

SCIENTIFIC AMERICAN

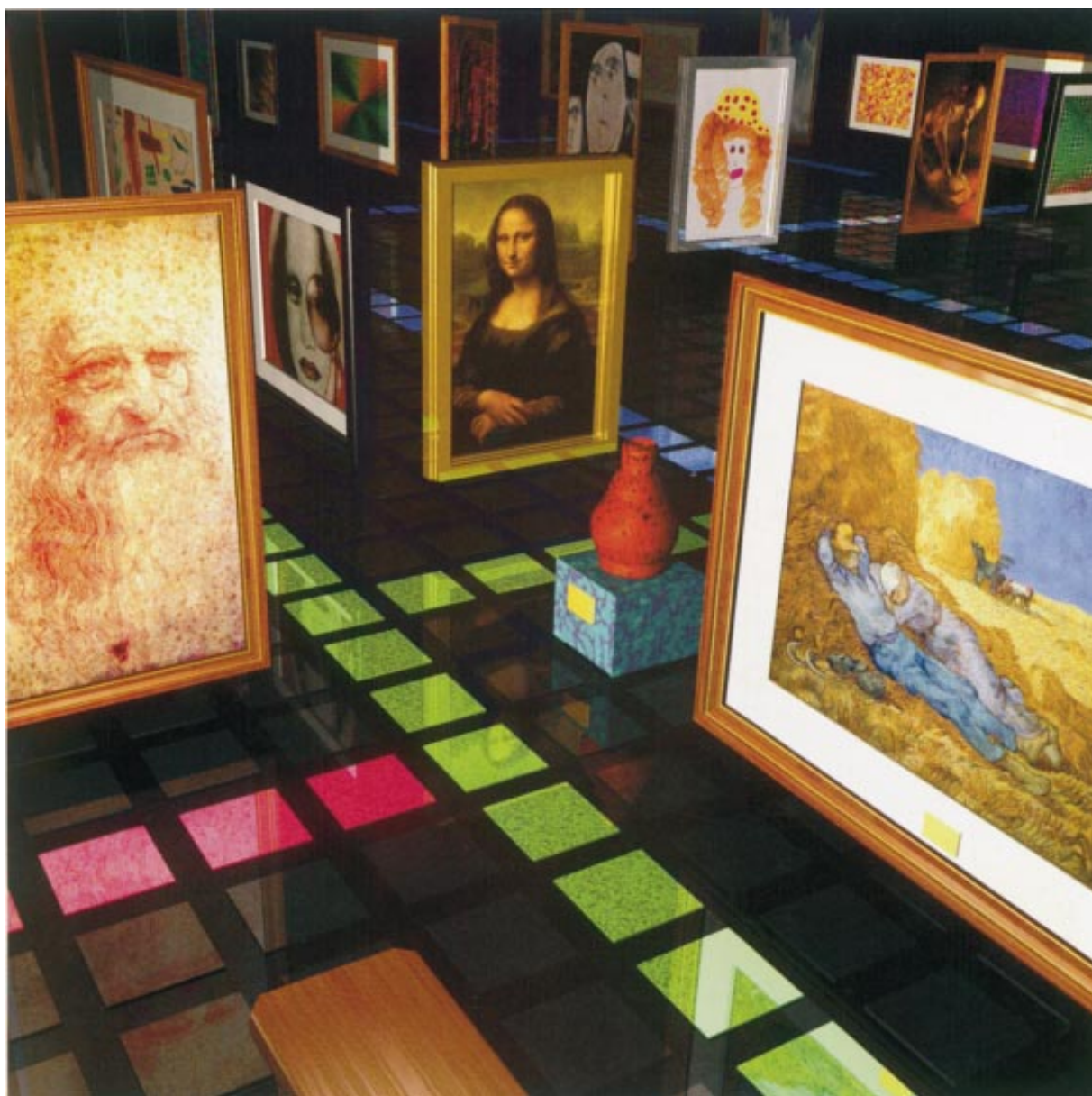
APRIL 1995

\$3.95

Machines that learn from hints.

Why frogs are vanishing.

Where the solar wind ends.



*Virtual museum of digitized art
exists only inside a computer.*

52

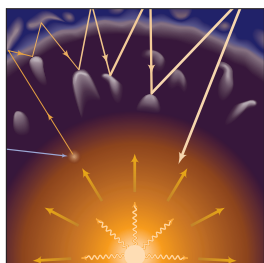


The Puzzle of Declining Amphibian Populations

Andrew R. Blaustein and David B. Wake

Frogs, toads and salamanders survived even the catastrophes that finished the dinosaurs, yet recent censuses suggest that many species are now mysteriously dwindling or disappearing. The destruction of their natural habitats, pollution, disease, changes in the ozone layer and even tastes in haute cuisine may be at the bottom of this ominous development.

58



Quest for the Limits of the Heliosphere

J. R. Jokipii and Frank B. McDonald

Far beyond Pluto—no one yet knows how far—there is a discontinuity in the near-vacuum of space. It marks the edge of the heliosphere, where the diffuse solar wind collides sharply with the tenuous interstellar medium. Some unusual cosmic rays that bathe our planet originate in this region. Now the *Pioneer* and *Voyager* spacecraft, their original missions completed, are heading there.

64



Machines That Learn from Hints

Yaser S. Abu-Mostafa

Computer scientists know how to build machines that can learn from examples, but how can those machines learn more efficiently? Here's a hint: give more of the right background information. Although computers, unlike humans, do not intuitively understand much about the real world, hints in the form of instructive examples can teach them important principles.

72



Understanding the Genetic Construction of Behavior

Ralph J. Greenspan

Beware of simplistic statements about the genes for complex human traits; the actual state of knowledge about behavioral genetics is crude. Consider what has been laboriously discovered about one well-defined behavior—courtship—in the fruit fly *Drosophila melanogaster*. All the results suggest that even in relatively simple organisms, behavior is influenced by a multitude of genes.

106



SCIENCE IN PICTURES

The Art Historian's Computer

Lillian Schwartz

Did Leonardo da Vinci complete the *Mona Lisa* as a self-portrait? Is Queen Elizabeth I hiding inside an engraving of Shakespeare? Computer graphics can sometimes find the answers to questions that confound more traditional analyses.

112



A Brief History of Infinity

A. W. Moore

The concept of infinity has been boggling minds for at least 2,000 years. The philosophers Zeno and Aristotle did their best to sidestep it; the Pythagoreans were horrified by its inescapability; mathematician Georg Cantor came closest to taming it, but some questions remain. The truly boundless may be beyond comprehension.

118



The Tapestry of Power in a Mesopotamian City

Elizabeth C. Stone and Paul Zimansky

Archaeologists have usually assumed that the world's first cities had centralized, authoritarian social structures. But detailed studies of Mashkan-shapir, a site in Iraq, indicate that its rich and poor citizens lived cheek by jowl and that politics and religion were peripheral.

124



TRENDS IN HEALTH CARE

The Price of Prevention

Kristin Leutwyler, staff writer

Health care policymakers on the lookout for medical cost-savings are in for a rude shock: an ounce of prevention is not *always* worth a pound of cure. The grim truth is that treating the general population to prevent disease now is usually more expensive than paying to treat the sick later. That fact leaves physicians, politicians and the public facing some uncomfortable choices.

DEPARTMENTS

14



Science and the Citizen

The Kobe quake's engineering lessons.... Oil rigs take to deeper waters.... Muscle maladies.... An anti-addiction drug.... Homosexuals in science.... X-ray of solar flares.... Atoms that act like light.

The Analytical Economist

A little statistics can be dangerous.

Technology and Business

The uncertainty of gene patents.... New aviation technology: still up in the air.... Diamonds for micro-machines.... Fish food in Calcutta.

Profile

Engineer Arati Prabhakar takes on Congress.

10



Letters to the Editors

This is no joke: real mail from real readers.

12



50 and 100 Years Ago

1945: High-tech metallurgy.
1895: The first fax machine.

130



The Amateur Scientist

Computers rescue art of the Young Masters.

135



Book Review: Daniel L. Schacter

Critical looks at "repressed memories" of abuse.

140



Essay: Jeremy Bernstein

Making Quot a fuss over intelligence.



THE COVER depicts a virtual museum, an infinitely expandable gallery of great art that can exist only within a computer. The 3-D image was constructed using Infini-D™ 2.6. Four Macintoshes took approximately 86 hours to render the image. Among the paintings are *Self-Portrait* and the *Mona Lisa*, by Leonardo da Vinci; *The Siesta after Millet*, by Vincent van Gogh; *Self-Portrait*, by Lillian Schwartz; and *Portrait of My Sister*, by Jamie S. Feigenbaum. (See "The Art Historian's Computer," by Lillian Schwartz, page 106.) Image by Slim Films.

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; Philip Morrison, *Book Editor*; 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, *Associate Art Director*; Johnny Johnson, *Assistant Art Director*; Nisa Geller, *Photography Editor*; Lisa Burnett, *Production Editor*

COPY: Maria-Christina Keller, *Copy Chief*; Nancy L. Freireich; 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, *Special Projects*; Carol Hansen, *Composition*; Madelyn Keyes, *Systems*; *Ad Traffic*: Carl Cherebin; Carey S. Ballard; Kelly Ann Mercado

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

ADVERTISING: Kate Dobson, *Associate Publisher/Advertising Director*. OFFICES: NEW YORK: Meryle Lowenthal, *New York Advertising Manager*; Randy James, Rick Johnson, 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*

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*; R. Vincent Barger, *Chief Financial Officer*; Robert L. Biewen, *Vice President*

DIRECTOR, ELECTRONIC PUBLISHING: Martin Paul

POST-POLIO SYNDROME

EDITED BY

Lauro S. Halstead, MD
Director, Post-Polio Program
National Rehabilitation Hospital
and Georgetown University
School of Medicine
Washington, D.C.

Gunnar Grimby, MD, PhD
Chairman, Department of
Rehabilitation Medicine
University of Goteborg
Goteborg, Sweden

Many people who survived the paralytic poliomyelitis epidemic of the 1950s are now being stalked by the post-polio syndrome. This syndrome is a diagnostic and treatment challenge to physicians taking care of post-polio patients. Symptoms vary. Pathogenesis is elusive. Its course is unpredictable. Indeed, it is so imprecise a condition that some challenge its very existence; but those who suffer from it are not among the challengers.

POST-POLIO SYNDROME is a significant contribution to our understanding of this often misdiagnosed syndrome and covers all aspects of clinical assessment and management of the patient. The editors are widely recognized as pioneers in the recognition and treatment of post-polio syndrome.

CONTENTS

CLINICAL PATHOLOGIC CORRELATIONS

Muscle Function, Muscle Structure and Electrophysiology in a Dynamic Perspective in Late Polio

Muscle Fiber Morphology in Post-Polio Patients

Immunological Role in Post-Polio Syndrome
Local Muscle and Total Body Fatigue

CLINICAL ASSESSMENT AND MANAGEMENT ISSUES

Prognosis and Differential Diagnosis

Evaluation and Management of Post-Polio Respiratory Sequelae: Noninvasive Options

Respiratory Management in Late Post-Polio

Long Term Effects of Post-Polio on Oral-Motor and Swallowing Function

Strategies for Exercise Prescription in Post-Polio Patients

Functional Limitations and Disability in Post-Polio

PERSONAL AND PSYCHOSOCIAL ISSUES

Psychosocial Issues and Post-Polio: A Literature Review of the Past Thirteen Years

The Lessons and Legacies of Polio

Growing Old with Polio: A Personal Perspective

January 1995 • 230 pages • hardcover
\$32.95 (U.S.) • \$37.95 (outside U.S.)

Send orders to: **HANLEY & BELFUS, INC., Medical Publishers**
210 South 13th Street • Philadelphia, PA 19107
Or Call: 1.800.962.1892 • 215.546.7293 • FAX 215.790.9330



LETTERS TO THE EDITORS

Breaking Research

If you believe that at some point in the future mankind will develop a virtual-reality technology that allows a user to experience a virtual life while believing it to be 100 percent genuine, how can you know whether the life you are living now is real or virtual? You can't. But I've created a formula that determines the exact chance that your life is either real or virtual. Remarkably, the answers, using conservative values, have generally ranged from 25 to 50 percent—a one-quarter to one-half chance that our lives are virtual, not real. I look forward to working with you on this project.

DAVE PACHECO
Richmond, Calif.

For every evolutionary advance, there is an equal and opposite regressive step to be offset. At present, the *L. bifidus* in mother's milk provides a turning point in evolution, protecting us from returning to something like Neanderthal man or even the monkey. The existence of "negative evolution," which might broadly be described as "inertia," is being ignored.

J. GORDON ROBERTS
Clearwater, Fla.

Having discovered the "Harmonic Cube," I will wager \$10,000 that I am the wisest human of all time.

GENE RAY, Cubic
St. Petersburg, Fla.

Out in Space

I am of the opinion that cosmological attraction, or gravity, does not exist. I believe that a cosmological repulsion force exists and that what we perceive as gravity is merely a shadowing of this force by one celestial body on another. The expanding universe and the Hubble constant are thereby more readily explained.

BRIAN DAVIDSON
Donegal, Ireland

The news story "Gone with a Bang," by Corey S. Powell [SCIENTIFIC AMERICAN, September 1994], describes "run-

away pulsars" with an average velocity of 450 kilometers per second, enough to escape from the Milky Way. Scientists cannot explain this phenomenon. The obvious explanation: space travel by intelligent beings. By using a pulsar for a giant spaceship, an entire solar system can travel anywhere in space. Entire civilizations can have all the comforts of their home planet during the journey.

LEROY PETERSON
Mesopotamia, Ohio

What's Past Is Prologue

If you and some others with a dictionary go into a huddle with the intent to find and publish some proof that time travel is not impossible, then the dictionary will prevent anything from getting out, unless the dictionary contradicts itself.

MARC CUNNINGHAM
Baton Rouge, La.

After reading your article "The Quantum Physics of Time Travel," by David Deutsch and Michael Lockwood [SCIENTIFIC AMERICAN, March 1994], I thought you would be interested in the following information. A well-known psychic artist has been placed in charge of an experimental group that is supposed to be able to record on videotape consistently clear scenes and sound from either past, present or potential future. The locations of these incidents can be closeups of anywhere in our Milky Way galaxy, even inside spacecraft and inside dwellings on any planet.

JEFFREY BLUNT
Springfield, Mass.

The Tough Questions

I am researching the effects of musical notes on the chemical elements, relating how certain sounds of the musical scale could influence chemicals. Example: What note on the musical scale could change or alter the molecular structure of the element sulfur?

DON DREIS
Bloomington, Ind.

We have heard that humans are a fluke of the universe. But what I want to know is: Does charge-parity violation mean that the universe is a fluke of itself?

JOHN W. WALL
San Francisco, Calif.

I failed to understand David Z. Albert's article "Bohm's Alternative to Quantum Mechanics" [SCIENTIFIC AMERICAN, May 1994]. It does not appear to contain a description of the proposed alternative. Do you have editors?

MICHAEL WALSH
St. Genis, France

Adam's Rib Revisited

As a sometime recipient of the Male Chauvinist Pig of the Year Award in the state of Victoria (where standards are high), may I felicitate you on having the courage and guile to publish the "Trends in Women's Health" article, "A Global View," by Marguerite Holloway [SCIENTIFIC AMERICAN, August 1994]. As providing evidence that women are incapable of objectivity or indeed any degree of rational thought, that they are credulous victims of fashion and wholly devoid of any critical capacity, that they confound sententious drivel with clear writing, that they are, in short, brainless bimbos down to the last individual, the article is faultless.

MICHAEL ALDER
Nedlands, Western Australia

Not Everybody's a Critic

For the life of me, I cannot understand why you people do not run an annual letters issue. It would be a better seller than the *Sports Illustrated* swimsuit issue.

MINAS ENSANIAN
Buffalo, N.Y.

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

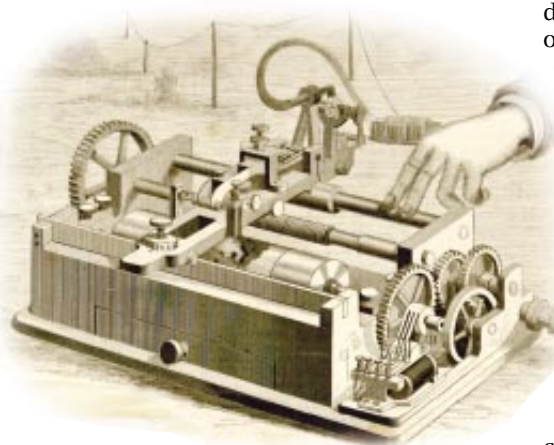


50 AND 100 YEARS AGO

APRIL 1945

Wool, which retains its original appearance, yet is protected against shrinkage, even after repeated laundering and dry-cleaning, is now commercially available through the use of a new synthetic resin known as Lanaset. When applied to the fabric, Lanaset stabilizes wool and wool blends without affecting the absorbency normally characteristic of wool. It also reduces felting and prevents fuzzing."

"Metallurgists, in the examination of metal surfaces with the optical microscope, have long recognized serious weaknesses in this procedure. The surface under observation very frequently does not offer adequate evidence of the true shape of the details and in many cases is even misleading. However, a newly developed microradiographic technique is able to give an indication of the three-dimensional contours of the metal structure and is able to indicate the distribution and identity of the chemical components of the metal in a precise manner."



The electro-artograph transmitter

"A recent development that increases flight safety by keeping airplanes' propellers free of ice consists in painting or spraying the propeller blades with a chemical lacquer called Icelac, which, black in color, has a consistency something like that of glycerine, and paints or sprays freely to give a shiny, tacky surface. A satisfactory icing protective surface is maintained for several thousand hours."

"Less than half the farms of the United States have electricity available. True, high-tension lines have brought this versatile servant to hundreds of thousands of farms, but the market has actually only been barely touched. High-tension lines are expensive and cannot be run everywhere."

"After standing overnight in zero degree weather, buses now receive a quick boost in temperature from a 'Janitrol' portable heater of the type developed to preheat airplanes at Alaskan air bases. A heat rise of 230 degrees enables the appliance to deliver positive heat in sub-zero temperatures."



APRIL 1895

Under the combined influences of great pressure and intense cold, hydrogen has at last surrendered and been liquefied. Hydrogen has hitherto most strenuously resisted all attempts at liquefaction, and the fact of its obduracy in this respect having now been overcome removes the only gaseous element known to us which has not been liquefied. Until, therefore, more attenuated gases are added to the list of chemical simplicities, no further discoveries on this particular line of research can be hoped for."

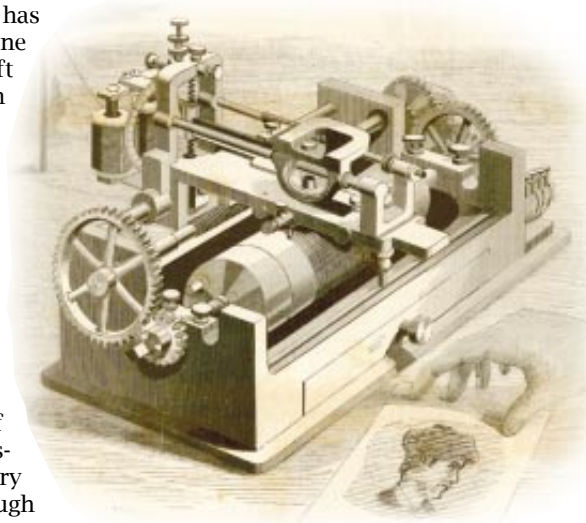
"A simple pneumatic cushion, with a soft touch to the ear, has been adapted to fit all telephone receivers. It is made of soft rubber, fitted into a metal rim which springs or clamps over the end of the receiver, forming a complete air chamber designed to effectually prevent the buzzing or clucking sounds so annoying to users of the telephone."

"Much has been written as to the picture that the compound eye of insects produces upon the brain or upon the nerve centers. It is obvious from the structure of these compound eyes that impressions through them must be very different from those received through our own. In point of fact, experiments have practically established that while

insects are shortsighted and perceive stationary objects imperfectly, their compound eyes are better fitted than the vertebrate eye for apprehending objects in motion, and they are likewise keenly sensitive to color."

"An ideal school room should provide fifteen square feet of floor space for each pupil and a supply of 200 cubic feet of air per minute for every person in the room. Such provisions would ensure the free movement of every child and a wholesome amount of air. In France, the perfect school room, it is thought, should have a window area equal to one-fourth the floor space. It is also thought best to have individual seats and desks for the pupils."

"When the telephone was introduced to the attention of the world, and the human voice was made audible miles away, there were dreamy visions of other combinations of natural forces by which even sight of distant scenes might be obtained through inanimate wire. It may be claimed, now, that this same inanimate wire and electrical current will transmit and engrave a copy of a photograph miles away from the original. As shown in the accompanying illustrations, the electro-artograph, named by its inventor, Mr. N. S. Amstutz, will transmit copies of photographs to any distance, and reproduce the same at the other end of the wire, in line engraving, ready for press printing."



The electro-artograph receiver



SCIENCE AND THE CITIZEN

Bracing for the Next Big One

Engineers grapple with retrofitting Japanese and U.S. buildings

The quake that struck Kobe, Japan, on January 17 could hardly have been better timed for impact, literally, on the structural engineering community. When it hit, dozens of engineers were jolted awake in nearby Osaka, where a Japanese-U.S. workshop on "Urban Earthquake Hazard Reduction" was to begin later that morning. Had they not been so rudely roused, the engineers would have heard case stud-

Mountain View, Calif., who was at the Osaka meeting. "There's enormous learning potential here."

Much has been made of the differences in philosophy between Japanese and American engineers: the former supposedly emphasize strength; the latter ductility, the quality that lets a structure bend and deform. But they share fundamental similarities. Whether in Tokyo or Los Angeles, steel-and-concrete con-

crete retaining hoops, known as confinement steel, encircle the longitudinal poles to hold the concrete in place. Similar changes occurred in Japan after a pair of earthquakes, one in Tokachi in 1968 and the other in Miyagi prefecture a decade later.

Before these improvements, though, tens if not hundreds of thousands of nonductile concrete buildings had been constructed in Japan, and Kobe's fared poorly in the earthquake, according to several U.S. engineers who toured the city. The good news is that most of the structures put up after the early 1980s—

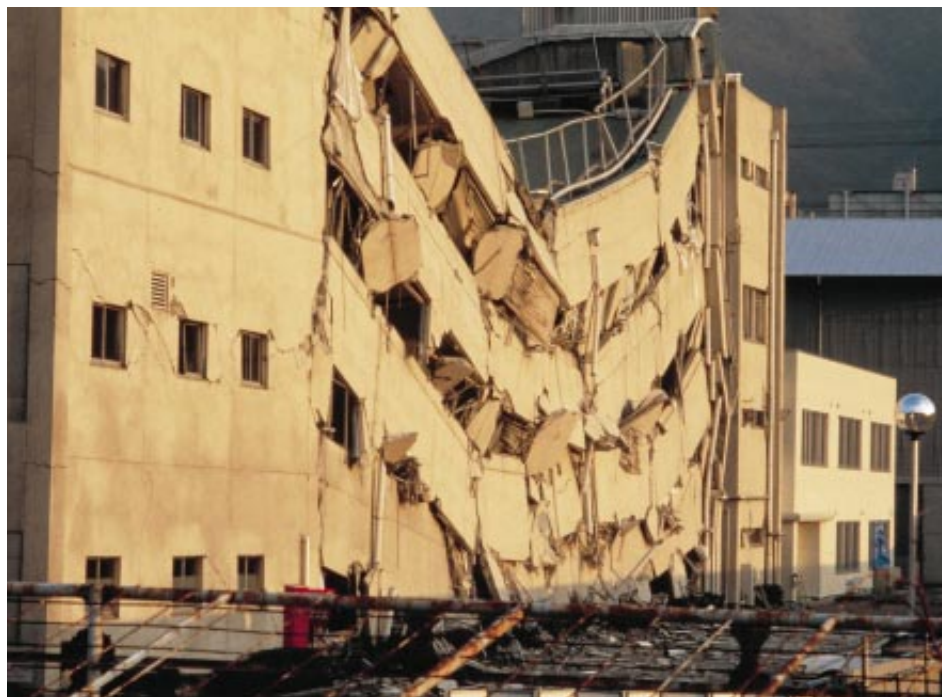
including ones based on steel-reinforced concrete and others erected around steel frames—survived without substantial visible destruction. "I saw dozens and dozens of those buildings that had no damage apparent from the outside," says Loring A. Wyllie, Jr., a senior principal with Degenkolb Engineers, a San Francisco structural engineering firm.

It will be months before a more complete picture of the damage emerges from Kobe, and there will undoubtedly be surprises. Only recently has it come to light, for example, that steel-frame buildings suffered more harm in the Northridge quake than had been thought. Such structures are considered more pliant than ones based on steel-reinforced concrete and therefore more resistant to powerful earthquakes.

A study by the Chicago-based American Institute of Steel Construction, however, has found that roughly a quarter of the

400 steel frames in the immediate area of the Northridge quake suffered cracks in the welds at the joints where columns meet girders. The buildings ranged in age from one to 31 years, and although none of them visibly listed or showed any other sign of distress, a number of them have been evacuated for repairs, and one was torn down.

Improvements are needed in the way steel frames are being built, states James O. Malley, a principal with Degenkolb Engineers. Exactly how this might be done is the point of a \$2.3-million proj-



BUNYO ISHIKAWA SYGMA

SAGGING STRUCTURE in Kobe, Japan, is testament to the January earthquake's ravages—and the need to combine architectural strength with flexibility.

ies from a comparable quake that had rocked Northridge, Calif., exactly one year earlier.

The Kobe quake was the second most devastating in Japanese history. Barreling through a densely populated port city that had not been thought to be especially at risk, it killed more than 5,000 people and destroyed or damaged some 50,000 buildings. "There is 10 times as much data as what we've gotten from California, because there's been 10 times the destruction," says Charles Kircher, a structural engineering consultant in

struction combines the metal's ability to withstand tension with the mortar's resistance to compression. This approach underlies nonductile concrete, in which steel bars run longitudinally through structural elements.

