The History of Synthetic Testosterone

Testosterone has long been banned in sports as a performance-enhancing drug. This use may soon be accepted in medicine alongside other legitimate hormonal therapies

by John M. Hoberman and Charles E. Yesalis



V. NOSKOV AND N. MAXIMOV *courtesy of Terry Todd*

On June 1, 1889, Charles Édouard Brown-Séquard, a prominent French physiologist, announced at the Société de Biologie in Paris that he had devised a rejuvenating therapy for the body and mind. The 72-year-old professor reported that he had drastically reversed his own decline by injecting himself with a liquid extract derived from the testicles of dogs and guinea pigs. These injections, he told his audience, had increased his physical strength and intellectual energy, relieved his constipation and even lengthened the arc of his urine.

Almost all experts, including some of Brown-Séquard's contemporaries, have agreed that these positive effects were induced by the power of suggestion, despite Brown-Séquard's claims to the contrary. Yet he was correct in proposing that the functions of the testicles might be enhanced or restored by replacing the substances they produce. His achievement was thus to make the idea of the "internal secretion," initially proposed by another well-known French physiologist, Claude Bernard, in 1855, the basis of an organotherapeutic "replacement" technique. Brown-Séquard's insight that internal secretions could act as physiological regulators (named hormones in 1905) makes him one of the founders of modern endocrinology. So began an era of increasingly sophisticated hormonal treatments that led to the synthesis in 1935 of testosterone, the primary male hormone produced by the testicles.

Since then, testosterone and its primary derivatives, the anabolic-androgenic steroids, have led a curious double life. Since the 1940s countless elite athletes and bodybuilders have taken these drugs to increase muscle mass and to intensify training regimens. For the past 25 years, this practice has been officially proscribed yet maintained by a $1-billion international black market. That testosterone products have served many therapeutic roles in legitimate clinical medicine for an even longer period is less well known. Fifty years ago, in fact, it appeared as though testosterone might become a common therapy for aging males, but for various reasons it did not gain this "legitimate" mass-market status. Perhaps most important, physicians were concerned that these drugs often caused virilizing side effects when administered to women, including a huskier voice and hirsutism.

SOVIET MIDDLEWEIGHT Vasily Stepanov is shown practicing in 1955. Many Soviet weightlifters began taking testosterone products to build strength before their American competitors did. At the 1954 World Championships the physician for the Soviet team told John Ziegler, his American counterpart, about the efficacy of these drugs. Ziegler then tested them on himself and several athletes in the U.S.

Today, however, there is compelling evidence that these spheres of "legitimate" and "illegitimate" testosterone use are fusing. Further research into the risks and the medical value of anabolic-androgenic steroids is under way. Indeed, scientists are now investigating the severity of such reported temporary short-term side effects as increased aggression, impaired liver function and reproductive problems. And some physicians are currently administering testosterone treatments to growing numbers of aging men to enhance their physical strength, libido and sense of well-being. Our purpose here is to describe the largely forgotten history of male hormone therapy that has culminated in the prospect of testosterone treatments for millions of people.

Organotherapy

Brown-Séquard provided samples of his *liquide testiculaire* free of charge to physicians willing to test them. The offer generated a wave of international experiments aimed at curing a very broad range of disorders, including tuberculosis, cancer, diabetes, paralysis, gangrene, anemia, arteriosclerosis, influenza, Addison's disease, hysteria and migraine. This new science of animal extracts had its roots in a primitive belief that came to be known as *similia similibus,* or treating an organ with itself. Over many centuries since ancient times, physicians had prescribed the ingestion of human or animal heart tissue to produce courage, brain matter to cure idiocy and an unappetizing array of other body parts and secretions—including bile, blood, bone, feces, feathers, horns, intestine, placenta and teeth—to ameliorate sundry ailments.

Sexual organs and their secretions held a prominent place in this bizarre therapeutic gallery. The ancient Egyptians accorded medicinal powers to the testicles, and the Roman scholar Pliny the Elder reports that the oil-soaked penis of a donkey or the honey-covered penis of a hyena served as sexual fetishes. The *Āyurveda* of Suśruta (circa 1000 B.C.) recommended the ingestion of testis tissue as a treatment for impotence. Johannes Mesuë the Elder (A.D. 777–857) prescribed a kind of testicular extract as an aphrodisiac. The *Pharmacopoea Wirtenbergica,* a compendium of remedies published in 1754 in Germany, mentions horse testicles and the phalluses of marine animals. These therapeutic exotica are significant because they dramatize the impossibility, for ancients and moderns alike, of separating sexual myth from sexual biology.

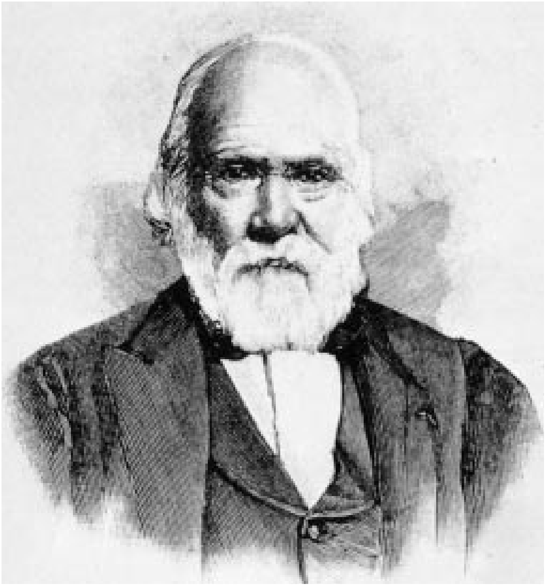
Two of the researchers inspired by Brown-Séquard's work were the Austrian physiologist Oskar Zoth and his compatriot Fritz Pregl, a physician who eventually turned to analytic chemistry and received a Nobel Prize in 1923. When sports physiology was in its infancy, these men investigated whether testicular extracts could increase muscle strength and possibly improve athletic performance. They injected themselves with a liquid extract of bull's testicles and then measured the strength of their middle fingers. A Mosso ergograph recorded the "fatigue curve" of each series of exercises.

Zoth's 1896 paper concluded that the "orchitic" extract had improved both muscular strength and the condition of the "neuromuscular apparatus." Most scientists now would say these were placebo effects, a possibility these experimenters considered and rejected. Yet the final sentence of this paper—"The training of athletes offers an opportunity for further research in this area and for a practical assessment of our experimental results"—can lay claim to a certain historical significance as the first proposal to inject athletes with a hormonal substance.

The growing popularity of male extracts prompted other scientists to search for their active ingredient. In 1891 the Russian chemist Alexander von Poehl singled out spermine phosphate crystals, first observed in human semen by the microscopist Anton van Leeuwenhoek in 1677 and again by European scientists in the 1860s and 1870s. Poehl claimed correctly that spermine occurs in both male and female tissues, and he concluded that it increased alkalinity in the bloodstream, thereby raising the blood's capacity to transport oxygen.

This was an interesting observation insofar as hemoglobin does pick up oxygen in a slightly alkaline environment

JOHN M. HOBERMAN and CHARLES E. YESALIS share an interest in the history of performance-enhancing drugs. Hoberman, a professor of Germanic languages at the University of Texas at Austin, has written often about the history of sports medicine and high-performance athletics. Yesalis is professor of health policy and administration and exercise and sports science at Pennsylvania State University. He studies the nonmedical uses of anabolic-androgenic steroids and other muscle-building drugs.


MUSÉE D'HISTOIRE DE LA MÉDICINE, PARIS

CHARLES ÉDUOARD BROWN-SÉQUARD, a 19th-century French physiologist, claimed to have reversed his own aging process by injecting himself with testicular extracts. The injections could not have rejuvenated him, as he declared, but for his insight he is considered a founder of modern endocrinology.

and releases it when the pH is slightly acidic. But he was incorrect in that no chemical mediates the binding of oxygen to hemoglobin. Still, Poehl believed he had improved on Brown-Séquard's work, for if spermine did accelerate oxygen transport, then it could claim status as a "dynamogenic" substance, having unlimited potential to enhance the vitality of the human organism. As it turned out, spermine's function remained unknown until 1992, when Ahsan U. Khan of Harvard Medical School and his colleagues proposed that it helps to protect DNA against the harmful effects of molecular oxygen.

Testicle Transplants

Between the flowering of spermine theory before World War I and the synthesis of testosterone two decades later, another sex gland therapy debuted in the medical literature and made wealthy men of a few practitioners. The transplantation of animal and human testicular material into patients suffering from damaged or dysfunctional sex glands appears to have begun in 1912, when two doctors in Philadelphia transplanted a human testicle into a patient with "apparent technical success," as a later experimenter reported. A year later Victor D. Lespinasse of Chicago removed a testicle from an anesthetized donor, fashioned three transverse slices and inserted them into a sexually dysfunctional patient who had lost both of his own testicles. Four days later "the patient had a strong erection accompanied by marked sexual desire. He insisted on leaving the hospital to satisfy this desire." Two years later the patient's sexual capacity was still intact, and Lespinasse described the operation as an "absolutely perfect" clinical intervention.

The most intrepid of these surgeons was Leo L. Stanley, resident physician of San Quentin prison in California. Stanley presided over a large and stable population of testicle donors and eager recipients. In 1918 he began transplanting testicles removed from recently executed prisoners into inmates of various ages, a number of whom reported the recovery of sexual potency. In 1920 "the scarcity of human material," Stanley wrote, prompted him to substitute ram, goat, deer and boar testes, which appeared to work equally well. He performed hundreds of operations, and favorable word-of-mouth testimony brought in many patients seeking treatment for an array of disorders: senility, asthma, epilepsy, diabetes, impotence, tuberculosis, paranoia, gangrene and more. Having found no ill effects, he concluded that "animal testicular substance injected into the human body does exert decided effects," including "relieving pain of obscure origin and promotion of bodily well-being."

Early organotherapy of this kind existed on the boundary separating legitimate medicine from quackery. Stanley's work, for example, was respectable enough to appear in the journal *Endocrinology.* Like Brown-Séquard, he complained about the "'lost manhood' charlatans" and "medical buccaneers" who navigated "this poorly charted sea of research" in a half-blind state and sometimes pursued financial gain rather than medical progress. Yet Stanley himself performed operations without hesitation and was persuaded by much ambiguous evidence. And the controversial "monkey gland" transplants performed by Serge Voronoff during the 1920s earned this Russian-French surgeon a considerable fortune.

In an appreciative retrospective monograph, the medical historian David Hamilton argues for Voronoff's sincerity at a time when endocrinology was a new field and medical ethics committees were few and far between. Although medical journals sounded regular warnings against "marvel mongering," "haphazard, pluriglandular dosing" and "extravagant therapeutic excursions," they expressed some cautious optimism as well. Given the limited knowledge and therapeutic temptations of this era, these treatments are better described as cutting-edge medicine than as fraud.

The Isolation of Testosterone

Before Stanley and his fellow surgeons started performing transplant operations, other scientists had begun searching for a specific substance having androgenic properties. In 1911 A. Pézard discovered that the comb of a male capon grew in direct proportion to the amount of animal testicular extracts he injected into the bird. Over the next two decades researchers used this and similar animal tests to determine the androgenic effects of various substances isolated from large quantities of animal testicles or human urine. Their quest entered its final stage in 1931, when Adolf Butenandt managed to derive 15 milligrams of androsterone, a nontesticular male hormone, from 15,000 liters of policemen's urine. Within the next few years, several workers confirmed that the testes contained a more powerful androgenic factor than did urine—testosterone.

Three research teams, subsidized by competing pharmaceutical companies, raced to isolate the hormone and publish their results. On May 27, 1935, Kàroly Gyula David and Ernst Laqueur and their colleagues, funded by the Organon company in Oss, the Netherlands (where Laqueur had long been the scientific adviser), submitted a now classic paper entitled "On Crystalline Male Hormone from Testicles (Testosterone)." On August 24 a German journal received from Butenandt and G. Hanisch, backed by Schering Corporation in Berlin, a paper describing "a method for preparing testosterone from cholesterol." And on August 31 the editors of *Helvetica Chimica Acta* received "On the Artificial Preparation of the Testicular Hormone Testosterone (Androsten-3-one-17-ol)" from Leopold Ružička and A. Wettstein, announcing a patent application in the name of Ciba. Butenandt and Ružička eventually shared the 1939 Nobel Prize for Chemistry for this discovery.

The struggle for the synthetic testosterone market had begun. By 1937 clinical trials in humans were already under way, employing injections of testosterone propionate, a slow-release derivative of testosterone, as well as oral doses of methyl testosterone, which is broken down in the body more slowly than is testosterone. These experiments were initially as haphazard and unregulated as the more primitive methods involving testicular extracts or transplants. In its early phase, however, synthetic testosterone therapy was reserved primarily for treating men with hypogonadism, allowing them to develop fully or maintain secondary sexual characteristics, and for those suffering from a poorly defined "male climacteric" that included impotence.

Testosterone, Women and Sports

Early synthetic testosterone products were also applied to a variety of female complaints, such as menorrhagia, painful breast conditions, dysmenorrhea and estrogen-driven breast cancers, on the grounds that testosterone neutralized estrogen. For about a century, physicians have recognized that altering the hormonal balance in certain women can cause their metastatic breast tumors to regress. Today it is accepted that about a third of all women with breast cancer have "hormone-dependent" tumors; androgen therapy serves as a second- or third-choice treatment for postmenopausal women with advanced breast cancers. In contrast, the androgen treatments of the 1940s were administered to women of various ages at a time when the mechanism of their antitumor effect was even less well understood than it is now. A clinically valid observation from this period, however, was that androgens could relieve pain, increase appetite and weight and promote a sense of well-being even if they failed to arrest tumor growth.

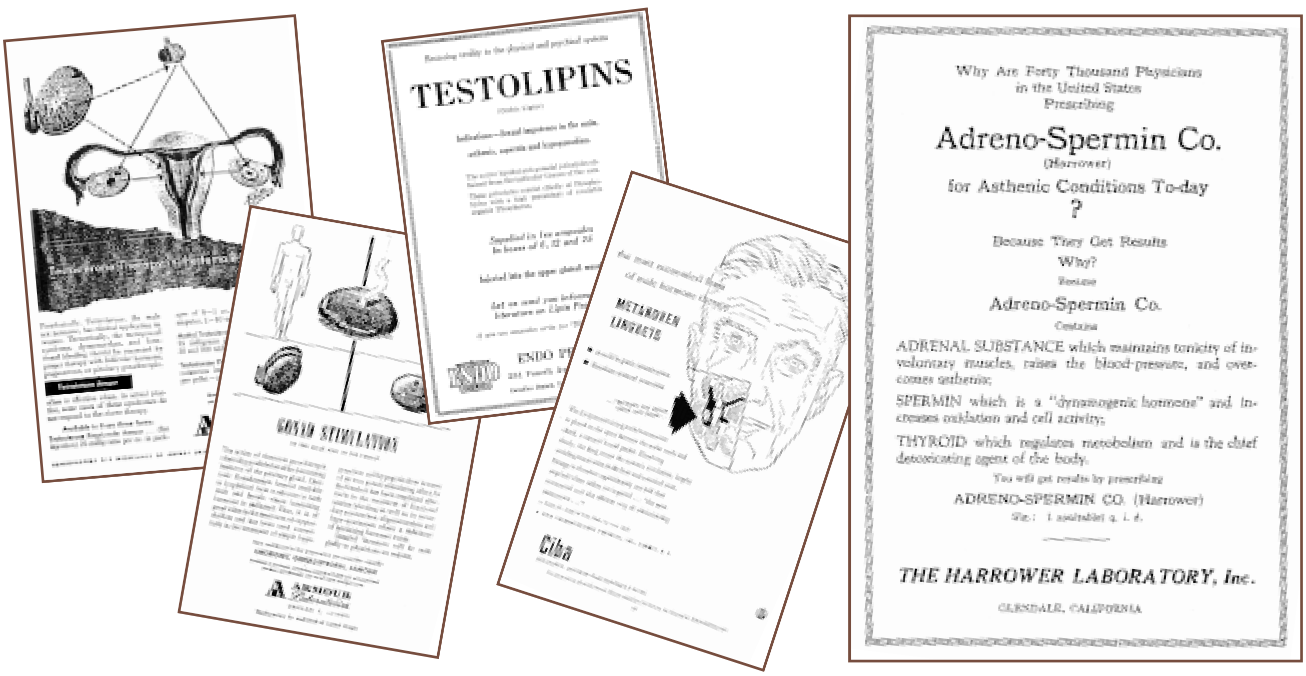
A consequence of treating women with testosterone was the discovery that androgens could renew or intensify female libido in most patients. One investigator reported in 1939 that the daily application of a testosterone ointment had enlarged the clitoris of a married woman who was then able to achieve orgasm. More commonly, subfascial pellets and injections were used to achieve similar effects, and the massive doses given to some breast cancer patients rarely failed to intensify their sex drive.

