JANUARY 1996
$4.95

SCIENTIFIC AMERICAN

The diet-aging connection.

Microchip progress: end in sight?

The ultimate physics theory.

***Nuclear theft** and smuggling could put weapons into terrorists' hands.*

SCIENTIFIC AMERICAN

January 1996 Volume 274 Number 1

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; QST No. Q1015332537. Subscription rates: one year $36 (outside U.S. and possessions add $11 per year for postage). 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. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.**

DEPARTMENTS

Letter from the Editor

Maybe life would just *seem* longer. That was my first reaction upon learning that we might be able to extend our life spans—and improve our general health—by putting ourselves on a tough diet. As Richard Weindruch explains in "Caloric Restriction and Aging" (page 46), a growing stack of evidence hints that cutting way back on our calories while still getting enough essential vitamins, minerals and other nutrients could add years to our lives. It works for rats. It works for guppies. Why not people?

Alas, this is not what we want to hear. Most of us have prayed that a lab-coated Ponce de León would discover a Soda Fountain of Youth to vindicate our guilty appetites. Chocolate, we would find, built strong bones. Crème brûlée improved eyesight and restored hair. A thick slab of barbecued ribs with extra sauce and a side order of french fries could cure whooping cough, erase wrinkles, lower blood pressure and make us better dancers. Instead we may be moving into an era when waiflike model Kate Moss will look unhealthy because she's a little too zaftig.

COVER art by Slim Films

Fortunately, there's hope. Weindruch notes that biomedical research may yet provide us with drugs or other interventions that can block the deleterious effects of an energy-rich diet. In the meantime, though, read up on the state of the research and mull the consequences before ordering your next ice cream cone.

This month's cover story—"The Real Threat of Nuclear Smuggling"—concerns a different threat, one that has perhaps been dismissed too readily by many policymakers and pundits. As Phil Williams and Paul N. Woessner argue, the possible rise of a thriving black market in radioactive materials could put at least a measure of the deadly force once restricted to the superpowers into the hands of unstable nations, gangsters and terrorists. Is there cause for alarm? Judge for yourself, starting on page 40.

On a brighter note, congratulations to Ian Stewart, author of our monthly "Mathematical Recreations" column. The Council of the Royal Society in London recently bestowed its Michael Faraday Award on Ian for his achievements in communicating mathematics to the general public. Few writers have ever done so with such charm or with such avidity—as in his books, including *Does God Play Dice?* and *The Collapse of Chaos,* on television and radio and, not least of all, in *Scientific American.*

JOHN RENNIE, *Editor in Chief*

SCIENTIFIC AMERICAN®

Established 1845

EDITOR IN CHIEF: John Rennie

BOARD OF EDITORS: Michelle Press, *Managing Editor;* Marguerite Holloway, *News Editor;* Ricki L. Rusting, *Associate Editor;* Timothy M. Beardsley; W. Wayt Gibbs; John Horgan, *Senior Writer;* Kristin Leutwyler; Madhusree Mukerjee; Sasha Nemecek; Corey S. Powell; David A. Schneider; Gary Stix; Paul Wallich; Philip M. Yam; Glenn Zorpette

ART: Edward Bell, *Art Director;* Jessie Nathans, *Senior Associate Art Director;* Jana Brenning, *Associate Art Director;* Johnny Johnson, *Assistant Art Director;* Carey S. Ballard, *Assistant Art Director;* Nisa Geller, *Photography Editor;* Lisa Burnett, *Production Editor*

COPY: Maria-Christina Keller, *Copy Chief;* Molly K. Frances; Daniel C. Schlenoff; Bridget Gerety

PRODUCTION: Richard Sasso, *Associate Publisher/Vice President, Production;* William Sherman, *Director, Production; Managers:* Carol Albert, *Print Production;* Janet Cermak, *Makeup & Quality Control;* Tanya DeSilva, *Prepress;* Silvia Di Placido, *Graphic Systems;* Carol Hansen, *Composition;* Madelyn Keyes, *Systems; Ad Traffic:* Carl Cherebin; Rolf Ebeling

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

ADVERTISING: Kate Dobson, *Associate Publisher/Advertising Director.* OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager;* Randy James, Thom Potratz, Elizabeth Ryan, Timothy Whiting. CHICAGO: 333 N. Michigan Ave., Suite 912, Chicago, IL 60601; Patrick Bachler, *Advertising Manager.* DETROIT: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, *Detroit Manager.* WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Lisa K. Carden, *Advertising Manager;* Tonia Wendt. 235 Montgomery St., Suite 724, San Francisco, CA 94104; Debra Silver. CANADA: Fenn Company, Inc. DALLAS: Griffith Group

MARKETING SERVICES: Laura Salant, *Marketing Director;* Diane Schube, *Promotion Manager;* Susan Spirakis, *Research Manager;* Nancy Mongelli, *Assistant Marketing Manager;* Ruth M. Mendum, *Communications Specialist*

INTERNATIONAL: EUROPE: Roy Edwards, *International Advertising Manager,* London; Vivienne Davidson, Linda Kaufman, Intermedia Ltd., Paris; Karin Ohff, Groupe Expansion, Frankfurt; Barth David Schwartz, *Director, Special Projects,* Amsterdam. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.; TAIPEI: Jennifer Wu, JR International Ltd.

ADMINISTRATION: John J. Moeling, Jr., *Publisher;* Marie M. Beaumonte, *General Manager;* Constance Holmes, *Manager, Advertising Accounting and Coordination*

SCIENTIFIC AMERICAN, INC.
415 Madison Avenue,
New York, NY 10017-1111

CHAIRMAN AND CHIEF EXECUTIVE OFFICER: John J. Hanley

CO-CHAIRMAN: Dr. Pierre Gerckens

CORPORATE OFFICERS: John J. Moeling, Jr., *President;* Robert L. Biewen, *Vice President;* Anthony C. Degutis, *Chief Financial Officer*

DIRECTOR, PROGRAM DEVELOPMENT: Linnéa C. Elliott

DIRECTOR, ELECTRONIC PUBLISHING: Martin Paul

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

Back to the Future

David A. Patterson predicted that the computer chips of tomorrow will merge memory and processors ["Microprocessors in 2020," SCIENTIFIC AMERICAN, September 1995]. That was actually done several years ago. The resulting product is embedded in plastic and called a smart card. They are currently used in France in banking and are the subject of the recently completed specifications jointly developed by Visa, MasterCard and Europay. The future often arrives much sooner than we think.

KENNETH R. AYER
Visa International
San Francisco, Calif.

Out of Gas?

In "Solar Energy" [SCIENTIFIC AMERICAN, September 1995], William Hoagland states that the solar energy reaching the earth yearly is 10 times the total energy stored on the earth, as well as 15,000 times the current annual consumption. This seems to mean that we have a 1,500-year supply of energy. But the usual estimate is that existing reserves will last 50 to 100 years.

RALPH M. POTTER
Pepper Pike, Ohio

Hoagland replies:

The 50-to-100-year figure is for fossil energy in the mix that is currently used (oil, coal, natural gas); it is also dependent on many estimates of the future demand for energy. It does not consider, for example, the broader use of coal and nuclear energy to meet these needs. The issue is really whether we want to incur the economic and environmental consequences of this route given the opportunities of solar energy.

Airport 2075

The engineers who dream of 800-passenger aircraft ["Evolution of the Commercial Airliner," by Eugene E. Covert; SCIENTIFIC AMERICAN, September 1995] must have had little experience in today's jets. The time spent checking in, loading passengers, stowing luggage, clearing for takeoff and then reversing the process at the destination may take as long as the actual flight. Logistical problems are roughly proportional to the square of the number of participants. Giant jets will spend most of their time on the ground with a mob of very unhappy passengers.

ROBERT GREENWOOD
Carmel, Calif.

L'Homme Machine

Simon Penny's essay, "The Pursuit of the Living Machine" [SCIENTIFIC AMERICAN, September 1995], reminded me of the mechanical chess player constructed in 1769 by the Hungarian inventor Wolfgang von Kempelen. This mysterious construction defeated many adversaries and could move its head and say, "Check!" in several languages. It was, of course, a technical joke, as there was a chess player hidden in the chess table, but the optical and mechanical construction was remarkable. Kempelen's lifelong aim was to construct a speaking machine for deaf-mutes and a writing machine for the blind.

J. DOBÓ
Budapest, Hungary

AIDS Concerns

In "How HIV Defeats the Immune System" [SCIENTIFIC AMERICAN, August 1995], Martin A. Nowak and Andrew J. McMichael observe that HIV infection "always, or almost always, destroys the immune system" and acknowledge that "chemical agents able to halt viral replication are probably most effective when delivered early." They fail to mention that such agents are available now and are being withheld. AZT, ddI, ddC, d4T and alpha-interferon slow viral replication but are commonly not prescribed until after the disease progresses.

KENNETH W. BLOTT
Toronto, Ontario

Nowak and McMichael reply:

Studies of early therapeutic intervention are yielding encouraging early results (see, for instance, papers in the *New England Journal of Medicine,* August 17, 1995). It is possible, however, that antiviral drugs will have serious toxic side effects or that they could select resistant viral strains that would preclude use of the drugs at a later stage. Therefore, it is essential to study these drugs first in controlled trials.

Tracking Tunneling

In "J. Robert Oppenheimer: Before the War" [SCIENTIFIC AMERICAN, July 1995], John S. Rigden credits Oppenheimer as "the first to recognize quantum-mechanical tunneling." Actually, the first treatment of tunneling was given by Friedrich Hund in a remarkable pair of papers submitted in November 1926 and May 1927 to *Zeitschrift für Physik.* This work came to our attention while preparing an article for a Festschrift celebrating the centennial of Hund's birth, which arrives in February.

BRETISLAV FRIEDRICH
DUDLEY HERSCHBACH
Harvard University

No Boys' Club

In "Magnificent Men (Mostly) and Their Flying Machines" ["Science and the Citizen," SCIENTIFIC AMERICAN, September 1995], David Schneider misrepresents my beliefs and those of the M.I.T.–Draper Aerial Robotics team. Schneider had remarked that there were no women on our team, and my response—"Well, we're M.I.T."—reflected the unfortunate demographics of our institute. For the record, our volunteer team was open to all. Both men and women contributed to the effort.

DAVID A. COHN
Massachusetts Institute of Technology

Letters may be edited for length and clarity. Because of the volume of mail, we cannot answer all correspondence.

ERRATUM

In "Down to Earth," by Tim Beardsley ["Science and the Citizen," SCIENTIFIC AMERICAN, August 1995], William H. Schlesinger of Duke University was mistakenly identified as W. Michael Schlesinger. We regret the error.

50, 100 AND 150 YEARS AGO

SCIENTIFIC AMERICAN

JANUARY 1946

The new multiplier phototube, called the image orthicon, picks up scenes by candle- and matchlight, and can even produce an image from a blacked-out room. The image orthicon tube has been a military secret until now, but as early as 1940, successful demonstrations of pilotless aircraft had been made with a torpedo plane which was radio-controlled and television-directed from 10 miles distant."

"Those gasoline fractions with low octane numbers have long been a problem to oil refiners. Researchers eventually determined that a mineral known as molybdenum oxide, when dispersed in activated alumina and used as a catalyst in an atmosphere of hydrogen, altered the molecular structure of the low-grade gasoline most effectively. The newly discovered process, 'Hydroforming,' doubled the octane rating of many low-grade gasolines, and guaranteed our war-time airplanes and those of our Allies vast quantities of high-octane gasoline, far superior to any in use by the enemy, and at reasonable cost."

JANUARY 1896

Many of our readers will have already been appraised of the death of Mr. Alfred Ely Beach, inventor, engineer and an editor of this journal. Our illustration shows one of his many inventions, the pneumatic system applied to an elevated railway. Visitors to the American Institute Fair, held in New York in 1867, will remember the pneumatic railroad suspended from the roof and running from Fourteenth to Fifteenth Streets."

"N. A. Langley has succeeded in obtaining helium perfectly free from nitrogen, argon, and hydrogen. This gas, when weighed, proves to be exactly twice as heavy as hydrogen, the usual standard. Guided by purely physical considerations, the experimenter arrived at the conclusion that the molecule of helium contains only one atom. Hence the atomic weight must be taken as 4."

"At a special meeting of the Anthropological Institute, held in London, Dr. Eugene Dubois, from Holland, read a paper describing his explorations in Java, and gave a demonstration of the interesting fossil remains discovered by him during six years' residence there. Most attention was attracted by the remains of a human-line femur, an anthropoid skull, and two molar teeth found in a Pliocene stratum on the banks of a river in Java. He holds that they form the strongest evidence yet adduced in favor of the doctrine of man's progressive development along with the apes from a common progenitor; for he asserts that these indicate a transitional form between man and an anthropoid ape, to which he has given the name Pithecanthropous erectus."

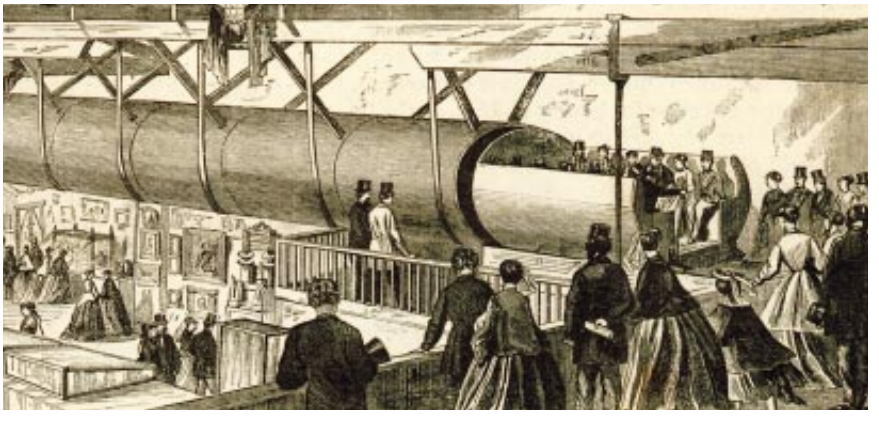
"Within a recent period cocaine has come into use on the race track, as a stimulant. Horses that are worn and exhausted are given ten to fifteen grains of cocaine by the needle under the skin at the time of starting, or a few moments before. The effects are very prominent, and a veritable muscular delirium follows, in which the horse displays unusual speed. The action of cocaine grows more transient as the use increases, and drivers may give a second dose secretly while in the saddle. Sometimes the horse becomes delirious and unmanageable, and leaves the track in a wild frenzy, often killing the driver, or he drops dead on the track from the cocaine, although the cause is unknown to any but the owner and driver."

JANUARY 1846

A new use of mesmerism has been recently put in requisition, at Oxford, Mass. A barn was destroyed by fire, last spring, and supposed to have been the work of an incendiary. A few weeks ago a professed mesmerizer was employed to put a subject to sleep, from whom such intelligence was elicited as to lead to the arrest of a person, who is now in prison awaiting trial. Should he be convicted, in consequence of the mesmeric relation, knaves may well dread the approach of mesmerism henceforth; and if this practice is successful, there will be no such thing as concealment of a crime, nor escape from detection."

"There are 90,000 slaves and 61,000 free blacks in Maryland. A member of the Maryland legislature lately proposed to seize and sell all the free blacks in the State, and apply the proceeds to the payment of the State debt. The bill would not pass."

"It is well known that a convex lens made of ice will converge the rays of the sun and produce heat. It may therefore be inferred that if a large cake of ice—say, twelve feet in diameter—be reduced to the convex form (which might readily be done by a carpenter's adz) and placed as a roof over a cabin, it would effectually warm the interior. And were the sun's rays admitted to pass through a trap-door into the cellar, and that of sufficient depth to bring the rays nearly to a focus, a sufficient heat would be produced to bake or roast provisions for a family."



Mr. Beach's pneumatic railway exhibit

Listening to Culture

Psychiatry takes a leaf from anthropology

Last April, a Bangladeshi woman who complained that she was possessed by a ghost arrived at the department of psychiatry at University College London. The woman, who had come to England through an arranged marriage, had at times begun to speak in a man's voice and to threaten and even attack her husband. The family's attempt to exorcise the spirit by means of a local Muslim imam had no effect.

Through interviews, Sushrut S. Jadhav, a psychiatrist and lecturer at the university, learned that the woman felt constrained by her husband's demands that she retain the traditional role of housebound wife; he even resented her requests to visit her sister, a longtime London resident. The woman's discontent took the form of a ghost, Jadhav speculated, an aggressive man who represented the opposite of the submissive spouse expected by her husband. By bringing the husband into the therapy, Jadhav made a series of subtle suggestions that succeeded in getting him to relent on his strictness. The specter's appearances have now begun to subside.

Jadhav specializes in cultural psychiatry, an approach to clinical practice that takes into account how ethnicity, religion, socioeconomic status, gender and other factors can influence manifestations of mental illness. Cultural psychiatry grows out of a body of theoretical work from the 1970s that crosses anthropology with psychiatry.

MIGUEL LUIS FAIRBANKS

***PARACHUTE GAME** is played by patients at a psychiatric unit at San Francisco General Hospital, which takes into account cultural background during the course of treatment.*

At that time, a number of practitioners from both disciplines launched an attack on the still prevailing notion that mental illnesses are universal phenomena stemming from identical underlying biological mechanisms, even though disease symptoms may vary from culture to culture. Practitioners of cultural psychiatry noted that although some diseases, such as schizophrenia, do appear in all cultures, a number of others do not. Moreover, the variants of an illness—and the courses they take—in different cultural settings may diverge so dramatically that a physician may as well be treating separate diseases.

Both theoretical and empirical work has translated into changes in clinical practice. An understanding of the impact of culture can be seen in Jadhav's approach to therapy. Possession and trance states are viewed in non-Western societies as part of the normal range of experience, a form of self-expression that the patient exhibits during tumultuous life events. So Jadhav did not rush to prescribe antipsychotic or antidepressive medications, with their often deadening side effects; neither did he oppose the intervention of a folk healer.

At the same time, he did not hew dogmatically to an approach that emphasized the couple's native culture. His suggestions to the husband, akin to those that might be made during any psychotherapy session, came in recognition of the woman's distinctly untraditional need for self-assertion in her newly adopted country.

The multicultural approach to psychiatry has spread beyond teaching hospitals in major urban centers such as London, New York City and Los Angeles. In 1994 the fourth edition of the American Psychiatric Association's handbook, the *Diagnostic and Statistical Manual of Mental Disorders,* referred to as the *DSM-IV,* emphasized the importance of cultural issues, which are mentioned in various sections throughout the manual. The manual contains a list of culture-specific syndromes, as well as suggestions for assessing a patient's background and illness within a cultural framework.

For many scholars and practitioners, however, the *DSM-IV* constitutes only a limited first step. Beginning in 1991, the National Institute of Mental Health sponsored a panel of prominent cultural psychiatrists, psychologists and anthropologists that brought together a series of sweeping recommendations for the manual that could have made culture a prominent feature of psychiatric practice. Many of the suggestions of the

Culture and Diagnosis Group, headed by Juan E. Mezzich of Mount Sinai School of Medicine of the City University of New York, were discarded. Moreover, the *DSM-IV*'s list of culture-related syndromes and its patient-evaluation guidelines were relegated to an appendix toward the back of the tome.

"It shows the ambivalence of the American Psychiatric Association [APA] in dumping it in the ninth appendix," says Arthur Kleinman, a psychiatrist and anthropologist who has been a pioneer in the field. The APA's approach of isolating these diagnostic categories "lends them an old-fashioned butterfly-collecting exoticism." A Western bias, Kleinman continues, could also be witnessed in the APA's decision to reject the recommendation of the NIMH committee that chronic fatigue syndrome and the eating disorder called anorexia nervosa, which are largely confined to the U.S. and Europe, be listed in the glossary of culture-specific syndromes. They would have joined maladies such as the Latin American *ataques de nervios,* which sometimes resemble hysteria, and the Japanese *tajin kyofusho,* akin to a social phobia, on the list of culture-related illnesses in the *DSM-IV.*

Eventually, all these syndromes may move from the back of the book as a result of a body of research that has begun to produce precise intercultural descriptions of mental distress. As an example, anthropologist Spero M. Manson and a number of his colleagues at the University of Colorado Health Sciences Center undertook a study of how Hopis perceive depression, one of the most frequently diagnosed psychiatric problems among Native American populations. The team translated and modified the terminology of a standard psychiatric interview to reflect the perspective of Hopi culture.

The investigation revealed five illness categories: *wa wan tu tu ya/wu ni wu* (worry sickness), *ka ha la yi* (unhappiness), *uu nung mo kiw ta* (heartbroken), *ho nak tu tu ya* (drunkenlike craziness with or without alcohol) and *qo vis ti* (disappointment and pouting). A comparison with categories in an earlier *DSM* showed that none of these classifications strictly conformed to the diagnostic criteria of Western depressive disorder, although the Hopi descriptions did overlap with psychiatric ones. From this investigation, Manson and his co-workers developed an interview technique that enables the differences between Hopi categories and the *DSM* to be made in clinical practice. Understanding these distinctions can dramatically alter an approach to treatment. "The goal is to provide a method for people to do research and clinical work without becoming fully trained anthropologists," comments Mitchell G. Weiss of the Swiss Tropical Institute, who developed a technique for ethnographic analysis of illness.

The importance of culture and ethnicity may even extend to something as basic as prescribing psychoactive drugs. Keh-Ming Lin of the Harbor–U.C.L.A. Medical Center has established the Research Center on the Psychobiology of Ethnicity to study the effects of medication on different ethnic groups. One widely discussed finding: whites appear to need higher doses of antipsychotic drugs than Asians do.

The prognosis for cross-cultural psychiatry is clouded by medical economics. The practice has taken hold at places such as San Francisco General Hospital, an affiliate of the University of California at San Francisco, where teams with training in language and culture focus on the needs of Asians and Latinos, among others (*photograph*). Increasingly common, though, is the assembly-line-like approach to care that prevails at some managed-care institutions.

"If a health care practitioner has 11 minutes to ask the patient about a new problem, conduct a physical examination, review lab tests and write prescriptions," Kleinman says, "how much time is left for the kinds of cross-cultural things we're talking about?" In an age when listening to Prozac has become more important than listening to patients, cultural psychiatry may be an endangered discipline. —*Gary Stix*

FIELD NOTES

Changing Their Image

On a cool October evening, troops of female journalists congregated at the august New York Academy of Sciences in Manhattan to appraise a group of blushing male scientists. The courageous men had modeled for the first-ever "Studmuffins of Science" calendar. "I want to change the image of science," explained "Dr. September," Bob Valentini of Brown University, with the wide-eyed earnestness of a Miss Universe desiring to eradicate world hunger. Karen Hopkin, who co-produces "Science Friday" for National Public Radio and is the calendar's creator, offered a more believable rationale for the enterprise: "It was an elaborate scheme for me to meet guys."

JAMES ARONOVSKY

Dr. March, ecologist Rob Kremer

To the disappointment of many in the audience, the studs turned out in modest suits and ties. Even the calendar featured only Dr. January, Brian Scottoline of Stanford University, in bathing trunks. "We wanted them to be wholesome, PG-13," said Nicolas Simon, the calendar's designer. "So we can sell to schoolgirls. It's educational." Dr. October, John Lovell of Anadrill Schlumberger, presented an alternative view of the creative process. He had offered to take off his shirt in the service of science, he declared, but "the photographer took one look at my chest and told me to put it back on." Still, three editorial assistants from *Working Mother* were suitably impressed. "All our readers will fall over their faces for these guys," one testified.

The truth is, surveys show that male scientists are not the ones who have trouble attracting mates, especially the kind who willingly follow wherever the scientific career leads. "I wish I had a wife" is the oft-heard sigh of female researchers who are not similarly blessed with portable (or culinarily capable) spouses. Some American women who are scientists even speak of how the decision to study mathematics and science, made in high school, was traumatic because it made them instantly unattractive to boys.

In addition to "Studmuffins," Hopkin's plans for 1997 include "Nobel Studs" (which one wag has redubbed "Octogenarian Pinups"). That should be as much of a hit. But her third venture, "Women in Science," may be the only one with a hope of offering a truly different image of scientists to schoolgirls and schoolboys. —*Madhusree Mukerjee*

Star Dreck

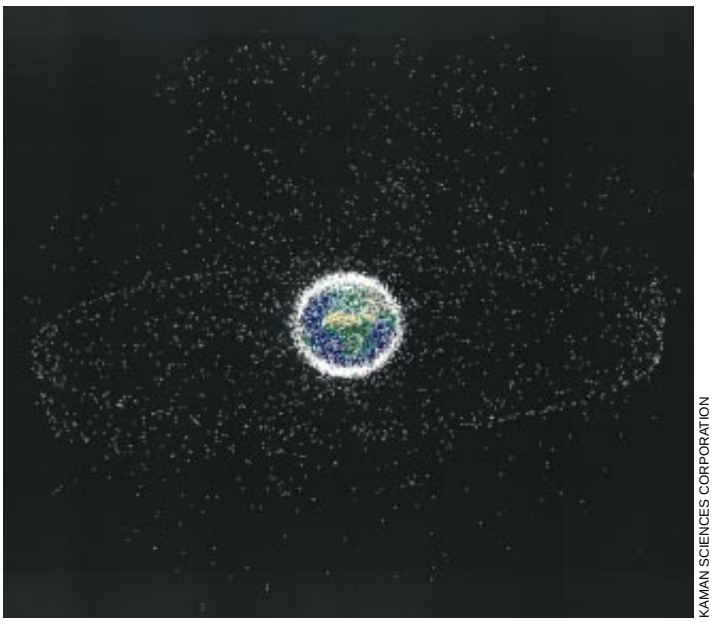
Conjuring images of "meteor storms" in bad science-fiction movies, the map below includes 7,800 of the larger man-made objects—including dead satellites—that are circling the earth. But contrary to appearances, "the sky is not falling just yet," says Nicholas L. Johnson of Kaman Sciences Corporation, which created the image. For clarity, the dots representing bits of debris are enormously exaggerated in size—which can give a false impression of the magnitude of the problem. Not a single functional satellite has been lost owing to space junk.

Nevertheless, the danger is real. Collisions in earth orbit occur at velocities of up to 15 kilometers per second, so a discarded bolt or lens cap could destroy a satellite or endanger astronauts. Objects as small as one centimeter across—hundreds of thousands of which lie in near-earth orbit—could knock out critical components on a spacecraft. And such tiny items cannot be tracked by current technology, so they strike without warning.

The most pressing concern, obviously, is loss of life, and here nature works in our favor. The density of debris at altitudes below 400 kilometers, where most manned space activities take place, is comparatively low because aerodynamic drag from the upper reaches of the atmosphere quickly causes little objects to spiral downward and burn up. Hence, for the space shuttles, orbiting junk "is not as serious a problem," Johnson comments.

Because of its large size and long intended life, the upcoming international space station faces a greater threat. But Johnson questions a claim, published in the *New York Times,* that because of space junk the station faces a 1-in-10 chance of incurring a "death or destruction of the craft" over its expected 10-year projected lifetime. "That's a misleading statement," he remarks dryly. Shielding will protect parts of the orbiting outpost, which is also designed to dodge oncoming objects. Still, undetectable, small items do pose a definite, if slight, hazard.

The greatest density of debris actually resides much higher, some 900 to 1,500 kilometers above the earth's surface. From a practical standpoint, however, the garbage problem may be most problematic in geosynchronous orbit, 35,785 kilometers up, where satellites' orbital periods match the 24-hour rotation of the earth. Real estate is tight at those heights, and orbits there may remain stable for millions of years, so inactive satellites and detritus are unwelcome.



KAMAN SCIENCES CORPORATION

***ORBITAL DEBRIS MAP** shows objects as tiny as 10 centimeters across. Space junk poses a small but growing risk.*

Cleaning up existing space pollution is no easy task, concludes a new report by a National Research Council panel (which included Johnson). But some simpler measures are under way: space-faring nations are reducing debris emanating from exploding rockets, and government and private users are moving old satellites out of geosynchronous orbit. Ultimately, all spacecrafts may be designed to crash back to the earth or to move to uncrowded orbital zones after they end their useful lives. For now, however, Johnson and his fellow NRC panelists are spreading the word that even if the current risk is small, space environmentalism makes sense. Johnson likens the situation to pollution of the oceans: for a long time the effects are invisible, but when they finally turn up, they are exceedingly difficult to reverse. *—Corey S. Powell*

Strange Places

An astronomical breakthrough reveals an odd new world

The universe became a slightly less lonely place last October 6, when Michel Mayor and Didier Queloz of the Geneva Observatory announced the detection of a planet around 51 Pegasi, a nearby star similar to our sun. The landmark discovery bolsters the belief that planetary systems—some of which may include habitable worlds—are a common result of the way that ordinary stars are born.

Mayor and Queloz inferred the presence of the planet by monitoring the light from 51 Pegasi, which is faintly visible to the naked eye in the constellation Pegasus. The two astronomers noted a slight, repeating shift in the star's spectrum, indicative of a back-and-forth motion having a period of 4.2 days. After 18 months of painstaking observations, Mayor and Queloz concluded that the star is being swung about by the gravitational pull of a small, unseen object—a planet. They reported that finding in Florence, Italy, at an otherwise quiet workshop on sunlike stars.

Astronomers initially greeted the announcement with skepticism, in part because the inferred planet around 51 Pegasi is so bizarrely unlike anything in our solar system. On the one hand, the planet is hefty, at least one half the mass of Jupiter. On the other hand, it orbits just seven million kilometers from 51 Pegasi, about one seventh the distance between the sun and Mercury; its surface must therefore be baking at a temperature of about 1,000 degrees Celsius. Theorists have believed that giant planets can form only in remote re-

gions where ice and chilled gases can gather in great abundance.

"My first reaction when I heard the report was 'Give me a break!'" laughs Geoffrey Marcy of San Francisco State University and the University of California at Berkeley. When Marcy and his co-worker Paul Butler checked 51 Pegasi themselves, however, they, too, detected a 4.2-day wobble; other observers have also confirmed the finding. "Our attitude has totally changed," Marcy says.

If the discovery is real, it confounds nearly all preconceptions of what a planetary system should look like. Computer simulations have suggested that other solar systems should broadly resemble ours, having small bodies close to the central star and Jupiter-like gas giants in the cold outlying areas. But the 51 Pegasi planet seems not to follow that pattern at all. "Nobody expected it," remarks Robert Stefanik of Harvard University. "It will change our views about how planets form."

Now the race is on to find additional planets. Marcy relates tentative evidence of a second body around 51 Pegasi, in a much more distant orbit; the exciting implication is that the star may possess a full system of planets. Mayor and Queloz are rumored to have similar evidence. Both teams of observers are tearing through their data to come up with decisive proof. "We've given up on sleep," says Butler, pleasantly weary. He and Marcy expect to have something fairly firm to report in the next few months.

All told, about half a dozen groups are performing similar high-resolution planetary searches, and the fierce competition is sure to yield more discoveries soon. Marcy's group alone has about eight years' worth of observations waiting for computer analysis, and "there are almost certainly planets in there," Butler claims. Indeed, the lack of previous results is itself significant. It suggests that "only a few percent of stars have Jupiter-like companions," Stefanik says, "but that does not mean there aren't Earth-like planets."

Finding Earth-size worlds lies beyond today's technology, although a sophisticated technique known as optical interferometry might bring them into view in the coming years. The current search techniques, in contrast, require patience more than they do money. "We can find other Jupiters for half a million dollars a year," Butler says. He has no doubt that the effort is worth the modest cost, especially given its philosophical implications. "As they say, it's been a million years since people looked up and wondered. Well, as of two weeks ago, we know." —*Corey S. Powell*

Virtual Pollution

Computers modeling the environment yield surprising results

Checking water quality was once a simple matter of sample jars and chemical tests. But these days many researchers no longer pull out the litmus paper—instead they just turn on their computers. Simulations of air, soil and water contamination are increasingly being hailed as cheap and efficient ways of studying the environment. And as recent findings regarding the Chesapeake Bay indicate, computers can demonstrate complex interactions that simply cannot be determined using other methods.

Computer modeling has revealed that approximately 25 percent of the nitrogen in the Chesapeake comes from air pollution wafting in from as far away as western Pennsylvania, Ohio and Kentucky. This finding alters the current perception that the bay's greatest problems stem from more local waterborne pollution, such as sewage and runoff from agriculture—which conservation efforts now seek to lessen.

To arrive at this conclusion, Robin Dennis of the Environmental Protection Agency and his colleagues digitally recreated the atmosphere above the eastern U.S. and combined this information with another model that examined how water flows into the Chesapeake. In particular, the group simulated how air moves across the country and how nitrogen pollution reacts with other airborne compounds and then drops to the ground directly or in rain.

Conventional wisdom has generally held that nitrogen pollution falls out fairly quickly. Thus, simple models had suggested that air pollution from local sources probably contributed to the bay's condition. But the more extensive model revealed that such pollution presents a much larger problem: 25 percent of nitrogen pollution is still being carried aloft 500 miles from its source.

Although water testing helps to monitor the state of the bay, models demonstrate how the pollution gets there. According to Dennis, "it can be difficult to disentangle measurements" to determine exactly where the pollution comes from—and which sources should be targeted. Despite several years of regulations on waterborne pollution, nitrogen levels have not decreased as much as expected. Dennis asserts that although controls on water pollution must not be abandoned, attempts to lower nitrogen levels in the bay may not be fully successful unless air pollution is also reduced.

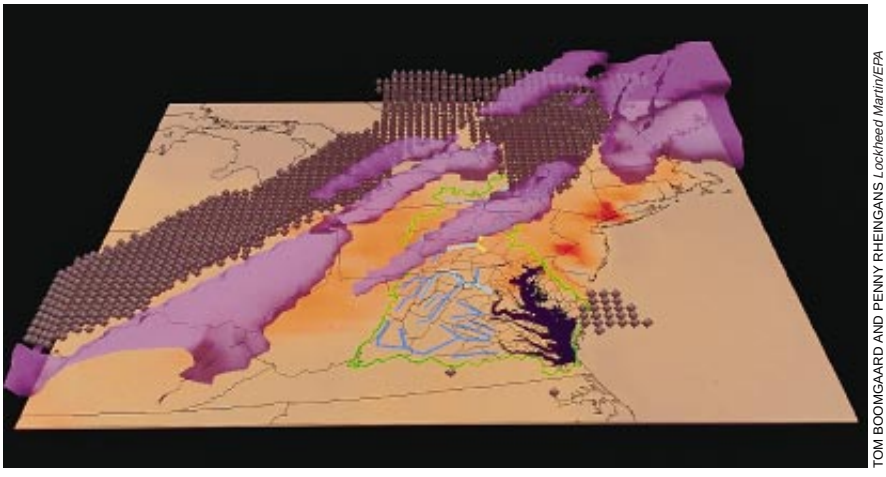
Much of the Chesapeake modeling was carried out at the EPA's three-year-old National Environmental Supercomputing Center, the world's only such facility devoted entirely to environmental issues. Currently the center provides computer time for about 40 different projects on topics such as urban air pollution or the effects of landfills. Instead of having to sample a huge region to determine where a toxic compound might end up if released by a factory, researchers using computers need only a few samples to establish original conditions. Then, intricate computer programs, which consider details down to the movement of atoms, fill in aspects such as how a compound will degrade in the environment, whether any secondary products will be toxic, how the chemicals might percolate down to the water table or how they might accumulate in wildlife. In some cases, the toxic compound being studied may not have been produced yet.

Such techniques can often save a great deal of money. In the early 1980s, researchers assessing the feasibility of a field experiment to study acid rain in the eastern U.S.—a project similar in scale to one that might test the findings from the Chesapeake model—put the price tag at $500 million. In contrast, Dennis estimates that the project to model the air pollution affecting the bay has cost around $500,000.

Hundreds of other major supercomputing centers offer services worldwide for research on topics ranging from ozone depletion to nuclear-waste disposal. Maureen I. McCarthy of Pacific Northwest Laboratories has used computers to predict how radioactive contaminants might behave in the soil around the Hanford, Wash., nuclear site. She argues that advances in theory and technology in the past five years have been so outstanding that researchers can now simulate chemical processes in the environment much more realistically. Realism is especially important in studies of hazardous-waste removal: experiments in the field can be expensive, time-consuming and difficult to carry out.

Yet for all their power, models cannot include every aspect of a natural system. And although experiments also cannot evaluate every detail, models in particular trigger complaints about accuracy. For instance, predictions about global warming have been controversial, because, as critics point out, various models, each with distinct assumptions, can give vastly different results.

If people will not believe a computer model that forecasts a rise in global temperature over the next century, it is unclear whether they will accept a computer's assessment of what is safe to put in drinking water. Having absolute faith in a simulation of an environmental problem can be tough, even for computer experts. Stephen E. Cabaniss of Kent State University notes that on a personal level, he might want to see results of toxicity experiments on animals before he would consider his tap-water safe. Indeed, Cabaniss and other high-tech types emphasize that for now, old-fashioned laboratory experiments as well as actual sampling of water, soil and air are still vital pieces of information needed to validate computer data to or nudge models in the right direction. Don't put away those lab coats yet. —*Sasha Nemecek*



TOM BOOMGAARD AND PENNY RHEINGANS *Lockheed Martin/EPA*

***COMPUTER MODEL** of the eastern U.S. reveals that pollution from western Pennsylvania, Ohio and Kentucky contributes to nitrogen levels in the Chesapeake Bay. Air pollution is shown in purple and rainfall in gray diamonds. Pollution that has fallen to the ground is represented in orange and water pollution in blue.*

ANTI GRAVITY

Into the Wild Green Yonder

The test of a first-rate intelligence, F. Scott Fitzgerald wrote between drinks, is the ability to hold two opposing ideas in the mind at the same time and still retain the ability to function. Such Fitzgeraldian thinking may help explain a U.S. Air Force program that was recently honored with an Innovations in American Government Award from the John F. Kennedy School of Government at Harvard University and the Ford Foundation.

Wanting to do its part to ensure that the earth remains blanketed by an unbroken, dependable layer of ultraviolet radiation–blocking ozone, the air force program phased out a particular use of ozone-depleting chemicals. The schizoid nature of this award-winning plan involves the ultimate purpose of the new, greener procedure: the air force is now employing environmentally friendly techniques to clean and repair ballistic-missile guidance systems.

On the environmental scoreboard, this is one of those good news–bad news stories, like the one about the Roman galley slaves who get extra food rations because the captain wants to go waterskiing:

• According to the air force, the phaseout cut the use of ozone-busting CFC-113 from two million pounds a year before 1988 to 18,000 in 1994. Good news.

• The new cleaning system uses only detergent and water, so it is actually much cheaper and faster than the old one. Good news.

• Methyl chloroform, an ozone depletor that posed a health risk to air force workers, has also been retired. Good news.

• The Aerospace Guidance and Metrology Center at Newark Air Force Base in Ohio, where the new cleaning procedure was developed, says the $100,000 award will be used to prepare and distribute a report on the program and to educate others using ozone-depleting solvents about the greener cleaning techniques. Good news.

• Ballistic-missile warheads can explode. Bad news.

Oddly enough, the air force was not behind another Kennedy School–Ford Foundation award winner: the Early Warning Program. It may sound like some strategy for intercepting bomb-carrying projectiles, but the U.S. Pension Benefit Guaranty Corporation actually came up with the program as a way to protect private pension plans. What with the cold war over, it may be that the greatest threats to national security include environmental damage and shaky investments.

So here's to the U.S. Air Force, whose environmental awareness is a promising sign that when the time comes to beat ballistic missiles into plowshares, we might still have an ozone layer under which to sow and reap. Provided, of course, that the earth isn't scorched first. —*Steve Mirsky*

MICHAEL CRAWFORD

Resisting Resistance

Experts worldwide mobilize against drug-resistant germs

The book-reading and moviegoing public has for some time been in a state of high anxiety about emerging infections. But the World Health Organization (WHO) finally put its seal of recognition on the topic last October, when it established a special bureau dedicated to fighting new and reemerging microbial threats. Although the amount of money set aside for the surveillance program so far is paltry—$1.5 million—hopes are high that the initiative will spur collaborations among existing research centers as well as stimulate other sources of funds.

The Rockefeller Foundation is considering expanding its support for infectious-disease laboratories in developing countries, according to Seth F. Berkley of the foundation's health sciences division, and the Federation of American Scientists is devising a global monitoring network. Member nations of the Pan-American Health Organization agreed last September on a plan of action, which was immediately activated to investigate an outbreak of leptospirosis in Nicaragua. The U.S. Centers for Disease Control and Prevention also has a new strategy for surveillance and prevention. Unfortunately, as several participants noted at a recent conference, the CDC program's budget last year was hardly more than Dustin Hoffman was paid for his performance in *Outbreak,* a recent movie about a deadly viral plague.

Alarm bells have been set ringing not only by the epidemics of pneumonic plague in India and of Ebola virus in Zaire but also by the continuing spread of resistance to antibiotics among many more familiar microbes. "The most frightening of these [microbial] threats is antibiotic resistance," declares June E. Osborn of the University of Michigan, a prominent public health expert. Malaria and tuberculosis, the infectious diseases that kill probably the most people worldwide, are now often resistant to standard drugs—and sometimes to second- and third-tier drugs, too.

In the U.S., pneumococci, which cause middle-ear infections and meningitis as well as pneumonia, are increasingly unfazed by many of the weapons in the pharmaceutical armory. Yet there are few organized and effective efforts to keep tabs on the new strains of germs. "Resistance has historically been a problem in hospitals, but it is now a problem equally in the community, and this is new this decade," notes Stuart B. Levy of Tufts University.

It is in hospitals that people are most vulnerable. *Staphylococcus aureus,* a cause of serious infections in wounds, is now resistant both to penicillin and, increasingly, to a semisynthetic form of the drug known as methicillin. The organism remains susceptible to a top-of-the-line antibiotic called vancomycin, but authorities fear that may not last.

KARI LOUNATMAA *Science Photo Library/Photo Researchers, Inc.*

STAPHYLOCOCCUS AUREUS, ***a common cause of infection in wounds, is becoming more resistant to penicillin.***

The American Society for Microbiology reported last year that between 1989 and 1993 there was a 20-fold increase in resistance to vancomycin among enterococci, a group of less dangerous bacteria that cause wound, urinary tract and other infections. But because the genes that confer vancomycin resistance can be carried on plasmids—small hoops of genetic material that occasionally cross the barrier between species—the resistance may yet jump to *S. aureus.* If so, surgery could become a markedly less safe proposition.

New drugs would be one solution. But despite some promising early-stage research, pharmaceutical companies do not expect to bring any new antimicrobial drugs to market until the end of the decade at the earliest. Meanwhile some workers, including Levy, are trying to devise other strategies to counter the threat. "Drug resistance is not inevitable if we use antibiotics wisely," he says. "It's sustained pressure that makes resistant strains predominate." Levy notes that a high proportion of the 150 million prescriptions for antibiotics written every year in the U.S. are for conditions that cannot be treated with such agents. Moreover, about half of the antibiotics used in the U.S. are fed to animals to prevent disease.

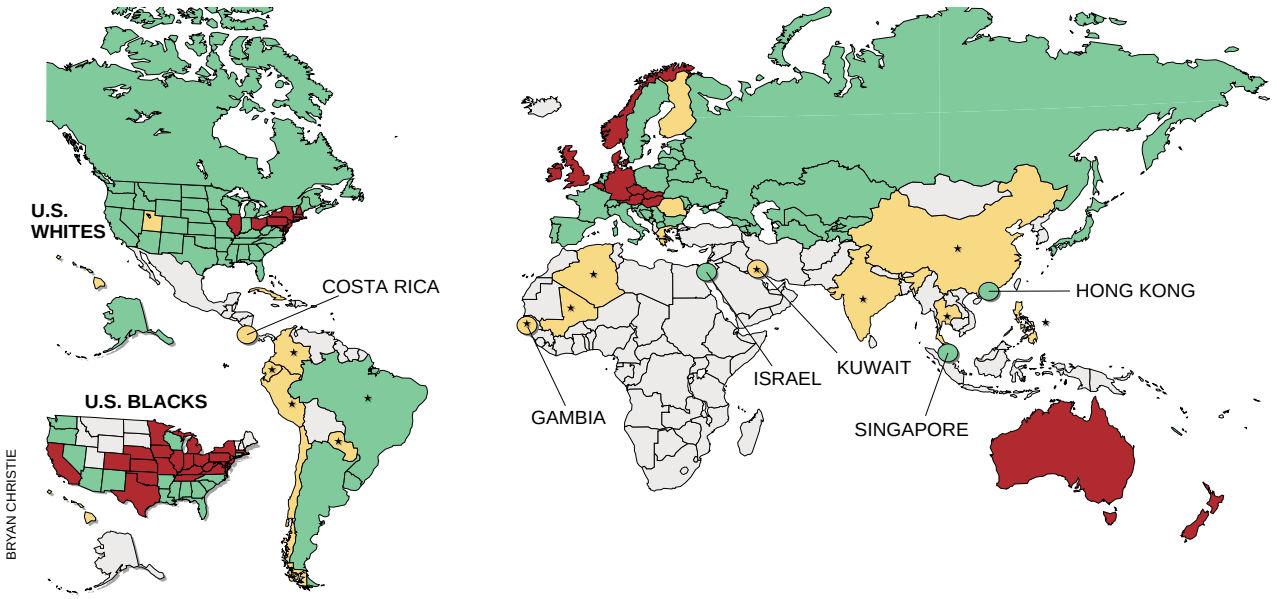
Levy has founded the Alliance for the Prudent Use of Antibiotics, which collects information about resistance and will try to counter it by recommending that drugs be rotated. "These are societal drugs," he maintains—meaning that their use has impacts beyond the patient for whom they are prescribed. Already some hospitals are putting restrictions on the use of vancomycin. But as Levy admits, the obstacles to countering resistance are formidable. Bacteria tend not to forget about drugs to which they have been exposed, so resistance declines only slowly after a drug is no longer used.

The WHO's official recognition of emerging infections, after decades of neglect, has been welcomed by most infectious-disease experts. But some question whether the agency has enough influence to make headway in countries that may be reluctant to admit that an epidemic is under way; after all, the WHO relies on voluntary cooperation from member countries.

Jonathan Mann, who resigned as head of the WHO's AIDS program in 1990, says the organization is not yet ready to assume a leadership role. Mann is promoting a binding international treaty to protect global health. The treaty would ensure that all worrisome outbreaks of disease are promptly investigated by an international team of qualified personnel. Mann has called for the U.S. to set an example the next time a mysterious outbreak of disease occurs within its own borders by inviting researchers from overseas to investigate the incident.

Mann's proposal is unlikely to be the last word on the subject. But the attention being focused on infectious disease indicates that a turning point may at least be in sight in one of humankind's oldest struggles. —*Tim Beardsley*

Colorectal Cancer Mortality among Men



Age-adjusted rates are shown for 36 industrial countries tabulated by the U.S. Centers for Disease Control and Prevention. Also shown are data for 13 developing countries (*stars*), as estimated from incidence data supplied by the World Health Organization. Data for most countries are for 1989, 1990 or 1991. Data for the individual successor states to the U.S.S.R., Yugoslavia and Czechoslovakia are not available separately.

A million people worldwide, about 145,000 of them in the U.S., will be diagnosed with colorectal cancer this year. Up to half a million, about 55,000 in the U.S., will die of the disease. Mortality from colorectal cancer rises progressively with age: in western Europe and English-speaking countries, it typically increases from fewer than one per 100,000 among those in the age 25 to 34 group to 170 or more in the age 75 to 84 group.

The highest rates are in Hungary and the former Czechoslovakia, which recorded, respectively, 46 and 47 deaths per 100,000. In the U.S., white males average 26 deaths per 100,000, whereas black men, who generally receive inferior medical care, average 32 per 100,000. The lowest mortality from colorectal cancer is in developing countries, such as India, which is estimated to have a rate less than one twentieth that of Western countries.

The large differences between developed and developing countries reflect differences in environment, genetic inheritance, way of life and, most important, diet. Countries such as the U.S. and Great Britain, where people typically eat meals rich in fat, meat, dairy products and protein, tend to have high rates of colorectal cancer; countries such as India and China, where diets are traditionally high in fiber, cereals and vegetables, tend to have low rates. People who migrate from a low-rate to a high-rate country—such as Greek migrants to Australia—tend to develop high rates of the disease as they acquire the habits of the host country, especially diet. The role of individual elements of diet in colorectal cancer is not clear, but dietary fat is perhaps the chief suspected culprit. There is also evidence supporting a beneficial effect from the consumption of vegetables.

In most industrial countries, mortality rates from colorectal cancer have declined in recent years because of greater use of early-detection methods and also, possibly, because of increasing awareness of the hazards of rich diets. A significant exception to the overall trend is in Japan, where rates have more than doubled since the 1950s as traditional diets were replaced by richer foods.

Exposure in the workplace to carcinogens such as asbestos may explain, at least in part, the high rates in Hungary and the former Czechoslovakia, where environmental safeguards have been lax, and in the northeastern U.S. Men have more exposure to workplace carcinogens than women do, which may help explain why rates for women are generally below those for men. —*Rodger Doyle*

The World According to RNA

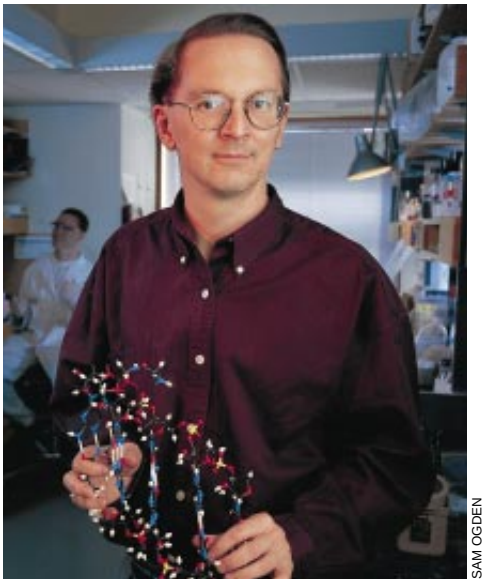
Experiments lend support to the leading theory of life's origin

In 1981 Francis Crick commented that "the origin of life appears to be almost a miracle, so many are the conditions which would have to be satisfied to get it going." Now, several findings have rendered life's conception somewhat less implausible. The results all bolster what is already the dominant theory of genesis: the RNA world.

The theory helped to solve what was once a classic chicken-or-egg problem. Which came first, proteins or DNA? Proteins are made according to instructions in DNA, but DNA cannot replicate itself or make proteins without the help of catalytic proteins called enzymes. In 1983 researchers found the solution to this conundrum in RNA, a single-strand molecule that helps DNA make protein.

Experiments revealed that certain types of naturally occurring RNA, now called ribozymes, could act as their own enzymes, snipping themselves in two and splicing themselves back together again. Biologists realized that ribozymes might have been the precursors of modern DNA-based organisms. Thus was the RNA-world concept born.

The first ribozymes discovered were relatively limited in their capability. But in recent years, Jack W. Szostak, a mo-



SAM OGDEN

JACK W. SZOSTAK is creating new forms of RNA, a model of which he holds.

lecular biologist at Massachusetts General Hospital, has shown just how versatile RNA can be. He has succeeded in "evolving" ribozymes with unexpected properties in a test tube.

Last April, Szostak and Charles Wilson of the University of California at Santa Cruz revealed in *Nature* that they had made ribozymes capable of a broad class of catalytic reactions. The catalysis of previous ribozymes tended to involve only the molecules' sugar-phosphate "backbone," but those found by Szostak and Wilson could also promote the formation of bonds between peptides (which link together to form proteins) and between carbon and nitrogen.