In 1971 an earthquake in the San Fernando Valley of California called attention to the method's basic shortcoming, namely, that powerful forces can cause the concrete to shatter and fall away from the inner steel. Within a few years, new building codes called for steel-reinforced concrete, in which sep-

ect nearing completion by a group called the SAC Joint Venture, consisting of the Structural Engineers Association of California, the Applied Technology Council and some California research universities.

With funding primarily from the Federal Emergency Management Agency, the collaboration has investigated such issues as joint design, welding practices and materials, frame design and the

minimum number of beams and columns, or "redundancy," needed for flexible support. It plans to issue interim guidelines in May. "A lot of stuff has to change," Malley notes.

Meanwhile California has ordered the retrofitting or at least investigation of thousands of unreinforced masonry buildings to ensure that they will not collapse in a moderately strong quake. Of course, that still leaves thousands of

nonductile concrete structures, which, so far, neither California nor Japan has mustered the political will to address.

"Our technological capabilities really exceed what society's willing to pay for," observes Craig Comartin, an engineering consultant in northern California. "In some cases, retrofitting buildings is extremely costly. I'm not whining; it's just a fact. It's a pay-me-now or pay-me-later kind of thing." —Glenn Zorpette

The End of the Road

Is a new malady afflicting elite athletes?

Announcing his retirement from competitive bicycle racing last December, Greg LeMond brought to a close a career marked by sublime athletic achievement. He had won the Tour de France three times and the world championship twice; some of these victories came after his near death in a 1987 hunting accident. LeMond, at age 33, has now also entered medical record books, becoming the first elite

myopathy should not prevent him from pursuing an active life.

The diagnosis raises questions about a disorder whose nature and prevalence are just now yielding to medical inquiry—and it underscores the difficulties physicians face when attempting to determine the maladies of athletes. Taube believes LeMond suffers from a previously unknown, yet not necessarily uncommon, form of this debilitating

changes in your metabolism and immune system if you overtrain. You may change your body's ability to adapt and leave it open to illness."

The condition LeMond is thought to have differs markedly from exertional myopathy, a more commonly known muscle disorder that can affect anyone who exercises beyond his or her capacity. High school football players and military recruits seem especially susceptible, perhaps because many of them start vigorous activity in hot weather without much advance training. Dehydration makes it tougher for muscles to recover and for the body to purge itself of wastes.

Reports have also tied exertional myopathy to HIV infection, as well as to the use of cocaine, methamphetamines, LSD, alcohol and various prescription drugs. Some of these drugs allow people to engage in repetitive activities for long periods without proper rest or fluid consumption. As people overexert, lactic acid builds up, and levels of muscle enzymes, such as creatine kinase—essential to muscle contraction—rise dramatically. Usually the resulting soreness passes, and enzyme levels return to normal in a few days, but in acute cases, the muscle cells rupture, flooding the bloodstream with myoglobin, enzymes and minerals.

Because exertional myopathy generally occurs when muscles are burning energy faster than they are being resupplied, the problem should vanish as a person's fitness improves. Mitochondria—which use oxygen to produce adenosine triphosphate (ATP), which, in turn, fuels cells—increase in so-called slow-twitch muscles during training, allowing them to process more oxygen more efficiently for longer periods without lactic acid buildup and muscle cell damage.

Although mitochondrial myopathy also affects the skeletal muscles, it is a far different disorder, notes John Shoffner of Emory University. Still a medical enigma, mitochondrial myopathy is one of a broad class of oxidative phosphory-



MICHEL GOUVERNEUR Gamma Liaison

CHAMPION GREG LEMOND retired this winter because of a muscle disorder.

athlete to be diagnosed with mitochondrial myopathy, a disorder that impairs muscle. He reported that, for mysterious reasons, his skeletal muscle cells could no longer use oxygen to produce the energy required for him to perform at peak capacity.

"Greg is the first trained athlete to be diagnosed with this condition," says LeMond's physician, Rochelle Taube of the Minneapolis Sports Medicine Center. "Usually people with mitochondrial myopathy can barely move, or they are children who die of the disease." Taube emphasizes, however, that LeMond's

disease. She is planning to study whether the onset of mitochondrial myopathy in athletes is related to particular sports and levels of exertion or to the individual's genetic makeup.

In general, it is difficult to explain why an athlete suddenly loses the ability to compete successfully or, worse, completely breaks down physically. For instance, a marathon runner complaining of fatigue might score far above normal on various exercise stress tests but actually be experiencing the first symptoms of a serious ailment. "It is possible," Taube says, "that there are

lation diseases that disrupt the cellular energy system in skeletal muscles, the liver, heart or brain and have been implicated in diabetes, aging and a number of neurodegenerative disorders. Basically, mitochondria fail to process enough oxygen to make sufficient ATP. Lacking aerobic capacity, some people with mitochondrial myopathy find climbing stairs or walking the length of a shopping mall fatiguing. Others might not even be aware that they have a problem until they begin to exert themselves. The condition is manifest clinically in exercise intolerance, muscle weakness, a type of cellular degeneration called ragged red fiber myopathy and increased numbers of abnormal mitochondria. A live muscle biopsy is required to observe the last two signs.

Shoffner claims that most cases are inherited (from the mother, the sole source of mitochondrial DNA). In some instances, environmental toxins such as high levels of lead or carbon monoxide from cigarette smoke or ingredients in certain medications—AZT, used for the treatment of AIDS is one—can induce myopathy by damaging the mitochon-

dria. Since 1988 over 30 genetic point deletions and even more mutations have been found in mitochondrial DNA.

By standard definition, it is inconceivable that an athlete engaged in endurance sports could suffer from mitochondrial myopathy. But Taube says that although LeMond shows none of the genetic or enzymatic deficiencies commonly associated with the disorder, "very subtle changes" in his muscle cells block their ability to use oxygen when he works hard. She speculates that the

lead pellets he still carries in his body from his hunting accident might relate to the onset of the condition.

While recognizing that something is adversely affecting LeMond's performance, certain specialists in the field have reacted with understandable caution to Taube's description of a new form of mitochondrial myopathy. Right or wrong, though, three other endurance athletes have contacted Taube since December saying they have the same symptoms. —Mark Derr

Astronomers in the Dark

There's more to empty space than meets the eye

Like viewers watching a Chinese shadow puppet play, some astronomers are finding themselves captivated not by light but by darkness. The luminous stars and galaxies sprinkled across the night sky are the obvious players in the cosmic drama. But it turns out that the vast stretches between galaxies have a story to tell as well.

Nadine Dinshaw of Steward Observa-

tory in Tucson, Ariz., and her co-workers have found that the thin gas permeating those voids is not a formless smear but rather is organized into huge clouds. Visible only in silhouette, these nonluminous clouds may be part of a ghostly network of sheets and filaments that fills the universe and traces the processes by which galaxies formed.

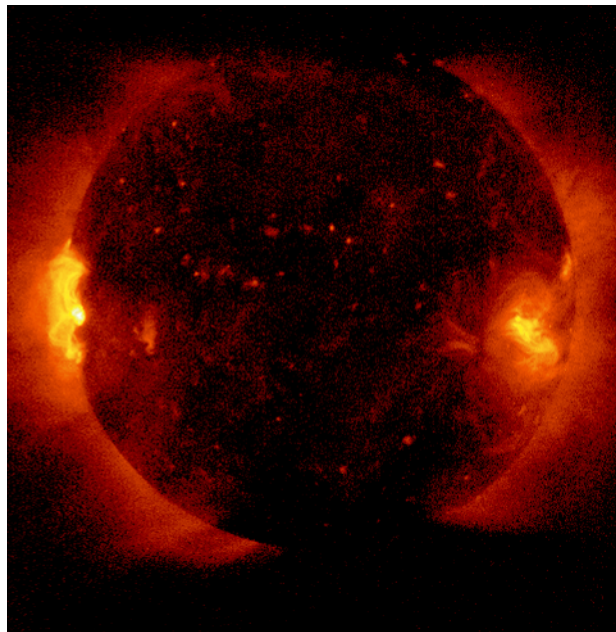
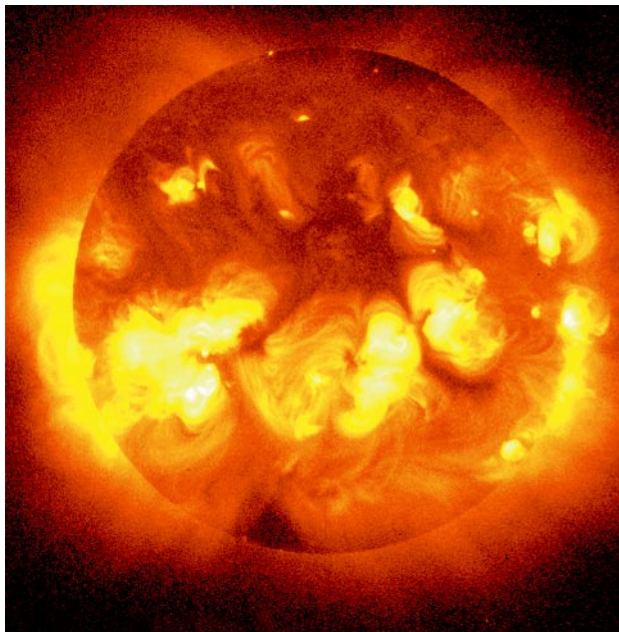
Astronomers deduced the presence of

Sun Spotting the Difference

X-ray images of the sun offer a new view of the nearby star and its cycles—one that differs markedly from the more familiar images made using visible light. Recent x-ray pictures from the Japanese *Yohkoh* spacecraft reveal a striking dimming of the sun's corona, its hot outer atmosphere, between 1992 (*left*) and this year (*right*). This change reflects the fluctuations of the 11-year solar cycle, as the star evolves from its period of maximum activity in

1991 to one of minimum action, which should occur by 1997. The cycle is characterized by shifts in the number of sunspots; these perturbations, which can be detected with standard imaging techniques, increase in number when solar activity is greatest. The x-ray data have shown that the corona changes as well: it is 30 times dimmer now than it was three years ago near the height of the solar cycle.

—Sasha Nemecek



YOHKOH TEAM

intergalactic clouds a quarter of a century ago, in the course of analyzing the radiation from faraway quasars (active galaxies whose brilliant central regions can easily be seen from across the universe). When that radiation is spread out into a spectrum, researchers found, certain characteristic wavelengths are absent. The pattern of the missing wavelengths revealed the culprit: clouds of hydrogen atoms scattered between the galaxies, which absorb some of the light from any object lying farther away from the earth. Each quasar shines through only a single part of a cloud, however, so the structure and extent of these

cosmic will-o'-the-wisps remained unknown. The absorbed radiation is mostly in the ultraviolet part of the spectrum, which does not penetrate the earth's atmosphere.

Dinshaw's team circumvented those problems by using the *Hubble Space Telescope* to look at not one but two light sources: two quasars that are separated in the sky by about one twentieth the angular diameter of the full moon (which, in this line of work, is a fairly wide separation). Because *Hubble* orbits above the earth's atmosphere, it can detect the ultraviolet rays invisible to ground-based observatories.

When Dinshaw and her collaborators looked at the results of their observation, she recalls, they were very surprised: they had found a monster. Previous research had hinted that intergalactic clouds were about 100,000 light-years across, or about the diameter of the Milky Way galaxy. Yet the spectra collected by *Hubble* indicated that a single cloud formation stretched across both quasars, giving it a minimum diameter of one million light-years. "Nobody expected the clouds to be so large," Dinshaw comments. The cloud also had remarkably little internal motion, suggesting it is a settled structure, not collapsing or flying apart.

"How can the cloud be so large and so quiescent?" Dinshaw wonders. Theorists are asking the same question. If the clouds are held together by the gravity from invisible "dark matter," they should collapse to a smaller size. If they are bound to a central galaxy, they should be moving faster. If they are held together by the pressure of the intergalactic medium, such large formations should have rapidly dissipated. "There's no well-developed theory to explain the kind of cloud we see," concludes Craig B. Foltz of Multiple Mirror Telescope Observatory, also in Tucson, who collaborated with Dinshaw.

Foltz suspects that he and his colleagues are watching many kinds of events happening at the same time, among them giant shock waves compressing gas clouds, hydrogen wisps collapsing around young galaxies and gas collecting around clumps of dark matter in regions where no galaxies exist. "We're seeing that the process of galaxy formation is very complex," and so the surrounding material takes on similarly complex properties, Foltz explains. "But that's okay with me—I don't mind complexity!"

Dinshaw's team has already examined two other quasar pairs and plans to look at a third. Such observations should help theorists fine-tune their cosmological models by revealing what is occurring in all parts of the universe, not just the well-lit corners. The early indications are that the dim regions are rich in unexplored details, "and it's only recently that we've been able to do observations like this," Dinshaw notes.

Such discoveries testify to the endurance of one of astronomy's most powerful but least glamorous tools—spectroscopy, which Foltz describes as "looking at the things that you don't see." A century ago the technique enabled astronomers to tell the poets what the stars are made of. Now it is broadening awareness of the hidden order of the universe.

—Corey S. Powell

A Ringside View of Stars

An unusual collision between galaxies has created a halo of stars ripe for a study—and the *Hubble Space Telescope* recorded it all. Although the crashing of galaxies and the subsequent formation of new stars is common, such impacts are difficult to decipher. "They often just leave behind a mixed-up mess," explains Kirk Borne of the Space Telescope Science Institute.

Yet when the Cartwheel galaxy (*left*)—located some 500 million light-years away—was jolted by one of two nearby galaxies (*right*), no mess ensued. "What makes this case unique is that the smaller galaxy basically hit the bull's-eye," Borne notes. "When the collision occurred, it sent a shock wave of energy outward, like a rock making a circular ripple when it's thrown into a pond." As the wave traveled, it compressed gas and matter in its wake, spewing billions of stars in an encircling band at the point of impact.

Because the ring around the galaxy is composed entirely of newly formed stars, cosmologists have an unprecedented chance to study a uniform population of massive stars—all born at about the same time under the same conditions. Identifying which of the two neighboring galaxies was responsible for the collision, however, will prove more tricky.

—Steven Vames



NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

Coming Out in the Sciences

I answered the telephone one night recently to hear the voice of an old friend from graduate school in physics. After chatting about research, he mentioned his reason for calling: he had come out as homosexual. During school, he had had an inkling of his orientation, he explained; only recently had he felt secure enough professionally to devote time to personal issues. "I thought you had been interested in me!" I blurted out, and he chuckled.

After I put the phone down, questions lingered. It had never occurred to me that a colleague was gay. Surely I should have known better. If, as surveys suggest, between 1 and 10 percent of the population is homosexual, how was it that I was aware, now, of just one physicist who was? I could think of great novelists who were gay but of only a sole scientist, Alan M. Turing. Were homosexuals underrepresented in physics, in science? Were they more closeted?

None of the researchers I called in subsequent months could identify a study of homosexual scientists. To conduct my own unscientific survey, I asked my friend to post questions on his electronic-mail network of some 50 gay astrophysicists. I polled two other bulletin boards—the 200-member Lesbians in Science, directed by physicist Elizabeth Zita, and the 800-member National Organization of Gay and Lesbian Scientists and Technical Professionals (NOGLSTP). I also interviewed other gay scientists. The responses were bewilderingly varied; only slowly did patterns emerge. "I know only two others in my field who

are openly gay," wrote Thomas Eads, a biophysicist at Purdue University. Yet everyone opined that the apparent rarity was misleading. "There just seem to be fewer of us who are out," remarked Ron Buckmire, a mathematician at Occidental College.

"Gays in science tend to be very discreet," agreed Mikel Susperregi, a cosmologist at the University of Oxford. "Those in the arts are much more conspicuous." Barbara Belmont of NOGLSTP noted that "science is a profession you can immerse yourself in, forsaking all social life. It provides a haven from society's pressures." Several researchers argued that science, in fact, attracts closeted gays.

Moreover, many of those who wanted to come out saw the sciences as unreceptive. "You look around, and you think, 'Gosh, this is a male-dominated field. I'd better keep my mouth shut,'" my friend said. "I find people in the hard sciences are much more intolerant," responded Julia George of the Jet Propulsion Laboratory in Pasadena, Calif. "Undoubtedly because there are many fewer women. Women, as a rule, seem to be more accepting."

That notion is supported by a 1982 poll of physicians conducted by William C. Matthews of the University of California at San Diego and others. Replies showed that women are better disposed toward homosexuals than are men. Indeed, NOGLSTP's "outlist" of 100 or so researchers (admittedly a small sample) showed few physicists and mathematicians, more astronomers and chemists and the

most biologists—in rough proportion to the number of women in each field.

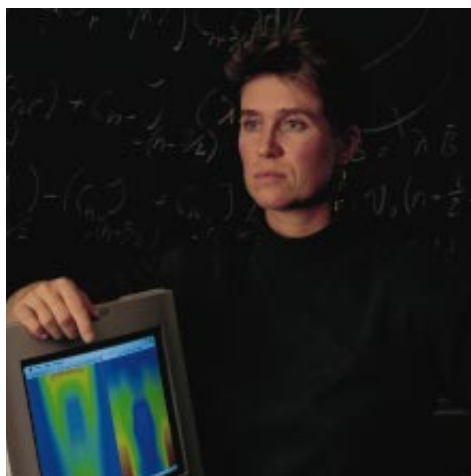
Gays who were out reported mixed experiences. "My colleagues' reception has been so chilling," Eads stated. One professor who is lesbian observed that male colleagues respect her talents more readily than other women's because "they don't think of me as a real woman, so it's not so weird that I can do science."

—Madhusree Mukerjee



Cosmologist Mikel Susperregi

DAVID LEVENSON/Black Star



Physicist Elizabeth Zita

JIM HEEMSTRA/SABA

I Get No Kick from CH₃CH₂OH

A new treatment for alcoholism receives FDA approval

THE more than one million alcoholics in this country who seek treatment every year have a new alternative. The U.S. Food and Drug Administration recently approved naltrexone as a medication for alcoholism; the drug is the first allowed for this purpose since the original medication for alcoholism, Antabuse, was introduced in 1948. The FDA's action reflects a growing belief that substance abuse has a strong biological component and is not

merely a character flaw. Yet experts emphasize that the drug will not obviate the need for conventional forms of therapy, such as counseling.

Two recent studies, carried out at the University of Pennsylvania and at Yale University, demonstrated the effectiveness of naltrexone—which was approved for the treatment of opiate addiction in 1984—in the rehabilitation of alcoholics. In one study of 70 patients, 23 percent of those given naltrexone relapsed dur-

ing the three-month-long study, whereas 54 percent of placebo recipients resumed drinking. Joseph R. Volpicelli, who led the team of researchers at the University of Pennsylvania, attributes the success of naltrexone to its ability to reduce the euphoria of alcohol and dampen the craving for another drink.

"Abusers respond to alcohol differently" than do other people, Volpicelli says. "As they drink, the motivation to have more increases." Volpicelli refers to this phenomenon as the "corn chip effect"—that is, you can't have just one. When alcoholics imbibe, their brains release unusually elevated levels of mole-

cules known as endorphins, which trigger a sensation of pleasure, enticing the drinker to indulge further.

Naltrexone mimics the shape of endorphins but not their action. By blocking the binding of these neurotransmitters to their receptors, it diminishes alcohol's pleasurable effects and the desire to drink more. "Naltrexone is a key that fits the lock of the endorphin receptor, but it doesn't open the door," Volpicelli explains. This outcome is significantly different from that of Antabuse, which can induce nausea and vomiting if patients drink during treatment.

Stephanie S. O'Malley, the lead researcher of the Yale study, speculates that naltrexone's gentler form of persuasion might be more helpful for some patients than is Antabuse. Because the drug seems to reduce the intense desire for alcohol, it actually helps patients to feel better and remain abstinent, thereby encouraging them to continue the regimen. In controlled studies of naltrexone, O'Malley notes that researchers are "finding good compliance" among

patients receiving the drug. She cautions against taking preliminary results too far: "In a broader population, compliance may not be as strong. Our patients tended to be fairly stable socially."

Furthermore, O'Malley points out that the recent studies have always combined medication with other types of treatment, such as counseling to develop coping skills. Indeed, she mentions a group of alcoholics who might particularly benefit from the integrated therapy. "Patients with high levels of craving, who are likely to do poorly otherwise, might do better with naltrexone as part of their treatment," O'Malley explains.

Because researchers are still in the early stages of studying the use of naltrexone for the treatment of alcoholism, Richard K. Fuller of the National Institute on Alcohol Abuse and Alcoholism expresses "cautious optimism" regarding this new application of the drug. Nevertheless, he says, "if further experience confirms its efficacy, the approval may be a milestone" in the treatment of substance abuse. Fuller adds that there

has been "an explosion of understanding in neuroscience" in the past few years, especially in the biology of addiction.

Indeed, the National Institute on Drug Abuse, which focuses primarily on abuse of illegal drugs, has screened close to two dozen different medications in clinical trials for their potential to treat crack-cocaine addiction alone, according to Charles Grudzinskas of the institute. Volpicelli hopes future approval of these types of drugs "will lead to a better understanding of substance abuse and lead to more funds being used in treatment programs."

In today's political climate, money is not likely to flood into such efforts. Unfortunately, "the funding of treatment has been a bipartisan failure," according to Herbert D. Kleber of Columbia University's Center on Addiction and Substance Abuse. The roughly \$700 million spent by the government every year for treatment and prevention research is dwarfed by the annual national costs of substance abuse, which estimates place at \$238 billion. —Sasha Nemecek

Fighting All the Time

Insights into HIV suggest ways to find better AIDS treatments

Until recently, many medical investigators believed the AIDS virus remains more or less dormant in asymptomatic persons, stepping up its assault on the immune system only in late stages of infection. Now two widely hailed reports published in January in *Nature* show that the virus wages a continuous, intense battle with the immune system from the start. The news is not all bad. Those results, and others from clinical investigations announced in February, point to novel strategies for combating the disease.

The studies that led to the new picture were conducted by two groups, one by George M. Shaw of the University of Alabama and the other by David D. Ho of the New York University School of Medicine. Both teams examined the effect on patients of experimental drugs that prevent the human immunodeficiency virus (HIV) from infecting cells in the body.

In all subjects the drugs caused HIV levels in the blood to drop rapidly by factors of 100 or more. At the same time, immune system cells of the CD4 type, which ordinarily slowly fall in number dur-

ing the course of HIV infection, surged in response. That observation suggests the virus has an immediate impact on CD4 cell depletion. It also indicates that "the immune system is quite resilient," notes Robert T. Schooley of the University of Colorado, one of the organizers of a conference on the AIDS virus held in late January.

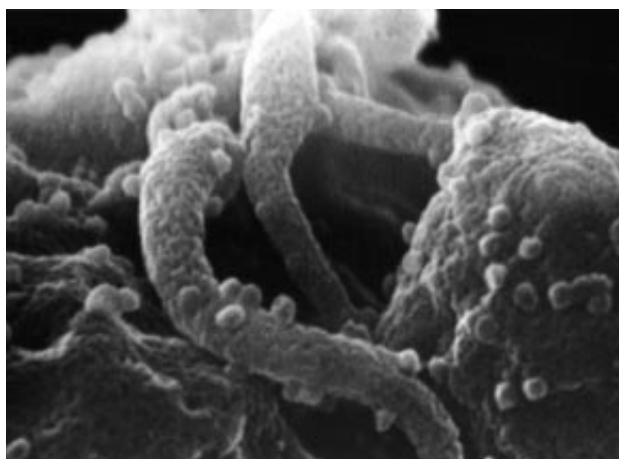
By studying how quickly virus levels declined when patients were treated, the investigators deduced that in untreated patients the rates of production and destruction of viruses must be extremely high. Individual particles survive only

about two days, but a billion or so new ones, about a third of the total in the body, are produced every day. Moreover, the virus destroys about a billion CD4 cells daily, which the body then attempts to replace.

Sadly, the effect of the drugs did not last. After two weeks, initially rare mutants started to multiply in the patients, like new heads sprouting from the Hydra of Greek mythology. The insights may nonetheless help researchers find drugs or drug combinations that have longer-lasting effects—and, with good fortune, clinical benefits.

The high turnover indicates that any effective antiviral drug should have an impact on virus levels within days, according to Ho's group. That understanding alone could save time that might have been wasted looking for long-term effects of ineffective drugs. The great rate of production of new viruses also suggests strongly that "the key is starting treatment early with potent drugs," says Martin Markowitz, one of Ho's team. The inevitable accumulation of mutations—a result of the rapid turnover—means that late treatment faces a monster with more heads.

Numerous anti-HIV drugs are now in early testing. Some block the viral enzyme known as reverse transcriptase; the widely prescribed



HIV PARTICLES (small spheres) bud from a lymphocyte.

ANDREW LEONARD/APL Microscopic

drug zidovudine (or AZT) works in the same way. Others attack different enzyme targets. Several of the experimental drugs appear to be substantially more effective than zidovudine against HIV, at least in the short term.

Interest at the January meeting focused on the combination of zidovudine and a drug called 3TC, developed by Glaxo. Patients given the combination still had 10-fold reductions in virus levels even after 24 weeks. The case is not proved, but such sustained lowered virus levels may well delay disease progression. "It was the most optimistic

AIDS meeting I'd been at for a long while," Markowitz notes.

The fervent hope is that by adding other drugs to the cocktail, it may be possible to hold down the level of viral replication for longer periods. "With combinations, you're seeing more pronounced immunologic effects," Schooley observes. "The earlier pessimism about antivirals was unfounded."

Nor is the idea of modifying the immune system's behavior with cytokines—naturally occurring molecules that modulate its activity—out of play. Research by Anthony S. Fauci, director

of the National Institute of Allergy and Infectious Diseases, and by others has shown that the cytokine interleukin-2, which stimulates immune responses, can increase CD4 cell levels in patients.

Schooley makes a plea for more clinical research on drug combinations that show promise in laboratory tests. "Of course, we'll need to prove that they have clinical benefits," he warns, "but I'm cautiously optimistic." Moreover, Schooley says, the finding that CD4 cells can rebound "offers hope to those so far along that they thought it didn't matter what we did." —*Tim Beardsley*

Down and Out in the Gulf of Mexico

Oil spewing offshore doesn't always signal pollution

People have long criticized oil companies for the accidental release of oil into the sea, but over the past decade a strange wrinkle has developed in the banner of environmental protection. Some scientists have begun to wonder whether drilling in the Gulf of Mexico could threaten marine life by, strangely enough, reducing oil leaks.