The use of testosterone to enhance female sexual response did not, however, become standard therapy. Currently it appears that only a small number of physicians in the U.S., and a greater proportion in Britain and Australia, use androgens for this purpose. As mentioned, testosterone therapy did not catch on in part because of certain side effects. Then, as now, some patients experienced reawakened sexual urges as emotionally disruptive and unwelcome. The most important impediment to a general testosterone therapy, though, was that clinicians wanted an anabolic steroid that would not virilize their female patients, giving them a deeper, husky voice, hair on the face and body, and an enlarged clitoris. Although not all physicians were alarmed by these symptoms, different assessments of them, including whether they might be irreversible, led to heated exchanges in professional journals.

The idea that testosterone could counteract the effects of estrogen led to its use as a therapy for male homosexuals (a goal the transplant surgeons had embraced in the early 1920s). "It is clearly evident that the estrogenic values are higher among the homosexuals," wrote one research team in *Endocrinology* in 1940, concluding that "the constitutional homosexual has different sex hormone chemistry [from] the normal male." In 1944 another group described "a series of clinical trials of organotherapy" involving 11 "overt homosexuals who applied for treatment for various reasons." In one Orwellian turn of phrase, they revealed that four subjects had "accepted organotherapy by compulsion"—a court order in one case and parental injunctions in the other three.

The organotherapy, which was uncontrolled by a placebo group, was a failure. Indeed, given that five subjects complained of increases in their sexual drive, the researchers conceded the likelihood that "the administration of androgen to the active (or aggressive) homosexual would rather regularly intensify his sex drive" instead of reducing it. Yet even this obstacle did not entirely extinguish their *furor therapeuticus.* "The results in appropriate cases," they wrote, "are too good to permit undue pessimism as to the value of this treatment."

Also during the 1940s, scientists dis-



ADVERTISEMENTS for male hormone products have appeared in American medical journals for many years. The earliest one shown (*far right*), which ran in the *Journal of the National Medical Association* in 1924, boasts that "Adreno-Spermin Co." is "dynamogenic," a term that Brown-Séquard associated with the animal testicular extracts he himself prepared.

covered that testosterone could facilitate the growth of muscle tissue. Charles D. Kochakian, a pioneer in synthetic hormone research, reported as early as 1935 that androgens stimulated the protein anabolic processes, offering the possibility that androgen therapy might restore protein tissue and stimulate growth in patients suffering from a spectrum of disorders. The clinical literature of the early 1940s often discussed the correlation between androgens and heightened muscularity, including speculations about the use of these drugs to boost athletic performance. One group of researchers decided in 1941 "to investigate whether the endurance in man for muscular work could be increased by testosterone" and obtained positive results. In 1944 another scientist wondered whether "the reduction of working capacity with age might proceed differently if the sex-hormone concentration could be artificially maintained at a higher level."

The writer Paul de Kruif popularized many of these findings in *The Male Hormone,* published in 1945. This widely read book may have helped promote testosterone use among athletes. According to anecdotal reports, West Coast bodybuilders began experimenting with testosterone preparations in the late 1940s and early 1950s. News of the efficacy of these drugs apparently spread during the early 1960s to other strength-intensive sports, from the throwing events of track and field to football.

Over the past 30 years anabolic steroid use has entered other Olympic sports, including hockey, swimming, cycling, skiing, volleyball, wrestling, handball, bobsledding and soccer. Steroid use is well documented among male athletes in college and high school. Of the estimated one million steroid abusers in the U.S., many take these drugs for noncompetitive bodybuilding. Drug-testing programs, designed to suppress steroid use in sports, have been seriously flawed since they were first implemented in the 1970s. These procedures often lack the sensitivity needed to catch drug users, and many elite athletes and corrupt sports officials have learned to avoid detection.

Clinical Uses of Testosterone

Some of the clinical uses of testosterone products date from the earliest period of androgen therapy. The most frequent and accepted application of anabolic steroids has been as a replacement therapy for men with hypogonadism. They have also been administered to treat impotence in patients with normal and below normal serum testosterone levels. Testosterone esters are frequently employed to stimulate growth and to initiate puberty in boys experiencing a significant developmental delay. Since the 1940s androgens have been used to treat wasting conditions associated with chronic debilitating illnesses (such as those suffered by victims of Nazi concentration camps) and trauma (including battle injuries), burns, surgery and radiation therapy.

Because anabolic steroids increase red blood cell production (erythropoiesis), they were the first-choice therapy for a variety of anemias before bone marrow transplantations and synthetic erythropoietin treatments became common. And from the late 1930s to the mid-1980s psychiatrists prescribed anabolic steroids for the treatment of depression, melancholia and involutional psychoses. Testosterone esters are now routinely used as an adjunct to human growth hormone (hGH) therapy for children who are hGH deficient. Most recently some physicians have begun testing anabolic steroids as a treatment for the weakness and muscle wasting that occurs during the progression of HIV infection and AIDS. Clinical case studies are promising and indicate that these patients experience an improved sense of well-being and an increase in strength, lean mass and appetite.

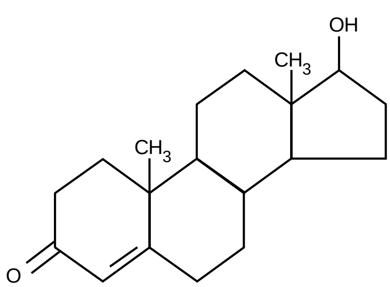
In addition, since the late 1970s testosterone esters have been evaluated as a possible method to regulate male fertility via the endocrine feedback loop. The hypothalamus reacts to high levels of testosterone in the blood by reducing the release of yet another hormone, luteinizing hormone–releasing hormone, which via the pituitary gland affects not only the body's production of testosterone but also of sperm. In 1990 the World Health Organization reported results from a 10-center, global trial that established the efficacy of anabolic steroids as a male contraceptive that produces minimal short-term physical side effects. It is interesting to note that the doses prescribed for these subjects exceeded those taken by the banned Olympic sprinter Ben Johnson. This comparison suggests that legitimizing anabolic steroids as male contraceptives would weaken the medical argument against their routine use by athletes.

During the late 1980s, researchers again began evaluating the effects of testosterone on "successful" aging, motivated in part by a graying society and favorable preliminary results of hGH supplementation in healthy older men. During the early 1990s, several scientists conducted pilot studies of the effects of testosterone supplementation in men over 54 years old who had either low or normal testosterone levels. The results were generally positive, including a gain in lean body mass and strength, a possible decline in bone resorption (with the potential to reverse or improve frailty), an increase in reported sexual desire and activity, and better spatial cognition and word memory.

Because most physicians intuitively accept the efficacy of hormonal replace-

Anabolic-Androgenic Steroids

The anabolic-androgenic steroids are all synthetic derivatives of testosterone (*molecular structure below*), the natural male hormone produced primarily by the testes. Women also produce testosterone, but in lower amounts than do men. The hormone is responsible for the androgenic, or masculinizing, and anabolic, or tissue-building, effects noted during male adolescence and adulthood. The main androgenic effects in males include the growth of the reproductive tract and the development of secondary sexual characteristics. In the pubertal male the latter is charted by increases in the length and diameter of the penis, development of the prostate and scrotum, and the appearance of the pubic, axillary and facial hair.



The anabolic effects are those that take place in the somatic and nonreproductive tract tissue, including thickening of the vocal cords, an acceleration of linear growth appearing before bony epiphyseal closure, enlargement of the larynx and development of libido and sexual potentia. An increase in muscle bulk and strength as well as a decrease in body fat also occurs.


RONALD C. MODRA *Sports Illustrated*

BEN JOHNSON was stripped of his gold medal in the 1988 Olympics in Seoul, South Korea, after drug tests revealed that he had taken anabolic steroids to enhance his performance. The doses this Canadian sprinter took, however, were allegedly lower than what the World Health Organization subsequently found to be safe to administer as a male contraceptive. The comparison draws into question the medical arguments against steroid use in sports.

ment therapy in women, they may readily adopt a comparable hormone therapy for men. Implicit cultural acceptance of mass male hormone therapy seems evident in the fact that over the past several years the lay press has broadcast and printed numerous reports on the potential benefits of both testosterone and estrogen therapy for the aging population. The Hormonal Healthcare Center in London administers testosterone injections to hundreds of men irrespective of age, and a gynecologist at Chelsea and Westminster Hospital in London currently prescribes testosterone pellets for about 25 percent of his postmenopausal patients. This trend is likely to continue, meaning that mass testosterone therapy could become standard medical practice within a decade.

This prediction is based on the fact that popular expectations and commercial motives can help define new medical "disorders." In 1992, for example, the National Institutes of Health requested proposals for research on whether testosterone therapy can prevent physical ailments and depression in older males, thereby raising the question of whether the aging process itself is about to be officially recognized as a treatable deficiency disease. John B. McKinlay, director of the New England Research Institute in Watertown, Mass., and a specialist on aging, has offered the following prognosis: "I don't believe in the male midlife crisis. But even though in my perspective there is no epidemiological, physiological or clinical evidence for such a syndrome, I think by the year 2000 the syndrome will exist. There is a very strong interest in treating aging men for profit, just as there is for menopausal women."

Commercial interest in response to the public's demand for androgens could cause physicians to overlook possible deleterious side effects and overestimate their clinical value. For example, in the January 1994 issue of the *Journal of Urology,* McKinlay and his colleagues stated that there was no correlation between any form of testosterone and impotence, a "major health concern" affecting a potential market of 18 million men for whom testosterone has long been prescribed on a much smaller scale. But failing to confirm the value of testosterone for one disorder is unlikely to deter its use to strengthen aging bodies or restore a waning interest in sex. Indeed, aging is increasingly being viewed as a medical problem, and this shift is leading to the recognition of a "male menopause" as treatable as its female counterpart. The official status of such a syndrome will signify new societal definitions of physiological normality and further legitimize ambitions to boost the human organism to higher levels of mental and physical performance.

FURTHER READING

THE MALE HORMONE. Paul de Kruif. Harcourt, Brace and Company, 1945.

ANABOLIC-ANDROGENIC STEROIDS. Charles D. Kochakian. Springer-Verlag, 1976.

THE MONKEY GLAND AFFAIR. David Hamilton. Chatto and Windus, 1986.

MORTAL ENGINES: THE SCIENCE OF PERFORMANCE AND THE DEHUMANIZATION OF SPORT. John Hoberman. Free Press, 1992.

ANABOLIC STEROIDS IN SPORT AND EXERCISE. Edited by Charles E. Yesalis. Human Kinetics Publishers, Champaign, Ill., 1993.

The Mid-Cretaceous Superplume Episode

The earth has an erratic “heartbeat” that can release vast amounts of heat from deep within the planet. The latest “pulse” of the earth occurred 120 million years ago

by Roger L. Larson

At one o’clock in the morning on December 13, 1989, I was awakened in my bunk on board the scientific drillship *JOIDES Resolution* by the sounds of celebration in the adjoining cabin. Since I had to relieve the watch at four anyway, I stumbled next door to join the party. The paleontologists in our expedition had just reported to my co–chief scientist, Yves Lancelot, now at the University of Aix-Marseilles, that microfossils of the Jurassic period had been recovered from the hole in the floor of the western Pacific Ocean that we were drilling more than three miles below us. Two days later the drill reached the volcanic basement—oceanic crust of Middle Jurassic age, about 165 million years old. A 20-year mystery was solved. At last, we had hard evidence of the world’s oldest deep-sea sediments and volcanic rocks that are still in place from eons ago.

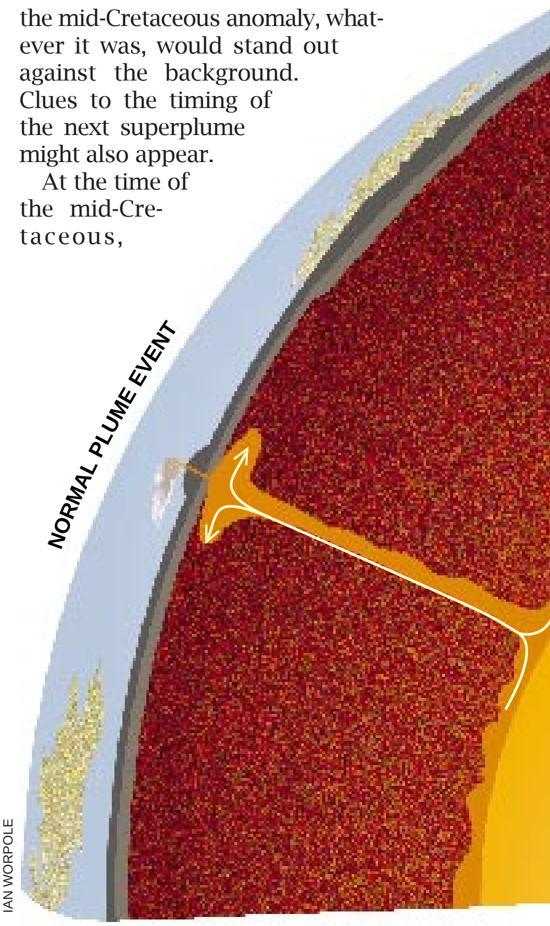
In succeeding days I reflected on why the quest had taken so long. My colleagues Clement G. Chase of the University of Arizona, Walter C. Pitman III of Lamont-Doherty Geological (Earth) Observatory, Thomas W. C. Hilde of Texas A&M University and I had first considered the problem in the 1970s. The target was not a small one. We had predicted from geophysical data that an area in the western Pacific the size of the continental U.S. should be Jurassic in age, somewhere between 145 and 200 million years old. But whenever we dredged or drilled in this area, we almost invariably recovered rocks called basalts, formed by volcanic eruptions during the mid-Cretaceous, generally ranging in age from 80 to 120 million years and no older. The first such basalt samples were dredged from the Mid-Pacific Mountains in 1950 by an early expedition of the Scripps Institution of Oceanography. Until the *JOIDES* discovery, however, geologists had not made much progress in answering the questions concerning the origin of the seemingly ever present mid-Cretaceous basalts or the possible existence of underlying Jurassic material.

The 1989 discovery provided some qualitative answers. The older sediments and oceanic crust were buried during the mid-Cretaceous epoch by what we now refer to as a “superplume” of volcanic material. Finally, our geophysical musings of the early 1970s could be supported with facts: the Jurassic existed in the western Pacific. We had samples of it locked away on board the *JOIDES Resolution.*

Because I am a geophysicist, I try to describe the earth and its processes quantitatively. I wanted to determine the size of the mid-Cretaceous superplume of the western Pacific, hoping to learn something of its origins. But saying that and doing it are two different things. What do you measure, and how do you measure it? I did not even know what “normal” was, so how could I describe the “anomalous” mid-Cretaceous superplume episode? The problem had to be expanded beyond the time and space framework of the mid-Cretaceous western Pacific. I decided to examine the rate of formation of oceanic crust—mainly volcanic rocks such as basalts that make up the solid basement underneath the seafloor—for all the ocean basins over their entire histories. Then the mid-Cretaceous anomaly, whatever it was, would stand out against the background. Clues to the timing of the next superplume might also appear.

At the time of the mid-Cretaceous,

ROGER L. LARSON first became acquainted with the oceans when he left Iowa State University with a bachelor’s degree in geology. He headed west in a new 1965 Ford Mustang and five years later earned a Ph.D. in oceanography from the Scripps Institution of Oceanography at the University of California, San Diego. He became interested in the tectonic history of the western Pacific in 1971 as a research associate at the Lamont-Doherty Geological (Earth) Observatory of Columbia University. That interest has continued during Larson’s work as a professor of oceanography at the University of Rhode Island. He has served as chief or co–chief scientist on a total of 13 oceanographic expeditions to the region. He still drives the Mustang.