One criticism of Szostak's work has been that nature, unassisted, was unlikely to have generated molecules as clever as those he has found; after all, he selects his ribozymes from a pool of trillions of different sequences of RNA. Szostak and two other colleagues, Eric H. Ekland and David P. Bartel of the Whitehead Institute for Biomedical Research in Cambridge, Mass., addressed this issue in *Science* last July. They acknowledged that it would indeed be unlikely for nature to produce the most versatile of the ribozymes isolated by Szostak's methods. But they argued that the ease with which these ribozymes were generated in the laboratory suggested that they were almost certainly part of a vastly larger class of similar molecules that nature was capable of producing.

Szostak's work still leaves a major question unanswered: How did RNA, self-catalyzing or not, arise in the first place? Two of RNA's crucial components, cytosine and uracil, have been difficult to synthesize under conditions that might have prevailed on the newborn earth four billion years ago. The origin of life "has to happen under easy conditions, not ones that are very special," says Stanley L. Miller of the University of California at San Diego, a pioneer in origin-of-life research.

Last June, however, Miller and his U.C.S.D. colleague Michael P. Robertson reported in *Nature* that they had synthesized cytosine and uracil under plausible "prebiotic" conditions. The workers placed urea and cyanoacetaldehyde, substances thought to have been common in the "primordial soup," in the equivalent of a warm tidal pool. As evaporation concentrated the chemicals, they reacted to form copious amounts of cytosine and uracil.

Nevertheless, even Miller believes that a molecule as complex as RNA did not arise from scratch but evolved from some simpler self-replicating molecule. Leslie E. Orgel of the Salk Institute for Biological Studies in San Diego agrees with Miller that RNA probably "took over" from some more primitive precursor. Orgel and two colleagues recently noted in *Nature* that they had observed something akin to "genetic takeover" in their laboratory.

Orgel's group studied a recently discovered compound called peptide nucleic acid, or PNA; it has the ability to replicate itself and catalyze reactions, as RNA does, but it is a much simpler molecule. Orgel's team showed that PNA can serve as a template both for its own replication and for the formation of RNA from its subcomponents. Orgel emphasizes that he and his colleagues are not claiming that PNA itself is the long-sought primordial replicator: it is not clear that PNA could have existed under plausible prebiotic conditions. What the experiments do suggest, Orgel says, is that the evolution of a simple, self-replicating molecule into a more complex one is, in principle, possible.

Szostak, Miller and Orgel all say that much more research needs to be done to show how the RNA world arose and gave way to the DNA world. Nevertheless, life's origin is looking less miraculous all the time. —*John Horgan*

Rubbed Out with the Quantum Eraser

Making quantum information reappear

Atoms, photons and other puny particles of the quantum world have long been known to behave in ways that defy common sense. In the latest demonstration of quantum weirdness, Thomas J. Herzog, Paul G. Kwiat and others at the University of Innsbruck in Austria have verified another prediction: that one can "erase" quantum information and recover a previously lost pattern.

Quantum erasure stems from the standard "two-slit" experiment. Send a laser beam through two narrow slits, and the waves emanating from each slit interfere with each other. A screen a short distance away reveals this interference as light and dark bands. Even particles such as atoms interfere in this way, for they, too, have a wave nature.

But something strange happens when you try to determine through which slit each particle passed: the interference pattern disappears. Imagine using excited atoms as interfering objects and, directly in front of each slit, having a special box that permits the atoms to travel through them. Each atom therefore has a choice of entering one of the boxes before passing through a slit. It would enter a box, drop to a lower energy state and in so doing leave behind a photon (the particle version of light). The box that contains a photon indicates the slit through which the atom passed. Obtaining this "which-way" information, however, eliminates any possibility of forming an interference pattern on the screen. The screen instead displays a random series of dots, as if sprayed by shotgun pellets. The Danish physicist Niels Bohr, a founder of quantum theory, summarized this kind of action under the term "complementarity": there is no way to have both which-way information and an interference pattern (or equivalently, to see an object's wave and particle natures simultaneously).

But what if you could "erase" that telltale photon, say, by absorbing it? Would the interference pattern come back? Yes, predicted Marlan O. Scully of the University of Texas and his co-workers in the 1980s, as long as one examines only those atoms whose photons disappeared [see "The Duality in Matter and Light," by Berthold-Georg Englert, Marlan O. Scully and Herbert Walther; SCIENTIFIC AMERICAN, December 1994].

Realizing quantum erasure in an experiment, however, has been difficult for many reasons (even though Scully offered a pizza for a convincing demonstration). Excited atoms are fragile and

easily destroyed. Moreover, some theorists raised certain technical objections, namely, that the release of a photon can disrupt an atom's forward momentum (Scully argues it does not).

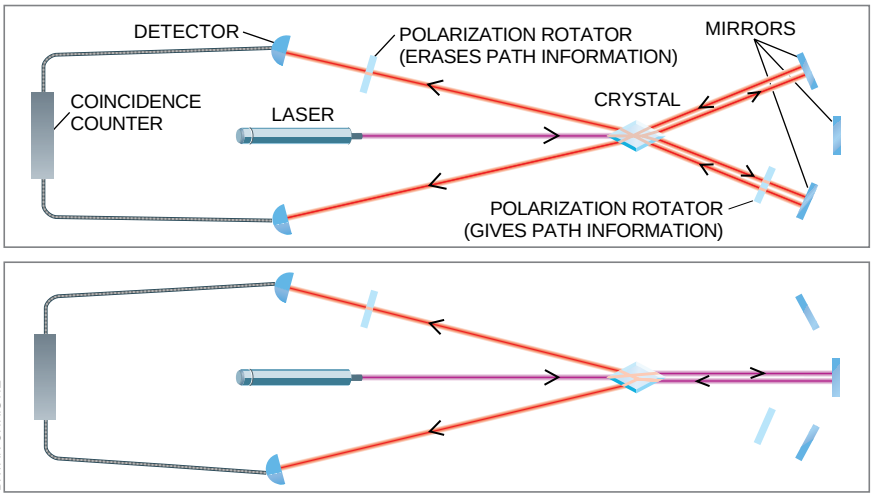
The Innsbruck researchers sidestepped the issues by using photons rather than atoms as interfering objects. In a complicated setup, the experimenters passed a laser photon through a crystal that could produce identical photon pairs, with each part of the pair having about half the frequency of the original photon (an ultraviolet photon became two red ones). A mirror behind the crystal reflected the laser beam back through the crystal, giving it another opportunity to create photon pairs. Each photon of the pair went off in separate directions, where both were ultimately recorded by a detector.

Interference comes about because of the two possible ways photon pairs can be created by the crystal: either when the laser passes directly through the crystal, or after the laser reflects off the mirror and back into it. Strategically placed mirrors reflect the photons in such a way that it is impossible to tell whether the direct or reflected laser beam created them. These two birthing possibilities are the "objects" that interfere. They correspond to the two paths that an atom traversing a double slit can take. Indeed, an interference pattern emerges at each detector. Specifically, it stems from the "phase difference" between photons at the two detectors. The phase essentially refers to slightly different travel times through the apparatus (accomplished by moving the mirrors). Photons arriving in phase at the detectors can be considered to be the bright fringes of an interference pattern; those out of phase can correspond to the dark bands.

To transform their experiment into the quantum eraser, the researchers tagged one of the photons of the pair (specifically, the one created by the laser's direct passage through the crystal). That way, they knew how the photon was created, which is equivalent to knowing through which slit an atom passed. The tag consisted of a rotation in polarization, which does not affect the momentum of the photon. (In Scully's thought experiment, the tag was the photon left behind in the box by the atom.) Tagging provides which-way information, so the interference pattern disappeared, as demanded by Bohr's complementarity.

The researchers then erased the tag by rotating the polarization again at a subsequent point in the tagged photon's path. When they compared the photon hits on both detectors (using a so-called coincidence counter) and correlated their arrival times and phases, they found the interference pattern had returned. Two other, more complicated variations produced similar results. "I think the present work is beautiful," remarks Scully, who had misgivings about a previous claim of a quantum-eraser experiment performed a few years ago.

More than just satisfying academic curiosity, the results could have some practical use. Quantum cryptography and quantum computing rely on the idea that a particle must exist in two states—say, excited or not—simultaneously. In other words, the two states must interfere with each other. The problem is that it is hard to keep the particle in such a superposed state until needed. Quantum erasure might solve that problem by helping to maintain the integrity of the interference. "You can still lose the particle, but you can lose it in such a way that you cannot tell which of two states it was in" and thus preserve the interference pattern, Kwiat remarks. But even if quantum computing never proves practical, the researchers still get Scully's pizza. —*Philip Yam*



QUANTUM ERASURE relies on a special crystal, which makes pairs of photons* (red) *from a laser beam* (purple) *in two ways: either when the beam goes through the crystal directly* (top) *or after reflection by a mirror* (bottom). *Devices that rotate polarization indicate—and can subsequently erase—a photon's path information.

Return of the Red Wolf

Controversy over taxonomy endangers protection efforts

The legislative corridors of Washington, D.C., have recently been resounding with howls about red wolves. At issue is the ongoing, federally sponsored program to reintroduce red wolves in parts of North Carolina and Tennessee. Some critics of the program question whether or not the red wolf is really a species—that is, biologically distinct from other groups of wolves and coyotes.

In early August 1995 the National Wilderness Institute (NWI), a wildlife management organization, submitted a petition to the Department of the Interior, recommending the removal of the red wolf from the Endangered Species List. Citing Robert K. Wayne of the University of California at Los Angeles and John L. Gittleman of the University of Tennessee [authors of "The Problematic Red Wolf," SCIENTIFIC AMERICAN, July 1995], the NWI petition contends that "the 'red wolf' is not a separate species but a hybrid" and therefore "cannot meet the Endangered Species Act's definition of species."

Drawing on genetic clues, Wayne and Gittleman suggest that the red wolf is probably a relatively recent hybrid of the coyote and a now extinct subspecies of the gray wolf. This hypothesis, however, contradicts research by Ronald M. Nowak of the U.S. Fish and Wildlife Service. Nowak points to fossil evidence indicating that "a small red wolf-like animal was present in the southern United States throughout the Pleistocene and right down to our times." Nowak states that fossils also reveal that ancestors of today's wolves and coyotes evolved into separate groups around one million years ago; red and gray wolves diverged about 300,000 years ago. Thus, Nowak proposes that the similar genetic makeup of the three species could

mean that the groups still retain many of their original genetic traits.

Although Wayne and Gittleman question the red wolf's status as a species, they have also argued that reintroduction efforts are still merited in locations where the wolf can serve an important ecological function. But according to James R. Streeter, policy director at the NWI, "the question is not whether it is ecologically useful to have a large carnivore [such as the red wolf] in the Southeast." The question, Streeter contends, is whether the animal qualifies for protection under the Endangered Species Act, given that it is not, in the opinion of some experts, a species.

In the past, lawyers from the Department of the Interior ruled that the act does, in some cases, cover hybrids, but now there is no official policy on the matter. So biologists at the Fish and Wildlife Service must decide how to respond to the NWI petition. Initially, Gary Henry, the service's red-wolf coordinator, recommended that the petition be denied because he felt it did not include any new information on the red wolf's taxonomic status.

"I had already finished this finding,"

MIDDLETON/LITTSCHWAGER

RED WOLF is endangered, but is it a species? Some critics of the federally sponsored protection program say no.

Henry comments, when he received word that officials at the Fish and Wildlife Service headquarters in Washington, D.C., also wanted to see arguments in favor of accepting the petition. Acceptance requires the service to evaluate data for one year before making a decision about whether the red wolf should be protected under the Endangered Species Act. Henry submitted both memorandums to the service's regional office in Atlanta; a decision is pending.

Coincidentally, shortly after the NWI petition was written, Senator Jesse Helms of North Carolina proposed legislation that would have canceled all funding for the red-wolf program. A spokesperson for Helms states that the senator was not aware of the NWI petition; instead Helms was responding to constituents' complaints about the supposed dangers presented by the red wolf. The legislation was defeated by a vote of 50 to 48. —*Sasha Nemecek*

THE ANALYTICAL ECONOMIST

Flying Blind

In an era when Congress may ask schoolchildren to skip lunch to help balance the budget, it sounds eminently reasonable that bureaucrats at arcane federal agencies such as the Bureau of Economic Analysis (BEA) or the Bureau of Labor Statistics (BLS) should share in the general pain. The same logic might lead a skipper trying to lighten an overburdened ship in the middle of the ocean to jettison sextant, chronometer and compass. Economists worry that, without social science data to measure their effects, there may be no way to tell whether the various policy experiments now being enacted are succeeding or failing.

The status of U.S. economic statistics is already "precarious," says Alan B. Krueger of Princeton University. He notes that the BLS has reduced the size of its statistical samples (thus compromising accuracy) and dropped many kinds of data entirely. Even such seemingly basic information as manufacturing turnover—the rate at which people quit factory jobs and companies hire replacements—is no longer available.

Krueger also laments the passing of the annual census of occupational fatalities. Although the BLS retains its general surveys of job safety, those figures rely on complaints filed with the Occupational Safety and Health Administration and so tend to be less reliable, he says. If proposed House and Senate budget cuts go through, international price, wage and productivity comparisons will have to be scrapped, forcing U.S. policymakers to rely on dead reckoning when they try to compare domestic workers with their European or Asian counterparts.

Important though such data may be, says William G. Barron of the BLS, no law requires its collection, and his agency will have its hands full preserving core programs such as the consumer price index and national unemployment statistics. The BLS would also eliminate its long-term economic projections, surveys of relative employment in 750 different occupations and its tracking of the employment status of older women.

The BEA, meanwhile, would stop collecting most of the data it currently acquires on multinational corporations, according to acting director J. Steven Landefeld. If anyone wants to know whether U.S. companies are producing an increasing proportion of their wares overseas or how many conglomerates from elsewhere are opening plants in the U.S., they could be out of luck. The same will hold for those who want to find out how much the U.S. is spending on pollution control or how much it is taking in from tourists.

Cutbacks at the bureau may also force it to delay publication of such basic statistics as quarterly estimates of the gross national product. The agency's 1996 budget calls for 40 days of "furloughs"—essentially distributing eight weeks' worth of temporary layoffs throughout the year.

Even more troubling, however, Landefeld says, is the probable elimination of state-by-state breakdowns for national income estimates. If Congress does not know who is earning how much (at least on average), it will have only a sketchy basis for apportioning more than $100 billion in federal transfers to the states. If plans to replace a raft of additional federal programs with block grants go forward, the amount of money to be allocated by guesswork could increase substantially.

Will a later Congress come to what economists would call its senses and restore some of the nation's financial instrumentation? Assuming that nothing has run seriously aground in the meantime, restoring the databases will still be difficult. Many of the numbers that economists depend on are cumulative, Krueger notes: each month builds on the preceding month's data, and any gap must be papered over by intellectually shaky estimates. Similarly, survey results become notoriously less accurate if respondents must reconstruct their behavior from months or years ago. At the moment, however, enforcing the "Contract with America" seems uppermost in the legislature's mind. Landefeld probably speaks for his profession when he observes, "We're not at the top of the agenda." —*Paul Wallich*

Return of the Breeder

Engineers are trying to teach an old reactor new tricks

Once upon a time, when fuel prices were high, nuclear fast-breeder reactors enjoyed brief fame based on a singular claim. While producing energy by splitting some uranium atoms, they could create an even larger number of plutonium atoms. This plutonium could then be turned into fuel to generate much more energy.

Economics and politics, though, have not been kind to breeder reactors. With oil prices at historic lows and former cold war adversaries awash in plutonium and uranium, the idea has seemed to lose considerable luster. In February 1994 Secretary of Energy Hazel R. O'Leary ended U.S. research into breeder technology—after some $9 billion had been spent on it.

Almost two years later, however, in a climate as hostile as ever to the technology, breeder reactors are—almost incredibly—resurging. In the past year the largest such reactor ever built, the 1,240-megawatt Superphénix near Lyons, France, was restarted after a long hiatus following some technical problems. A smaller breeder reactor in Japan generated electricity for the first time last August. And in recent months, engineers in India, which is pursuing two different breeder-reactor technologies, were preparing to connect a tiny experimental breeder reactor near Madras to the electricity grid.

These operational milestones were supplemented by a study and conference supporting the technology. Last August a panel led by Nobel laureate Glenn T. Seaborg issued a report calling on industrial countries to develop and use breeder and other reactors as a way of making more fossil fuels available to less developed countries, many of which are struggling to electrify. The panel, assembled by the American Nuclear Society, also chided the U.S. government for halting its breeder research. Then, in early October, a technical meeting, held in Madras under the auspices of the Vienna-based International Atomic Energy Agency, drew experts from Russia, Japan, China, the Republic of Korea, Brazil and India. According to an attendee, participants concluded that breeders have a high level of operational reliability and safety.

But others looking at the same data might call the record mixed. In the U.S., for example, three breeder reactors were built, two of which were significant. Argonne National Laboratory's Experimental Breeder Reactor II operated for three decades (until 1994) without any serious problems. On the other hand, a commercial, power-generating plant named Fermi began operating in 1963 near Detroit and suffered a partial core meltdown three years later. It was repaired but soon closed because of safety concerns. France's Superphénix, too, has had problems, mostly linked to flaws in its liquid-sodium cooling system. (Such a coolant is necessary in a fast-breeder reactor because the water used in conventional reactors would slow the neutrons liberated by fission, limiting the number that could cause breeding—the conversion of uranium 238 to useful plutonium 239.) In late October a steam leak forced a temporary shutdown of the plant.

There are several reasons for the interest in expensive, exotic plants to make fuel, even though there is plenty of it around. For Japan and India, especially, the impetus is national self-sufficiency. These countries have relatively few fuel resources and appear to be planning for a day when fuel is not so cheap. "In Japan, actually, we don't need a fast-breeder reactor in this century," says Toshiyuki Zama, a spokesman for Tokyo Electric Power Company, the largest Japanese utility. "But we have to develop technologies for the future." More pragmatically, Japanese officials spent some $6 billion on the 280-megawatt breeder reactor, named Monju after the Buddhist divinity of wisdom, and are eager to recoup some of this outlay by generating electricity.

France, which already has large amounts of plutonium from its extensive nuclear power program, plans to convert Superphénix so that it can destroy plutonium rather than produce it. According to engineering manager Patrick Prudhon, a reactor core and fuel rods are being designed for this purpose as part of a project budgeted at $200 million a year. The new hardware, to be tested after the year 2000, will let the reactor consume about 150 kilograms of plutonium per year, Prudhon figures. Unfortunately, France's power reactors add about 5,000 kilograms of plutonium every year to an already sizable stockpile. Growing accumulations of plutonium have fueled concerns that some of the poisonous, fissile element could fall into the wrong hands [see "The Real Threat of Nuclear Smuggling," by Phil Williams and Paul N. Woessner, page 40].

"Our aim is to demonstrate the capability, to let the politicians of the next century decide whether it is a good opportunity to use fast reactors to destroy plutonium," Prudhon comments. In fact, in this mode the reactor can destroy not just plutonium but virtually all the actinides present in nuclear waste. Actinides are isotopes with atomic numbers between 89 and 103. Some are radioactive for thousands of years, making them the most troublesome components of waste.

Far from the esoteric, futuristic notion it has become, this application was envisioned even before the nuclear power industry was born. "From the 1940s on, it was always [Enrico] Fermi's idea to use fast-spectrum reactors to consume all the actinides," says H. Peter Planchon, associate director of the engineering division of Argonne National Laboratory. This way "you would be faced with disposing of fission products with relatively short half-lives. That's still the view of the French and Japanese." —*Glenn Zorpette*

ERIC BOUVET *Gamma Liaison*

***FUEL-HANDLING** and coolant-pump machinery are visible in the dome over the Superphénix reactor vessel. The reactor is near Lyons, France.*

Making Free Software Pay

The Internet creates an alternative economics of innovation

According to conventional economic theory, the Internet is impossible. Textbooks say people will only innovate—or do anything, for that matter—if they are financially rewarded. Yet the software that has created and that runs the world's biggest computer network is for the most part given away. So the network should not have been built, let alone grown to be one of the most innovative realms in the fast-moving world of computing.

Part of the solution to this conundrum is, of course, that there are other ways of rewarding people. For many a hacker, the excitement of innovation and the prospect of being known as a "Net god" are enough reason to toil over free software. But as the Net becomes more commercialized, fame alone will pale beside riches. Which brings back the original puzzle: How can the cooperative Internet make room for individual gain without losing the shared core of technology and information that sustains it?

One way is to emulate Netscape, the most successful of the companies rushing to commercialize the Internet. When Marc Andreessen wrote Mosaic—which quickly became the most popular browser for reading the World Wide Web's vast array of text, sounds, pictures and just plain neat stuff—he sensed opportunity and joined with Jim Clark of Silicon Graphics. The two founded Netscape and developed a second, commercial generation of the software. Although the team made its first, barebones product available for free, it now sells more advanced products.

But Netscape's way is not the only one. Many of the basic tools used to build the Web are still given away. PERL (the Practical Extraction and Report Language) makes it easy to write programs that respond to text messages with specific actions—making it exactly the right tool for building Web sites. Written by Larry Wall, PERL is distributed for free. So is TCL—the Tool Control Language, pronounced "tickle"—which was created by John Ousterhout when he was at the University of California at Berkeley and which is used to build quick and easy programs on Unix workstations.

Now that big-money projects are coming to rely on this software, however, firms are worried. Users are dependent on the goodwill of others to fix bugs, provide technical support and make improvements. So far goodwill has worked wonders, but cynical executives are reassured only by clear-cut contractual responsibilities. Enter Michael Tiemann, who built Cygnus Support into a $10-million-a-year company by making "free software affordable."

Most of Cygnus's business centers on gcc—a compiler for the C++ programming language originally written by legendary hacker Richard Stallman of the Massachusetts Institute of Technology but since improved by Tiemann and others. At Stallman's insistence, the source code to gcc is free, but companies can pay Cygnus to modify gcc, adapt it to new hardware and answer their technical queries.

The key distinction between Cygnus and Netscape lies in the ownership of the product. Although Netscape may make public some of the technical standards to which its product complies, the product itself is a jealously guarded source of competitive advantage. Cygnus's competitive advantage lies in the expertise it brings to modifying its products. Cygnus approaches the software business as a service industry rather than a manufacturing one.

Although at first blush it sounds an unlikely way to make money, Cygnus's software-for-free, service-for-fee strategy could foster innovation. It creates an easy-to-cross bridge between the academic world and the commercial one. It removes software buyers' perennial fear of being held hostage to the success of their favorite supplier. It distributes, and thereby speeds, the work of adapting software to all the different bits of hardware used on the Internet. It enables rapid, continuous innovation that is directed by users and also benefits all users. It provides a straightforward mechanism for a group to innovate rapidly and yet remain united by a common core of technology.

The drawback is that Cygnus does not offer much incentive to invest in the original product. Nevertheless, there are plenty of products on the Internet to which the model is perfectly suited. Cygnus's Tiemann is thinking of offering a support package for PERL, TCL and other popular Web tools. Still languishing in academia are a variety of other useful tools—ranging from e-mail programs to programming languages and sophisticated modeling software.

Perhaps in the future some companies will band together to jump-start the process of creating free software in order to build common technology for their shared use and improvement. After all, networks are making a world in which machines share work across corporate boundaries as well as those of space and time. Giving away software to reach that end may be more profitable than it sounds. —*John Browning*

Freewheeling

Most people would think that a wheelchair with "legs" makes about as much sense as a fish with a bicycle. Vijay Kumar of the General Robotics and Active Sensory Perception Laboratory at the University of Pennsylvania thinks differently. His ungainly, motorized wheelchair can tackle all but the most difficult terrains—in fact, wheelchair-bound people may soon be able to roam the beach.

Unlike traditional wheelchairs, which are conveyed solely by their wheels, Kumar's creation uses two legs to help the wheels along. A computer detects whether the wheels or the legs are getting better traction and channels the motor's power accordingly. On a flat, smooth surface, the wheels work easily. But in sand or mud, "the wheels slip, so the legs dig in and pull," Kumar explains. The result is as graceful as a bionic sea turtle, but it works.

Inside the house, the new wheelchair can open doors, push debris aside and step over small objects. It can also climb over curbs a foot high; a small modification will allow it to climb stairs. Future models may use the "feet" to dip into a toolbox for attachments such as claws for turning doorknobs. Although the wheelchair is only a prototype, and there are no plans for mass production, Kumar's approach might open whole new worlds to people confined to wheelchairs. —*Charles Seife*

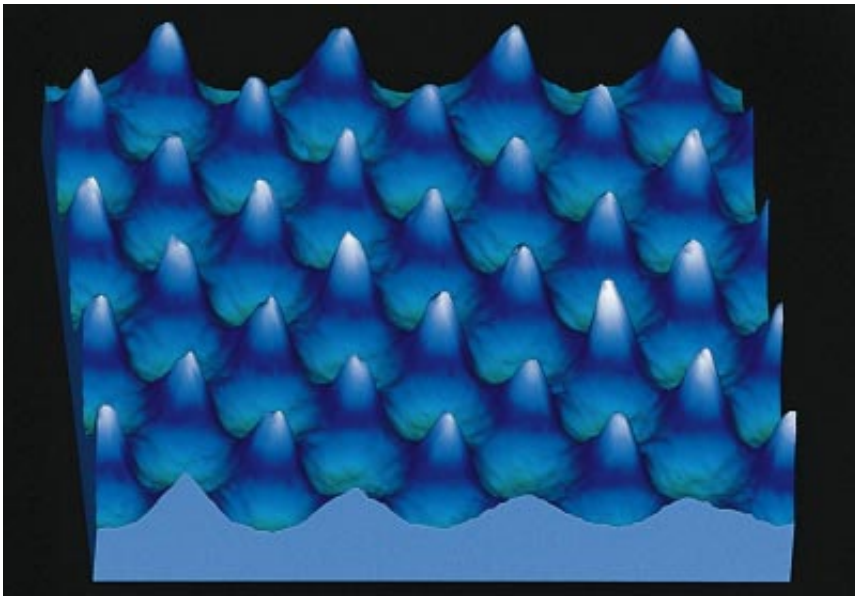
NAJLAH FEANNY *SABA*

Light over Matter

All mass-produced computer chips are etched from disks of silicon using flashes of light, projected through stencils, to draw circuit patterns. Cranking up the frequency of the light, from green to blue and recently to ultraviolet, engineers have shrunk circuits' size and boosted their speed. But the technique will soon hit its limits: at frequencies higher than ultraviolet, light turns into unwieldy x-rays that are hard to focus.

One alternative may be to use light as the stencil and to project the matter. Jabez J. McClelland and his colleagues at the National Institute of Standards and Technology recently used this strategy to draw a grid of chromium dots on a tiny slab of silicon. The dots are just 80 nanometers wide—significantly smaller than anything ultraviolet light can paint. With further development, the physicists believe, their technique could draw two billion circuit patterns on a centimeter-square chip in just a few minutes.

The trick to writing so small is to use lasers as lenses. McClelland and cohorts boiled atoms off a block of chromium in an oven, then focused them into a tight, tiny beam. They directed the stream through "optical molasses"—a laser beam set just below the frequency at which chromium atoms resonate like struck bells—which slowed the atoms. Just before the chilled particles hit the silicon slab, they ran into another laser beam skimming the silicon surface. This second beam was



JABEZ J. MCCLELLAND *National Institute of Standards and T*

***CRISSCROSSED LASERS** focused chromium atoms into tiny dots, each just 80 nanometers wide. The technique may one day draw circuits.*

The Midnight Hour

Japan ventures onto the Net in the dark of night

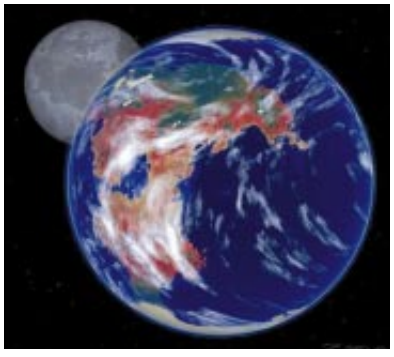
At midnight, computer screens across Tokyo light up as Internet users take advantage of cheaper telephone rates to surf the World Wide Web. In Japan, where telephone fees make Internet access five to 10 times costlier than in the U.S., the new night-owl pricing is bringing more people into cyberspace. "Two years ago I was the first commercial Internet provider in Japan," says Roger Boisvert, president of Global OnLine Japan. "Today there are 45 Internet service providers in Tokyo alone."

Driven by strong computer sales and popular fascination with the Web, Japan is just catching the Internet fever already widespread in the U.S. and Europe. "The Internet and the World Wide

Playing Slartibartfast with Fractals

Most computer artists use high-tech tools but old-fashioned techniques, such as painting with electronic brushes, sculpting with virtual chisels and altering with digital versions of darkroom tricks. Ken Musgrave, a computer scientist and landscape artist at George Washington University, produces his works in a way only computers can: he programs them. The results are spectacularly realistic vistas that can be explored from virtually any perspective and distance.

At the heart of Musgrave's art are fractals, surfaces drawn by repeating certain relatively simple equations over and over. Fractal surfaces have infinite detail—the closer you look, the more you see—yet require only a few lines of computer code.



KEN MUSGRAVE

Recently Musgrave has been writing a system to produce an entire virtual planet, called Gaea, accompanied by a moonlike satellite named Selene. Gaea looks similar to an earth barren of life, not only from outer space (*far left*) but also up close. Musgrave has "discovered" some of his favorite landscapes (*near left* and *right*) wandering the surface of Gaea.

"I call my program the Slartibartfast system, after the character in *The Hitchhiker's Guide to the Galaxy* who created Earth," Musgrave says. The system's power is rivaled by its simplicity: it comprises about

aimed at a mirror and carefully tuned to form a standing wave (one whose troughs and peaks stay fixed in space), again just below the resonant frequency of chromium.

Stumbling upon a standing wave, atoms feel a strong urge to surf up onto a crest or down into a trough. The wave thus acts like a lens, deflecting the atoms passing through it into neat lines spaced a half wavelength apart on the silicon. Cross two lasers over the substrate, as McClelland did, and the lines split into a grid of dots. The next step is scanning the lasers across the surface to draw arbitrary patterns: nanocircuits.

Laser-focused atomic deposition, physicists' catchy name for this technique, still has many hurdles to clear on the way to factory floors. Not all the atoms get focused, for instance, so the peaks are connected by a bed of metal that would short any circuit. It may not be possible to etch material away without destroying the pattern. But because the technology could theoretically produce wires 10 times smaller than those made by the photolithography processes used today, it is likely at least to focus attention. *—W. Wayt Gibbs*

Web have become popular trends, especially with the young people," says Masaya Nakayama of Japan's Network Information Center in Tokyo.

Shinichi Maeda, Tokyo marketing manager for Cisco Systems, a top vendor of network routers, says Internet hosts are growing at the rate of 300 percent a year in Japan. That compares with the 50 percent annual growth of the country's on-line services. Such services—Japan's versions of CompuServe and America Online—have three million subscribers. Although the number of Internet users is as elusive in Japan as elsewhere, Maeda says that at current rates, the number may surpass on-line service subscribers by late 1997.

As in the U.S. and Europe, Japanese on-line services are racing to stay ahead. Many are trying to provide Internet access lest their subscribers desert them for the Web. Yet they are somewhat insulated from direct competition because their services are in Japanese, whereas most Internet resources are in English.

Internet providers are, of course, trying to come out ahead as well. So intense is the competition that one Tokyo provider is even giving away accounts in hopes of building a following. Justnet is a division of Just System Corporation, which controls 50 percent of the word-processing market with a program called Ichitaro. Justnet built a Web browser into the version of Ichitaro released in August and began offering free Internet usage. Justnet's Timothy Gleeson states 100,000 users signed up in the first four months. "The target is a million users by 1997," says Gleeson, who is not sure when Justnet will start charging for Internet service.

But even these free Internet accounts cost money. Unlike in the U.S., where consumers pay a flat rate for local calls, Japanese telephone customers pay seven to 10 yen for each three minutes on a local call: a surcharge of $1.40 to $2.00 per hour. "The lack of a flat rate for local calls is limiting Internet development in Japan," comments Naoki Yamamoto, editor of *Digital Highway Report,* a newsletter for Japanese information managers.

Under government pressure, Nippon Telephone & Telegraph recently began offering a flat rate for accessing the Internet between 11:00 P.M. and 8:00 A.M. "We get our heaviest usage at midnight," says Boisvert of Global OnLine Japan. But observers say real price competition must wait until Japan undergoes telephone industry deregulation similar to that in the U.S. and Europe.

Indeed, Japan's legendary industrial planners seem to have misjudged Internet policy so far. While the U.S. Defense Advanced Research Projects Agency and National Science Foundation pushed the Internet's technology through in the 1970s and 1980s, various Japanese government agencies backed competing network protocols—none of which became popular. Fortunately, Jun Murai of Keio University won corporate backing for his Japan University Network, the precursor of today's academic and commercial Internet services in Japan.

Regardless of the regulatory hurdles and the high cost, it is clear that the psychology of getting connected has taken root in Japan. The tendrils of the Internet are spreading beyond the big cities, deep into traditional land. Noriyosi Yoshida, vice president of Hiroshima City University, is helping to create a local network to give Hiroshima residents access to municipal and academic resources as well as to the Internet. "We sincerely expect to maintain an eternally peaceful world," Yoshida says. "I believe one of the most effective ways to promote this is through the Internet and the World Wide Web." *—Tom Abate*

1,000 lines of computer code (fewer than a typical Nintendo game). Musgrave's aesthetic, which he calls proceduralism, dictates that the programs should be as short and fast as possible. As a result, Musgrave's worlds look right for all the wrong reasons: the models have nothing whatsoever to do with the laws of physics. His rings of Saturn, for example (*far right*), consist of a fractal line, varying in transparency along its length, swept around a circle.

"My goal is to get an interactive renderer in a $200 box," Musgrave chuckles. "I figure that every kid who can afford it will have to have one, because it will let them explore an infinite universe of detailed planets. Of course, game makers will inevitably infest all these lovely worlds with hostile aliens."

Musgrave will have to accelerate his program by several orders of magnitude for that to happen. A Gaean landscape still takes several minutes to render. Of course, just a few years ago it would have taken hours. It may not be long before anyone can play Slartibartfast in a virtual universe. *—W. Wayt Gibbs*

Additional images and a video clip zooming from outer space to the mountains of Gaea are available from Scientific American *on America Online.*

PROFILE: JOSEPH ROTBLAT

From Fission Research to a Prize for Peace

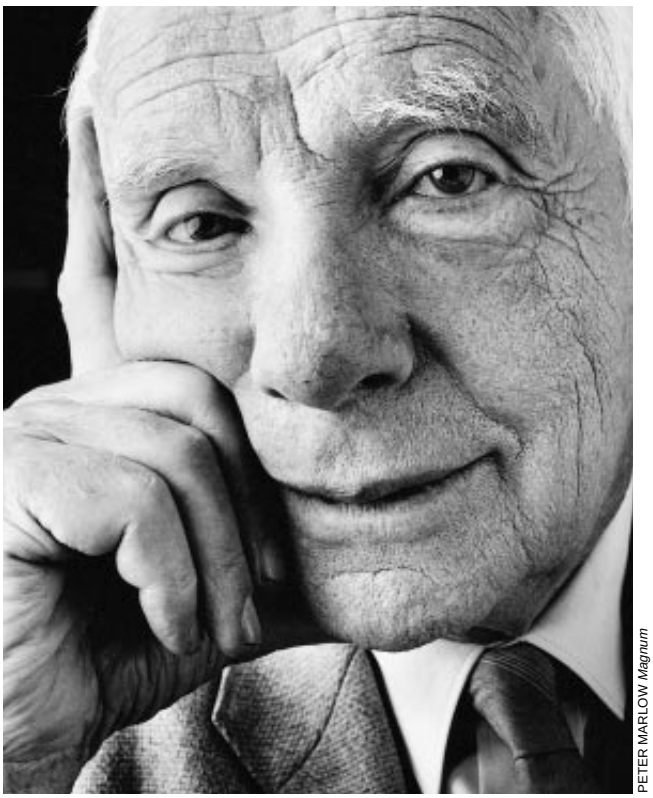
The building of the atomic bomb is the tale of the century. From that experience have come many stories of scientists ensnared in the web of national politics or entranced by the search for the fundamentals of the universe. There was one physicist, however, who marched to a different drummer, who left the Manhattan Project when it was discovered the Germans were not building a bomb.

"The one who paused was Joseph Rotblat," the physicist Freeman Dyson once wrote, "who, to his everlasting credit, resigned his position at Los Alamos." Joseph Rotblat left Los Alamos National Laboratory in New Mexico in 1944, while there was still time to write a different history for this century. A nuclear physicist, Rotblat transformed his career to medical physics and passionately pursued disarmament. Last year Rotblat was awarded the Nobel Peace Prize for his efforts to eliminate nuclear weapons from the planet.

A vigorous man with thinning white hair, Rotblat spoke about his decision several years ago at a meeting of Physicians for Social Responsibility in Chicago: "This was truly a choice between the devil and the deep blue sea," he said. "The very idea of working on a weapon of mass destruction is abhorrent to a true scientist; it goes against the basic ideals of science. On the other hand these very ideals were in danger of being uprooted, if—by refusing to develop the bomb—a most vile regime were enabled to acquire world domination. I do not know of any other case in history when scientists were faced with such an agonizing quandary.

"Four years after I started work on the bomb, serious doubts began to occupy my mind about this work. It became daily clearer to me that Germany, with its vastly extended military operations and crippling damage to its industry, was most unlikely to be able to build the bomb, even if its scientists hit on the right idea of how to make it. The reasons for which I sacrificed my principles were rapidly wearing off. This led me to the decision to resign from the project."

Rotblat lost his wife, his home, his world, to the Nazis. Many people suffering such losses would have retreated into themselves. Instead, from reserves that few can fathom, he took on a very public career and began working for nuclear disarmament. An intensely private man, he agreed to tell me his story.



PETER MARLOW *Magnum*

JOSEPH ROTBLAT, who received the 1995 Nobel Peace Prize, left the Manhattan Project in 1944, before its completion.

Rotblat was born in Warsaw in 1908. Turn-of-the-century Poland was a peasant nation with a veneer of sophisticated city gentry. Rotblat's parents were Jewish; his father was in the paper-transport business. Life included a pony and summers in the country. World War I ended that idyll. The family business failed. "In the basement in the house in which we were living we distilled *somogonka*—illicit vodka—as a way of earning a living," Rotblat recalled. "One had to fight for one's survival."

Rotblat obtained his degree in 1932 and began research at the Radiological Laboratory of Warsaw. Working in Poland in the 1930s with few of the amenities of his Western European colleagues, Rotblat asked the right questions and found some of the answers. During this period, Rotblat married Tola Gryn, a student of Polish literature. In 1939 he accepted an invitation from James Chadwick to work at the University of Liverpool. Liverpool's cyclotron was part of the attraction; Rotblat hoped to build one in Warsaw upon his return. Just as Rotblat was planning his trip to England, nuclear physics was thrown into turmoil. Two German chemists, Otto Hahn and Fritz Strassmann, split the uranium atom by firing neutrons at it; the process of nuclear fission resulted. The experiment had not converted lead into gold, but its consequences were as significant. A large amount of energy is released during fission. So are some neutrons.

How many was crucial. If it were just a single neutron, there was little chance that the new neutron would also hit a uranium nucleus and so continue the process. But if two or more were products of the splitting, then the probability of a chain reaction would increase. A number of physicists around the globe, including Rotblat, set out to find the answer. He soon discovered that several surplus neutrons are released from each fissioning uranium atom—but he was beaten to publication by France's Frédéric Joliot-Curie.

"I began to think about the consequences and the possibility that a chain reaction can proceed at a very fast rate," Rotblat said. "Then, of course, there could be an explosion because of the enormous amount of energy produced in a short time." Rotblat traveled to England on his own; his fellowship gave him too little money for two. Six months later he received additional funds, and in late August 1939, Rotblat returned to Poland to make arrangements for his wife to join him in Liverpool.

He left Poland first; Tola was to join him shortly. Because there had been a partial news blackout in Poland, Rotblat and his wife were unaware of how serious the situation had become. The Nazis invaded Poland on the first of September, and the conflict was over within a few weeks. Rotblat sought transit

visas for his wife through Belgium, Denmark and Italy; each time, borders closed before his wife could leave Poland. Rotblat would never see her again.

Back in England, Rotblat decided the immediate danger from the Nazis was so great that "one had to put aside one's moral scruples regarding the bomb." With Chadwick's help, Rotblat began experiments in Liverpool to investigate the potential for an atomic bomb. Conditions were not exactly easy. "Almost every night, I was doing several hours of firewatching, for incendiary bombs." Nevertheless, by 1941 British researchers had established that the bomb was theoretically possible.

Although U.S. researchers had made much progress toward a self-sustaining nuclear reaction—a reactor—their efforts toward an explosive device had been stymied. The British restored the Americans' belief in the bomb. Churchill and Roosevelt agreed to set up a joint research facility in the U.S. The British team, including Rotblat, would work with the Americans. After moving to Los Alamos, Rotblat learned of American plans for the bomb. He recalled that one night at dinner General Leslie Groves, military commander of the Manhattan Project, "mentioned that the real purpose in making the bomb was to subdue the Soviets." Rotblat began to speak with other Los Alamos physicists about not using the bomb, but the usual response was that "we started an experiment; we must see it through."

Events in Europe were overtaking the researchers. Rotblat continued, "In late 1944 Chadwick told me that an intelligence report indicated that the Germans weren't working on the bomb. A few days later I told him I wanted to leave."

Threatening him with arrest should he speak about it, Los Alamos security agents kept Rotblat from discussing his decision with the other scientists. Instead he told his colleagues that he was returning to Europe in order to be closer to his family (although he had heard nothing from them during the war). After the war ended, he discovered that his wife had perished, while his mother, sister and two brothers had survived.

Rotblat returned to Liverpool at the beginning of 1945. He kept his silence until the dropping of the bombs on Japan that August. He realized that the atomic bomb "was a small beginning of something much larger. I could foresee the coming of the hydrogen bomb." He began to give talks across Britain, attempting to convince his fellow physicists to call a moratorium on nuclear research.

Rotblat also began a transition to medical applications of physics, and within several years he moved to Saint Bartholomew's Hospital in London. His investigations of treatments for cancer led Rotblat to studying the effects of radiation on healthy subjects with Patricia J. Lindop, a physiologist. Ironically, this work led him back to the bomb. "Even in 1957, which was 12 years after the bomb, many people did not believe that cancer results from radiation," Rotblat said. "They used to say that only leukemia is induced by radiation, not other cancers. From the work I did with Lindop on mice, I could see that all sorts of cancer were produced."

In 1954 Rotblat met Bertrand Russell, who had been growing increasingly concerned about the dangers of the nuclear arms race. The British philosopher suggested that a group of scientists be convened for the purpose of discussing nuclear disarmament. And so Pugwash—the movement of scientists with which Rotblat shared the Nobel Peace Prize—was born. Pugwash got its name from the Nova Scotia town where the first meeting was held. "It was very small, with 22 people," Rotblat reminisced. But what 22 people! The participants included three Nobel laureates, the vice president of the Soviet Academy of Sciences and a former director-general of the World Health Organization.

Rotblat believes scientists must bear a moral responsibility for their discoveries.

It was an extraordinary undertaking, at a complicated time. "Anyone in the West, to come to such a meeting, to talk peace with the Russians, was condemned as a Communist dupe," Rotblat noted. "It was a risk, a gamble. It could have just broken up in disarray. As it turned out, people really spoke up and argued—but argued as scientists." The conference's brief report detailed the radiation hazards of nuclear testing, made recommendations on arms control and stated several principles of scientists' social responsibility. The world's leaders listened. Pugwash meetings continued.

In 1961, a year of high tension between East and West, a Pugwash conference brought together the vice president of the Soviet Academy of Sciences and the U.S. presidential science adviser. Afterward, they met with President John F. Kennedy and discussed a nuclear test ban. A treaty banning aboveground testing of nuclear weapons was signed in 1963. Subsequent Pugwash meetings helped to pave the way for peace negotiations between the U.S. and North Vietnam in the late 1960s and for the 1972 Treaty on Anti-Ballistic Missile Systems between the U.S. and the U.S.S.R. For many years, Rotblat's office at Saint Bartholomew's Hospital served as Pugwash's headquarters. Rotblat organized the conferences, wrote histories of the movement and served as the secretary-general for 14 years. In 1988 he was elected president of Pugwash, a position he still holds. Some call him "Mr. Pugwash."

It is easy to believe that with the end of the cold war and reductions of nuclear arsenals, Pugwash's objectives have been achieved. Rotblat knows well that the world is not so simple. The new situation has new instabilities—Russia is a prime example. Nor has the end of the cold war diminished the desire of Iraq and North Korea, for instance, to join the nuclear club.

"I do not believe that a permanent division into those who are allowed to have nuclear weapons and those who are not is any basis for stability in the world," Rotblat declared. "Therefore, the ultimate solution is the elimination of nuclear weapons. How can we prevent one nation from secreting a few weapons away? This is a task for scientists, primarily a technological problem ensuring that no one is cheating." Economic considerations are also important, Rotblat said: "If we are to have disarmament, we have to see that the transition from military industries to peaceful industry—the problem of conversion—can be arranged so as not to cause economic upheavals."

Perhaps the greatest task for Pugwash, and for all of humanity, is creating "a climate of trust and goodwill" among all the world's people. "We have to develop in each of us a sense of loyalty to mankind that will be an extension of our present loyalties to our family, our city, our nation." Scientists, who "are to a large extent citizens of the world," can and should lead this educational effort, Rotblat said.

Rotblat has a large classical record collection waiting for his retirement. That time has not yet come. At 87, his energy is that of a man half his age; he continues to lecture and attend meetings worldwide. In December he was scheduled to travel to Oslo, Norway, for the awarding of the Nobel Prize for Peace. He has come a long way for someone whose first venture outside Poland was at the age of 30 in the spring of 1939. —*Susan Landau*

The Real Threat of Nuclear Smuggling

Although many widely publicized incidents have been staged or overblown, the dangers of even a single successful diversion are too great to ignore

by Phil Williams and Paul N. Woessner

During past centuries, most people who thought of smuggling at all considered it a somewhat esoteric profession—a way of avoiding taxes and supplying goods that could not be obtained through licit channels. Drugs added a more insidious dimension to the problem during the 1970s and 1980s, but trade in uranium and plutonium during the past five years has given smuggling unprecedented relevance to international security.

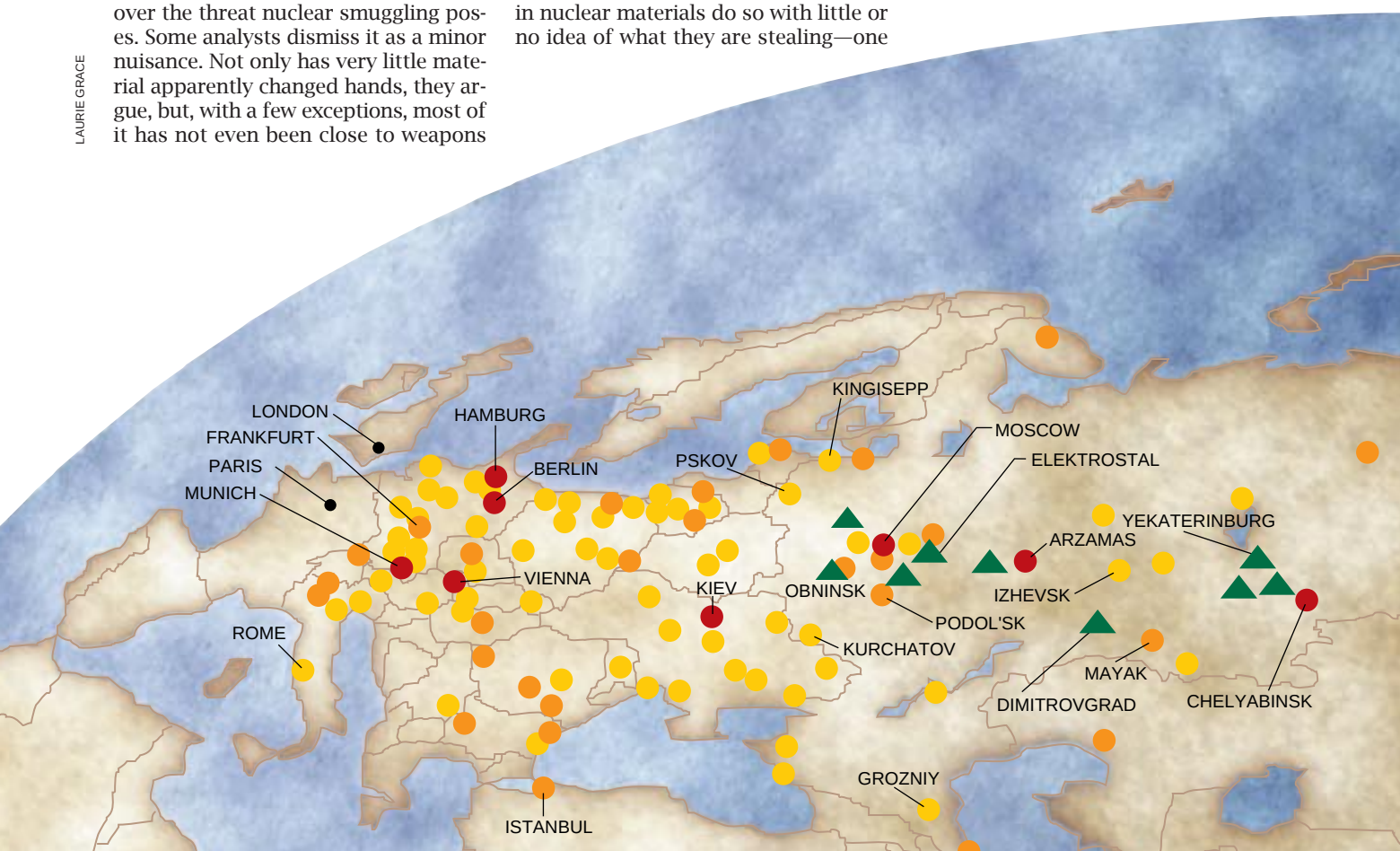
Yet there is considerable controversy over the threat nuclear smuggling poses. Some analysts dismiss it as a minor nuisance. Not only has very little material apparently changed hands, they argue, but, with a few exceptions, most of it has not even been close to weapons grade. None of the radioactive contraband that has been confiscated by Western authorities has been traced unequivocally to weapons stockpiles. Some of the plutonium that smugglers try to peddle comes from smoke detectors.

In addition to these amateur smugglers, there are many scam artists who sell stable elements that have been rendered temporarily radioactive by exposing them to radiation or who obtain large advances based on minute samples. Indeed, many of those who traffic in nuclear materials do so with little or no idea of what they are stealing—one Pole died of radiation poisoning after carrying cesium in his shirt pocket, and a butcher in St. Petersburg kept uranium in a pickle jar in his refrigerator.

The Danger Is Real

Based on the bumbling nature of most of the smuggling plots uncovered so far, some well-informed observers have suggested that, in Germany at least, the only buyers are journalists, un-

LAURIE GRACE



dercover police and intelligence agents. Some go even further and contend that pariah states such as Iraq, Iran, Libya or North Korea may not be interested in acquiring illicit nuclear arsenals at a time when they are in the process of trying to reestablish normal relations with the West.