This curious twist results from two separate advances. Oil and gas developers have been moving drilling operations farther offshore, enticed by immense oil and gas fields previously thought to lie in water too deep to be economically tapped. At the same time, researchers have been uncovering unexpected richness and diversity in deep-dwelling marine life.

The work of both these groups has recently been focused on the Gulf floor's many natural oil seeps. Although sensitive instruments are sometimes needed to detect their subtle chemical traces, ambient effusions in the Gulf can be so great that they leave markings on the surface. Researchers have even discovered that some of these slicks are visible from space.

Most leaks, however, are hard to find. In the mid-1980s oil companies hired scientists from Texas A&M University to survey native petroleum seeps on the continental slope. The researchers stumbled on a surprisingly plentiful biota—fields of huge tube worms, giant mussels and deep-sea crabs. These groupings resemble the dense seafloor communities found several years earlier during exploration of the hydrothermal vents that form at sites of submarine volcanism where tectonic plates separate.

Like vent assemblages, petroleum-seep communities are isolated from sunlight, and their discoverers quickly realized that life there must rely on a food chain that begins with energy from the constant petrochemical bath. Bacteria living within the gill cells of the giant mussels, for example, provide for their hosts by metabolizing methane.

Soon after their discovery, protection of seep dwellers became a priority. "Initially we worried they would be so rare that we might have an endangered-species situation," remarks Ken Graham of the Minerals Management Service, the agency regulating oil and gas develop-

ment in the Gulf. Because such development involves erecting platforms, setting huge anchors and laying vast stretches of undersea pipeline, the Minerals Management Service issued special guidelines in 1989 for work that might impinge on chemosynthetic fauna.

But continuing study revealed that seep organisms are not rare after all. Chemosynthetic communities seem to develop wherever substantial leaks occur on the deep seafloor, and there is little fear now that offshore drilling could accidentally destroy some unique species. But could extraction of petroleum offshore reduce pressure to nearby seeps, thereby robbing these deep-sea enclaves of their basic foodstuffs?

Although somewhat far-fetched, the question is not completely academic.

Graham reports that Texas wildcatters often found that the extraction of oil on land could cause nearby seeps to dry; there might be similar results offshore. Ian R. MacDonald of Texas A&M believes degradation of a seep by exploitation of its source reservoir is indeed possible, but "it has to be considered on a case-by-case basis."

Conoco, for instance, has a platform sited close to one of the most extensive chemosynthetic communities found so far in the Gulf, a place known as Bush Hill. "They are producing oil a mile away from it," MacDonald remarks, "but they believe they are not tapping the reservoir" used by the organisms. Geologists at Shell, a company active in exploiting deepwater fields, also feel confident that the reservoirs being drilled in the Gulf are not directly connected to any seafloor seeps.

So it would seem, for the time being at least, that the giant mussels and giant tubeworms as well as the giant corporations can continue to feed together happily in the Gulf. —*David Schneider*



BOB THOMASON/Tony Stone Images

OIL RIG and deep-dwelling marine life both feed off petroleum reservoirs.

Catching That Wave

Atoms act like light—and get bent out of shape

Physicists may not be regarded as people who enjoy making waves, but some of them want to change that. By inducing atoms to act like light—that is, by turning particles into waves—and passing them through an interferometer, researchers at the Massachusetts Institute of Technology have explored how atoms act toward one another. They have, in a way, split an atom into two halves, using one half to probe a gaseous medium and the other as a point of reference. The results promise to lead to better measurements of atomic behavior and improved navigational instruments.

That particles behave like waves—and vice versa, for that matter—is a fundamental truth in physics. In fact, all objects, including beach balls and graduate students, have a wave nature. The undulations are not noticeable, because the wavelengths of such hefty objects are too small to be relevant. In contrast, light particles such as electrons and neutrons have proved more amenable to manipulation as light [see “The Duality in Matter and Light,” by Bert-

hold-Georg Englert, Marlan O. Scully and Herbert Walther; *SCIENTIFIC AMERICAN*, December 1994]. Indeed, the ready appearance of their waves has made these subatomic particles indispensable in microscopy.

The advent of skilled nanofabrication and other related technologies in recent years has enabled physicists to demonstrate the wave nature of a comparatively heavier object. “Atoms are more complex than neutrons or photons,” remarks Steven Chu of Stanford University. “They offer another way of making measurements that are potentially more useful” than those achieved with subatomic particles. Researchers have been especially interested in interferometers—devices that split a wave and then recombine it later so that the wave interferes with itself.

Now physicists have gone beyond merely showcasing the atom’s interfering wave properties. Jörg Schmiedmayer, Michael S. Chapman and David E. Pritchard, with their M.I.T. colleagues, are using the interferometer as a tool to probe how much an atom wave bends

when it passes through a particular medium. In other words, they are studying the force exerted between different types of atoms. “A whole bunch of people are starting to rethink these long-range forces,” Pritchard points out. “Theories are pretty unclear here.”

The workers passed sodium atoms through a finely etched diffraction grating. The grating generated sodium waves that were then each split into two parts. The parts moved down separate paths. One route traveled through a gas—such as xenon, helium or ammonia—while the other part of the wave bypassed the gas. When the sections recombined, the waves interfered with themselves. The interference produced a fringe pattern—an alternating series of bright and dark bands. The bright bands indicate that the waves combine there to achieve maximum intensity; the dark areas mark the places that the waves cancel one another. The pattern shifted according to the type of gas through which the wave had moved.

The researchers found that xenon, the heaviest gas studied, behaves most in the way theory suggests in terms of how its force acts over distance. The other gases, however, had variations that sometimes produced results sub-

SHOWERED BY PRAISE.



CHRYSLER CIRRUS LXI

© 1996 MOTOROLA, INC. MOTOROLA AND * ARE REGISTERED TRADEMARKS OF MOTOROLA, INC. ALL RIGHTS RESERVED. CIRRUS IS A REGISTERED TRADEMARK OF CHRYSLER CORP. MOTOR TEND CAR OF THE YEAR AWARDED FOR 1996 MODEL YEAR.

Copyright 1996 Scientific American, Inc.

stantially different from those expected. Resolving the discrepancies may be helpful to theoreticians trying to predict the properties of the long-sought Bose-Einstein condensate, a collection of atoms that supposedly acts as a single giant atom when cooled close enough to absolute zero.

Atom interferometers might have less obscure uses, too—namely, as gyroscopes. The interference pattern is extremely sensitive to inertial effects. “We expect our interferometer to be about as good as those on commercial jet airliners,” Pritchard says, “and we should be able to build one 500 times better than that.”

Pritchard’s method is not the only way to make atom waves interfere. Chu’s group, for instance, relies on laser bursts to excite the same atom to two different quantum states, which then interfere. With his apparatus, Chu has been refining measurements of various physical quantities, such as Planck’s constant. Other workers have been exploiting the principles behind interfering quantum states to control molecular dynamics as well atomic behavior [see “Laser Control of Chemical Reactions,” by Paul Brumer and Moshe Shapiro; SCIENTIFIC AMERICAN, March].

MICHAEL CRAWFORD



“AND BY APPROPRIATING QUOTIDIAN DOMESTIC ICONS, THESE CRO-MAGNON EXPRESSIONISTS ESTABLISHED AN APT ART-HISTORICAL REFERENCE POINT FROM WHICH TO RAISE GEO-POLITICAL, HEALTH CARE AND CONTEMPORARY SOCIAL ISSUES INDEPENDENT OF FORMALIST IDEOLOGY.”

Although there is no theoretical limit to the size of the object that can be made wavy, there are practical limitations. Pritchard did some calculations to see if he could get the wave nature of a

graduate student to interfere but found that the transit time through the system would be longer than the student’s lifetime. It would have been a rough way to earn a degree. —Philip Yam

POWERED BY MOTOROLA.

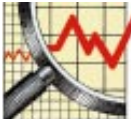


HERE’S A CLOUD WITH A SILVER LINING. CHRYSLER CIRRUS, MOTOR TREND’S 1995 “CAR OF THE YEAR,” DELIVERS SKY’S-THE-LIMIT FEATURES AT A DOWN-TO-EARTH PRICE, WITH PERFORMANCE HANDLED BY OVER 60 MOTOROLA

SEMICONDUCTORS GUIDING ALL ITS SYSTEMS. NO SURPRISE, REALLY—WE’RE THE AUTO INDUSTRY CHIP-OF-CHOICE WORLDWIDE. FROM AUTOMOTIVE ELECTRONICS TO WIRELESS COMMUNICATIONS, PRODUCTS POWERED BY MOTOROLA ARE FAST BECOMING A WAY OF LIFE.

Copyright 1995 Scientific American, Inc.

 **MOTOROLA**
Semiconductor Products Sector
FOR INFORMATION CALL 1-800-521-6274



The Insignificance of Statistical Significance

Economists, astrophysicists, sociologists, geologists as well as some medical researchers spend a lot of time looking at experiments that God has already performed. If God had not arranged things so that some stars were young and some were old, the astrophysicists would not know much about stellar evolution. Likewise, if God had not arranged things so that the minimum wage varied relative to the average wage for unskilled labor from decade to decade and state to state, economists would have a hard time convincing anyone that the minimum wage puts poor people out of work.

Economists and astrophysicists come to their knowledge by finding regularities of some kind in the world; one crucial part of their task is figuring out whether particular correlations point to an important law or to the fickle hand of coincidence. As a matter of fact, economists are having a hard time convincing people that the minimum wage

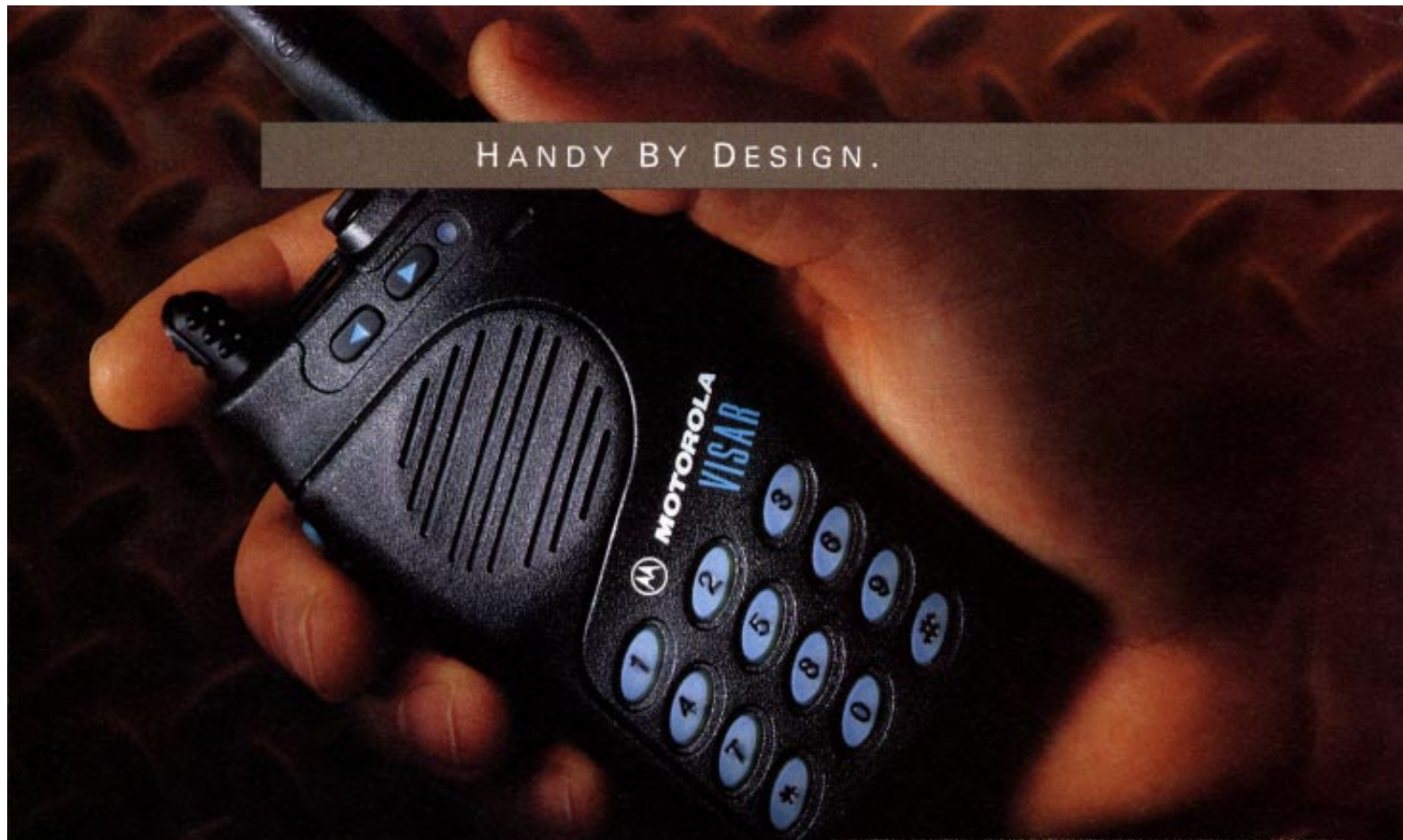
contributes to unemployment because recent studies show no “statistically significant” effect on jobs. When Congress takes the issue up later this year, the livelihoods of thousands of people could hang in the balance.

But just what does that phrase mean, and what does it have to do with the debate? Go back two centuries, to Pierre Simon, Marquis de Laplace, the first person to apply the notion of statistical significance to a serious scientific problem. In 1773 Laplace wanted to know where comets came from. He reasoned that if they originated inside the solar system, they would orbit in the same plane as the planets, whereas if they came from the far reaches of space, their paths would have no correlation with those of bodies circling the sun. Laplace checked the motions of the last 12 comets to be discovered and firmly rejected the hypothesis that comets came from inside the solar system. If the comets were of local origin, one

might by chance travel at some weird angle to the plane. But the odds of getting two anomalies would be lower, of three lower yet—the probability, so to speak, of rolling snake eyes three times in a row. This was a very smart idea.

In the succeeding two centuries, statisticians have refined Laplace’s simple notion into “statistical significance” and developed an arsenal of formulas for determining whether the phenomena that researchers observe are caused by sampling error (accidentally picking unrepresentative subjects) or “real” effects. The gold standard for most studies is the “95 percent confidence level,” which indicates odds of only one in 20 that a result arises from chance. Economists use it to test whether the minimum wage has a “significant” effect on employment. Medical researchers use it to decide whether half an aspirin a day keeps the cardiologist away.

Gradually, however, it has dawned on a few scientists that something is screwy. An obvious problem is that with so many people doing so many studies, some of them are going to run into that one-in-20 chance of believing in a mirage. The converse mistake is more subtle: scientists care about whether a result is statistically significant, but they should care much more about whether



HANDY BY DESIGN.



LOST JOBS? Some economists say low-paid workers will be fired if the minimum wage rises; others claim the evidence is statistically insignificant.

it is meaningful—whether it has, to use a technical term, oomph.

Sadly, many scientists have started thinking that statistical significance measures oomph. If an answer meets the 95 percent confidence criteria, it must be important; if it doesn't, it isn't.

The clearest refutation of this notion came in the study that established the lifesaving effect of aspirin in men who had already had a heart attack. Researchers stopped the experiment before their numbers reached “real” statistical significance because the effect of a mere half an aspirin a day was so

obvious that they considered it unethical to go on giving placebos to anyone.

Is this messy state of affairs Laplace's fault? He was right about comets because the relevant scale for measuring the oomph in orbits was obvious. Furthermore, a sample of a dozen could yield results that were scientifically as well as statistically significant. But the scale for measuring the effects of aspirin or of changes in the minimum wage is not so clear: you may get statistically impeccable answers that make little difference to anyone or “insignificant” ones that are absolutely crucial.

That conundrum is sharpest now in the debate among economists about the minimum wage. David Card and Alan B. Krueger of Princeton University have used tests of statistical significance to argue there is no convincing evidence that the minimum wage has a strong effect. Most other economists disagree, both because their theory tells them otherwise and because they think Card and Krueger are asking for too much certainty. But because both sides are muddled about the difference between oomph and statistical significance, the disagreement is not likely to get resolved in time to help Congress. Depending on what legislators decide, many poor people (not to mention teenagers on summer vacation) might lose their jobs. Ironically, even if they do, economic samplers may not be able to prove how many jobs were lost or that the minimum wage really had an effect.

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

POWERED BY MOTOROLA.



MOTOROLA'S LATEST BRAINCHILD IS SO SMALL IT'S BIG NEWS. VISAR PUTS FULL-SIZE TWO-WAY RADIO PERFORMANCE LITERALLY IN THE PALM OF YOUR HAND. THIS SMALL WONDER IS POWERED BY ANOTHER

MINIATURE MOTOROLA MARVEL—A 68HC11 MICROCONTROLLER, MAKING ALL SYSTEMS GO, FROM WIRELESS COMMUNICATIONS TO AUTOMOTIVE ELECTRONICS, PRODUCTS POWERED BY MOTOROLA ARE FAST BECOMING A WAY OF LIFE.



MOTOROLA
Semiconductor Products Sector

FOR INFORMATION CALL 1-800-521-6274



Coming in for a Landing

Should satellites or microwaves direct airplanes in bad weather?

When the London fog rolls in, pilots flying into Heathrow Airport turn on their instrument landing system receivers to pick up the radio beacons that will guide them safely through the soup and onto the runway. But those beams, steady for 50 years so far, may soon begin to rock and roll. In January 1998 European FM radio broadcasters will turn up the volume on their transmitters, some of which are located uncomfortably close—both on

Federal Aviation Administration, will try to persuade other nations to abandon MLS development, as the U.S. did last summer. “We would like to eliminate the mandatory transition to MLS on international runways,” Morgan says.

Heads turned from MLS when the U.S. Department of Defense encircled the earth with the Global Positioning System (GPS), 24 satellites that broadcast their location and a timing signal. With any four in sight, a GPS receiver can

proved that point. United Parcel Service autopiloted a Boeing 757-200 through 50 landings using a ground station that measured errors in the satellite signal and beamed the corrections to the incoming aircraft. Stanford University research associate Clark E. Cohen hit on an even better approach while working on a relativity experiment. Cohen’s solution—to boost precision, add precise satellites—is more feasible than it sounds, because the additional satellites can be on the ground.

The FAA tested Cohen’s idea on a United Airlines 737-300. Two transmitters, each the size of a credit card and powered by a nine-volt battery, were placed about four kilometers from the end of the runway. As the test plane passed over the pseudosatellites, a small antenna on its belly picked up weak but accurate signals and used them to refine to a few centimeters location data coming from four real GPS satellites. Over several days the jet’s autopilot made 110 successful touch-and-go landings.

The real test of Stanford’s system came on flight number 37, when data uploaded to one GPS satellite caused it to blink out for two seconds just as the 737 was descending. How quickly and reliably the system warns the pilot not to trust the instruments in such situations—what regulators call the system’s “integrity”—is at the crux of the debate over the GPS as a possible replacement for the instrument landing system. Although the U.S. has assured other nations that the GPS will be free and reliable to all users for at least 10 years, “we have to understand that in times of war, things change,” says Joseph F. Dorfler, the FAA’s satellite program manager. Software glitches and hardware failures could also shut satellites down without warning.

To reinforce GPS integrity in the U.S., the FAA hopes to build a \$500-million network of 24 ground stations that will relay GPS corrections to pilots and alert them the moment something goes wrong. Other countries are looking at Glonass, a Russian counterpart to GPS, and Inmarsat, which maintains a group of marine navigation satellites, as possible backups. But most observers doubt that these systems will achieve the integrity needed for low-visibility landings by the 1998 deadline.

Stanford’s experimental system may offer help here. A computer on board the aircraft compares the six corrected



RAINER GROSSKOPF/Tony Stone Images

LANDING SYSTEM that guides airplanes down through foul weather must be replaced soon. The U.S. wants the world to rely on its military satellites, but other nations raise safety issues.

the ground and in the electromagnetic spectrum—to landing systems at major airports.

The world’s aviation agencies planned for this contingency 20 years ago. The instrument landing system, everyone agreed, would be replaced in the mid-1990s with a microwave landing system (better known as MLS). In addition to dodging interference by moving to a less crowded part of the spectrum, MLS would allow pilots to make steeper and curved descents, cutting noise and boosting airport capacity.

But in recent years that agreement has crumbled, and the future of poor-weather landing is up in the air. At a special international meeting this month to settle the issue, Ronald E. Morgan, director of system architecture for the

work out its position (plus or minus 100 meters) anywhere on or above the planet. U.S. airlines were immediately starstruck by the technology, which could allow their fleets to break out of intercity airways and fly more direct routes to save time and fuel. Last December the FAA approved their requests to use GPS to navigate over the oceans.

Low-visibility descents, of course, are a different matter. A 100-meter error could land a jumbo jet in a parking lot. If the GPS is to replace the instrument landing system, its errors must be held to less than a meter and safety-threatening system failures to one in a billion. “We’ve solved the accuracy problem,” announces Robert D. Till, an FAA test program manager.

Two demonstrations last October

signals it receives from four satellites and two beacons. If any one is too far out of line, it disengages the autopilot and sounds the alarm. On flight 37, Cohen says, all that happened within a quarter of a second.

The FAA plans further integrity tests this spring to bolster its case at the upcoming international meeting in Montreal. But U.S. airlines, which are loath to buy expensive MLS equipment, nonetheless "anticipate that we will have to have some limited MLS capability," says J. Roger Fleming, a senior official at the Air Transport Association. Otherwise, when Heathrow fogs in, Americans could be shut out. —W. Wayt Gibbs

A Widget's Best Friend

Diamonds may bring a new facet to motors and sensors

Tiny electric engines built out of silicon chips have been making wonderful illustrations for gee-whiz technology stories since the early 1980s. They have not been good for much else. Friction, fragility and corrosion are among the ailments silicon suffers. The element's major advantage is that it is easy to form into complex microscopic shapes because integrated-circuit engineers have spent billions of dollars developing tools to grow it, etch it, dope it and alloy it with other materials. Micromachinists have just tagged along for the ride.

Now a team of microfabricators appears to be taking the wheel. Researchers at Oak Ridge National Laboratories report in *Applied Physics Letters* that they have made minute mechanical devices out of diamond. Diamond has a much lower coefficient of friction than silicon; it is also stronger and stiffer and more resistant to chemical attack.

John D. Hunn and his colleagues have developed a technique for etching patterns in diamond films. First, they bombard a diamond surface with high-energy oxygen ions to produce a layer of graphite inside the crystal (the ions do not produce defects at the surface, where they are traveling quickly, but rather cause disruption when they come to a stop). Then the researchers grow a layer of diamond film on top of the prepared diamond. Third, an ultraviolet laser cuts a trench through the film down to the graphite layer, outlining the shape of the parts to be made. Finally, the team places the diamond in a furnace, where high-temperature oxygen burns away the graphite, freeing each piece from the substrate.

The process is yielding gears as small as four tenths of a millimeter across. Such a gear could easily serve as a rotor in a small electric motor. The novel diamond part could be dropped into one of the silicon frames that has already been built, Hunn says: "You could do it tomorrow."

Flashy though such a motor might be, the real payoff of the technology will be in diamond sensors, Hunn comments. A diamond-film membrane etched thin by a technique that uses implanted ions could signal pressure

changes inside a corrosive liquid that would destroy a silicon sensor.

But these microsensors will not be available tomorrow or even next year. The engineers must figure out how to integrate electronic manufacturing activities with their machining technology. And to date, Hunn explains, neither diamond-film researchers nor micromachinists have followed the team's lead. Nevertheless, the Oak Ridge project has already given the "jewelry phase" of high-tech research and development a new meaning. —Paul Wallich

KUWAIT PRIZE 1995

Invitation to Nominations

The Kuwait Prize was institutionalized to recognize distinguished accomplishments in the arts, humanities and sciences.

The Prizes are awarded annually in the following categories:

- A. Basic Sciences
- B. Applied Sciences
- C. Economics and Social Sciences
- D. Arts and Letters
- E. Arabic and Islamic Scientific Heritage

The Prizes for 1995 will be awarded in the following fields:

- | | |
|--|---|
| A. Basic Sciences: | Mathematical Statistics |
| B. Applied Sciences: | Allergy |
| C. Economics and Social Sciences: | Energy and Development in the Gulf Cooperation Council Countries |
| D. Arts and Letters: | Theatre and the Arab Social Issues |
| E. Arabic and Islamic Scientific Heritage: | Chemistry and Pharmacology |

Foreground and Conditions of the Prize:

1. Two prizes are awarded in each category:
 - * A Prize to recognize the distinguished scientific research of a Kuwaiti, and,
 - * A Prize to recognize the distinguished scientific research of an Arab citizen.
2. The candidate should not have been awarded a Prize for the submitted work by any other institution.
3. Nominations for these Prizes are accepted from individuals, academic and scientific centres, learned societies, past recipients of the Prize, and peers of the nominees. No nominations are accepted from political entities.
4. The scientific research submitted must have been published during the last ten years.
5. Each Prize consists of a cash sum of K.D. 30,000/- (U.S. \$100,000/- approx.), a Gold medal, a KFAS Shield and a Certificate of Recognition.
6. Nominators must clearly indicate the distinguished work that qualifies their candidate for consideration.
7. The results of KFAS decisions regarding selection of winners are final.
8. The papers submitted for nominations will not be returned regardless of the outcome of the decision.
9. Each winner is expected to deliver a lecture concerning the contribution for which he was awarded the Prize.

Inquiries concerning the Kuwait Prize and nominations including complete curriculum vitae and updated lists of publications by the candidate with four copies of each of the published papers should be received before 31/10/1995 and addressed to:

The Director General

The Kuwait Foundation for the Advancement of Sciences

P.O. Box: 25263, Safat-13113, Kuwait

Tel: +965 2429780 Fax: +965 2403891 / Telex: 44160 KEFAS

Simply, the Best

Energy-efficient cookstove technology makes a comeback

E. F. Schumacher's *Small Is Beautiful*, the bible of the so-called appropriate technology movement, first came into print more than 20 years ago. By most accounts, the small-scale technology movement that Schumacher championed initially failed to achieve its promise. The developing world is filled with rundown wind pumps and solar panels that were never properly connected to a generator. But although the original implementation may have been flawed, appropriate technology has staged a revival.