SUPERPLUMES build vast areas of oceanic plateaus and seamounts (*right*), compared with the small region affected by normal plumes (*left*). The plumes are shown in a progressive sequence of

widespread volcanic eruptions covered or created vast amounts of ocean floor very quickly. Typically, though, seafloor spreading generates most of the oceanic crust in a slower, more regular way. In this process the crust becomes older symmetrically away from mid-ocean ridges where molten magma rises up out of the earth's mantle and then cools and solidifies. As new magma continues to rise, the older oceanic crust is rafted away from the eruption center and onto the flanks of the ridge. Thus, any particular parcel of crust is transported as if it were on one of two identical conveyor belts moving away from the mid-ocean ridge in opposite directions [see "The Mid-Ocean Ridge," by Kenneth C. Macdonald and Paul J. Fox; SCIENTIFIC AMERICAN, June 1990].

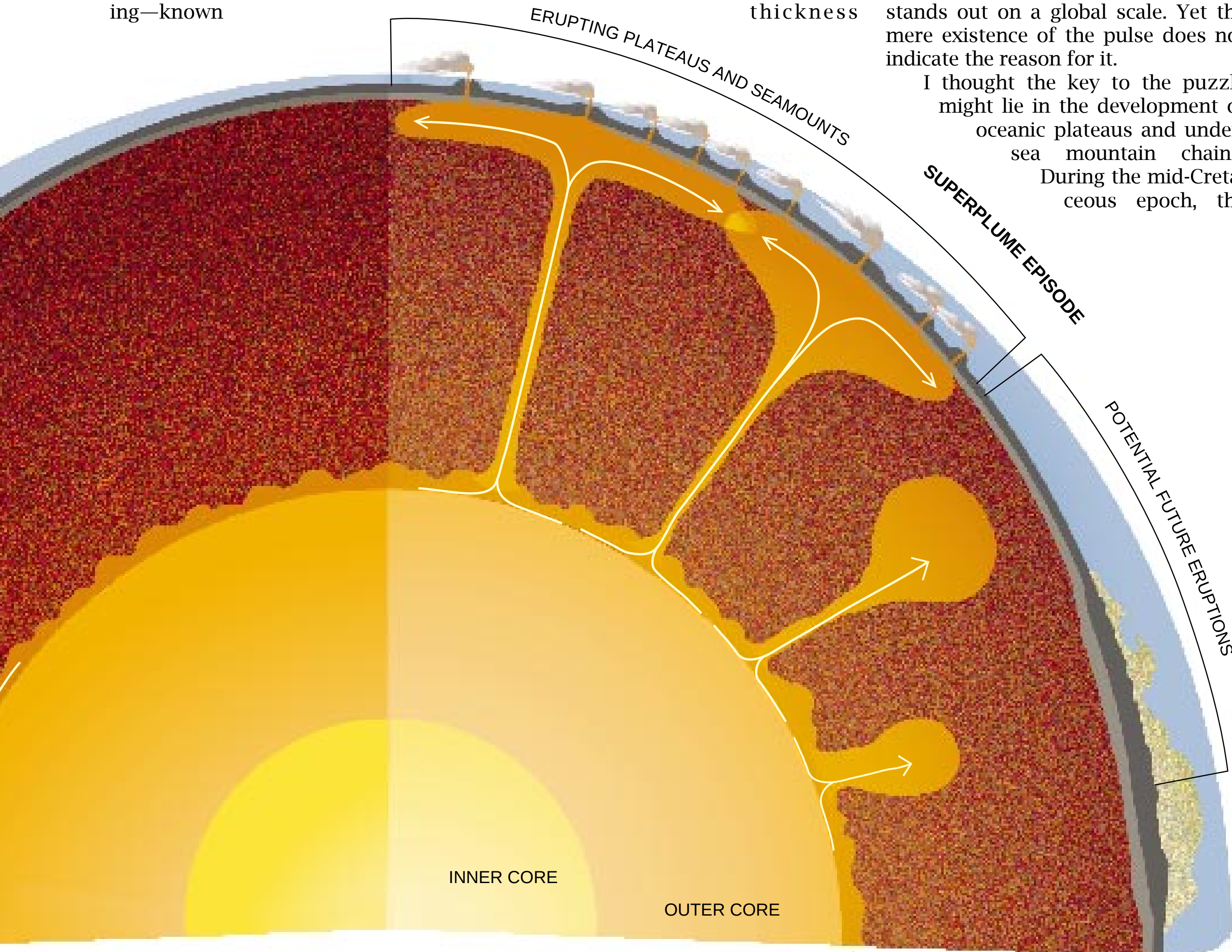
Areas of the ocean floor formed by spreading—known as abyssal plains—are covered with organized processions of abyssal hills and fracture zones running perpendicular to the mid-ocean ridges. Yet the western Pacific looks nothing like this. Its physiography is more like a muddy New England road in March. The seemingly randomly oriented chains of seamounts, taller than abyssal hills, and the oceanic plateaus that make up the "muddy road" of the western Pacific have no systematic age gradients across them. The only characteristic they share is that they are almost all from the mid-Cretaceous, to the extent that we even know their ages.

The first step in my investigation was to measure the changing rate of production of oceanic crust. In order to do this, I compiled information on the areas and ages of ocean floor and estimated the thickness of the crust. I was able to calculate this rate for the past 150 million years, nearly back to the maximum age of the world's ocean basins. These calculations of overall crustal production clearly show the mid-Cretaceous superplume [*see illustration on page 85*].

The Onset of the Pulse

The world-total histogram shows the plume's sudden onset 120 to 125 million years ago, when formation of ocean crust doubled in about five million years. Crustal production peaked soon after the onset of the pulse and then tapered more or less linearly over the next 70 to 80 million years. It returned to values nearly the same as those before the episode 30 to 40 million years ago. The mid-Cretaceous superplume in ocean crustal production stands out on a global scale. Yet the mere existence of the pulse does not indicate the reason for it.

I thought the key to the puzzle might lie in the development of oceanic plateaus and undersea mountain chains. During the mid-Cretaceous epoch, the



events, which occurs for both superplumes and normal plumes: birth at the thermal boundary layer, ascent through the mantle, flattening at the base of the lithosphere and, finally, eruption at the surface. The actual geography of the mid-Cretaceous superplume event in the western Pacific would show the plumes in a more irregular order.

"MUDDY ROAD" of the western Pacific seafloor results from the intense volcanic activity of the mid-Cretaceous superplume event, which produced randomly oriented plateaus and undersea mountain chains. The ocean floor of the eastern Pacific, in contrast, shows the smooth, lineated physiography that is characteristic of crust formed by seafloor spreading. (From *World Ocean Floor Map,* by Bruce C. Heezen and Marie Tharp, 1977.)

rate of production of these formations jumped at the same time as the overall rate did, with a similar impulsive onset and a long succeeding taper back to normal values. Although the maximum amplitude at the height of this pulse was less than that for the world-total curve, the relative increase was much larger. Whereas total output of oceanic crust initially doubled, plateau and seamount production increased by a factor of five. So whatever produced the superplume episode also had the strongest effect on plateau and mountain chain generation.

What causes these undersea plateaus and seamount chains? Independently, other investigators have converged on the notion that they result from plumes of material from deep in the earth's mantle that have been overheated and thus rise buoyantly because of their reduced density [see "Large Igneous Provinces," by Millard F. Coffin and Olav Eldholm; SCIENTIFIC AMERICAN, October 1993]. In particular, oceanic plateaus result from the initial massive, rapid eruptions caused by these rising plumes. Such upwellings occasionally occur on the continents where we can study them directly. Exotically named regions such as the Paraná Basalts of Brazil, the Deccan Traps of western India and the Siberian Traps of northern Russia consist of vast fields of basalt flows, several hundred kilometers across and one or two kilometers thick [see "Volcanism at Rifts," by Robert S. White and Dan P. McKenzie; SCIENTIFIC AMERICAN, July 1989]. The oceanic plateaus are features similar to their continental cousins, but they are even larger. For instance, the largest of the oceanic plateaus (the Ontong-Java Plateau of the western Pacific) is estimated to be 25 times bigger than the largest continental one (the Deccan Traps).

Seamount chains trail away from oceanic plateaus and result from material behind and below the head of the rising plume material. Because the plumes are relatively fixed, and the overlying tectonic plates drift horizontally on the surface of the planet, the subsequent eruptions in the mountain chains record the motions of the plates. Thus, these seamount chains should be oldest next to their parent oceanic plateau and trace a path of younger and younger seamounts that ends in an active volcano if the "chimney pipe" to the deep mantle is still alive. The best known of these seamount chains is the Hawaiian Islands, which extends underwater far to the northwest of the islands themselves. Its rising plume system exists today below the island of Hawaii, where volcanic eruptions continue to rumble. The islands and seamounts become successively older to the northwest as they are rafted away on the Pacific plate, which moves in a northwesterly direction over a fixed plume location.

Once I realized that the features of the ocean crust most affected by the mid-Cretaceous volcanic activity—seamount chains and plateaus—were each formed by plumes of mantle material, it was a small logical step to suppose that the entire anomaly resulted from plume activity on a much larger than normal scale. Because I live in a superlative-prone society, I named this a "superplume episode." The initial pulse of the superplume reached the earth's surface around 120 million years ago; the intense volcanic activity started suddenly and continued through the mid-Cretaceous, lasting tens of millions of years, gradually tapering off after that.

Overheated Plumes

The superplume episode was most likely caused by the upwelling of one or perhaps several enormous plumes that ascended through the easily deformed mantle, spread out at the base of the earth's more rigid outer shell, known as the lithosphere, and erupted onto the ocean floor. Although the Pacific was most strongly affected, evidence of the superplume event is also present in the Indian, South Atlantic and Caribbean oceans. The area of the Pacific involved may have been several thousand kilometers across, in sharp contrast with the size of regions affected by today's plume activity, which are usually one tenth the size in area.

I suspect the overheated plumes rise from the very base of the mantle and affect the process that causes reversals of the earth's magnetic field in the underlying outer core. There is a general inverse correlation between the production rate of crust formed by plumes and the frequency of reversals of the earth's magnetic field. For example, during periods of intense plume activity, including during the mid-Cretaceous, almost no magnetic reversals take place. Conversely, as is the case today, when plume activity is low, magnetic reversals occur at record pace. How the earth's magnetic field actually reverses its polarity is a mystery. Peter L. Olson of Johns Hopkins University and I think the correlation between crustal formation and magnetic-field reversals may provide a clue to understanding how the reversals take place and to determining the source of the mantle plume material. We believe an increase in the "boiling rate" of the core somehow causes magnetic reversals to become more infrequent. Additionally, the connection may reveal information about the advent of the next superplume.

Boiling iron within the outer core is almost certainly the source of the earth's magnetic field. Such molten iron is an excellent electrical conductor, and the convective motion of the iron and its associated electrical field generates the earth's magnetic field. The heat given off by the molten iron percolates through the core-mantle boundary, the lid to this boiling pot, by the process of conduction. The heat becomes trapped just above the boundary in the lowermost 100 to 200 kilometers of solid silicate rock of the mantle. This process continues until enough excess heat accumulates. Then the buoyancy of the overheated, less dense lower mantle overcomes the viscosity of the overlying cooler, more dense mantle rock. Huge plumes of mantle material rise nearly 3,000 kilometers through the mantle and eventually trigger volcanic eruptions at the surface. Ascending material removes heat from the lowermost mantle, allowing the outer core to boil even more vigorously than before.

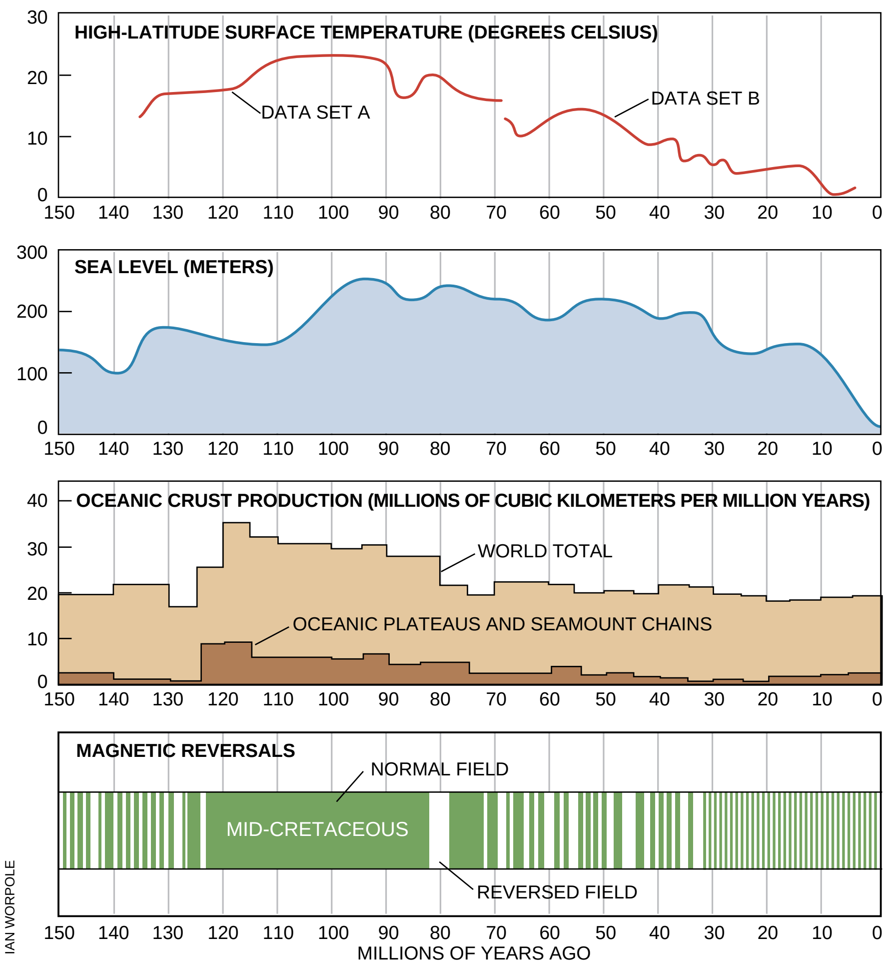
Global Effects

The most recent of these major overturns erupted 120 to 125 million years ago as the mid-Cretaceous superplume episode. Much of the material that surfaced at this time left the "muddy road" effect seen today on the western Pacific seafloor. Such an episode that doubles the world-total rate of oceanic crustal production in a short period must have staggering geologic consequences. The mid-Cretaceous was characterized by several profound anomalies resulting from the superplume.

First and probably least controversial is the rise in worldwide sea level to an elevation 250 meters or so higher than it is today. Assuming that the total amount of seawater in the planet's oceans is constant, a rise in the level of the sea surface is simply a reflection of a corresponding rise in the level of the seafloor. Ocean above newly formed crust is abnormally shallow because the crust and underlying lithosphere are still relatively warm, less dense and therefore expanded. As the two cool, they contract, allowing the seafloor to deepen. This phenomenon of expansion and contraction explains why oceanic ridges, where new crust is being formed, are raised above the older, deeper crust found on the flanks. If an abnormal amount of new crust is formed rapidly—as it was at the beginning of the mid-Cretaceous pulse—then the average seafloor level will be elevated, and the sea surface will rise accordingly. In the mid-Cretaceous, rising sea levels drowned much of what is dry land today; for example, my birthplace in Iowa was then at the bottom of the ocean. When the water receded, it left deposits of limestone and chalk, including the famous White Cliffs of Dover in England.

The earth's surface temperature also increased as a result of the superplume episode. When molten lava erupts, it releases certain chemicals, including carbon dioxide. Higher amounts of carbon dioxide in the mid-Cretaceous atmosphere led to a natural greenhouse effect that raised global temperatures by roughly 10 degrees Celsius. Studying the effects of elevated carbon dioxide levels during this period could reveal possible scenarios for the earth's climate in the future. Massive burning of fossil fuels and large-scale deforestation continue to increase the level of carbon dioxide in the modern atmosphere.