Nevertheless, nuclear smuggling presents a grave challenge. In almost all illicit markets, only the tip of the iceberg is visible, and there is no reason why the nuclear-materials black market should be an exception. Police seize at most 40 percent of the drugs coming into the U.S. and probably a smaller percentage of those entering western Europe. The supply of nuclear materials is obviously much smaller, but law-enforcement agents are also less experienced at stopping shipments of uranium than they are in seizing marijuana or hashish. To believe that authorities are stopping more than 80 percent of the trade would be foolish.

Moreover, even a small leakage rate could have vast consequences. Although secrecy rules make precise numbers impossible to get, Thomas B. Cochran of the Natural Resources Defense Council in Washington, D.C., estimates that a bomb requires between three and 25 kilograms of enriched uranium or between one and eight kilograms of plutonium. A kilogram of plutonium occupies about 50.4 cubic centimeters, or one seventh the volume of a standard aluminum soft-drink can.

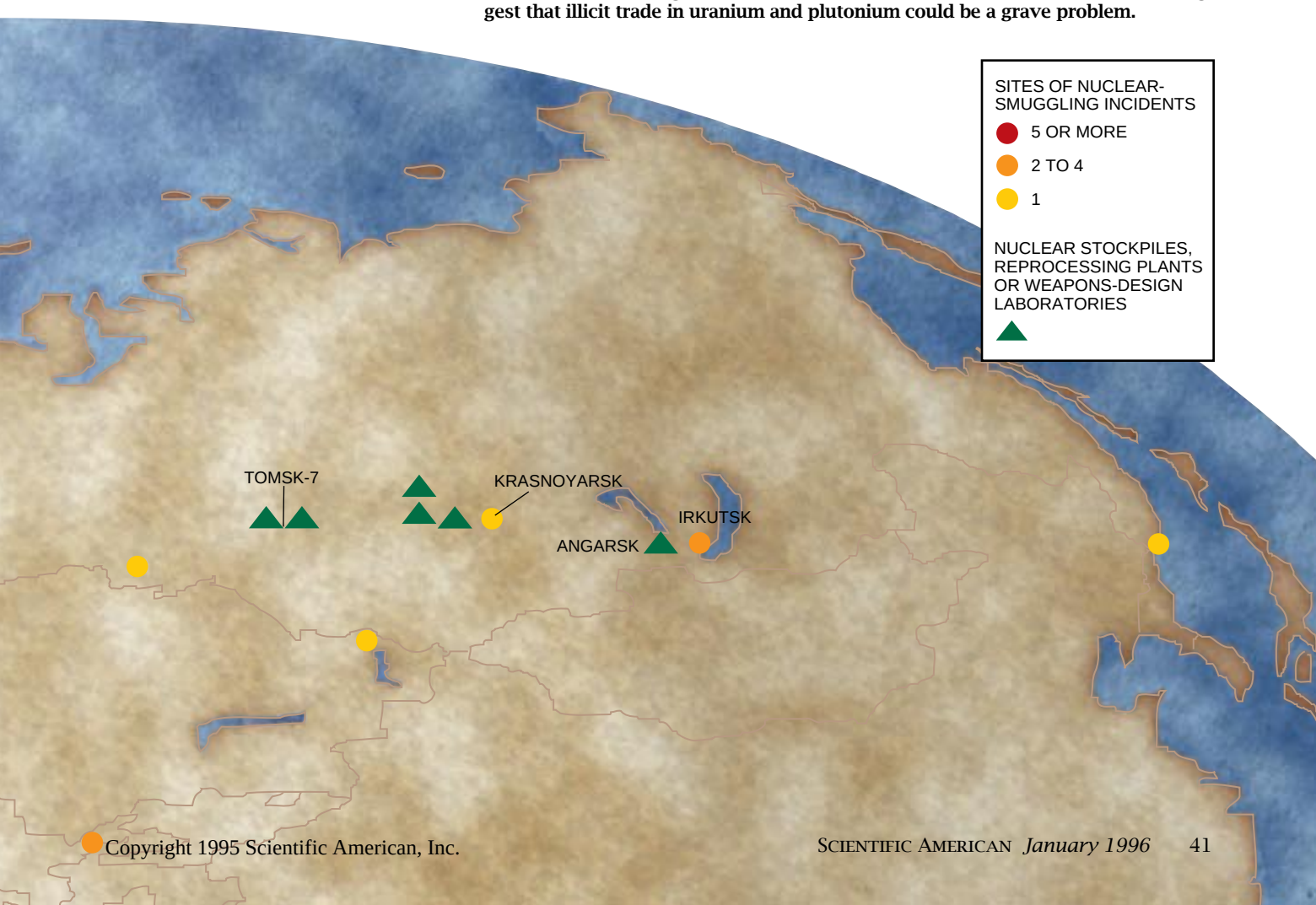
Although rigorous screening of all international shipments could catch some radioactive transfers, several of the most dangerous isotopes, such as uranium 235 and plutonium 239, are only weakly radioactive and so could be easily shielded from detection by Geiger counters or similar equipment. X-ray and neutron-scattering equipment, such as that in place at airports to detect chemical explosives, could uncover illicit radionuclides as well, but because it is not designed for the task its practical effectiveness is limited.

If the amounts of material needed for nuclear weapons are small in absolute terms, they are minuscule in comparison to the huge stockpiles of highly enriched uranium and plutonium, especially in Russia, where both inventory control and security remain quite problematic. World stocks of plutonium, which totaled almost 1,100 tons in 1992, will reach between 1,600 and 1,700 tons by the year 2000, enough to make as many as 200,000 10-kiloton bombs. As disarmament agreements are implemented, another 100 tons of refined weapons-grade plutonium will become available in the U.S. and Russia—ironically, in the post–cold war world, one of the safest places for plutonium may well be on top of a missile.

Security Is Lax

In addition, the U.S. and former Soviet states each hold about 650 tons of highly enriched uranium. These large stockpiles are all the more disturbing because control over them is fragile and incomplete. The Russian stores in particular suffer from sloppy security, poor inventory management and inadequate measurements. Equipment for determining the amount of plutonium that has been produced is primitive—yet without a clear baseline, it is impos-

NUCLEAR-SMUGGLING INCIDENTS have been reported across central Europe to the Pacific coast of Russia (dots show sites of seizure, origin or transfer). Security at some stockpiles is being upgraded, but unsettled economic and political conditions are undermining morale. Hundreds of incidents over the past five years suggest that illicit trade in uranium and plutonium could be a grave problem.



A Nuclear Bestiary

	Licit Use	Illicit Use
Americium 241	Alpha-particle source for smoke detectors and other devices	Fraud (substitute for more desired elements)
Beryllium	Neutron reflector in reactors or bombs	Illicit reactors; nuclear weapons
Cesium 137	Radiation source for industrial or medical applications; present in radioactive waste from reactors	Fraud; murder by radiation
Cobalt 60	Gamma-radiation source for industrial or medical applications	Fraud; murder by radiation
Lithium 6	Thermonuclear weapons	Thermonuclear weapons
Plutonium	Alpha-particle source for smoke detectors; nuclear weapons; nuclear reactor fuel	Fraud; nuclear weapons
Polonium 210	Alpha-particle and neutron source for industrial applications	Nuclear weapons
Uranium	Nuclear reactor fuel; nuclear weapons	Fraud; nuclear weapons
Zirconium	Structural material for nuclear reactors	Illicit reactors

KARL GUDE (*drawings*)

sible to know what may be missing.

Virtually nonexistent security at nuclear installations compounds the problem. The collapse of the KGB took with it much of the nuclear-control system. Ironically, under the Soviet regime security was tight but often superfluous. Nuclear workers were loyal and well paid and enjoyed high status. As pay and conditions worsen, however, disaffection has become widespread. With an alienated workforce suffering from low and often late wages, the incentives for nuclear theft have become far greater at the very time that restrictions and controls have deteriorated.

In November 1993 a thief climbed through a hole in a fence and entered a supposedly secure area in the Sevmorput shipyard near Murmansk. He used a hacksaw to cut through a padlock on a storage compartment that held fuel for nuclear submarines and stole parts of three fuel assemblies, each containing 4.5 kilograms of enriched uranium. Although the uranium was eventually recovered, Mikhail Kulik, the official who conducted the investigation of the theft, was scathing in his report: there were no alarm systems, no lighting and few guards. Kulik noted: "Even potatoes are probably much better guarded today than radioactive materials." Improvements in security at the base since the incident have been very modest.

The situation is not entirely gloomy. According to reports, some of the nuclear cities—formerly secret sites where bombs were designed and built—are well secured, and the controls on weapons-grade materials are generally more stringent than those on lower-quality items. Although efforts to enter the closed zone at Arzamas-16 (the Russian weapons-design laboratory that is a rough counterpart to Los Alamos National Laboratory in the U.S.) have reportedly doubled during the past year, the security system there appears to remain effective. Moscow is also making efforts to reestablish security throughout its nuclear industry—in some cases, such as the Kurchatov Institute of Atomic Energy in Moscow, with direct assistance from the U.S. Yet the task is formidable. Nearly 1,000 stores of enriched uranium and plutonium are scattered throughout the former Soviet Union.

The Rise of Smuggling Networks

Against this background, it is hardly surprising that the number of nuclear-smuggling incidents—both real and fake—has increased during the past few years. German authorities, for example, reported 41 in 1991, 158 in 1992, 241 in 1993 and 267 in 1994. Although the vast majority of cases do not involve material suitable for bombs, as the number of incidents increases so does the likelihood that at least a few will include weapons-grade alloys.

In March 1993, according to a report from Istanbul, six kilograms of enriched uranium entered Turkey through the Aralik border gate in Kars Province. The material had apparently been brought from Tashkent to Grozniy, where Chechen crime groups entered the picture, then to Nakhichevan via Georgia, before arriving in Istanbul. Although confirmation of neither the incident nor the degree of the uranium's enrichment was forthcoming, it raised fears that Chechen "Mafia" groups had obtained access to enriched uranium in Kazakhstan. Kazakhstan's agreement in 1994 to transfer enriched uranium to the U.S. suggests that such speculation may have had some basis.

In October 1993 police in Istanbul seized 2.5 kilograms of uranium 238 and detained four Turkish businessmen, along with four suspected agents of Iran's secret service. A Munich magazine later reported that the uranium may have gone to Turkey via Germany. According to one of the Turkish detainees (a professor who had previously

been involved in the smuggling of antiquities), accomplices flew the uranium by Cessna to Istanbul from Hartenholm, a private airfield near Hamburg owned by Iranian arms dealers.

Significantly, 1994 saw several incidents involving material that was either weapons grade or very close to it. On May 10 police in Tengen, Germany, found six grams of plutonium 239 while searching the home of businessman Adolf Jaekle for other contraband. The plutonium, which was in a container in the garage, was discovered only by accident. Jaekle had widespread connections, including links with former officers of the KGB and the Stasi (the East German secret police) and with Kintex, a Bulgarian arms company that has long been suspected of a wide range of nefarious activities. Much of the initial speculation has dissipated, but important questions about the Jaekle case remain unanswered. It would be unwise to exclude the possibility that the plutonium was simply a sample for a much larger delivery.

On August 10 authorities in Munich arrested a Colombian dentist and two Spaniards in possession of 363.4 grams of high-grade plutonium and 201 grams of lithium 6 (a component of hydrogen bombs). They had brought their contraband to Munich from Moscow on a Lufthansa flight and were captured amid much fanfare. It later turned out that agents from the German federal intelligence body, the BND, had induced the three men to bring in the material.

The operation caused great controversy in Germany; BND agents were accused of helping to create rather than control the nuclear-smuggling problem. The three men were connected neither with Colombian drug gangs nor Basque terrorists; there was no evidence that they were experienced smugglers. They simply had financial problems and had been trying to solve them by selling the lithium and plutonium.

In all the controversy over the propriety of the BND's actions, however, an important point was lost. Even as amateurs, the three men succeeded in obtaining a significant amount of high-grade plutonium.

Then, on December 14, police in Prague arrested three men in a car with 2.7 kilograms of highly enriched (87.7 percent) uranium 235. Two were nuclear workers who had come to the Czech Republic in 1994: a Russian from a town near Obninsk and a Belarusian from Minsk. The third was a Czech nuclear physicist, Jaroslav Vagner, who had not been officially employed in the nuclear industry for several years. In mid-1994 a similarly enriched sample of uranium had apparently turned up in Landshut, Bavaria, and on March 22, 1995, two more men, one of them a police officer, were arrested in connection with the December incident.

KATHERINE LAMBERT

DEPLETED URANIUM SLUG, weighing roughly seven kilograms, fits comfortably in Thomas B. Cochran's hand. The physicist, who works for the Natural Resources Defense Council in Washington, D.C., estimates that a similar amount of weapons-grade material would be enough to construct a bomb capable of destroying a small city.

The number of smuggling cases in Germany, at least, has declined since these highly publicized arrests. Traffickers appear to be going elsewhere. Some have gone through Switzerland and Austria and into Italy. More may be taking the routes to the south through the Central Asian republics and the Black Sea. As former International Atomic Energy Agency inspector David Kay has pointed out, these paths in effect reverse those used by the KGB to smuggle Western goods into the former Soviet Union. Border controls in these areas are much weaker than those going into western Europe, and the potential clients are closer.

Some of the seizures in Germany and Turkey make it fairly clear that outlaw states such as Iran may in fact be looking for high-quality nuclear material. It appears, indeed, that some of them have set up their own networks. Both Libya and Iraq have experience with such methods, since each nation set up front companies to facilitate the illicit diversion of precursor chemicals and equipment to develop chemical weapons.

Furthermore, as the Jaekle case implies, smugglers are not all blundering amateurs. Although there is no monolithic nuclear Mafia, ex-spies from the former Soviet bloc countries appear to be taking a leading role in the professional networks. They have apparently been joined by entrepreneurs, often involved in the arms business, whose dealings span a continuum from the licit, through the shady, to the illicit.

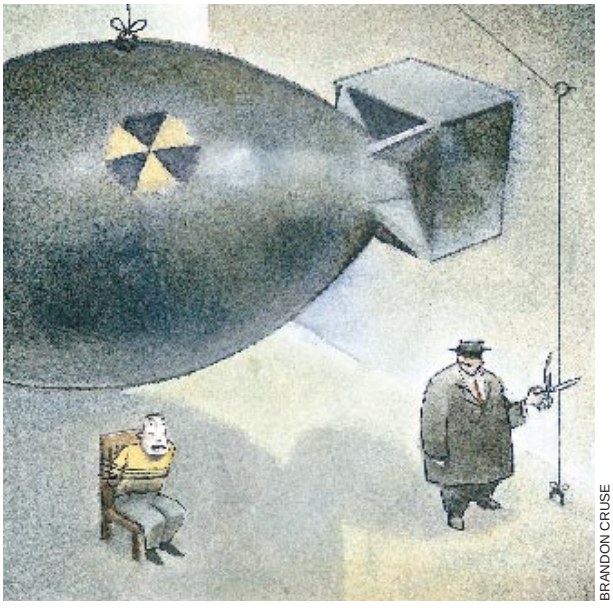
Not surprisingly, because nuclear smuggling is a potentially profitable business, organized crime groups have also become involved. Some Turkish gangs appear to be engaged in the trade—having graduated from clandestine export of antiquities, they treat uranium as just another commodity.

In Italy, Romano Dolce, a magistrate investigating the nuclear trade, was arrested for participation in the very crimes he was pursuing. This scandal aroused considerable speculation that he had focused on some cases in order to divert attention from other, more significant transactions.

Bombs and the Mob

Although outlaw nations probably make up most of the market for nuclear weapons, there is a clear danger that organized crime groups or terrorists could also join the nuclear club. The transition from transporting nuclear contraband to using it directly is apparently an easy one: radioactive isotopes have already been used for murder. In late 1993 Russian "Mafia" assassins allegedly planted gamma-ray-emitting pellets in the office of a Moscow businessman, killing him within months. At least half a dozen similar incidents have been reported in Russia since then.

A criminal organization could also use radionuclides for large-scale extortion against a government or corporation. It would be fairly easy for a nuclear blackmailer to establish credibility by leaving a sample for analysis. Subsequent threats to pollute air or water supplies, or even to detonate a small nuclear weapon, could have considerable leverage. Nor can the possibility be entirely excluded that a terrorist organization or an extremist cult such as the one that allegedly carried out nerve-gas attacks in 1995 in the Tokyo subway might acquire nuclear materials.



BRANDON CRUSE

Even if such a group did not acquire enough uranium or plutonium to make a fission weapon, they might be able to mix unstable isotopes with conventional explosives to create widespread contamination. If the terrorists who bombed the World Trade Center in New York City had made use of radioactive substances, for example, they might have killed thousands and rendered a major business district uninhabitable. *—P.W. and P.N.W.*

Perhaps the most insistent question, however, concerns the involvement of Russian organized crime. Although nuclear trafficking does not seem to be a priority for these groups—other activities are both more immediately lucrative and less risky—there is growing evidence that some Russian criminal groups are diversifying into trade in radioactives.

Enforcement Efforts Lag

Even though serious efforts are being made to attack the problem at the source, the international community has been slow to respond to the dangers that nuclear smuggling presents. The Russian nuclear regulatory agency, GAN, now has 1,200 employees, but the degree of authority it can actually wield over the old nuclear bureaucracy—both civilian and military—is uncertain.

Furthermore, even if GAN is successful, it will take several years to upgrade safeguards, and smugglers are not going to sit by idly in the meantime. As a result, there will be a premium on good intelligence and law enforcement during the remainder of the 1990s. Unfortunately, international agencies with nuclear expertise are not yet cooperating effectively with those whose responsibility is to stop illicit trade. The IAEA and the United Nations Crime Prevention and Criminal Justice Branch are both located in Vienna's International Center, but the IAEA's mandate does not allow it to engage in investigative activity. As a result, contacts between the two have been little more than desultory.

In Washington, meanwhile, early responses to the smuggling problem have been ill conceived and poorly coordinated. Since 1994, the Federal Bureau of Investigations has taken the lead and has been working closely with the Defense Nuclear Agency and the Defense Intelligence Agency, but the U.S. remains some distance from a comprehensive policy.

We suggest that systematic multinational measures be taken as soon as possible to inhibit theft at the source, to disrupt trafficking and to deter buyers. The U.S., Germany, Russia and other nations with an interest in the nuclear problem should set up a "flying squad" with an investigative arm, facilities for counterterrorist and counterextortion actions and a disaster management team.

Such an idea seems very far-fetched at the moment, at least in part because of a continuing reluctance to recognize the severity of the threat. It would be a tragedy if governments were to accept the need for a more substantive program only after a nuclear catastrophe.

The Authors

PHIL WILLIAMS and PAUL N. WOESSNER work at the Ridgway Center for International Security Studies at the University of Pittsburgh. Williams, who directs the center, is a professor in the graduate school of public and international affairs. During the past three years, his research has focused on transnational criminal organizations and drug trafficking, and he is the editor of a new journal, *Transnational Organized Crime.* Woessner, a research assistant at the Ridgway Center, received his master's degree in international affairs in 1994. He also earned an M.S. in planetary science and a B.S. in astronomy and physics, the latter at the University of Maryland.

Further Reading

"POTATOES WERE GUARDED BETTER." Oleg Bukharin and William Potter in *Bulletin of the Atomic Scientists,* Vol. 51, No. 3, pages 46–50; May–June 1995.

CHRONOLOGY OF NUCLEAR SMUGGLING INCIDENTS: JULY 1994–JUNE 1995. Paul N. Woessner in *Transnational Organized Crime,* Vol. 1, No. 2, pages 288–329; Summer 1995.

NUCLEAR MATERIAL TRAFFICKING: AN INTERIM ASSESSMENT. Phil Williams and Paul N. Woessner in *Transnational Organized Crime,* Vol. 1, No. 2, pages 206–238; Summer 1995.

Caloric Restriction and Aging

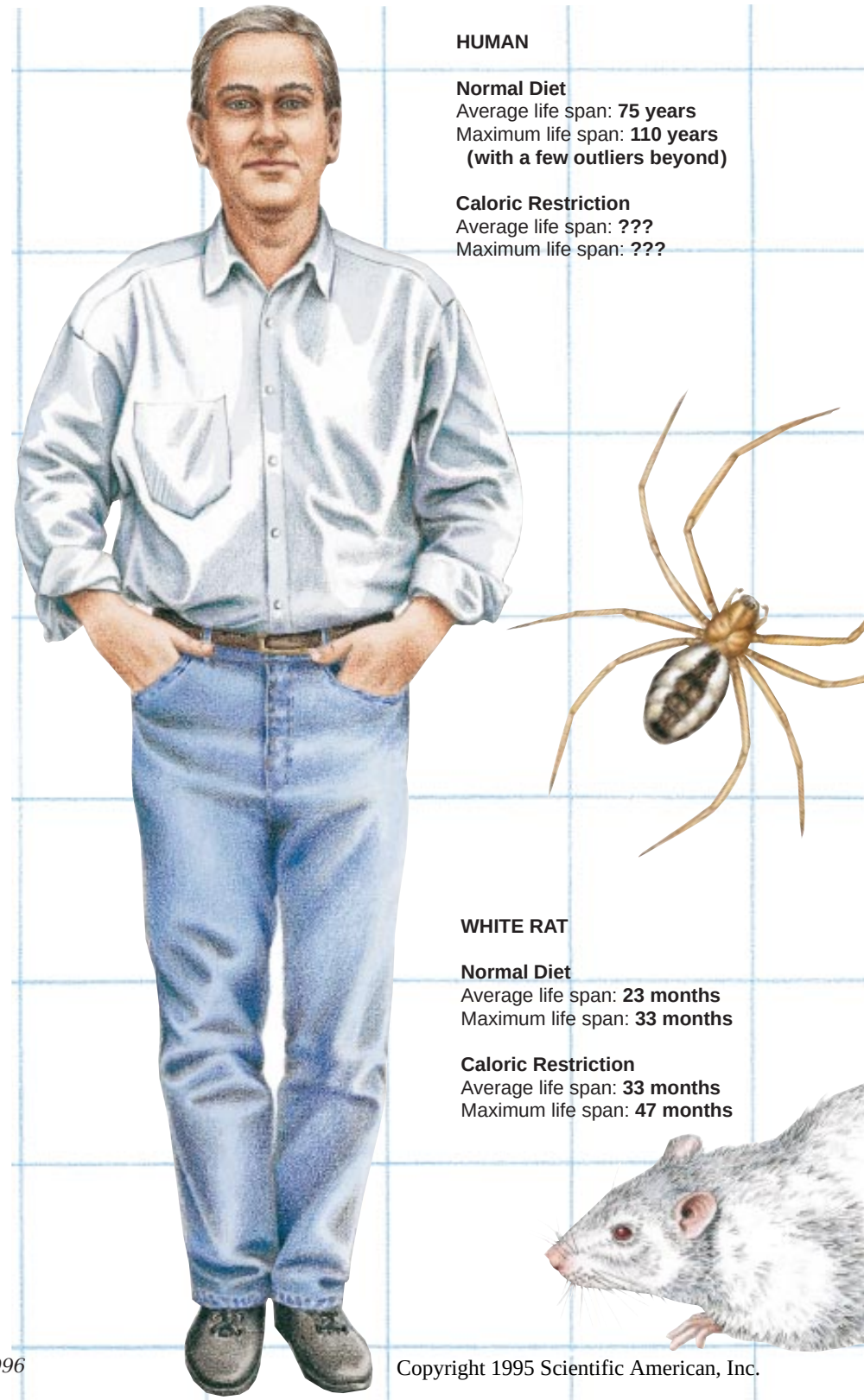
Eat less, but be sure to have enough protein, fat, vitamins and minerals. This prescription does wonders for the health and longevity of rodents. Might it help humans as well?

by Richard Weindruch

Sixty years ago scientists at Cornell University made an extraordinary discovery. By placing rats on a very low calorie diet, Clive M. McCay and his colleagues extended the outer limit of the animals' life span by 33 percent, from three years to four. They subsequently found that rats on low-calorie diets stayed youthful longer and suffered fewer late-life diseases than did their normally fed counterparts. Since the 1930s, caloric restriction has been the only intervention shown convincingly to slow aging in rodents (which are mammals, like us) and in creatures ranging from single-celled protozoans to roundworms, fruit flies and fish.

Naturally, the great power of the method raises the question of whether it can extend survival and good health in people. That issue is very much open, but the fact that the approach works in an array of organisms suggests the answer could well be yes. Some intriguing clues from monkeys and humans support the idea, too.

Of course, even if caloric austerity turns out to be a fountain of youth for humans, it might never catch on. After all, our track record for adhering to severe diets is poor. But scientists may one day develop drugs that will safely control our appetite over the long term or will mimic the beneficial influences of caloric control on the body's tissues. This last approach could enable people to consume fairly regular diets while still reaping the healthful effects of limiting their food intake. Many laboratories, including mine at the University of Wisconsin–Madison, are working to understand the cellular and molecular ba-



LIFE HAS BEEN EXTENDED, often substantially, by very low calorie diets in a range of animals, some of which are depicted here. Whether caloric restriction will increase survival in people remains to be seen. Such diets are successful only if the animals receive an adequate supply of nutrients.

sis of how caloric restriction retards aging in animals. Our efforts may yield useful alternatives to strict dieting, although at the moment most of us are focused primarily on understanding the aging process (or processes) itself.

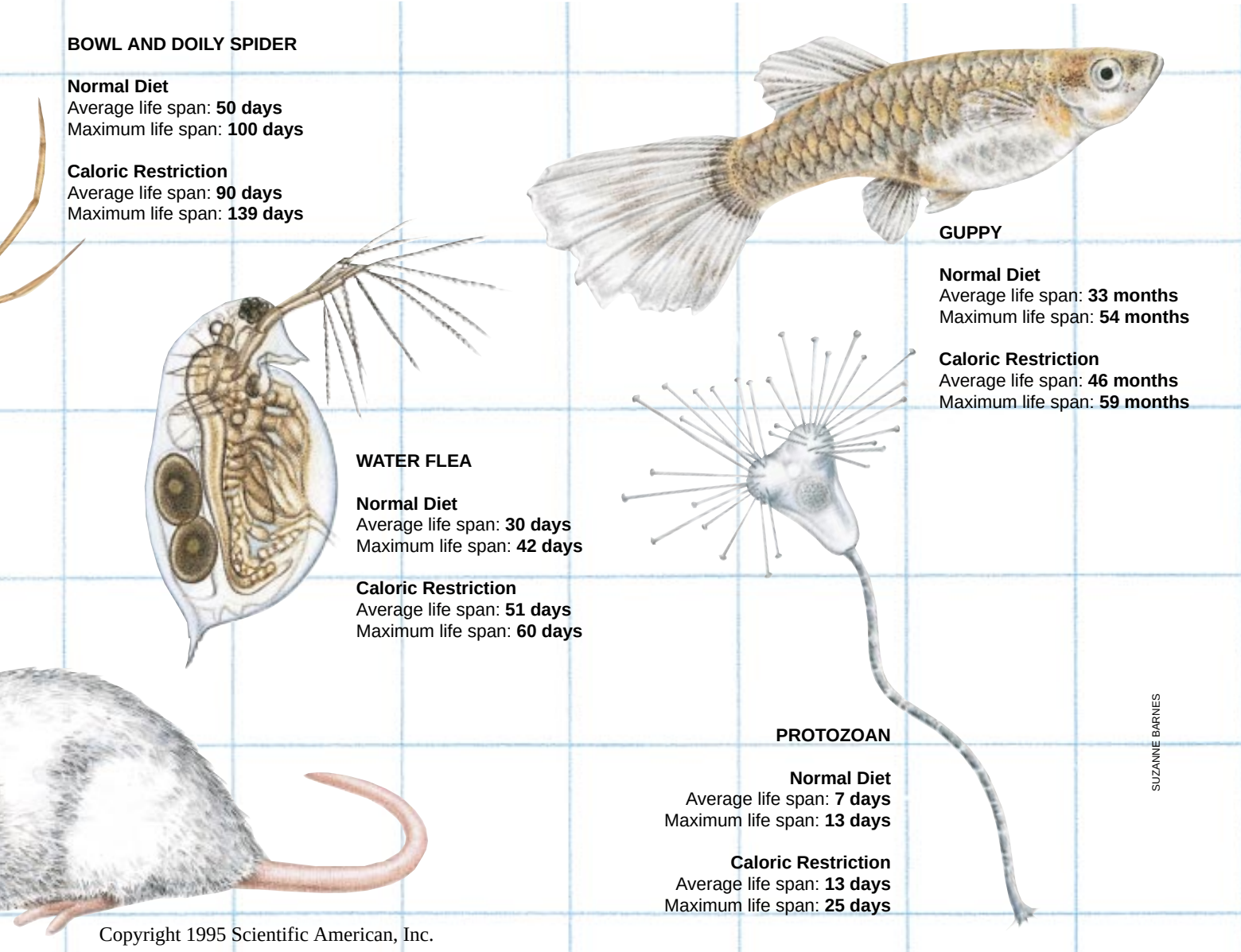
Less Is More for Rodents

Research into caloric restriction has now uncovered an astonishing range of benefits in animals—provided that the nutrient needs of the dieters are guarded carefully. In most studies the test animals, usually mice or rats, consume 30 to 50 percent fewer calories than are ingested by control subjects, and they weigh 30 to 50 percent less as well. At the same time, they receive enough protein, fat, vitamins and minerals to maintain efficient operation of their tissues. In other words, the animals follow an exaggerated form of a prudent diet, in which they consume minimal calories without becoming malnourished.

If the nutrient needs of the animals are protected, caloric restriction will consistently increase not only the average life span of a population but also the maximum span—that is, the lifetime of the longest-surviving members of the group. This last outcome means that caloric restriction tinkers with some basic aging process. Anything that forestalls premature death, such as is caused by a preventable or treatable disease or by an accident, will increase the average life span of a population. But one must truly slow the rate of aging in order for the hardiest individuals to surpass the existing maximum.

Beyond altering survival, low-calorie diets in rodents have postponed most major diseases that are common late in life [*see box on next page*], including cancers of the breast, prostate, immune system and gastrointestinal tract. Moreover, of the 300 or so measures of aging that have been studied, some 90 percent stay "younger" longer in calorie-restricted rodents than in well-fed ones. For example, certain immune responses decrease in normal mice at one year of age (middle age) but do not decline in slimmer but genetically identical mice until age two. Similarly, as rodents grow older they generally clear glucose, a simple sugar, from their blood less efficiently than they did in youth (a change that can progress to diabetes); they also synthesize needed proteins more slowly, undergo increased cross-linking (and thus stiffening) of long-lived proteins in tissues, lose muscle mass and learn less rapidly. In calorie-restricted animals, all these changes are delayed.

Not surprisingly, investigators have wondered whether caloric (energy) restriction per se is responsible for the advantages reaped from low-calorie diets or whether limiting fat or some other component of the diet accounts for the success. It turns out the first possibility is correct. Restriction of fat, protein or carbohydrate without caloric reduction does not increase the maximum life span of rodents. Supplementation



alone with multivitamins or high doses of antioxidants does not work, and neither does variation in the type of dietary fat, carbohydrate or protein.

The studies also suggest, heartteningly, that caloric restriction can be useful even if it is not started until middle age. Indeed, the most exciting discovery of my career has been that caloric restriction initiated in mice at early middle age can extend the maximum life span by 10 to 20 percent and can oppose the development of cancer. Further, although limiting the caloric intake to about half of that consumed by free-feeding animals increases the maximum life span the most, less severe restriction, whether begun early in life or later, also provides some benefit.

Naturally, scientists would be more confident that diet restriction could routinely postpone aging in men and women if the results in rodents could be confirmed in studies of monkeys (which more closely resemble people) or in members of our own species. To be most informative, such investigations would have to follow subjects for many years—an expensive and logistically difficult undertaking. Nevertheless, two major trials of monkeys are in progress.

Lean, but Striking, Primate Data

It is too early to tell whether low-calorie diets will prolong life or youthfulness in the monkeys over time. The projects have, however, been able to measure the effects of caloric restriction on so-called biomarkers of aging: attributes that generally change with age and may help predict the future span of health or life. For example, as primates grow older, their blood pressure and their blood levels of both insulin and glucose rise; at the same time, insulin sensitivity (the ability of cells to take up glucose in response to signals from insulin) declines. Postponement of these changes would imply that the experimental diet was probably slowing at least some aspects of aging.

One of the monkey studies, led by George S. Roth of the National Institute on Aging, began in 1987. It is examining rhesus monkeys, which typically live to about 30 years and sometimes reach 40 years, and squirrel monkeys, which rarely survive beyond 20 years.

Benefits of Caloric Restriction

Since 1900, advances in health practices have greatly increased the average life span of Americans (*inset in a*), mainly by improving prevention and treatment of diseases that end life prematurely. But those interventions have not substantially affected the maximum life span (*far right in a*), which is thought to be determined by intrinsic aging processes. (The curves and the data in the inset show projections for people born in the years indicated and assume conditions influencing survival do not change.) Caloric restriction, in contrast, has markedly increased the maximum as well as the average life span in rodents (*b*) and is, in fact, the only intervention so far shown to slow aging in mammals—a sign that aging in humans might be retarded as well.

Although severe diets extend survival more than moderate ones, a study of mice fed a reduced-calorie diet from early in life (three weeks of age) demonstrates that even mild restriction offers some benefit (*c*). This finding is potentially good news for people. Also encouraging is the discovery that caloric restriction in rodents does more than prolong life; it enables animals to remain youthful longer (*table*). The calorie-restricted mouse in the corner lived unusually long; most normally fed mice of her ilk die by 40 months. She was 53 months old when this photograph was taken and died of unknown causes about a month later.

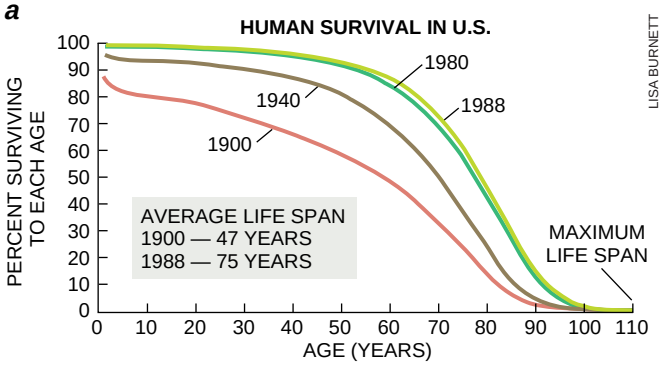
RESTRICTION IN RODENTS: SELECTED EFFECTS

Postpones age-related declines in:
Blood glucose control; female reproductive capacity; DNA repair; immunity; learning ability; muscle mass; protein synthesis

Slows age-related increases in:
Cross-linking of long-lived proteins; free-radical production by mitochondria; unrepaired oxidative damage to tissues

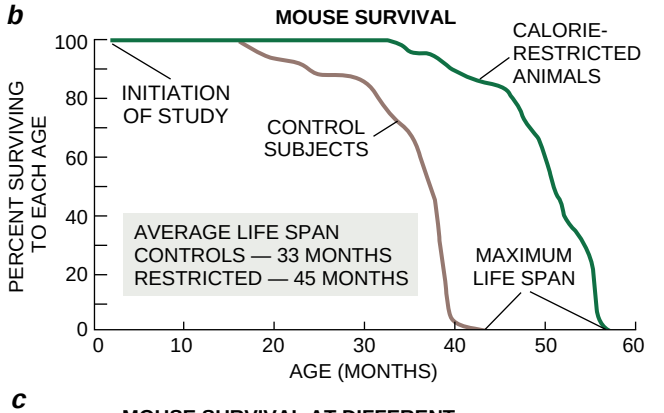
Delays onset of late-life diseases, including:
Autoimmune disorders; cancers; cataracts; diabetes; hypertension; kidney failure

a

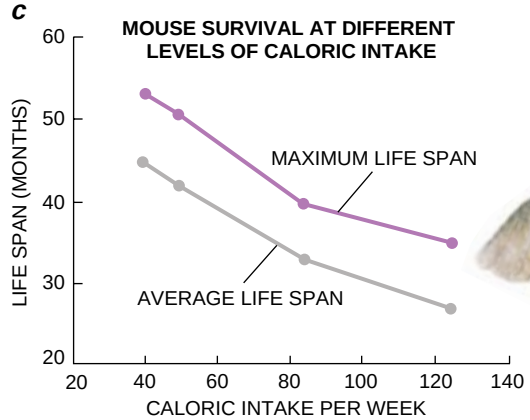


SOURCE: U.S. Bureau of the Census; National Center for Health Statistics

b



c



Some animals began diet restriction in youth (at one to two years), others after reaching puberty. The second project, involving only rhesus monkeys, was initiated in 1989 by William B. Ershler, Joseph W. Kemnitz and Ellen B. Roecker of the University of Wisconsin–Madison; I joined the team a year later. Our monkeys began caloric restriction as young adults, at eight to 14 years old. Both studies enforce a level of caloric restriction that is about 30 percent below the intake of normally fed controls.

So far the preliminary results are encouraging. The dieting animals in both projects seem healthy and happy, albeit eager for their meals, and their bodies seem to be responding to the regimen much as those of rodents do. Blood pressure and glucose levels are lower than in control animals, and insulin sensitivity is greater. The levels of insulin in the blood are lower as well.

No one has yet performed carefully controlled studies of long-term caloric restriction in average-weight humans over time. And data from populations forced by poverty to live on relatively few calories are uninformative, because such groups generally cannot attain adequate amounts of essential nutrients. Still, some human studies offer indirect evidence that caloric restriction could be of value. Consider the people of Okinawa, many of whom consume diets that are low in calories but provide needed nutrients. The incidence of centenarians there is high—up to 40 times greater than that of any other Japanese island. In addition, epidemiological surveys in the U.S. and elsewhere indicate that certain cancers, notably those of the breast, colon and stomach, occur less frequently in people reporting small caloric intakes.

Intriguing results were also obtained after eight people living in a self-contained environment—Biosphere 2, near Tucson, Ariz.—were forced to curtail their food intake sharply for two years because of poorer than expected yields from their food-producing efforts. The scientific merits of the overall project have been questioned, but those of us interested in the effects of low-calorie diets were fortunate that Roy L. Walford of the University of California at Los Angeles, who is an expert on caloric restriction and aging (and was my scientific mentor), was the team's physician. Walford helped his colleagues avoid malnutrition and monitored various aspects of the group's physiology. His analyses reveal that caloric restriction led to lowered blood pressure and glucose levels—just as it does in rodents and monkeys. Total serum cholesterol declined as well.

RICHARD WEINDRUCH

MICE ARE THE SAME AGE—40 months. Yet compared with the normally fed animal at the right, the one at the left, which has been reared on a low-calorie diet since 12 months of age (early middle age), looks younger and is healthier.

The results in monkeys and humans may be preliminary, but the rodent data show unequivocally that caloric restriction can exert a variety of beneficial effects. This variety raises something of a problem for researchers: Which of the many documented changes (if any) contribute most to increased longevity and youthfulness? Scientists have not yet reached a consensus, but they have ruled out a few once viable proposals. For instance, it is known that a low intake of energy retards growth and also shrinks the amount of fat in the body. Both these effects were once prime contenders as the main changes that lead to longevity but have now been discounted.

Several other hypotheses remain under consideration, however, and all of them have at least some experimental support. One such hypothesis holds that caloric restriction slows the rate of cell division in many tissues. Because the uncontrolled proliferation of cells is a hallmark of cancer, that change could potentially explain why the incidence of several late-life cancers is reduced in animals fed low-calorie diets. Another proposal is based on the finding that caloric restriction tends to lower glucose levels. Less glucose in the circulation would slow the accumulation of sugar on long-lived proteins and would thus moderate the disruptive effects of this buildup.

A Radical Explanation

The view that has so far garnered the most convincing support, though, holds that caloric restriction extends survival and vitality primarily by limiting injury of mitochondria by free radicals. Mitochondria are the tiny intracellular structures that serve as the power plants of cells. Free radicals are highly reactive molecules (usually derived from oxygen) that carry an unpaired electron at their surface. Molecules in this state are prone to destructively oxidizing, or snatching electrons from, any compound they encounter. Free radicals have been suspected of contributing to aging since the 1950s, when Denham Harman of the University of Nebraska Medical School suggested that their generation in the course of normal metabolism gradually disrupts cells. But it was not until the 1980s that scientists began to realize that mitochondria were probably the targets hit hardest.

The mitochondrial free-radical hypothesis of aging derives in part from an understanding of how mitochondria produce ATP (adenosine triphosphate)—the molecule that provides the energy for most cellular processes, such as pumping ions across cell membranes, contracting muscle fibers and constructing proteins. ATP synthesis occurs by a very complicated sequence of reactions, but essentially it involves activity by a series of molecular complexes embedded in an internal membrane—the inner membrane—of mitochondria. With help from oxygen, the complexes extract energy from nutrients and use that energy to manufacture ATP.

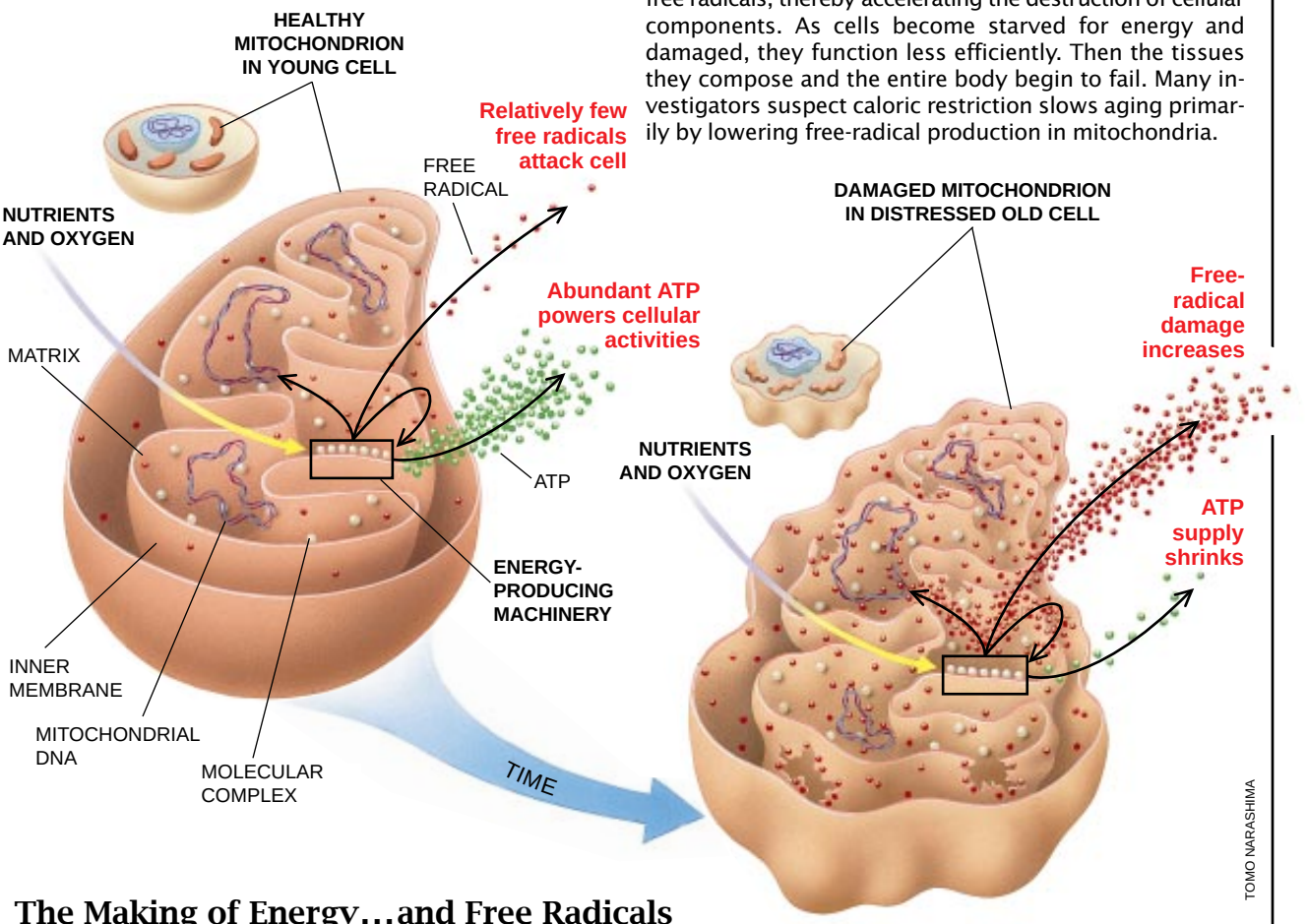
Unfortunately, the mitochondrial machinery that draws energy from nutrients also produces free radicals as a by-product. Indeed, mitochondria are thought to be responsible for creating most of the free radicals in cells. One such by-product is the superoxide radical ($O_2^{\cdot -}$). (The dot in the formula represents the unpaired electron.) This renegade is destructive in its own right but can also be converted into hydrogen peroxide (H_2O_2), which technically is not a free radical but can readily form the extremely aggressive hydroxyl free radical ($OH^{\cdot -}$).

Once formed, free radicals can dam-

A Theory of Aging

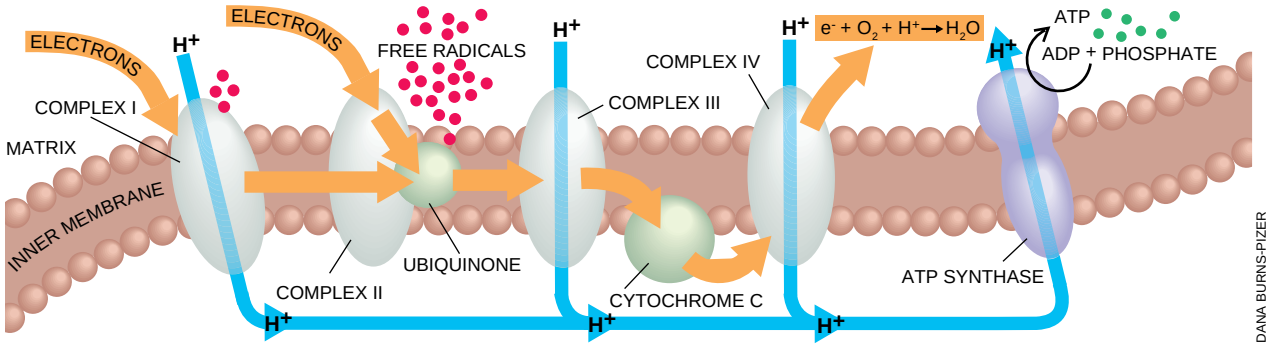
A leading explanation for why we age places much of the blame on destructive free radicals (*red*) generated in mitochondria, the cell's energy factories. The radicals form (*left*) when the energy-producing machinery in mitochondria (*boxed in black*) uses oxygen and nutrients to synthesize ATP (adenosine triphosphate)—the molecule (*green*) that powers most other activities in cells. Those radicals attack, and may permanently injure, the machinery itself and the mitochondrial DNA that is needed to construct parts of it. They can also harm other components of mitochondria and cells.

The theory suggests that over time (*right*) the accumulated damage to mitochondria precipitates a decline in ATP production. It also engenders increased production of free radicals, thereby accelerating the destruction of cellular components. As cells become starved for energy and damaged, they function less efficiently. Then the tissues they compose and the entire body begin to fail. Many investigators suspect caloric restriction slows aging primarily by lowering free-radical production in mitochondria.



The Making of Energy…and Free Radicals

The energy-producing machinery in mitochondria consists mainly of the electron-transport chain: a series of four large (*gray*) and two smaller (*light green*) molecular complexes. Complexes I and II (*far left*) take up electrons (*gold arrows*) from food and relay them to ubiquinone, the site of greatest free-radical (*red*) generation. Ubiquinone sends the electrons down the rest of the chain to complex IV, where they interact with oxygen and hydrogen to form water. The electron flow induces protons (H+) to stream (*blue arrows*) to yet another complex—ATP synthase (*purple*)—which draws on energy supplied by the protons to manufacture ATP (*dark green*). Free radicals form when electrons escape from the transport chain and combine with oxygen in their vicinity.



age proteins, lipids (fats) and DNA anywhere in the cell. But the components of mitochondria—including the ATP-synthesizing machinery and the mitochondrial DNA that gives rise to some of that machinery—are believed to be most vulnerable. Presumably they are at risk in part because they reside at or near the "ground zero" site of free-radical generation and so are constantly bombarded by the oxidizing agents. Moreover, mitochondrial DNA lacks the protein shield that helps to protect nuclear DNA from destructive agents. Consistent with this view is that mitochondrial DNA suffers much more oxidative damage than does nuclear DNA drawn from the same tissue.

Proponents of the mitochondrial free-radical hypothesis of aging suggest that damage to mitochondria by free radicals eventually interferes with the efficiency of ATP production and increases the output of free radicals. The rise in free radicals, in turn, accelerates the oxidative injury of mitochondrial components, which inhibits ATP production even more. At the same time, free radicals attack cellular components outside the mitochondria, further impairing cell functioning. As cells become less efficient, so do the tissues and organs they compose, and the body itself becomes less able to cope with challenges to its stability. The body does try to counteract the noxious effects of the oxidizing agents. Cells possess antioxidant enzymes that detoxify free radicals, and they make other enzymes that repair oxidative damage. Neither of these systems is 100 percent effective, though, and so such injury is likely to accumulate over time.

Experimental Support

The proposal that aging stems to a great extent from free-radical-induced damage to mitochondria and other cellular components has recently been buttressed by a number of findings. In one striking example, Rajindar S. Sohal, William C. Orr and their colleagues at Southern Methodist University in Dallas investigated rodents and several other organisms, including fruit flies, houseflies, pigs and cows. They noted increases with age in free-radical generation by mitochondria and in oxidative changes to the inner mitochondrial membrane (where ATP is synthesized) and to mitochondrial proteins and DNA. They also observed that greater rates of free-radical production correlate with shortened average and maximum life spans in several of the species.

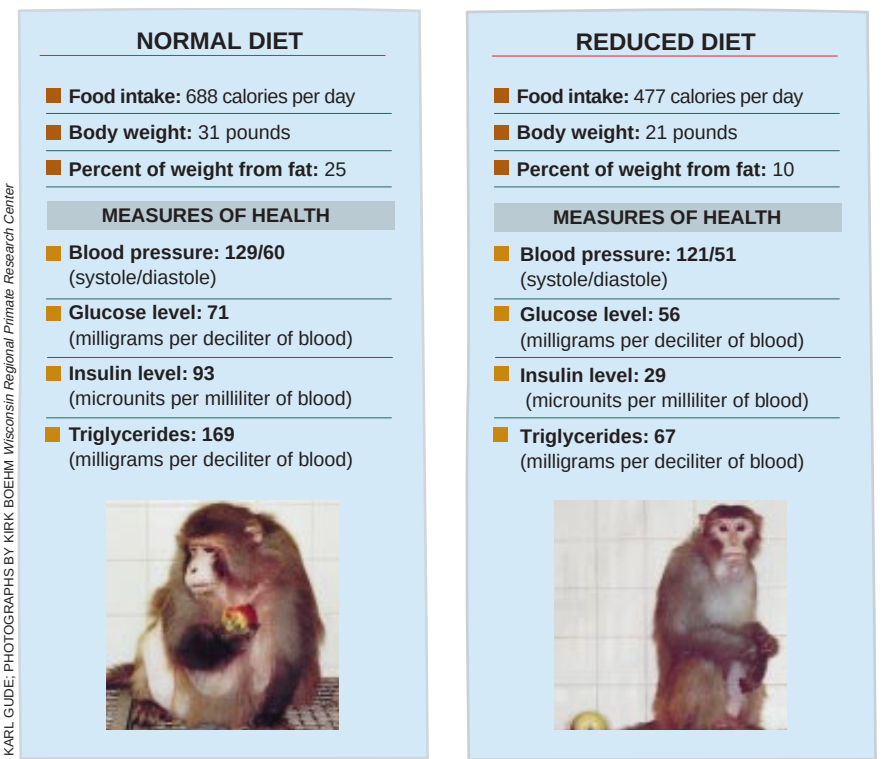
It turns out, too, that ATP manufacture decreases with age in the brain, heart and skeletal muscle, as would be expected if mitochondrial proteins and DNA in those tissues were irreparably impaired by free radicals. Similar decreases also occur in human tissues and may help explain why degenerative diseases of the nervous system and heart are common late in life and why muscles lose mass and weaken.