Proof of change can be seen on the windmill-dotted Namibian coast or in the Kenyan countryside, where tens of thousands of photovoltaic systems help to cope with an unreliable or nonexistent electric power grid. In sheer numbers, however, the biggest gains have come inside the home, where more than a third of the world's population—some two billion people—cook over smoky fires. In excess of 120 million households in China, eight million in India and 700,000 in East Africa have begun to adopt stoves that more efficiently burn wood, charcoal, coal or livestock dung. Worldwide several hundred improved cookstove programs in more than 50 countries got their start during the 1980s.

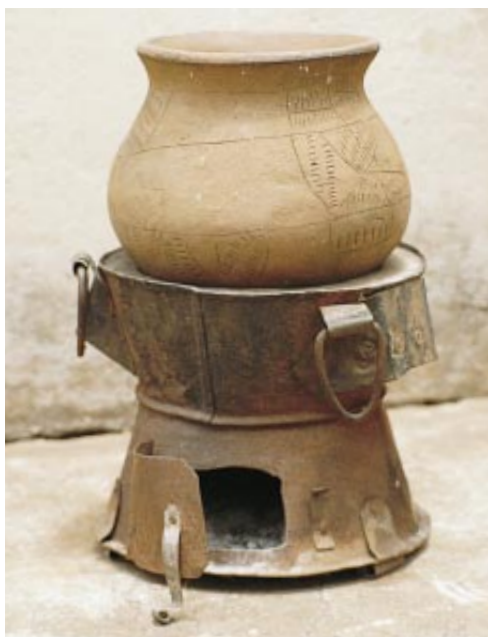
Slight improvements in such stove technology can have long-ranging impacts on quality of life and on the environment. Poor city dwellers in developing countries may spend up to a quarter of their \$300 to \$400 annual earnings on wood or charcoal. A stove that costs less than \$5 but uses 30 percent less wood can pay for itself in a matter of months. Rural villagers who adopt enhanced stoves can cut down on time spent (up to one full day a week) cutting and carrying wood.

Breathing also comes easier. Traditional stoves produce a noxious mix of hydrocarbons, carbon monoxide and particulates that help to make acute respiratory diseases, such as pneumonia, the leading health hazard in the developing world. Such illnesses cause an estimated 4.3 million deaths every year. Pollution levels from indoor cooking can make the air in some rural homes rival or exceed that of the dirtiest of industrial cities.

Over the past two decades physicists, engineers and development outreach groups have labored to create stove designs that achieve efficiencies of be-

tween 35 to 50 percent, several times that of a cooking fire set between three stones on which a pot is placed. The fires of a traditional cookstove—a metal canlike enclosure stoked with wood is one example—are only marginally better than such three-stone fires.

Following Schumacher to the letter, stove designers have emphasized use of local materials. One method of bettering efficiency (defined as the maximum amount of heat delivered to a



IMPROVED STOVE with clay pot, in Kenya

pot) comes from making ceramic stove liners out of sand and clay. The sheaths promote high cooking temperatures.

An earlier generation of appropriate technology proselytizers made a number of blunders, both technical and social. They would, for example, fail to match the circumference of the stove

top to the dimensions of common cooking utensils. Or they would train only men in how to care for the stoves. They were also smug about how much they needed to know about the physics of seemingly simple cooking implements. "Getting the thermodynamics right and understanding the heat-transfer characteristics are not things you do as a weekend project," says Daniel M. Kammen, a physicist by training and a professor in the Woodrow Wilson School of Public and International Affairs at Princeton University.

Selling improved cookstoves in the developing world requires as much market research as does plying a new dishwashing detergent to consumers in suburban Chicago. The cookstove program in China, for one, gained acceptance only after design alterations gave the stoves a modern look. Careful administration and pricing are also essential. The Chinese government removed many of the bureaucratic constraints to setting up stove programs. It provided few subsidies, yet most rural households still decided to purchase the stoves.

A World Bank report published last year found that India, in contrast to China, established a cumbersome administrative structure for its national program, relied on heavy subsidies and failed to concentrate on households with the greatest need. It is estimated that half of the more than eight million stoves purveyed through the Indian program lie unused.

Schumacher made as many mistakes as his followers did. The visionary recommended that large-scale enterprises should end up being run by the public sector, an idea that appears somewhat dated in light of current events. But his notion that a revolution in simple technology can transform individual lives may have gained credence even in this post-Marxist era. —Gary Stix

Genes in the Not So Public Domain

Human DNA databanks open for business—and vie for users

Two commercially funded databases of DNA sequences that will identify most human genes are unlocking their computer files to researchers worldwide. This boon for biomedical work could pave the way for powerful new pharmaceuticals in the next century. Yet the projects, one bankrolled by SmithKline Beecham and the other by Merck & Co., have been the focus of rancorous disputes.

Both databases employ so-called expressed sequence tags (ESTs), short genetic sequences that can help workers find entire genes. But the similarities between the databases end there. Merck says all the information it produces will be put immediately into the public domain, with no restrictions on access. The first sequences generated by the effort, which is being conducted at Washington University, were deposited

DANIEL M. KAMMEN Princeton University

in a public-access database in February. It will be 18 months before the database contains enough ESTs to tag the majority of human genes.

In contrast, the SmithKline Beecham project, which began two years ago at the nonprofit Institute for Genomic Research (TIGR) in Gaithersburg, Md., has already tagged about half the estimated 70,000 human genes. TIGR, which is also supported by Human Genome Sciences in Rockville, Md., is now testing its database at a dozen or so research institutions. Founder of TIGR and the pioneer of the EST technique, J. Craig Venter, along with his team, has submitted a long paper describing and categorizing TIGR's sequences to a scientific journal (rumored to be *Nature*). When the paper is published, the database will be made available to academic scientists who are not employees of for-profit corporations, Venter says.

The information, however, has strings attached. The U.S. Patent and Trademark Office has ruled that ESTs cannot be patented. As a result, TIGR's sponsors have insisted that the most valuable sequences be provided only under

certain conditions. The main stipulation is that if a sequence is used to develop a profitable drug, the researcher's institution must agree to negotiate with TIGR on a share of the royalties. Moreover, although TIGR has dropped a proposal to limit the number of queries that investigators can send to TIGR's computers, the database as a whole remains under wraps. And a few genes that TIGR has identified have been withheld.

Investigators accustomed to working with public information are balking at using TIGR's vast database in an international effort to map the sequences. "The problem is that the sequences are secret, and we have not been able to work out how to put them on a map open to everyone," says Francis S. Collins, director of the National Center for Human Genome Research at the National Institutes of Health. Recently, however, Venter has quelled criticism of TIGR by entering into a collaboration to create a public map of ESTs.

Merck's public-domain database was announced in response to TIGR's decision last fall to put restrictions on its information. The clash has led to some

bad blood. Venter disparages the Merck effort as an "attempt to hurt us" and questions the extent of Merck's financial commitment—a topic the drug company refuses to discuss. For his part, Keith O. Elliston, an associate director for bioinformatics at Merck, queries whether TIGR's methods generate accurate results. The Merck initiative diverges enough in its technical details that it will not duplicate the TIGR work, according to Elliston. DNA fragments will be characterized at both ends—unlike sequences in TIGR's database. The extra information will, Elliston says, make it far easier to construct highly accurate gene maps. Moreover, the collection of sequences should contain fewer duplicate entries than TIGR's, he adds. Venter counters that only TIGR's database identifies all its ESTs.

Nevertheless, "the Merck plan is a good one," comments Collins of the NIH. Despite the corporate public-relations battle, the important business of tracking down human genes and discovering exactly what they do is moving full-steam ahead—with the help of drug company money. —Tim Beardsley

Voting for a Cure

The world's toughest question blared in 117-point type across page B5 of the *New York Times* on January 19. Readers moving between a profile of Senator Alfonse D'Amato of New York and the latest on the O. J. Simpson murder trial had their first—and probably their last—chance to cast a ballot for a remedy for cancer. For that day, a question that Ph.D.'s have pondered for decades was reduced to a matter of near-religious faith: "I believe in tumor antigens," read a checkoff box at the bottom of the ad. "I believe in cancer vaccines," read another.

Madison Avenue advertising executive Steve Fenton created the full-page advertisement to raise the profile of his pro bono client, the Cancer Research Institute (CRI). This \$5.4-million-a-year charity has, since 1953, raised money to find ways to marshal the human immune system's fight against malignant cells. Instead of the usual low-key, anonymous public-service approach, Fenton decided to combine the highbrow with the mundane. He placed words such as "tumor antigens" and "cytokines" in headline print. And in a direct steal from A&P store specials on asparagus

spears, he put these titles, along with a brief explanatory text, in their own clip-out coupons that could be mailed in with a donation.

Asking people to vote on a cure for cancer arouses some skepticism among serious scientists, especially given the decidedly mixed history of the war against the disease. "It has the inference that the cure for cancer is around the corner and you can select from a menu," comments Samuel S. Epstein, a professor at the University of Illinois at Chicago.

Still, for the CRI, tumor antigens, cancer vaccines and cytokines have proved a big hit. "We're getting more response to this ad than any we've ever run," says Jill O'Donnell-Tormey, one of the organization's executive directors, of the not overwhelming 40 replies received. Most respondents merely filled in a generic box that leaves it up to the CRI's scientific advisory council to decide where the money goes. Some did cast a ballot. Of the six immunotherapies listed, cancer vaccines have taken the lead in the polls.

Fenton, who is an executive with D'Arcy Masius Benton & Bowles, says he was merely trying to provoke readers into thinking, "'Gee, this sounds interesting, and the money isn't going into a black hole.'" But the author of the slogan "America's Getting into Training," for an Amtrak ad, may still have his work cut out for him before tumor antigens and cytokines become as familiar as Metroliners and asparagus spears. —Gary Stix

Where Do You Think The Cure For Cancer Will Come From?

EXCERPTS from a Cancer Research Institute advertisement

☐ I believe in cancer vaccines

Just like with other diseases, a vaccine could stimulate an immune response that protects against cancer recurring—or even prevents it completely.

Experimental cancer vaccines are already being tested in patients. Your help will push these studies along.

Enclosed is my tax deductible gift of \$_____ to Cancer Research Institute, 681 Fifth Ave., NY, NY 10022. 1-800-99-CANCER

Name _____

Address _____

City _____

State _____ Zip _____

The Fishy Business of Waste

Development near Calcutta may thwart age-old recycling

Fish love sewage, and the inhabitants of Calcutta love fish. These city dwellers also produce sewage: 700 million liters of it a day. By growing fish on this waste, fisherfolk to the east of the Indian metropolis discovered how to link these needs at the turn of the century. But this natural recycling system—the largest of its kind and possibly the oldest—is threatened

by a state government that sees profitable real estate in the expanse of shallow waters.

The fisheries flourish on elaborate folk technology. Diverted into a series of ponds, the sewage sheds its putrescent solids. Air-breathing fish and water plants survive in pools nearest the sewage inlet. Further downstream, algae bloom on nutrients in the clear wa-

ter. Tilapia and carp, in turn, thrive on the algae. More than 20 tons of these fish feed the Calcutta market every day. (Bioassays of these fish show that their levels of heavy metals and coliform bacteria are low.) The algae also provide much needed oxygen to the heavily polluted city.

Further, sludge from the distillation ponds fertilizes extensive “garbage gardens.” Calcutta’s solid waste, stripped of its paper, plastics and metals by some 25,000 ragpickers who roam the city’s noxious dumps, is ultimately almost entirely organic. Irrigated by treated sewage and then composted, the waste becomes soil on which farmers grow a variety of vegetables. Their harvest supplies the city with 150 tons of produce a day.

This recycling expanse, however, has been shrinking as Calcutta expands eastward. In the 1960s northeastern water bodies were filled in to create a township. The Dutch consultants who drew up the plans apparently overlooked what British engineers had noted in the 1880s: the land naturally slopes to the east. Once these pond beds were gone, the eastbound sewage canals stopped flowing during the rains of the monsoon season, and now they regularly flood the city.

In addition, thugs—or antisocials, as Indians call them—associated with the communist state government are reported to have seized many fisheries on the pretext of distributing land to impoverished laborers. Instead the parcels have been sold to various real-estate developers.

Recently citizens groups sued to stop the government from building a shiny new World Trade Center, funded by a multinational group, on such “reclaimed” terrain. The case is still being disputed. Meanwhile politicians are touting a “Megacity 2011” expansion project, the logistics of which, observes Asish K. Ghosh of the Zoological Survey of India, have not been divulged to environmentalists. One thing is clear, though: the plan involves grand designs on the fisheries.

At the same time, these water bodies are becoming ever more essential. In February the Indian Supreme Court threatened to close 30 factories that were discharging pollutants into the Hoogly River, west of Calcutta. Dhrubajyoti Ghosh of the Institute for Wetlands Management has suggested that these effluents could be treated by expanding the fisheries.

But the government is clearly not listening. And Calcutta’s leap into the 21st century could well become a plunge into miasmic waters. —*Madhusree Mukerjee*

THAT DOLPHIN OFF YOUR BOW COULD BE ONE YOU HELPED SAVE.

Rescuing stranded dolphins--and learning how to save them from disease and pollution -- is a major effort at Mote Marine Laboratory. It's important work, but it doesn't come cheap. We need your support to do more. Please donate to help us continue this important research. Because the dolphin dancing off your bow could be one you personally helped save.



MOTE MARINE LABORATORY

1600 Thompson Parkway, Sarasota, Florida 34236

1-800-691-MOTE

Changing the Image

Looking to MRI for diagnosing breast cancer

Few would quibble with the value of screening for breast cancer, but even at its current level of maturity and appreciation, the art of x-ray mammography still shows a glaring flaw. It often fails to detect malignancies. In women younger than 50 years, for example, it misses existing cancers nearly half the time. So some medical researchers are cautiously contemplating screening for breast tumors with magnetic resonance imaging (MRI), a technique that probes the human body with a combination of magnetic fields and radio waves.

Certain women stand to benefit most from MRI: their breasts contain more fibrous or glandular material and less fat than is typical, and this constitution makes their breast tissues prone to scatter or severely attenuate x-rays. For these women (perhaps 40 percent of the female population), mammography renders images that are little more than a cloudy blur that prevents physicians from readily seeing the subtle architectural distortions indicative of cancer. Although some breast cancers signal their presence through the fog by creating microcalcifications that, like bone, leave bright spots on x-ray negatives, many malignancies do not contain these tiny radiographic beacons. Hence, for a woman with "dense" breasts, a doctor's normally assuring statement that no evidence of cancer appears on her mammogram has little weight.

MRI could, at least in principle, provide the ideal diagnostic alternative because it relies on a completely different and perhaps more promising combination of physics and physiology. Practitioners generally agree that MRI is superior to x-rays for imaging soft tissue,

but its chief attraction for use in the breast is that it can be applied in conjunction with tracers. Images taken before and after one of these magnetic compounds is infused into a patient's bloodstream can delineate tumors amazingly well because the compound is preferentially absorbed. And, unlike contrast agents employed for x-rays, those used for MRI almost never give rise to dangerous allergic reactions.

Steven E. Harms and his colleagues at Baylor University Medical Center have been carefully examining the effectiveness of MRI for women with breast cancer. In one of their recent studies the patients with cancer later underwent mastectomies; thus, the exact state of the excised breast tissue could be determined. Harms and his co-workers found that mammography, even combined with ultrasound, depicted no cancers that were not also detected with MRI scanning. But the scans showed many cancers not found by mammography: more than a third of the breasts with negative mammograms showed tumors with MRI.

Women with radiographically dense breasts should find such results particularly tantalizing. And should they eventually benefit from it, they might also find MRI screening relatively comfortable. The procedure only requires them to lie face downward, with their breasts held still below them, not painfully compressed as in mammographic examinations. The patient's stay within the imaging machine is also guaranteed to be quite short, because the contrast agent marks tumors for only a few minutes after injection.

No full-scale clinical studies, however, have yet established the seemingly reasonable proposition that MRI screening would save some women's lives. "Right now we don't know," remarks Jeffrey C. Weinreb of New York University Medical Center. Screening breasts with MRI "could be so great and revolutionary that everybody's getting it done," he notes. "Then again it could flop. We're just scratching the surface and learning how to do it."

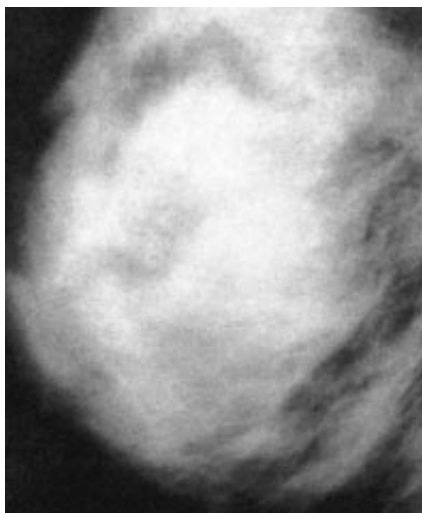
Although a broad, long-term study comparing MRI with mammographic screening is needed, it would take many years to complete and would be immensely expensive. A less costly and

more effective approach might be to focus such studies on women who are genetically predisposed to the development of breast cancer. No formal analysis of this kind has yet been started, but according to Weinreb, these efforts are probably only a few years away.

Even if MRI screening proves clinically valuable, the cost could prevent it from finding widespread use. Weinreb is, nonetheless, optimistic: "If it does work, breast MRI can probably be done cheaply." He estimates that the cost for a screening exam could eventually decline to within a factor of two of x-ray mammography. Indeed, he states, "if I were doing 40 [breast MRI scans] a day, I could do it [for that amount] now."

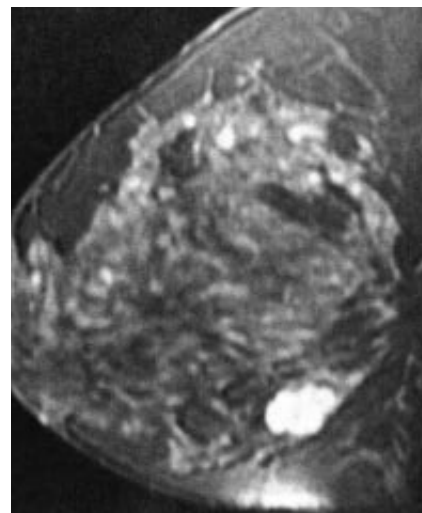
For some people, price is less of a concern than the threat of cancer, and many women are understandably reluctant to wait for results of long-term studies before taking advantage of this technique. Many manufacturers seem to be anticipating such demand and are producing specially designed pickup coils for MRI scans of the breast. One new company, Advanced Mammography Systems in Wilmington, Mass., has developed a relatively low-cost MRI machine that is solely for breast exams.

Weinreb fears that the proliferation of dedicated hardware means that "medical entrepreneurs" may not wait for the completion of the proper trials before marketing breast MRI services to an increasingly worried public. Noting that certain suburbs of New York City have some of the highest rates of breast cancer in the country, he speculates that before long "somebody is going to open up one of these facilities on Long Island and will clean up." So, as promising as this technique now appears, it stands as anyone's guess whether the future of breast MRI screening will see women's cancers, or just their fears, exposed. —David Schneider



STEVEN E. HARMS courtesy of Magnetic Resonance Quarterly

IMAGES OF CANCER within a "dense" breast are foggy in an x-ray mammogram (left) but clear with MRI (right).



STEVEN E. HARMS courtesy of Magnetic Resonance Quarterly



PROFILE: ARATI PRABHAKAR

Engineering the Future

Gusts of sleet whip through the air as I drive past unattended guard stations and by the landscaped fields that form a miniature greenbelt around the National Institute of Standards and Technology in Gaithersburg, Md. The first big snow of 1995 landed on the nation's capital just 36 hours ago, throwing this Monday morning rush hour into confusion. As I slog through the parking lot toward the ivory-colored administrative tower rising from a complex of low, blocky labs, I imagine the institute's director staring out the top-floor window, morosely pondering the weather as a metaphor for the Clinton administration's political fortunes.

Such a scene may have fit many past NIST directors, most of them middle-aged bench scientists who grew comfortable with their limited, inconspicuous mission: to help industrial competitors agree on what constitutes a volt, a meter, a second. But from the moment I walk into the director's modest office, it is apparent that its newest inhabitant, Arati Prabhakar (pronounced AR-ah-thee Pra-BOCK-er), does not fit the archetype for this agency—and that the agency is changing to adapt.

It is not just the outward appearances that are different. To be sure, those are striking: Prabhakar is the first female, the first foreign-born citizen and the youngest person to direct the institute. Gregarious and amiable, she seems remarkably at ease for someone who will defend her programs before the Senate science committee the next day. ("It's not so difficult after the first few times," she confides.) And the snow is hardly making her morose. "My boyfriend and I had a snowball fight Saturday night," she says with a chuckle in her deep voice. "It was great."

The most germane characteristic that distinguishes Prabhakar from her predecessors, it turns out, lies in another first. "I am the first director of NIST or NBS [the National Bureau of Standards, its former name] who was trained as

an engineer, rather than as a scientist," she explains. "That is both symbolic and meaningful, because it's representative of what NIST is trying to do today"—or, to be more precise, of what the White House is trying to do.

"It was very obvious when this administration came in that NIST was part of its agenda," Prabhakar notes. "That's why I'm here." President Bill Clinton has consistently argued that because of a flaw in the free-market system, U.S. companies invest so little in long-term re-



ARATI PRABHAKAR is a different kind of director for the National Institute of Standards and Technology.

search that they risk losing their technological edge to government-subsidized competitors overseas. The public sector, he has asserted, can compensate for this private-sector failing by putting up seed money in certain high-risk, high-payoff research areas, such as digital data storage and DNA diagnostic tools.

In NIST's tiny Advanced Technology Program (ATP), launched during the Bush administration in 1990, Clinton saw a vehicle for his vision. He plotted

a course that would nearly quadruple NIST's budget to \$1.4 billion by 1997 and has stuck to it, requesting \$1.023 billion from Congress for 1996. About half that money is set aside for ATP grants to companies willing to put an equivalent amount of their own funds into a long-term development project.

From its genesis, the ATP has faced conservative critics who wonder why underpaid government analysts should rush in where venture capitalists fear to tread—and why the crown collects no royalty when federally funded projects do lead to successful products. As power shifted across the aisles of Congress this winter, ATP supporters felt a chill creep into discussions of the program's future, which can no longer be taken for granted. Even its present is under debate: a bill put to the House of

Representatives in February would take back \$107 million of the \$853.8 million Congress gave to NIST for 1995.

As the debate unfolds, all eyes will be on Prabhakar. "We're on a very steep ramp-up to expand these pilot programs into national-scale efforts," she says. Any slip could be politically perilous. And this afternoon she bids farewell to the ATP manager, who is retiring with no replacement at hand. The challenge of juggling these tasks with mandatory house calls to Capitol Hill could seem overwhelming for someone who in three days will turn 36. Yet Prabhakar exudes only confident enthusiasm, perhaps because her life seems to have prepared her precisely for this job.

Prabhakar is certainly accustomed to being unusual. Born in New Delhi, she was only three when her family emigrated to America, eventually settling in Texas. "When you're an Indian kid in Lubbock, you're inherently different anyway," she says with a laugh. Pursuing her interest in math and science was just another way to

stand out.

Prabhakar raced through an electrical engineering degree at Texas Tech University in three years, then headed to the California Institute of Technology for graduate school. Money was no problem—a pleasantly recurring pattern in her career—thanks to a full scholarship and stipend from the Bell Laboratories Graduate Research Program for Women. More important than the cash, she states, was the offer of two research-

MARTIN H. SIMON/SABA

ers to act as mentors. Robert E. Nahory, who later moved on to Bellcore, and Martin A. Pollack visited her once every year during her graduate studies and employed her during one summer.

Nahory recalls Prabhakar struggling with her classes and with the novelty of being the only woman in her department: "In class the men would always sit in the same seats, and they would leave an empty chair between themselves and Arati. So she began to experiment by moving to different seats. It would cause some general confusion, but everybody would move so that again she was surrounded by empty seats. You can imagine how isolated she must have felt."

"I did not have a good time in graduate school," Prabhakar admits, "but it did force me to think through what I liked and what I didn't. Recently I found some notes I made back then, when I was in the throes of deciding what to do with my life. Looking back, it turns out to be a perfect road map for what I've done since. What I really like is exploring and communicating the connections between science, technology and business. I love that a whole lot more than working on laboratory equipment at three o'clock in the morning."

By 1984, when she received the first Ph.D. in applied physics that Caltech had ever given to a woman, Prabhakar was convinced that she wanted "something orthogonal" to a life of academic research, something that felt more like engineering. When Pollack pointed out an advertisement for a fellowship at the Office of Technology Assessment, Prabhakar hesitated for only a second. "I had never conceived of working in the government as a career," she says, grinning. "I still feel that way."

Nevertheless, she packed her bags for Washington. Her initial assignment could hardly have been more formative: to report on the state of semiconductor research in the U.S. at a time when Japan was marshaling its manufacturers to overtake American dominance in computer chips. Japan's rapid success prompted Congress to create Sematech, a federally backed consortium that provided a seminal test of government's ability to invigorate a flagging industry. Years later U.S. companies won back a leading share of the market. As Prabha-

kar oversaw the public's investment in Sematech, candidate Clinton held up the consortium as a model for his technology policy, even as critics derided the tax dollars spent helping an industry they said would have helped itself.

While Prabhakar met with industry experts around the country to educate herself about their semiconductor research, she inevitably educated them about Arati Prabhakar. Her report thus led directly to a job offer from Richard A. Reynolds, then science director at the Department of Defense's Advanced Research Projects Agency (ARPA), who asked her to run a \$6-million program funding research on gallium arsenide electronics. Prabhakar attributes her big break to "a culture at ARPA that celebrates taking risk—very similar to the culture that is growing now in our ATP. Dick took such a risk when he hired me. I was just a year out of graduate school. I didn't know from nothing. But there were good, smart people to show me the

while the White House was frowning on intervention. ARPA became an ideological battleground.