An excess amount of organic carbon and inorganic carbonate was also deposited during the mid-Cretaceous. The enhanced deposition is related to the elevations in sea level and air tempera-



GEOLOGIC CONSEQUENCES of the mid-Cretaceous superplume event include rising surface temperature and sea level. The superplume itself can be seen by the increase in the world-total rate of oceanic crust production; it is particularly evident in the rate of formation of oceanic plateaus and seamount chains. Additionally, reversals of the earth's magnetic field ceased during the superplume. At present, reversals occur frequently, indicating that plume activity is low.

ture, which, we have seen, resulted from the superplume episode. Tiny plants and animals, known as phytoplankton and zooplankton, make their living floating at shallow levels in the ocean where light can penetrate. Plankton apparently thrived during the mid-Cretaceous in the abnormally warm oceans that accompanied the natural warming of the atmosphere. Normally, when these organisms die, their bodies sink in the deep sea and quickly dissolve because of the extreme pressure of the overlying seawater. But during the mid-Cretaceous, many of the dead organisms fell instead on the drowned continents. The carbon from the skeletons did not dissolve in the shallow waters but was preserved. Some of it formed the White Cliffs, and some was buried more deeply and eventually turned to oil. The resulting oil constitutes up to 50 percent of the world's oil supply. Ironically, this outcome of the mid-Cretaceous greenhouse event may have created the fuel for the next greenhouse episode.

SAM OGDEN

KIMBERLITE DIAMONDS, such as these from West Africa, were transported to the earth's surface during the mid-Cretaceous superplume episode.

Other geologic anomalies associated with the mid-Cretaceous superplume include the placement of a very large percentage of the earth's diamond deposits. Diamonds are made of pure carbon atoms, squashed into the tightest, densest conceivable packing order by pressures that exist at least 200 to 300 kilometers below the earth's surface. Most diamonds are ancient even on geologic timescales, having formed more than a billion years ago, but, according to Stephen E. Haggerty of the University of Massachusetts at Amherst, many of them were brought to the surface during the mid-Cretaceous. They were transported up volcanic structures called kimberlite diamond pipes (after a mining area in Kimberley, South Africa) that extend deep down into the crust and presumably into the upper mantle. The diamonds were probably torn loose from their sources within the mantle by rising plumes and brought up in their solid, original state.

The formation of most of the mountain ranges that edge the western coasts of North and South America was strongly controlled by the superplume episode. The Sierra Nevada Mountains of western North America and the Andes Mountains of western South America were created during the mid-Cretaceous by increased subduction of Pacific crust underneath western North and South America. Subduction occurs close to the continents when the oceanic lithosphere is thrust below the adjacent landmass and recycled into the mantle below. Remember that because of the proximity of the erupting plumes rates of seafloor spreading in the Pacific increased dramatically. What comes up must go down if the earth's diameter remains constant, so as production of ocean floor increased, so did subduction rates. Abnormally large amounts of oceanic crust were thrust deep under the western coastlines of North and South America. As the crust and accompanying oceanic sediments sank several hundred kilometers below the earth's surface, the minerals with the lowest melting points became semiliquid as temperatures and pressures rose. Some of the adjacent continental crust also melted from frictional heating. This molten rock combination rose back to near the surface as its density lessened and then solidified to form the granite cores of the mountain ranges that are the spine of the west coast of the Americas.

The Next Pulse

Much of the earth's geologic history is controlled by events that have their true origins deep within the planet, some 3,000 kilometers below our feet. The story presented here hints at only a few of the dynamic processes and interactions that cause material from the core-mantle boundary to rise in sporadic pulses and affect the surface environment. The most recent of these bursts dramatically altered the terrestrial climate, surface structure, and fossil-fuel and mineral supply.

We are only now beginning to consider the evidence for previous superplume episodes, and an argument rages concerning the rate of the earth's past "heartbeat." Our planet has clearly settled down from the effects of the most recent superplume event, but when the next episode will occur is a matter of speculation. During the past 40 million years, creation of oceanic plateaus and seamount chains has proceeded at a much slower pace. Sea level has dropped to a near-record low. Because we are now living at a pause in the middle of an ice age, global temperature today could be described more reasonably as that of an "ice house" instead of a greenhouse. As expected during periods of little plume activity, the earth's magnetic field is reversing more frequently than ever, and a large temperature anomaly exists at the base of the mantle, indicating a slow boiling rate in the core below. Current estimates are that temperature increases by 1,000 to 1,500 degrees C through the lowermost 100 to 200 kilometers of the mantle.

It has been 120 million years since the last superplume event, and it is merely a matter of time before the next one. Yet we cannot predict when it will happen. As scientists, we are in a situation somewhat similar to that of a farmer trying to forecast the coming of spring when the groundhog does not see its shadow on the second of February. We can only say that the next superplume episode is "just around the corner."

FURTHER READING

THE MID-CRETACEOUS SUPER PLUME, CARBON DIOXIDE, AND GLOBAL WARMING. Ken Caldeira and Michael R. Rampino in *Geophysical Research Letters,* Vol. 18, No. 6, pages 987–990; June 1991.

A SUPERPLUME IN THE MANTLE. K. G. Cox in *Nature,* Vol. 352, No. 6336, pages 564–565; August 15, 1991.

SUPERPLUMES AND SUPERCHRONS. Mike Fuller and Robin Weeks in *Nature,* Vol. 356, No. 6364, pages 16–17; March 5, 1992.

THE MESOZOIC PACIFIC: GEOLOGY, TECTONICS, AND VOLCANISM. Edited by Malcolm S. Pringle, William W. Sager, William V. Sliter and Seth Stein. American Geophysical Union, 1993.

SUPERKIMBERLITES: A GEODYNAMIC DIAMOND WINDOW TO THE EARTH'S CORE. Stephen E. Haggerty in *Earth and Planetary Science Letters,* Vol. 122, No. 1–2, pages 57–69; March 1994.

Toward "Point One"

by Gary Stix, *staff writer*

At a technical conference this month, engineers will present papers describing a memory chip that can hold a billion bits of information and a microprocessor that performs in excess of a billion instructions per second, a 1,000-fold increase in capacity and processing power since the early 1980s. Organizers of the IEEE International Solid-State Circuits Conference expect the relentless progression of semiconductor technology to continue: they contend that papers on terachips (capable of handling a trillion bits or instructions) will arrive by the end of the next decade.

The optimism of the research establishment conceals the growing difficulties of converting laboratory samples into real manufacturing items. A pocket videophone or high-resolution teleconferencing system will depend on continuing advances in the greatest mass-production technique of all time. "An electronic memory circuit will have gone from $10 in the 1950s down to a hundred thousandth of a cent a few years from now," says Alec N. Broers, head of the engineering department at the University of Cambridge. "It's gone further than any technology in history."

But by all accounts, it has become more challenging to make chips faster and smarter by shrinking the size of transistors squeezed onto a few square centimeters of silicon. "We've never had the physics confronted so dramatically as now," says Richard R. Freeman, head of advanced lithography research at AT&T Bell Laboratories.

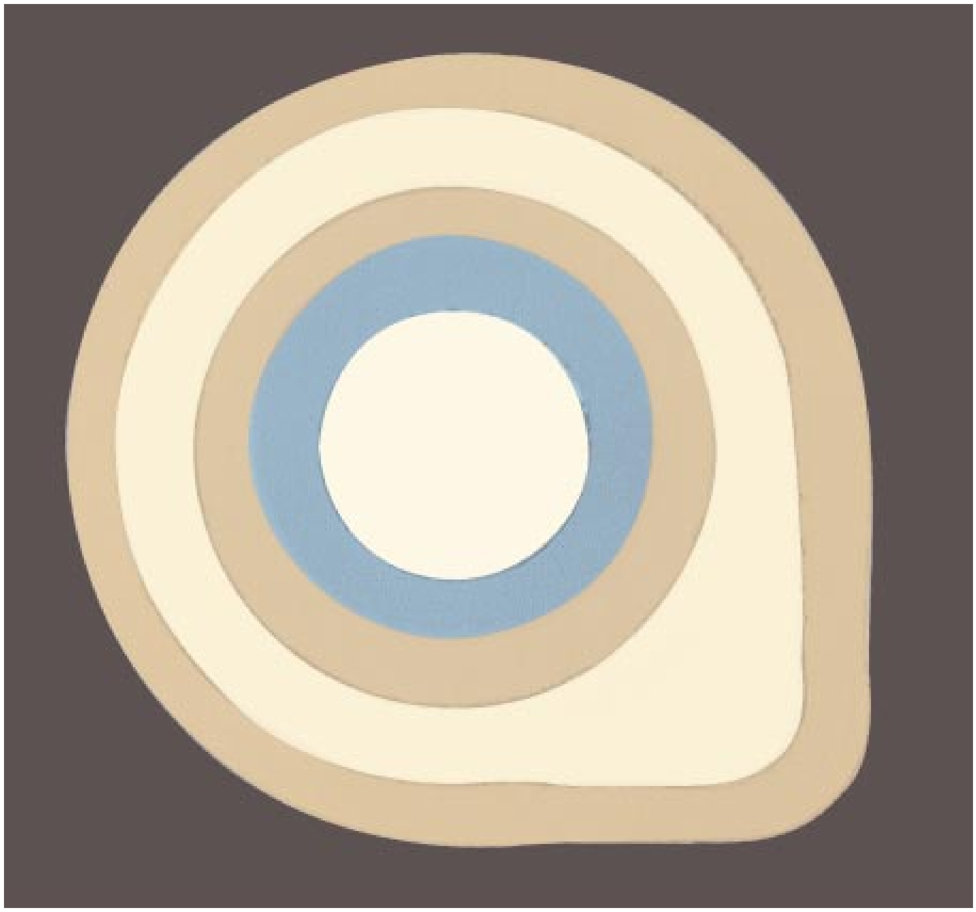
The physicists, chemists and engineers who study the lithographic-imaging process at the heart of chip manufacturing are having trouble deciding how to advance the technology. The smallest structural dimensions in sophisticated commercial memory chips being readied for the marketplace measure 0.35 micron—and they are getting smaller. (Memory chips have traditionally led in miniaturization, although microprocessors have begun to close the gap in recent years.)

Early in the next decade the semiconductor industry plans to sell gigabit chips now in the laboratory with a transistor electrical channel, or gate, that measures between 0.1 and 0.2 micron in length. "Point one," as the lesser decimal is known by the engineering cognoscenti, is small by any standard: it is about the width of a DNA coil, or a thousandth the width of a human hair.

At these tiny dimensions the photographic process for developing a circuit image on the surface of a chip starts to falter. Circuit patterns projected through giant $1-million lenses blur easily on the chip's surface. The ultraviolet light focused on the chip gets absorbed before it can print images of submicron transistors, capacitors and wires.

Photolithography has progressed beyond the most optimistic predictions. But if the light finally fades, lithographers may grudgingly have to consider a technology on which numerous corporate and university research careers have turned during more than two decades. Lithography using the short, nanometer wavelengths of x-rays may be the only means of fashioning circuits with billions of transistors. Last fall several U.S. companies, including IBM, AT&T and Motorola, formally launched a collaborative effort to share the development costs needed to bring x-ray lithography into the factory.

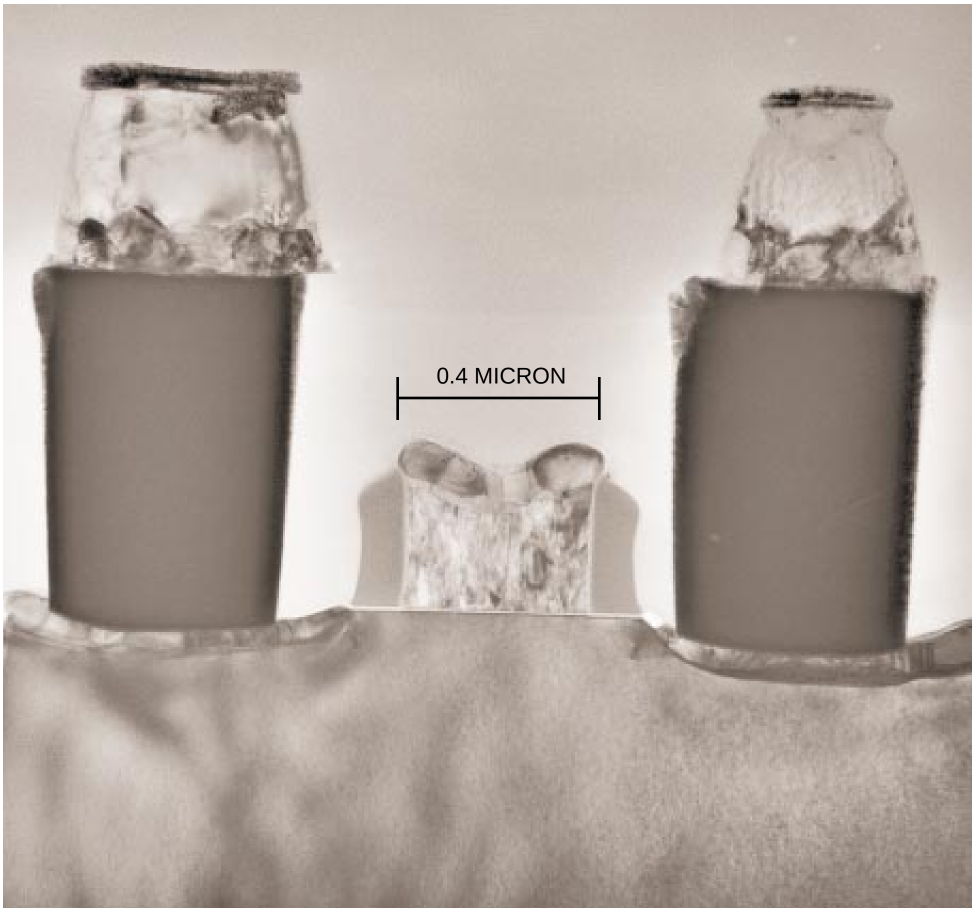
Champions of x-rays for making chips have not won many converts, even among lithographers within their own companies. In fact, the high costs and technical uncertainties of x-ray lithog-



INTEL CORPORATION

SMALL GETS SMALLER as witnessed in a then and now comparison of two transistors. The first planar transistor, vintage 1959, measured 764 microns in diameter and could be seen by the naked eye **(viewed from**

Gigabit chips are now in the laboratory. But the critical technology needed for manufacturing smaller circuits confronts diminishing returns



above, left). ***A contemporary transistor, shown in profile through a transmission electron microscope*** **(right),** ***measures about two microns across and has elements as small as 0.4 micron.***

raphy have prompted U.S. industry groups to contemplate spending $300 million to $500 million on a crash campaign to bring more conventional optical lithography to the point that it might be used in manufacturing a gigabit memory chip with a billion transistors or a microprocessor that cycles at billions of times per second. "I won't quote anybody, but I was in a meeting where people said that when we get out of optics we're out of the business," says Karen H. Brown, director of lithography for Sematech, the U.S. industry's research and development consortium.

Strains between these factions have set off a tug-of-war to obtain increasingly scarce government funding. "Money spent on x-ray lithography has come at the expense of advanced optical lithography," says John R. Carruthers, director of components research at Intel. If some means can be found to let optical technology prevail, the huge x-ray lithography development effort may be written off without its ever having produced a transistor sold on the commercial market.

Tensions between advocates of the two technologies—and inherent doubts about the myriad other technical problems to make circuits with such small dimensions—mean that the relentless three-year cycles for introducing a new generation of memory circuits and microprocessors may start to slow. Worldwide this trend could have a dramatic impact on a $100-billion semiconductor industry that is projected to double in revenues by the year 2000. The industry may have to find ways of achieving productivity gains beyond making tinier components. Otherwise, the dramatic decrease in costs for a unit of memory or logic in each successive chip generation could disappear.

From the standpoint of basic physics, the dominant type of chip, the metal oxide semiconductor, might continue to operate down to dimensions of 0.03 micron, less than a tenth the size of the most advanced circuitry being readied for use in the factory. Below that scale it may be difficult to turn off the tiny switches called transistors. They would act less like switches than leaky faucets: an electron may move uncontrollably from one side of a transistor to another.