Some of the strongest support for the proposition that caloric restriction retards aging by slowing oxidative injury of mitochondria comes from Sohal's group. When the workers looked at mitochondria harvested from the brain, heart and kidney of mice, they discovered that the levels of the superoxide radical and of hydrogen peroxide were markedly lower in animals subjected to long-term caloric restriction than in normally fed controls. In addition, a significant increase of free-radical production with age seen in the control groups was blunted by caloric restriction in the experimental group. This blunted increase was, moreover, accompanied by lessened amounts of oxidative insult to mitochondrial proteins and DNA. Other work indicates that caloric restriction helps to prevent age-related changes in the activities of some antioxidant enzymes—although many investigators, including me, suspect that strict dieting ameliorates oxidative damage mainly through slowing free-radical production.

By what mechanism might caloric restriction reduce the generation of free radicals? No one yet knows. One proposal holds that a lowered intake of calories may somehow lead to slower consumption of oxygen by mitochondria—either overall or in selected cell types. Alternatively, low-calorie diets may increase the efficiency with which mitochondria use oxygen, so that fewer free radicals are made per unit of oxygen consumed. Less use of oxygen or more efficient use would presumably result in the formation of fewer free radicals. Recent findings also intimate that caloric control may minimize free-radical generation in mitochondria by reducing levels of a circulating thyroid hormone known as triiodothyronine, or T_3, through unknown mechanisms.

Applications to Humans?

Until research into primates has progressed further, few scientists would be prepared to recommend that large numbers of people embark on a severe caloric-restriction regimen. Nevertheless, the accumulated findings do offer



KARL GUDE; PHOTOGRAPHS BY KIRK BOEHM *Wisconsin Regional Primate Research Center*

RESULTS FROM ONGOING TRIAL of caloric restriction in rhesus monkeys cannot yet reveal whether limiting calories will prolong survival. But comparison of a control group (*left*) with animals on a strict diet (*right*) after five years indicates that at least some biological measures that typically rise with age are changing more slowly in the test animals. Blood pressure is only slightly lower in the restricted group now, but has been markedly lower for much of the study period.

some concrete lessons for those who wonder how such programs might be implemented in humans.

One implication is that sharp curtailment of food intake would probably be detrimental to children, considering that it retards growth in young rodents. Also, because children cannot tolerate starvation as well as adults can, they would presumably be more susceptible to any as yet unrecognized negative effects of a low-calorie diet (even though caloric restriction is not equivalent to starvation). An onset at about 20 years of age in humans should avoid such drawbacks and would probably provide the greatest extension of life.

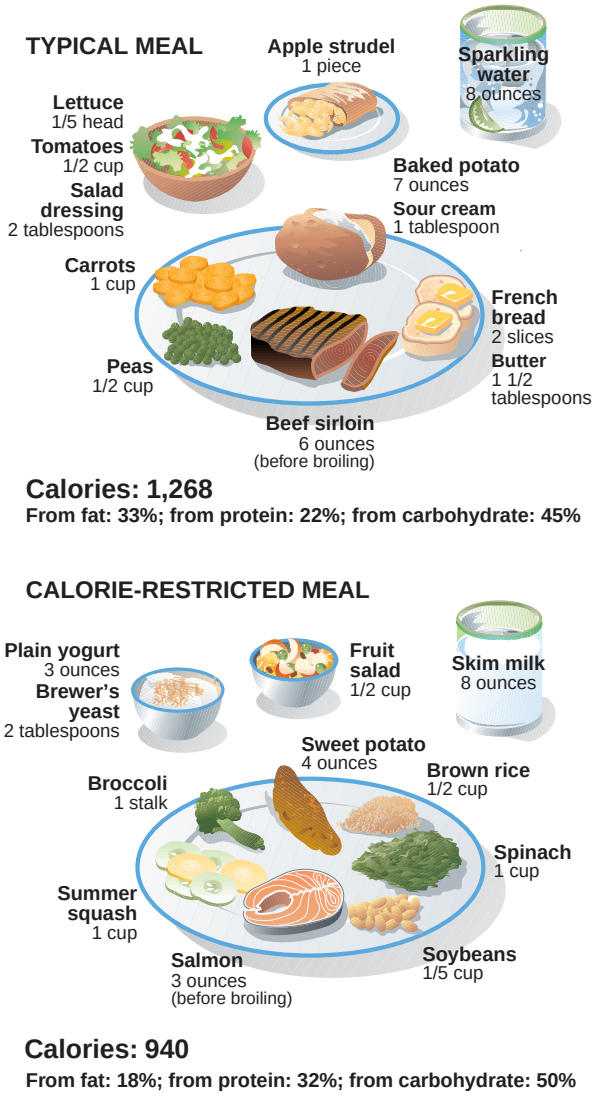
The speed with which calories are reduced needs to be considered, too. Early researchers were unable to prolong survival of rats when diet control was instituted in adulthood. I suspect the failure arose because the animals were put on the regimen too suddenly or were given too few calories, or both. Working with year-old mice, my colleagues and I have found that a gradual tapering of calories to about 65 percent of normal did increase survival.

How might one determine the appropriate caloric intake for a human being? Extrapolating from rodents is difficult, but some findings imply that many people would do best by consuming an amount that enabled them to weigh 10 to 25 percent less than their personal set point. The set point is essentially the weight the body is "programmed" to maintain, if one does not eat in response to external cues, such as television commercials. The problem with this guideline is that determining an individual's set point is tricky. Instead of trying to identify their set point, dieters (with assistance from their health advisers) might engage in some trial and error to find the caloric level that reduces the blood glucose or cholesterol level, or some other measures of health, by a predetermined amount.

The research in animals further implies that a reasonable caloric-restriction regimen for humans might involve a daily intake of roughly one gram (0.04 ounce) of protein and no more than about half a gram of fat for each kilogram (2.2 pounds) of current body weight. The diet would also include enough complex carbohydrate (the long chains of sugars abundant in fruits and vegetables) to reach the desired level of calories. To attain the standard recommended daily allowances for all essential nutrients, an individual would have to select foods with extreme care and probably take vitamins or other supplements.

Anyone who contemplated following a caloric-restriction regimen would also have to consider potential disadvantages beyond hunger pangs and would certainly want to undertake the program with the guidance of a physician. Depending on the severity of the diet, the weight loss that inevitably results might impede fertility in females. Also, a prolonged anovulatory state, if accompanied by a diminution of estrogen production, might increase the risk of osteoporosis (bone loss) and loss of muscle mass later in life. It is also possible that caloric restriction will compromise a person's ability to withstand stress, such as injury, infection or exposure to extreme temperatures. Oddly enough, stress resistance has been little studied in rodents on low-calorie diets, and so they have little to teach about this issue.

It may take another 10 or 20 years before scientists have a firm idea of whether caloric restriction can be as beneficial for humans as it clearly is for rats, mice and a variety of other creatures. Meanwhile investigators studying this intervention are sure to learn much about the nature of aging and to gain ideas about how to slow it—whether through caloric restriction, through drugs that reproduce the effects of dieting or by methods awaiting discovery.



KARL GUDE; SOURCE: ROY L. WALFORD *University of California, Los Angeles*

DINNER of a person on a roughly 2,000-calorie diet (*top*) might be reduced considerably—by about a third of the calories (*bottom*)—for someone on a caloric-restriction regimen. To avoid malnutrition, people on such programs would choose nutrient-dense foods such as those shown.

The Author

RICHARD WEINDRUCH, who earned his Ph.D. in experimental pathology at the University of California, Los Angeles, is associate professor of medicine at the University of Wisconsin–Madison, associate director of the university's Institute on Aging and a researcher at the Veterans Administration Geriatric Research, Education and Clinical Center in Madison. He has devoted his career to the study of caloric restriction and its effects on the body and practices mild restriction himself. He has not, however, attempted to put his family or his two cats on the regimen.

Further Reading

THE RETARDATION OF AGING AND DISEASE BY DIETARY RESTRICTION. Richard Weindruch and Roy L. Walford. Charles C. Thomas, 1988.

FREE RADICALS IN AGING. Edited by Byung P. Yu. CRC Press, 1993.

MODULATION OF AGING PROCESSES BY DIETARY RESTRICTION. Edited by Byung P. Yu. CRC Press, 1994.

Technology and Economics in the Semiconductor Industry

Although the days of runaway growth may be numbered, their passing may force chipmakers to offer more variety

by G. Dan Hutcheson and Jerry D. Hutcheson

The ability to store and process information in new ways has been essential to humankind's progress. From early Sumerian clay tokens through the Gutenberg printing press, the Dewey decimal system and, eventually, the semiconductor, information storage has been the catalyst for increasingly complex legal, political and societal systems. Modern science, too, is inextricably bound to information processing, with which it exists in a form of symbiosis. Scientific advances have enabled the storage, retrieval and processing of ever more information, which has, in turn, helped generate the insights needed for further advances.

Over the past few decades, semiconductor electronics has become the driving force in this crucial endeavor, ushering in a remarkable epoch. Integrated circuits made possible the personal computers that have transformed the world of business, as well as the controls that make engines and machines run more cleanly and efficiently and the medical systems that save lives. In so doing, they spawned industries that are able to generate hundreds of billions of dollars in revenues and provide jobs for millions of people. All these benefits, and far too many more to list here, accrue in no small measure from the fact that the semiconductor industry has been able to integrate more and more transistors onto chips, at ever lower costs.

This ability, largely unprecedented in industrial history, is so fundamental in the semiconductor business that it is literally regarded as a law. Nevertheless, from time to time, fears that technical and economic obstacles might soon slow the pace of advances in semiconductor technology have cropped up. Groups of scientists and engineers have often predicted the imminence of so-called showstopping problems, only to see those predictions foiled by the creativity and ingenuity of their peers.

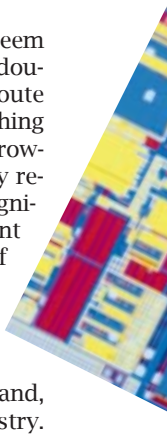
To paraphrase a former U.S. president, here we go again. With the cost of building a new semiconductor facility now into 10 figures, and with the densities of transistors close to the theoretical limits for the technologies being used, an unsettling question is once more being asked in some quarters. What will happen to the industry when it finally must confront technical barriers that are truly impassable?

Moore and More Transistors

In 1964, six years after the integrated circuit was invented, Gordon Moore observed that the number of transistors that semiconductor makers could put on a chip was doubling every year. Moore, who cofounded Intel Corporation in 1968 and is now an industry sage, correctly predicted that this pace would continue into at least the near future. The phenomenon became known as Moore's Law, and it has had far-reaching implications.

Because the doublings in density were not accompanied by an increase in cost, the expense per transistor was halved with each doubling. With twice as many transistors, a memory chip can store twice as much data. Higher levels of integration mean greater numbers of functional units can be integrated onto the chip, and more closely spaced devices, such as transistors, can interact with less delay. Thus, the advances gave users increased computing power for the same money, spurring both sales of chips and demand for yet more power.

To the amazement of many experts—including Moore himself—integration continued to increase at an astounding rate. True, in the late 1970s, the pace slowed to a doubling of transistors every 18 months. But it has held to this rate ever since, leading to commercial integrated circuits today with more than six million transistors. The electronic components in these chips measure 0.35 micron across. Chips with 10 million or more transistors measuring 0.25 or even 0.16 micron are expected to become commercially available soon.



In stark contrast to what would seem to be implied by the dependable doubling of transistor densities, the route that led to today's chips was anything but smooth. It was more like a harrowing obstacle course that repeatedly required chipmakers to overcome significant limitations in their equipment and production processes. None of those problems turned out to be the dreaded showstopper whose solution would be so costly that it would slow or even halt the pace of advances in semiconductors and, therefore, the growth of the industry. Successive roadblocks, however, have become increasingly imposing, for reasons tied to the underlying technologies of semiconductor manufacturing.

Chips are made by creating and interconnecting transistors to form complex electronic systems on a sliver of silicon. The fabrication process is based on a series of steps, called mask layers, in which films of various materials—some sensitive to light—are placed on the silicon and exposed to light. After these deposition and lithographic procedures, the layers are processed to "etch" the patterns that, when precisely aligned and combined with those on successive layers, produce the transistors and connections. Typically, 200 or more chips are fabricated simultaneously on a thin disk, or wafer, of silicon [*see illustration on page 58*].

In the first set of mask layers, insulating oxide films are deposited to make the transistors. Then a photosensitive coating, called the photoresist, is spun over these films. The photoresist is ex-

CIRCUIT LAYOUT helps designers keep track of the design for a chip. Different layers of the chip are shown in different colors. This image shows part of the layout for Motorola's forthcoming Power PC 620 microprocessor.

posed with a stepper, which is similar to an enlarger used to make photographic prints. Instead of a negative, however, the stepper uses a reticle, or mask, to project a pattern onto the photoresist. After being exposed, the photoresist is developed, which delineates the spaces, known as contact windows, where the different conducting layers interconnect. An etcher then cuts through the oxide film so that electrical contacts to transistors can be made, and the photoresist is removed.

More sets of mask layers, based on much the same deposition, lithography and etching steps, create the conducting films of metal or polysilicon needed to link transistors. All told, about 19 mask layers are required to make a chip.

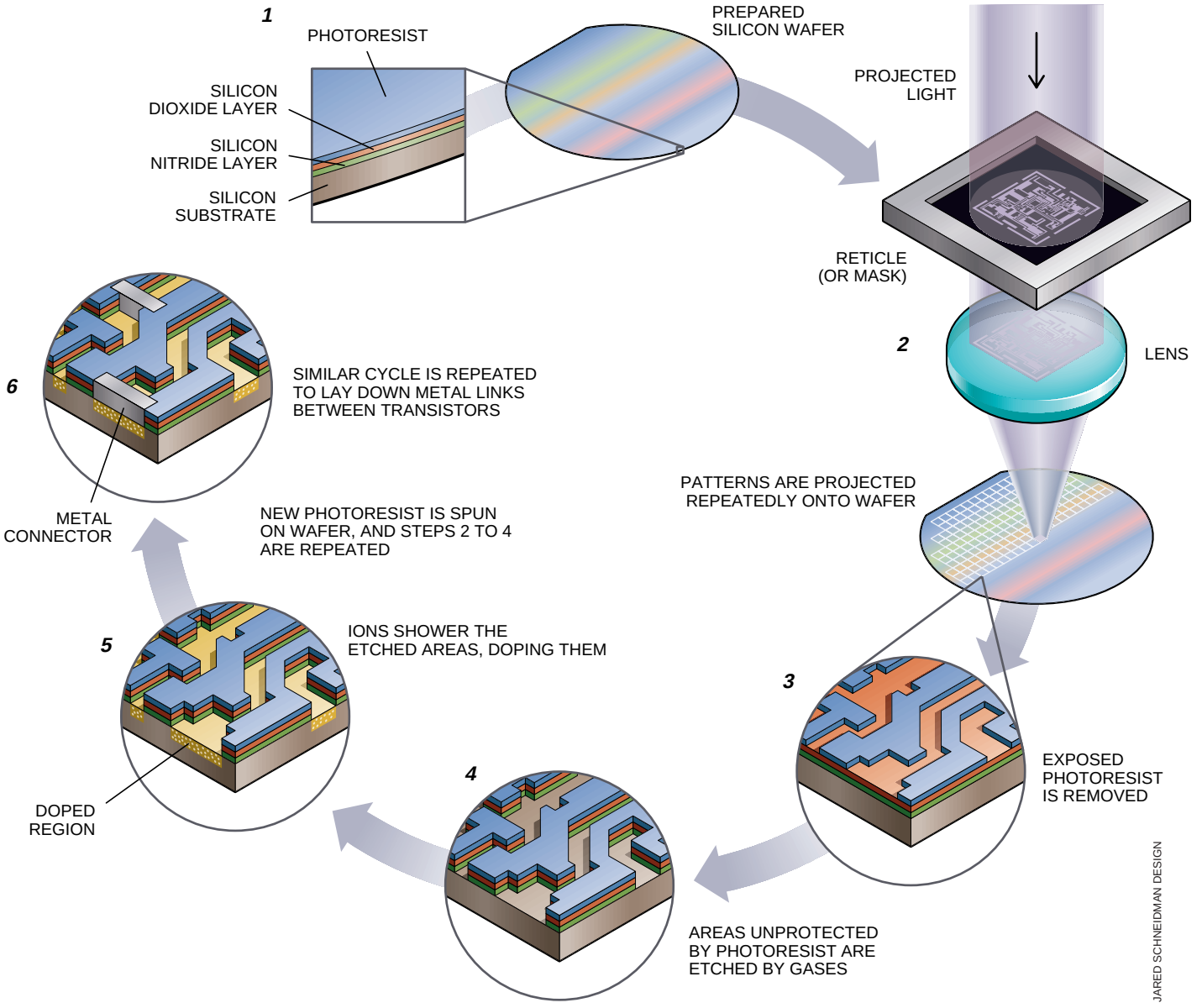
The physics underlying these manufacturing steps suggests several potential obstacles to continued technical progress. One follows from Rayleigh's resolution limit, named after John William Strutt, the third Baron of Rayleigh, who won the 1904 Nobel Prize for Physics. According to this limit, the size of the smallest features that can be resolved by an optical system with a circular aperture is proportional to the wavelength of the light source divided by the diameter of the aperture of the objective lens. In other words, the shorter the wavelengths and larger the aperture, the finer the resolution.

The limit is a cardinal law in the semiconductor industry because it can be used to determine the size of the smallest transistors that can be put on a chip. In the lithography of integrated circuits, the most commonly used light source is the mercury lamp. Its most useful line spectra for this purpose occur at 436 and 365 nanometers, the so-called mercury g and i lines. The former is visible to the human eye; the latter is just beyond visibility in the ultraviolet. The numerical apertures used range from a low of about 0.28 micron for run-of-the-mill industrial lenses to a high of about 0.65 for those in leading-edge lithography tools. These values, taken together with other considerations arising from demands of high-volume manufacturing, give a limiting resolution of about 0.54 micron for g-line lenses and about 0.48 for i-line ones.

Until the mid-1980s, it was believed that g-line operation was the practical limit. But one by one, obstacles to i-line operation were eliminated in a manner that well illustrates the complex relations between economics and technology in the industry. Technical barriers were surmounted, and, more important, others were found to be mere by-products of the level of risk the enterprise was willing to tolerate. This history is quite relevant to the situation the industry now finds itself in—close to what appear to be the practical limits of i-line operation.

Must the Show Go On?

One of the impediments to i-line operation was the fact that most of the glasses used in lenses are opaque at i-line frequencies, necessitating the use of quartz. Even if practical quartz



CHIP FABRICATION occurs as a cycle of steps carried out as many as 20 times. Many chips are made simultaneously on a silicon wafer, to which has been applied a light-sensitive coating (*1*). Each cycle starts with a different pattern, which is projected repeatedly onto the wafer (*2*). In each place where the image falls, a chip is made. The photosensitive coating is removed (*3*), and the light-exposed areas are etched by gases (*4*). These areas are then showered with ions (or "doped"), creating transistors (*5*). The transistors are then connected as successive cycles add layers of metal and insulator (*6*).

lenses could be made, it was reasoned, verifying the alignment of patterns that could not be seen would be difficult. Moreover, only about 70 percent of i-line radiation passes through the quartz; the rest is converted to heat in the lens, which can distort the image.

Nor do these represent the extent of the difficulties. Rayleigh's limit also establishes the interval within which the pattern projected by the lens is in focus. Restricted depth of focus can work against resolution limits: the better the resolution, the shallower the depth of focus. For a lens as described above, the depth of focus is about 0.52 micron for the best g-line lenses and about 0.50 for i-line ones. Such shallow depths demand extremely flat wafer surfaces—much flatter than what could be maintained across the diagonal of a large chip with the best available equipment just several years ago.

Innovative solutions overcame these limitations. Planarizing methods were developed to ensure optically flat surfaces. Fine adjustments to the edges of the patterns in the reticle were used to shift the phase of the incoming i-line radiation, permitting crisper edge definitions and therefore smaller features—in effect, circumventing Rayleigh's limit. One of the last adjustments was the simple acceptance of a lower value of the proportionality constant, which is related to the degree of contrast in the image projected onto the wafer during lithography. For i-line operation, manufacturers gritted their teeth and accepted a lower proportionality constant than was previously thought practical. Use of the lower value meant that the margins during fabrication would be lower, requiring tighter controls over processes—lithography, deposition and etching—to keep the number of acceptable chips per wafer (the yield) high. As a result of these innovations, i-line steppers are now routinely used to expose 0.35-micron features.

In this last instance, what was really at issue was the loss in contrast ratio that a company was willing to tolerate. With perfect contrast, the image that is created on the photoresist is sharp. Like so many of the limitations in the industry, contrast ratio was perceived to be a

technical barrier, but it was actually a risk decision. Lower contrast ratios did not lower yields, it was found, if there were tighter controls elsewhere in the process.

It has been difficult to predict when—or if—this stream of creative improvements will dry up. Nevertheless, as the stream becomes a trickle, the economic consequences of approaching technical barriers will be felt before the barriers themselves are reached. For example, the costs of achieving higher levels of chip performance rise very rapidly as the limits of a manufacturing technology are approached and then surpassed. Increasing costs may drive prices beyond what buyers are willing to pay, causing the market to stagnate before the actual barriers are encountered.

Eventually, though, as a new manufacturing technology takes hold, the costs of fabricating chips begin to decline. At this point, the industry has jumped from a cost-performance curve associated with the old technology to a new curve for the new process. In effect, the breakthrough from one manufacturing technology to another forces the cost curve to bend downward, pushing technical limits farther out [*see illustration at right*]. When this happens, higher levels of performance are obtainable without an increase in cost, prompting buyers to replace older equipment. This is important in the electronics industry, because products seldom wear out before becoming obsolete.

The principles outlined so far apply to all kinds of chips, but memory is the highest-volume business and is in some ways the most significant. From about $550,000 25 years ago, the price of a megabyte of semiconductor memory has declined to just $38 today. But over the same period, the cost of building a factory to manufacture such memory chips has risen from less than $4 million to a little more than $1.2 billion, putting the business beyond the reach of all but a few very large firms. Such skyrocketing costs, propelled mainly by the expense of having to achieve ever more imposing technical breakthroughs, have once again focused attention on limits in the semiconductor industry.

Breakthroughs Needed

The semiconductor industry is not likely to come screeching to a halt anytime soon. But the barriers now being approached are so high that getting beyond them will probably cause more far-reaching changes than did previous cycles of this kind. To understand why requires outlining some details about the obstacles themselves.

Most have to do with the thin-film structures composing the integrated circuit or with the light sources needed to make the extremely thin conducting lines or with the line widths themselves. Two examples concern the dielectric constant of the insulating thin films. The dielectric constant is an electrical property that indicates, among other things, the ability of an insulating film to keep signals from straying between the narrowly spaced conducting lines on a chip. Yet as more transistors are integrated onto a chip, these films are packed closer together, and cross-talk between signal lines becomes worse.

One possible solution is to reduce the value of the dielectric constant, making the insulator more impermeable to cross-talk. This, in turn, initiates a twofold search, one for new materials with lower dielectric constants, the other for new film structures that can reduce further the overall dielectric constant. Some engineers are even looking for ways to riddle the insulating film with small voids, to take advantage of the very low dielectric constant of air or a vacuum.

Elsewhere on the chip, materials with the opposite property—a high dielectric constant—are needed. Most integrated circuits require capacitors. In a semiconductor dynamic random-access memory (DRAM), for instance, each bit is actually stored in a capacitor, a device capable of retaining an electrical charge. (A charged capacitor represents binary 1, and an uncharged capacitor is 0.) Typically, the amount of capacitance available on a chip is never enough. Capacitance is proportional to the dielectric constant, so DRAMs and similar chips need materials of a high dielectric constant.

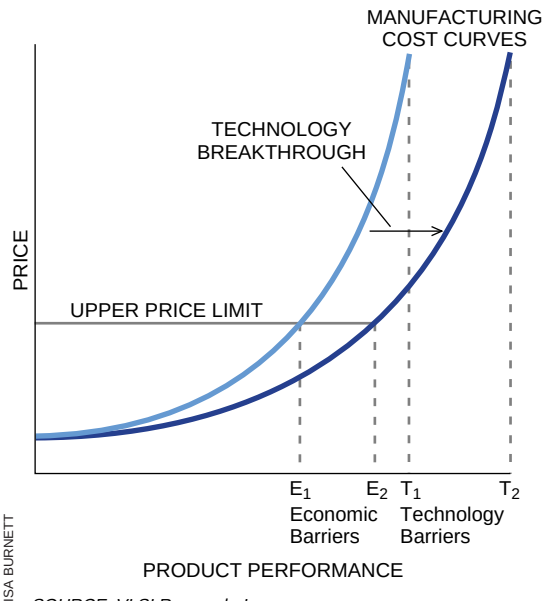
The quest for more advanced light sources for lithography is also daunting. Finer resolution demands shorter wavelengths. But the most popular mercury light sources in use today emit very little energy at wavelengths shorter than the i line's 365 nanometers. Excimer lasers are useful down to about 193 nanometers but generate little energy below that wavelength. In recent years, excimer-laser lithography has been used to fabricate some special-purpose, high-performance chips in small batches. For still shorter wavelengths, x-ray sources are the last resort. Nevertheless, 20 years of research on x-ray lithography has produced only modest results. No commercially available chips have been made with x-rays.

Billion-Dollar Factories

Economic barriers also rise with increasing technical hurdles and usually make themselves evident in the form of higher costs for equipment, particularly for lithography. Advances in lithography equipment are especially important because they determine the smallest features that can be created on chips. Although the size of these smallest possible features has shrunk at roughly 14 percent annually since the earliest days of the industry, the price of lithography equipment has risen at 28 percent a year.

In the early days, each new generation of lithography equipment cost 10 times as much as the previous one did. Since then, the intergenerational development of stepping aligners has reduced these steep price increases to a mere doubling of price with each new significant development. The price of other kinds of semiconductor-fabrica-

PRICE VERSUS PERFORMANCE



LISA BURNETT

SOURCE: VLSI Research, Inc.

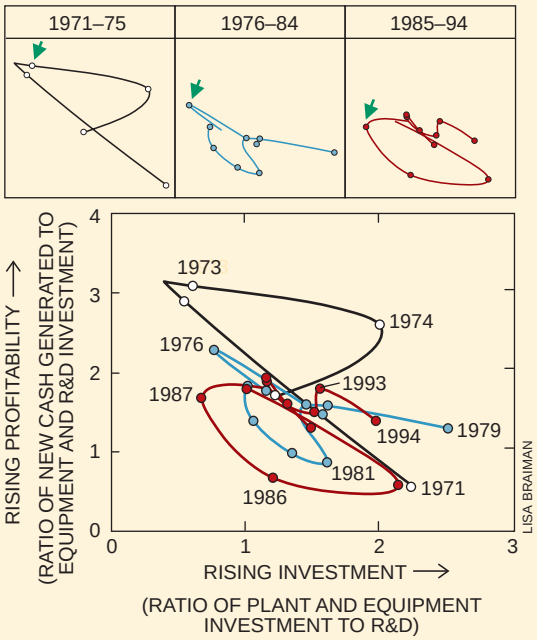
COST CURVE is associated with each chip-manufacturing system. Technology barriers, T_1 and T_2, are where minute increases in chip performance can be achieved only at a huge cost. Economic barriers are encountered well before the technological ones, however. These occur where the line representing the maximum price customers are willing to pay intersects with the curves (at E_1 and E_2). Technology breakthroughs have the effect of bending the curve downward, to the position of the darker plot. When this happens, performance improves, shifting the barriers to E_2 and T_2.

How Much Bang for the Buck?

For about 60 years, almost all industrial companies have used basically the same model to keep track of financial returns from their investments in equipment, research, marketing and all other categories. Developed just before World War I by Donaldson Brown of Du Pont, the model was brought into the business mainstream by General Motors during its effort to surpass Ford Motor Company as the dominant maker of automobiles.

Since its universal adaptation, this return-on-investment (ROI) model has held up well in industries in which the rates of growth and technological advance are relatively small. To our knowledge, however, the model has never been shown to work well in a sector such as the semiconductor industry, in which many key rates of change—of product performance and the cost of manufacturing equipment, to name just two—are in fact nonlinear. From an economic viewpoint, it is this nonlinearity that makes the semiconductor industry essentially unlike all other large industries and therefore renders all other business models unsuitable.

In the semiconductor industry, relatively large infusions of capital must be periodically bestowed on equipment and research, with each infusion exponentially larger than the one before. Moreover, as is true for any company, investments in research, new equipment and the like must eventually generate a healthy profit. At present, however, semiconductor companies have no way of determining precisely the proportion of their financial returns that comes from their technology investments. This inability poses a serious problem for the semiconductor industry. So for several years we have been working on methods of characterizing the industry that take into account these nonlinear elements, with an eye toward modifying the ROI model.



PHASE CHART shows the relation between Intel's profits and investments in technology throughout the company's history. Plotted points trace loops that each correspond to roughly a six-year cycle. (Each is shown in a different color.) During each of them, Intel roams from a period of unprofitable operations caused by heavy investment to an interval of very good cash generation stemming from much lighter investment. Green arrows indicate the year in each cycle when Intel made the most money and spent lightly on equipment.

In the conventional model, additional capital investments are made only when gaps occur between a manufacturer's actual and anticipated capacity (the latter is the capacity a company thinks it will need to meet demand in the near future). Such gaps usually result from the aging of equipment and the departure of experienced personnel. In industries such as semiconductors, on the other hand, not only must increases in capacity be constantly anticipated, but also great advances in the manufacturing technology itself must be foreseen and planned for.

To account for this technology-drag effect, we began by considering the ratio of cash generated during any given year to investments made in new technology the year before. New technology, in this context, consists of both new manufacturing equipment and research and development. Cash generated during the year is the gross profit generated by operations, including money earmarked for reinvestment in R&D. (For tax reasons, the standard practice in the industry is not to include R&D funds in this category but rather to treat them as an operating expense.)

What this ratio indicates are incremental profits per incremental investment, one year removed. It shows, in effect, how high a company is keeping its head above water, with respect to profits, thanks to its investment in ever more costly technology. ROI, in contrast, measures the incremental profits over a year coming from all investments, rather than just those of the previous year.

So far we have merely lumped new manufacturing equipment and R&D together as new technology. But the effect of technology drag becomes more striking when the two categories are separated, and the ebb and flow between them is elucidated. One way of doing this is to compute the ratio of these two investments year by year and then plot it against our old standby: the ratio of cash generated during a given year to investments made in new technology during the previous year. The results for Intel over most of its history are plotted in the chart at left.

Several interesting aspects of Intel's financial history emerge in this diagram, called a phase chart. Connecting the plotted points traces loops that each correspond to roughly a six-year cycle, during which Intel roams from a period of unprofitable operations caused by heavy capital investment to an interval of very good cash generation stemming from much lighter capital investment. From the chart, it is clear that Intel is now entering another period of heavy capital investment. Other semiconductor (and comparable high-technology) companies go through similar cycles. Of course, the timing of the periods of profitability and heavy investment varies from company to company.

Each loop's lower portion is lower than the one that preceded it. This insight is perhaps the most significant that the illustration has to offer, because it means that Intel's profits, relative to the capital expenditures generating them, are declining with each successive cycle. Because it shows the full cycle between investment in technology and its payoff, this phase chart is a powerful tool for observing and managing the investment cycles peculiar to this unique, dynamic industry. *—G.D.H. and J.D.H.*

tion equipment has gone up similarly.

Such increases have boosted the overall costs of building semiconductor plants at about half the rate of Moore's Law, doubling every three years. Intel is spending $1.1 billion on its new factory in Hillsboro, Ore., and $1.3 billion on another one in Chandler, Ariz. Samsung and Siemens are each building plants that will cost $1.5 billion to finish; Motorola has plans for a factory that may cost as much as $2.4 billion. Smaller factories can be built for less, but they tend to be less efficient.

That factories now cost so much is one piece of widely cited evidence that formidable technical barriers are close. But the fear that the barriers might be insurmountable, bringing the industry to a halt, seems to us to be unfounded. Rather the prices of semiconductors may increase, and the rate of change in the industry may slow.

Such an occurrence would not be entirely without precedent. The cost per bit of memories rose by 279 percent between 1985 and 1988 without dire consequences. In fact, 1988 was one of the semiconductor industry's best years. When the cost per bit begins to rise permanently, the most likely result will be an industrial phase change that significantly alters business models.

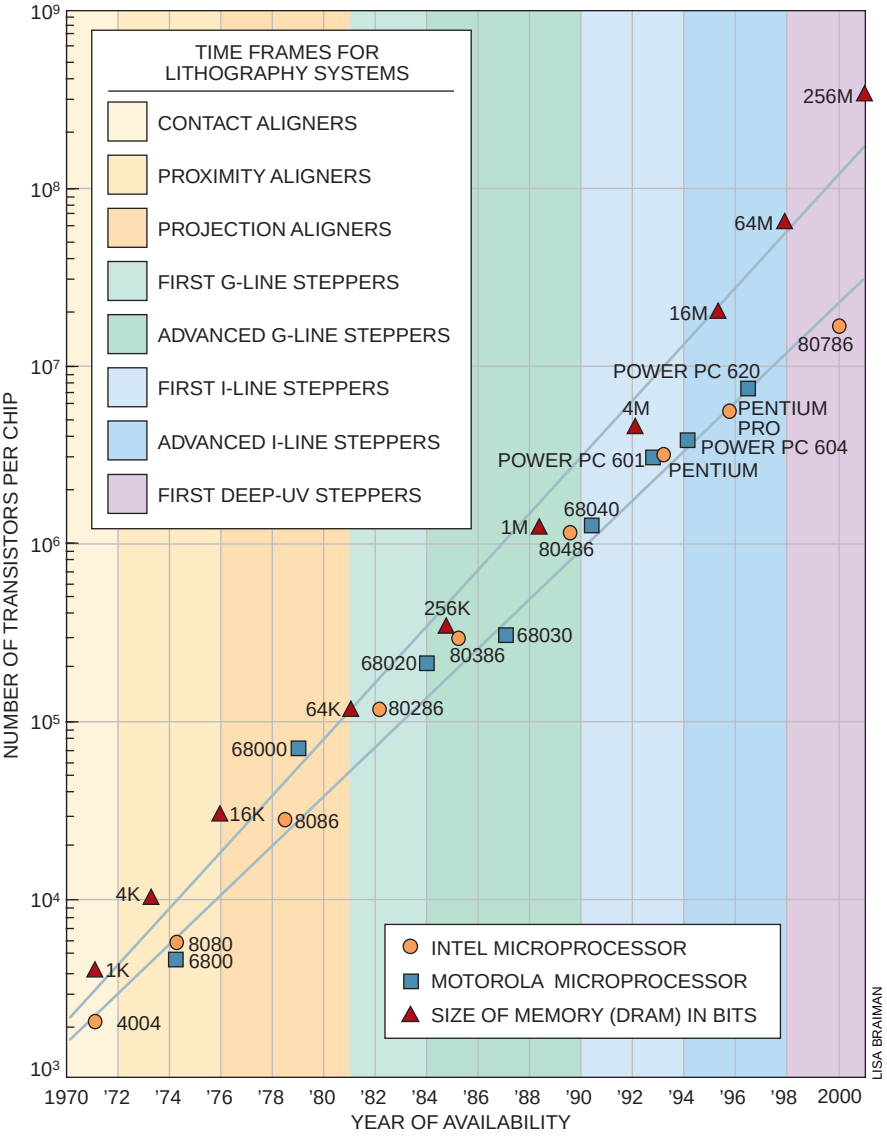
Up, Up and Away

Virtually every industry more than a few decades old has had to endure such phase changes. Although the semiconductor industry is obviously unique, it is still subject to the principles of economics and of supply and demand. So the history of older, technical industries, such as aviation, railroads and automobiles, would seem to have episodes that could act as pointers about what to expect.

Like the semiconductor industry, aviation had a fast start. In less than 40 years the industry went from the Wright brothers' monoplane to the Pan Am Clipper, the Flying Fortress and the Superfortress. Also like the semiconductor industry, aviation initially served mainly military markets before moving on to nonmilitary ones (mail and passenger transport). The aviation industry sustained growth by lowering the costs per passenger-mile traveled, while also reducing transit times. The dual missions are comparable to the semiconductor industry's steadfast efforts to increase the density of transistors on chips and therefore boost performance, while lowering chip costs.

For several decades, aviation grew by focusing its research and development on increasing passenger capacity and airspeed. Eventually, the trends peaked with Boeing's 747 as a benchmark for capacity and the Concorde as one for speed. Although the 747 was a successful aircraft, filling its many seats was often difficult on all but the longest routes. The Concorde, on the other hand, was an economic failure because noise pollution limited its use. Both represented high-water marks, in the sense that technology could not realistically provide more capacity or speed. Nevertheless, aviation did not go into a tailspin. It entered a second phase in which a greater diversity of smaller airplanes were designed and built for more specific markets. The focus of research and development shifted from speed and size to more efficient and quieter operations and more passenger comfort.

In railroads, the trends were similar. From the 19th century until well into the 1970s, the pulling power of locomotives was continually increased in order to lower the costs of moving freight. Locomotives were a significant capital expense, but gains in pulling power occurred more rapidly than increases in cost did. Eventually, however, locomotive development costs became so great that suppliers and users teamed up. The Union Pacific Railroad, the largest railroad of its time, joined with General Motors's Electro-Motive Division to create the EMD DD-40, a monster that turned out to be too big and inflexible for any purpose other than hauling freight clear across the U.S. Its failure

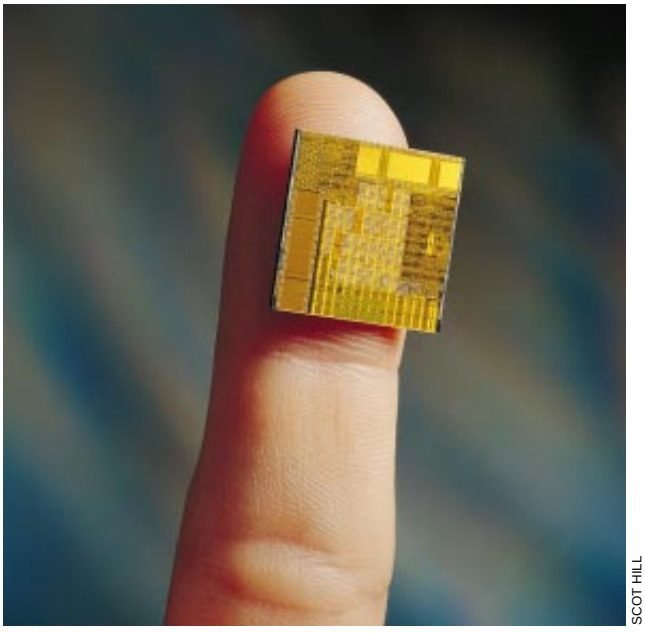


SOURCES: VLSI Research, Inc.; Integrated Circuit Engineering Corporation

TRANSISTOR DENSITIES on integrated circuits have increased at an exponential rate, as shown on this logarithmic plot. The rate has been sustained by a succession of lithography systems, which are used in chipmaking to project patterns onto wafers. Higher densities have been achieved in memory chips because of their more regular and straightforward design.

led the railroad industry back to the use of smaller engines that could be operated separately for small loads but hitched together for big ones.

Today the semiconductor industry finds itself in a position not unlike that of the railroad companies just before the EMD DD-40. The costs of developing the factories for future generations of memory chips are so high that companies have begun banding together into different groups, each to attack in its own way the enormous problems posed by fabricating these extremely dense chips economically.



SCOT HILL

INTEGRATED CIRCUIT, or die, for Motorola's Power PC 620 microprocessor has nearly seven million transistors. Housed in a ceramic package, it will be used mainly in computer workstations and file servers.

Big Plants, Little Variety

From automobile manufacturing, too, come important lessons. In the 1920s Henry Ford built increasingly more efficient factories, culminating with the vast Rouge plant, which first built Model A cars in 1928. Starting with iron ore, the facility manufactured most of the parts needed for automobiles. But the automobile industry had already changed, and Ford's efforts to drive down manufacturing costs by building larger and more efficient plants came at the price of product variety. The old joke about Ford was that the buyer could have a car in any color he or she wanted, as long as it was in black.

Trends in auto manufacturing shifted to permit more conveniences, features and models. As the industry matured, Alfred E. Sloan of General Motors recognized that efficiency was no longer increasing with factory size and that big factories were good mainly for building large numbers of the same product. He therefore split the company into divisions with clearly defined markets and dedicated factories to support them. Customers preferred the resulting wider variation in designs, and General Motors was soon gaining market share at Ford's expense.

Similar scenarios are being played out in chips. Intel split its 486 microprocessor offerings into more than 30 variations. During the early 1980s, in contrast, the company offered just three versions of its 8086 microprocessor and only two versions of its 8088. Dynamic memory chips are being similarly diversified. Toshiba, for example, currently has more than 15 times as many four-megabit DRAM configurations as it had 64-kilobit ones in 1984.

The common theme in all these industries, from railroads to semiconductors, is that their initial phase was dominated by efforts to improve performance and to lower cost. In the three transportation industries, which are considerably more mature, a second phase was characterized by product refinement and diversity—similar to what is now starting to happen in chipmaking. Companies are shifting their use of technology from lowering manufacturing costs to enhancing product lines. The important point is that all these industries continued to thrive in spite of higher manufacturing costs.

It may not be long before the semiconductor industry plateaus. The pace of transistor integration will decline, and manufacturing costs will begin to soar. But as the histories of the aviation, railroad and automobile industries suggest, the semiconductor industry could flourish as it encounters unprecedented and even largely impassable economic and technical barriers. In a more mature industry, growth will almost certainly come from refined products in more diversified lines.

Information storage, and those societal functions that depend on it, will keep moving forward. In fact, slowing the rate of progress in semiconductors could turn out to have unexpected advantages, such as giving computer architectures and software time to begin assimilating the great leaps in chip performance. Even in the semiconductor industry, maturity can be a splendid asset.

The Authors

G. DAN HUTCHESON and JERRY D. HUTCHESON have devoted their careers to advancing semiconductor manufacturing. In 1976 Jerry founded VLSI Research, Inc., as a technical consulting firm. Dan, his son, joined in 1979. They have focused on analyzing how the interplay of technology and economics affects the business of semiconductor manufacture. Before founding VLSI Research, Jerry, who entered the industry in 1959 as a device physicist, held various positions in research, manufacturing and marketing at RCA, Motorola, Signetics, Fairchild and Teledyne Semiconductor. Dan started his career at VLSI Research as an economist, building several simulation models of the manufacturing process. In 1981 he developed the first cost-based model to guide semiconductor companies in their choice of manufacturing equipment. He currently serves on the University of California's Berkeley Extension Advisory Council and the Semiconductor Industry Association's Technology Roadmap Council.

Further Reading

IS SEMICONDUCTOR MANUFACTURING EQUIPMENT STILL AFFORDABLE? Jerry D. Hutcheson and G. Dan Hutcheson in *Proceedings of the 1993 International Symposium on Semiconductor Manufacturing.* Institute of Electrical and Electronics Engineers, September 1993.

SIA 1994 NATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS. Semiconductor Industry Association, 1994.

LITHOGRAPHY AND THE FUTURE OF MOORE'S LAW. Gordon E. Moore in *SPIE Proceedings on Electron-Beam, X-Ray, EUV, and Ion Beam Lithographies for Manufacturing,* Vol. 2437; February 1995.

TOWARD POINT ONE. Gary Stix in *Scientific American,* Vol. 272, No. 2, pages 90–95; February 1995.

AFFORDABILITY CONCERNS IN ADVANCED SEMICONDUCTOR MANUFACTURING: THE NATURE OF INDUSTRIAL LINKAGE. Donald A. Hicks and Steven Brown in *Proceedings of the 1995 International Symposium on Semiconductor Manufacturing.* Institute of Electrical and Electronics Engineers, September 1995.

Neural Networks for Vertebrate Locomotion

The motions animals use to swim, run and fly are controlled by specialized neural networks. For a jawless fish known as the lamprey, the circuitry has been worked out

by Sten Grillner

It is difficult to grasp how the human brain is able to keep up with the requirements of running or even walking: deciding what joint needs to be moved, exactly when it should bend and by how much, and then sending the proper series of impulses along nerves to activate the appropriate combination of muscles. The dexterity that even lowly creatures display as they swim, fly, run or otherwise propel their bodies through their surrounds is truly marvelous. Even the most sophisticated mobile robots perform poorly in comparison.

Although many mysteries of animal locomotion are yet unsolved, scientists are beginning to comprehend the way vertebrates (creatures with backbones, including humans) can almost effortlessly coordinate complicated movements that may involve hundreds of muscles. The formidable task of managing the body's various motions is simplified by a remarkable form of neural organization, one that distributes the responsibility for coordinating such acts to distinct networks of nerve cells. Some of these specialized circuits, such as the one that keeps a person constantly breathing, are ready to operate flawlessly from birth. Others, such as those that control crawling, walking or running, can take time to mature.

The neural networks that govern specific, oft-repeated motions are sometimes called central pattern generators. They can steadfastly execute a particular action over and over again without the need for conscious effort. The key neural-control circuits that humans use for breathing, swallowing, chewing and certain eye movements are contained within the brain stem, which surrounds the uppermost spinal cord. Oddly enough, the circuits for walking and running (as well as some protective reflexes) are not located in the brain at all but reside in the spinal cord itself.

Since the late 1960s, my colleagues and I have been attempting to unravel the design of the neural systems that coordinate locomotion in various experimental animals in hopes that this research will help scientists understand some of the intricacies of the human nervous system. Much is yet to be learned, but we have finally produced a blueprint for the neural networks responsible for movement in a simple vertebrate, a type of jawless fish known as a lamprey.

Of Mice and Men

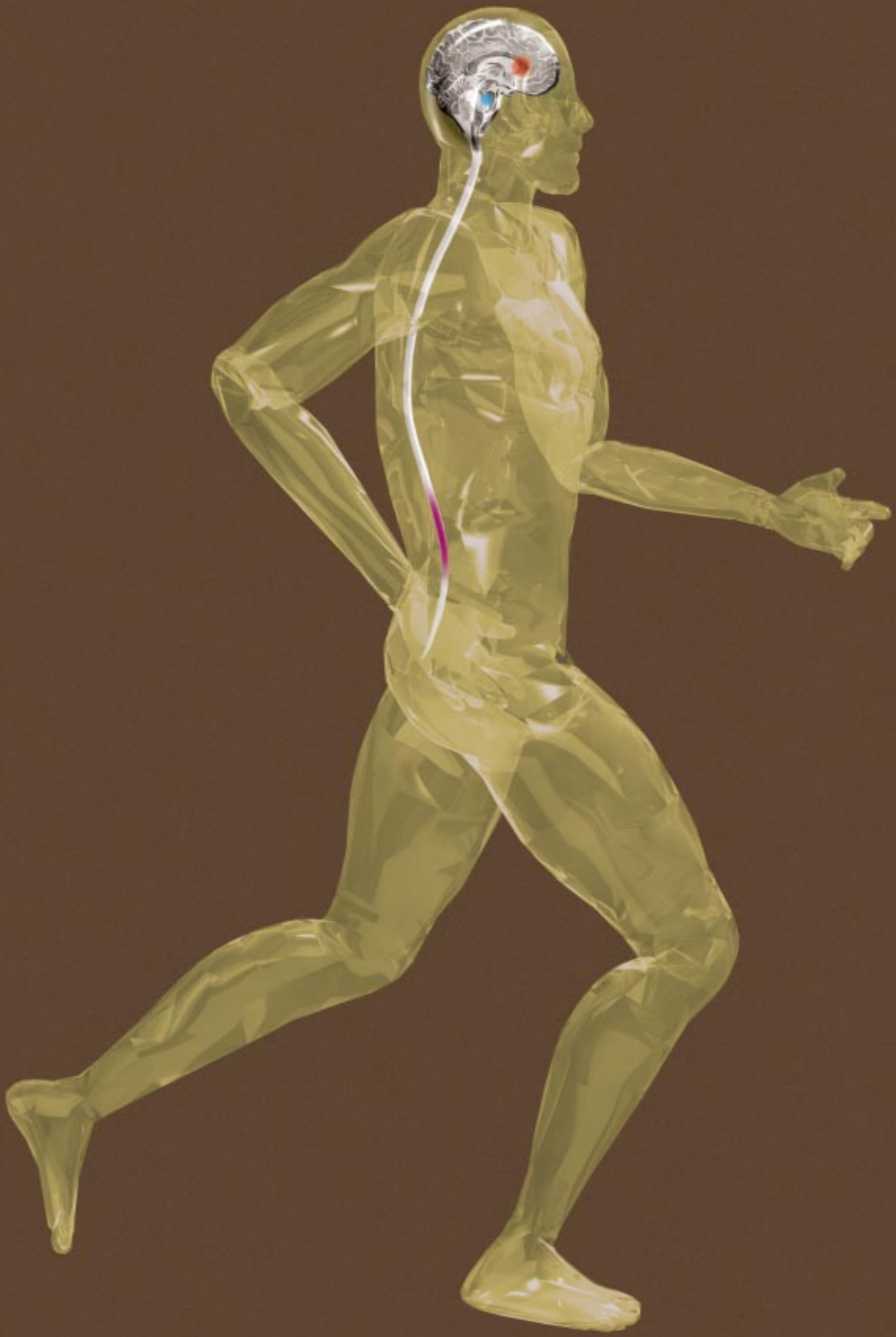
Scientists have deduced much about the organization of the human central nervous system from studies of laboratory animals. Appreciation for the significance of the spinal cord to locomotion first came just after the turn of the century, when pioneering British neurophysiologists Charles S. Sherrington and T. Graham Brown observed that mammals with severed spinal cords could produce alternating leg movements even though the connection to the brain had been cut. Much later my colleagues and I were able to show definitively that such motions corresponded to those movements used for locomotion. Thus, we could conclude that the essential nerve signal patterns for locomotion are generated completely within the spinal cord.

Yet it remained a question how the brain controls these circuits and chooses which should be active at a given instant. Much insight into this process came during the 1960s, when the Russian investigators Grigori N. Orlovski and Mark L. Shik, then at the Academy of Science in Moscow, demonstrated that the more that specific parts of the brain stem of a cat were activated, the faster the animal under study would move. With increasing stimulation, the cat would proceed from a slow walk to a trot and finally to a gallop. A very simple control signal from a restricted area of the brain stem could thus generate intricate patterns involving a large number of muscles in the trunk and limbs by activating the pattern generators for locomotion housed within the animal's spinal cord.