Prabhakar recounts the lessons of that war. "The Soviet empire was breaking up, and I remember watching the little Baltic countries leading the charge toward independence. They faced a critical question: How far ahead should you get? Lithuania and the others needed to be far enough out that they could drive the agenda. But if they pushed too far, the tanks would roll in, and it would be Prague and 1968 all over again. That's what it was like at ARPA at that point. It was absolutely mandatory that we rethink the traditional approach of working in isolation from the commercial industry. It wasn't cost-effective, it wasn't the best way to do our job, and it wasn't the best way to deliver national security. And yet if you got too far out from the mind-set of the Pentagon, it was clear that the tanks would roll in. And guess what—the tanks rolled in."

In 1990 then ARPA director Craig Fields was rolled out, ostensibly for approving a \$4-million research project that Prabhakar had arranged with a gallium arsenide firm. Critics charged that it was pure venture capitalism, but a later Defense Department review vindicated Prabhakar's decision.

As battle lines again form, this time over the ATP, Prabhakar has more—and more powerful—allies than before. "All my bosses firmly believe that this is the way we should be going," Prabhakar exclaims with conviction.

Republican chairs of key subcommittees in both houses of Congress have voiced more fondness for the program than animosity toward it. But whether the industrial beneficiaries of the ATP will come to its rescue is less certain.

Prabhakar acknowledges that life in the political spotlight might make some of her 3,200 employees uncomfortably warm. "The days are over when NIST was invisible," she comments. "That's the price you pay for stepping up to these big challenges." But she for one is unwilling to retreat to the former isolation of her ivory tower. "You can't allow yourself to get co-opted by the system," she says. "There's really no point in coming to work in the morning unless you can push the boundaries of what's possible." —W. Wayt Gibbs



PRABHAKAR defends her programs at a Senate hearing.

MARTIN H. SIMON/SABA

ropes. The secret to success in these jobs is to work with a broad community and to listen and add value in a collaborative effort rather than trying to do things in an arbitrary way."

Another secret of hers is to be at the right place at the right time. Within several years of her arrival, Congress shifted money and power from the Pentagon to ARPA, raising its budget for microelectronics 10-fold. Prabhakar rode the wave past many senior colleagues, and by 1991 she commanded a budget of \$300 million and all of ARPA's chip research. Prabhakar admits to a sense of déjà vu when she looks back to that time, especially now that Republicans are threatening to rein in NIST's growth. In the late 1980s it was Congress that was pushing the government to promote commercial research actively,

The Puzzle of Declining Amphibian Populations

The number of frogs, toads and salamanders is dropping in many areas of the world. The causes range from destruction of their local habitats to global depletion of the ozone layer

by Andrew R. Blaustein and David B. Wake

Perhaps our fascination with frogs and other amphibians starts in childhood, with the discovery of tadpoles and the observation of their metamorphosis. But for many adults today, interest stems more from observation of another type of change: amphibian populations in many parts of the world seem to be dwindling, and some groups are disappearing from their native habitats completely. The loss—first recognized as a global phenomenon in 1990—deserves attention not only because it is disturbing in its own right but also because frogs and their kin (mainly toads and salamanders) may serve as indicators of the overall condition of the environment.

Amphibians are valuable as gauges of the planet's health for a few reasons. First, they are in intimate contact with

many components of their natural surroundings. For example, as larvae, frogs live in water, but as adults most find themselves at least partially on land. Their moist, delicate skins are thin enough to allow respiration, and their unshelled eggs are directly exposed to soil, water and sunlight. As larvae, they are herbivores and as adults, carnivores. Because amphibians sample many parts of the environment, their health reflects the combined effects of many separate influences in their ecosystem. Second, these animals are good monitors of local conditions because they are homebodies, remaining in fairly confined regions for their entire lives. What happens to frogs and their brethren is happening where humans live and might affect our species as well.

Finally, amphibians are so varied that any single characteristic, unique to the class, can be dismissed as the cause of the dwindling numbers; hence, we suspect that environmental factors are indeed the main cause for their decline. Amphibians are diverse in color, form, behavior and natural history. They vary in physical size, reproductive capacity and population density. And they are found in many ecosystems and habitats, including deserts, grasslands and forests, from sea level to high mountaintops. Although these creatures are most abundant in the tropics, they are also common in temperate zones and can even be found at higher latitudes, such as in Alaska and northern Canada.

Which environmental factors might

account for the rapid decline of animals that have managed, over hundreds of millions of years, to survive events that led to the mass extinction of many species, including the dinosaurs? The explanations that have been proposed are almost as diverse as the amphibian species in jeopardy, ranging from destruction of habitat to natural fluctuations in population size.

One or more of the suggestions do seem to explain the shrinkage of many populations. But in other cases, the reasons for the declines are not obvious. In those instances, the damage may be caused by subtle, interacting aspects of regional or even global conditions. In particular, recent work, completed last spring, has led to the surprising discovery that stratospheric ozone depletion may well be harming amphibian species in some parts of the world.

We began to suspect that the ozone problem might play a role as a result of studies that one of us (Blaustein) and

ANDREW R. BLAUSTEIN and DAVID B. WAKE combine their interests in behavioral ecology and evolutionary biology in their studies of amphibians. Blaustein, an ecologist, is a professor at Oregon State University. Wake, an evolutionary biologist, is a professor at the University of California, Berkeley. Both Blaustein and Wake belong to the Species Survival Commission of the World Conservation Union. Blaustein is co-chairman of the Pacific Northwest Section of the Task Force on Declining Amphibian Populations. Wake was the founding chairman of the International Task Force on Declining Amphibian Populations.

ULTRAVIOLET RADIATION breaking through the dwindling ozone shield in the stratosphere is a newly discovered danger to amphibians who lay their eggs in the open. The radiation, which can damage DNA (*inset*), has recently been shown to account for severe losses of fertilized eggs in at least two such species in the Cascade Mountains of Oregon, including the Cascades frog (*Rana cascadae*) depicted here. The egg losses, in turn, may have led to declines in overall population sizes.

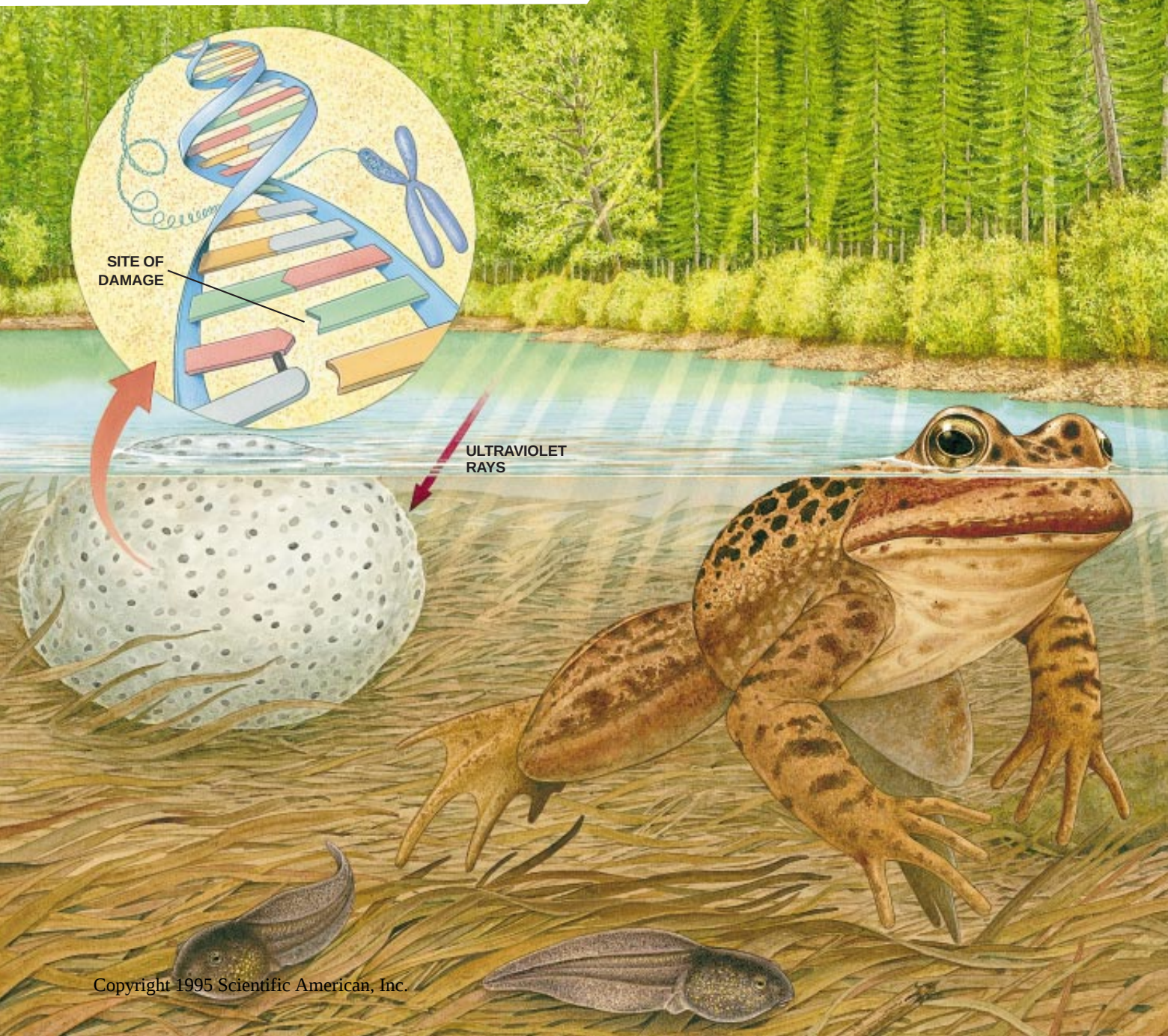
his students began in Oregon in 1979. The most recent experiments, often conducted at relatively high elevations (above 4,000 feet) in remote, undisturbed parts of the Cascade Mountains, examined various aspects of life for several species of amphibians monitored from the egg, or embryonic, stage through the tadpole phase and into adulthood. Although the group intended to carry out a straightforward survey of amphibian behavior and ecology, it discovered some unexpected results.

The Role of Ultraviolet Rays

As part of the research, the team documented massive die-offs of fertilized eggs in two species in particular: the Cascades frog (*Rana cascadae*) and

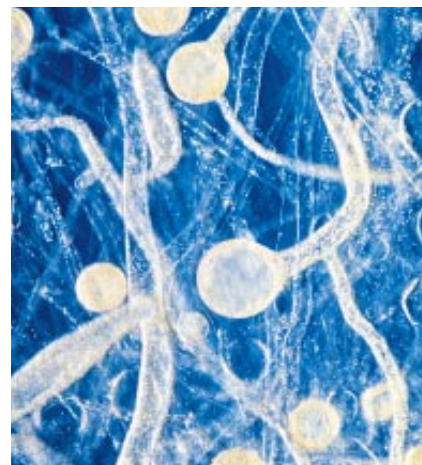
the western toad (*Bufo boreas*). Additionally, over the course of 10 years, the group noticed that the numbers of adults of these species were dropping. The investigators guessed that the shrinking numbers of adult frogs and toads could result from the fact that so few fertilized eggs survived, and thus they began to explore the reasons for the damage to the eggs.

The researchers quickly ruled out the possibility that the chemistry of the water where the animals were laying their eggs was at fault. They brought eggs into the laboratory and reared the resulting embryos in a sample of the same lake water in which other eggs left be-





KEVIN HORGAN



CABISCO Visuals Unlimited

CAUSES OF AMPHIBIAN DECLINES are varied. The most significant threat to populations remains habitat destruction, such as the burning of Brazilian rain forests (far left). Other proposed causes include diseases, such as infection by the *Saprolegnia* fungus (left),

hind had perished. The embryonic frogs and toads developed and hatched normally in the laboratory. Furthermore, chemical analyses of the lakes and ponds where eggs died revealed no obvious pollution or excess acidity.

By the late 1980s another possible cause of egg destruction had presented itself. Scientists in several disciplines documented a decrease in the stratospheric ozone shield that blocks most ultraviolet rays from reaching the ground. These observations led Blaustein and his co-workers to wonder whether increased exposure to ultraviolet radiation could explain the reproductive problems they had seen. They also thought it might explain why many of the amphibian species known to be in decline were mountain dwellers that lay their eggs in open, often shallow, water. Such eggs undergo prolonged exposure to sunlight and thus to any ultraviolet radiation that passes through the ozone shield.

The researchers speculated that excessive exposure to ultraviolet radiation could be contributing to the problems of the Cascades frog and western toad because they were aware of evidence showing that ultraviolet rays can damage plant and animal life. In particular, ultraviolet-B radiation (with a wavelength ranging from 280 to 320 nanometers), in the middle of the ultraviolet spectrum, is especially harmful to living organisms. In humans, for example, it can suppress the immune system, cause cataracts and contribute to skin cancer. What is more, as early as the mid-1970s, Robert C. Worrest of Oregon State University had shown that ultraviolet-B rays

could cause amphibian embryos to develop abnormally in the laboratory.

Few experiments, however, had considered the consequences of ultraviolet radiation on amphibians or other animals in nature. Blaustein and his colleagues therefore set out to determine whether increasing levels of ultraviolet-B radiation could play a role in the decline of amphibian populations in the wild. The team for this enterprise included ecologist Susan C. Walls and molecular geneticists John B. Hays and Peter D. Hoffman, as well as graduate students D. Grant Hokit and Joseph M. Kiesecker, all then at Oregon State University.

The crew based its procedure on an understanding of how ultraviolet radiation affects DNA. When DNA absorbs energy from such rays, the bonds that hold the molecule together break, and new structures are formed. The changes in DNA can disrupt the functioning of cells and may even kill them. But many organisms have the capacity to repair DNA damage caused by ultraviolet radiation. As part of this process, some of those organisms—including certain species of algae, plants, fish, marsupials and amphibians—activate an enzyme known as photolyase, which removes the harmful structures.

By measuring the amount of photolyase produced in the eggs of various amphibians, the workers found that levels varied among species. Most important, they determined that species with falling populations were generally those with eggs that produced low levels of photolyase—and therefore had little protection from ultraviolet radiation.

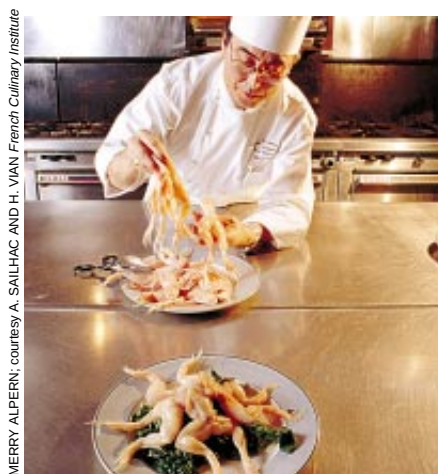
The species with the most photolyase,

the Pacific treefrog (*Hyla regilla*), was not suffering from a decrease in population. Pacific treefrog eggs have about three times as much photolyase as do Cascades frog eggs and six times as much as western toad eggs. Hence, it seems that because the embryonic Cascades frogs and western toads produce low levels of photolyase, they do not make enough of the enzyme to counteract exposure to unusually high amounts of ultraviolet radiation. This lack of protection in turn may lead to the high mortality observed for the eggs and explain why these two species are candidates for threatened status in some states.

Once the group noticed a correlation between lack of protection from ultraviolet radiation and declining population, the next challenge was to find supporting evidence that the rays were actually at fault. The team collected freshly laid eggs of Cascades frogs, western toads, Pacific treefrogs and northwestern salamanders (*Ambystoma gracile*). The salamanders, like the other three groups, lay their fertilized eggs in open, shallow water. Additionally, these salamanders produce extremely low levels of photolyase.

Exposed Eggs Fail to Hatch

The researchers placed the eggs in the bottom of screened enclosures. Atop one third of the containers they placed a cover of clear plastic (Mylar) that shielded the eggs from ultraviolet-B radiation. A second set remained open, fully exposing the eggs. On the remaining third of the boxes, they placed a clear cover of plastic acetate



MERRY ALPERN; courtesy A. SALLHAC AND H. VAN French Culinary Institute

and human consumption of frogs (right). Such consumption was a particular problem in the U.S. in the early part of this century, before protective measures began taking effect. Pollution (far right) of waterways and of the air also threatens amphibians.



BADZIC Sipa Press

that allowed transmission of radiation. This treatment served as a control to ensure that the outcome observed in shaded boxes was not caused by the covers.

The workers placed a total of 48 boxes randomly around lakes and ponds at several different sites where each species normally lays its eggs. The experiments on frog and toad eggs were conducted in the spring of 1993 at relatively high altitudes (greater than 4,000 feet) in the Cascade Range of Oregon. The team studied the eggs of the northwestern salamanders in the foothills of the Oregon Coast Range (600-foot elevation) during 1994. The research continued until all the eggs either hatched or perished, a process that took from one to two weeks because of varying weather conditions.

If it were true that an inability to combat the harm caused by excessive exposure to ultraviolet radiation was destroying the eggs of many amphibian species, the producers of the lower amounts of photolyase would be expected to fare worse, and the producers of higher levels, better. The results of the field experiments were dramatic. More than 90 percent of the northwestern salamander eggs exposed to ultraviolet-B radiation died (compared with 45 percent of eggs protected from the rays). More than 40 percent of the exposed western toad and Cascades frog eggs died (compared with 10 to 20 percent of the shielded eggs). In contrast, almost all the eggs of Pacific treefrogs in all three experimental treatments hatched successfully.

Clearly, amphibian eggs in wild populations were dying from exposure to

ultraviolet-B radiation. And this damage to the eggs was very possibly contributing to the decline in adult populations that had been observed earlier. Investigators do not know whether northwestern salamanders are disappearing, but if these experiments are any indication, chances are good that those creatures, too, are in jeopardy.

By what mechanism does ultraviolet radiation lead to the destruction of amphibian eggs and embryos? Other research by Blaustein and his colleagues may have uncovered a partial explanation. It turns out that since the late 1980s, increasing numbers of amphibians in Oregon have been sickened by the fungus *Saprolegnia*, which is found naturally in lakes and ponds. The fungus is also known to infect hatchery-reared fishes, especially salmon and trout. Perhaps fish that have been released into lakes and are infected with *Saprolegnia* contaminate amphibian eggs in those waters. Because ultraviolet rays can impair immune function in many animals, it seems reasonable to guess that some amount of egg damage in amphibians is caused by an ultraviolet-induced breakdown in the ability of amphibian embryos to resist infection by the fungus.

Aside from harming fertilized eggs, ultraviolet radiation may contribute to declines in amphibian populations by reducing the supply of aquatic insects on which frogs and their relatives feed. High levels of such radiation have been known to kill insect larvae as well as aquatic algae.

The work in Oregon has provided one potentially important clue to the mys-

tery of amphibian disappearance. But many questions still remain. How many eggs can fail to hatch before a population itself begins to decline? Does ultraviolet radiation harm growing tadpoles that congregate in shallow water? And are adults that bask in sunlight affected directly by ultraviolet radiation? The two of us are now beginning to focus on these issues.

The Threat of Habitat Destruction

As worrisome as the increase in ultraviolet radiation seems to be, it is not the only potentially significant cause of shrinkage of amphibian populations. In the Monteverde cloud forest of Costa Rica and in the Australian rain forests, for example, amphibians typically live under a dense foliage canopy and hide their eggs. Yet many of their numbers are also in decline.

One of us (Wake) has been investigating causes of dwindling amphibian populations since the 1970s, when the first hints of a problem began to emerge. The issue is indeed compelling, for although evidence of falling numbers is strong in various parts of the world, in other areas amphibians appear to be doing well. This puzzling situation has prompted us and others to examine closely the possible reasons for the declines we have seen.

No single explanation fits every case, but all seem to be important to one degree or another. Destruction and modification of habitat are probably the most serious causes of falling amphibian populations. Like other animals, amphibians are threatened when forests are de-

Sharp-Snouted Torrent Frog (*Taudactylus acutirostris*), found in Australia
Cause of decline unknown



KLAUS UHLENHUTH Animals Animals

Common Toad (*Bufo bufo*), found in Europe
Cause of decline unknown



KIM TAYLOR Bruce Coleman Inc.

Tiger Salamander (*Ambystoma tigrinum*), found in North America
Pollution, possibly acid precipitation



Z. LESZCZYNSKI Animals Animals

AMPHIBIAN SPECIES shown here are among the many with dwindling populations—or are already extinct—for reasons that are now being investigated in

Additional causes may explain isolated cases of dropping numbers of amphibians. Some populations may be decreasing because they are collected for human consumption. In France, for instance, the demand for frog legs is tremendous: the French eat 3,000 to 4,000 metric tons of them a year. Some 20,000 frogs must be sacrificed in order to supply a single metric ton of legs. And before the turn of the century, red-legged frogs (*R. aurora*) were probably overharvested as a food source in Oregon and California.

Ironically, efforts to boost amphibian populations in the western U.S. probably created more problems for the native amphibians there and provided an illustration of yet another possible cause of population declines: the introduction of nonnative species to an area.

To make up for decreases in the number of red-legged frogs, inhabitants of Oregon and California introduced the bullfrog (*R. catesbeiana*). This animal, with its voracious appetite, competed with or preyed on native amphibians in its new habitat. Noting that the introduced bullfrogs have become quite abundant in some places where the original frog species have declined, many biologists have recently suggested that bullfrogs are a major cause of falling numbers. And at least two ongoing studies have directly linked the introduction of bullfrogs to the dwindling of native frog species.

Similarly, introduction of fish into an ecosystem may hurt amphibians, especially in regions with few species of fish, low numbers of individual fish or no fish at all. In the southern Sierra Nevada Mountains of California, the introduction of salmon and trout into streams has been implicated in the demise of mountain yellow-legged frogs (*R. muscosa*). These fish species directly harm



JAMES HANKEN Bruce Coleman Inc.

Mountain Yellow-Legged Frog (*Rana muscosa*), found in California
Introduction of fish to habitat



MICHAEL FOGDEN Bruce Coleman Inc.

Harlequin Frog (*Atelopus varius*), found in Costa Rica
Cause of decline unknown

stroyed and wetlands are filled in or paved. Indeed, such activities probably account for the decrease in a majority of species threatened today.

In one striking example of this phenomenon, a recent survey in western North Carolina showed that clear-cutting of national forests leads to the deaths of enormous numbers of salamanders every year. Although most of the species involved have relatively large geographic ranges and are not in danger of extinction, the findings have distressing implications for amphibians living in tropical America, Africa and Asia. There many amphibians are more vulnerable because they have very limited geographic ranges.

Pollution Plays a Part

Pollutants, too, may have altered amphibian populations in some parts of the world, although data on the effects of pollution on these creatures are sparse. Some evidence suggests that acid rain and snow, fungicides, herbicides, insecticides and industrial chemicals may all act by impairing the reproduction and development of amphibians. Certain synthetic compounds can mimic the activity of naturally occurring hormones. Examination of birds, fish and reptiles indicates that these substances can have drastic consequences, such as a reduction in sperm count

and the alteration of male genitalia.

Diseases—possibly related to environmental pollution—seem to jeopardize some amphibians as well. Recall, for instance, that eggs of the Cascades frog and western toad are vulnerable to the fungus *Saprolegnia* and that susceptibility to the fungal infection is probably increased by exposure to excessive ultraviolet radiation. Further, the late Arthur N. Bragg, when he was at the University of Oklahoma, showed that *Saprolegnia* can destroy whole populations of tadpoles, although this discovery has been largely overlooked as a cause of amphibian deaths.

So far only a few studies have linked a disease to the extinction of an entire population of amphibians. Investigators have found, however, that the bacterium *Aeromonas hydrophila* may have triggered the disappearance of several populations of western toads in Colorado. The bacterium is highly contagious and has been implicated as well in the death of adult frogs, toads and salamanders in several other states.

Some scientists attribute the apparent shrinkage of amphibian populations to natural fluctuations in population size. Yet certain long-term investigations show a more or less steady decline in the number of amphibians over the past 20 to 30 years—an indication that in some populations other forces are at work.

Treefrog (*Nyctimystes dayi*),
found in Australia
Cause of decline unknown



STEPHEN J. RICHARDS

Natterjack Toad (*Bufo calamita*),
found in Great Britain
Pollution, particularly acid rain



HANS REINHARD Bruce Coleman Inc.

Golden Toad (*Bufo periglenes*),
found in Costa Rica (possibly extinct)
Cause of decline unknown



MICHAEL FOGDEN Bruce Coleman Inc.

detail. In many cases, the declines remain puzzling. For each species, the causes of trouble implicated are given next to each picture.

amphibians by eating eggs, tadpoles and even adults, but they also have a broader and potentially more profound result.

Many separate amphibian populations are linked to one another by streams patrolled by few or no fish. These links are important because frogs and their relatives are extremely vulnerable to changes in their local habitat, and they rely on the appearance of occasional migrants to help them rebuild diminished communities. The addition of new fish species into an area can block migration between communities and thus prevent the reconstitution of endangered populations.

Dangerous Consequences

The disappearance of amphibians represents more than just a loss of esthetically and behaviorally appealing creatures. These animals are crucial components of many ecological communities, and they can directly benefit humans. In some ecosystems, amphibians are the most abundant vertebrates, and so their absence can seriously disrupt the functioning of the rest of the ecological community. Adult amphibians are hunters of various animals, including mosquitoes, flies, fish, birds and even small mammals. Also, amphibian larvae serve as a food supply for aquatic insects, fish, mammals and birds. Destruction of frogs, toads and salamanders thus has repercussions elsewhere in the food chain.

From the perspective of humans, amphibians represent a storehouse of pharmaceutical products waiting to be tapped fully. Hundreds of chemical secretions have been isolated from amphibian skin, and scientists are just beginning to learn how valuable these substances may be. Some of these compounds are already used as painkillers



BOB AND CLARA CALHOUN Bruce Coleman Inc.