But manufacturing difficulties could cause the technology to expire before then. When millions of transistors are linked to one another, wires must be stacked like a multitiered freeway—electrical resistance in these small-diameter wires and the distance a signal must travel slow operating speeds. Engineers must stand guard for 60-nanometer-size particles, so-called killer defects that can ruin the memory or brain of a "smart" device. Building chips this small requires very large factories: state-of-the-art manufacturing facilities are headed toward a price tag of $2 billion. Of all these hurdles, however, one of the most daunting is posed by the attempt to wring more out of lithography.

Limits of Light

Photolithography is a hand-me-down from a printing process invented in the 18th century by a German map inspector. It channels light through a mask, a plate of quartz covered with chromium lines that trace a circuit pattern. The light subsequently moves through one of the world's most sophisticated optical devices, a series of 30 or more lens elements that retail for over $1 million. These demagnifying lenses reduce the image to one quarter or one fifth its original size and project it onto a few square centimeters of a wafer, a silicon disk roughly 200 millimeters across. The light exposes a micron-thick photosensitive polymer—a photoresist—that is spread over the surface of the wafer. The table on which the wafer sits then "steps" to position another area below the beam. (For that reason, the lithography equipment is called a step-and-repeat tool or, simply, a stepper.)

In the next stage of processing, developing chemicals wash away either the light-exposed or the unexposed parts of the photoresist (depending on the needs of the chip maker). The circuit pattern drawn on the resist gets transferred to the surface of the wafer by chemical etching. This lithographic patterning delineates where impurities, such as boron or arsenic, should be implanted into a chip to alter the electrical conductivity of circuit elements, a process called doping. Lithography can also define the areas to place metal wires to connect circuit elements. The finished wafer is then cut up into individual chips that are packaged in a ceramic or plastic covering.

The gigabit-chip generation may finally force technologists up against the limits of optical lithography. To make these chips, the industry has targeted lithography that uses a pulsed (excimer) laser that emits light at a wavelength as small as 0.193 micron, in the deep ultraviolet segment of the electromagnetic spectrum. But for wavelengths below 0.2 micron, the photoresists absorb so much of the light that it takes more time to transfer a pattern to a chip. It may be impossible to process economically the many hundreds of wafers an hour produced in a large factory.

The few materials that have been identified for 0.193-micron lens manufacturing perform poorly. The fused silica glass for lenses tends to absorb the illumination and to heat up, which can degrade an image by changing the angle at which the lens refracts the light.

Lithographers confront the formidable task of building structures smaller than the wavelength of light. Wavelengths of 0.193 or even 0.248 micron have been targeted to make the 0.180-micron structures for gigabit chips. "Think of it as trying to paint a line that is smaller than the width of the paintbrush," says Steven D. Berger, a researcher at Bell Labs. "There are ways of doing it, but not very many ways of doing it in a controlled manner. The thing with lithography is not doing it once. It's doing it 10^{10} times in one second."

Keeping Focused

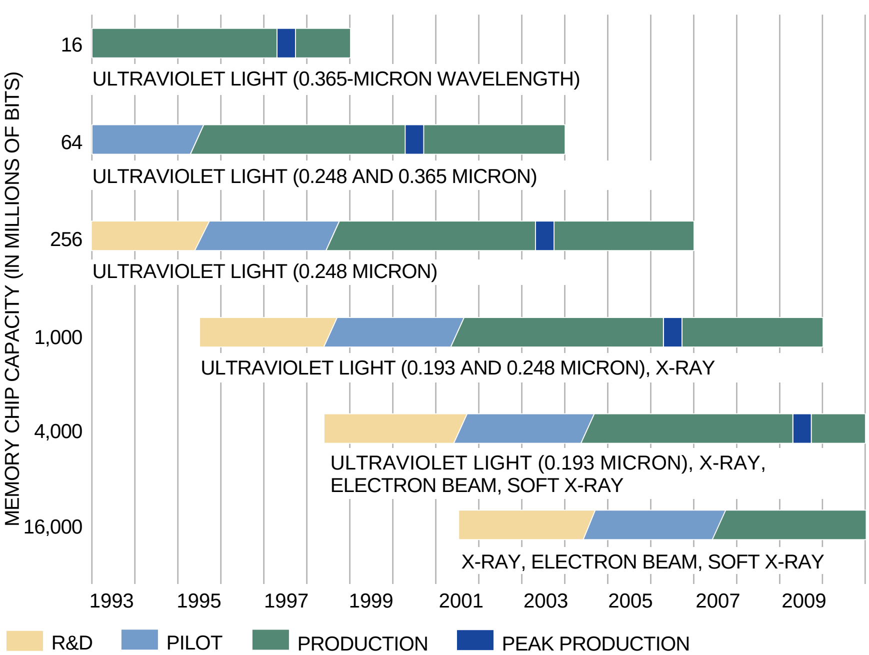
Other problems loom when working at these short wavelengths. Chip making follows a basic law of optics known by any photographer. A wide lens aperture increases resolution—the strands of a child's hair or tinier chip components stay clearly imaged. At the same time, depth of field decreases. For the amateur photographer, this trade-off means that the cabin in the background gets a little fuzzy. For the photolithographer, watching over lenses with giant apertures, the focus of the projected image starts to fade at distances well below a micron. As a result, the focus does not remain sharp down into the jagged Himalayan microtopography of a gigabit chip. The number of defective chips could skyrocket. "If the yield falls off, you wouldn't have a cost-effective system," says David C. Shaver, head of the solid-state division at the Massachusetts Institute of Technology Lincoln Laboratory.

Light specialists contemplate using what they call tricks, which allow printing of smaller features without reducing depth of field. One scheme employs masks that alter the phase of the light passing through them, which can improve line resolution by 50 to 100 percent. These phase-shift masks are expensive and difficult to make and cannot be used for all the geometric patterns printed on the resist. Intel, however, is contemplating using some of these techniques, beginning in 2001, to make a new microprocessor whose smallest dimensions measure 0.18 micron.

This depth-of-field problem has caused chemists to consider novel approaches for the photoresist. Surface imaging allows only 0.2 micron or so of the top layer of resist to be exposed, instead of the more typical one micron depth. After the resist polymer is exposed, it is put into a vacuum chamber and comes into contact with a gaseous compound that contains silicon. In one type of resist the nonexposed areas absorb the silicon, which acts as a barrier that protects the underlying layer from a gas of oxygen ions. The gas etches away the photoexposed sections on the chip. It also etches more deeply into the exposed areas of the resist than just the thin layer that was initially exposed. Besides improving resists, chip manufacturers also try to deal with the depth-of-field problem by polishing, or planarizing, the top layer on a chip with a chemical slurry—it is easier to focus on a flat surface.

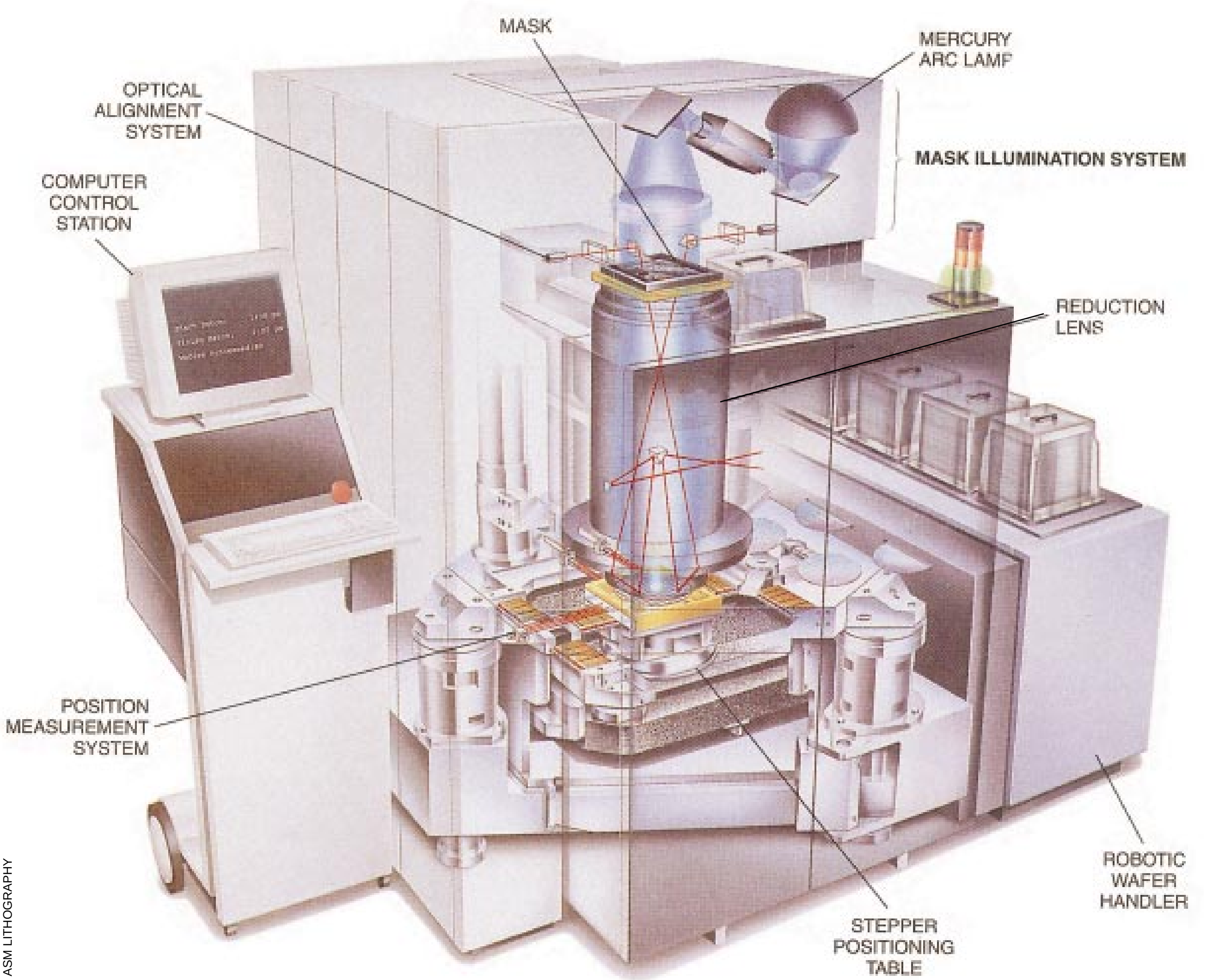
Researchers at the M.I.T. Lincoln Laboratory have demonstrated a prototype of a lithography system using light with a wavelength of 0.193 micron. But prototypes are not enough. "They make images; they print stuff," says Sematech's Brown. "We have to go past that to make commercial masks and resists you can buy." For that reason, Sematech has banded with industry, academia and government to explore the building of a manufacturing base to make chips with light from the outer reaches of the ultraviolet spectrum.

Research on 0.193-micron lithogra-



SEMICONDUCTOR INDUSTRY ASSOCIATION; STEVEN STANKIEWICZ

LITHOGRAPHY DEVELOPMENT in the U.S. calls for increasingly smaller wavelengths of light and other forms of electromagnetic energy to produce chips with ever larger memory capacity. The procession of chip generations will require moving to ever smaller wavelengths of ultraviolet light and then perhaps to x-rays or electron beams.



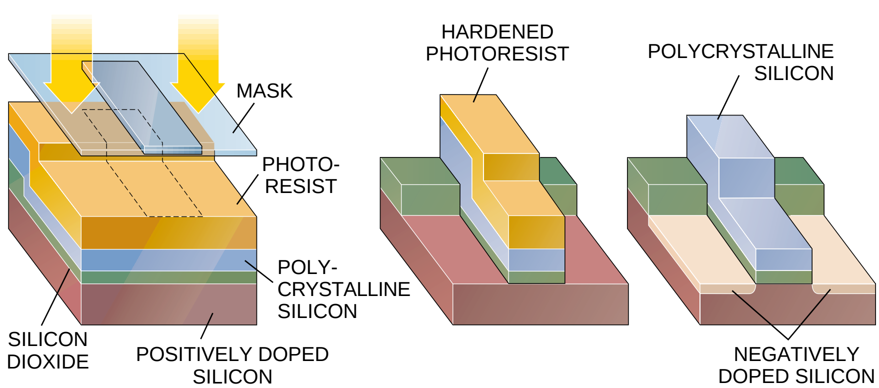
ASM LITHOGRAPHY

STEPPER, or photolithography machine, imprints circuit patterns on silicon wafers. Ultraviolet light from an arc lamp (or from a laser) passes through a mask bearing the image of the circuit. A sophisticated lens apparatus reduces the image and projects it onto part of a wafer. The table then moves, or "steps," to expose other chips.

phy lags by about three years the schedule for bringing gigabit chips to market in the early part of the next decade. Whether the process can produce millions of chips a year remains unknown. "There's a chance that 0.193 may prove to be impractical," says David A. Markle, vice president of advanced technology for Ultratech Stepper, a California equipment maker. "We need to introduce a new technology that ultimately supplants conventional optical technology, and it's not clear what it will be."

A group of researchers who have been closeted in laboratories for decades hope that lithography involving x-rays can succeed if conventional optics fail. X-ray lithography has been nurtured by the Department of Defense's desire for high-performance chips and by IBM, a company that once took it on itself to develop new semiconductor manufacturing technology. For its part, IBM has acknowledged that it can no longer tread this route alone. It joined with AT&T, Motorola and Loral Federal Systems (IBM's former government con-

PATTERNING A TRANSISTOR involves photolithography in which light is projected through the clear parts of a quartz mask (*left*). The photoresist, a polymer coating on the silicon wafer, reacts to the light; exposed areas are then removed with a solvent. A plasma of ions etches through the unprotected polycrystalline silicon conductive layer and the silicon dioxide insulating layer (*center*); the rest of the photoresist is removed. The silicon (*pink area*) is implanted with impurities, such as arsenic. Free electrons in this "negatively doped" area conduct current (*right*).



STEVEN STANKIEWICZ

tracting unit) in setting up a collaborative effort to move x-ray lithography beyond the laboratory.

In principle, the technology should be a natural choice for drawing finer circuit elements. At roughly one nanometer (a billionth of a meter), x-rays have a wavelength about one four hundredth that of the light used in today's most advanced commercial systems.

The technology for producing and harnessing x-rays is considerably different, however—and that is where the debate over the feasibility of x-rays begins. Whereas the radiation for optical lithography can be generated by lasers, the necessary x-rays may emanate from a synchrotron, an energy source usually deployed for physics experiments. IBM owns the only commercial synchrotron storage ring in the U.S. It consists of two superconducting magnets whose field confines electrons within a closed orbit. Electrons emit x-rays as they circulate within the storage ring. (Sortec, a Japanese government and industry consortium, is also working on x-ray lithography development.)

The price tag of $20 million to $50 million for such x-ray generators should not deter their purchase. Those amounts are 3 percent or less of the cost of the newest semiconductor plants. Moreover, a synchrotron can supply x-rays to 16 steppers simultaneously. For a sizable increase in plant capacity, a $20-million outlay is not unreasonable.

The problem is that if a company wants to make a *small* increase in plant capacity, that, too, costs $20 million. Through its collaboration with IBM, AT&T is exploring development of x-rays from a laser-generated plasma that supplies energy to a single stepper. Because there is no commercial supplier of the entire suite of equipment, AT&T is using a machine made by Hampshire Instruments—a company that went out of business in 1993.

Another technical obstacle springs from the lack of a commercially feasible way to focus x-rays. Given that x-ray steppers lack the equivalent of lenses (or equivalently demagnifying mirrors), the masks must bear more of the engineering burden: each circuit element outlined on the mask has to be the same small size as the structure to be created on the chip.

The inability to reduce the image also complicates the process of aligning one mask image atop another. Making a gigabit chip requires 20 or so lithographic steps, each with a separate mask. Up to eight of these steps require that the x-ray mask alignments be precise to within a few tens of nanometers—a difficult mechanical tolerance to meet.

Nevertheless, IBM has set about tackling the most imposing challenge to the technology's success. It has received funding from the Advanced Research Projects Agency (ARPA) to set up a facility to fabricate commercial x-ray masks. Materials that absorb x-rays are hard to find. Gold and tungsten will do the job. But these metals must be laid down in awkward dimensions atop a membrane of silicon through which the radiation is transmitted. A gold circuit feature may have to be 0.4 or 0.5 micron high to absorb the x-rays but need be, say, only 0.10 micron wide. "It looks like the New York City skyline," says Larry F. Thompson, vice president of technology development for Integrated Solutions, a Massachusetts stepper manufacturer.