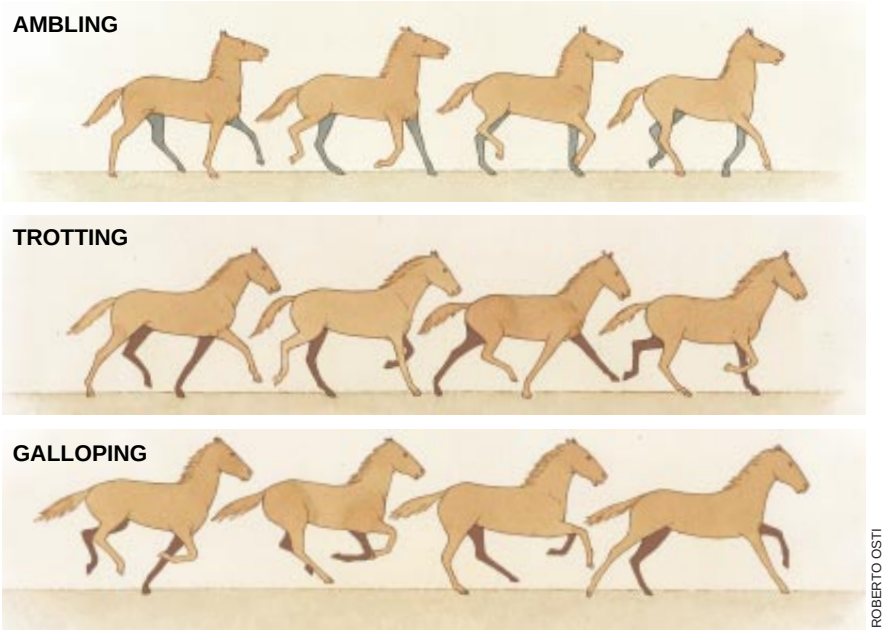
Beyond providing clues to the interactions between brain and spinal cord, this experiment helped to explain how animals can move about even after much of their brain is surgically removed. Some mammals (such as the common laboratory rat) can have their entire forebrain excised and are still able to walk, run and even maintain their balance to some extent. Although they move with a robotic stride, without making any attempt to avoid obstacles placed in their path, these animals are fully able to operate their leg muscles and to coordinate their steps.

The details of how the brain activates neural networks in the spinal cord took years to lay out. It is now known that large groups of nerve cells in the fore-

LOCOMOTION for humans, like all vertebrate animals, is orchestrated by the central nervous system. Specialized neural circuits in the forebrain (*red*) select among an array of "motor programs" by activating specific parts of the brain stem (*blue*). The brain stem in turn initiates locomotion and controls the speed of these movements by exciting neural networks (called central pattern generators) located within the spinal cord (*purple*). These local networks contain the necessary control circuitry to start and stop the muscular contractions involved in locomotion at the appropriate times. Networks of neurons in the brain stem also control breathing, chewing, swallowing, eye movements and other frequently repeated motor patterns.



SLIM FILMS



ROBERTO OSTI

PATTERN GENERATORS, separate neural networks that control each limb, can interact in different ways to produce various gaits, such as the amble, trot and gallop of a horse. In ambling (*top*), the animal must move the foreleg and hind leg of one flank in parallel. Trotting (*middle*) requires movement of diagonal limbs (front right and back left, or front left and back right) in unison. Galloping (*bottom*) involves the forelegs, and then the hind legs, acting together.

brain, called basal ganglia, connect (either directly or through relay cells) to target neurons in the brain stem that in turn can initiate different "motor programs." Under resting conditions, the basal ganglia continuously inhibit the brain's sundry motor centers so that no movements occur. But when the active inhibition is released, coordinated motions may begin. The basal ganglia thus function to keep the various motor programs of the nervous system under strict control. This suppression is essential: renegade operation of a motor program could be disastrous for most any animal.

In humans, for instance, diseases of the basal ganglia can cause involuntary facial expressions and hand or limb movements. Such hyperkinesis occurs commonly in cerebral palsy and Huntington's disease and as a side effect of some medications. Other diseases of the basal ganglia can lead to the opposite situation, with more inhibition than desired being applied; victims then have difficulty initiating movements. The best known example of such a disability is Parkinson's disease.

Ferrari or Model T?

Although medical researchers keenly desire to understand how such neurological disorders arise and what might be done to correct them, progress has been difficult to achieve because the human nervous system (which houses nearly a trillion neurons) is extraordinarily complicated. It is not yet feasible to examine the neural circuits in humans, or indeed in any mammal, in much detail. My colleagues and I have therefore focused our studies on much simpler vertebrates. We sought an experimental animal with the same basic neural organization as humans but with far fewer components.

Our fundamental approach has been similar to something an imaginary researcher from outer space might undertake to deduce the basic mechanics of an automobile. An extraterrestrial scientist would fare best by beginning such an analysis with a Model T Ford (if one could be obtained), because that vintage vehicle has all the essential components of a car—internal-combustion engine, transmission, brakes and steering—manufactured from a simple design and arranged for easy inspection. Investigations that began by directly probing a more advanced model, such as a modern turbocharged Ferrari, might prove far more frustrating. One presumes that knowledge of a Model T would serve as the foundation needed to understand the anatomy of the more elegant and sophisticated car.

We investigated several possible subjects before settling finally on the lamprey—an elongate, jawless fish with a large mouth adapted for sucking. The lamprey is a primitive vertebrate with a nervous system composed of comparatively few cells (only about 1,000 in each segment of the spinal cord), making it ideal for our purposes. The lamprey also suited us because Carl M. Rovainen of Washington University had shown that the fish's central nervous system could be maintained in a glass dish and studied for several days after it is removed from the animal. Moreover, motor networks in the isolated nervous system remain active.

The strategy of choosing a simple but relevant experimental animal for study has yielded key insights into many different biological processes. For example, examination of invertebrate nerve cells, such as those of the squid and lobster, provided the first important clues to how nerve impulses are generated and how networks of nerve cells function.

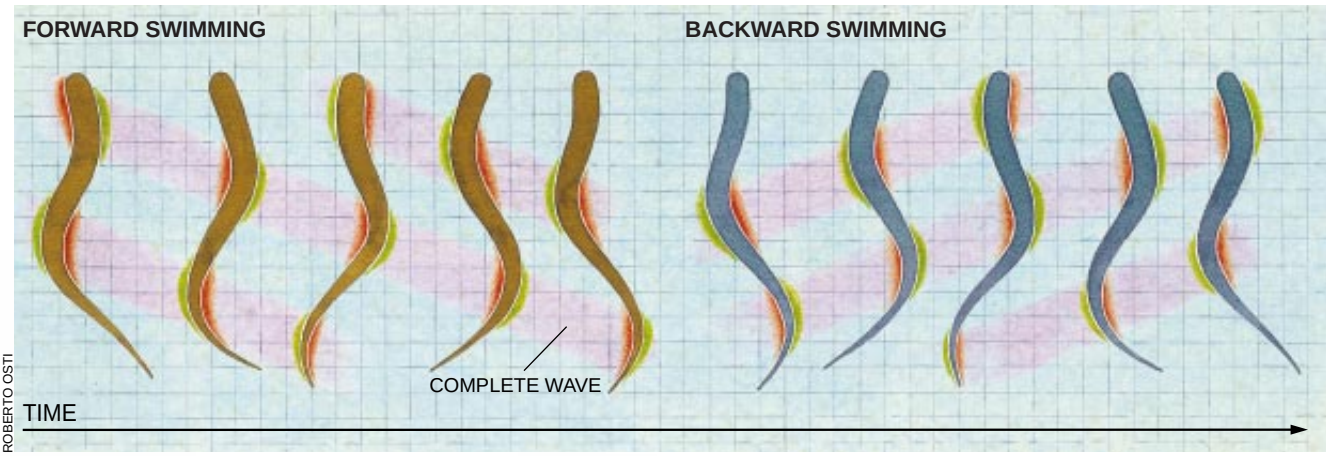
A Hardwired Fish

From the beginning of our studies with the lamprey in the late 1970s, my colleague Peter Wallén and I, along with a number of collaborators, have concentrated on understanding the fundamental features of the animal's swimming. Like other fish, the lamprey propels itself forward through the water by contracting its muscles in an undulating wave that passes along the creature's body from head to tail.

To produce a propulsive wave, the animal must generate bursts of muscle activity that bend each section of the spine toward one side and then the other in rhythmic alternation. But the lamprey also needs to coordinate the contractions of consecutive segments along its body so that a smooth wave forms. We soon discovered that the neural controls for both these abilities are distributed throughout the spinal cord. If a lamprey's spinal cord is isolated and separated into several pieces, each length can be made to show the characteristic alternating pattern, and within any given portion the activity between adjoining segments stays coordinated.

Further observations showed that the lag between activation of adjacent segments remains fixed during a given wave, as the undulatory motion propagates down the body of the lamprey. But the lag time changes with the fish's speed, so that the overall period of that wave (the time it takes for the wave to travel the entire length of the body) can vary from about three seconds during very slow swimming to as little as one tenth of a second for sudden sprints. Exactly the same characteristic contractions occur in reverse order when the fish swims backward.

To understand how the lamprey ner-



ROBERTO OSTI

UNDULATORY SWIMMING in the eellike lamprey constitutes a relatively simple form of vertebrate locomotion that neuroscientists can examine effectively. In response to signals emitted by the brain, wave after wave of muscle contraction (*red*) and extension (*green*) pass from head to tail down the body of a fish, propelling it forward through the water (*left*). Similar waves traveling from tail to head can drive the creature backward (*right*).

vous system could orchestrate such motions, my colleagues and I needed to determine exactly which nerve cells contributed to locomotion and how they interacted. So we devised experiments using electrodes with tips that were less than a thousandth of a millimeter wide. With these sensors we could map out distant connections by placing one electrode inside an individual cell in the brain stem and, at the same time, using another electrode to probe various target cells in the spinal cord. To find a pair of nerve cells that could communicate with each other among the hundreds of possibilities required considerable skill and patience. But the diligent labor of many people finally made it possible to identify the neurons that controlled locomotion and to trace how they were wired together.

We ultimately discovered that nerve cells in the brain stem have long extensions (axons) that are in contact with the neurons involved with locomotion throughout the spinal cord. In response to signals from the brain, local networks of cells within discrete parts of the cord generate bursts of neural activity. These networks act as specialized circuits, exciting the neurons on one side of a given segment of the lamprey's body while suppressing similar nerve cells on the

ROBERTO OSTI

TONY SICA *Gamma Liaison*

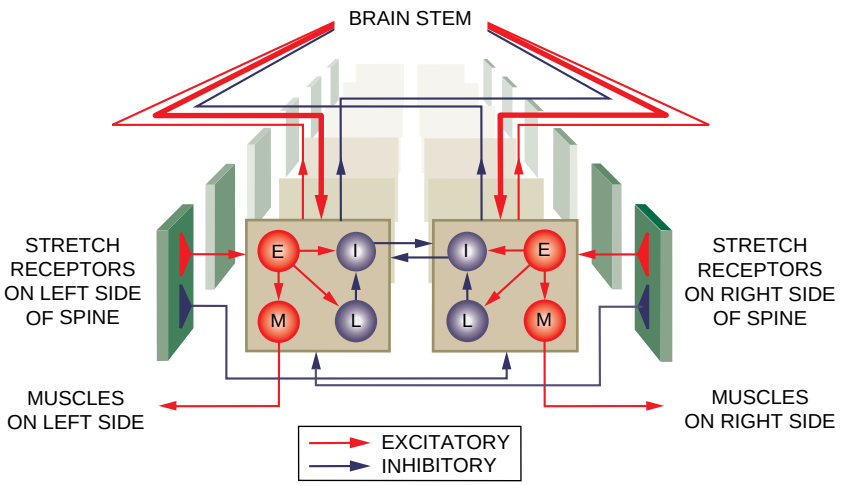
SEA LAMPREY (*Petromyzon marinus*), which can be up to a meter in length, has served as an ideal experimental animal for the author's studies of vertebrate locomotion. Because the fish's nervous system is comprised of relatively few cells, the brain stem (*brown*) and spinal cord (*beige*) can be isolated and probed in the laboratory (*inset*).

Parallel Processing

Within a single segment of the lamprey's spinal cord lies an intricate network of interconnected nerve cells. Groups of neurons (*boxes*) on the left and right sides of the cord are excited by signals sent from the animal's brain stem. Specialized neurons within these groupings respond by sending either excitatory (*red*) or inhibitory (*purple*) signals to neighboring cells. Neurons known as E cells (for excitatory) on one side of the spinal segment will activate motoneurons (M) that in turn cause the muscles on that side of the fish to contract. These E cells also induce inhibitory (I) neurons to reduce the level of excitation in the group of neurons on the opposite side of the spine, ensuring that the opposing muscles relax.

The bursts of excitation that cause one side to contract are terminated in a number of ways. Certain stretch receptor neurons (*purple triangles*) on the opposite side of the spine emit signals that inhibit the contraction. At the same time, other activated stretch receptors (*red triangles*) excite the neurons on the extended side to initiate contraction there. In addition, large (L) inhibitory nerve cells on the contracting side can be induced by the brain stem to inhibit the I cells. This allows the opposite side to become active and send inhibitory signals back. Finally, there are several electrochemical mechanisms inside cells that can force a pulse of excitation to subside, helping to control the timing of the network.

Although these local spinal cord circuits can operate autonomously, they normally feed back information to the brain about the ongoing network activity. These signals can then be combined with other forms of sensory input, such as cues from vision or from the balance system in the inner ear, to modify the animal's movements.



CAREY BALLARD

opposite side. Thus, when one flank of a given section becomes active, the other is automatically inhibited. Other specialized nerve cells, called motoneurons, link the nerves of the spinal cord to the muscle fibers that actually do the job of moving the lamprey through water.

But these spinal networks are not simply passing signals sent down from the brain of the animal. Although the brain stem issues the overall command for the fish to swim, it delegates the task of coordinating the muscle movements to these local teams, which can process incoming sensory data and adjust their own behavior accordingly. In particular, they react to specific "stretch receptor neurons" that sense the bending of the lamprey's spine as it swims.

As one side of the body is contracting, the other is extending—and it is this extension that triggers the stretch receptors. These activated nerve cells then take one of two complementary actions: they either excite neurons on the extended side (inducing muscles there to contract), or they inhibit neurons on the opposite side, causing them to halt contraction. By such processes (as well as several rather complex cellular mechanisms), the fundamental oscillatory movements of the lamprey's neuromuscular system are maintained.

As we further followed the neural circuitry of the lamprey's spinal cord, we determined that some of the local neural networks extend axons along the spine. Special inhibitory cells in each segment send signals through these axons in the direction of the tail for as much as one fifth of the length of the spine. So-called excitatory cells contain somewhat shorter axons that extend in both directions. Thus, the activity at one location on the spinal cord can affect adjacent regions.

But how exactly might signals linking different segments create the characteristic wavelike motion? After much thought, we proposed that nerve signals could excite the leading segment (near the lamprey's head) so that the contractions there alternate back and forth faster than the spine would otherwise tend to oscillate. The second section behind the head would follow the quickened motions of the first (because the two segments are coupled by nerve cells), but with a slight lag as the inherently slower section tried to catch up with the leader. By similar reasoning, the third section should then follow the second with a slight delay—and so forth down the line. The series of incremental delays, we surmised, allowed the lamprey to produce a uniform wave.

Virtual Reality

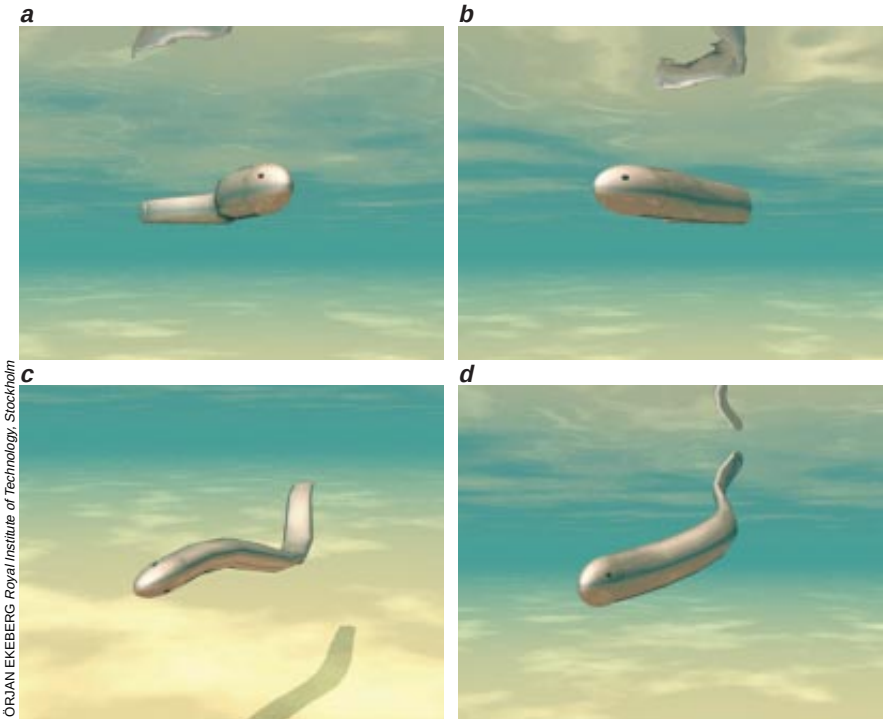
Even with our newly developed wiring diagrams and a mass of other detailed information about the properties of the different types of nerve cells involved, we were long challenged to make more than modest, general statements about how these complex neural circuits operated. To test whether the information we had gleaned truly explained how the lamprey could swim, my colleagues and I joined with Anders Lansner and Örjan Ekeberg of the Royal Institute of Technology in Stockholm to create various computer models of the process.

First, we developed schemes that could reproduce the behavior of the different neurons used for locomotion. Then we succeeded in simulating on our computer the entire ensemble of interacting cells. These numerical exercises allowed us to test a variety of possible mechanisms, and they have proved to be indispensable tools in the analysis of the lamprey's neural organization. Because the computer models can generate a signal pattern that is quite similar to that occurring during actual locomotion, we can finally say with some confidence that the circuits we have deciphered do indeed capture essential parts of an extensive biological-control network.

Our computer simulations not only showed alternating contractions on ei-

ther side of the spine but also refined our conception of the lag between the activation of adjoining segments. This delay arises from the neurons that reach along the cord and inhibit segments in the tailward direction. These connections ensure that the overall level of excitation will typically be highest at the head end of the animal—a condition that leads to delayed activation of the segments, one after the other, all along the animal's body. We also found that the normal pattern could be reversed by increasing the excitability in the most tailward part of the spinal cord, thereby enabling backward swimming. The "hardwired" spinal network of the lamprey thus retains a considerable degree of flexibility.

For the most part, we considered computer simulations that mimic only the lamprey's neural activity. But recent efforts led by Ekeberg have succeeded in modeling the entire lamprey, from the muscle fibers controlling the different segments to the viscous properties of the surrounding water. The neural-control circuits we had previously charted provided everything this virtual lamprey needed to swim.



ÖRJAN EKEBERG *Royal Institute of Technology, Stockholm*

VIRTUAL SWIMMING by a simulated lamprey suggests that neural models developed in the laboratory can portray how the real creature maneuvers itself through the water. These computer-generated images show the lamprey swimming straight (*a*), turning (*b*), rolling to one side (*c*) and pitching downward (*d*).

Crawling Out of the Water

We can now be satisfied that the lamprey's capacity for locomotion can be understood in terms of the interactions of spinal nerve cells. But how certain is it that these mechanisms operate in higher forms of life? The lamprey diverged from the main vertebrate line quite early during the course of evolution, about 450 million years ago, at a time when vertebrates had not yet developed limbs. So it was not immediately obvious whether our results were relevant to other animals.

But the mechanism for locomotion in one other vertebrate—the tadpole—has also been revealed at a cellular level by Alan Roberts and his colleagues at the University of Bristol. It has been especially comforting for me to see that the tadpole's nervous system resembles in most aspects the lamprey network. For other vertebrates as well, from fish to primates, the overall neural organization is arranged along a similar plan. Discrete regions of the brain stem initiate locomotion, and the spinal cord processes the signals with specialized circuits.

Yet the cellular mechanisms used for locomotion in these other animals are still largely unknown. Researchers have shown that pattern generators are present and have probed some of their neural components, but so far it has not been possible to unravel their inner architecture. During the past few years, however, new techniques have been developed to isolate the spinal cords of the other classes of vertebrates (mammals, birds and reptiles), and it seems likely that in the next few years investigators may uncover how these animals control walking, running and flying.

Because the earliest vertebrates used only undulatory swimming for locomotion, the networks that later evolved to control fins, legs and wings may not be all that different from what my colleagues and I have already studied. Evolution rarely throws out a good design but instead modifies and embellishes on what already exists. It would be most surprising to discover that there were few similarities between lampreys and humans in the organization of control systems for locomotion. Scientists may yet devise ways to map out and to activate dormant pattern-generating circuits in people with severed spinal cords. Indeed, such miraculous medical advances might not be that far away: a turbocharged Ferrari is, after all, just another kind of car.

The Author

STEN GRILLNER received an M.D.–Ph.D. degree in 1969 from the University of Göteborg in Sweden, where he then joined the faculty. In 1975 Grillner moved to the department of physiology at the Karolinska Institute in Stockholm. He joined the Nobel Institute for Neurophysiology in 1987 and now serves both as chairman of the department of neuroscience at Karolinska and as director of the Nobel Institute. Grillner has also been a member of the Nobel Committee for Physiology or Medicine since 1987.

Further Reading

NEUROBIOLOGICAL BASES OF RHYTHMIC MOTOR ACTS IN VERTEBRATES. S. Grillner in *Science,* Vol. 228, pages 143–149; April 12, 1985.

NEURAL CONTROL OF RHYTHMIC MOVEMENTS IN VERTEBRATES. Avis H. Cohen, Serge Rossignol and Sten Grillner. Jon Wiley & Sons, 1988.

THE NEURAL CONTROL OF FISH SWIMMING STUDIED THROUGH NUMERICAL SIMULATIONS. Ö. Ekeberg, A. Lansner and S. Grillner in *Adaptive Behavior,* Vol. 3, No. 4, pages 363–384; Spring 1995.

NEURAL NETWORKS THAT CO-ORDINATE LOCOMOTION AND BODY ORIENTATION IN LAMPREY. S. Grillner, T. Deliagina, Ö. Ekeberg, A. El Manira, R. H. Hill, A. Lansner, G. N. Orlovski and P. Wallén in *Trends in Neuroscience,* Vol. 18, No. 6, pages 270–279; June 1995.

Cleaning Up the River Rhine

Intensive international efforts are reclaiming the most important river in Europe

by Karl-Geert Malle

VOLKER HINZ

In Köln, a town of monks and bones,
and pavements fanged with murderous stones
And rags, and hags, and hideous wenches;
I counted two and seventy stenches,
All well defined, and several stinks!
Ye Nymphs that reign o'er sewers and sinks,
The river Rhine, it is well known,
Doth wash your city of Cologne;
But tell me, Nymphs, what power divine,
Shall henceforth wash the river Rhine?

So wrote Samuel Taylor Coleridge in 1828, after a visit to the Rhine with William Wordsworth. Just how prophetic those lines were, Coleridge could hardly have guessed. In the 1960s and early 1970s, as central Europe resurged economically, organic and inorganic pollution in the Rhine reached levels high enough to decimate or wipe out dozens of fish species and other creatures that had existed in the river for many thousands of years. Most of the river became unsuitable for swimming or bathing, and the production of drinking water was threatened.

After the 1970s, however, international attempts to clean up the Rhine, including some dating to the 1950s, finally began reclaiming long stretches of the river. Today, although much remains to be done, the work is emerging as a success story in cooperation among nations for the sake of pollution control and as a model for the many other places where transborder flows of contaminants have strained relations.

Most important, the reclamation efforts are giving new life to one of the world's great rivers. Countless poets, prose writers and lyricists have praised (or lamented) this fabled waterway, which winds through or along five countries before emptying finally through a Dutch delta into the North Sea. From Alpine headwaters in east central Switzerland, the Rhine flows 1,320 kilometers north and west through rugged mountains, lovely Lake Constance in Switzerland, the Black Forest, broad Alsatian valleys and such cities as Strasbourg, Bonn, Düsseldorf and Rotterdam. Both the Volga and the Danube are mightier and longer, but neither can match the Rhine's relatively constant flow and utility as an artery into the heart of Europe.

The catchment area, within which precipitation collects and feeds the river, covers some 190,000 square kilometers and has a population of 50 million. More than eight million of these people get their drinking water from the river, and another 10 million from Lake Constance. All told, more than 20 percent of the average flow of the Rhine is diverted and used, either for human consumption or for industrial rinsing and cooling. In addition, 10 hydroelectric plants, built after World War I and clustered in French territory below Alsace, generate a total of 8.7 million megawatt-hours every year.

The river is navigable for 800 kilometers, from the North Sea to the waterfalls below Lake Constance. Each day more than 500 barges ply its waters, laden with gasoline and other petroleum-based fuels, salt, phosphate rock, coal, gravel, new automobiles and other cargo, which adds up to about 150 million tons a year. Overall, use of the Rhine has been estimated to be five times greater than that of the Mississippi, when measured in terms of gross national product and accounting for the two rivers' flow rates and catchment populations.

One of the main reasons for the Rhine's importance is its relatively reliable water supply. In winter and spring, the flow is fed by precipitation in the plains and uplands of the catchment area; in summer, melting snow from the Alps takes over. Constance and other Swiss lakes act as convenient reservoirs. The average flow rate between 1925 and 1992 was 2,350 cubic meters per second, and the ratio between the greatest yearly flow (3,170 cubic meters per second in 1966) and the lowest (1,510 in 1971) is 2.1—less extreme than for most other rivers.

Reclamation Begins

Before the end of World War II, hardly any large sewage treatment plants were situated on the Rhine. The odors so poignantly described by Coleridge were at one time confined to the major cities, whose sewage was released untreated into the river. After the war, however, as central Europe revived its economies, the pollution problem intensified to such an extent that it could no longer be ignored. In 1953 five nations—Switzerland, France, Luxembourg, Germany and the Netherlands—formed the International Commission for the Protection of the Rhine against Pollution, to coordinate multinational efforts and to monitor, at least, the levels of contaminants in the river.

The commission, known by the acronym IKSR, now monitors water quality at several fixed points, mostly at inter-

national borders. The most significant of these lies between Germany and the Netherlands, because this area is not far downstream from the heavily industrial German Ruhrgebiet. Named after the Ruhr, a small tributary of the Rhine, this region's coal, steel and chemical plants constitute one of the major sources of pollution. Moreover, after entering the Netherlands, the river passes through a fairly rural area, with no significant sources of pollution until it encounters the big Dutch harbors at Rotterdam and Amsterdam, some 150 kilometers from Germany and not far from the delta on the North Sea.

Another notable factor in this regard is the river's flow, which slows considerably in the Netherlands. After making its way from Switzerland through France and Germany within seven or eight days, the water remains for 70 or 80 days in the Netherlands because of the country's sophisticated system of dikes, canals and other water-management facilities (including two large artificial lakes, the IJsselmeer and the Haringvliet). Shortly after entering the Netherlands, the Rhine becomes a delta with three separate estuaries: the Waal, Lek and IJssel.

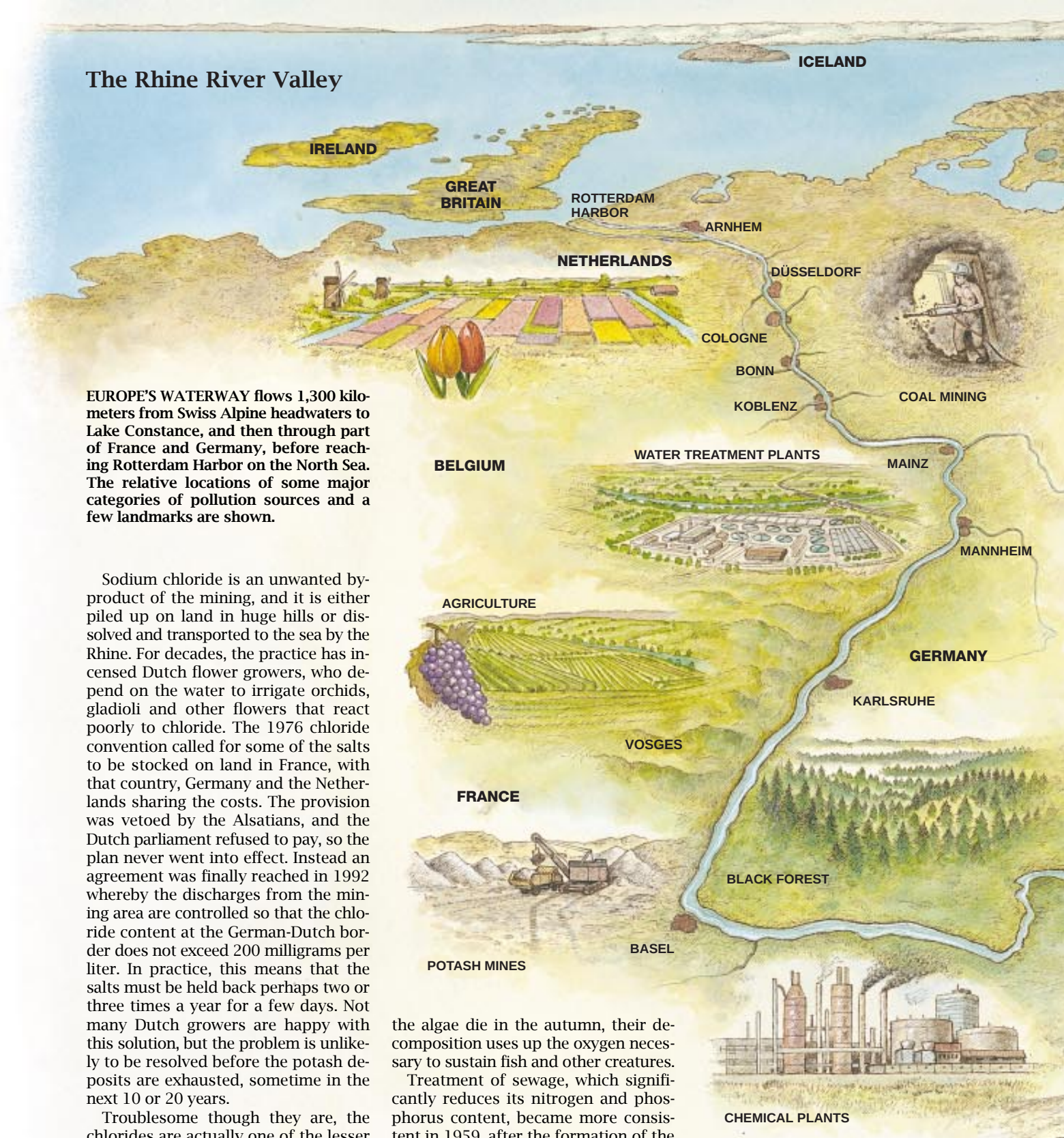
The IKSR issues an annual report with extensive analytical data, reviewed and confirmed by the countries within the commission. But its main business is proposing and implementing international projects to improve the river. The two most significant of these are the Convention on the Protection of the Rhine against Chemical Pollution and the Convention on the Protection of the Rhine against Chloride Pollution. Both were put into effect in 1976. Another convention, which would have addressed thermal pollution, was drafted but never ratified by all the member states. Although the river is being warmed gradually by natural trends and human activities, the degree of warming—1.8 degrees Celsius since 1925—is not considered onerous enough to merit extensive attention.

INDUSTRY AND FARMING have contributed significantly to the Rhine's pollution. A combination of tighter regulations, voluntary controls and subsidies for farmers who limit their use of certain chemicals has lately lessened impacts on the river.

Inorganic Contaminants

Rivers typically contain far more inorganic salts than organic substances. Some of these salts are leached naturally out of the soil by rainwater and carried to the river. Through sewage, industrial activity and agriculture, however, humankind has significantly augmented the natural content. The concentration of salts—including anions such as chloride and carbonate, as well as cations such as sodium and calcium—was measured at 581 milligrams per liter at the German-Dutch border in 1992. There have been no striking changes in this level of concentration over the past two decades. Reducing the concentration of fairly dilute salts is extremely expensive, so the water is desalinated only when necessary—before being used for irrigation, for example.

Human activities have been especially productive of chloride ions. Based on the 1992 measurements, it is believed that they contribute some 318 kilograms of chloride to the river every second, as compared with the 15 to 75 kilograms per second from nature. The largest sources are the French potash mines in Alsace, which alone add 130 kilograms per second.



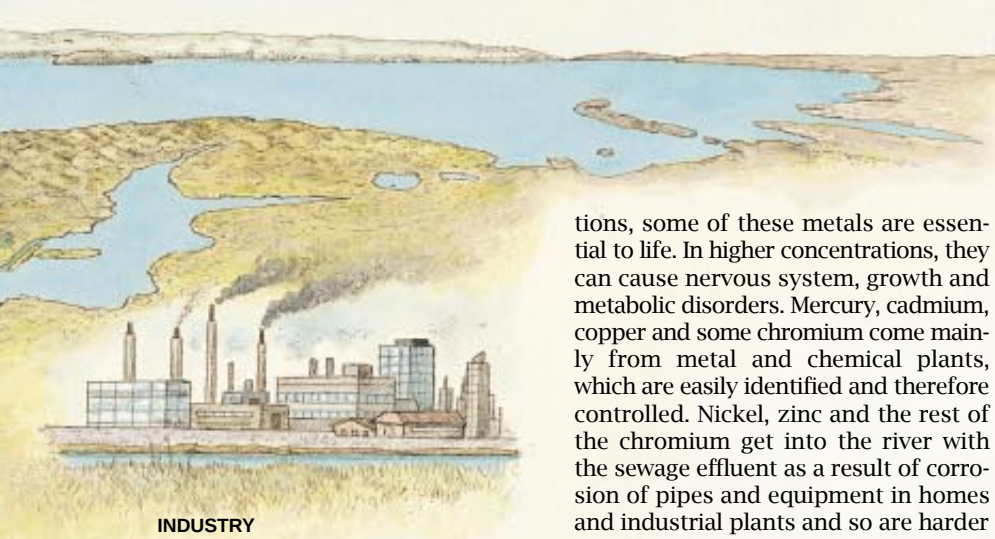
EUROPE'S WATERWAY flows 1,300 kilometers from Swiss Alpine headwaters to Lake Constance, and then through part of France and Germany, before reaching Rotterdam Harbor on the North Sea. The relative locations of some major categories of pollution sources and a few landmarks are shown.

Sodium chloride is an unwanted by-product of the mining, and it is either piled up on land in huge hills or dissolved and transported to the sea by the Rhine. For decades, the practice has incensed Dutch flower growers, who depend on the water to irrigate orchids, gladioli and other flowers that react poorly to chloride. The 1976 chloride convention called for some of the salts to be stocked on land in France, with that country, Germany and the Netherlands sharing the costs. The provision was vetoed by the Alsatians, and the Dutch parliament refused to pay, so the plan never went into effect. Instead an agreement was finally reached in 1992 whereby the discharges from the mining area are controlled so that the chloride content at the German-Dutch border does not exceed 200 milligrams per liter. In practice, this means that the salts must be held back perhaps two or three times a year for a few days. Not many Dutch growers are happy with this solution, but the problem is unlikely to be resolved before the potash deposits are exhausted, sometime in the next 10 or 20 years.

Troublesome though they are, the chlorides are actually one of the lesser problems at the Dutch end of the river. Nitrogen and phosphorus, which come mostly from sewage, boost algae growth to extreme levels and artificially stimulate the food chain—a process called eutrophication. The phenomenon is of concern mainly in the slow-moving Dutch waters, where the algae can become thick enough to hinder shipping and to clog pumps. In addition, when the algae die in the autumn, their decomposition uses up the oxygen necessary to sustain fish and other creatures.

Treatment of sewage, which significantly reduces its nitrogen and phosphorus content, became more consistent in 1959, after the formation of the International Water Conservation Commission for Lake Constance. Since the early 1970s, sewage treatment plants, along with the use of phosphate-free detergents, have steadily reduced the phosphorus in the river. In 1992 an average concentration of 0.21 milligram of phosphorus per liter was measured at the German-Dutch border, which, though an improvement over previous years, is still 10 times higher than the level in Lake Constance.

Nitrogen levels, however, have not fallen significantly, and the reasons are somewhat complicated. Nitrogen in the river has two main forms: ammonium and nitrate. Over the past 20 years, treatment has reduced the ammonium to about 0.27 milligram of nitrogen per



liter. Nitrate levels, on the other hand, have increased and were found in 1992 to be 3.8 milligrams of nitrogen per liter. Most of the nitrate is believed to come from fertilizers used to grow crops along the banks. Hopes that nitrate levels will not increase are pinned to various subsidy programs. For example, faced with a surplus of certain crops, the European Union gives subsidies to farmers who take land out of cultivation. Germany, meanwhile, subsidizes farmers who avoid using fertilizers in the bank zones of the river and who reduce the total use of fertilizers.

Another class of inorganic chemicals of great physiological significance are the heavy metals. In trace concentrations, some of these metals are essential to life. In higher concentrations, they can cause nervous system, growth and metabolic disorders. Mercury, cadmium, copper and some chromium come mainly from metal and chemical plants, which are easily identified and therefore controlled. Nickel, zinc and the rest of the chromium get into the river with the sewage effluent as a result of corrosion of pipes and equipment in homes and industrial plants and so are harder to restrict. Lead has been minimized through its removal from gasoline. The river's minute traces of gold do not endanger human health. (In the early 19th century untold numbers of prospectors panned for the metal along the banks of the upper Rhine, between the Swiss border and Worms in Germany.)

Overall, the amount of these metals in the Rhine has declined by more than 90 percent since the early 1970s. Sewage treatment plants have also helped by immobilizing large amounts in the sludge. In addition, programs have been instituted at industrial dischargers, by which the metals are selectively retained for reuse.

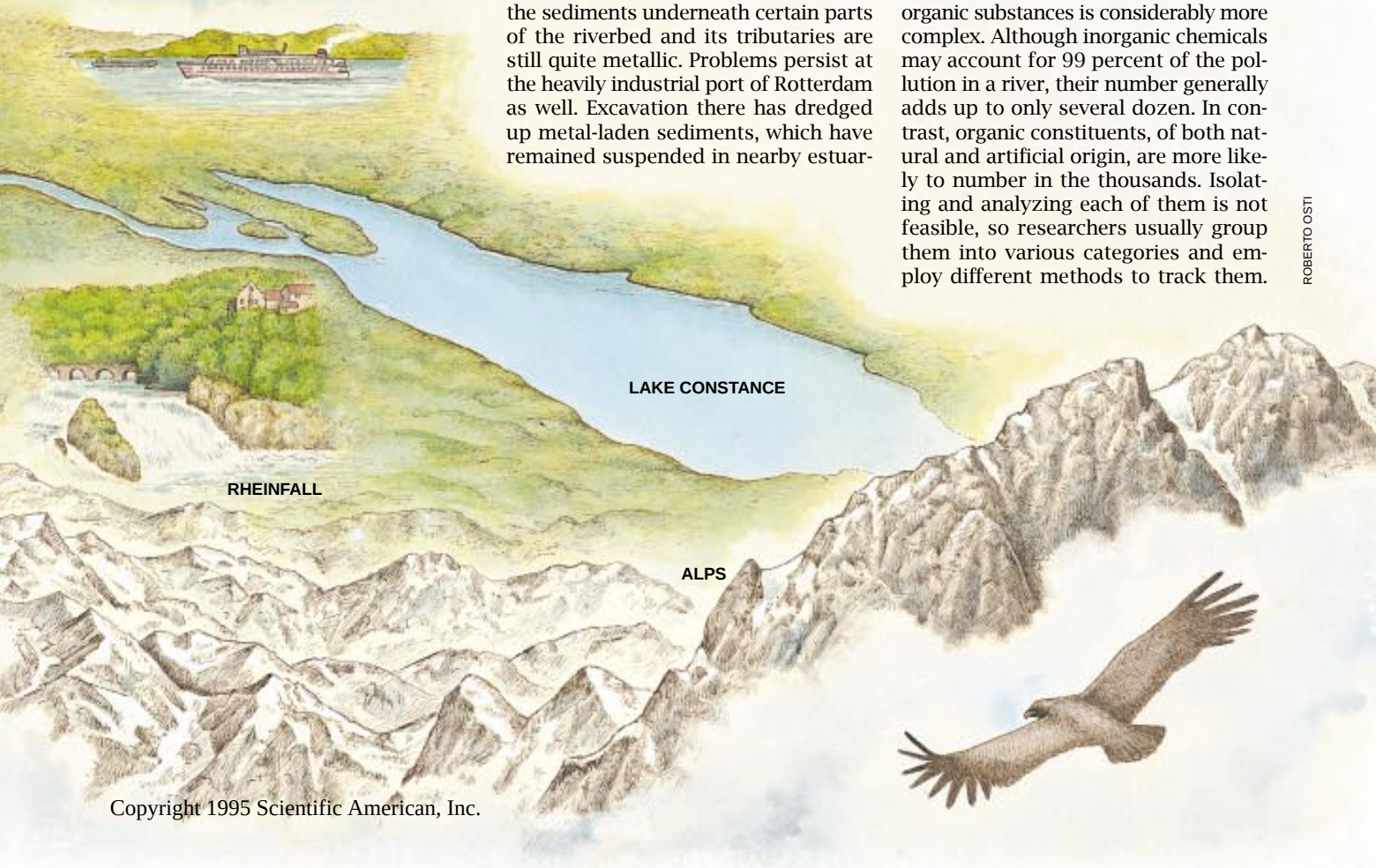
For the most part, the metal content in the Rhine's waters is no longer sufficient to harm people or marine life. But the sediments underneath certain parts of the riverbed and its tributaries are still quite metallic. Problems persist at the heavily industrial port of Rotterdam as well. Excavation there has dredged up metal-laden sediments, which have remained suspended in nearby estuaries. Protracted negotiations between the port authorities and some upstream metal and chemical industries have led to private-law contracts intended to reduce further the amount of metals that are released.

Organic Ogres

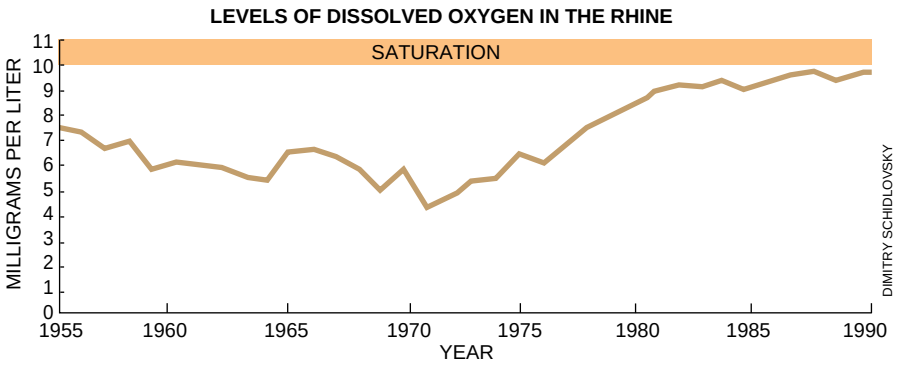
Whereas the monitoring and control of inorganic substances are useful in any river, water quality overall is generally much more sensitive to organic pollutants. Although such organics are usually no more than 1 percent of the pollution in a river, they tend to use up its dissolved oxygen, making the water unfit for sustaining life.

Between 1969 and 1976, organic pollution peaked in the Rhine, frequently sending dissolved oxygen levels below two milligrams per liter in some parts of the middle and lower river during the summer months. Such levels are not high enough to sustain many organisms. Since then, Germany alone has spent some $55 billion on sewage treatment plants, which retain about 90 percent of the organic pollutants. Dissolved oxygen has returned to healthier levels of about nine or 10 milligrams per liter (about 90 percent of the amount the water can physically contain in solution).

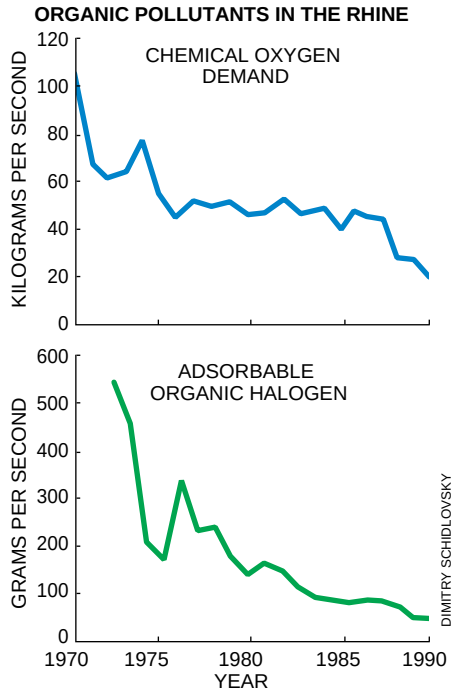
In comparison with the monitoring of inorganic pollutants, keeping tabs on organic substances is considerably more complex. Although inorganic chemicals may account for 99 percent of the pollution in a river, their number generally adds up to only several dozen. In contrast, organic constituents, of both natural and artificial origin, are more likely to number in the thousands. Isolating and analyzing each of them is not feasible, so researchers usually group them into various categories and employ different methods to track them.



ROBERTO OSTI



LIFE-SUSTAINING OXYGEN levels reached a nadir around the summer of 1970, when amounts in certain parts of the Rhine were too low to keep many creatures alive.



OXYGEN CONSUMPTION by organic pollutants, as well as the influx of such compounds containing halogens (mainly chlorine), has fallen dramatically since 1970.

One commonly used technique establishes the effects or characteristics of all organic substances in the water. Another determines the concentrations of groups of similar compounds. Together these techniques can give a useful estimate of the organic state of a body of water and can be supplemented by measurements of single organic substances as needed.

One of the most common measurements in the first category is the biological oxygen demand, typically within a five-day period (abbreviated BOD_5). Bacteria and nutrients are added to the water, and their consumption of oxygen is recorded, generally in milligrams per liter of water. Another good measurement is known as chemical oxygen demand (COD), in which concentrated sulfuric acid and chromium are used to establish the maximum possible oxygen consumption of the sample. In 1992 at the German-Dutch border, the BOD_5 was measured at an average of three milligrams per liter; the COD was 10 milligrams per liter.

Substances commonly grouped together fall into four broad categories: adsorbable organic halogens (AOX, which refers primarily to compounds that contain chlorine); detergents; hydrocarbons; and humic acids. Even as part of larger molecules, chlorine and other halogens are especially worrisome because of their toxicity and persistence. Chlorine compounds come from several sources, including cellulose factories, which bleach raw cellulose to whiten it for making paper. Most of these factories have been converted to low- or even no-chlorine bleaching processes. Chlorine has also entered the river in the form of insecticides, such as DDT (dichlorodiphenyltrichloroethane), HCH (hexachlorocyclohexane), HCB (hexachlorobenzene) and PCP (pentachlorophenol). Germany no longer produces or permits the use of these chemicals. DDT and HCB are similarly banned by other countries along the Rhine, and the use of HCH, PCP and other persistent chlorinated compounds is tightly restricted by these and other countries. Since the mid-1970s, such measures have helped reduce the levels of some organic chlorine compounds at the German-Dutch border by factors between 5 and 15.

Substitution of less persistent pollutants has also been effective in controlling surfactants, the active ingredients of detergents. Since 1964, Germany has permitted only detergents that are readily biodegradable. This restriction has been key in reducing anionic surfactants (the most common kind), as measured in the waters flowing by Düsseldorf, from 650 grams per second in 1964 to 80 grams per second in 1987. In 1992 all measurements for anionic surfactants at the German-Dutch border were below 0.05 milligram per liter.

Hydrocarbons are more readily biodegradable than the halogenated compounds. Petroleum products such as gasoline, kerosene and naphtha account for about 20 percent of all upstream traffic carried on the Rhine across the Dutch border. In addition, the bilges of the thousands of vessels that use the Rhine every year collect some 20,000 cubic meters of an oil-and-water admixture, most of which is collected and removed by special-purpose boats. Such measures are keeping hydrocarbon concentrations down to 0.01 milligram per liter at the Dutch border.

The fourth grouping of organic substances—the humic acids—are the short-term products of biodegradation, one of the most potent tools of reclaimers. Yet whether it is induced in a sewage treatment plant or takes place naturally in the river itself, biodegradation is a never-ending process. Only part of the organic substances consumed by bacteria is fully metabolized and respirated as carbon dioxide. The rest are only partly oxidized and thereby converted into humic acids. Although produced in sewage treatment plants, they have always been in the river. In 1973 humic acids were estimated to account for about 25 percent of the residual organic pollution of the Rhine. More recently, this fraction has increased, although the percentage is difficult to determine because of the lack of analytical methods. Humic acids are considered essentially harmless, however, because they are produced early on in the natural, gradual oxidation of organic material in any river.

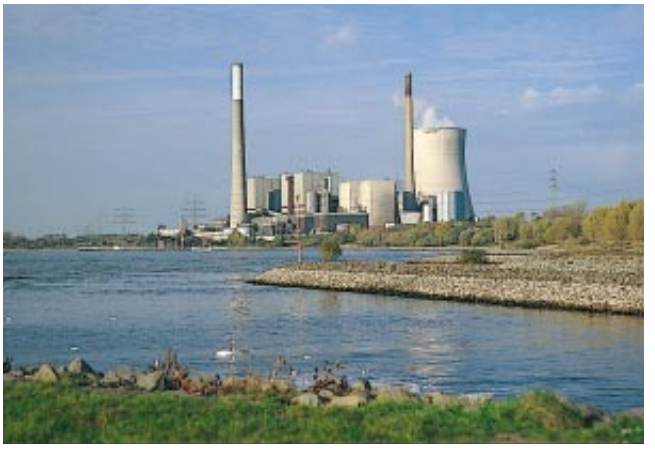
A Seminal Accident

No matter how carefully certain pollutants are controlled, the Rhine, like any heavily trafficked waterway, remains vulnerable to the occasional accident. At one time, accidents went mostly unnoticed amid the high background level of pollution. Today's cleaner river, however, reacts more profoundly to such events, and a sensitive early-warning system has been put in place to alert authorities when accidents occur.

A few unfortunate episodes were pivotal in getting the monitoring system up and running. In one of the worst, on November 1, 1986, a Sandoz Ltd. warehouse full of pesticides caught fire near Basel. Water sprayed on the fire washed the chemicals into the river,

THOMAS MAYER *Das Fotoarchiv*

UNHEALTHFUL FROTH and even dead creatures were periodic appurtenances of the river Rhine as late as the 1980s.



JOCHEN TACK *Das Fotoarchiv*

Today the same region, seen in front of the Voerde power plant in the industrial Ruhr area of Germany, is much cleaner.

where one of them—disulfoton—proved especially toxic to eels. Many thousands were killed downstream, all the way to Karlsruhe.

The disaster triggered an effort of unprecedented scope to follow the Rhine's recovery and to assess the biological state of the river in general. During the project, biologists from a German government research institute used a diving bell to study fish, macroinvertebrates and other creatures systematically in the riverbed. To the surprise of many, the river's fauna completely recovered by October 1988, less than two years after the fire.