Western Toad (*Bufo boreas*),
found in western North America
Disease related to ultraviolet radiation



W. DUELLMANN U. Kansas Natural Hist. Museum

Gastric Brooding Frog (*Rheobatrachus silus*), found in Australia (possibly extinct)
Cause of decline unknown

and in treatment of victims of traumas ranging from burns to heart attacks. Others are being investigated for their antibacterial and antiviral properties. As amphibians disappear, potential cures for a number of maladies go with them.

The evidence that depletion of the ozone shield in the stratosphere can harm the developing embryos of amphibians highlights the complexity of the forces leading to the elimination of species. Nevertheless, habitat degradation and destruction clearly remain the most powerful causes of amphibian disappearance around the world. If habitat modification occurs slowly enough—as it did for 3,000 years in western Europe—amphibians can adjust and even

adapt to human-induced alterations. But many of the changes we have discussed, such as rises in ultraviolet levels and in the amounts of pollutants in the environment, have occurred so rapidly that species with long generation times often cannot adapt quickly enough.

There are a lot more species of amphibians than scientists studying them. Of those that are known, many have been seen only once, at the time of their discovery. The number of species described continues to increase at a rate of 1 to 2 percent a year. If, as we believe, many of these species are at risk, a wonderfully diverse group of creatures is vanishing from the planet at a time when study of them has just begun.

FURTHER READING

DECLINING AMPHIBIAN POPULATIONS. David B. Wake in *Science*, Vol. 253, page 860; August 23, 1991.
AMPHIBIAN DECLINES: JUDGING STABILITY, PERSISTENCE, AND SUSCEPTIBILITY OF POPULATION TO LOCAL AND GLOBAL EXTINCTION. Andrew R. Blaustein, David B. Wake and Wayne P. Sousa in *Conservation Biology*, Vol. 8, No. 1, pages 60-71; March 1, 1994.
AMPHIBIAN DECLINES AND CLIMATE DISTURBANCE: THE CASE OF THE GOLDEN TOAD AND THE HARLEQUIN FROG. J. Alan Pounds and Martha L. Crump in *Conservation Biology*, Vol. 8, No. 1, pages 72-85;

March 1, 1994.
PUTTING DECLINING AMPHIBIAN POPULATIONS IN PERSPECTIVE: NATURAL FLUCTUATIONS AND HUMAN IMPACTS. Joseph H. K. Pechmann and Henry M. Wilbur in *Herpetologica*, Vol. 50, No. 1, pages 65-84; March 1994.
UV REPAIR AND RESISTANCE TO SOLAR UV-B IN AMPHIBIAN EGGS: A LINK TO POPULATION DECLINES? Andrew R. Blaustein, Peter D. Hoffman, D. Grant Hokit, Joseph M. Kiesecker, Susan C. Walls and John B. Hays in *Proceedings of the National Academy of Sciences*, Vol. 91, No. 5, pages 1791-1795; March 1, 1994.

Quest for the Limits of the Heliosphere

Four aging spacecraft are racing to the outer reaches of the solar system. Soon they may break through the last barriers to interstellar space

by J. R. Jokipii and Frank B. McDonald

Glowing comets and brilliant auroras are visible reminders that space in the inner solar system is far from empty. This region is permeated by swiftly flowing charged particles emanating from the sun, a continuous torrent of solar wind that often blows in sudden gusts. The fetch of this wind extends well past the orbit of the earth or the range of visible comets. The outward rush of particles and the solar magnetic field carried with them carve an enormous spherical cavity in the interstellar medium that reaches far beyond the orbit of the most distant planets of the solar system. This immense region, a bubble of solar dominance within the vastness of space, is called the heliosphere.

One might imagine that with increasing distance from the sun the heliosphere gradually fades to a diffuse boundary wherein particles of the solar wind gently mix with the interstellar

breeze of dust and gas. But this is not at all the case: near the limits of the outer heliosphere lies an abrupt discontinuity at which a myriad of intriguing physical phenomena are thought likely to occur. As of yet, however, astrophysicists have no direct measurements of the heliosphere's outer margins and so must infer, theorize or simply speculate on its exact nature. We do not even know with any certainty how far from the sun this boundary forms. But our ignorance of the distant reaches of the heliosphere may last only a few more years, when space probes finally break through this first barrier toward interstellar space.

Past the orbits of Neptune and Pluto, on trajectories taking them beyond the edges of the solar system, drifts a small flotilla of spacecraft. This modest scientific armada consists of *Pioneer 10* and *11*, along with *Voyager 1* and *2*, all of which were launched about two decades ago. If we were to look back at the solar system from any of these spacecraft today, the sun would be the brightest object in view, but it would nonetheless appear more than 1,000 times dimmer than as seen from the earth. Even at these great distances, though, the four spacecraft remain well within the heliosphere. Onboard instruments continue to register disturbances originating on the sun's surface that propagate outward at about 400 kilometers per second. Despite this enormous velocity, these sudden gusts still take many months to reach the probes.

The original mission of the *Pioneer* and *Voyager* spacecraft—to study the giant planets Jupiter, Saturn, Uranus and Neptune—stands as one of the enduring triumphs of space exploration. But the continuing vitality of these four probes after the celebrated planetary fly-bys and our growing awareness of the complex and dynamic behavior of

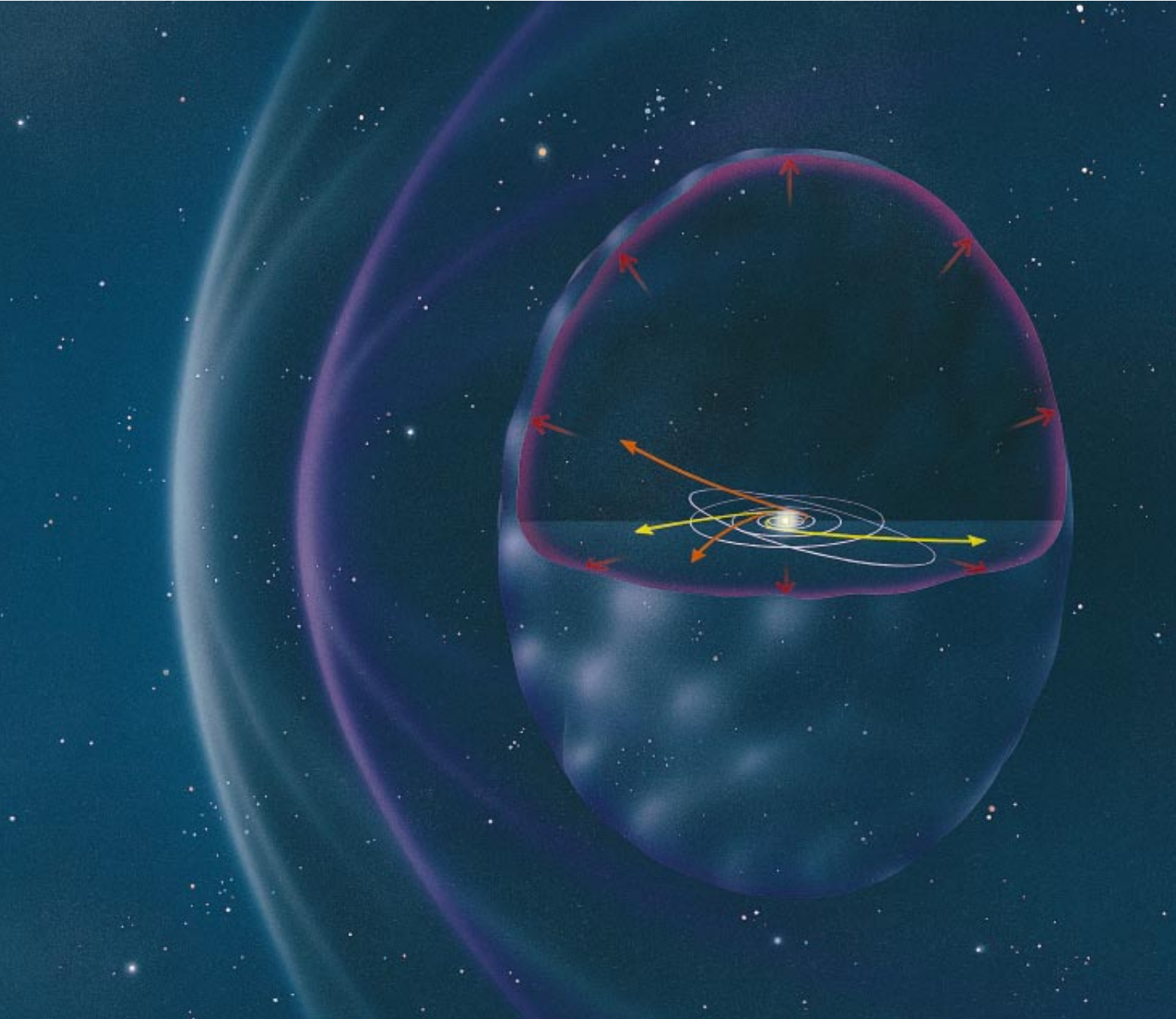
the distant solar wind have engendered an important second mission for these versatile scientific minions: to study the most remote parts of the heliosphere and its interface with the interstellar medium. The success of this newly established mission depends not only on the technical capabilities of the spacecraft and their earth-based controllers but also on the nature of the heliosphere itself.

The Outer Heliosphere

The general structure of the solar wind and the heliosphere was first outlined three decades ago by Eugene N. Parker of the University of Chicago in a series of revolutionary theoretical papers. Observations have since verified the structure he predicted, at least out to the position of the farthest spacecraft, *Pioneer 10*. We now have confirmed that the solar wind, as it moves radially outward from the sun, brings with it the imprint of the solar atmosphere: certain parts of the sun emit high-speed wind in vast streams that flow hundreds of kilometers per second faster than the typical wind. If the sun were stationary, those streams would simply form linear rays, but because it rotates every 27 days, fast streams originating near the solar equator can overtake slower streams emanating from adjacent areas on the surface. Sometimes this configuration lasts for several solar rotations, setting up regions in space where the interface between fast and slow streams of solar plasma becomes spiral in shape. These irregularities rotate with the sun and are known to space physicists as co-rotating interaction regions.

In addition, some parts of the solar atmosphere can eject irregular puffs and gusts that travel as transient disturbances in the wind. Occasionally, huge

J. R. JOKIPII and FRANK B. McDONALD have been involved in the study of energetic particles in the heliosphere for many years. Jokipii has been professor of planetary sciences and astronomy at the University of Arizona since 1974 and has served as an interdisciplinary investigator on the Ulysses mission. His major scientific interest is the origin and propagation of energetic particles in space. McDonald is a senior research scientist at the University of Maryland's Institute for Physical Science and Technology and has been a principal investigator on many National Aeronautics and Space Administration projects, in particular the cosmic-ray experiments on the *Pioneer* and *Voyager* spacecraft. He has also been active in planning other scientific missions, such as the *Interplanetary Monitoring Platform* and *High Energy Astronomical Observatory* series as well as the *Compton Gamma Ray Observatory*.



HELIOSPHERE encompasses that part of space dominated by the solar wind. Near its outer margins, outflowing solar plasma (red arrows) is deflected by the flow of interstellar gas at the heliopause (purple), but only after slowing abruptly at

the spheroidal termination-shock front. A bow shock (white) may also form in the interstellar gas. Data from *Pioneer* (yellow arrows) and *Voyager* (orange arrows) should improve understanding of the heliosphere's farthest boundaries.

eruptions on the sun produce blast waves that severely disrupt the solar wind's more steady currents.

The many varied interactions of the solar-wind plasma produce shock waves, which heat the wind and also generate energetic particles. Co-rotating interaction regions and their associated shock waves are a major feature of the solar wind out to more than 10 astronomical units. (One astronomical unit, or AU, is the radius of the earth's orbit around the sun, some 150 million kilometers, or 93 million miles.) Farther out, such interaction regions combine, forming

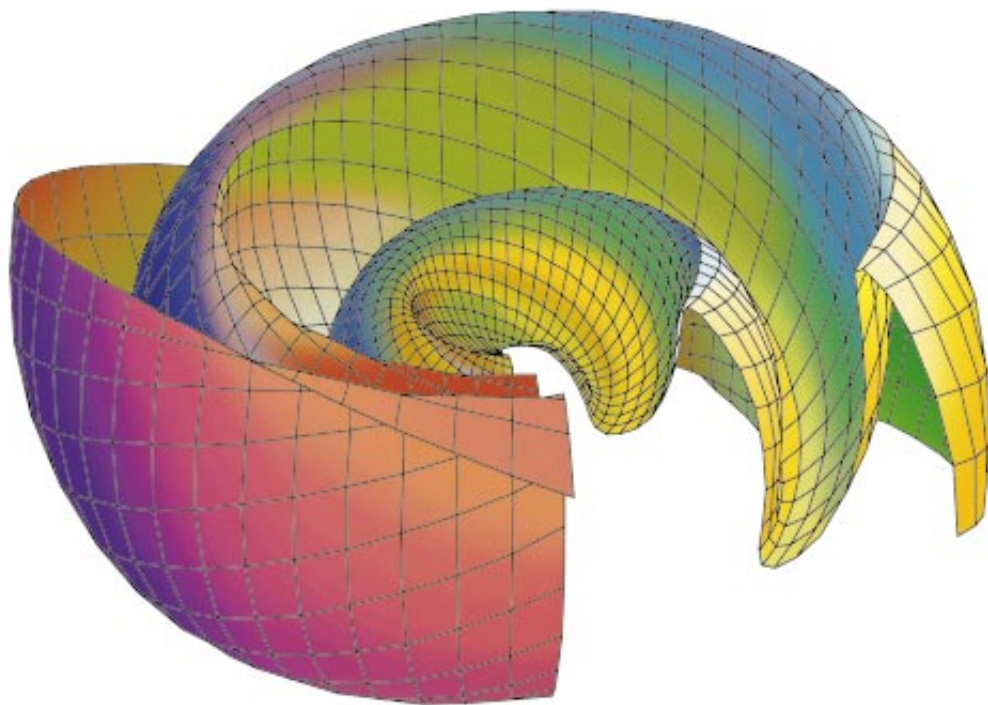
so-called global merged interaction regions, which populate space to the outer reaches of the heliosphere.

The Interplanetary Magnetic Field

Embedded within the stream structure of the heliosphere lies a complex interplanetary magnetic field. The interplay of the magnetic field and solar wind can be rather complex; some of this behavior, however, can be readily visualized in terms of the familiar concept of magnetic lines of force and the properties these field lines give to

the solar wind as it expands outward.

By earthly standards the plasma that constitutes the solar wind might seem rather insubstantial and formless. Yet because it is a good electrical conductor and because the kinetic energy of the flow is so much greater than the energy of the magnetic field, the magnetic-field lines in the heliosphere can be treated as though they move with the solar wind, being effectively "frozen in." This frozen magnetic flux lends the tenuous plasma added pressure and viscosity. These properties develop from magnetic forces rather than from



NEUTRAL CURRENT SHEET forms in the solar wind under the influence of the oppositely directed magnetic fields above and below the sun's magnetic equator. During the solar cycle, the tilt of the magnetic equatorial plane changes along with rising and falling solar activity. This tilt, combined with the rotation of the sun, creates a spiral geometry (shown here in idealized form) as the solar wind carries the trapped magnetic field outward.

J. R. JOSEPH

the more familiar molecular interactions found in denser fluids. So the heliosphere contains a tangle of magnetic-field lines, stretched out by the wind into an enormous spiral whorl, which until recently had hardly been explored.

Early observations indicated that the magnetic-field lines alternated between being directed inward or outward from the sun at different solar longitudes. The first chance to measure the field at relatively high heliographic latitudes came in 1974, after *Pioneer 11* deflected its trajectory out of the ecliptic plane using Jupiter's large gravitational attraction to exchange momentum. It came as a surprise (although Michael Schulz, a space physicist at Aerospace Corporation, had predicted it) when *Pioneer 11* and *Voyager 1* reached the heliographic latitude of 16 degrees north and discovered that the magnetic field was nearly always directed outward. Edward J. Smith of the Jet Propulsion Laboratory in Pasadena, Calif., concluded that the spacecraft were observing magnetic fields carried by the solar wind and that at northern solar latitudes this field was oriented away from the sun.

The *Pioneer* and *Voyager* missions showed in 1976 that the sun's magnetic field was organized such that the field lines in the northern hemisphere generally pointed outward from the sun; those in the opposite hemisphere pointed inward. Because the polarity of the

sun's field changes every 11 years (at the time of the sunspot maximum), a magnetic cycle lasting 22 years results. So in 1986 the *Pioneer 11* and *Voyager 1* space probes detected that the northern field was pointed duly inward.

In the heliosphere the transition between the inward- and outward-directed magnetic fields has the shape of a very thin, warped surface that is carried outward by the solar wind to form a vast interplanetary current sheet. Solar rotation twists the sheet so that the wrinkles lie along spiral magnetic-field lines and rotate with the sun. Space probes near the solar equatorial plane detect magnetic fields that are alternately directed inward and outward as the current sheet rotates past them in space. The sheet is least wrinkled during solar minima, the periods of fewer sunspots and lowered activity that occur every 11 years. Its geometry becomes so convoluted during solar maxima that the normal sheetlike structure becomes entirely unrecognizable.

But in whatever configuration, the magnetic-field pattern originating at the sun's surface is carried to the distant margins of the heliosphere by the solar wind over about a year's time. During this period, fast plasma streams continue to merge with slower ones, spawning regions of enhanced plasma density and magnetic-field strength. Out to some great but as yet unknown

distance, the large-scale structure of the solar wind and the magnetic field is fundamentally preserved.

The Termination Shock

As it travels, the solar wind expands over an increasingly large volume. Eventually the solar-wind plasma is spread so thinly that it can no longer push outward against the small inward pressure of the local interstellar medium. The wind does not slow down gradually at this point, because its velocity is greater than that at which disturbances can move within it. Instead the solar wind undergoes a sudden, violent change in speed.

This behavior follows from the fundamentals of supersonic fluid flow. For the motion to diminish incrementally, the downstream material must signal the upstream fluid to slow. These signals must be carried by sound waves moving through the medium. But such waves cannot propagate against flows moving faster than sound. As a result, the upstream fluid crashes into the fluid ahead,

setting up a confrontation called a shock wave. Something similar occurs in a highway accident when cars behind cannot slow down fast enough to avoid hitting those ahead.

Much as with a multicar accident, we expect the solar-wind termination shock to be irregular and turbulent. As the solar-wind gas passes through the shock, its outward velocity should slow to about one quarter of its original value. Some of the wind's kinetic energy is converted to heat, raising the temperature of the interstellar gas to more than a million degrees Celsius. Some kinetic energy goes into compressing the magnetic field: we expect that field strength should jump to about four times its value inside the shock. So at this boundary, where the solar wind trades outward velocity for heat and turbulence, we expect to find a giant, spheroidal shock front with a complex but still somewhat mysterious structure.

Evidence from Cosmic Rays

Before spacecraft offered direct measurements, astrophysicists relied mainly on the study of cosmic rays to deduce something of the nature of the outer heliosphere. Galactic cosmic rays are subatomic particles (electrons, protons, all the heavier nuclei from helium to uranium, positrons and a small num-

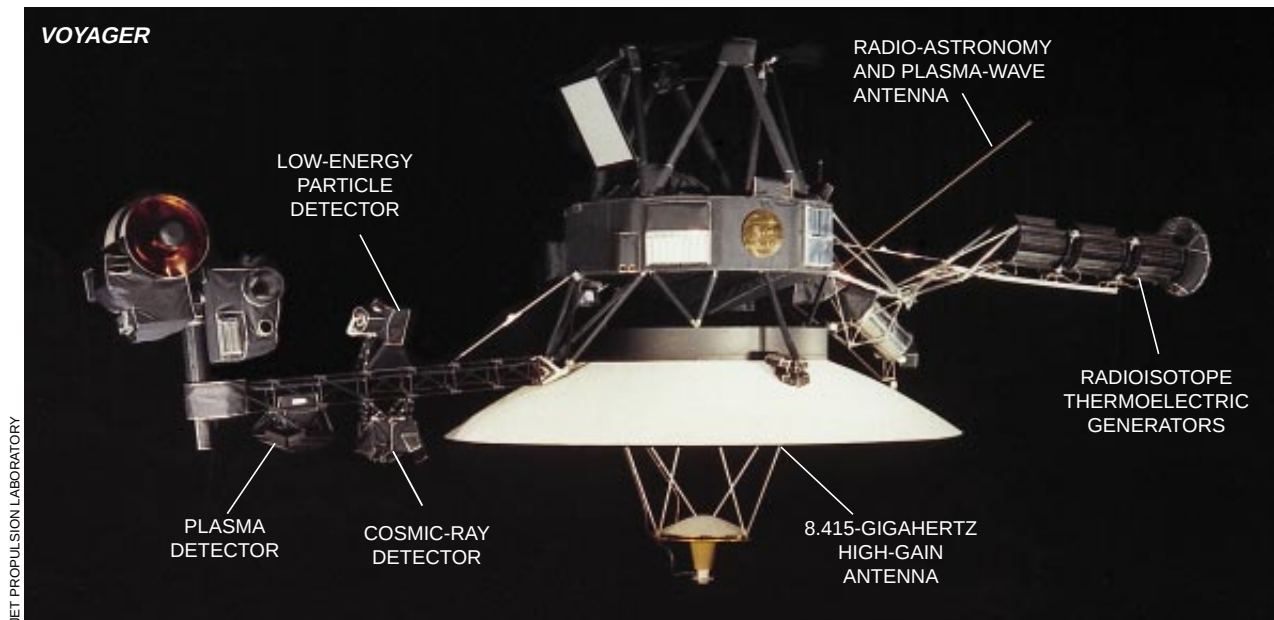
Probes Built to Go the Distance

Deep-space missions present major technical challenges to spacecraft designers in the areas of weight, power and communications. The Pioneer missions called not only for escape from the earth's gravity but also for sufficient energy to reach Jupiter with a travel time of two years or less. The solution was to use a high-energy rocket booster and to keep the spacecraft as light as possible—at launch *Pioneer 10* weighed only 250 kilograms. Reliable sources of electricity were also critical: far from the sun, the only practical ones are radioisotope thermoelectric generators (RTGs), which use the decay of radioactive materials to produce electricity from heat. Each *Pioneer* spacecraft has four RTG units, which initially generated a total of 155 watts. The ultimate lifetime of these missions will probably be defined by the radioactive decay of the plutonium oxide fuel and the degradation of the conversion elements within the RTG units.

The *Voyager* probes were designed in the mid-1970s using what had been learned from the earlier Pioneer missions. A more powerful launch vehicle made it possible to deploy heavier (825 kilograms) and more complex space-

craft: the *Voyager* design includes significant onboard computer capability, an experiment platform with precision pointing and improved RTG units that supplied 470 watts at launch. The sophistication and flexibility of the *Voyager* system have been demonstrated by the considerable reengineering that was done in flight to prepare *Voyager 2* for its late-scheduled encounter with Uranus [see "Engineering *Voyager 2*'s Encounter with Uranus," by Richard P. Laeser, William I. McLaughlin and Donna M. Wolff; SCIENTIFIC AMERICAN, November 1986].

Communication with the *Pioneer* and *Voyager* probes demands a large onboard antenna, the dominant feature. The *Pioneer* craft support a 2.7-meter parabolic dish antenna, somewhat smaller than the 3.7-meter dish found on the *Voyager* probes. The *Pioneer* antenna spins about the spacecraft's axis, which is kept pointed toward the earth by the occasional use of small thrusters. The *Voyager* spacecraft do not spin but are stabilized about all three axes so that the high-gain antenna can be kept directed toward giant antennae on the earth.



JET PROPULSION LABORATORY

GOLDSTONE 70-METER ANTENNA



JET PROPULSION LABORATORY

The three sites of the National Aeronautics and Space Administration's Deep Space Network (in California, Australia and Spain) are among the most critical components of the Pioneer and Voyager programs. During the 20 years since the Pioneer 10 mission, the Jet Propulsion Laboratory in Pasadena, Calif., has significantly upgraded this system by adding new receivers and increasing antenna size. But such improvements cannot compensate for the faltering signal levels caused by the ever increasing distance to the four space probes. For more reliable communication, the total rate of data transmission from *Pioneer 10* has been slowed to its minimum value of 16 bits a second—about as fast as one might send messages by Morse code. The *Voyagers'* larger antenna and higher transmission frequency make possible the use of the more available 34-meter antenna of the Deep Space Network and allow communication at 160 bits a second—slower than most computer modems but still better than the telegraph from *Pioneer*.

ber of antiprotons) that travel at close to the speed of light and appear to populate all parts of the universe. Their ubiquitous presence in the cosmos is inferred from their ability to produce high-energy gamma rays and radio waves (which can be detected on the earth). Within our galaxy, cosmic rays commonly originate where the shock-wave remnants of supernova explosions are thought to accelerate the particles to extremely high energies. In addition, during periods of heightened activity, the sun, too, can occasionally produce significant numbers of solar "cosmic" rays of lower energy.

The heliosphere is constantly bathed by galactic cosmic rays. These cosmic-ray particles can diffuse upwind against the solar plasma because of their extremely high speeds and the presence of irregularities in the field. Because of their electrical charge, the particles gyrate tightly around magnetic-field lines, and as a result, cosmic rays also tend to travel out of the heliosphere along with the frozen-in magnetic flux. In general, the solar wind acts to modulate the intensity of cosmic rays impinging on the earth, making it difficult for them to reach the inner heliosphere. This exclu-

sion is most effective at low energies; highly energetic cosmic rays proceed largely unaffected. Because the fraction excluded varies with solar activity, cosmic-ray intensity follows the 11-year sunspot cycle: it peaks when solar activity is at a minimum.

Galactic cosmic rays pass through the outer heliosphere, and so they can provide valuable information about this unexplored region. Much of our understanding of the outer reaches of the solar system has been derived by comparing models of how cosmic rays traverse the heliosphere with observations. For example, data from the four deep-space probes have shown that the cosmic-ray gradient—the rate at which the intensity of galactic cosmic rays increases with heliocentric distance—is much smaller than was expected. This finding indicates that the heliosphere is larger than was predicted before the launch of the *Pioneer* and *Voyager* probes.