Uncertainties about optical lithography may keep x-rays going. But time may be running out. AT&T, Motorola and Loral Federal Systems have lent a lingering legitimacy to a development effort carried for years by IBM, which had started to lose support from government funding agencies. Even the financially strapped IBM, after spending a few hundred million dollars on x-rays, looks back with some regret. "In hindsight, our work in x-ray lithography was done much too early," wrote Eberhard Spiller, a pioneer of the technology, in a 1993 issue of the *IBM Journal of Research and Development.*

Later this year the alliance of companies will assess whether it is worthwhile to take the costly next step of gearing up for production, which could require an investment of hundreds of millions of dollars.

Alternative Technology

Lithographers might consider x-ray technology more favorably if this form of radiation could be demagnified through a lens. A program to devise an x-ray lens is the goal of one of the most advanced research efforts in lithography. This "projection" x-ray system tries to apply technologies developed, in part, for the Strategic Defense Initiative to chip making.

X-ray lithography research is also intended to meet the policy goal of having the government's nuclear weapons and energy laboratories assume a post–cold war role that includes some of the basic research duties once fulfilled by Bell Labs and IBM's laboratories. The laser and measurement expertise of the national labs might be adapted to lithography.

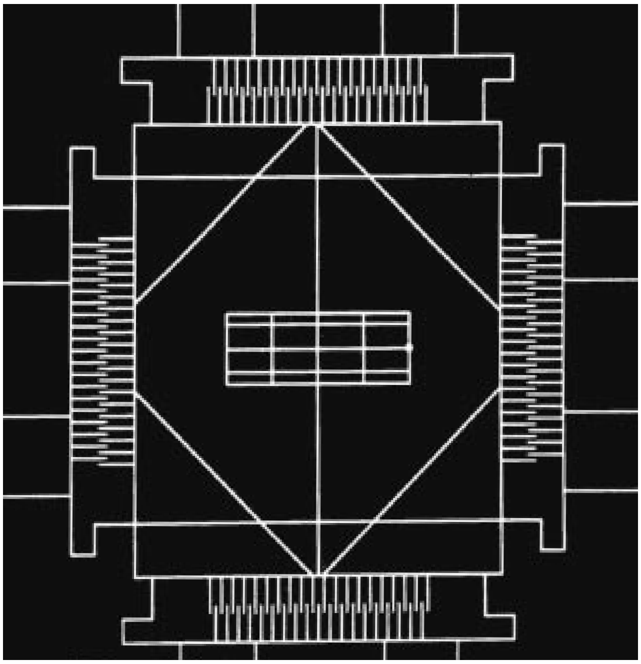
Three national laboratories—Sandia, Lawrence Berkeley and Lawrence Livermore—have spearheaded an effort to develop a projection x-ray system for exposing chip elements of 0.1 micron or less. They have recruited Intel and AT&T, among other companies, to provide support, expertise and testing facilities for this program. The consortium's approach is to train a high-powered laser onto a metal target to generate x-rays, which then illuminates a reflective mask. The resulting energy bounces among a series of mirrors that reduce the image to the size of the chip on the wafer. (Nikon and Hitachi are also researching this technology.)

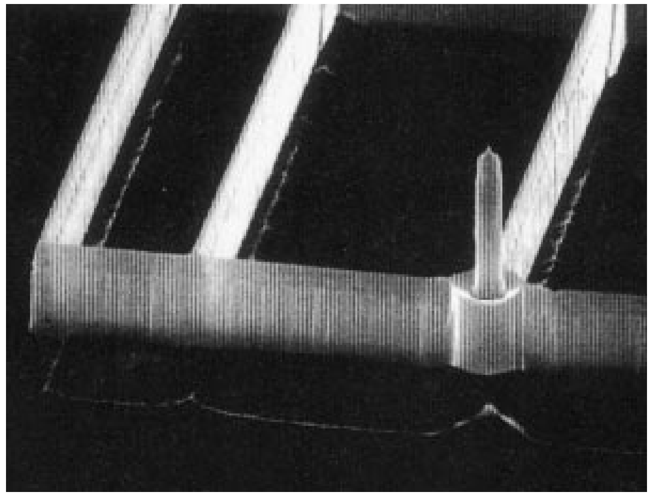
Making microprocessors commercially with this system may be as onerous as tracking and shooting down an

ALAN D. WILSON *IBM*

SYNCHROTRON, the only one of its kind in the U.S. designed for commercial manufacturing, is unloaded into IBM's East Fishkill, N.Y., facility on March 29, 1991.

NATIONAL NANOFABRICATION FACILITY, CORNELL UNIVERSITY





FUTURE LITHOGRAPHY is presaged by this 200-micron motor (*left*) from Cornell University that incorporates a 20-nanometer-diameter silicon tip (*closeup at right*). The machine's ability to move the tip up, down or sideways could make it a forerunner of lithographic tools with many tips for patterning a chip surface.

incoming ballistic missile. By alternating layers of silicon and molybdenum, researchers have successfully created materials that reflect x-rays instead of absorbing them. They are nonetheless left with the burdensome task of polishing and coating the mirrors to angstrom-level specifications. They must maintain this level of smoothness for each mirror over an area of many square centimeters.

It is a sign of the times that whereas researchers formerly referred to the radiation used in these experiments as "soft" x-rays, they now call it "extreme ultraviolet." The name change reflects the stigma that has come to be attached to x-ray lithography.

Surprisingly enough, the technology that can make the smallest chip structures is already used every day in manufacturing. Electron-beam lithography employs a focused "pencil" of charged particles to draw lines directly on a photoresist. Indeed, companies sketch circuit patterns onto photolithographic masks with this technique. For 20 years, engineers have dreamed of marshaling it for high-volume lithography. Unfortunately, electron beams are achingly slow: a pencil beam must draw each element of a circuit pattern individually, rather than exposing the entire chip surface in a flash of light. It can take hours to make single chips—too long for mass production, although some high-performance electronics that use nonsilicon materials are made this way.

Since the late 1980s Bell Labs has studied a method that scans a broad electron beam across a chip. As in photolithography, the radiation gets projected through a mask, and the image is reduced with a lens. AT&T considers this scanning-beam technique to be the most promising for long-term lithography. Still far ahead is a lithographic technique that could promote the embryonic science of nanotechnology. In theory, microscopic tools might fashion the tiniest silicon transistors, those whose smallest dimensions measure only a few tens of nanometers. They might also fashion wholly new types of electronics that store or process information by sensing the position of individual atoms.

Conventional optical lithography can make such tools. It sketches the outlines for hundreds or even thousands of cathodes on silicon. When a voltage is applied to the cathodes, they generate beams of electrons, which can draw circuit lines less than 0.05 micron wide. Noel C. MacDonald, a professor of electrical engineering at Cornell University, has built an array of 1,000 cathodes, providing the makings for an electron-beam machine on a chip.

MacDonald foresees employing the technology for making masks—and perhaps later for actually building small chips. MacDonald, with students Yang Xu and Scott A. Miller, has also demonstrated how a scanning tunneling microscope can be integrated with motors 200 microns wide, which are also fabricated with photolithographic methods. The sharpened tip of the scanning tunneling microscope has been used in research laboratories to push around atoms. The micromotors might let an array of tips—thousands or even a million—pattern a surface rapidly enough for commercial manufacture of circuit lines of but a few nanometers.

Arrays of cathodes or scanning microscopes are examples of the most advanced research projects in lithography anywhere. But they are still graduate-level research projects, not equipment that can be bought by leading chip manufacturers.

Perhaps only one or two of these technologies will make it to the factory floor. It simply costs too much to fund any more than that. Expenses for big-ticket lithography may require industry collaborations of competing suppliers, whether their corporate headquarters is in Tokyo or Silicon Valley.

Design ingenuity may overtake the drive to craft diminutive physical attributes. Chips can still be made bigger to hold more components. Horizontal layers of transistors could be stacked one atop the other on the same chip to increase memory or logic density. All the while the size of individual chip elements—a transistor or capacitor—may remain the same.

The art and science of lithography may be reaching maturity. Motorola or Intel might learn by consulting with retired Boeing executives. In constant dollars, air travel is practically free compared with its cost 35 years ago, yet the expense of building and flying supersonic transports means there are very few of these airplanes carrying passengers today. "We're flying at the same speed that the 707 flew in 1958," says AT&T lithography expert Richard Freeman. "I think the same kind of thing could conceivably happen here." "Point one" may be the chip industry's Mach 1.

 by Robert A. Hiller and Bradley P. Barber

Producing Light from a Bubble of Air

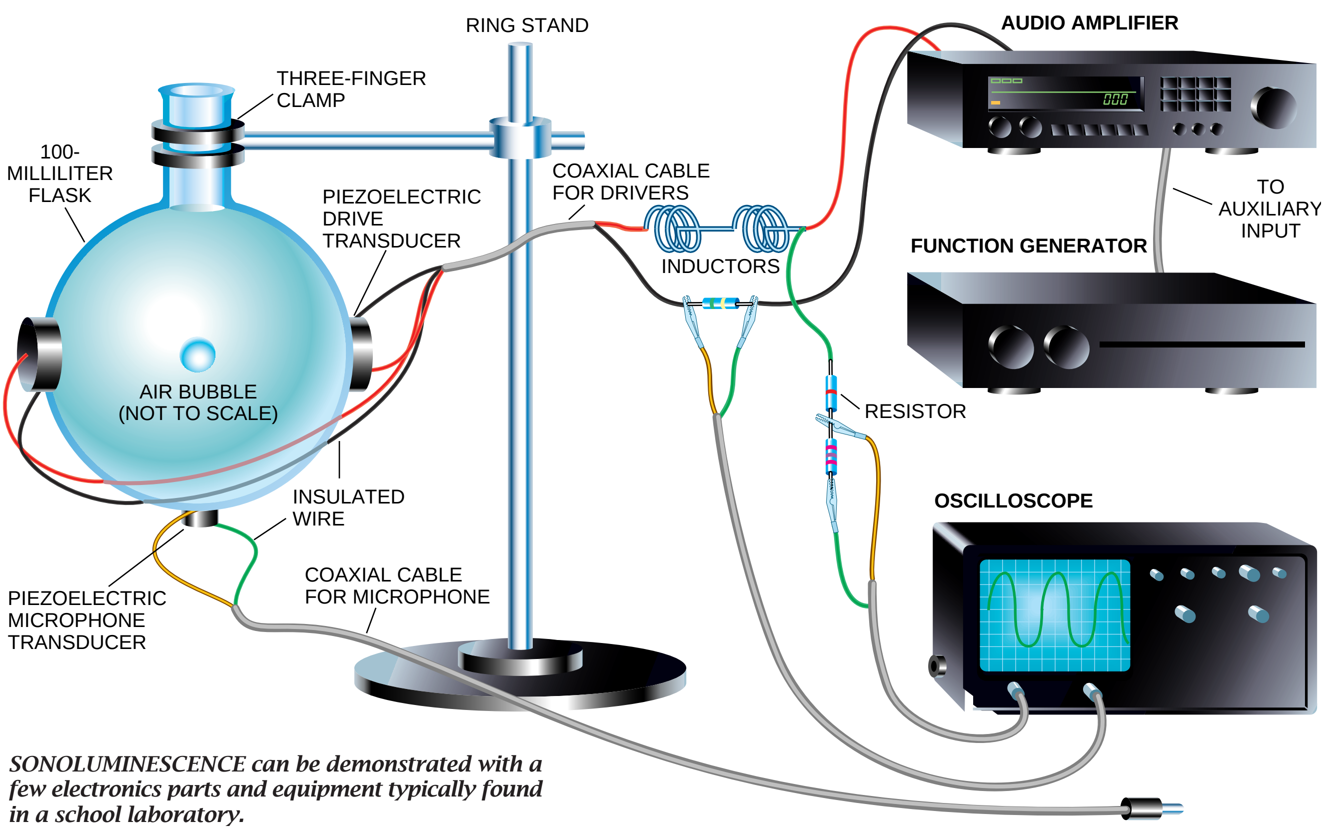
A glowing bubble of air cannot be bought anywhere at any price. But with an oscilloscope, a moderately precise sound generator, a home stereo amplifier and about $100, readers can turn sound into light through a process called sonoluminescence [see "Sonoluminescence: Sound into Light," by Seth J. Putterman; page 70]. The apparatus is relatively simple. A glass spherical flask filled with water serves as the resonator—the cavity in which sound is created to trap and drive the bubble. Small speakers, called piezoelectric transducers, are cemented to the flask and powered by an audio generator and amplifier. Bubbles introduced into the water coalesce at the center of the flask and produce a dim light visible to the unaided eye in a darkened room.

The filled flask must be vibrated at its resonant frequency—that is, at the sound frequency at which it responds most intensely. The resonant frequency equals the speed of sound in water (1,500 meters per second) divided by the diameter of the sphere. The glass will cause the actual resonance frequency to be about 10 percent higher. We used a 100-milliliter Pyrex spherical boiling flask with a diameter of 6.5 centimeters. Filled with water, the container resonated at about 25 kilohertz. A small-necked flask will produce the best results. Grease and oil can interfere with the bubble, so the glassware should be thoroughly washed with soap and water and rinsed well.

You will need three ceramic piezoelectric transducers: two to create the acoustic wave and one to act as a microphone to monitor the sound of the collapsing bubble. We used disks 15 millimeters in diameter and six millimeters thick for the driving transducers. The microphone was three millimeters in diameter and one millimeter thick. As a courtesy to readers of *Scientific American,* the three transducers are offered as a set for $95 from Channel Industries, Inc., 839 Ward Drive, Santa Barbara, Calif.; telephone (805) 967-0171; fax (805) 683-3420.

Connect fine wire (about 36 gauge) to the piezoelectric ceramics to serve as leads (thin wire minimizes sound loss). The wire is soldered to the silver electrodes on the ceramic. Remove the oxide layer on the transducers by rubbing them lightly with a pencil eraser. Working quickly with a cool soldering iron, place a small dot of solder on the silver sides of each piezoelectric transducer. Remove six millimeters of insulation from the end of the wire. Tin the copper lead (that is, melt some solder on it) and, after briefly heating the solder, place it on the solder dot. A wise move is to attach three leads to each disk, spaced equidistantly in the form of a triangle. This pattern ensures that each disk rests evenly on the curved surface of the flask. The other leads will also act as spares in case the first breaks.

Attach the transducers to the flask with epoxy. The quick-drying, five-minute type is recommended because it allows the transducers to be broken off



SONOLUMINESCENCE can be demonstrated with a few electronics parts and equipment typically found in a school laboratory.

the glassware without damage. Use just enough epoxy to fill the space between the flask and the transducer. For symmetry, place the two drive transducers on opposite sides on the equator of the flask and the microphone ceramic on the bottom. The transducers are polarized—one side will be identified with a plus sign or a dot. Make sure the two drivers are attached to the flask and wired in the same way: both should have plus signs toward the flask, or vice versa.

Solder a short lead to the outside of each transducer. Wire the drive transducers in parallel so they will expand and contract at the same time. Connect the wires to coaxial cables, which reduces electrical cross talk between the components. The microphone wires in particular should be short, extending no more than 10 millimeters before being connected to coaxial cables. Make the leads long enough so that they will not be under tension when connected. Suspend the flask either by clamping its neck to a laboratory stand or by hanging it with wires tied to the neck. Fasten all cables to the stand to prevent wire breakage.

The piezoelectric speakers act electrically as capacitors. To drive them with an audio amplifier (typically a low-voltage, low-impedance source), an inductor must be wired in series with them. The inductance is chosen so that it is in electrical resonance with the piezoelectric capacitance at about 25 kilohertz—that is, at the same frequency at which acoustic resonance occurs. The drivers described here will have a capacitance of about two nanofarads, so the inductance required is about 20 millihenrys. A good trick for adjusting the inductance is to use two or more inductors in series. By changing the distance between them, the total inductance may be raised or lowered by up to 50 percent. Two 10-millihenry inductors spaced about five centimeters apart would make a reasonable starting point.