The survey documented 155 species of macroinvertebrates, which tend to cluster near the banks, between Basel and Düsseldorf. Some of the most common were freshwater sponges, leeches, zebra mussels, benthic amphipods, mayflies, caddis flies and worms, and chironomid larvae. Some species, such as the mayfly *Ephoron virgo* and the snail *Theodosus fluviatilis,* had been thought to be virtually extinct in the Rhine but were found in large quantities. Other once common species, such as the mussel *Spaericum solidum* or the stonefly *Euleuctea geniculata,* were present as individual specimens and are only now starting to reestablish themselves in greater numbers. Somewhat disconcertingly, however, an amphipod, *Corophium curvispinum,* a relatively recent arrival from the Caspian and Black seas, is proliferating extensively enough to drive out certain species of sponges and mollusks.

Some fish, too, are making a comeback. Of the 47 endemic species known to have inhabited the river, a recent survey found 40. About 75 percent of the individual fish identified were hardy, unspecialized creatures, including roach, bleak and bream. Researchers also spotted carp, perch, eel, pike, chub and dace. In addition, they tallied 15 species that had been introduced into the river, including pike-perch, rainbow trout and sunfish. In 1992, for the first time in decades, sexually mature salmon (*Salmo salar*) were caught in the Rhine. Released as hatchlings into tributaries of the river a year or more previously, they had survived a migration to the North Sea. Even sturgeon—believed extinct from the river for 40 years—have been seen occasionally.

A Model River

The river survey was not the only legacy of the 1986 warehouse fire in Basel. The states bordering the Rhine launched a joint Rhine Action Program, with four objectives: long-term safeguarding of the drinking water; decontamination of sediments; reestablishment of higher species of fish (salmon and so on); and protection of the North Sea.

As a first step, in 1989 the program compiled an inventory of all discharges of 30 different hazardous substances. By 1995 the rates of discharge for all had been reduced by 50 percent or more. In addition, improved safeguards prevent or limit discharges into the river after industrial accidents. In coming years, the construction of fish ladders to help creatures such as salmon get back to their upstream spawning grounds and the improvement of those grounds will begin to restore runs.

In 1991, at a conference on the waterworks in the Rhine's catchment area, the Dutch minister for transport and public works, Hanja Maij-Weggen, called for the experience acquired on the Rhine to be applied to the Meuse and Scheldt rivers. With the end of the cold war, international commissions have been set up to reclaim the Elbe, which flows from the Czech Republic and Poland through Germany to the North Sea, and the Oder, which forms part of the border between Poland and Germany. Even the Volga has benefited from lessons learned on the Rhine, which were passed on from German experts to their Russian counterparts in a recent series of meetings. For the Danube, reclamation will have to wait: work on the river has been suspended because of the war in the former Yugoslavia.

Amid sweeping changes in Europe it is fitting that the Rhine has become a tribute to the great things that can happen when countries cooperate.

The Author

KARL-GEERT MALLE recently retired from BASF in Ludwigshafen, Germany. After joining that company in 1960, he worked in inorganic research and in various production units before taking on the responsibility of monitoring all wastewater discharges. In 1984 he headed special tasks in environmental protection and was appointed a director. He is a member of a working committee of the German Commission for the Protection of the Rhine and of the Commission for the Evaluation of Water-Endangering Substances. He lectures on ecological chemistry at the Technical College in Mannheim.

Further Reading

HEAVY METAL POLLUTION IN THE RHINE BASIN. William M. Stigliani, Peter R. Jaffé and Stefan Anderberg in *Environmental Science and Technology,* Vol. 27, No. 5, pages 786–793; May 1993.

REHABILITATION OF THE RIVER RHINE. Edited by J.A. van de Kraats. Special Issue of *Water Science and Technology,* Vol. 29, Issue 3; 1994.

The Evolution of Continental Crust

The high-standing continents owe their existence to the earth's long history of plate-tectonic activity

by S. Ross Taylor and Scott M. McLennan

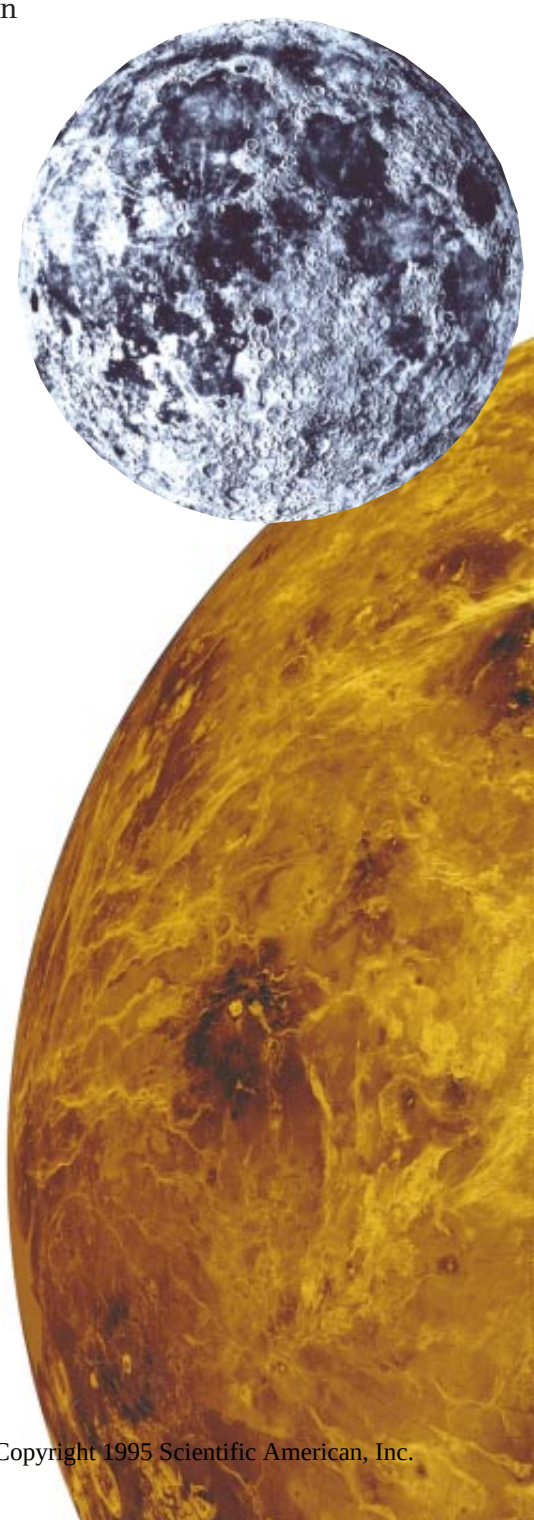
Except perhaps for some remote island dwellers, most people have a natural tendency to view continents as fundamental, permanent and even characteristic features of the earth. One easily forgets that the world's continental platforms amount only to scattered and isolated masses on a planet that is largely covered by water. But when viewed from space, the correct picture of the earth becomes immediately clear. It is a blue planet. From this perspective it seems quite extraordinary that over its long history the earth could manage to hold a small fraction of its surface always above the sea, enabling, among other things, human evolution to proceed on dry land.

Is the persistence of high-standing continents just fortuitous? How did the earth's complicated crust come into existence? Has it been there all the time, like some primeval icing on a planetary cake, or has it evolved through the ages? Such questions had engendered debates that divided scientists for many decades, but the fascinating story of how the terrestrial surface came to take its present form is now essentially resolved. That understanding shows, remarkably enough, that the conditions required to form the continents of the earth may be unmatched in the rest of the solar system.

The earth and Venus, being roughly the same size and distance from the sun, are often regarded as twin planets. So it is natural to wonder how the crust of Venus compares with that of our own world. Although centuries of telescopic observations from the earth could give no insight, beginning in 1990 the *Magellan* space probe's orbiting radar penetrated the thick clouds that enshroud Venus and revealed its surface with stunning clarity. From the detailed images of landforms, planetary scientists can surmise the type of rock that covers Venus.

Comparison with the Neighbors

Our sister planet appears to be blanketed by rock of basaltic composition—much like the dark, fine-grained rocks that line the ocean basins on the earth. *Magellan*'s mapping, however, failed to find extensive areas analogous to the terrestrial continental crust. Elevated regions named Aphrodite Terra and Ishtar Terra appear to be remnants of crumpled basaltic lavas. Smaller, dome-shaped mounds are found on Venus, and these forms might indicate that a substantially different bedrock composition does exist in some places;



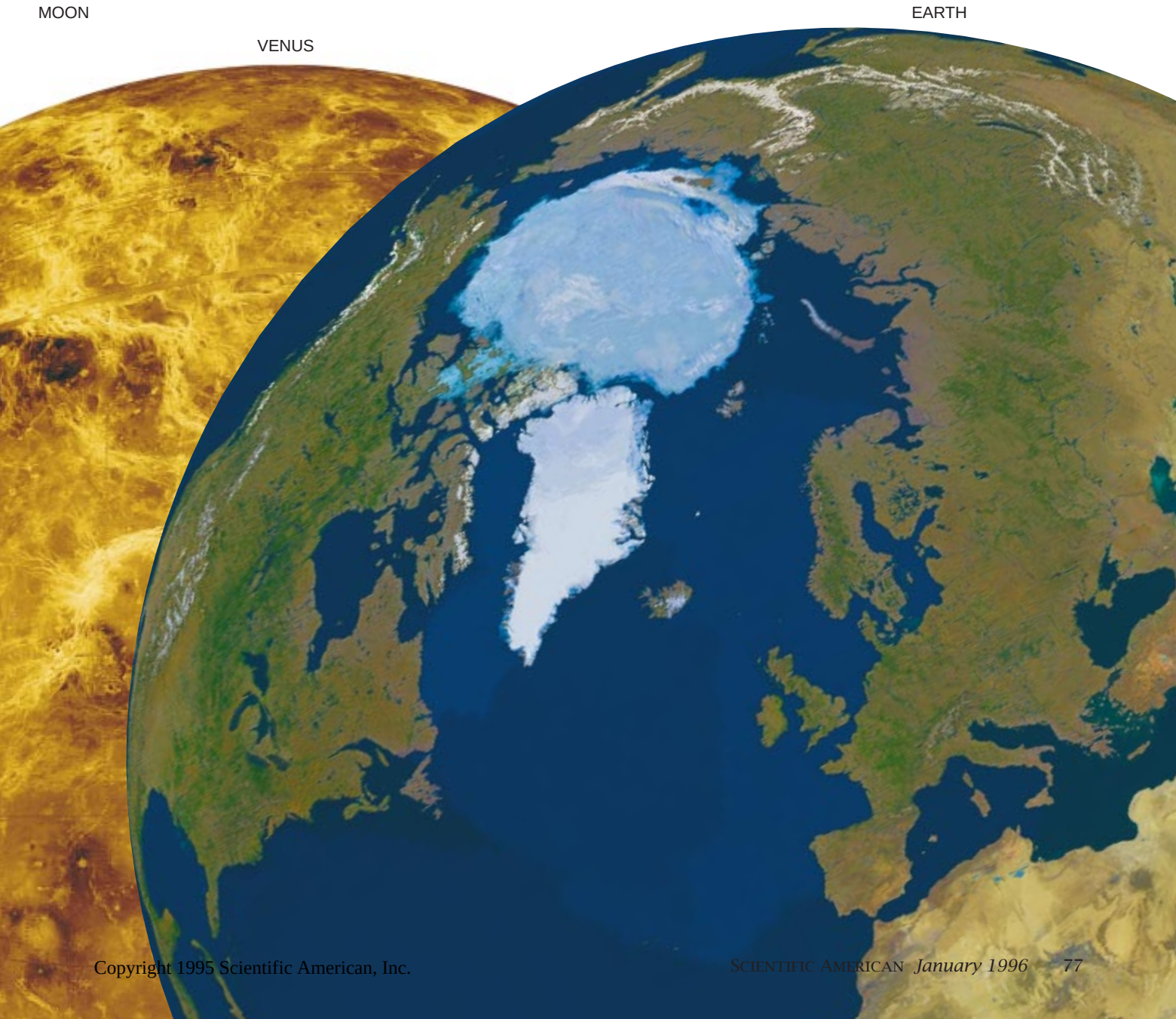
EARTH'S CRUST is composed primarily of the basaltic rocks that line the ocean basins. Granitic rocks constitute the high-standing continental platforms. Venus is nearly the same size as the earth, yet radar imagery indicates that it is encrusted almost entirely by basalt. Only a tiny fraction of that planet's surface exhibits pancake-shaped plateaus (*detail above*) that might, like the earth's continents, be built of granitic material. The crust of the earth's moon is largely covered by white highlands formed as that body first cooled from a molten state; volcanic eruptions later created dark so-called seas of basalt.

it is also possible that these pancake-like features may be composed merely of more basalt.

After analyzing the wealth of radar data provided by *Magellan,* scientists have concluded that plate tectonics (that is, the continual creation, motion and destruction of parts of the planet's surface) does not seem to operate on Venus. There are no obvious equivalents to the extensive mid-ocean ridges or to the great trench systems of the earth. Thus, it is unlikely that the crust of Venus regularly recycles back into that planet's mantle. Nor would there seem to be much need to make room for new crust: the amount of lava currently erupting on Venus is roughly equivalent to the output of one Hawaiian volcano, Kilauea—a mere dribble for the planet as a whole.

These findings from Venus and similar surveys of other solid bodies in the solar system show that planetary crusts can be conveniently divided into three fundamental types. So-called primary crusts date back to the beginnings of the solar system. They emerged after large chunks of primordial material came crashing into a growing planet, releasing enough energy to cause the original protoplanet to melt. As the molten rock began to cool, crystals of some types of minerals solidified relatively early and could separate from the body of magma. This process, for example, probably created the white highlands of the moon after low-density grains of the mineral feldspar floated to the top of an early lunar "ocean" of molten basalt. The crusts of many satellites of the giant outer planets, composed of mixtures of rock with water, methane and ammonia ices, may also have arisen from catastrophic melting during initial accretion.

In contrast to the product of such sudden, large-scale episodes of melting, secondary crusts form after heat from the decay of radioactive elements gradually accumulates within a planetary body. Such slow heating causes a small fraction of the planet's rocky interior to melt and usually results in the eruption of basaltic lavas. The surfaces of Mars and Venus and the earth's ocean floors are covered by secondary crusts created in this way. The lunar maria (the "seas" of the ancient astronomers) also formed from basaltic lavas that originated deep in the moon's interior. Heat from radioactivity—or perhaps from the flexing induced by tidal forces—on



JET PROPULSION LABORATORY/NASA (*detail*); JPL/NASA (*Venus*); TOM VAN SANT *The Geosphere Project* (*Earth*); NASA (*Moon*)

some icy moons of the outer solar system may, too, have generated secondary crusts.

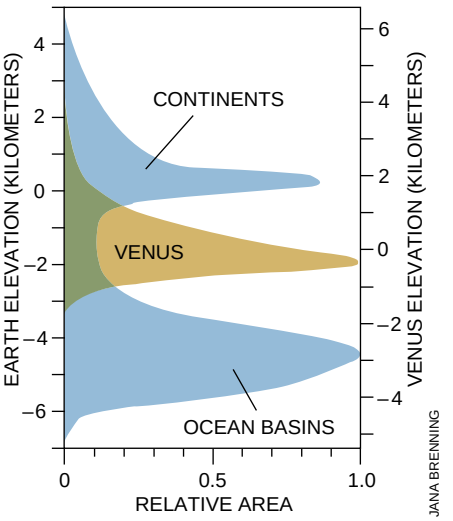
Unlike these comparatively common types, so-called tertiary crust may form if surface layers are returned back into the mantle of a geologically active planet. Like a form of continuous distillation, volcanism can then lead to the production of highly differentiated magma of a composition that is distinct from basalt—closer to that of the light-colored igneous rock granite. Because the recycling necessary to generate granitic magmas can occur only on a planet where plate tectonics operates, such a composition is rare in the solar system. The formation of continental crust on the earth may be its sole demonstration.

Despite the small number of examples within each category, one generalization about the genesis of planetary surfaces seems easy to make: there are clear differences in the rates at which primary, secondary and tertiary crusts form. The moon, for instance, generated its white, feldspar-rich primary crust—about 12 percent of lunar volume—in only a few million years. Secondary crusts evolve much more slowly. The moon's basalt maria (secondary crust) are just a few hundred meters thick and make up a mere one tenth of 1 percent of the moon's volume, and yet these so-called seas required more than a billion years to form. Another example of secondary crust, the basaltic oceanic basins of our planet (which constitute about one tenth of 1 percent of the earth's mass) formed over a period of about 200 million years. Slow as these rates are, the creation of tertiary crust is even less efficient. The earth has taken several billion years to produce its tertiary crust—the continents. Yet these features amount to just about one half of 1 percent of the mass of the planet.

Floating Continents

Many elements that are otherwise rarely found on the earth are enriched in granitic rocks, and this phenomenon gives the continental crust an importance out of proportion to its tiny mass. But geologists have not been able to estimate the overall composition of crust—a necessary starting point for any investigation of its origin and evolution—by direct observation. One conceivable method might be to compile existing descriptions of rocks that outcrop at the surface. But even this large body of information might well prove insufficient. A large-scale exploration program that could reach deeply enough into the crust for a meaningful sample would press the limits of modern drilling technology and would, in any event, be prohibitively expensive.

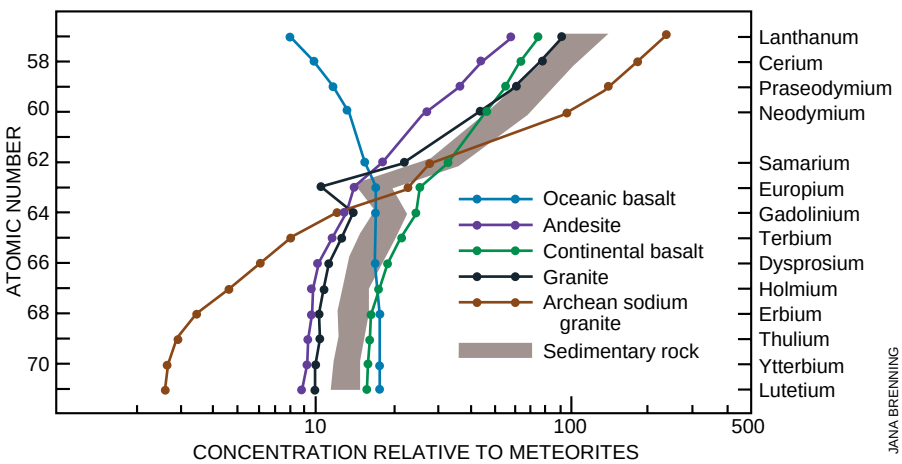
Fortunately, a simpler solution is at hand. Nature has already accomplished a widespread sampling through the erosion and deposition of sediments. Lowly muds, now turned into solid rock, give a surprisingly good average composition for the exposed continental crust. These samples are, however, missing those elements that are soluble in water, such as sodium and calcium. Among the insoluble elements that are transferred from the crust into sediments without distortion in their relative abundances are the 14 rare-earth elements, known to geochemists as REEs. These elemental tags are uniquely useful in deciphering crustal composition because their atoms do not fit neatly into the crystal structure of most common minerals. They tend instead to be concentrated in the late-forming granitic products of a cooling magma that make up most of the continental crust.



SURFACE ELEVATIONS are distributed quite differently on the earth (*blue*) and on Venus (*gold*). Most places on the earth stand near one of two prevailing levels. In contrast, a single height characterizes most of the surface of Venus. (Elevation on Venus is given with respect to the planet's mean radius.)

Because the REE patterns found in a variety of sediments are so similar, geochemists surmise that weathering, erosion and sedimentation must mix different igneous source rocks efficiently enough to create an overall sample of the continental crust. All the members of the REE group establish a signature of upper crustal composition and preserve, in the shapes of the elemental abundance patterns, a record of igneous events that may have influenced the makeup of the crust.

Using these geochemical tracers, geologists have, for example, determined that the composition of the upper part of the continental crust approximates that of granodiorite, an ordinary igneous rock that consists largely of light-colored quartz and feldspar, along with a peppering of various dark minerals. Deep within the continental crust, below about 10 to 15 kilometers, rock of a more basaltic composition is probably common. The exact nature of this material remains controversial, and geologists are currently testing their ideas using measurements of the heat produced within the crust by the important radioactive elements potassium, uranium and thorium. But it seems reasonable that at least parts of this inac-



RARE-EARTH ELEMENT abundance patterns provide characteristic chemical markers for the types of rock that have formed the earth's crust. Although igneous rocks (those that solidify from magma) can have highly variable rare-earth element signatures (*dotted lines*), the pattern for most sedimentary rocks falls within a narrow range (*gray band*). That uniformity arises because sediments effectively record the average composition of the upper continental crust.

cessible and enigmatic region may consist of basalt trapped and underplated beneath the lower-density continents.

It is this physical property of granitic rock—low density—that explains why most of the continents are not submerged. Continental crust rises on average 125 meters above sea level, and some 15 percent of the continental area extends over two kilometers in elevation. These great heights contrast markedly with the depths of ocean floors, which average about four kilometers below sea level—a direct consequence of their being lined by dense oceanic crust composed mostly of basalt and a thin veneer of sediment.

At the base of the crust lies the so-called Mohorovicic discontinuity (a tongue-twisting name geologists invariably shorten to "Moho"). This deep surface marks a radical change in composition to an extremely dense rock rich in the mineral olivine that everywhere underlies both oceans and continents. Geophysical studies using seismic waves have traced the Moho worldwide. Such research has also indicated that the mantle below the continents may be permanently attached at the top. These relatively cool subcrustal "keels" can be as much as 400 kilometers thick and appear to ride with the continents during their plate-tectonic wanderings. Support for this notion comes from the analysis of tiny mineral inclusions found within diamonds, which are thought to originate deep in this subcrustal region. Measurements show that diamonds can be up to three billion years old and thus demonstrate the antiquity of the deep continental roots.

It is curious to reflect that less than 40 years ago, there was no evidence that the rocks lining ocean basins differed in any fundamental way from those found on land. The oceans were simply thought to be floored with foundered or sunken continents. This perception grew naturally enough from the concept that the continental crust was a world-encircling feature that had arisen as a kind of scum on an initially molten planet. Although it now appears certain that the earth did in fact melt very early, it seems that a primary granitic crust, of the type presumed decades ago, never actually existed.

The Evolution of Geodiversity

How was it that two such distinct kinds of crust, continental and oceanic, managed to arise on the earth? To answer this question, one needs to consider the earliest history of the solar system. In the region of the primordial solar nebula occupied by the earth's orbit, gas was mostly swept away, and only rocky debris large enough to survive intense early solar activity accumulated. These objects themselves must have grown by accretion, before finally falling together to form our planet, a process that required about 50 million to 100 million years.

Late in this stage of formation, a massive planetesimal, perhaps one the size of Mars, crashed into the nearly fully formed earth. The rocky mantle of the impactor was ejected into orbit and became the moon while the metallic core of the body fell into the earth [see "The Scientific Legacy of Apollo," by G. Jeffrey Taylor; SCIENTIFIC AMERICAN, July 1994]. As might be expected, this event proved catastrophic: it totally melted the newly formed planet. As the earth later cooled and solidified, an early basaltic crust probably formed.

It is likely that at this stage the surface of the earth resembled the current appearance of Venus. None of this primary crust has, however, survived. Whether it sank into the mantle in a manner similar to to that taking place on the earth or piled up in localized masses until it was thick enough to transform into a denser rock and sink remains uncertain. In any event, there is no evidence of substantial granitic crust at this early stage. Telltale evidence of such a crust should have survived in the form of scattered grains of the mineral zircon, which forms within granite and is enormously resistant to erosion. Although a few ancient zircons dating from near this time have been

MICHAEL GOODMAN

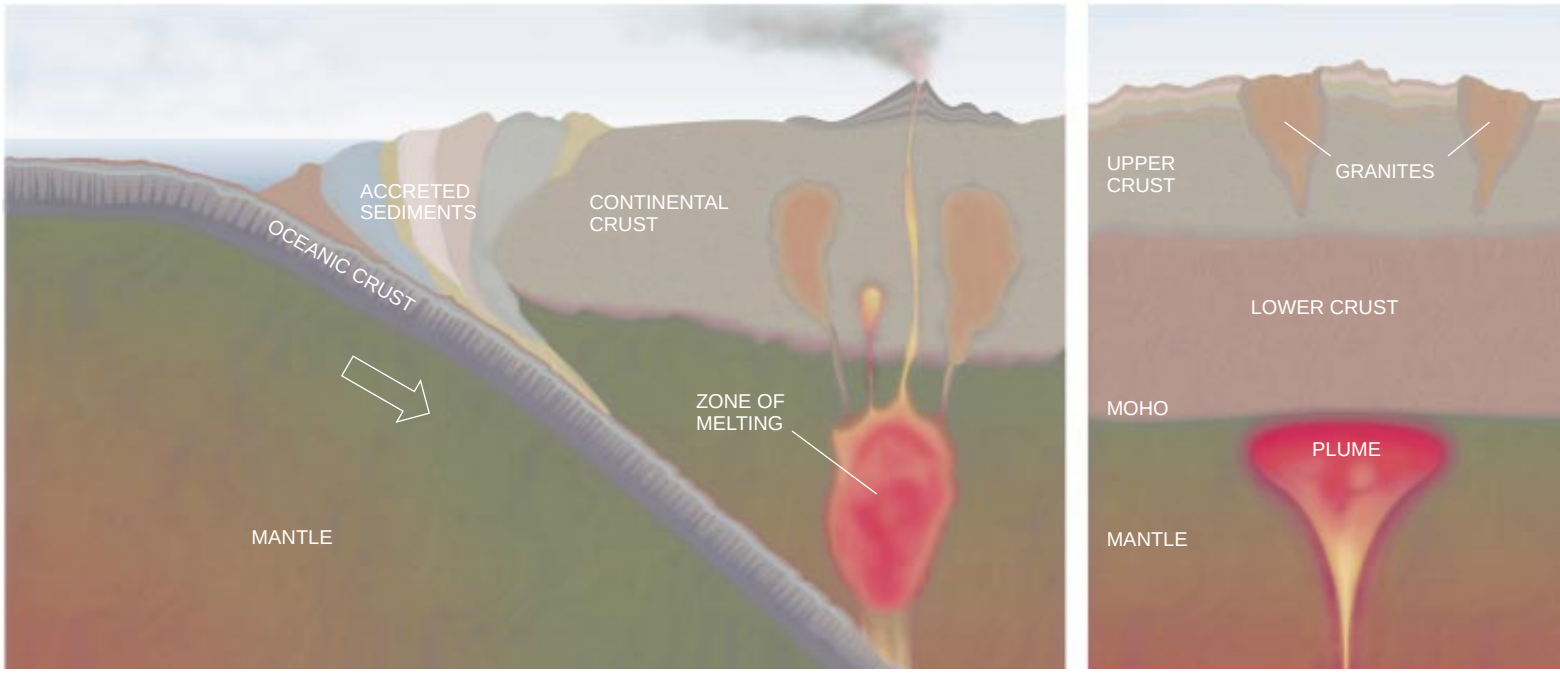
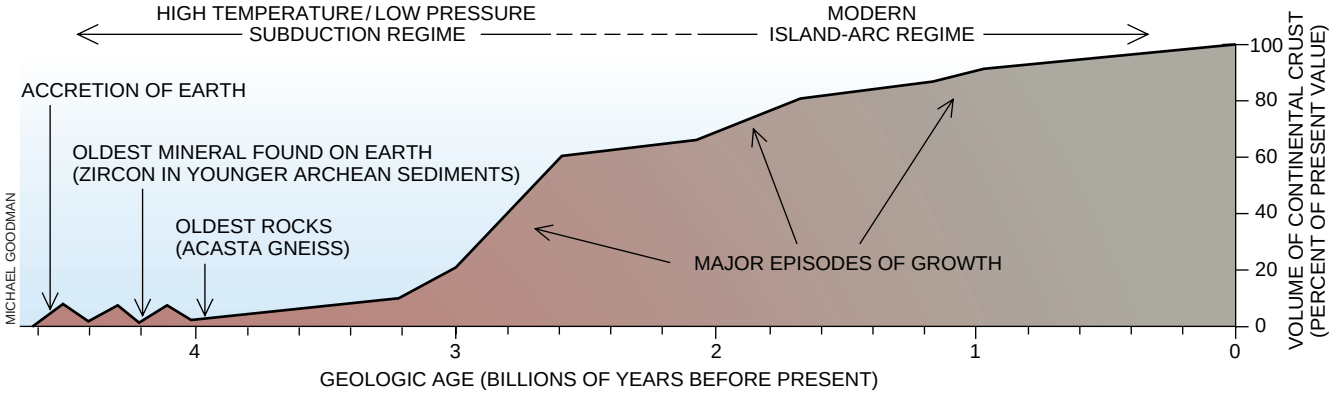


PLATE-TECTONIC ACTIVITY carries oceanic crust deep into the earth (*left*), burying wet sediments along with the descending slab. At a depth of 80 kilometers, high temperatures drive water from the sediments, inducing the overlying rock to melt. The magma generated in this process rises buoyantly and forms new continental material near the surface. As this crust matures (*right*), heat from radioactivity (or from rising plumes of basaltic magma) may subsequently trigger melting at shallow levels. Such episodes create an upper layer that is made up largely of granite.



CRUSTAL GROWTH has proceeded in episodic fashion for billions of years. An important growth spurt lasted from about 3.0 to 2.5 billion years ago, the transition between the Archean and Proterozoic eons. Widespread melting at this time formed the granite bodies that now constitute much of the upper layer of the continental crust.

found (the oldest examples are from sedimentary rocks in Australia and are about 4.2 billion years old), these grains are exceedingly scarce.

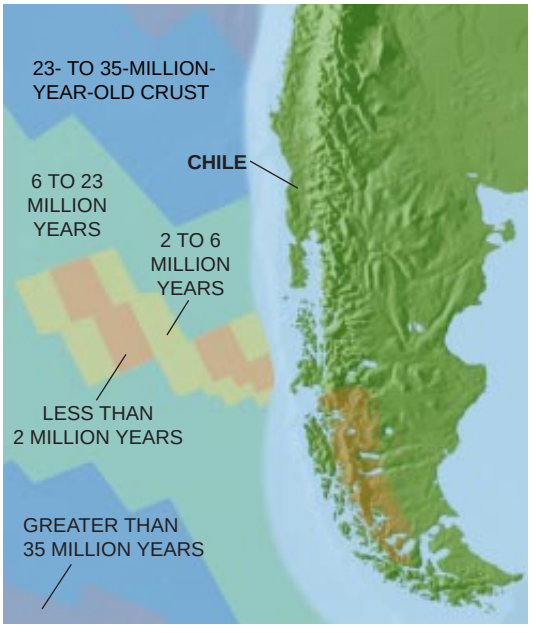
More information about the early crust comes from the most ancient rocks to have survived intact. These rocks formed deep within the crust just less than four billion years ago and now outcrop at the surface in northwest Canada. This rock formation is called the Acasta Gneiss. Slightly younger examples of early crust have been documented at several locations throughout the world, although the best studied of these ancient formations is in western Greenland. The abundance of sedimentary rock there attests to the presence of running water and to the existence of what were probably true oceans during this remote epoch. But even these extraordinarily old rocks from Canada and Greenland date from some 400 million to 500 million years after the initial accretion of the earth, a gap in the geologic record caused, no doubt, by massive impacts that severely disrupted the earth's earliest crust.

From the record preserved in sedimentary rocks, geologists know that the formation of continental crust has been an ongoing process throughout the earth's long history. But the creation of crust has not always had the same character. For example, at the boundary between the Archean and Proterozoic eons, around 2.5 billion years ago, a distinct change in the rock record occurs. The composition of the upper crust before this break contained less evolved constituents, composed of a mixture of basalt and sodium-rich granites. These rocks make up the so-called tonalite-trondjemite-granodiorite, or TTG suite. This composition differs considerably from the present upper crust, which is dominated by potassium-rich granites.

The reason for the profound change in crustal composition 2.5 billion years ago appears to be linked to the earth's plate-tectonic churnings. Before this time, oceanic crust was being recycled rapidly because higher levels of radioactivity existed and more intense heating tended to drive a faster plate-tectonic engine. There may have been more than 100 separate plates during the Archean, compared with the dozen that currently exist. Unlike the present oceanic crust—which travels a long distance and cools significantly before sinking back into the mantle—oceanic crust during this earlier epoch survived for a far shorter time. It was thus still relatively hot when it entered the mantle and so began to melt at shallower depths than typically occurs now. This difference accounts for the formation of sodium-rich igneous rocks of the TTG suite; such rocks now form only in those few places where young, hot oceanic crust subducts.

The early tendency for magma to form with a TTG composition explains why crust grew as a mixture of basalt and tonalite during the Archean eon. Large amounts—at least 50 percent and perhaps as much as 70 percent of the continental crust—emerged at this time, with a major episode of growth between 3.0 and 2.5 billion years ago. Since that time, the relative height of ocean basins and continental platforms has remained comparatively stable. With the onset of the Proterozoic eon 2.5 billion years

Snapshot of the Past



Sodium-rich granites are characteristic of the Archean eon, a time when the earth's plate-tectonic engine operated more quickly than it does today. Such tonalite-trondjemite-granodiorite (TTG) granites form only if young (and hence hot) oceanic crust enters the mantle, sparking melting at shallow levels. These conditions still exist along the southern coast of Chile, where relatively new oceanic crust has subducted underneath the South American plate. The region containing TTG granites is marked in brown.

ago, the crust had already assumed much of its present makeup, and modern plate-tectonic cycling began.

Currently oceanic crust forms by the eruption of basaltic lava along a globe-encircling network of mid-ocean ridges. More than 18 cubic kilometers of rock are produced every year by this process. The slab of newly formed crust rides on top of an outer layer of the mantle, which together make up the rigid lithosphere. The oceanic lithosphere sinks back into the mantle at so-called subduction zones, which leave conspicuous scars on the ocean floor in the form of deep trenches. At these sites the descending slab of lithosphere carries wet marine sediments as well as basalt plunging into the mantle.

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

VOLCANOES, such as this one erupting on the Kamchatka Peninsula, mark where new continental material forms above subducting oceanic crust. In time, such geologically active terranes will evolve into stable continental crust.

At a depth of about 80 kilometers, heat drives water and other volatile components from the subducted sediments into the overlying mantle. These substances then act as a flux does at a foundry, inducing melting in the surrounding material at reduced temperature. The magma produced in this fashion eventually reaches the surface, where it causes spectacular, explosive eruptions. Pinatubo and Mount St. Helens are two recent examples of such geologic cataclysms. Great chains of volcanoes—such as the Andes—powered by boiling volatiles add on average about two cubic kilometers of lava and ash to the continents every year.

But subduction-induced volcanism is not the only source of new granitic rock. The accumulation of heat deep within the continental crust itself can cause melting, and the resultant magma will ultimately migrate to the surface. Although some of this necessary heat might come from the decay of radioactive elements, a more likely source is basaltic magma that rises from deeper in the mantle and becomes trapped under the granitic lid; the molten rock then acts like a burner under a frying pan.

Crustal Growth Spurts

Although the most dramatic shift in the generation of continental crust happened at the end of the Archean eon, 2.5 billion years ago, the continents appear to have experienced episodic changes throughout all of geologic time. For example, sizable, later additions to the continental crust occurred from 2.0 to 1.7, from 1.3 to 1.1 and from 0.5 to 0.3 billion years ago. That the earth's continents experienced such a punctuated evolution might appear at first to be counterintuitive. Why, after all, should crust form in spurts if the generation of internal heat—and its liberation through crustal recycling—is a continuous process?

A more detailed understanding of plate tectonics helps to solve this puzzle. During the Permian period (about 250 million years ago), the major continents of the earth converged to create one enormous landmass called Pangea [see "Earth before Pangea," by Ian W. D. Dalziel; SCIENTIFIC AMERICAN, January 1995]. This configuration was not unique. The formation of such "supercontinents" appears to recur at intervals of about 600 million years. Major tectonic cycles driving the continents apart and together have been documented as far back as the Early Proterozoic, and there are even suggestions that the first supercontinent may have formed earlier, during the Archean.

Such large-scale tectonic cycles serve to modulate the tempo of crustal growth. When a supercontinent breaks itself apart, oceanic crust is at its oldest and hence most likely to form new continental crust after it subducts. As the individual continents reconverge, volcanic arcs (curved chains of volcanoes created near subduction zones) collide with continental platforms. Such episodes preserve new crust as the arc rocks are added to the margins of the continents.

For more than four billion years, the peripatetic continents have assembled themselves in fits and starts from many disparate terranes. Buried in the resulting amalgam is the last remaining testament available for the bulk of the earth's history. That story, assembled from rocks that are like so many jumbled pieces of a puzzle, has taken some time to sort out. But the understanding of crustal origin and evolution is now sufficient to show that of all the planets the earth appears truly exceptional. By a fortunate accident of nature—the ability to maintain plate-tectonic activity—one planet alone has been able to generate the sizable patches of stable continental crust that we find so convenient to live on.

The Authors

S. ROSS TAYLOR and SCOTT M. MCLENNAN have worked together since 1977 examining the earth's crustal evolution. Taylor has also actively pursued lunar and planetary studies and has published four books on planetology. He is a foreign associate of the National Academy of Sciences. Taylor is currently with the Research School of Physical Sciences at the Australian National University. McLennan is a professor in the department of earth and space sciences at the State University of New York at Stony Brook. He recently received a Presidential Young Investigator Award from the National Science Foundation. His research applies the geochemistry of sedimentary rocks to studies of crustal evolution, plate tectonics and ancient climates.

Further Reading

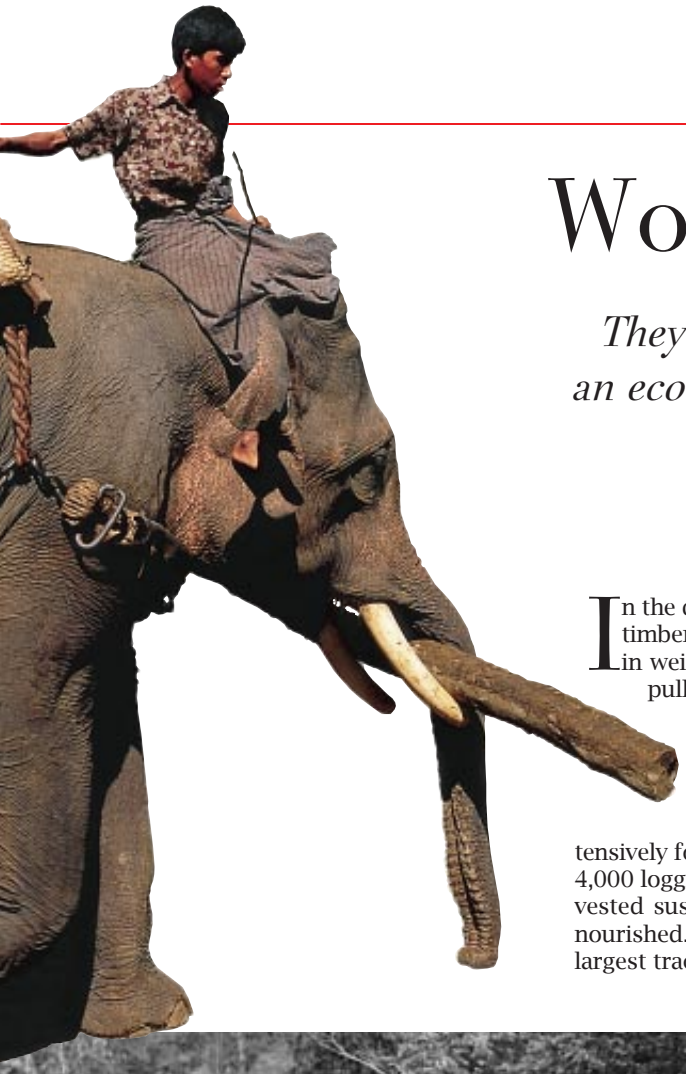
NO WATER, NO GRANITES—NO OCEANS, NO CONTINENTS. I. H. Campbell and S. R. Taylor in *Geophysical Research Letters,* Vol. 10, No. 11, pages 1061–1064; November 1983.

THE CONTINENTAL CRUST: ITS COMPOSITION AND EVOLUTION. S. Ross Taylor and Scott M. McLennan. Blackwell Scientific Publications, 1985.

OASIS IN SPACE: EARTH HISTORY FROM THE BEGINNING. Preston Cloud. W. W. Norton, 1988.

THE GEOCHEMICAL EVOLUTION OF THE CONTINENTAL CRUST. S. Ross Taylor and Scott M. McLennan in *Reviews of Geophysics,* Vol. 33, No. 2, pages 241–265; May 1995.

NATURE AND COMPOSITION OF THE CONTINENTAL CRUST: A LOWER CRUSTAL PERSPECTIVE. Roberta L. Rudnick and David M. Fountain in *Reviews of Geophysics,* Vol. 33, No. 3, pages 267–309; August 1995.



Working Elephants

They earn their keep in Asia by providing an ecologically benign way to harvest forests

by Michael J. Schmidt
Photography by Richard Ross

In the dense forests of Myanmar, men and elephants labor together to extract timber as they have done for more than 100 years. A gesture, a word, a shift in weight is all it takes for an "oozie" to direct his elephant to carry, push, pull or stack massive logs that elsewhere are manipulated by machines. If kept viable, the tradition can guarantee the survival not only of Asian elephants but also of the forests. If the tradition is lost, a magnificent species might become extinct, and some of the oldest natural forests in Asia could become monocultured tree farms.

Myanmar, formerly Burma, is the last country to use elephants extensively for logging. Just two decades ago Thailand had a vigorous population of 4,000 logging elephants. But the Thai forests were clear-cut instead of being harvested sustainably, and now many of the beasts are underemployed and malnourished. Only in Myanmar has a reverence for heritage allowed some of the largest tracts of forest on the earth to flourish unspoiled. A century-old policy of

harvesting selected trees and transporting the logs by teams of men and elephants has kept vast sections of forest robust and highly productive.

Elephants offer several ecological benefits over machines. They enter and exit the forest on narrow paths, obviating the need for an extensive network of logging roads. These roads damage far more than the ecosystem around the roadbeds. They provide access for humans who engage in slash-and-burn agriculture, leaving behind infertile, bare earth. Logging roads also allow poachers to penetrate deep within a forest, where they devastate endangered creatures. Elephants run on renewable and cheap green fuel, whereas mechanized log skidders require expensive, polluting petroleum. And the log skidders roar through the forest on tank treads, crushing everything in their path. The vegetation never recovers.

Although man-elephant teams are ecologically benign, they are not as efficient as mechanical devices—which many foreign timber companies are eager to fund. The elephants can make up the difference only if enough of them are employed to get the required thousands of logs out of the forests every year. Currently 5,700 elephants are in use; the challenge is to keep their numbers that high.

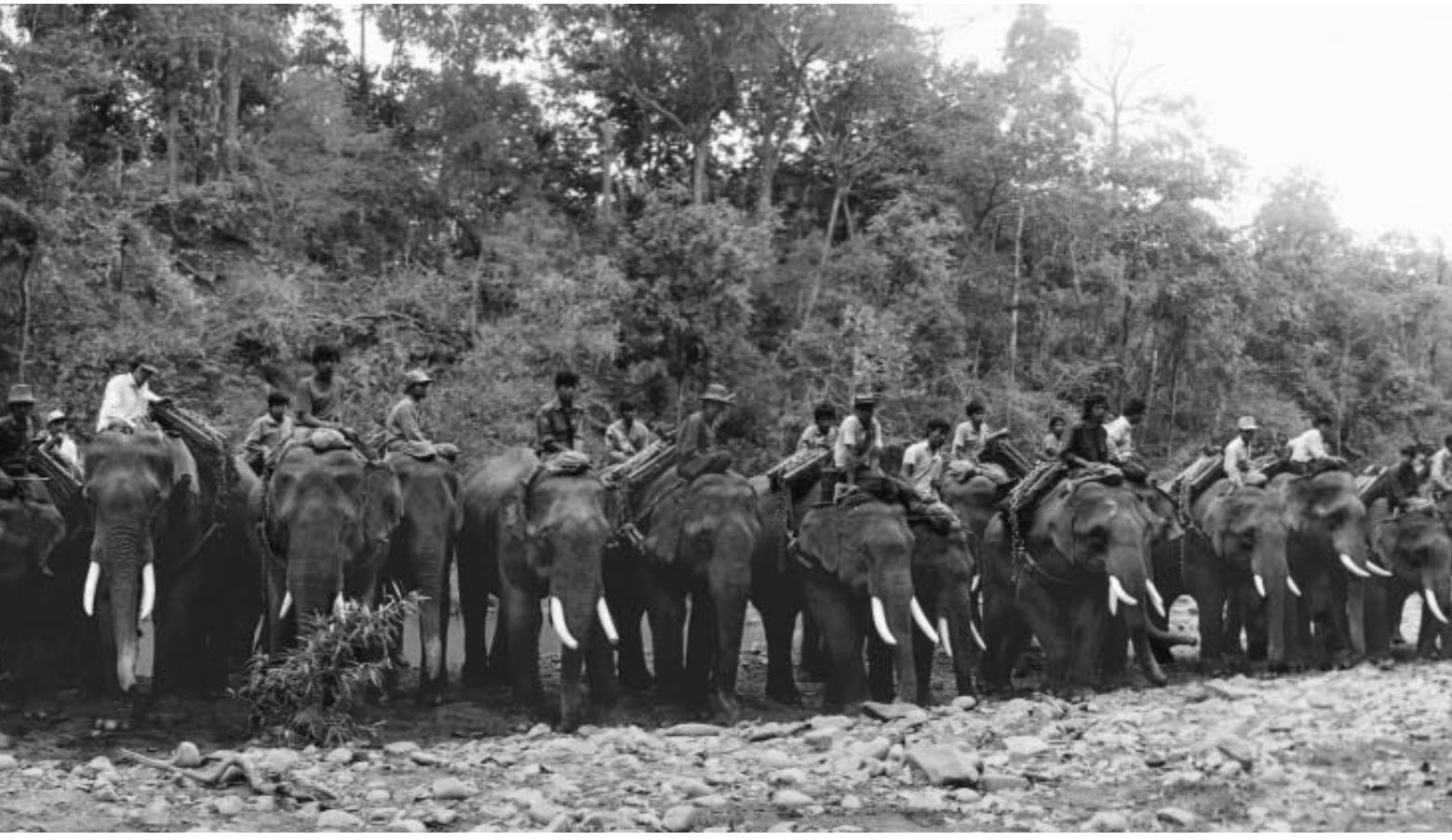
In 1992 Khyne U Mar, a veterinarian at the Myanmar Timber Enterprise, contacted the Metro Washington Park Zoo in Portland, Ore., where I work. She was seeking help in maintaining the size of the working elephant troops. Traditionally, new recruits for logging—adolescent elephants between 12 and 18 years of age—were taken from wild herds. But today there are too few wild elephants left even in Myanmar's vast forests (which may hold 5,000) to allow the animals to be captured for logging.

About 200 to 300 working elephants are lost every year to retirement, injury or death. But the working elephants give birth to no more than 200 calves a year. The only means of replacing the retirees is for Myanmar to breed elephants.

At the Metro Washington Park Zoo we have a 33-year-old program for breeding captive elephants. Using our techniques on Myanmar's working elephants provides a unique opportunity to conserve an endangered species. The loggers constitute almost 15 percent of all Asian elephants, wild and domestic. If their numbers can be kept intact, there will always be a population of Asian elephants on the earth that is large enough and genetically diverse enough to yield many generations of healthy offspring.

Even better, working elephants are gentle—a very helpful trait in handling an endangered species. But breeding even these tame elephants in their forest home has its challenges. The five inseminations we have managed in Myanmar so far have not borne fruit. Still, we are hopeful that we will be able to help Myanmar's elephants reproduce, thus saving the creatures and the logging tradition. Within the next 20 years the fate of these giants will be decided.

EMPLOYEES of the Myanmar Timber Enterprise line up for a group portrait (*below*). One technique used by the elephants in their work is to balance a small log on the tusks (*opposite page*), holding it in place with the trunk.



PULLING TEAK LOGS through forests (*above*) and across streams (*right*), elephants make do with trails rather than intrusive roads. Some logs are floated down mountain streams to the Irrawaddy River, where they are bound into giant rafts for transport to Yangôn (formerly Rangoon). Elsewhere, the logs are dragged to collection points and loaded onto trucks.

The animals need to be extremely alert mentally to avoid injury while working with these massive beams. Besides, elephants are susceptible to heat stroke and can labor only in the morning or early evening; they do not work at all through the hot months of March, April and May.

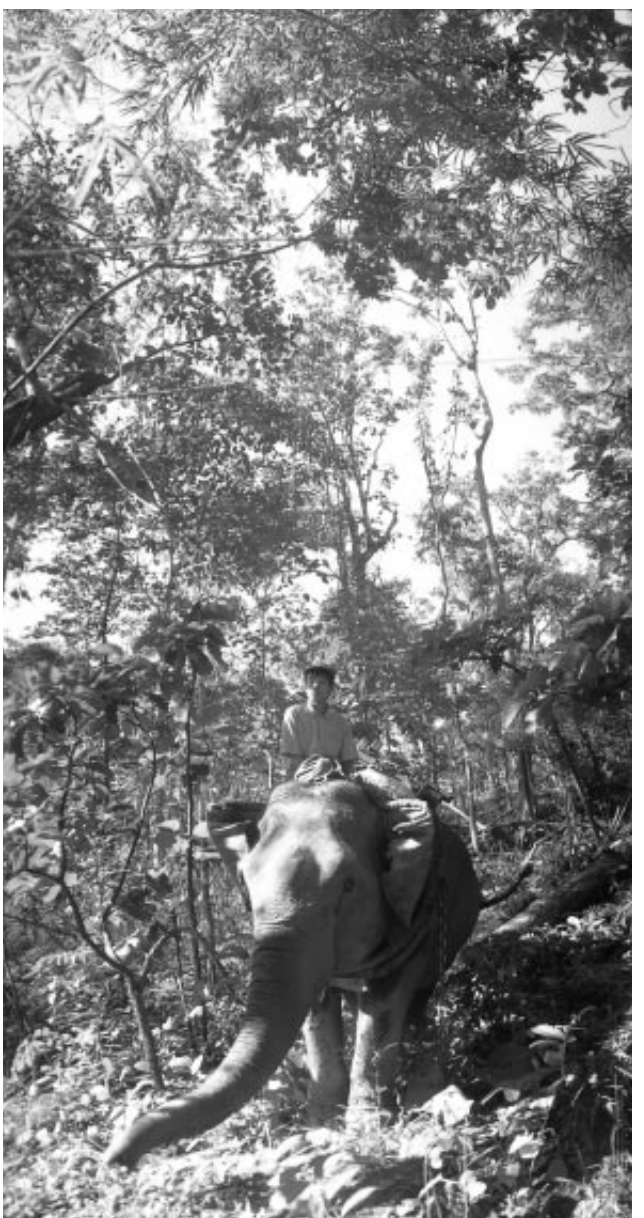
Inseminating an Elephant

Female elephants can be successfully inseminated only during the three times a year they are in estrus. But it is impossible to tell from external signals when estrus occurs; minute changes in progesterone levels in a female's blood have to be monitored weekly or even daily. During the monsoon season in Myanmar, swimmers sometimes have had to carry the blood samples across flooded rivers to Yangôn, where technicians at the Medical Research Institute measure the progesterone.

When estrous cycles are known, artificial insemination becomes possible. Tame cow elephants are readily habituated to accept a five-foot-long, flexible pneumatic insemination tube matching their unique anatomy. (In the photograph at the left, the author inspects an elephant's vagina with a fiber-optic scope.) And trained bull elephants voluntarily donate semen to an elephant-size artificial vagina. Work is under way to determine a reliable method of deep-freezing elephant semen in liquid nitrogen, so that a sperm bank can be set up—potentially including the contributions of hundreds, perhaps thousands, of bull elephants.