Anomalous Cosmic Rays

In the early 1970s, as the *Pioneer* spacecraft moved toward Jupiter, detectors on a number of spacecraft revealed the existence of an unexpected

low-energy cosmic-ray component. Continuing study has demonstrated an enhancement at low energies for rays consisting of helium, nitrogen, oxygen, neon, argon and, most recently, hydrogen nuclei. This peculiar composition and energy spectrum define the anomalous cosmic-ray component. Observations from the *Pioneer* and *Voyager* spacecraft have shown that the intensity of the anomalous cosmic rays increases with distance from the sun.

What is the origin of this mysterious cosmic-ray component? Work over the past two decades has painted a compelling picture of how these cosmic rays are generated, although its accuracy is not completely proved.

In 1974 Lennard A. Fisk, Benzion Kozlovsky and Reuven Ramaty, while at the National Aeronautics and Space Administration Goddard Space Flight Center, suggested that the anomalous component originates as neutral atoms in interstellar space. As the heliosphere moves through the interstellar gas, neutral atoms, which are not affected by magnetic fields or other forces of the plasma, stream freely into the inner heliosphere. Those that pass near the sun are ionized by solar radiation or by the solar wind itself to become singly charged ions. Once the neutral atoms become ions, the magnetic-field lines in the solar wind snare them and convect these particles outward. Fisk and his colleagues speculated that subsequent acceleration to higher energies turns these ions into the anomalous cosmic rays.

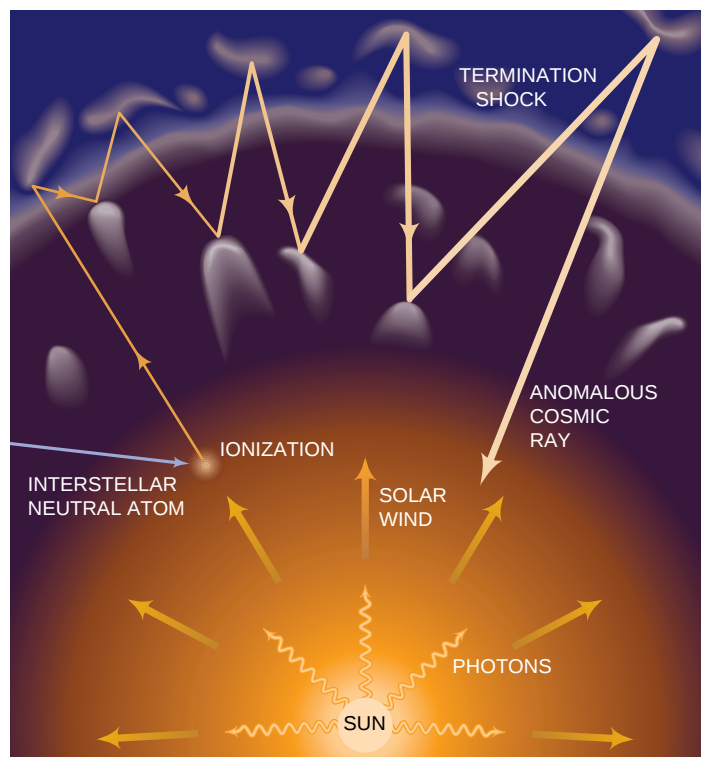
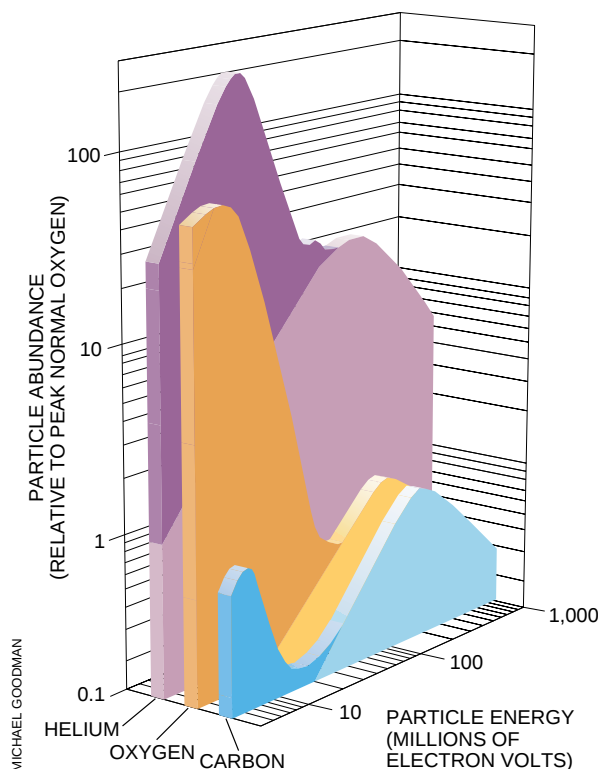
The original basis for this suggestion was that most of the carbon in the interstellar medium cannot take part in this process, because carbon is almost completely ionized in interstellar space (which explains its very low abundance in the anomalous component). Recent observations near the earth by a number of space missions have demonstrated that the anomalous oxygen (and presumably also the other components) is singly charged. This result supports the model of Fisk and his co-workers: cosmic-ray nuclei from a nearby source (within the heliosphere) can retain some electrons, whereas normal cosmic rays are fully stripped of all their electrons during their passage through the galaxy.

But how were these newly formed ions accelerated to the observed cosmic-ray energies? During the 1970s, a number of proposals were put forth. None, however, successfully predicted the steady increase in the intensity of the anomalous cosmic rays registered by the *Pioneer* and *Voyager* probes as they moved far out into the heliosphere. Then, in 1981, one of us (Jokipii), along with Mark E. Pesses and David Eichler,



J. R. JOKIPII

TERMINATION SHOCKS are not unique to the heliosphere; one can be seen in water running in a kitchen sink. The streaming water initially radiates outward faster than waves can propagate through it. As a result, the surrounding fluid cannot send an inward signal that its motion has slowed. A shock front forms where the fast- and slow-moving parts of the fluid abruptly collide. The termination-shock boundary of the heliosphere, like its water analogue, is likely to be irregular and turbulent.



COSMIC-RAY PARTICLES generally show a smoothly varying distribution of energies (left), but certain nuclei are strangely abundant at low energies (darker shading). This anomaly arises from neutral particles that stream into the heliosphere,

become ionized and convect outward. Particles accelerate to higher energies by scattering off magnetic irregularities on both sides of the shock front (right). Accelerated particles deflect inward and are detected as anomalous cosmic rays.

both then at the University of Maryland, suggested that the acceleration of singly charged ions occurs at the termination-shock boundary. Plasma shocks can accelerate charged particles, and this location seemed a likely site for energizing the anomalous cosmic rays—it contains the strongest, most long-lived shock anywhere in the heliosphere. Detailed computer modeling has since shown that most observed features of the anomalous component follow naturally from this notion.

Observations of the Shock Front

Important clues about the nature of the termination region have been collected by Donald A. Gurnett and William S. Kurth of the University of Iowa. Since 1983 they have registered low-frequency bursts of radio noise (at two to three kilohertz) using detectors on board both the *Voyager* spacecraft. The signals persist for many months and then gradually drift to higher frequencies. In July 1992 these researchers observed the onset of a particularly strong radio event and noted that it occurred more than 400 days after an unusually intense period of solar activity. This sequence followed the same pattern as another large noise burst in 1983. These remarkable radio signals probably orig-

inate just beyond the termination shock and, along with the anomalous cosmic rays, provide tantalizing information about this vast unexplored frontier.

As the *Pioneer* and *Voyager* space probes speed farther and farther from the sun, there is an increasing likelihood that they will soon encounter the termination shock. Estimates based on what was then known about the interstellar medium had originally put the termination-shock boundary anywhere from 75 to 150 AU from the sun, but data collected so far from the probes would suggest considerably smaller values. So it is entirely possible that one or more of the probes will reach the shock within the next decade.

NASA scientists have therefore taken steps to ensure that the proper measurements will be made during passage through the shock. Indeed, they believe they may have several opportunities to observe it, as gusts and turbulence in the solar wind move the termination shock in and out—perhaps leading to multiple crossings as the front moves back and forth past the spacecraft. Once the spacecraft finally pass beyond the shock, the wind will slacken, and, for the first time, an artifact of humanity will begin to experience directly the effects of the interstellar plasma. Perhaps then the true nature of the inter-

stellar medium will finally be clarified.

Sometime in the 21st century, after having reported the physical conditions of the outer heliosphere and possibly the termination shock itself, the four spacecraft will continue their journey to the stars. *Pioneer 10* should remain operational until the turn of the century (at about 70 AU), and *Voyager 2* has enough consumables to last until about 2015 (at about 130 AU). But even after steamage and communication are lost, for eons to come these probes will follow a well-charted course through our galaxy as four small man-made objects added to the gaseous clouds of interstellar space. They go as the first voyagers from planet Earth, like small bottles tossed into an infinite sea.

FURTHER READING

- THE HELIOPAUSE. Steven T. Suess in *Reviews of Geophysics*, Vol. 28, No. 1, pages 97–115; February 1990.
- PHYSICS OF THE OUTER HELIOSPHERE. Edited by S. Grzedzielski and D. E. Page. COSPAR Colloquia Series, Vol. 1. Pergamon Press, 1990.
- EXPLORING THE SUN. Karl Hufbauer. Johns Hopkins University Press, 1991.
- SOLAR WIND SEVEN. Edited by E. Marsch and R. Schwenn. COSPAR Colloquia Series, Vol. 3. Pergamon Press, 1992.

Machines That Learn from Hints

*Machine learning improves significantly
by taking advantage of information
available from intelligent hints*

by Yaser S. Abu-Mostafa

Why is an elephant big, dark and strangely shaped?" the question goes. "Because if it was small, white and round, it would be an aspirin." This answer may ring funny to human ears, but it could well prove informative to a computer trying to identify such objects as elephants or aspirin. Knowledge we commonly take for granted is not available to machines unless carefully spelled out. For machines, learning is not at all simple.

Despite the challenges, machine learning is one of the fastest-growing technologies today. The past few years have witnessed an explosion of applications, ranging from automated reading of handwritten zip codes at the post office to predicting seat demand in the airline industry. Indeed, the last time you received a credit card from a bank, chances are it was approved by a machine that learned on its own how to evaluate credit risk. And the future of machine learning is on the rise.

Designing a computer program to handle a particular job almost invariably demands a thorough understanding of

that task and its solution. Machine learning therefore has a fundamental appeal. Instead of devising a specialized program, one could merely provide training examples to a versatile machine that would learn on its own.

A self-learning credit-card approval system would, for instance, use historical data about "good" and "bad" customers to judge applicants. The machine does not care about the details of the problem. All it does is take matched pairs of inputs (in this case, personal information) and outputs (credit behavior) and absorb whatever information their relation contains. The trained machine then serves to evaluate new applicants. This kind of procedure takes automation one step further than normally envisaged. It not only applies a computer to a repetitive task, it automates the very problem of designing a system to perform that task.

One can, in principle, apply the methodology of machine learning to a wide array of problems. If, however, the input-output examples available lack vital information, the machine may fail to acquire proficiency. Fortunately, one can often append the needed information in the form of an intelligent hint. The hints used in machine learning range from simple observations to sophisticated knowledge.

In computer-vision applications, for instance, in which the goal is to recognize objects, there are many invariance hints. These assert that an object remains the same object when it shifts position in the range of view or changes in size. In financial-market applications, there are many monotonicity hints, which state that if an input consistently shifts in one sense or direction, the output must also consistently move just one way. Each particular application has its own hints that can aid the learning process.

If one knows enough about a given application to offer hints, why bother with machine learning in the first place? Why not employ this knowledge to design a specialized machine for the job? In some instances one can do so, but the fact of the matter is that usually too little is known about a problem to specify a method for its solution according to a well-defined set of rules.

Applications range between two extremes: structured problems that are totally defined and require no examples, and random problems that are completely undefined and depend entirely on training examples for their solution. Machine learning using intelligent hints is the way to handle the vast middle ground.

Machine-Learning Paradigm

How do machines learn? Many different models for machine learning have been devised. Typically the implementation used will have a general structure that is broadly tailored to the problem, but it will also have many free parameters—these might be thought of as the knobs and dials for tuning the machine. The values given to these adjustments determine how the machine will ultimately act; different settings will produce completely different results.

The behavior of a machine can be viewed mathematically as a function that associates input values (the specifics of a problem to be solved) with corresponding output values (the decision or action to be made). The goal in machine learning is to make the machine emulate the target function, the desired mapping of inputs to outputs. We can use training examples from the target function to guide the selection of values for the machine's free parameters. With each example, the machine refines its internal settings so that it matches the

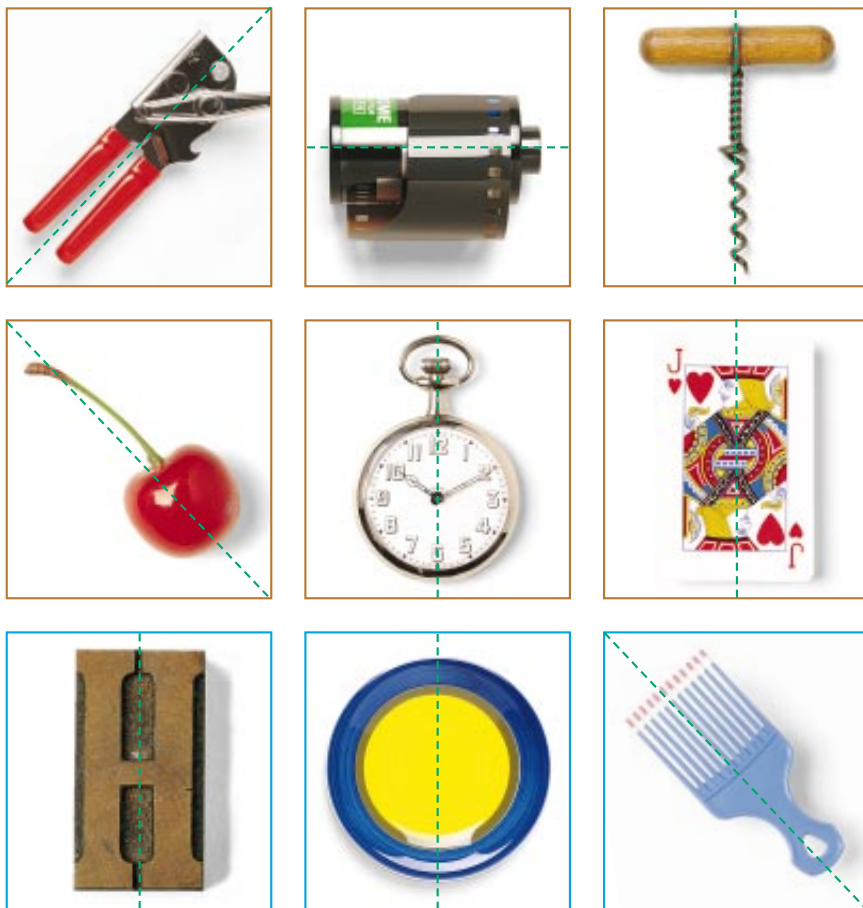
YASER S. ABU-MOSTAFA is professor of electrical engineering and computer science at the California Institute of Technology and chairman of NeuroDollars, a California-based corporation. He received a B.Sc. from Cairo University in 1979, an M.S.E.E. from the Georgia Institute of Technology in 1981 and a Ph.D. from Caltech in 1983, where he now heads the Learning Systems Group and serves as one of the principal investigators for the National Science Foundation Center for Neuromorphic Systems Engineering. Abu-Mostafa has been a technical consultant for Citibank since 1988. He has published widely in the areas of learning theory, neural networks, pattern recognition, information theory and computational complexity.



DAN WAGNER

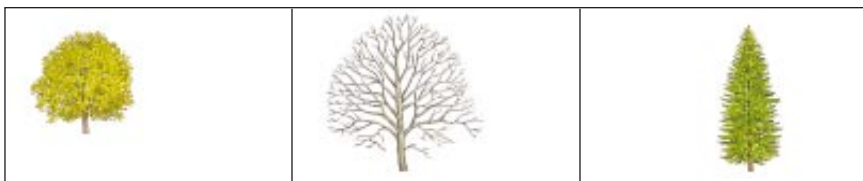
CAN YOU SOLVE IT? These objects have been sorted into two classes, indicated by either a blue or brown border. Which characteristic distinguishes them? Computers programmed

to learn from examples often face similar puzzles. Providing the machine with hints can make learning faster and easier. For a hint to help with this puzzle, turn the page.



VISUAL HINT aids both machines and people in solving the puzzle. Drawing the axis makes it clear that the top six objects lack the mirror-image symmetry exhibited by the bottom three. This characteristic distinguishes brown and blue categories.

RECOGNITION TASK



SHIFT IN VARIANCE



SCALE IN VARIANCE



INVARIANCE HINTS can help machines recognize that objects do not lose their identity when viewed in a new way. A machine attempting to identify trees, for instance, would not inherently know that size and position did not matter (*top*). Training on “virtual” examples of quite different subjects—such as a face or a chair—could prompt the machine to grasp these principles.

inputs and outputs appropriately. When the machine reaches a setting that corresponds as closely as possible to the target function, it will have in effect “learned” it. Machine learning is simply the search for the right positions for the knobs. Because the search is guided by the training examples, this paradigm is called, naturally enough, learning from examples.

The most widely applied form of such machine learning is the neural network [see “How Neural Networks Learn from Experience,” by Geoffrey E. Hinton; *SCIENTIFIC AMERICAN*, September 1992]. Neural networks were inspired by the power of real neurobiological systems. They consist of many computational elements interconnected in such a way that each element’s output reflects inputs from a number of other elements. The adjustable parameters of a neural network are called synaptic weights after their biological counterparts, the synapses that connect nerve cells in the brain. The flexibility of neural networks and the simplicity of their training have made them the machine-learning model of choice for the past 10 years; neural networks now find uses in a broad range of machine-learning applications. Although specialized electronic and even optical networks have been built [see “Optical Neural Computers,” by Yaser S. Abu-Mostafa and Demetri Psaltis; *SCIENTIFIC AMERICAN*, March 1987], in most cases, one implements a neural network simply as a program running on a personal computer or workstation.

With all the training required, we might imagine the need for tedious late-night sessions at the computer, supervising the machine as it learns. Fortunately, responsibility for finding the optimal adjustments usually falls on a learning algorithm, a method that reduces the process to a series of simple, repetitive steps that the computer can perform independently. One of the most common learning systems in use today is the back-propagation algorithm for training neural networks. This technique was popularized primarily by David E. Rumelhart while at the University of California at San Diego.

Back-propagation uses simple calculus to decide how to change the parameters of the neural network. It takes a training example—an input and its corresponding output—and makes small modifications to the network parameters to minimize the difference between the current response of the network and the target response. This step is repeated over and over, each time nudging the network a bit closer to the desired effect. After going through all the examples

DAN WAGNER

ROBERTO OSTI

several times, the network can replicate the target function reasonably well.

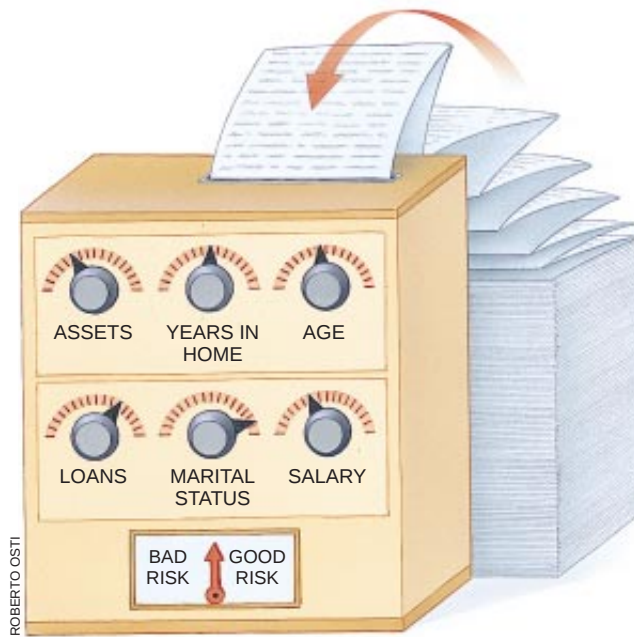
Let us see how this might work for credit-card approval. An input-output example in this case would be the data supplied by an applicant (such as age, salary and marital status) and her eventual credit performance (whether the bank profited or lost in extending credit). A neural network that approves credit cards would ideally predict a person's later behavior just by looking at the data provided by the individual. To learn this function, the network takes thousands of cases of real people's applications and credit behavior, and it keeps modifying its internal parameters in an attempt to match its output with these historical records.

Some of the variable parameters may be concerned with the salary and would move the network toward approving people who earn more. Others may be concerned with a combination of age or marital status, favoring particular combinations of these attributes. Eventually, a final setting is reached that makes the network produce the right response with as many training examples as possible. Now the network can be used to grade a fresh application by extrapolating from its "experience" to predict how the new customer will behave with credit.

Learning with Intelligent Hints

Whether implemented as a neural network or in some other way, all machine-learning methods share this same fundamental premise of learning from examples. For the machine to learn successfully, it must be able to generalize from the limited input-output samples on which it was trained. Do the training examples convey enough information for the machine to respond properly to novel inputs? Perhaps not. Because the machine does well on the training examples does not necessarily imply it will do equally well on something it has never before encountered.

Remember that a machine knows nothing about the function it is trying to learn except what it sees in the training data. If the data are deficient—there may be too few training examples or too much irrelevant information contained in them—the machine will not generalize properly. Or the examples



MACHINE LEARNING involves adjusting a system's internal parameters such that it makes the proper associations between data inputs and desired outputs. A credit-approval system, for example, would be trained to link applicants' personal data with their known credit behavior. In effect, the learning process "tunes the dials" until the machine can duplicate the input-output relations in the training examples.

may not encompass all the important information. Suppose, for instance, that I want to train a machine-vision system to recognize trees. I may not be able to specify how to identify a tree in exact mathematical terms, so I cannot structure the problem and give the machine rigid rules to apply. If I simply show the machine pictures of trees and objects that are not trees, I am giving it information, but I am still not telling it everything I know. For instance, I know that a tree remains a tree if it is shifted a little or rescaled. People intuitively realize that much, but the machine does not—unless specifically told so. Without hints, the machine might take a very long time, if ever, to reach that form of "understanding."

Even the simplest hint can boost learning. As in a game of 20 questions, in which the answers to some elementary questions can narrow the search significantly, a few hints may make the difference between learning a function and not learning it at all. To take advantage of this situation, I introduced a formalism called learning from hints some six years ago, and it has since become a feature of many learning systems. The most notable achievements of this approach are in automated trading systems for financial markets and in systems for handwritten character recognition.

The credit-card-application problem can also benefit from intelligent hints. Admittedly, it is difficult to define ex-

actly what makes a person a good credit risk, but one hint is obvious: if two people are identical except that one earns less money than the other, and the machine approves credit for the lower-paid person, it must also approve credit for the higher-paid one. This is one of many possible monotonicity hints. While the machine is learning, it should set its free parameters in such a way that it matches inputs and outputs according to the target function but simultaneously satisfies such hints.

One application of learning from hints that my colleagues and I conducted in our Learning Systems Group at the California Institute of Technology is in the area of foreign-exchange trading. We ran a machine-learning experiment to forecast the exchange rates for the U.S. dollar against four major foreign currencies. We wanted to test whether the

resulting trading system would be more profitable when we injected a common-sense hint into the learning process. The hint we used reflected a symmetry that was obvious to us: if a given pattern in the price history implies a certain movement (up or down) in U.S. dollars compared with a foreign currency, the foreign currency should move the same way if that pattern emerges in its own price history. The results of the experiment were quite successful. In all four markets, the symmetry hint brought the system a consistent increase in profit.

To assure that the improvement in our neural-network program came from information contained in the symmetry hint, we tried to fool the machine with two alternatives. The first was an uninformative hint, giving the machine random pieces of information. To our satisfaction, the machine did not benefit. Performance was about the same as when there was no hint at all. Next we fed the machine a hint that provided deliberately erroneous information. Performance then deteriorated rapidly, as would be expected. The intelligent hint had truly helped.

Implementing the Hints

The main challenge to using hints in machine learning is in automating the process. Hints come in various guises; they range in character from subtle to glaringly obvious. How can one algo-

rhythm orchestrate learning from all the varied pieces of information encompassed by such hints?

First, all the representations of hints must be standardized to enable the learning algorithm to deal with them on equal footing. The clue for the proper way to accomplish this end comes from the representation of the target function itself: as input-output examples. What one is telling the machine is, "When we input so and so, you should output such and such." The identity of the input-output examples completely distinguishes one target function from another. Similarly, if we could represent each of the desired hints by a set of examples, it would not matter what type of hint we wanted to introduce.

To represent the monotonicity hint in credit-card approval, an example might take the form of two persons with identical data except for salary. When the applications of both are presented to the machine, its responses may agree with the hint (by approving or denying credit to both or by approving only the higher-salaried applicant), or it may disagree (by approving only the lower-salaried person). The learning algorithm can adjust the machine's parameters to satisfy the hint, exactly as though it were incorporating one more example of the target function. Learning from hints in this way can therefore piggyback other learning mechanisms.

Remarkably, the examples representing the hint need not be real. The two applicants with different salaries could be hypothetical, or "virtual," cases. We can use virtual examples because we are

FOREIGN-EXCHANGE trading systems improve with symmetry hints. For instance, if the value of the dollar rises relative to the deutsche mark when a certain pattern appears in its price history, the mark should rise in the same way when its history shows that pattern. A neural-network program for foreign exchange that was provided with this hint performed more profitably than did machines supplied with either an uninformative hint, an erroneous hint or no hint.

not requiring the machine to make the correct decision about a real person but rather to act in a way that is consistent with the hint. This principle can also apply to the symmetry hint in foreign exchange; virtual examples can be constructed from price patterns that never occurred in history. For the task of computer vision, we can represent invariance hints using pictures of objects that are completely unrelated to the real target function. Our training for this hint does not require that the machine's output be right or wrong, only that it remain consistent as the input pattern shifts or changes in size.