To find the correct inductance, you will need to measure the voltage and current from the circuit [*see illustration above*]. Use a two-channel oscilloscope to display both quantities simultaneously. Get them in phase (their patterns on the oscilloscope should line up) by adjusting the inductance. Although the current from a home stereo system is low, the voltage may give a mild shock, so be sure all exposed connections and external wiring are insulated, covered with electrical tape or painted over with nail polish. The piezoelectric microphone transducer will typically produce about one volt; its output may be sent directly to the high-impedance input of the oscilloscope.

A sonoluminescent bubble can be created only in water in which the naturally dissolved air has been removed. A simple way to degas water is to boil it. Use a 500- to 1,000-milliliter Pyrex Erlenmeyer flask with an airtight stopper. Fit a hollow tube about six millimeters in diameter and about 10 centimeters long through the stopper and attach a short piece of rubber tubing to it. The tubing allows steam to vent and slows the diffusion of air back into the flask.

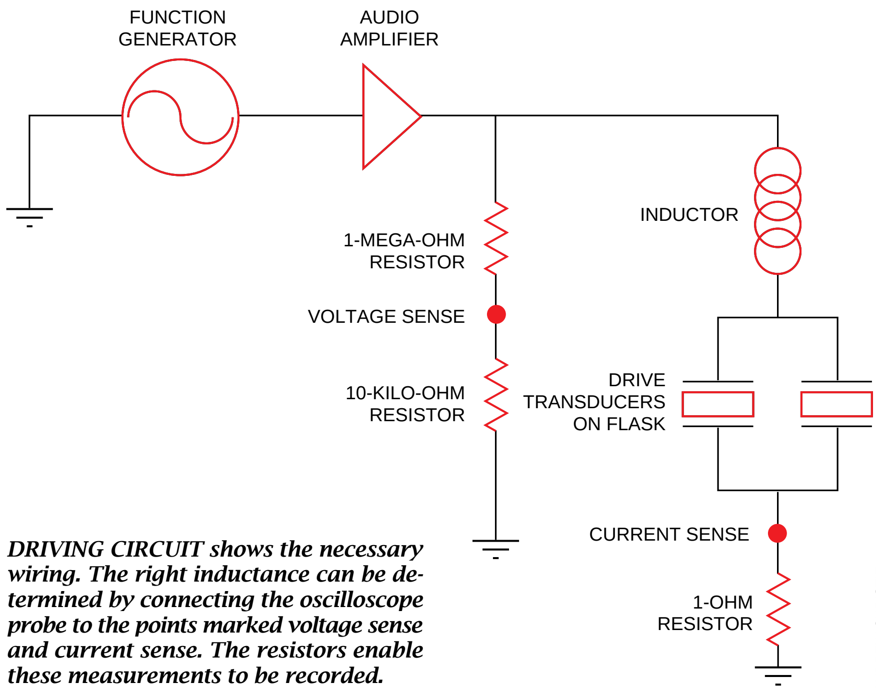
Fill the flask halfway with distilled water. Slowly heat the water and keep it at a rolling boil for 15 minutes. Then remove the flask from the heat, clamp the rubber tubing to prevent air from entering and allow the flask to cool (refrigeration will speed things up). After cooling, the flask will be under a strong vacuum, and the water will be well degassed. Keep sealed until ready to use, as the liquid will reabsorb air in a few hours when the container is opened.

Carefully pour the degassed water into the resonator flask, letting it run down the wall. Doing so will introduce a little air, but that actually brings the amount in the water to about one fifth the atmospheric concentration, which is the correct level for sonoluminescence. The water will slowly regas but will remain usable for several hours. Fill the flask with water up to the bottom of the neck so that the fluid level makes the mass of water approximately spherical.

For a 100-milliliter spherical boiling flask, tune the audio generator to the approximate resonance frequency of 25 kilohertz. Set the oscilloscope to display simultaneously the output voltage of the amplifier and the current through the drivers. Turn the volume control so that the amplifier output voltage reads about one volt peak to peak and adjust the inductance so that the current is in phase with the voltage. Monitor the current, because it can exceed the limit for the coil, causing it to overheat. Also, periodically check that the frequency is set to the resonance peak, as it may change with the water level and temperature.

Now, on the oscilloscope, display the piezoelectric microphone output. As you vary the generator frequency, you will notice a broad peak in the microphone signal, about one to two kilohertz wide. This peak corresponds to the electrical resonance between the inductor and the capacitance of the drivers. The acoustical resonance shows up as a much sharper peak in the microphone signal, less than 100 hertz wide, and as a slight dip in the drive current.

An easy way to find the resonance for the first time is to examine bubbles in the flask. For best viewing, position a bright lamp just behind the flask, because small bubbles scatter light much more efficiently in the forward direc-



***DRIVING CIRCUIT** shows the necessary wiring. The right inductance can be determined by connecting the oscilloscope probe to the points marked voltage sense and current sense. The resistors enable these measurements to be recorded.*

ANDREW CHRISTIE

ROBERT A. HILLER and BRADLEY P. BARBER study sonoluminescence in Seth J. Putterman's laboratory at the University of California, Los Angeles. Hiller is writing his thesis on the spectrum of sonoluminescence. Barber, a postdoctoral fellow, discovered that the light from bubbles emerges in regular picosecond intervals synchronized to the sound field.

tion. A dark background improves visibility. With an eyedropper, extract a small amount of water. While looking into the backlit flask, squirt the eyedropper at the surface hard enough to create about 10 to 30 bubbles. Adjust the generator to find the frequency at which the bubbles move toward the center and eventually coalesce into one. When everything is tuned correctly, it is usually possible to create a bubble just by poking a wire at the surface.

When you have a bubble in the center of the flask, slowly increase the amplitude. The bubble will be stable at first, and then it will "dance" over a few millimeters. Still greater amplitude will cause the bubble to stabilize again and shrink, becoming almost invisible, before growing again. Above a certain sound intensity, the bubble will disintegrate. Best light emission is obtained just below this upper amplitude threshold.

Small ripples should be visible on the oscilloscope trace from the microphone. This signal is high-frequency sound emitted by the bubble as it collapses with each cycle. Watching the ripples is an easier way to monitor the status of the bubble than is looking at the bubble itself. An electrical high-pass filter can be used to attenuate the driving sound, making the ripples on the oscilloscope more apparent. To view the glow emitted by a bubble, turn off the room lights and let your eyes adjust to the darkness. You should see a blue dot, somewhat like a star in the night sky, near the center of the flask. The bubble may be made brighter and more stable by fine-tuning the frequency and amplitude of the driving sound. If the glowing bubble is moving in the flask or varying in brightness over a few seconds, the water contains too much air. Try substituting freshly degassed water.

If the bubbles do not move at all or if they move toward the side of the flask, you are probably at the wrong frequency. Set the generator frequency to 24 or 26 kilohertz, readjust the inductance so the current and voltage are in phase and try again. Changing the water level or the way the flask is hung may improve the acoustics. As a last resort, carefully remove the transducers with a razor blade and try a different flask.

Having produced sonoluminescence, you can explore many questions about the phenomenon. For example, how do magnetic and electrical fields affect the light emission? How will various substances dissolved in the water change the behavior of the bubble? What new ways are there for probing the transduction of sound into light? With this home setup, you are ready to research at the frontiers of science.

BOOK REVIEW by Leon J. Kamin

Behind the Curve

THE BELL CURVE: INTELLIGENCE AND CLASS STRUCTURE IN AMERICAN LIFE, by Richard J. Herrnstein and Charles Murray. Free Press, 1994 ($30).

This book, with 400,000 copies in print just two months after its publication, has created an enormous stir. The authors unabashedly assert that scientific evidence demonstrates the existence of genetically based differences in intelligence among social classes and races. They maintain further that data from some 1,000 publications in the social and biological sciences show that attributes such as employment, income, welfare dependency, divorce and quality of parental behavior are determined by an individual's intelligence. These claims—another eruption of the crude biological determinism that permeates the history of IQ testing—lead Herrnstein and Murray to a number of social policy recommendations. The policies would not be necessary, or humane, even if the cited evidence were valid. But the caliber of the data in *The Bell Curve* is, at many critical points, pathetic. Further, the authors repeatedly fail to distinguish between correlation and causation and thus draw many inappropriate conclusions.

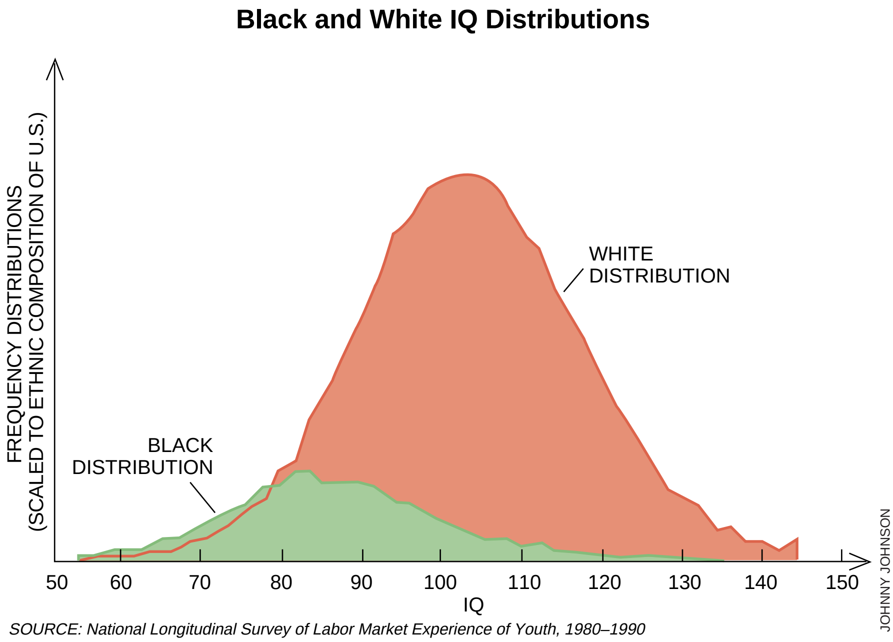
I will deal first with an especially troubling example of the quality of the data on which Herrnstein and Murray rely. They ask, "How do African-Americans compare with blacks in Africa on cognitive tests?" They reason that low African-American IQ scores might be the result either of a history of slavery and discrimination or of genetic factors. Herrnstein and Murray evidently assume that blacks reared in colonial Africa have not been subjected to discrimination. In their view, if low IQ scores of African-Americans are a product of discrimination, rather than genes, black Africans should have higher IQs than African-Americans.

To answer the question they have posed, Herrnstein and Murray call on the authority of Richard Lynn of the University of Ulster in Ireland, described as "a leading scholar of racial and ethnic differences," from whose advice they have "benefited especially." They state that Lynn, who in 1991 reviewed 11 Af-

LEON J. KAMIN is professor of psychology at Northeastern University in Boston. His more extensive critique of this work will appear in *The Bell Curve Debate,* edited by Russell Jacoby and Naomi Glauberman (Times Books/Random House, 1995).

ROBERTO CAPUTO *Aurora*

AFRICAN SCHOOLCHILDREN live and learn under very different circumstances than do children in the West. But the claims made by the authors of* The Bell Curve *assume that a single test can measure "intelligence" in all cultures.



BELL CURVE, adapted from an illustration in the book, shows the distribution of white and black IQ scores. What it does not show is* why *the scores are different.

rican IQ studies, "estimated the median black African IQ to be 75...about 10 points lower than the current figure for American blacks." Herrnstein and Murray conclude that the "special circumstances" of African-Americans cannot explain their low average IQ relative to whites. That leaves genetics free to explain the black-white difference.

But why do black Americans have higher scores than black Africans? Herrnstein and Murray, citing "Owen 1992," write that "the IQ of 'coloured' students in South Africa—of mixed racial background—has been found to be similar to that of American blacks." The implication is clear: the admixture of Caucasian and African genes, both in South Africa and in the U.S., boosts "coloured" IQ 10 points above that of native Africans. But the claims made regarding African and coloured IQs cannot withstand critical scrutiny.

Lynn's 1991 paper describes a 1989 publication by Ken Owen as "the best single study of the Negroid intelligence." The study compared white, Indian and black pupils on the Junior Aptitude Tests; no coloured pupils were included. The mean "Negroid" IQ in that study, according to Lynn, was 69. But Owen did not in fact assign IQs to any of the groups he tested; he merely reported test-score differences between groups, expressed in terms of standard deviation units. The IQ figure was concocted by Lynn out of those data. There is, as Owen made clear, no reason to suppose that low scores of blacks had much to do with genetics: "the knowledge of English of the majority of black testees was so poor that certain [of the] tests...proved to be virtually unusable." Further, the tests assumed that Zulu pupils were familiar with electrical appliances, microscopes and "Western type of ladies' accessories."

In 1992 Owen reported on a sample of coloured students that had been added to the groups he had tested earlier. The footnote in *The Bell Curve* seems to credit this report as proving that South African coloured students have an IQ "similar to that of American blacks," that is, about 85 (the actual reference does not appear in the book's bibliography). That statement does not correctly characterize Owen's work. The test used by Owen in 1992 was the "nonverbal" Raven's Progressive Matrices, which is thought to be less culturally biased than other IQ tests. He was able to compare the performance of coloured students with that of the whites, blacks and Indians in his 1989 study because the earlier set of pupils had taken the Progressive Matrices in addition to the Junior Aptitude Tests. The black pupils, recall, had poor knowledge of English, but Owen felt that the instructions for the Matrices "are so easy that they can be explained with gestures."

Owen's 1992 paper again does not assign IQs to the pupils. Rather he gives the mean number of correct responses on the Progressive Matrices (out of a possible 60) for each group: 45 for whites, 42 for Indians, 37 for coloureds and 28 for blacks. The test's developer, John Raven, repeatedly insisted that results on the Progressive Matrices tests cannot be converted into IQs. Matrices scores, unlike IQs, are not symmetrical around their mean (no "bell curve" here). There is thus no meaningful way to convert an average of raw Matrices scores into an IQ, and no comparison with American black IQs is possible.

The remaining studies cited by Lynn, and accepted as valid by Herrnstein and Murray, tell us little about African intelligence but do tell us something about Lynn's scholarship. One of the 11 entries in Lynn's table of the intelligence of "pure Negroids" indicates that 1,011 Zambians who were given the Progressive Matrices had a lamentably low average IQ of 75. The source for this quantitative claim is given as "Pons 1974; Crawford-Nutt 1976."

A. L. Pons did test 1,011 Zambian copper miners, whose average number of correct responses was 34. Pons reported on this work orally; his data were summarized in tabular form in a paper by D. H. Crawford-Nutt. Lynn took the Pons data from Crawford-Nutt's paper and converted the number of correct responses into a bogus average "IQ" of 75. Lynn chose to ignore the substance of Crawford-Nutt's paper, which reported that 228 black high school students in Soweto scored an average of 45 correct responses on the Matrices—*higher* than the mean of 44 achieved by the same-age white sample on whom the test's norms had been established and well above the mean of Owen's coloured pupils.

Seven of the 11 studies selected by Lynn for inclusion in his "Negroid" table reported only average Matrices scores, not IQs; the other studies used tests clearly dependent on cultural content. Lynn had earlier, in a 1978 paper, summarized six studies of African pupils, most using the Matrices. The arbitrary IQs concocted by Lynn for those studies ranged between 75 and 88, with a median of 84. Five of those six studies were omitted from Lynn's 1991 summary, by which time African IQ had, in his judgment, plummeted to 69.

Lynn's distortions and misrepresentations of the data constitute a truly venomous racism, combined with scandalous disregard for scientific objectivity. Lynn is widely known among academics to be an associate editor of the racist journal *Mankind Quarterly* and a major recipient of financial support from the nativist, eugenically oriented Pioneer Fund. It is a matter of shame and disgrace that two eminent social scientists, fully aware of the sensitivity of the issues they address, take Lynn as their scientific tutor and uncritically accept his surveys of research.