With such a sperm bank in place, it will be possible to invigorate isolated groups of elephants throughout Asia by inserting diverse new genes. Shipping elephant sperm, after all, is easier than shipping elephants.

PHOTOGRAPH COURTESY OF MICHAEL J. SCHMIDT



OOZIE, or trainer (*right*), shares a vocabulary of as many as 30 different commands with his elephant. Most elephants will obey most oozies, but a few elephants respond only to a particular man. Among oozies, riding the most unpredictable animal—one that is potentially dangerous—is a status symbol.

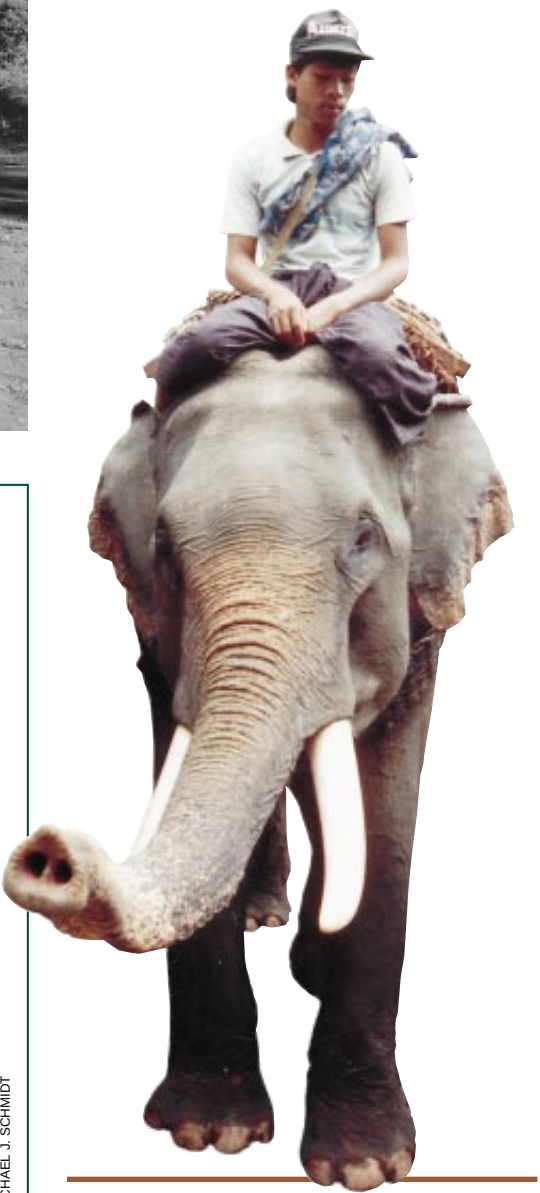
TACK SHED (*left*), where the gear for dragging logs is stored, is often the place where an elephant is saddled. At night, the elephants are hobbled and let loose in the forest to feed. The elephants are also fed salted cooked rice at the end of the workday. The oozie finds his elephant at 4 A.M. by following the distinct tones of its bell. (In earlier times, a wild bull elephant often mated with a female working elephant during her night out.)

Saddling an Elephant

An elephant's sensitive spine has to be protected from injury while the animal drags logs up to 3,000 pounds in weight. A big pad of leaves and bark make a thick, soft layer on which the dragging gear is balanced. This "saddle" prevents the chain from digging into the elephant's back; the oozie sits on the neck. These photographs, taken by the author at a logging camp in Thailand (*right*), show elephants being saddled for logging as they have been for hundreds of years.

BATHING, an intimate daily ritual that can last an hour, is a time for the oozie to scrub, clean and bond with his elephant. Because elephants live an average of 50 to 60 years, and sometimes as long as 80, the relationship can become a lifelong one.

PHOTOGRAPHS COURTESY OF MICHAEL J. SCHMIDT



The Author

MICHAEL J. SCHMIDT, a veterinarian, researches elephant-breeding techniques at the Metro Washington Park Zoo in Portland, Ore. He obtained a doctorate in veterinary medicine in 1973 from the University of Minnesota and is a member of the IUCN Asian Elephant Subspecialties Group. Schmidt and his wife, Anne, live on an island north of Portland, where they breed Hanoverian horses.

Explaining Everything

by Madhusree Mukerjee, *staff writer*

A slight breeze stirs the hot air under the canopy, bringing scant relief to the scientists arrayed in its shade. A magpie calls raucously, punctuating the even tones of the seminar speaker's voice. "I want my four-dimensional canonical gravity to be one at infinity," says Renata Kallosh of Stanford University, as she chalks out equations on the makeshift blackboard. Strains of music from a distant concert reach a crescendo and suddenly stop. Jeffrey A. Harvey of the University of Chicago is asking a question:

"What does it mean that your black holes have zero mass? Do they move at the speed of light?"

"No, they have nothing, no momentum," Gary T. Horowitz of the University of California at Santa Barbara turns to reply.

"Oh, baloney!" That was Leonard Susskind of Stanford.

"They have no energy, no momentum—there's nothing there!" Harvey protests.

The heady debate shifts, unresolved—one of many that sporadically erupt among the theoretical physicists gathered at the Aspen Center for Physics in the Colorado Rockies. A sense of barely suppressed excitement fills the air. The Theory of Everything, or TOE, the theorists believe, is hovering right around the corner.

When finally grasped—the fantasy goes—the TOE will be simple enough to write down as a single equation and to solve. The solution will describe a universe that is unmistakably ours: with three spatial dimensions and one time dimension; with quarks, electrons and the other particles that make up chairs, magpies and stars; with gravity, nuclear forces and electromagnetism to hold it all together; with even the big bang from which everything began. The major paradigms of physics—including quantum mechanics and Einstein's gravity—will be revealed as intimately related. "Concepts of physics as we know them today will be completely changed as the story unfolds," predicts Edward Witten of the Institute for Advanced Study in Princeton, N.J.

Grand promises were also heard a decade ago, when "string theory" gained favor as a TOE. Physicists crafted the theory from the idea that the most elementary object in the universe is an unimaginably tiny string. The

A new symmetry, duality, is changing the way physicists think about fundamental particles—or strings. It is also leading the way to a Theory of Everything

undulations of such strings were posited to yield all the particles and forces in the universe. These loops or segments of string are about 10^{-33} centimeter long and vibrate in many different modes, just as a violin string can. Each vibrational mode has a fixed energy and so by the laws of quantum mechanics can be thought of as a particle. But string theory soon ran into mathematical barriers: it frayed into five competing theories. "It's unaesthetic to have five unified theories," wryly comments Andrew Strominger of the University of California at Santa Barbara. Worse, the theories had thousands of solutions, most of which looked nothing like our universe. Asked in 1986 to summarize the TOE in no more than seven words, Sheldon L. Glashow of Harvard University, a longtime critic, exclaimed in mock anguish: "Oh, Lord, why have you forsaken me?"

The "Lord," it would appear, has heard. A peculiar new symmetry, called duality, is making all the different strings twine into one another. Indeed, duality is redefining what physicists consider a fundamental particle—or string. Elementary objects now seem to be made of the very particles they create. Witten believes duality not only will lead to a TOE but also may illuminate why the universe is the way it is. "I think we are heading for an explanation of quantum mechanics," he asserts. Few critics of the theory's current claims can be heard: string mathematics is so complex that it has left behind the vast majority of physicists and mathematicians.

At the same time, the world according to duality is getting even more bizarre. Strings mutate with ease into black holes, and vice versa; new dimensions blow up in different realms; and not only strings but bubbles and other membranes shimmer down the byways of the universe. The multitude of links, the researchers believe, points to a deeper entity—presumably the TOE—that explains it all. "It's like aspen trees," offers Michael J. Duff of Texas A&M University, waving at a nearby stand. "There is a root system that spreads under the ground. You see only the little bits that poke up above the surface."

A New Symmetry

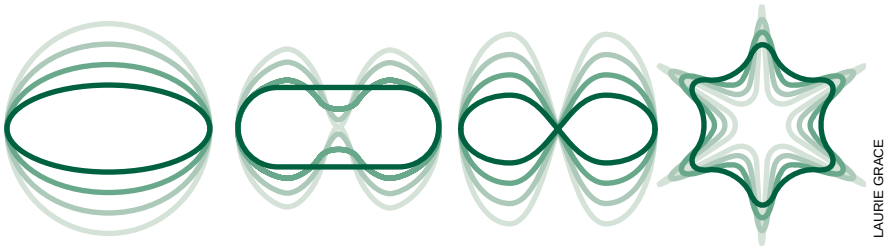
The word "dual"—fast replacing "super" as the most overused word in particle theory—has many different connotations for physicists. Broadly, two theories are said to be dual if they are apparently dissimilar but make the same physical predictions. For example, if all the electrical and magnetic quantities in Maxwell's equations for electromagnetism are interchanged, one nominally obtains a different theory. But if in addition to electrical charges, the world is presumed to contain magnetic charges (such as the isolated north pole of a bar magnet), the two theories become exactly the same—or dual.

More specifically, duality makes elementary and composite objects interchangeable: whether a particle or other entity is irreducibly fundamental or is itself made up of even more fundamental entities depends on your point of view. Either perspective ultimately yields the same physical results.

The first signs of duality appeared while physicists were working on quantum-field

ELEMENTARY STRING, on closer inspection, turns out to be an intricate, composite object, woven from the very particles and strings to which it gives rise.

AL KAMAJIAN



LAURIE GRACE

DIVERSE MODES of vibration can be induced in any string. Quantum mechanics allows the waves to be interpreted as particles. If loops of string about 10^{-33} centimeter long are fundamental constituents of matter, then their vibrational energies are the masses of elementary particles such as electrons, quarks and photons.

theories, theories that describe particles as quantum-mechanical waves spread out in space-time. In the field theory called quantum chromodynamics, or QCD, quarks are elementary particles that have a kind of charge, much like electrical charge, called color. Color makes quarks attract one another very strongly, clumping into pairs and triads to form larger, composite particles such as protons.

Just as in the familiar world there are no particles with magnetic charge, there are no particles with color magnetic charge. But in 1974 Gerard 't Hooft of Utrecht University in the Netherlands and Alexander Polyakov, then at the Landau Institute near Moscow, described how fields making up quarks might knot into small balls endowed with color magnetic charge. Such clumps—which physicists visualize as hedgehoglike spheres studded with arrows representing vectors—are generically called solitons and behave like particles. Thus, a theory of quarks with color charge might also imply the existence of solitons with color magnetic charge, otherwise known as monopoles. The monopoles would be composite particles, derived from the fields of more elementary quarks.

In 1977 David Olive and Claus Montonen, working at CERN near Geneva, speculated that field theories involving color might be dual. That is, instead of quarks being elementary and monopoles composite, perhaps one could think of the monopoles as being elementary. Then one might start with a field theory of interacting monopoles, finding that it gave rise to solitons that looked like quarks. Either the quark or the monopole approach to the theory should give the same physical results.

Most theorists were skeptical. Even if duality did exist, it was thought impossible to establish: the mathematics of QCD is extremely hard, and it would be necessary to calculate two sets of predictions for comparison. "In physics it's very rare that you can calculate something exactly," remarks Nathan Seiberg of Rutgers University. In February 1994, however, Ashoke Sen of the Tata Institute in Bombay, India, showed that on occasion, predictions of duality could be precisely tested—and were correct.

The calculation converted the string community. "Witten went from telling everyone this was a waste of time to telling them this was the most important thing to work on," Harvey chuckles. Witten, often referred to as "the Pope" by detractors of string theory, has initiated many trends in particle physics during the past two decades.

Meanwhile Seiberg was developing an extremely helpful calculational shortcut for studying QCD. His work was based on supersymmetry. Supersymmetry is the idea that for each kind of particle that constitutes matter, there should be a related particle that transmits force, and vice versa. The symmetry has yet to be found in nature, but theorists frequently invoke its powers.

Seiberg was able to show, by using supersymmetry to constrain the interaction between particles, how some hitherto impossible calculations in QCD might be done. He and Witten went on to demonstrate that versions of QCD that include supersymmetry are dual.

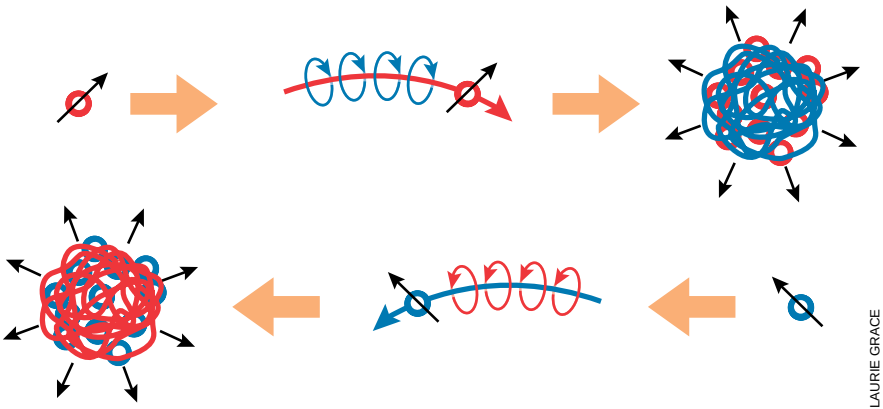
There is an immediate, startling benefit. QCD is difficult to calculate with because quarks interact, or "couple," strongly. But monopoles interact weakly, and calculations with these are easy. Duality would allow theorists to deal with monopoles—and automatically know all the answers to QCD. "It's some kind of magical trick," Harvey says. "We don't understand yet why it should work." Armed with duality, Seiberg and Witten went on to calculate in great detail why free quarks are never observed in nature, verifying a mechanism put forth in the 1970s by 't Hooft and Stanley Mandelstam of the University of California at Berkeley.

Of course, the validity of all this work hinges on the assumption that supersymmetry exists. Still, Seiberg hopes that, ultimately, duality will prevail even if supersymmetry is absent, so that "the qualitative results will be true even if the quantitative results depend on supersymmetry."

Duality is, however, much more than a calculational tool: it is a new way of looking at the world. "Something thought of as composite becomes fundamental," Harvey points out. And vice versa. Even the normally conservative Seiberg has not been able to resist speculating that perhaps the quarks are solitons, duals of some other truly elementary particles that are even smaller.

Stringing Strings Together

The concept of duality may have grown out of field theories, but as Sen observes, "duality is much more natural in string theory." It is also more versatile. Duality can unite strings of



LAURIE GRACE

DUALITY, a type of symmetry, makes it possible to view composite entities as equivalent to fundamental particles, and vice versa. For example, a quark has a kind of charge called color (*red*). Moving electrical charges generate magnetic fields; likewise, moving quarks generate color magnetic fields (*blue*). Sometimes many quarks can tangle into a composite object, called a monopole, that has a color magnetic charge (*top right*). Because of duality, however, it is possible to think of a monopole as a fundamental particle (*bottom right*). Monopoles in turn can clump to form quarks—which are now composite objects (*bottom left*). The arrows (*black*) signify properties of the particles that are vectors, such as angular momentum.

different kinds, existing in different dimensions and in space-times of different shapes. All these feats are allowing string theory to overcome its limitations and rise to the status of a TOE.

Earlier in its evolution, string theory had failed as a unified theory because of the many types of strings that were posited, as well as the embarrassing multiplicity of answers that it gave. This plenitude has its source in yet another peculiarity of string theory—it is consistent only if strings originally inhabit a 10-dimensional space-time. The real world, of course, has four dimensions, three of space and one of time. The extra six dimensions are assumed to curl up so tight that they pass undetected by large objects such as humans—or even quarks. "Think of a garden hose," suggests Brian R. Greene of Cornell University. "From a distance it looks one-dimensional, like a line. If you get close, you see it's actually a two-dimensional surface, with one dimension curled up tight."

Unhappily for string theorists, the extra six dimensions can curl up in very many different ways: "Tens of thousands is the official estimate," Strominger quips. Each of these crumpled spaces yields a different solution to string theory, with its own picture of the four-dimensional world—not exactly what one wants from a TOE.

A type of duality called mirror symmetry found in the late 1980s has helped lessen this problem by merging some of the alternative solutions. Mirror symmetry revealed that strings in two different curled spaces sometimes yield the same particles. For example, if one dimension becomes very small, a string looped around that dimension—like a rubber band around a hose—might create the same particles as a string moving around a "fat" dimension.

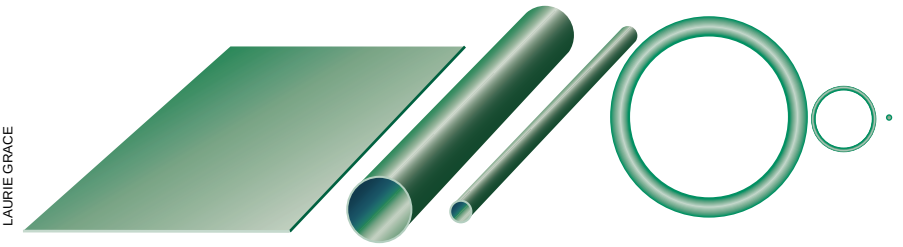
The size to which a dimension shrinks is rather similar, in string theory, to another parameter: the strength with which particles interact. In 1990 Anamaria Font, Luis E. Ibáñez, Dieter Lüst and Fernando Quevedo, collaborating at CERN, suggested that something like mirror symmetry also exists for coupling strengths. Just as large spaces can have the same physics as small ones, perhaps a string theory with large coupling could give the same results as another having small coupling.

This conjecture related string theories in the same way that duality worked for field theory. Moreover, from afar, strings look like particles, so that duality in string theory implies duality in field theory, and vice versa. Each time duality was tested in either case, it passed with flying colors and helped to draw the two realms closer together.

Meanwhile duality was emerging from a completely different quarter—supergravity. This unified theory was an attempt to stretch Einstein's gravity to include supersymmetry. (In contrast, string theory tried to modify particle theory to include gravity.) In 1986 Duff, then at Imperial College, London, was able to derive a picture for supergravity that involved vibrations of an entirely new fundamental entity: a bubble. Whereas strings wiggled through 10 dimensions, this bubble floated in 11.

"The vast majority of the string community was not the least interested," Duff recalls—most likely because no one knew how to do calculations with this bubble. Still, he continued to work on diverse theories involving closed membranes. He found that a five-dimensional membrane, or a "five-brane," that moved through a 10-dimensional space could serve as an alternative description of string theory.

The five-brane could wrap itself around an internal curled space, like a skin around a sausage. But if this internal space shrank to nothing, the bubble ended up looking like a string. Duff suggested that this convoluted string was actually the same as the ones in string theory, positing a "string-string" duality. At the same time, Christopher M. Hull of Queen Mary and Westfield College and Paul K. Townsend of the University of Cambridge were hypothesizing about many generalizations of duality in string theory. "Neither group paid much attention to the other's paper," Duff says, with a gleam in his eye.



REDUCING DIMENSIONS of a space can be achieved by pasting its edges together and shrinking it. For example, a two-dimensional sheet of rubber is first curled into a cylinder, and the curled dimension is then shrunk. When thin enough, the cylinder looks like a (one-dimensional) line. Twisting around this length of "hose" and sticking its ends together, one gets a doughnut shape. The radius of the doughnut can be shrunk until it is small enough to approximate a point—a zero-dimensional space. Such changes could explain why the extra dimensions of space-time that string theory says must exist are too small to be detectable.

Explosion of Dualities

That is, until March 1995, when matters came to a head at a conference at the University of Southern California. Witten gave the first talk of the day, pulling together evidence for duality from diverse realms. He recognized that Hull, Townsend and Duff were all talking about the same idea and went on to conjecture that Duff's bubbles in 11 dimensions were solitons of a particular string in 10 dimensions. After Witten, Seiberg spoke. "Natty [Seiberg] was so impressed by Witten's talk," chuckles John H. Schwarz of the California Institute of Technology. "He said, 'I should become a truck driver.' " But Seiberg also presented many new results, prompting Schwarz—one of the founders of string theory—to begin his talk with, "I'll get a tricycle."

An explosion of activity followed and has continued unabated. Every day scientists log on to the electronic preprint library at Los Alamos National Laboratory to find some 10 new papers in the field. "It's the first thing you do every morning," remarks Anna Ceresole of the Polytechnic of Turin. "Like reading the newspaper." Scattered and curious evidence for duality is turning up, relating strings and bubbles to solitons of all kinds and shapes.

One soliton, which resembles a hairy caterpillar, with vector arrows pointing out all along a line, turned out to be dual to a fundamental string. (It is also similar to a cosmic string, a fad in cosmology begun by Witten a decade ago.) Different kinds of strings squeezed into the real world—four dimensions—also proved dual. "Things happen for different reasons, yet they agree," Seiberg remarks. "It feels like magic."

There is a method behind the mad hunt for dualities. "Many string theories are not realistic," Sen points out. "We need to understand all of them to find the real one." Duality serves to connect, and therefore to reduce, the number of options. Witten believes the five string theories involving 10 dimensions that now prevail will all turn out

to be reflections of an ultimate, supreme, quantum string.

Duff has even proposed a "duality of dualities"—the duality between spaces, and that between elementary and composite objects, might turn out to be connected. Among the most peculiar predictions of such ideas is that the size of a curled space influences the strength with which particles interact, and vice versa. So if an internal dimension is big, coupling between particles might also be large.

Besides, Susskind explains, "As you go from place to place, the size of the internal dimension may vary." If a curled dimension blows up, in some far corner of the universe, space-time acquires a new, fifth dimension. Where it squeezes tight, as in our immediate environment, quantum effects appear. Indeed, the fundamental scale associated with quantum theory, called Planck's constant, is intimately entwined with duality: it relates, for example, the mass of a particle or string with that of its dual. "That is the most compelling evidence for me that string theory might teach us about quantum mechanics," remarks Stephen H. Shenker of Rutgers.

"Suddenly, dimensions are changing, dimensions of fundamental objects are changing, wrapping around, anything goes," Duff shakes his head in wonder. One more suggestion from Townsend is a kind of "democracy"—the membranes turning up as solitons of string theory might all be fundamental objects, having the same status as strings. That idea has yet to catch on with the Americans, who point out that calculations with membranes still do not make sense. As Cumrun Vafa of Harvard University notes doubtfully, "It's kind of coming in sideways. You never know."

Black Holes

As if that were not enough, a connection emerged last April between strings and black holes—promising to overcome the second major embarrassment in string theory. Strominger, Greene and David R. Morrison of Duke University found that black holes help to connect perhaps thousands of the tens of thousands of solutions to string theory in a complex web. The connections make the problem of finding the "right" solution to string theory—that describing our universe—much easier.

In a sense, black holes have been lurking at the edges of string theory all along. If enough mass accumulates in one place, it collapses under its own gravitational pull to create a black hole. But as Stephen W. Hawking of the University of Cambridge has argued, a black hole—which usually absorbs everything, even light—may also radiate particles, slowly losing mass and shrinking. If the original mass were made up of strings, the decay would ultimately lead to an object with zero size—an "extremal" black hole, looking in fact rather like a particle.

Susskind protests that these tiny black holes are nothing like the collapsed stars that astrophysicists search for: "Andy's [Strominger's] work is great, but calling these things black holes is I think a bit of hype." (Susskind's own latest paper is entitled "The World as a Hologram.") In fact, extremal black holes—or black bubbles or black sheets—are simply clumps of string fields, otherwise known as solitons.

Strominger was investigating how extremal black holes behave when a dimension of space-time curls up very tight. Imagine taking an infinitely long hose, looping it around and sticking the ends together so that it resembles a doughnut. In this way, both dimensions of the surface of the hose can be shrunk, creating a much smaller space (that still has no boundaries). Now suppose that the doughnut becomes very thin at one point. As it pinches in, Strominger found that some black holes, made of membranes wrapped around the scrunched dimension, become massless. He decided to include these objects in his calculations, as quantum-mechanical waves.

Two miraculous things happened. Earlier calculations in string theory had always failed when the hose thinned to a line, but the quantum-mechanical black holes made the mathematics work out fine even in this extreme case. The real savior, Horowitz explains, is quantum physics: "In classical physics, an electron falling into the point charge of a proton gives you infinities. Only when you add quantum mechanics do you

The Middle Ages

The revival of string theory is unique in one respect—the age of the scientists involved. Physicists, like fashion models, tend to think they are over the hill at 25. "It's often said mathematics and physics are young men's subjects," notes Michael J. Duff of Texas A&M University (with unconscious sexism). But unlike earlier revolutions in physics, this spurt in particle theory is led by scientists in their late thirties and early forties.

One reason might be the vast array of subjects in addition to string theory—field theory, supersymmetry, gravity, solitons and topology—that the researchers need to have at their fingertips. "It's hard for young people to master all of these fields fast enough to make a contribution," says Jeffrey A. Harvey of the University of Chicago. Most of the leaders of this revival are those who made string theory prominent in the 1980s—and they are now 10 years older.

They also got their tenured faculty positions 10 years ago. But few of the students who were trained then in string theory have made it through the system. "The field was overhyped, and there was a backlash," Duff explains. And science funding fell pre-

PHOTOGRAPHS BY MADHUSREE MUKERJEE

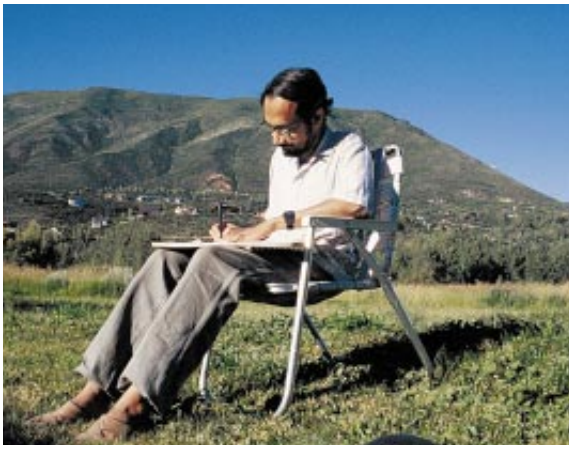
cipitously, so that the majority of the younger physicists were not able to find jobs.

Those who did manage to hang on were under tremendous pressure to publish. That took a toll on creativity. "Young people can't easily strike out on their own for a few years," Harvey reports. "If they don't do anything that anybody takes an interest in, they can't get another job." Leonard Susskind of Stanford University agrees: "The postdoc system leaves little time for reflection." And there were few new students—no one saw the field as having a future.

At the same time, an entirely different generation of physicists—the old, famous men—seem to have been pushed out of the picture. Sidney R. Coleman of Harvard University, for instance, declined to comment on the new developments: "At my age, you tend to emit a lot of gas," he protested. "I'd rather not." His Harvard colleague Sheldon L. Glashow, whose barbs still rankle with some string theorists, was entirely unaware that something had changed.

Susskind, who at age 55 falls between the generations, takes an optimistic view of this turnover: "It's a good sign there is a generation in the process of giving up. Means the field is moving in directions that older people can't follow." He complains, however, that the 40-year-olds, though undoubtedly brilliant, are too normal to be interesting. Indeed, the friendly, well-adjusted individuals at a string-theory workshop in Aspen, Colo. (*photographs*), seem a far cry from the arrogant, eccentric geniuses of yesteryear epitomized by Richard P. Feynman.

But when the bright young things of a new generation come in, Nathan Seiberg of Rutgers University predicts, "we have to be squeezed out. They will take over." No matter what changes, faith in the magic of youth seems destined to go unchallenged.



see that the electron goes into orbit." Another consequence was that large numbers of the massless black holes appeared: the system underwent a phase transition, much like vapor condensing to water.

The phase transition mirrored a change in the doughnut itself. It tore open at the thinnest part—violence that physicists and mathematicians have always shrunk from—and remolded into a sphere, an alternative way of curling up a two-dimensional sheet. Thus, two very different curled spaces in string theory were connected. "Mathematicians don't like it, because it involves tearing," Strominger admits. "But quantum effects smooth it out."

Different kinds of tears may ultimately turn out to relate thousands of solutions to string theory. With the internal spaces thus linked, strings can then find the "special" one by moving around among them. Just as water freezes in the Arctic and vaporizes in the Sahara, strings can choose a configuration suited to their environment. Finding the right solution then becomes a dynamical problem.

Somewhere in the universe, Strominger speculates, there might be a droplet in which strings have found a different internal space. On entering the droplet, black holes would turn into strings. And strings into black holes. In our immediate surroundings, such droplets might appear fleetingly as virtual universes, which exist for microscopic fractions of time and die away before they become evident.

The Theory

Despite these flights of fancy, the physicists come down to earth long enough to caution that the ultimate theory is still far off. Even the optimist Vafa, who has bet Witten a scoop of ice cream that string theory will be solved by the end of the century, believes that a true understanding will take decades to emerge. "By the time we find a beautiful formulation, it might not be called a string theory anymore," Schwarz muses. "Maybe we'll just call it 'the Theory.'" (Claims of finding the TOE met with so much ridicule in the 1980s that string theorists are now allergic to that sobriquet.)

Not everyone is convinced that the Theory is around the corner. "Coming from the string-theory clan, the reports are as usual loaded with overstatements," 't Hooft acidly retorts. An immense problem is that there may never be any experimental tests for strings. No one can even conceive of a test for something so minute: modern equipment cannot probe anything smaller than 10^{-16} centimeter. Theorists pray that when the Large Hadron Collider at CERN starts operating in 2005, supersymmetry, at least, will be discovered. "It will be one of the nicest ways for nature to have chosen to be kind," says Witten (echoing Einstein's faith that God is not malicious).

But even if supersymmetry shows up, another nagging problem will remain. In the real world, the familiar four-dimensional space-time is flat; the kind of imperfect supersymmetry that theorists attribute to nature, however, makes space-time curl up impossibly tight in all dimensions.

Witten has a fantasy for getting around this impasse, which relies on duality between theories in different dimensions. Perhaps one can begin with a universe in which only three dimensions are initially flat—one of the four we know is still curled up. Such space-times have peculiar but pleasant properties that allow the problems with supersymmetry to be fixed. Ultimately, the fourth dimension might be induced to expand, leading to a world like the one we know. "Witten's suggestion is pretty wild," Schwarz grins, "but he might be right."

The peculiarity of gravity also raises many difficult questions. Einstein found that gravity arises from the curvature of space-time. Therefore, to quantize gravity is to quantize space and time. In that case, Horowitz argues, "maybe there is no meaning to space and time, and maybe these emerge as some approximate structure at large distances."

String theory is a long way from meeting such expectations. Besides, the Theory will need to be able to describe the most extreme circumstances, such as the genesis of the universe or the environment inside a black hole. "String theorists tend to trust their theory

The Mathematics of Duality

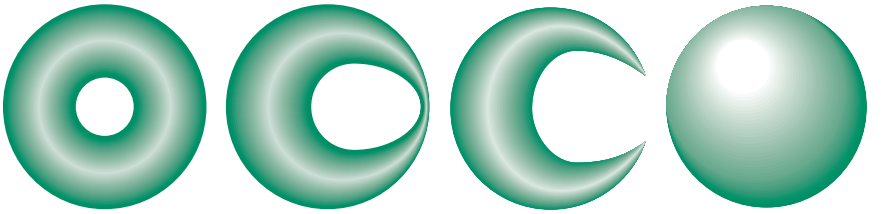
Using intuition, analogies and a kind of free-flowing mathematics inspired by nature, physicists have solved some long-standing problems in classical mathematics. They are also forcing open a new branch of mathematics, called quantum geometry. "The physicists are telling us where to look," remarks John Morgan, a mathematician at Columbia University. "It's frustrating. We don't have the access they do to this kind of thinking."

In 1990 Edward Witten of the Institute for Advanced Study in Princeton, N.J., was awarded the Fields Medal—the Nobel of mathematics—for the manifold ways in which he had used theoretical physics to unravel mathematical puzzles. A key concept from physics, supersymmetry, turns out to connect intimately with modern geometry. "It's very surprising," remarks David R. Morrison of Duke University. The latest triumph of supersymmetry is a means of classifying four-dimensional spaces. These dimensions, which pertain to the real world, are curiously also the most complex.

Simon K. Donaldson of the University of Oxford had shown in 1982 how to use quantum-field theories to count the number of holes in a four-dimensional space and thus to classify it topologically. (For example, a sphere, a doughnut and a pretzel all belong in different categories of two-dimensional surfaces because they contain different numbers of holes.) But the calculations were horrendous because of the intractable nature of the field theories. In 1994 Nathan Seiberg of Rutgers University and Witten pointed out that the results of one supersymmetric quantum-field theory could be provided by another, via a symmetry called duality. Thus, easy calculations might suffice to obtain the results of very difficult ones. Witten provided an equivalent set of numbers that could be calculated almost 100 times faster than the "Donaldson numbers." "Seiberg-Witten theory opened up the field and allowed us to answer most of the outstanding questions completely," Morgan says.

Duality of a different kind, called mirror symmetry, has illuminated another vexing question. Mathematicians want to know how many curves of a given complexity can be drawn in a particular space. The problem is especially difficult to solve for convoluted curves. But Brian R. Greene of Cornell University and Ronen Plesser of Hebrew University of Jerusalem found that strings inhabiting two apparently unrelated spaces can yield the same results. Using this mirror symmetry, Philip Candelas of the University of Texas at Austin and others were able to tell the results of virtually impossible calculations in one space by looking to the mirror space—thus deriving the long-sought numbers.

Indeed, string theory yields many more insights than classical mathematics can accommodate. The contributions that are commonly cited are only those that appear when strings are shorn of quantum mechanics. Quantum strings undulate in a host of spaces that mathematicians have yet to construct. Moreover, Greene, Morrison and Andrew Strominger of the University of California at Santa Barbara have shown that quantum effects make it possible for spaces with different numbers of holes—such as a doughnut and a sphere—to transform smoothly into one another, a no-no for mathematicians. (The standard rules for manipulating spaces allow them to be stretched or compressed, but no holes can be opened or closed in them.) The study of such spaces is becoming the brand-new field of quantum geometry.



LAURIE GRACE

ELIMINATING HOLES in closed spaces was thought impossible in mathematics, but physicists have found a way. A doughnut and a sphere are both ways of curling up a two-dimensional surface, but they differ in the number of holes they contain. (A doughnut has one; a sphere has none.) If part of the doughnut thins to a point, however, the rest of it can be separated. The doughnut can then be remolded into a sphere.

The findings have rejuvenated the venerable disciplines of algebraic geometry and number theory. "These are core subjects in mathematics," states Shing-Tung Yau of Harvard University (another recipient of the Fields Medal). "If you open up a new domain here, you expect to have a lot of influence on the rest of mathematics." A major stumbling block is that mathematicians have not proved the results from string theory to their satisfaction.

Yet the mathematicians agree that physicists with their questionable methods are getting at mathematical truths. "We can't free ourselves of rigor, or the field will fall apart," Morgan explains. But rigor can also be a burden, keeping mathematicians from the leaps of faith that physicists blithely make. "Are we going to wait for physicists to tell us again where to look?" he asks. "Or are we going to get to a state where we have access to that intuition?"

blindly, claiming it can deal with everything," states 't Hooft with finality. "In reality, they don't understand gravitational collapse any better than anybody else."

But string theorists, dazzled by the mathematical riches glinting within reach, seem undeterred by any criticism. Pierre M. Ramond of the University of Florida tries to explain: "It's as if you are wandering in the valley of a king, push aside a rock and find an enchanted staircase. We are just brushing off the steps." Where the steps lead is unknown—so the adventure is all the more thrilling.

Evening falls in Aspen. As the setting sun lights up the tree trunks and leaves in clear yellow, the physicists continue an argument they have started over dinner. This time, it is about the wave function of the universe, a direct attempt to describe the latter as a quantum-mechanical object. "In my own opinionated, uneducated, ignorant view, I personally think it's a lot of crap," Susskind vents. Horowitz, who along with others, has constructed such wave functions, laughs out loud. The air starts to chill, and the quaint streetlamps glow brighter in the gathering darkness. But the physicists seem in no hurry to retire.

Recording Nature's Sounds

Evolution is the greatest composer—just stand on a lonely beach, and you'll hear. From the wind and waves, from the voices of gulls, nature conducts a symphony grander and more pleasing than any played in a concert hall. And the concert does not end at the seashore. Over all the earth's surface there may be 10 million unique sounds produced by birds, amphibians, mammals and insects. Yet despite their pull on our psyche, only about 1 percent of these natural rhapsodies has been properly recorded and archived. Far fewer have been systematically studied. With a little care, you can capture sounds never before recorded and study their patterns.

You might even be able to contribute them to a growing national registry. The Library of Natural Sounds at the Cornell Laboratory of Ornithology holds the world's largest collection of nature's voices. The archive contains songs from more than half of all bird species, not to mention insect chirps, amphibian croaks and mammal bleats. Contributing to this library of course demands high-performance devices, which can run into the thousands of dollars. Fortunately, discoveries can come more cheaply. With the help of some special software, you can do original research with a modest recording system.

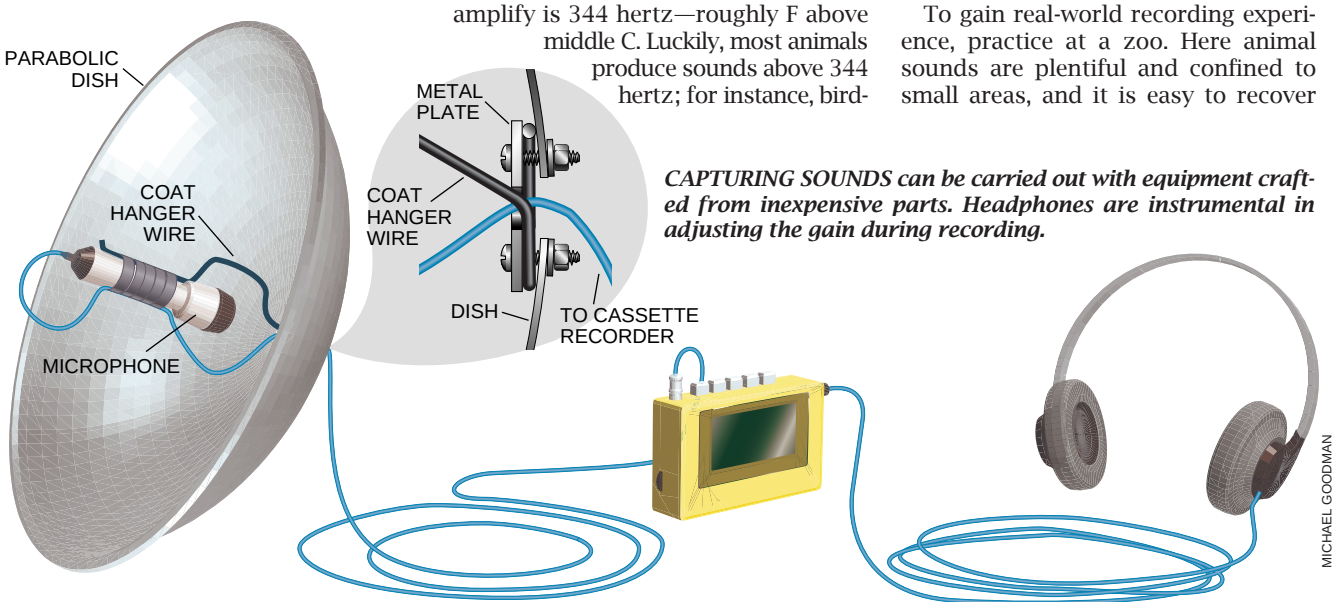
Before collecting data, you must first learn how to recognize animals by their calls. Read field guides, talk to naturalists and, most important, listen to recordings. The best resource for studying natural sounds is the Cornell library, but it charges a minimum of $22.50 for sounds. Fortunately, many collections of natural sounds are now compiled on CD. *The Guide to Bird Sounds* is the National Geographic Society's audio companion to its printed field guide, and the *Peterson Field Guides* offer calls from nearly every North American bird. If you have a CD-ROM player, check out *Bird Song Master,* which uses your computer to help you learn to identify birdcalls. Frog fans might try *Voices of the Night,* produced by the Cornell lab, which eavesdropped on lovesick amphibians. All these products can be purchased from the Library of Natural Sounds and other specialty retail outlets.

Recording nature successfully demands that you be able to isolate one particular sound and screen out all the rest. A parabolic dish does this task quite well. With a microphone at its focus, it will collect only sounds that fall parallel to its center. The dish, however, must be small enough to be carried easily in the field, limiting its diameter to about one meter. To be reflected, a sound wave must fit inside the dish. Unfortunately, that means it will not amplify a sound wave longer than one meter. So the lowest pitch this dish will amplify is 344 hertz—roughly F above middle C. Luckily, most animals produce sounds above 344 hertz; for instance, birdcalls characteristically range from 2,000 to 6,000 hertz.

Parabolic reflectors are not hard to come by. Edmund Scientific (telephone: 609-547-8880) carries an 18-inch-diameter, polished-aluminum reflector (No. 80254) for about $35. Radio dishes are also parabolic; a military surplus shop may have a few of the right size. And remember, any dish approximately parabolic will provide some amplification and directionality—some woks and large salad bowls work surprisingly well. You can also experiment with tightly stretched umbrellas, plastic snow-slides, decorative housings for light fixtures and even the bottom of a barbecue grill.

Once you have a parabolic amplifier, you will need a microphone, a tape recorder and a set of headphones. Drill a hole at the base of the parabolic dish large enough to accommodate the microphone wire. A stiff metal-wire coat hanger makes a good support for the microphone. Coil up one end of the hanger wire at the base of the dish and bolt a metal plate on top of the coiled area to hold the wire to the dish [*see illustration below*]. With duct tape, secure the free end of the hanger wire to the microphone, which should face the dish. You will have to experiment to find the best position for the microphone. The wire from the microphone can be hooked directly to a portable cassette recorder. The electronic equipment does not have to be fancy—you can learn a lot by stalking around with a $30 microphone and $50 tape machine.

To gain real-world recording experience, practice at a zoo. Here animal sounds are plentiful and confined to small areas, and it is easy to recover



***CAPTURING SOUNDS** can be carried out with equipment crafted from inexpensive parts. Headphones are instrumental in adjusting the gain during recording.*

MICHAEL GOODMAN

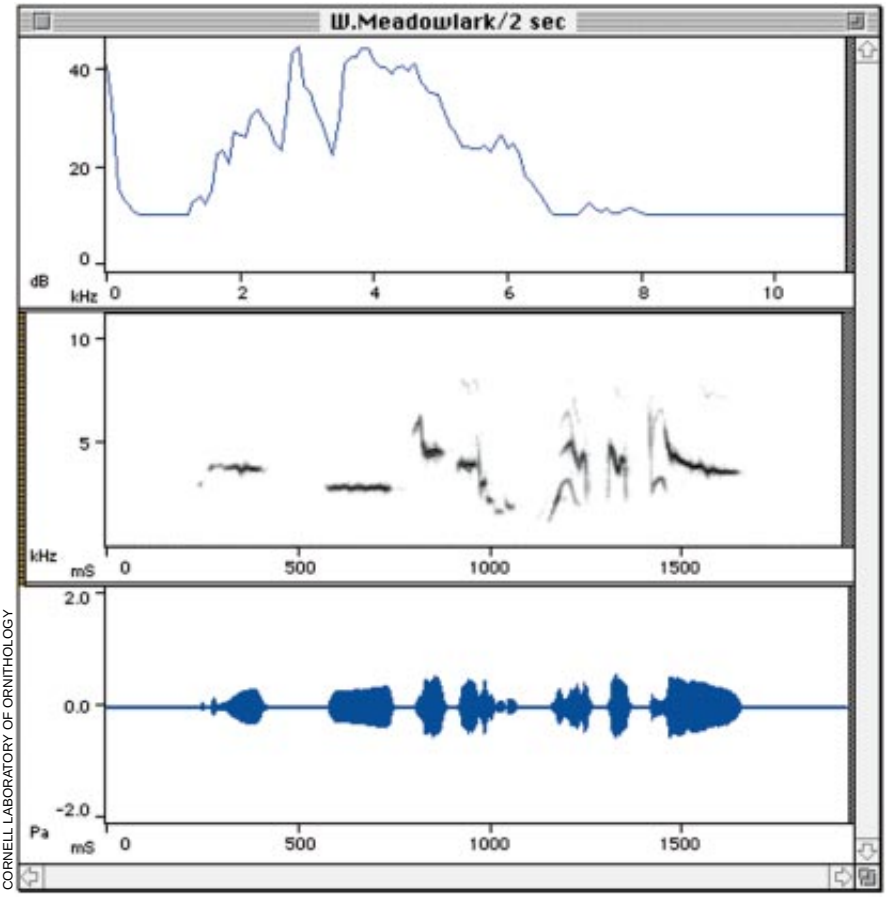
from equipment mishaps. When in the wild, be aware of what is happening behind your target. Like a telephoto lens, a parabolic microphone will compress the distance between foreground and background. Noises coming from behind your target will sound louder on the recording than they did to your ear. So position yourself to aim either upward or downward at your target. Then your background will be only the quiet sky or ground.

A few years ago only professionals could dissect sounds. Today you just need a little software. Canary, a Macintosh program designed at the Cornell ornithology lab, is ideal for studying natural voices. It lets you instantly see how any tone is constructed [*see illustration at right*]. Canary lets you zero in on any part of the melody and analyze it in detail. The program also helps to hide sins committed in the field: unwanted sounds, like distant rumblings from a truck, can be deleted.

But it is Canary's ability to compare different sounds that makes it so powerful. Canary tests sounds mathematically to see how alike they are. For identical sounds, the result is 1; for completely different ones, the result is 0. The program shifts one melody past the other in time and compares them. Then it plots these numbers so you can observe the difference graphically.

This feature makes all kinds of explorations possible. You can measure the difference between two consecutive calls from the same bird, compare one bird's song with that of another from the same species and test your ideas about the notes from closely related species. With a bit of cleverness, you can measure the time delay between when the sound arrives at each of three widely separated microphones and hence locate the bird's position when it called. (Sound travels about 344 meters per second and arrives first at the microphone closest to the animal.) You could piece together an individual bird's movements and determine the boundaries of its territory. Or try mapping the search pattern of a hunting bat by its sonar—the possibilities are endless. Given its power, Canary is a bargain at $250.

If a Windows or DOS machine lives on your desk, consider Wave for Windows (Turtle Beach, $149), Sound Forge (Sonic Foundry, $495) or Software Audio Workshop (Innovative Quality Software, $599). The drawback is that these programs were designed for sound engineers, not scientists, so they lack some of Canary's most useful features. I tip my hat to Wave. It has all the power necessary to view sound and decode its frequencies. Although musicians may enjoy the features of Wave's more expensive competitors, scientists will not see much difference. All these packages should be available from large software catalogues.



CORNELL LABORATORY OF ORNITHOLOGY

***WESTERN MEADOWLARK SONG can be viewed in three ways with the Canary program: the amount of energy the bird emits in each frequency of its song* (top), *the frequencies in the song over time* (middle) *and the sound intensity (measured in pascals, a unit of pressure) over time* (bottom).**

If you have your heart set on contributing original data to the Library of Natural Sounds, you will need professional equipment. But a good microphone could set you back $1,000, and high-quality tape recorders run upward of $2,000. So what is an amateur to do? One way is to find other people interested in recording natural sounds and pool your resources. Contact local nature organizations or post a message on an electronic bulletin board on the Internet—try the swap-meet site on the Web page of the Society for Amateur Scientists. And don't forget sound studios and local colleges; there may well be nature buffs already lurking among the equipment you need. If you are really ambitious, form a nonprofit corporation, which is eligible to receive tax-deductible donations (get a copy of *The Nonprofit Corporation Handbook,* Nolo Press, 1995). Audio manufacturers and supply houses are sometimes happy to donate outdated but perfectly functional equipment.

For more information about recording natural sounds, contact the Library of Natural Sounds and request the free equipment sheet, which describes what you will need to get serious about this hobby. Reach them at Cornell Laboratory of Ornithology, 159 Sapsucker Woods Road, Ithaca, NY 14850; e-mail: libnatsounds@cornell.edu or http://www.ornith.cornell.edu/birdlab.html; telephone: (607) 254-2404. To learn more about software and novel experiments you can do, send $2 to the Society for Amateur Scientists, 4951 D Clairemont Square, Suite 179, San Diego, CA 92117, or download the information free from http://www.thesphere.com/SAS/ or from Scientific American*'s area on America Online. Some sample birdsongs are posted on the* Scientific American *area and on the Cornell Web page.*

SHAWN CARLSON, besides being the executive director for the Society for Amateur Scientists, is also adjunct professor of physics at San Diego State University.

Mother Worm's Blanket

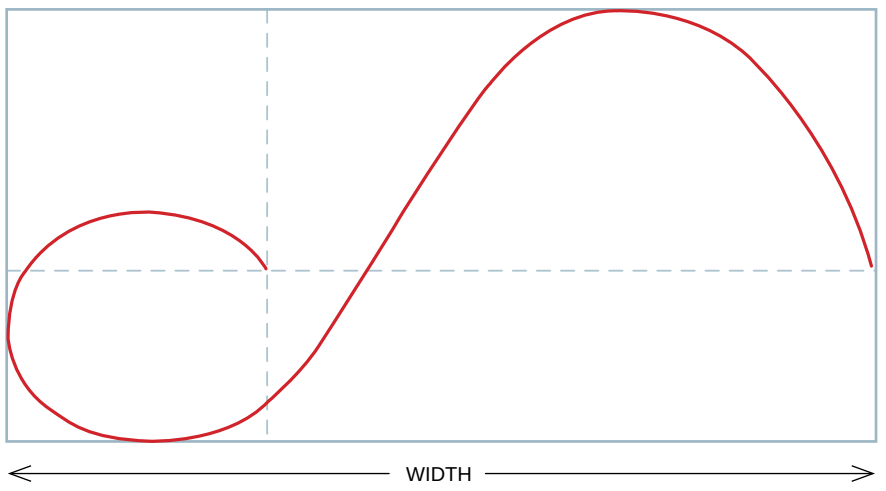
The problem known as Mother Worm's Blanket asks for the smallest two-dimensional region that can cover any curve that is one unit in length. It is an odd name, but imagine the region as a blanket and the curve as a slumbering worm. Its mother needs what we will call a universal blanket—one that covers Baby Worm, whatever position it takes. Many decades ago the mathematician Leo Moser of the University of Alberta in Canada posed a baffling, and as yet unsolved, question: What shape is the universal blanket having the smallest area?

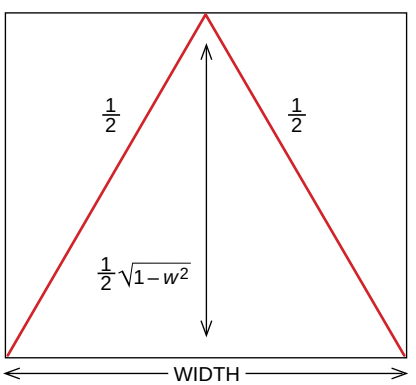
One obvious solution to the general problem is a circular disk of unit radius, having a total area of π, or 3.141. If Baby Worm's head rests at the center, its tail is at most one unit of measure away. In 1973 C. J. Gerriets and George D. Poole of Emporia State University described a pentagonal blanket only 0.286 unit in area. Until recently, no one had found anything smaller. But a few months ago David Reynolds of Credence Systems Corporation in Beaverton, Ore., sent me proof that a quadrilateral blanket, having an area of $(4+\sqrt{3})/24$, or roughly 0.239, is universal. In this realm of mathematics, Reynolds's result is dramatic. Some variation on his method might lead to smaller blankets still. Any solutions will be considered for "Feedback," but please include a proof—or strong experimental evidence—that your shape is universal.

To illustrate Reynolds's method, I'll first use it to prove that a semicircle of unit diameter—having an area of $\pi/8$, or roughly 0.392—is universal. Given a position for Baby Worm, draw a straight line joining its head and tail. (If it has curled into a closed loop, draw any line that crosses its body.) Now find the smallest rectangle enclosing Baby Worm; two sides of the rectangle should be parallel to the first line you drew [*see illustration below*]. Call the rectangle a patch. The width of the patch, w, lies between 0 and 1. For any w, we can estimate the height. We know that the greatest height for any patch, $\frac{1}{2}\sqrt{(1-w^2)}$, is needed when Baby Worm falls asleep in the form of an isosceles triangle.

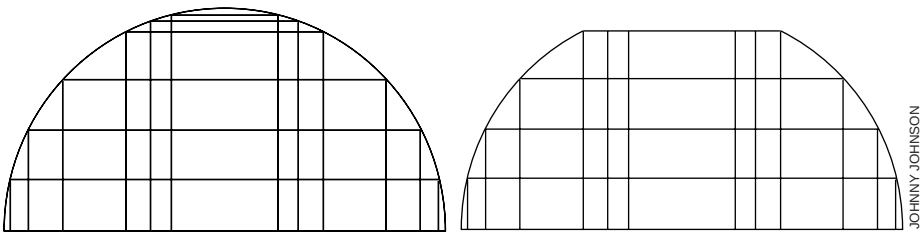
We have thus identified a family of patches, one for each w, that are "collectively" universal. Regardless of how Baby Worm curls up, at least one patch from this family covers it. We can build a universal blanket by overlaying these patches. Reynolds calls this construction Baby Worm's Patchwork Quilt, although it is more like an appliqué: the edges need not join and, in fact, the more overlap, the better. No matter how the patches are arranged, the region they cover is definitely universal. To see why, imagine that Baby Worm has curled up; find which patch covers it. The patchwork quilt consists of that patch, plus additional material; it can therefore be arranged so that the selected patch covers Baby Worm.

Our next task is to overlay the patches to obtain the smallest possible quilt. If they are arranged symmetrically, having their bases aligned, they cover a semicircle of unit diameter. As promised, then, a semicircle is indeed universal. But there is no special reason for choosing this arrangement. Nor must the patches be rectangular. And we do not actually need the full range of widths from 0 to 1 used above. When w equals $\sqrt{5}/5$, the patch is square. Any patch whose height is greater than $\sqrt{5}/5$ can be rotated 90 degrees, after which its height will be less than $\sqrt{5}/5$. Omitting these patches from the semicircle appliqué lets us trim a flat piece

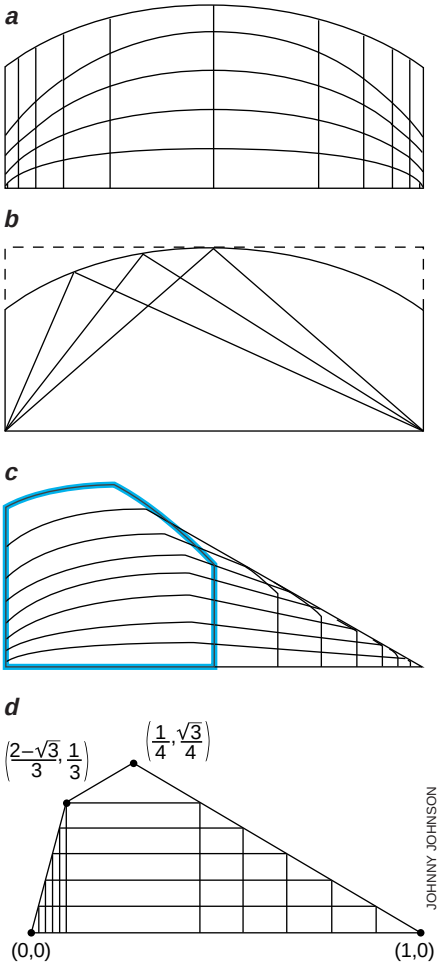




RECTANGULAR PATCHES can be created to cover any unit curve **(top).** ***The tallest patch is needed when the curve forms two sides of an isosceles triangle*** **(left).** ***The entire family of such patches can be arranged to fill a semicircle*** **(center).** ***Trimming the top yields an even smaller universal blanket*** **(right).**



JOHNNY JOHNSON



SMALLER BLANKETS **(a)** ***can be made from patches having one elliptical border*** **(b).** ***Considering how Baby Worm might be moved underneath the blanket yields further improvements*** **(c).** ***The smallest known blanket has an area of 0.239*** **(d).**

from its top, reducing its area but maintaining its universality.

A slightly more complicated argument, involving rotating Baby Worm itself, leads Reynolds to conclude that in fact we can ignore all patches for which w is less than $1/2$. Moreover, the maximum height of the patch is needed only at the middle. Otherwise the height is determined by a segment of an ellipse. Incorporating these improvements yields the blanket shown in illustration *a* at the right. At first glance, it seems we have taken this method as far as it will go, but consider that the blanket can be flipped over. If Baby Worm reaches the elliptical boundary of the patch—say, to the left of center—then it cannot also reach the top edge at the right of center. Thus, we can shave away part of the right upper edge of each patch. Overlaying the resulting patches creates a blanket bounded by part of an ellipse and three straight lines.

In constructing these patches, we assumed that Baby Worm's head and tail were somehow pinned down. The elliptical part of the boundary came from just one patch—the one of width $1/2$ [*shown in blue in illustration* c *at right*]. But any worm inside that patch can be moved toward the right end of the blanket, where there is plenty of extra room. By thinking about a few key shapes for the worm, Reynolds shows that the elliptical edge and the left vertical edge of our earlier blanket can be replaced by two straight lines. The product has an area of only 0.239, as stated earlier.

When investigating questions of this kind—geometric optimization problems—it is worth bearing in mind that there may not be one ultimate solution. Even if you find a sequence of solutions, each improving on the previous one, the sequence may not converge. Richard Courant and Herbert E. Robbins introduced a good example of what I mean in their classic book, *What Is Mathematics?* It is a problem I call Mother Gnat's Tent. Baby Gnat sleeps hovering one foot above a flat floor. Mother Gnat covers her baby with a blanket that, like a tent, touches the floor on all sides. Which tent has the smallest surface area?

A conical tent having a circular base works, and the smaller the base, the smaller the surface area. In fact, you can find suitable tents whose area is any number greater than zero. But improving on this sequence of solutions leads to a "best" tent having an area of zero, which is impossible. A cone having a base of zero is not a surface but a line segment. As such, it does not enclose Baby Gnat in any meaningful sense. So for Mother Gnat's Tent, there is no optimal solution—if we insist that the tent should be a surface.

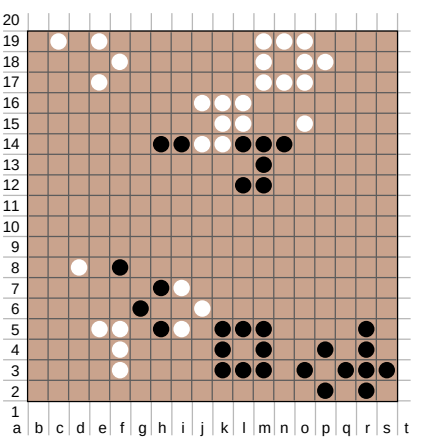
So, too, it is not clear that Mother Worm's Blanket has a definitive solution. To some extent, it depends on what types of blankets you consider. All polygonal worms of unit length, for example, can be covered by a blanket having zero area. Many such blankets exist, but they are mostly holes; the problem might be better referred to as Mother Eel's Net. In contrast, John M. Marstrand of the University of Bristol in England proved in 1979 that no blanket having zero area can cover every smooth worm. (By smooth, I mean a curve that has a tangent at every point.) Marstrand's result suggests that Mother Worm's Blanket does have a well-defined solution—and an interesting one. Reynolds's may well be it.

But amateur mathematicians can still have fun trying to find better ones. Moreover, there are many variations to this problem, nearly all of which are unsolved: What is the universal worm blanket having the smallest perimeter? How about Baby Snake's Sleeping Bag, a sac that must contain a unit-length snake however it coils up in three dimensions? And what about worms that live on the surface of a sphere?

Feedback

A number of readers wrote to point out a supposed error in "Playing Chess on a Go Board," in the November 1994 issue. The column introduced the game of gess and ended by posing a puzzle, in which white would move and mate in one. The answer given was o11-j6, using the notation explained in the article. The readers observed that black can then move m3-l4, arriving at the position shown at the right. "Footprints centered at k7 or j7 or j8 do not permit a diagonal thrust that damages the black ring," one wrote. "Therefore, black has not yet been mated."

This is correct. There are, however, other moves available. The footprint centered at j7 can move one space down to j6, wiping out both the white stone at i5 and the black stone at k5. Although white sacrifices one stone, this player still wins because the move damages black's ring. —*I.S.*

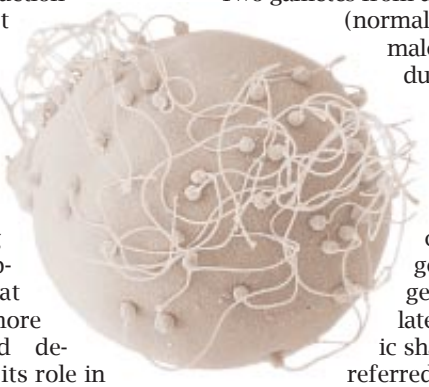


Mating Games

Review by Peter D. Sozou

EROS AND EVOLUTION: A NATURAL PHILOSOPHY OF SEX, by Richard E. Michod. Addison-Wesley Publishing, 1995 ($25). **HUMAN SPERM COMPETITION: COPULATION, MASTURBATION AND INFIDELITY,** by R. Robin Baker and Mark A. Bellis. Chapman & Hall, 1995 ($78.95). **WITH PLEASURE: THOUGHTS ON THE NATURE OF HUMAN SEXUALITY,** by Paul R. Abramson and Steven D. Pinkerton. Oxford University Press, 1995 ($25). **WHAT'S LOVE GOT TO DO WITH IT: THE EVOLUTION OF HUMAN MATING,** by Meredith F. Small. Anchor Books, 1995 ($24.95).

Sex—the fusion of genetic material from different individuals—is widespread in the living world. It is the mode of reproduction practiced by the vast majority of higher organisms. But scientists have no sure explanation of why living things go to the trouble of having sex, rather than simply reproducing asexually—which appears, on the surface at least, to be much more straightforward. And debates still rage about its role in shaping the way humans think and act. The books reviewed here all testify to the enduring hunger to understand the reproductive process that made us who we are.



Richard E. Michod starts by posing the most basic question—"Why sex?" The origin of sex opened up new evolutionary pathways that led to multicellular organisms and distinct male and female genders. These developments are a consequence of sex, but they have also had a profound bearing on its costs and benefits. Hence, to ask "Why sex?" is really to ask two distinct questions: "Why did sex originate?" and "Why does sexual reproduction exist among modern, complex plants and animals?"

To the first question, Michod gives a simple answer: sex repairs genes. The genome consists of two strands of DNA bound together in a double helix. If a single strand is damaged, the adjoining strand can be used as a template for repair. But if both strands are damaged at the same site, the required sequence of DNA must be obtained from somewhere else. In primitive single-celled organisms that have just a single genome, such genetic repair necessitates getting DNA from a different individual—sex. This explanation is supported by the observation that simple life-forms such as viruses are more likely to have sex if they are damaged.

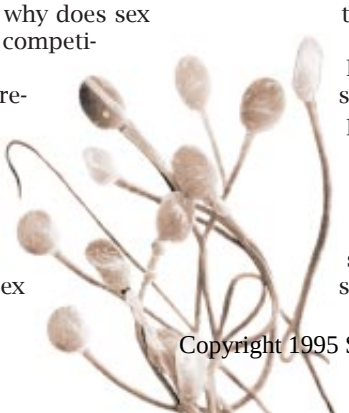
The second question poses a greater challenge. Most higher organisms are "diploid," meaning that they have a double genome made up of a number of pairs of chromosomes. Sexual reproduction involves the production of "gametes," containing a single set of chromosomes from the parental germ cell. Before separation, the chromosome pairs usually exchange long sequences of DNA, a process known as crossover. Two gametes from different individuals (normally a male and a female) then fuse to produce a new organism. There is a cost to this sexual process—the cost that the female germ cells pay for accepting half the genes for their progeny from an unrelated male. This genetic sharing is sometimes referred to as the "cost of males," because in most species males do not contribute resources toward their offspring. A female who reproduced asexually, giving virgin birth to clonal daughters, would enjoy a doubling of the representation of her genes in the next generation. What of species in which the males make a substantial contribution to parental care? Even in this case, a female could still double her genetic output by fooling a male into copulating with her and providing resources, while in reality giving birth to a genetic copy of herself. So why does sex persist in a relentlessly competitive world?

Most current theories regarding the advantage of sex are concerned with the benefits of providing a good complement of genes to offspring. Among the early proposals was the idea that sex can bring together beneficial mutations from different individuals, or that it can help eliminate bad mutations from a lineage. More recently, some researchers have argued that, in an uncertain environment, there are benefits to producing a diverse crop of offspring, or that sex enables hosts to keep changing defenses in their battle against parasites.

Michod believes these ideas are not adequate to the task of explaining why sex is so widespread. His alternative solution is again based on gene repair. In diploids, genetic information is duplicated in the chromosome pairs. A sister chromosome could therefore provide a template for repairing double-strand damage. Why not avoid the cost of sex by directly producing diploid offspring? The answer, according to Michod, lies with the mechanism for repairing genes. A damaged chromosome will join with the sister chromosome; when the two separate chromosomes are re-formed, there is a roughly 50–50 chance that crossover will occur. An offspring inheriting all its chromosomes from the same parent after such repair would inherit a double dose of some genes, and consequently, recessive harmful mutations would be expressed. Sex keeps them masked.

A distinctive feature of this theory is that crossover is regarded not as a deliberate method of shuffling genes but as an incidental consequence of a mechanism for repairing gene damage. But Michod does not explain why there is normally no crossover during sperm production in the fruit fly *Drosophila melanogaster.* Are males of this species somehow protected from gene damage?

Much of the ecological evidence about sex is open to sharply differing interpretations. A case in point concerns the "haplodiploid" sex-determining system of ants, bees and wasps. In these animals, fertilized eggs develop into diploid females, whereas unfertilized eggs develop into "haploid" males—that is, they have only a single set of chromosomes. Michod argues that this system facilitates the purging of harmful mutations in males, which explains why many of these animals are highly inbred. Other researchers believe instead that the unusual



DAVID M. PHILLIPS *Photo Researchers, Inc.*

CNRI *Science Photo Library*

genetic relationships between haplodiploid siblings facilitate certain life histories, such as the formation of social colonies, which result in inbreeding. It is worth noting that some social animals that are not haplodiploid—such as termites and naked mole rats—also exhibit inbreeding.

Despite the evident problems with Michod's ideas, much of *Eros and Evolution* is compellingly written. An early chapter gives an excellent introduction to the basic ideas of evolutionary biology. One of the appendices explains with great clarity how changes in the mating system have an effect over several generations: I found this particularly illuminating. It is a pity, however, that Michod has made no space for a discussion of the fascinating evolutionary questions associated with the sex chromosomes and with cytoplasmic DNA. An evolutionary chronology would have been useful, as would a glossary of technical terms. There are also places where Michod muddies the waters, such as a tortuous chapter attempting to clarify "survival of the fittest" and a baffling attempt to make sense of Freud's theory of the "death instinct."

Michod tries to invest his thesis with emotional appeal by casting it as a reincarnation of arguments in Plato's *Symposium.* In this view, sex enables mature adults to produce youthful offspring, because "the losses caused by age are repaired." This is a charming idea, but Michod does not explain how certain asexual creatures, such as whiptail lizards, cope with gene damage in the germ line. The overall evidence is too inconclusive to accept or reject firmly the gene-repair theory of sex.

The remaining books move on to more familiar turf by considering one of the consequences of the evolution of sex—human sexuality. Not surprisingly, popular culture intertwines with science. Of the three, only *Human Sperm Competition* does not refer to any Woody Allen film. Perhaps R. Robin Baker and Mark A. Bellis were more inspired by the World War II film, *Tora! Tora! Tora!,* for they advance the astonishing idea that the majority of human spermatozoa are designed for warfare.

This conclusion is preceded by a broad discussion of how evolution has molded sex and reproduction. The received wisdom is that natural selection will favor males whose sperm can successfully compete for the prize of fertilization. It will also favor females who can maximize the chances of their eggs being fertilized by sperm from the best males.

Sperm competition occurs when a female mates with more than one male during a single fertile cycle. Under these circumstances, the male is expected to produce a large number of sperm to try to swamp the opposition. In chimpanzees, which are highly promiscuous, males have large testes for prodigious sperm production and an external scrotum that facilitates sperm storage. Gorillas, in contrast, are not subject to sperm competition and hence have comparatively small, nonscrotal testes.

The genital anatomy of human males—external scrotum and moderately sized testes—suggests that there has been significant sperm competition in our ancestral line. Baker and Bellis suggest that the human penis is shaped to function as a piston, with copulatory thrusting movements flushing out sperm from earlier rivals. The discussion of this and other anatomical points in *Human Sperm Competition* is accompanied by line drawings that leave little to the imagination.

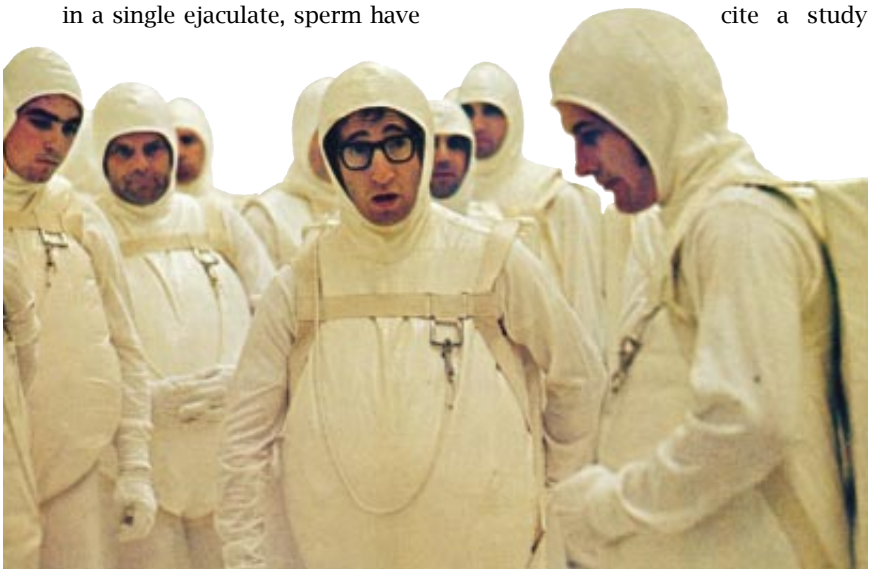
Some researchers have proposed that the female orgasm increases sperm retention, enabling women to influence the chances of a given partner's sperm achieving fertilization. Baker and Bellis describe their tests of this hypothesis, in which volunteers collected "flowbacks" that emerge after intercourse. The authors conclude that an orgasm affects sperm retention not only from the current copulation but also from the following one. This influence on future encounters extends to "noncopulatory" orgasms, hinting at a role for female masturbation in controlling fertilization.

Baker and Bellis's most original idea is that sperm have evolved specific features for combat, which the authors unblushingly call the "Kamikaze Sperm Hypothesis." This notion stems from the researchers' observation that, within a single ejaculate, sperm have a variety of shapes and sizes, many of which appear patently unsuited to penetrating an egg. The authors believe some of these sperm are "blockers" that obstruct passage of future inseminates through the female tract. Others have a "seek-and-destroy" function, actively attacking sperm from rival males, or a "family-planning" role, taking out both the enemy and the home team. This division of labor leaves only a small number of "egg-getters" for fertilization.

On the basis of this theory, Baker and Bellis suggest that a man may vary the number and composition of sperm ejaculated according to circumstances. If he has been separated from his mate for a long time, he should deposit more sperm—and a higher proportion of seek-and-destroy or family-planning types—to guard against possible recent cuckoldry. In this light, male masturbation may be a means of discarding older sperm, with the composition of the fresh cohort's being customized to suit current requirements.

These ideas are intriguing, but the supporting evidence remains equivocal. The study of flowbacks, for instance, provides only an indirect method of determining sperm retention. Such observations constitute less than compelling proof of the role of the female orgasm on future sperm retention.

Baker and Bellis support their kamikaze hypothesis with studies of numbers of different sperm forms in ejaculates, in flowbacks and in different parts of the female tract. The classification of sperm was performed manually, however, which makes the results difficult to replicate. Demonstrating the existence of specific seek-and-destroy sperm types is particularly problematic. Baker and Bellis cite a study



EVERETT COLLECTION

***WOODY ALLEN** portrays a sperm dressed for reproductive battle in the 1972 film* **Everything You Always Wanted to Know about Sex (But Were Afraid to Ask).**

by R. A. Beatty showing that some bulls sire substantially more calves than others when their sperm is mixed with that from other bulls. Beatty found, however, that bulls that perform well in sperm competition also perform well in its absence. To bolster their view that very few sperm are egg-getters, Baker and Bellis refer to a study by Dina Ralt and her colleagues indicating that a small proportion of sperm will swim toward chemical attractants. Recent research by Amnon Makler of the Israel Institute of Technology casts doubt on the reality of this effect, however.

Reproductive biologists will need to see stronger evidence before they embrace the kamikaze hypothesis, but Baker and Bellis are to be commended for raising such provocative ideas in a forthright manner.

Paul R. Abramson and Steven D. Pinkerton offer lighter reading in *With Pleasure.* This book is essentially a celebration of sex—in the colloquial sense, meaning genital activity—from a liberal, hedonistic perspective. The authors' central argument is that sex is for pleasure, not procreation, because it is usually pleasure that provides the motivating force for human sexual activity. This line of reasoning would seem to lead to the tautological conclusion that the purpose of any action is to satisfy the motivation for carrying it out.

Philosophical caveats aside, much of the book is stimulating and informative and written with ample wit. The ethnographic and historical references illustrate the enormous cultural variability of human sexual practices. We learn that in the Innis Beag community of Ireland "intercourse is invariably completed quickly, with the man falling asleep shortly after achieving orgasm," whereas among the Mangaians of Polynesia "a 'good' man is able to bring his partner to climax two or three times for every one of his." The Sambia of Papua New Guinea have practiced ritualized fellatio, believing that "the ingestion of older men's semen is essential to masculine development."

Abramson and Pinkerton offer an impassioned defense of pornography, arguing that it represents a moral counterculture and hence deserves the same protection as the scientific thoughts of Galileo and Darwin. But it is surely simplistic to claim that all opponents of pornography have "a shared opposition to nonprocreative sexuality."

Of greater concern is a misleading treatment of HIV transmission dynamics in an otherwise instructive discussion of AIDS. We are advised that "provided that condoms are used consistently, engaging in 100 one-night stands is actually safer than having 100 unprotected sexual contacts with a single partner of unknown HIV status." This conclusion is based on the assumption that there is a constant probability of HIV transmission for a given type of sexual act for the whole population. In reality, some HIV carriers are more infectious than others, and there may be couple-dependent factors that affect the probability of transmission. Furthermore, a highly promiscuous partner is likely to have a greater prior probability of carrying the AIDS virus. Allowing for these considerations, promiscuity with protection is not necessarily the safer of the two scenarios.

In *What's Love Got to Do With It?,* anthropologist Meredith F. Small presents a personal, feminist take on the mating game. She adopts an intimate style, peppered with numerous anecdotes. The book succeeds in conveying a flavor of what participating in scientific research is all about.

Small's fascination with primate and human sexuality is clearly evident, and she goes into considerable detail about the physiological stages of sexual arousal. She argues that the traditional Victorian view of women as passive participants in sex is being overturned by recent findings, including those laid out in *Human Sperm Competition.* In fact, the view that women have an active sexual role dates back to the writings of Hippocrates and Galen. An unfortunate consequence of this belief, as Abramson and Pinkerton point out, was that women who conceived after being raped were sometimes branded as harlots.

The main thesis of *What's Love Got to Do With It?* is that women are just as motivated as men to have sex and just as promiscuous by nature. Small rejects evolutionary psychologists' findings to the contrary, arguing that these results are a consequence of cultural conditioning among both subjects and researchers. Her own analysis, however, is hardly objective, nor does she help her cause with a poorly informed discussion of the effects of AIDS on the preponderance of "gay" genes in the population.

By openly revealing her political perspective, Small is perhaps being more transparent and honest than many other researchers. And her book points to one of the central difficulties in studying sex and sexuality: they are so intimate a part of our lives that it is often difficult to separate scientific thesis from subjective belief.

PETER D. SOZOU is in the department of biology at Birkbeck College, University of London.

Walking in Water

Review by N. Katherine Hayles

OF TWO MINDS: HYPERTEXT PEDAGOGY AND POETICS, by Michael Joyce. University of Michigan Press, 1995 ($29.95).

Walking in water, catching a few sentences from the murmur of many voices, bicycling through a landscape that is also a text—these are some of the images that Michael Joyce, a professor of English at Vassar College, uses to describe the emerging form of electronic writing known as hypertext. Hypertext is the underpinning of CD-ROM multimedia and the World Wide Web; its intricate, nonlinear nature is revising the way everyone from research scientists to preschoolers interact with the written word. Writers since Gutenberg have dreamed of achieving "definitive" texts; editors have sought to determine which one of many versions of a manuscript was the most authentic; generations of students have learned to navigate canonical texts by means of chapter titles, page numbers and indexes. Hypertext challenges these traditional practices. In *Of Two Minds,* Joyce reflects on the new technology—how we will use it, and on what.

Stepping back from metaphor, Joyce quotes Theodor H. Nelson's definition of hypertext as "non-sequential writing with reader controlled links." In an electronic hypertext, a block of text on the computer screen contains embedded interactive elements—an icon, a word or phrase, or a concealed "hot spot" the reader finds, sometimes by trial and error. Clicking the cursor on an interactive element brings up another block of text, which in turn has other links leading from it. The text exists not as pages bound in a linear sequence but as a network of screens that the reader activates as he or she chooses.

A variety of elements can fit within that electronic mesh, customizing it to different applications. Hypertext may be easily combined with digitized images and sounds, as is becoming common practice on the World Wide Web. In literary studies departments, classic works are being encoded into hypertext format and integrated with critical commentary, historical information and graphics. Meanwhile modern writers (including Joyce) concoct interactive fictions, which resemble the "Choose Your Own Adventure" stories many of us read as children.

Hypertext destabilizes such fundamental notions as the text, the author and the reader. In a hypertext fiction, for example, there is not one story but a series of different narratives that

emerge in conjunction with the reader's choices. Because the reader actively collaborates with the author in bringing the narrative into existence, the distinction between the two fades away. For the author, hypertext means ceding control over the text, accepting the reader as partner, finding one's voice blended with a chorus of others.

The unity of the text also disperses. The best way to understand this process is to experience it. For instance, in Jon Lanestedt and George P. Landow's pedagogical hypertext *The 'In Memoriam' Web* (Eastgate Systems, 1992), certain sections from Alfred, Lord Tennyson's poem of that name are linked with other blocks in the poem that resonate with them, as well as with critical commentary—some of it written by students—interpreting the significance of the links. As one clicks from the poem to a student essay to another section of the poem to a bit of biographical information, the "text" ceases to be just the poem by itself and becomes instead the entire interconnected network.

More than one scholar has blanched at the prospect of turning the classic texts of Western culture into hypertexts. Joyce recounts a walk he took with a philosopher, who when he discovered what Joyce was up to asked plaintively, "You can't let the students change Plato, can you? Surely you can't let them do that." Joyce contemplates responding, "Which Plato?"—suggesting that "Plato" is already a hypertextual fiction, a composite that never existed in original purity. I find that answer somewhat disingenuous, for it underplays the transformative force of hypertext that elsewhere Joyce eloquently defends.

Better, to my mind, is when *Of Two Minds* confronts the meaning of hypertext head-on. Joyce's language soars when he writes of hypertext as a metaphor for the interconnectedness of contemporary life: "A constant murmur surrounds us and becomes palpable… as charged as the lives of those unfortunate souls we read about who dwell under high-voltage transmission wires. Surrounded by a surge of information, we spend our days, our hair standing on end, the fillings of our teeth complaining like the red-wing blackbirds perched on the thrumming wires above us; even at the center of our cells the proteins vibrate and mutate into some new and terrible variety of information."

In Joyce's book, phrases, sentences, even entire passages repeat from one chapter to the next, sometimes printed in italics to alert the reader to the repetition, sometimes cycled through without warning. The practice reminds us that this is a text written *on* a computer as well as written *about* computers.

JEFFREY SHAW "LEGIBLE CITY" (1989), *from collection of ZKM Karlsruhe*

VIRTUAL-REALITY INSTALLATION simulates the feel of hypertext.

When the University of Michigan Press invited me to read the manuscript of Joyce's book, my impression then was that there was too much repetition, and I voiced these reservations to the press. Because I was not in direct communication with the author and had no hand in soliciting the manuscript, I heard only indirectly, after the book was out, how strongly he felt that the repetition is essential to conveying through a printed text the feel of hypertext.

This difference of opinion highlights the fact that hypertext, compared with print, embodies a different rhetoric, a different aesthetic and different kinds of conceptual structures. Because virtually everyone older than 25 years in our culture has been raised on print rather than hypertext, it is inevitable that we come to hypertext with expectations formed from print. *Of Two Minds* attempts to bridge the gap between the print linearity and the electronic networking by using rhetorical looping to simulate, in paper form, the repetitions that can occur in hypertext.

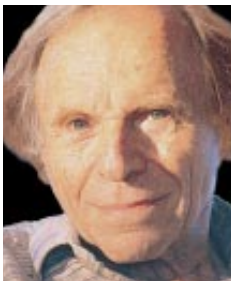
In an actual hypertext, variations in wording and differences of context quickly become significant in establishing fresh patterns and offshoots. If we think of the linear flow of text as its warp, the repeated sections are its woof; the idea is to weave strand into strand until the interconnections grow as dense and supple as silk.

Another way in which hypertext differs from print, Joyce argues, is through its topology, the virtual space created during the reading process. Unlike print, with its flat surfaces and linear sequences, the "two and a half dimensions" of hypertext mapping resonate both with how we know the world and with how we know our own bodies. Joyce quotes from Hélène Cixous, a French feminist critic: "I don't write," Cixous proclaims. "Life becomes text starting out from my body. I am already text. History, love, violence, time, work, desire inscribe it in my body." Hypertext creates a virtual expanse within the computer that (far more directly than words on paper) parallels the flow of our perceptions of external and internal space.

Anybody who has spent time on-line will instantly recognize the sensation that a world of connections—the fabled "cyberspace"—lies behind the computer screen. Joyce wants to call hypertext a "city of text," as though it were an urban landscape we negotiate through bodily movements. That vision is realized in "Legible City," a virtual-reality simulation created by the German artist Jeffrey Shaw. Shaw started with a model of a city block in Manhattan and replaced the buildings with letters of the same size. The user moves through the simulation by riding a virtual bicycle, reading the text as it goes by and choosing which path of text to follow.

Is "Legible City" a hypertext, a metaphor for a hypertext or simply a metaphor for life? In the same vein, we might wonder whether *Of Two Minds* is about computers, about hypertexts or about the increasingly common experience of writing and reading on computers. Perhaps, as Joyce repeatedly urges us to do, we ought to rephrase the thought as hypertext would have us do: " 'rejecting the objective paradigm of reality as the great 'either/or' and embracing, instead, the 'and/and/and.' "

N. KATHERINE HAYLES is in the department of English at the University of California, Los Angeles.



WONDERS

by Philip Morrison

Head Start on the 20th Century

As the millennium approaches, there is an irresistible impulse to look back and review the exhilarating twists and turns of scientific progress over the past 100 years. But real events seldom neatly follow the turn of the calendar. The revolution we call 20th-century physics, for instance, began about five years ahead of the century itself; the more accurate centennial is this very month.

I was lucky enough once to hear firsthand from a participant just how it was. About 20 years ago I was a fascinated luncheon guest of a man celebrating his 100th birthday. He was a lively, articulate, assured person, a true media magnate and a benefactor to the university of his hometown, the biggest city in his oil- and gas-rich prairie state. He had studied physics in his college days and was eager to tell me of his unforgettable encounter with a mysterious new kind of ray in Colorado Springs some 80 years before, on a cold day in January 1896, when he first heard tell of the German scientist Wilhelm Röntgen and his brand-new discovery of "x-rays."

The man had gone as usual to his morning physics lecture. His professor was genuinely elated. (I shall use quotation marks here, but I do not pretend that these statements are verbatim—not at all. They are only a reasonable retelling of the old story as I heard it.)

"Gentlemen," the lecturer opened, "this morning's paper carries an astonishing story. A German professor has found a new radiation able to reveal the bones within the living hand! More than that, we have right here in our laboratory all that is needed for the effect. If the story is true, we should be able to duplicate this marvel by lunchtime. The regular lab is canceled. Let's all work together, and perhaps we shall be the first in all America to see x-ray images."

The students were galvanized. The preparations went better than they had hoped. Pretty soon they had wired the induction coil that sparked the Crookes vacuum tube and brought together fluorescent screens, photographic film, and developer. My host recalled that as noon neared, he himself had run off to the college chapel to fetch a good-sized Bible, into which they would insert a coin whose shadow picture they would take through inches of paper.

Whether that professor and his students were the first in America to do such things one may doubt, for we do know that at Columbia, Yale, Harvard, Dartmouth, Cornell, Penn and Chicago, at least, the physicists had also heard the news and promptly sought to make pictures by x-ray. No precise times and dates are known to me to support anyone's claim to be first. (The Colorado College lab faced one intrinsic handicap—its longitude; the workday starts two hours earlier under Eastern Standard Time.)

Twentieth-century physics began about five years ahead of the century itself.

Of course, the point is not which was trivially the very first lab in America to follow Röntgen's lead but rather the great fact that this incredible discovery was not only irresistible to the public mind but latent and quickly achievable within most of the modest physics labs in the world. It took only the brief report of success to nucleate the rediscovery, just as a tiny seed crystal nucleates headlong crystal growth in a waiting supersaturated solution.

That sense of a long overdue, widely replicable discovery was something I myself lived through, four decades later in January 1939. We at Berkeley were told by telephone—soon enough every other nuclear physics lab would heed the news as well—how to see the just announced, energy-rich signature of the nuclear fission of uranium. The next day we all stared at the telltale pulses. World War II would begin that very autumn; one consequence of that chance historical juxtaposition was the nuclear bomb. (In the August 1995 issue of *Scientific American,* I wrote a fuller narrative of my own life with fission, then and ever since.)

X-rays opened a new path for physics, bringing with them new puzzles. Where the cathode rays hit the wall of the x-ray tube, they excited a glowing phosphorescent spot on the glass, and most of the x-rays came from there. That mysterious luminescence intrigued A. Henri Becquerel, who was, like his father and grandfather before him, a Parisian physics professor of talent. He had as a legacy his father's own study of glowing minerals and some excellent samples the elder Becquerel had prepared. Would similar glows that arose from sources other than the glass wall of the tube also emit x-rays? He tried out many phosphors, materials that glow eerily for a while after exposure to strong sunlight. None of them made any mark on a photographic plate through opaque black wrapping, as the x-rays had done.

Then he tried some "beautiful lamellas" of a uranium salt, unusual crystals that his father had grown 15 years earlier. On March 2, 1896, he reported that the uranium compound had indeed darkened the shielded photographic plate, even though the sample had by an accident of cloudy weather not first been exposed to sunlight. Many other uranium salts showed no phosphorescence and yet darkened the plate. Even uranium ores and metal sent out the penetrating rays. Before the end of 1896, Becquerel knew that uranium was somehow responsible, so he wrote of the "uranic rays." He did not yet grasp that the energy stored in uranium was unrelated to the phosphorescence of his father's uranyl sulfate crystals. That was a mere pun of nature; the radioactivity of uranium is quite independent of the splendid yellow-green glow of the sun-exposed crystals. But nuclear physics grew from that first coincidence.

Marie Curie, a young Parisian physicist, newly and happily married, was drawn late in 1897 to study Becquerel's rays, although by then nearly everyone else had chosen to work on the more controllable x-rays. She persuaded her physicist husband, Pierre, to let her try this new path to the doctorate. Within months, her promising finds induced him to abandon his own work in crystal physics to join her in a search well beyond uranium for the nature of "radioactivity," the word the two Curies would coin in their joint paper of summer 1898; a sample of the new element,

radium, was visible in their laboratory by the end of that year.

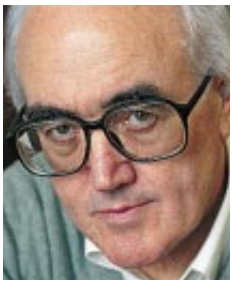
Radioactivity, x-rays and electrons lie at the base of 20th-century physics. The name of the electron somewhat predates the other two, because the idea of corpuscles of electrical charge was already pretty plain from electrochemistry. Many labs were active in tracking down the electrically charged particle during the final years of the 1800s. But the capstone was set by J. J. Thomson, who out of his and others' pioneering results was able for the first time in 1899 to specify both the mass and the charge of the electron. In our time, fast-shifting beams of electrons are present in nearly every household around the world, streaming down the length of a television picture tube. This device, which draws pictures by means of the swiftest of pencils, an electron beam, was called a cathode-ray oscilloscope in 1897 by its inventor, Ferdinand Braun of the University of Strasbourg.

Two key technological innovations also were born in the physics labs before the century turned. But they both depended more on the old physics, the magnificent classical physics founded by James Clerk Maxwell and Isaac Newton and so well mastered by the opticians and the engineers of the time; these innovations made no great call on elementary particles or quanta.

Movies are such a hallmark of our present century that we may be a little surprised to learn that the brothers Lumière opened their first projection cinema in Paris to paying audiences in the final week of 1895. Thomas Edison had his own movie success the previous year, but his version did not evolve from the personal peepshow to screen projection in front of an audience. And radio—as wireless telegraphy—grew vigorously under many inventors during that last decade of the 1800s.

For me, the clearest point of beginning of our technologically modern century is Guglielmo Marconi's first regular commercial link for wireless messages across the English Channel—early, naturally, for it started operation in the spring of 1896. Another now ubiquitous technology showed up even earlier: the automobile, developed in the 1880s.

But the case for the truly new 20th-century physics starting before its time is surely made by x-rays, electrons and nuclear physics, all begun before the century opened, before the rapid succession of other discoveries that define our era: Max Planck's 1900 energy quantum, Ernest Rutherford and Frederick Soddy's 1902 atomic transformations, Albert Einstein's 1905 photon.



CONNECTIONS

by James Burke

Breakfast Thoughts

I was rinsing dishes at the kitchen sink the other morning and thinking about this column when it occurred to me that what I was doing was (like everything, if you look long enough) a perfect example of the strange way things in the modern world are linked by events that happened along the great web of change. Things like the jet of high-pressure water I was using to wash the cornflakes out of my cereal bowl.

There was once a very early high-pressure water system set up outside 17th-century Paris, supplied by the river Seine and driven by a contraption of such humongous proportions that the local village name was changed from Marly to Marly-la-Machine. The machine in question was a river-spanning line of water-mill-powered pumps built to feed the ornamental fountains at the Palace of Versailles located a few miles away; water would shoot up into the air at considerable expense in order to amuse the king and his various mistresses. Now, extravagances such as fountains and mistresses are fine when your economy is in good shape. Which, by the late 18th century, France's was not. Come to think of it, neither was the king.

But everybody's cash flow improved in 1797, when Joseph Montgolfier, a balloonist papermaker (job descriptions could be like that back then) showed members of the new Republican government a device he had invented that would deliver the water to Versailles (and, more democratically, also to canals, irrigation networks and city water supplies) virtually free of cost or maintenance, on account of its having almost no moving parts.

Montgolfier's "Hydraulic Ram," along with its various escape valves, used the current of water in a river to compress air, which then drove spurts of water up, down or sideways as much as 120 times a minute. By the time of Montgolfier's death there were some 700 rams at work, all over Europe. None of these were delivering water to Versailles, however—no king, no fountains.

With all that high-pressure water, rams could deliver a lot of lifting power. In 1850 an English engineer called William Fairbairn used a variant on the ram to raise into place, at a rate of two inches a minute, a series of massive 1,200-ton, rectangular-section iron tubes (through which trains would pass) for the Britannia railroad bridge across the Menai Straits in Wales.

Fairbairn also hired a fellow by the name of Richard Roberts, who had invented an automatic riveting-control machine. The machine utilized perforated cards that Roberts had seen controlling silk-weaving looms (which I recently discussed in this column in quite a different connection). The cards were designed to block or permit the passage of sprung wire hooks. The hooks passing through the holes would then lift the particular threads relevant to that part of the pattern, so that the weaver's shuttle could pass underneath them. Roberts used his cards in a similar way to control the choice of size, number and position of rivet-hole punches to be used on a certain stretch of girder.

The same perforated-card idea was picked up again, to very different ends, later in the century by Herman Hollerith in the U.S., who ended up in business with some people who later changed their company's name to IBM.

Meanwhile Roberts's riveting technique worked so well that it caught the eye of a hotshot engineer, Isambard Kingdom Brunel. Brunel knew his new monster iron ship, the *Great Eastern* (in which he was planning to use the same tube-girder structures that had performed so successfully in the Menai Bridge), would need at least three million rivets for the hull alone.

As it turned out, riveting was about the only thing that went well for the ship. The engineers built her parallel to the banks of the Thames, only to find that the river was too far away for a proper launch, and so at a cost of months and megabucks (well, megapounds), they were obliged to move her using rams, slides, winches, cradles and all sorts of kit. Even then, it took six attempts before she was finally in the open water. On the *Great Eastern*'s maiden run to America, so many catastrophes had already befallen the ship that

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: $8.95 each ($9.95 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.

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. We occasionally make subscribers' names available to reputable companies whose products and services we think will interest you. If you would like your name excluded from these mailings, send your request with your mailing label to us at the Harlan, IA address. E-mail: SCAinquiry@aol.com.

only 38 out of 300 first-class passenger berths were taken—and even then there was a final day's delay due to drunkenness among the crew. Things went from bad to worse, so that by 1865, the biggest ship in the world ended up with nothing better to do than to lay the latest transatlantic submarine telegraph cable. And then to lose it when the roll of the ship snapped the cable. And then to find it again the following year.

The cable was still in good working order when the engineers finally joined the broken ends back together, because its sheathing had been made of the latest wonder-gunk: gutta-percha, a kind of latex sap from Malaysian trees. But gutta-percha did more serious things than allow President James Buchanan to open the transatlantic telegraphic link with a few statesmanlike remarks to Queen Victoria. It also provided the first electric light insulation in homes and put chewing gum on the market. And it revolutionized the game of golf.

Up to this time, golf balls had been stuffed with feathers. The problem was, most of these "featheries" lasted about as far as the third green. Then, in 1848, along came gutta-percha and a solid, moldable ball, spherical enough to go where you hit it and hard enough to stay in shape for longer than one game. The new balls were very grudgingly accepted by the jealously traditional owners of Royal St. Andrew's (the world's foremost golf club, and they knew it), where the balls changed the shape of golf clubs in more senses than one.

Gutty balls, as they were called, were so cheap that, by the 1850s, the game was attracting trainloads of holiday-making Scottish workingmen. To accommodate the crowd, St. Andrew's was obliged to split each of its (very wide) fairways longitudinally into two, each half to be played in opposite directions. Which is why there are now 18 holes on a standard golf course instead of the original nine.

An early game at St. Andrew's must have been one of those rare occasions when the hot air you hear so much on golf courses was literally all about hot air—hot air being the reason why there were suddenly so many working-class wannabe golf champs keen to improve their swing. At the time, the river Clyde at Glasgow was trying to become the world's greatest shipyard, and failing. The one thing any would-be industrial region needed was coal, but although there were thousands of tons of it buried below the Scottish Lowlands, the stuff was of such a low grade that it would hardly make toast, let alone smelt iron.

Enter James B. Nielson, boss of the Glasgow Gasworks. In 1827 he invented the gas-fired, hot-blast furnace, which could generate hearth temperatures high enough to make any fuel burn. Using the Nielson process, you could not only make iron using the local junk-quality coal, you were able to make three times more of it than you could using the good stuff. Hence, within a few years came the Scottish industrial revolution, Clyde-side shipbuilders and the sootiest cities in the world. And, because of the new, iron-driven wealth, some of the best golfers anywhere.

Apart from opening up Scottish coal mines, Nielson's technique was also hot news for the previously impoverished owners of the Pennsylvania anthracite coal deposits, similarly hard to burn but now suddenly profitable. In no time at all, Pittsburgh was becoming Steel Town U.S.A., and railroads were being built to bring iron ore down from the Great Lakes (in the process revolutionizing the entire business world by means of new railroad administration tricks, such as cost accounting, monthly returns, divisional structures and departmental management).

The burgeoning anthracite-fired iron and steel industry was soon littering the Pennsylvania landscape with large amounts of used coke. As it happened, Pennsylvania was where the great English chemist and American Revolution sympathizer Joseph Priestley had spent his declining refugee years. And it was he who had noted that coke is a superb conductor of electricity. Drawing on Priestley's fortuitous discovery, a resident of Pittsburgh called Edward Acheson experimented with putting coke and clay into an electric furnace. By 1885 Acheson was coming up with the second-hardest stuff in the world, a material he named carborundum.

Not surprisingly, Acheson soon went into the abrasive business. The bits of carborundum that he stuck on grinding wheels created quite a stir, which helped Acheson win the contract to manufacture the lights with which George Westinghouse would illuminate the 1893 World's Colombian Exposition in Chicago. Those same grinding surfaces are still used, only now they are usually resin-bonded to their wheel by a process involving a solvent called furfural.

Furfural is a chemical you get when you add sulfuric acid and water at high pressure to a mixture of throwaway plant by-products: oat husks, bagasse, rice hulls—and also the cobs you have left over, once you've made cornflakes like the ones that I was so deftly cleaning off my dishes when I set out along this set of connections.

ESSAY by Christian de Duve

The Constraints of Chance

It has become fashionable for biologists to emphasize the role of contingency in the origin and evolution of life on the earth, including the advent of humankind and the development of mind. Those momentous events are said to be the products of highly improbable combinations of chance occurrences. As the late Jacques Monod wrote in his 1970 best-seller *Chance and Necessity,* "the universe was not pregnant with life, nor the biosphere with man." Often presented as established fact, such affirmations are seen as driving the final nail into the coffin of whatever illusion we still entertain about the human condition and its significance in the universe. When examined critically, however, the science behind this view emerges as less conclusive than is commonly believed.

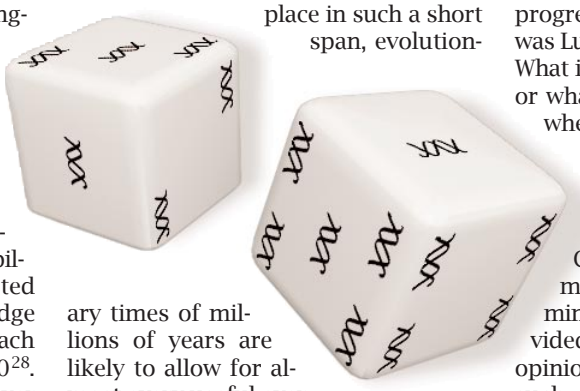
The thesis that the origin of life was highly improbable is demonstrably false. Life did not arise in a single shot. Only a miracle could have done so. If life appeared by way of scientifically explainable events, it must have followed a very long succession of chemical steps leading to the formation of increasingly complex molecular assemblages. Being chemical, those steps must have been strongly deterministic and reproducible, imposed by the physical and chemical conditions under which they took place.

The involvement of many steps reinforces their deterministic character. Single events of very low probability readily take place, but a connected string of such events does not. Bridge hands are being dealt all the time, each with a probability of one in 5×10^{28}. But the same hand is essentially never dealt twice, let alone many times, in succession. Given the nature of matter and given the conditions that existed on the earth four billion years ago, life was bound to arise in a form not very different, at least in its basic molecular properties, from its present form.

What now of the probability of evolution producing conscious beings? Here the proponents of contingency seem to stand on safe Darwinian ground. Hardly any biologist today doubts that every evolutionary step starts with a fortuitous heritable change, the outcome of which is then tested by natural selection. The obvious implication is that chance governs the directions of evolution. This is the majority opinion—justified, but in need of qualification.

Chance does not exclude inevitability. Of critical importance are the constraints within which chance operates. One is the number of options. There are only two possibilities when a coin is flipped, six when a die is cast, 36 when a roulette wheel is spun and 5×10^{28} when a hand of bridge is dealt. The number may be large, but it is always finite. So it is with possible mutations. Their number is not only limited, it is not even extremely large, relatively speaking. This point is readily corroborated by experience.

Antibiotic-resistant bacteria, chloroquine-resistant malarial parasites, DDT-resistant mosquitoes and herbicide-resistant weeds all have appeared in the course of a few decades—not thanks to fluke mutations but because the spread of the drugs has suddenly given banal mutations an opportunity to prove beneficial and be selected. If wide-ranging changes of this kind can take place in such a short span, evolutionary times of millions of years are likely to allow for almost every useful eventuality. Contrary to a widespread notion, evolution does not so much follow the vagaries of chance mutations—although this may occasionally happen—as do mutations wait, so to speak, for an opportunity to affect the course of evolution.



In multicellular organisms, existing body plans impose additional constraints on evolution. Effective mutations are restricted to the small number of genes that control the development of an organism—such as homeotic genes—and must be such as to modify the developmental blueprint in a manner conducive to evolutionary success or at least compatible with it. Most of the changes that meet these conditions do not alter the basic body plan. They characterize what I call "horizontal" evolution and lead to biodiversity. There are probably more than one million species of insects, but all are insects. It is in this kind of diversification that contingency plays its leading role, mostly in the form of environmental conditions that happen to provide some mutation with a selective advantage.

Much fewer, because far more constrained, are the changes that significantly increase the complexity of body plans ("vertical" evolution). The constraints no doubt leave room for developments that failed to happen on the earth but could happen elsewhere, and vice versa. But some directions could be compelling. The emergence of thinking beings, for instance, appears much less improbable than is often intimated. Once neurons emerged and started interconnecting, life progressed toward the formation of increasingly complex networks, no doubt furthered by the associated selective advantages. Six million years ago a chimpanzee's brain represented the apex of this evolutionary progression. Three million years ago it was Lucy's. Today it is the human mind. What it will be six million years hence—or what has already materialized elsewhere—is anybody's guess.

Life and mind appear as cosmic imperatives, written into the fabric of the universe. Given the opportunity, matter must give rise to life, and life to mind. Conditions on our planet provided this opportunity. The prevalent opinion among cosmologists is that such conditions may prevail on many other sites in the universe. If so, and if the views defended in this essay are correct, there must be many other living planets, at least a fraction of which have evolved or shall evolve toward the formation of conscious beings—some perhaps more advanced than we.

CHRISTIAN DE DUVE, co-winner of the 1974 Nobel Prize for Physiology or Medicine, is professor emeritus at the University of Louvain, Belgium, and also at the Rockefeller University. He is the author of Vital Dust: Life as a Cosmic Imperative *(BasicBooks, 1995).*