I turn now to a revealing example of Herrnstein and Murray's tendency to ignore the difference between mere statistical associations (correlations) and cause-and-effect relationships. The authors lament that "private complaints about the incompetent affirmative-action hiree are much more common than scholarly examination of the issue." They then proceed to a scholarly and public discussion of "teacher competency examinations." They report that such exams have had "generally beneficial effects," presumably by weeding out incompetent affirmative-action hirees. That view of tests for teachers is not shared by those who argue that because blacks tend to get lower scores, the tests are a way of eliminating competent black teachers. But Herrnstein and Murray assure us that "teachers who score higher on the tests have greater success with their students."

To support that statement, they cite a single study by two economists who analyzed data from a large number of North Carolina school districts. The researchers obtained average teacher test scores (a measure of "teacher quality") and pupil failure rates for each district. They reported that a "1% increase in teacher quality…is accompanied by a 5% decline in the…rate of failure of students"—that is, there were fewer student failures in districts where teachers had higher test scores. It does not follow from such a correlation, however, that hiring teachers with higher test scores will reduce the rate of student failure. The same researchers found, to their surprise, that "larger class size tends to lead to improved average [pupil] performance." Does it follow that increasing the pupil-to-teacher ratio would further improve student performance? That policy might please many taxpayers, just as firing teachers with lower test scores would please some. But neither policy derives logically from the observed correlations.

To understand why, consider the following. The average proportion of black students across the North Carolina school districts was 31 percent. Suppose—it does not stretch credulity—that black teachers (who have lower test scores) tend to work in districts with large proportions of black pupils (who have higher failure rates). Such nonrandom assignment of teachers would produce a correlation between teacher test scores and pupil failure rates, but one cannot then conclude that the teachers' test scores have any causal relation to student failure. To argue that, one would have to show that for a group of black teachers and for a separate group of white teachers, teachers' test scores predicted the failure rates of their students. No such information was available to the original researchers or to Herrnstein and Murray.

What about the finding that high pupil-to-teacher ratios are associated with good pupil performance? There is no way to be certain, but suppose deprived black children tended to be in small, de facto segregated rural schools, whereas more privileged whites were in larger classrooms. Would cramming more pupils into the rural schools promote academic excellence?

There is an important and general lesson buried in this example: the arithmetical complexity of the multitude of correlations and logistic regressions stuffed into *The Bell Curve* does not elevate their status from mere associations to causes and effects.

The confusion between correlation and causation permeates the book's largest section, which consists of an interminable series of analyses of data taken from the National Longitudinal Survey of Labor Market Experience of Youth (NLSY). Those data, not surprisingly, indicate that there is an association within each race between IQ and socioeconomic status. Herrnstein and Murray labor mightily to show that low IQ is the cause of low socioeconomic status, and not vice versa. The argument is decked out in all the trappings of science—a veritable barrage of charts, graphs, tables, appendices and appeals to statistical techniques that are unknown to many readers. But on close examination, this scientific emperor is wearing no clothes.

The NLSY survey included more than 12,000 youngsters, who were aged 14 to 22 when the continuing study began in 1979. At that time the respondents or their parents

NEW YORK MEDICAL JOURNAL

ARCHIVES OF THE HISTORY OF AMERICAN PSYCHOLOGY *University of Akron*

SUSPECTED MENTAL DEFECTIVES, including the men in this 1913 photograph, were examined at Ellis Island using tests such as the one at the right. Testees had to identify the four sad faces within 20 seconds—and to understand instructions possibly given in an unfamiliar tongue. Such questionable studies helped to encourage immigration restrictions in the U.S.

gave information about their educations, occupations and incomes and answered questions about themselves. Those reports are the basis for classifying the childhood socioeconomic status of the respondents. The teenagers also took the Armed Forces Qualification Test, regarded by psychometricians as essentially an IQ test. As they have grown older, the respondents have provided more information about their own schooling, unemployment, poverty, marital status, childbearing, welfare dependency, criminality, parenting behavior and so on.

Herrnstein and Murray pick over these data, trying to show that it is overwhelmingly IQ—not childhood or adult socioeconomic status—that determines worldly success and the moral praiseworthiness of one's social behaviors. But their dismissal of socioeconomic status rests ultimately on the self-reports of youngsters, which do not form an entirely firm basis.

I do not suggest that such self-reports are entirely unrelated to reality. We know from many sources that children from differing social class backgrounds do indeed differ in measured IQ. And in the NLSY study, after all, the respondents' self-reports are correlated with the objective facts of their IQ scores. But comparing the predictive value of those self-reports with that of test scores is playing with loaded dice.

Further, the fact that self-reports are correlated with IQ scores is, like all correlations, ambiguous. For Herrnstein and Murray, the relation of their index of parental socioeconomic status to the child's IQ means that parents of high status—the "cream floating on the surface of American society"—have transmitted high-quality genes to their offspring. But other interpretations are possible. Perhaps the kinds of people who get high test scores are precisely those who are vain enough to claim exaggerated social status for themselves. That tendency could artificially inflate correlations of IQ both with parental socioeconomic status and with self-reports of success, distorting all tests of the relative predictive power of socioeconomic status and IQ. Such an explanation may seem far-fetched to some readers, but it is clearly a logical possibility. The choice between such alternative interpretations of statistical associations cannot be based on logic alone. There is plenty of elbow room for ideological bias in social science.

The core of the Herrnstein-Murray message is phrased with a beguiling simplicity: "Putting it all together, success and failure in the American economy, and all that goes with it, are increasingly a matter of the genes that people inherit." Income is a "family trait" because IQ, "a major predictor of income, passes on sufficiently from one generation to the next to constrain economic mobility." Those at the bottom of the economic heap were unlucky when the genes were passed out, and they will remain there.

The correlations with which Herrnstein and Murray are obsessed are of course real: the children of day laborers are less likely than the children of stockbrokers to acquire fortunes or to go to college. They are more likely to be delinquent, to receive welfare, to have children outside of marriage, to be unemployed and to have low-birth-weight babies. The children of laborers have lower average IQs than do the children of brokers, and so IQ is also related to all these phenomena. Herrnstein and Murray's intent is to convince us that low IQ causes poverty and its attendant evils—not, as others hold, vice versa.

For eight dense chapters, the authors of *The Bell Curve* wrestle with data from the NLSY survey, attempting to disentangle the roles of IQ and of socioeconomic status. They employ a number of quantitative tools, most prominently logistic regression, a technique that purports to specify what would happen if one variable were "held constant" while another variable were left free to vary. When socioeconomic status is statistically held constant by Herrnstein and Murray, IQ remains related to all the phenomena described. When IQ is held constant, the effect of socioeconomic status is invariably reduced, usually substantially, and sometimes eliminated.

There are a number of criticisms to be made regarding the ways in which Herrnstein and Murray analyze these data. But for argument's sake, let us suppose that their analyses are appropriate and accurate. We can also grant that, rightly or wrongly, disproportionate salaries and wealth accrue to those with high IQ scores. What then do the Herrnstein-Murray analyses tell us?

The socioeconomic status of one's parents cannot in any immediate sense "cause" one's IQ to be high or low. Family income obviously cannot directly determine a child's performance on an IQ test. But income and the other components of an index of socioeconomic status can serve as rough indicators of the rearing environment to which a child has been exposed. With exceptions, a child of a well-to-do broker is more likely to be exposed to book learning earlier and more intensively than is a child of a laborer. And extensive practice at reading and calculating does affect, very directly, one's IQ score. That is one plausible way of interpreting the statistical link between parental socioeconomic status and a child's IQ.

The significant question is not whether socioeconomic status, as defined by Herrnstein and Murray, is more or less statistically associated with success than is their measure of IQ. Different measures of socioeconomic status, or different IQ tests, might substantially affect the results they obtained; other scholars, using other indices and tests, have in fact achieved quite different results. The significant question is, why don't the children of laborers acquire the skills that are tapped by IQ tests?

Herrnstein and Murray answer that the children of the poor, like their laborer parents before them, have been born with poor genes. Armed with that conviction, the authors hail as "a great American success story" that after "controlling for IQ," ethnic and racial discrepancies in education, wages and so forth are "strikingly diminished." They reach this happy conclusion on the questionable basis of their regression analyses. But the data, even if true, allow another reading. We can view it as a tragic failure of American society that so few black and low-socioeconomic status children are lucky enough to be reared in environments that nurture the skills needed to obtain high IQ scores. For Herrnstein and Murray, it is only fair that the race should go to the swift, who are blessed with good genes and high IQs. The conception that our society hobbles most of the contestants at the starting line does not occur to them.

On close examination, this scientific emperor is wearing no clothes.

In the world of *The Bell Curve,* the explanatory power of IQ is ubiquitous. The authors note that among blue-collar workers who tell researchers that they have dropped out of the labor force because of physical disability or injury, low IQ is common. Why? "An answer leaps to mind: The smarter you are, the less likely that you will have accidents." That answer leapt to mind before the thought that low-IQ workers, in minimum wage jobs, have little incentive to remain in the labor force. Dull young women lack the "foresight and intelligence" to understand that the welfare system offers them a bad deal. Welfare might be a bad deal for Herrnstein and Murray, but I am not so sure that single mothers on welfare have not figured out *their* odds pretty accurately.

People who have low IQs, according to *The Bell Curve,* commit crimes because they lack foresight, and so the threat of prison does not deter them. Further, they cannot "understand why robbing someone is wrong." Then what is to be made of the fact that although "very dull" young males are stopped by the police, booked for an offense and convicted less often than "normal" males, they are nevertheless jailed more than twice as often? "It may be...that they are less competent in getting favorable treatment from the criminal justice system. The data give us no way to tell." Perhaps not, but some hints are available. There is no doubt that O. J. Simpson is "competent," but his ability to hire high-priced lawyers is not irrelevant to the treatment he will receive from the criminal justice system.

The Bell Curve, near its closing tail, contains two chapters concerned with affirmative action, both in higher education and in the workplace. To read those chapters is to hear the second shoe drop. The rest of the book, I believe, was written merely as a prelude to its assault on affirmative action. The vigor of the attack is astonishing.

Affirmative action "cannot survive public scrutiny." It is based on "the explicit assumption that ethnic groups do not differ in...abilities." Hiring and promotion procedures "that are truly fair... will produce...racial disparities," and "employers are using dual standards for black and white job applicants because someone or something...is making them do so." That behavior has resulted in the "degradation of intellectual requirements" in recruiting police, which has affected "police performance on the street." We learn that a veteran of the Washington, D.C., police force has heard "about people in the academy who could not read or write." And a former instructor saw "people diagnosed as borderline retarded graduate from the police academy." These anecdotes take their place among the politically potent folktales about welfare queens driving Cadillacs.

At long last, Herrnstein and Murray let it all hang out: "Affirmative action, in education and the workplace alike, is leaking a poison into the American soul." Having examined the American condition at the close of the 20th century, these two philosopher-kings conclude, "It is time for America once again to try living with inequality, as life is lived...." This kind of sentiment, I imagine, is what led *New York Times* columnist Bob Herbert to the conclusion that *The Bell Curve* "is just a genteel way of calling somebody a nigger." Herbert is right. The book has nothing to do with science.

ESSAY by Anne Eisenberg

Scientists and Their CD-ROMs

The age of CD-ROMs is upon us. Scientists have their reservations about the silvery discs—"I know a few four-year-olds who like them; I myself prefer books," one math professor told me. But they are starting to buy these digital media anyway, the better to pursue their research, their businesses and even the occasional browse through *The Haldeman Diaries* or the NASA *Viking* photographs of the surface of Mars.

That CD-ROMs and technically minded people would eventually find one another was destiny, given the attractive combination of the practical and the pleasurable the discs offer. Still, the union had a slow beginning. Compact-disc technology was originally developed for audio recordings. By the mid-1980s it had been adapted as a dense storage medium for other types of data: a single disc holds hundreds of times the data of a floppy diskette. (Unlike floppies, however, these discs have read-only memory, meaning that once the data are recorded, they cannot be changed.) By the start of 1994 roughly 8,000 CD-ROM titles were available—very few of them appealing to scientists.

The merchandising of CD-ROMs using the term "multimedia" was part of the problem. Multimedia is a word scientists associate more with games such as Hell Cab and Virtual Tarot than, say, the *Journal of the American Chemical Society.* (Multimedia software is industry jargon for personal computer programs that include animation, sound, text, graphics and bits of video. The low quality of the initial offerings led cynics to dub them examples of "multimediocrity.") If the term "multimedia" were not suspect enough, it soon had a companion sales word, "hypermedia"—from hypertext, in which the user points at a word and clicks to learn more about the topic. To achieve the layered effect of hypermedia, the user points to link words, sounds and images.

Many multimedia CD-ROMs were designed for those who, in the words of one endorser, "wanted more than just print." That was a group scientists were not at all sure they cared to join. But by the fall of 1994 even the *Journal of the American Chemical Society* was available on CD-ROM, complete with typeset text and equations and full-color illustrations. It joined a myriad of highly respectable reference books and periodicals in which the multimedia turned out to be more than window dressing. For instance, the 1994 edition of Microsoft Bookshelf contained on one easily searchable CD-ROM the complete text of the *American Heritage Dictionary, Roget's Thesaurus,* the Columbia *Dictionary of Quotations*, the *Concise Columbia Encyclopedia* (15,000 entries and 1,300 images), the *Hammond Intermediate World Atlas* and the *World Almanac and Book of Facts 1994*—not to mention tricky pronunciations of foreign places in digital stereo.

By the end of 1994 many scientists unmoved by the beguiling ability of CD-ROMs to produce sound and images had come to appreciate them for their prodigious data storage. The discs come as a particular relief to researchers confronting overflowing shelves of journals (10 years of *JACS* on CD-ROM will take up only a few inches of shelf space) or to those whose floppies or slide collections were starting to run them out of the house. The new medium has also been embraced by users about to install a new computer operating system, a chore that is far handier from one CD-ROM than from dozens of floppies. For convenience alone, scientists joined the flocks of people upgrading their computers through "multimedia kits," packages of add-ons that include a drive, a 16-bit sound card and a set of speakers to perch atop the monitor.

After a long and bumpy trip, the CD-ROM at last became acceptable, especially if the speakers in the kit were properly austere (few scientists seem ready for Sound Blasters in the lab).

This year amateur and professional astronomers will be loading CD-ROMs onto their personal computers and, far from Mount Palomar, poring over the billion stars and several million galaxies of the entire Northern Hemisphere night sky. They will be looking at digitized plates from the National Geographic Society–Palomar Observatory *Sky Survey.* The quality of the images on the CD-ROMs is said to be virtually indistinguishable from that of the original data. Scientists studying galactic structure and galaxy clustering, whose research has previously been limited to relatively small regions of the sky, expect that the massive, unified data set will alter their approach both to astronomical problems as well as to the actual questions they ask.

CD-ROMs appeal not only to researchers but to many who seek efficient solutions to a host of technical problems. For instance, when U.S. nautical charts become available on inexpensive CD-ROMs this year, weekend sailors will go below deck, insert the discs into their color notebook computers and hook up their Global Positioning System to the computer to feed in latitude and longitude. The program will scroll so that the boat stays centered on the chart displayed on-screen. Banished will be clock, sextant, compass and all those cumbersome, slippery charts.

Many scientists who have started using CD-ROMs in their reading have finessed entirely the issue of disc versus books by using both disc and book—typically the CD-ROM for on-screen searching, then the book when they have settled into an easy chair. The double-barreled technique works well in a variety of situations: the *CD-ROM Mayo Clinic Family Health Book,* for instance, to find the name of a particular disease, then the *New England Journal of Medicine* for a more careful contemplation of what doctors are doing about it.

It appears that CD-ROMs have finally shed their reputation as a childish medium among scientists, a good number of whom may be found nowadays transfixed in front of their monitors, sampling the pleasures of the disc. They watch and listen to the CD-ROM score of *The Rite of Spring,* clicking occasionally for a technical explanation of a passage ("The bassoon enters here"). Or they brush up on their French using a CD-ROM tutorial—it has a perfect accent. And many of them decide that despite the shovelware (from "shoveling" any available title onto disc), the seedyware (from a play on the letters "CD" and pornography) and the mindless games, many CD-ROMs are actually turning out to be just as the French 1 disc promised—*très charmant.*

ANNE EISENBERG (aeisenb@photon.poly.edu) answers her E-mail from Polytechnic University in Brooklyn, N.Y., where she is a professor and writer.