Virtual examples are extremely important in many applications because they can add substantial information to what may be a meager set of training examples. In foreign-exchange forecasting, for instance, the real data are limited to a small set of recent price patterns. The ability to supplement this scant historical data with virtual examples is thus of great value. The remain-



JEAN MIELE The Stock Market

ing challenge is to find a learning algorithm that can achieve the proper balance between the hints and the real input-output examples.

The Balancing Act

A learning algorithm will strive to adjust the parameters of the machine to agree simultaneously with the entire training set and with all the examples of each hint. But a perfect solution is normally impossible, so some compro-



OBLIQUE VIEW

MIRROR IMAGE

FRONTAL VIEW

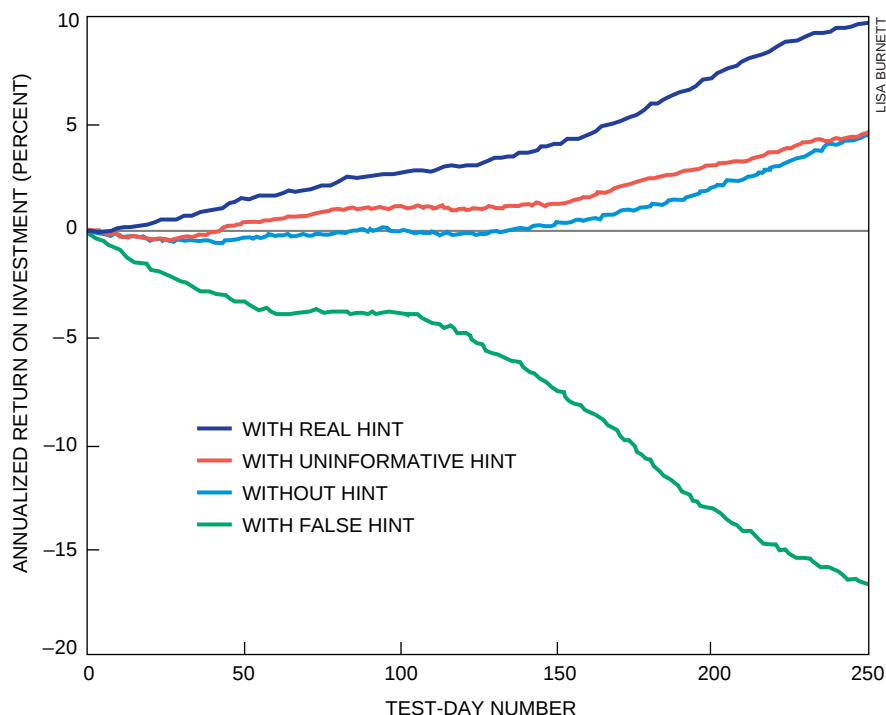
MIRROR IMAGE



DAN WAGNER

RECOGNIZING A FACE from any angle becomes easier with a symmetry hint. A naive machine, knowing nothing about human faces, would need many views to learn how to master this task (*top*). A hint that faces are symmetrical (*bottom*) al-

lows generalization from a smaller number of angles. Animals may rely on an innate understanding of such a hint: monkeys can be taught to recognize a face more reliably if they memorize it first from an oblique angle rather than a frontal view.



LISA BURNETT

mise must be found. To achieve this goal, the machine needs to assess how much agreement exists at each step. During the training process, some hints may be learned better than others. But if the computer can determine which hint is least well learned, it can pay more attention to it in the next iteration. This is the concept behind Adaptive Minimization, an algorithm I developed that schedules different hints for learning in a way that achieves a balance among them and with the training set.

The Adaptive Minimization algorithm is "adaptive" in the sense that it constantly evaluates how well the machine is satisfying both the hints and training examples, and it continually modifies the adjustable parameters. The term "minimization" reflects that the algorithm is trying to minimize a quantitative measure of the error between the current actions of the machine and the behavior ultimately desired for it.

Although I began work in this area with many of the basic ideas in mind, I have to admit that my first breakthrough came from necessity rather than spontaneous insight. I had been invited to present the hints framework at a scientific meeting, and only the day before did I find out that the seminar was scheduled to be 10 minutes longer than I had anticipated. The thought of facing my audience with an embarrassingly short lecture kept me up all night trying to see if I could expand on the algorithmic part of my talk. That evening the main idea of the Adaptive Minimization algorithm came to me; the next day I

delivered it in a well-received lecture. I doubt my audience appreciated just how recent those results were!

Learning Impediments

There are many challenges still facing the technology of machine learning. Perhaps the most severe faults stem from the tendency of machines to "overlearn" from training examples—something that can compromise a machine's ability to function correctly. Overlearning takes place when the machine memorizes the training examples at the expense of generalization. You might encounter a similarly ironic situation if you travel to Egypt and take a tour of the Pyramids. Some local guides provide an elaborate narrative for the tour, in English, and can answer common questions about the pharaohs perfectly. If you are encouraged by this performance and further query them, you will be surprised to find out that they do not speak English! They have memorized the necessary English sentences for a tour, but they have not "generalized" at all. Machine learning can suffer the same fate.

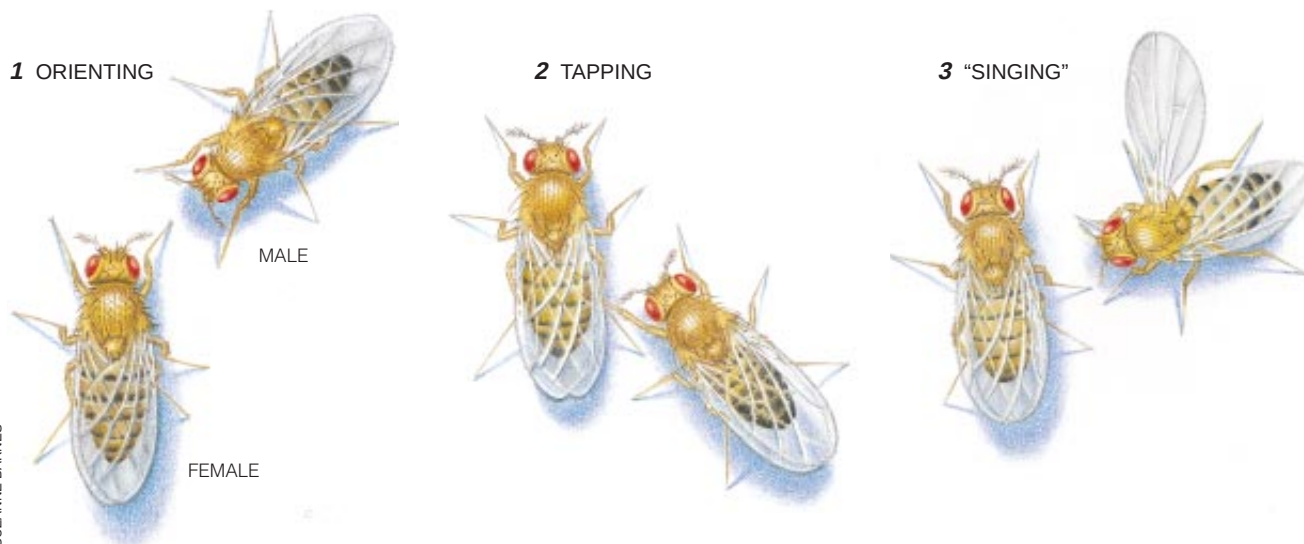
Another common pitfall in more complex machine-learning problems is a requirement for excessive computation time. As the learning algorithm searches for the optimal settings of free parameters (called the global optimum), it sometimes gets trapped in a poorer configuration (called a local optimum) that is better than similar solutions but still not the best that is theoretically

possible. There is no efficient way for avoiding local optima in general. Some learning tasks have been shown to be NP-complete, a technical term that characterizes a class of computational problems believed to require excessive amounts of computer time to find the global optimum. In practice, however, the problem has not been debilitating. Satisfactory performance usually requires only that the machine reach a good local optimum.

Despite the existence of such difficulties, machine learning has proved itself worthy in solving a wide array of real-world problems. It is a classical subject rooted in research carried out many decades ago, but it has been rejuvenated and expanded in recent years. With the addition of procedures for learning from hints and other technical advances yet to come, machine learning will undoubtedly continue to make its way into our daily lives.

FURTHER READING

LEARNING FROM HINTS IN NEURAL NETWORKS. Y. S. Abu-Mostafa in *Journal of Complexity*, Vol. 6, No. 2, pages 192–198; June 1990.
 ARTIFICIAL INTELLIGENCE. Special issue of *Communications of the ACM*, Vol. 37, No. 3; March 1994.
 LEARNING FROM HINTS. Y. S. Abu-Mostafa in *Journal of Complexity*, Vol. 10, No. 1, pages 165–178; March 1994.
 NEURAL NETWORKS IN THE CAPITAL MARKETS. Edited by Paul Refenes. John Wiley, 1995.



Understanding the Genetic Construction of Behavior

*Studies of courtship and mating in the fruit fly
offer a window on the ways genes
influence the execution of complex behaviors*

by Ralph J. Greenspan

Within the first 15 years of this century, the newborn science of genetics had begun to give people their first glimpse of how heredity might work. Studies of such traits as flower color in plants and wing shape in fruit flies had confirmed Gregor Mendel's once obscure 1865 proposal that physical characteristics are passed from parents to progeny by discrete units of inherited material, or genes (the name these mysterious units were given in 1911). As commonly happens when a new discipline experiences its first flush of success, scientists and others soon began to apply the understandings of the budding field more broadly, and sometimes less carefully, to explain other phenomena—notably, the behavior of human beings. Often they claimed that complex behaviors were directed by single genes.

Yet even careful researchers have failed to link specific human behaviors convincingly to solitary genes or to small sets of genes. The reason could lie with methodology. When it comes to human behavior, there is virtually no way to disentangle unequivocally the

influences of genes from those of culture and upbringing. On the other hand, if scientists could somehow manage to mask the effects of environment and focus solely on the genetic aspects of a behavior, they might still find the old assumptions flawed. Well-controlled investigations in simpler organisms suggest that a multitude of genes, some acting quite subtly, probably contribute to most behaviors.

Early Thinking on Humans

The question of whether human behavior is hereditary was initially asked more than a century ago. Francis Galton, a pioneer in the use of statistics, was among the first scientists to take up this issue. In the 1880s he analyzed various physical and behavioral traits in parents and their grown children. Using his newly invented "coefficient of correlation," he argued that behavioral traits are inherited. By comparing the distribution of traits in different generations, he concluded that each characteristic was the product of multiple donations from hereditary material.

A rather different view gained a following in the early 1900s on the heels of the rediscovery of Mendel's work, and it was embraced by such influential geneticists as Charles B. Davenport, a founder of Cold Spring Harbor Laboratory on Long Island. In the extreme, these researchers ascribed such ill-defined characteristics as musical ability, temperament or "feeble-mindedness" to individual genes. In 1921, for instance, Davenport asserted that "it appears probable, from extensive pedigrees that have been analyzed, that feeble-mindedness of the middle and higher grades is inherited as a simple recessive [trait], or approximately so." (In spite of their divergent views on the mechanisms of inheritance, both men regrettably drew similar, dangerous conclusions from their observations. Galton, who coined the term "eugenics," became a strong advocate of improving the human race by selective breeding between people having desirable traits. Davenport ardently supported that practice as well.)

Some of the first experiments designed to assess the impact of genes on behavior were carried out in the 1920s



on dogs. Those investigations examined, among other traits, pointing (indicating the location of prey) and vocalizing during hunting.

Dog breeds are as distinctive in their behavior as they are in appearance. The early studies crossbred dogs that differed in some behavioral characteristics and then mated their offspring to one another. If one or just a few genes controlled a chosen behavioral trait, investigators would expect to find that the animals of the final generation divided into discrete groups in which one group closely resembled the mother, a second closely resembled the father and perhaps one or a few groups behaved in an intermediate manner. If many genes were involved, workers would expect to find no discrete classes and a broad range of behavior in the offspring. The results were consistent with the last pattern, indicating that many genes underlay the appearance of each trait. Similar conclusions came from studies of maze running by laboratory rats.

Such analyses were informative but had major limitations. Breeding experiments cannot meaningfully address the genetic basis of behaviors that are relatively invariant in all members of a species. To delve into that problem and others, scientists needed ways to identify the specific genes involved in behaviors. Unfortunately, they would not have those techniques until many years had passed.

By the 1960s, however, many of the technical obstacles to the genetic dissection of behavior in animals had begun to fall. The structure of DNA had been deciphered in 1953. Studies of microorganisms had revealed that genes specify the makeup of proteins. When a gene is activated, it leads to the synthesis of the encoded protein. That protein, in turn, carries out some needed function in the body—such as helping to build and operate the nervous system (which itself ultimately shapes behavior). Such research had also clarified the steps by which genes give rise to proteins and had laid the foundation for development in the 1980s of many useful tools for isolating individual genes and deter-



MALE FRUIT FLY COURTS a female by executing a programmed sequence of steps. In the early stages he orients toward the female (1) and taps her abdomen with his foreleg (2). Next, he extends one wing and vibrates it to produce a “love song” (3). Then he licks the genitals of his partner (4), attempts to mount her (5) and, finally, mates with her (6). Analyses of this sequence suggest the genetic contributions to behavior are often surprisingly subtle.

mining the functions of their corresponding proteins.

Seymour Benzer of the California Institute of Technology was a leader in establishing that genes are linear segments of DNA. In the mid-1960s he also became one of the first investigators to go beyond linking genes to physical traits. Benzer began, in detailed studies of the fruit fly *Drosophila melanogaster*, to identify genes that affect behaviors. That effort is ongoing, particularly in the laboratory of Jeffrey C. Hall of Brandeis University, who was among the earliest researchers to work with Benzer in this new field. It also continues in my laboratory at New York University and elsewhere. I got involved in the mid-1970s, when I became Hall's first graduate student at Brandeis.

Spotlight on Fruit Flies

Among the behaviors receiving the most attention is the one the flies seem to do best: courting. This process consists of a series of actions, each of which is accompanied by the exchange of visual, auditory and chemosensory signals between males and females. The male is the more active of the dancers in this intricate ballet and has therefore been the focus of much of the research.

The ritual begins with a step called orientation. The male, who needs no instruction in this process, stands facing the female, about 0.2 millimeter away. Then he taps her on the abdomen with a foreleg and follows her if she moves away. Next, he displays one wing and flutters it to execute his form of a “love song.” Depending on the female's level



of interest at this point, he may go back and repeat his actions. But if all is going well, he unfurls his proboscis (a tubular appendage carrying the mouthparts at the tip) and licks the female's genitals. He may then mount her and, if she is receptive, copulate with her. Fruit flies will not mate unless the males have gone through this entire routine and the female has become receptive. Rape is uncommon in the fruit-fly world.

As a first step to finding the genes that might participate in courtship, Hall, initially working in Benzer's laboratory, set out to identify the parts of the central nervous system that control each element in the courting routine. He did so by producing extraordinary flies, called genetic mosaics, that carried mixtures of male and female cells.

The technique was based on an understanding of sexual development in fruit flies. In fly embryos, such development is controlled by the complement of X chromosomes within each cell. Cells that have one X chromosome give rise to male anatomical structures and behaviors in the fully formed fly; cells that have two X chromosomes lead to female anatomy and behavior. These differences arise because single-X (male) and double-X (female) cells activate separate, albeit overlapping, sets of so-called sex-determining genes. Hall knew that if a fly carried mainly female cells but harbored male cells in a particular site of the brain, any typically male

RALPH J. GREENSPAN, who earned his Ph.D. in biology from Brandeis University in 1979, is professor of biology and neural science at New York University and head of the W. M. Keck Laboratory of Molecular Neurobiology there. Outside the laboratory and N.Y.U.'s classrooms, Greenspan pursues an interest in making science accessible to the lay public. He teaches adult education courses at the New School for Social Research in New York City and helps to organize symposia for the Thrivers' Network of the Cancer Center at the University of California, San Diego, a group of cancer survivors and their families.

Early Views of Human Behavior

In the first decades of this century the opinions of two men represented opposite poles of thinking on the question of whether single genes or many lie at the root of any given behavior. Beginning in the late 1800s Francis Galton, a pioneering statistician, argued that human traits, including behaviors, are controlled by a multitude of the hereditary units that later came to be called genes. Charles B. Davenport, a respected geneticist, subsequently asserted that single genes were in control. Studies of fruit flies and other animals indicate Galton's view was probably the more accurate of the two.



Charles B. Davenport

If the names of Galton and Davenport are familiar, it is because both are now notorious for their advocacy of eugenics, which in Galton's words involves checking "the birth-rate of the Unfit" and improving the human race by "furthering the productivity of the Fit." Galton introduced the term in the 1880s, and Davenport, who established a research center in human eugenics at Cold Spring Harbor Laboratory on Long Island, pushed the program forward. Davenport is sitting at the right on the bottom step in the photograph below, which depicts a class of students he trained to carry out eugenics research.



Francis Galton



Eugenics training class of 1914

courtship activities it displayed could be attributed to a male pattern of gene expression, or activation, in that site.

Once mosaics were produced, he monitored the animals' attempts at courtship. Then he froze the flies and painstakingly sliced the diminutive creatures (measuring just 1.5 millimeters long) into 20 thin sections, noting (with the help of a clever colorization technique) the distribution of male and female cells. The experiments were particularly nerve-racking in the 1970s because the method for creating mosaics had an inconvenient drawback: no two individuals ended up with exactly the same clusters of male and female cells. Each fly had to survive a battery of behav-

ioral tests, and all 20 sections had to be analyzable. The uniqueness of each animal meant that the experimenter had no second chance.

After examining many of these mosaics, Hall concluded that initiation of courtship (orienting toward the female, tapping her abdomen, following, and extending a wing) required male cells in one side or the other of a relatively small region near the top and toward the back of the fly's brain. This region integrates signals from the fly's various sensory systems. In other words, male cells at that site somehow give rise to a trigger mechanism for courtship that is present in males but not in females. Later steps in courtship, especially those

demanding precise motor coordination, require male tissue in additional parts of the nervous system. To perform a proper courtship song, for example, flies must have male cells in the "trigger" region as well as in parts of the thoracic ganglion, which is the fly's version of a spinal cord.

More recently my colleagues and I have also identified the region of the brain involved in determining sexual preference in fruit flies. We did so almost inadvertently, after Jean-François Ferveur in my laboratory (now at the University of Paris in Orsay) created entire strains of mosaic flies that were mainly male but that had female cells in selected areas of the brain. Before studying the courting behavior of these insects, we wanted to see whether full-fledged males would mistake our mosaics for females. The mosaics were not perceived to be female. To our surprise, however, a few strains displayed an odd behavior of their own: they courted males as vigorously as they courted females.

Examination of the brains in these unusual insects, undertaken in conjunction with Klemens F. Störtkuhl and Reinhard F. Stocker of the University of Fribourg, revealed that sexual discernment was altered when either of two parts of the central nervous system was female: the antennal lobe or the mushroom body of the brain. Both these regions, the second of which lies close to the trigger site for courtship, participate in processing olfactory signals. If either or both of these centers for analyzing odors were female, the fly lost the ability to distinguish males from females and became equally interested in both.

Genetic Influences on Courtship

The discovery that so many different regions of the central nervous system are involved in male courtship suggests that a variety of genes also participate in the process. Indeed, more than a dozen have been discovered, mainly by Hall and his colleagues. For instance, the *fruitless* gene influences sexual preference. Mutation in this gene affects male flies in much the same way as having female cells in the antennal lobe or mushroom body does: it causes males to court other males as avidly as they court females. The gene is also needed in the late stages of courtship; males carrying a mutant gene never attempt to copulate with females.

Hence, the picture beginning to emerge is more consistent with Galton's view than with Davenport's. Oddly enough, no one has yet identified any gene involved in courtship that is dedi-

cated solely to that behavior. Growing evidence suggests an explanation neither Galton nor Davenport would have predicted. It may be that most genes underlying courtship (and other behaviors) serve more than one function in the body. Identical genes may also be used for somewhat different purposes in males and females.

Consider, for example, one of the three genes known to participate in the male's courtship song. It is called *period* and has been studied most extensively by Hall and Charalambos P. Kyriacou of the University of Leicester.

Hall and Kyriacou decided to examine *period* when they discovered, in 1980, that the male song has a distinct rhythm to it. They already knew, from research done by Benzer's graduate student Ronald J. Konopka, that the gene affects the fly's circadian rhythms—the timed cycles, such as for waking and sleeping, that are characteristic of all living things. This property led them to wonder whether *period* might also affect the rhythm of the courtship song.

The song performed by fluttering the wing is not very musical to our ears, but it does have a detectable pattern. As the insect raises and lowers its wing once, the up-and-down motion produces a characteristic sound, or pulse, that can be picked up with a recording de-

vice. For approximately 27 seconds, the male gradually increases the interval between each successive pulse. Then, over another 27 or 28 seconds, he gradually decreases the interval, so that a plot of the intervals over time yields a smooth sinusoidal curve.

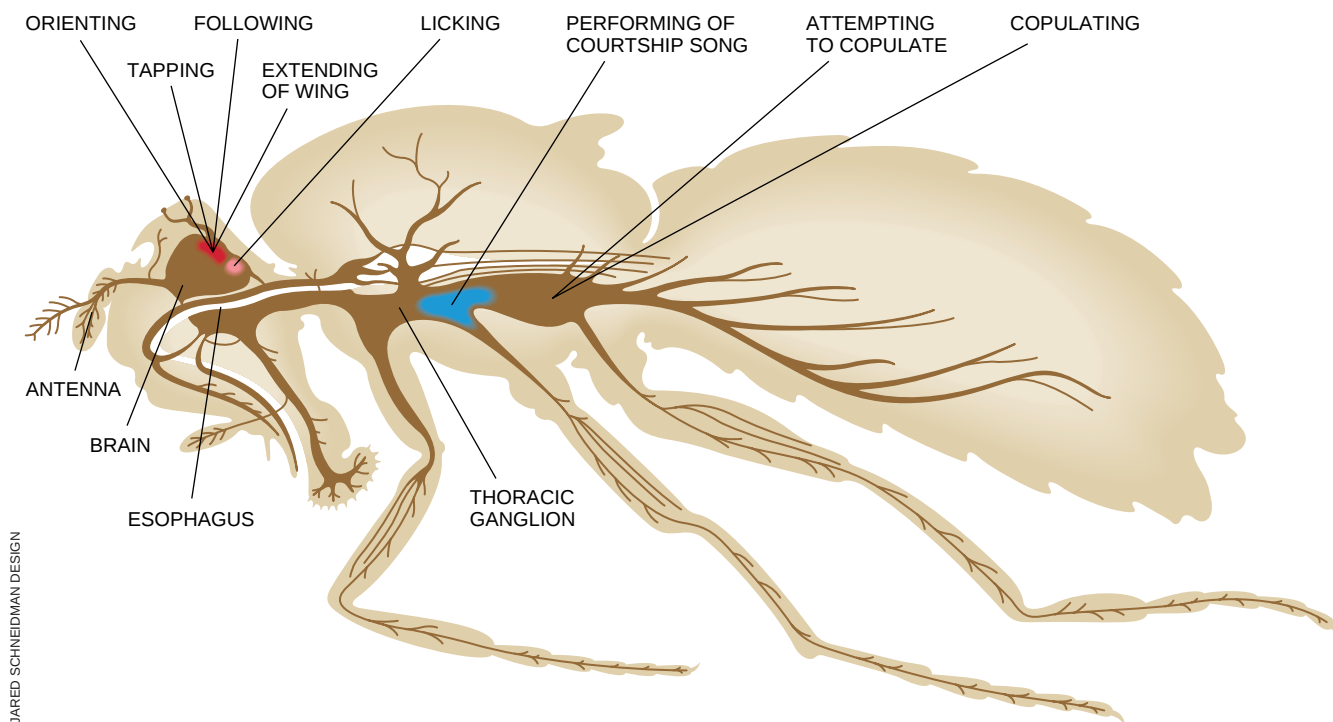
Hall and Kyriacou found that males carrying a normal *period* gene produce a normal song that makes females more receptive to their advances. In contrast, males carrying an inactive gene generate a song that lacks the usual smooth rhythm and is apparently less effective at stimulating females: when computer-generated simulations of normal and rhythmless songs were played for lone females who were then paired with a male, the females exposed to the aberrant song proved less receptive to the male's advances. Similarly, less drastic mutations in the gene allow rhythmicity to be retained but stretch or shrink the sine curve; in the process, they reduce the song's power over a female.

The subtlety of *period*'s effects on the overall courtship routine, and on the song itself, adds credence to the notion that courtship—and other complex behavior—is regulated by multiple genes acting together. And the fact that the *period* gene participates in setting other clocks, and is also expressed in many parts of the central nervous system,

supports the idea that any given gene may affect more than one behavior.

In a fascinating turn of events, Hall, Kyriacou and Michael Rosbash, also at Brandeis, have recently pinpointed the exact part of the gene that controls song rhythm. A small region in the middle is devoted to the song, and the balance of the gene controls other rhythms. That division of labor was deduced in part from the fact that a different species of fruit fly, *D. simulans*, has the same 24-hour cycle of activity and rest as is found in *D. melanogaster* but performs a song that differs in the intervals between pulses. The *period* gene in both species is similar, except for small differences in the middle region. What is more, genetically engineered flies that carry a hybrid *period* gene made by replacing the middle region of the *D. melanogaster* gene with the corresponding segment of *D. simulans* will "sing" just like *D. simulans*.

Although sexual preference and courtship behavior are certainly programmed in fruit flies, males and females have the ability to modulate their activity in response to one another's reactions. In other words, they can learn. Just as the ability to carry out courtship is directed by genes, so, too, is the ability to learn during the experience. Studies of this phenomenon lend further



SITES IN THE CENTRAL NERVOUS SYSTEM (brown) that control the steps of courtship in male fruit flies have been mapped by studying flies that consist of a mixture of genetically male and female cells. To perform the initial steps of the mating routine (orienting, tapping and wing extension) and

to follow peripatetic females, flies must have male cells in a small trigger zone (red) at the back of the brain. They also need male cells nearby (pink) to perform licking, in part of the thoracic ganglion (blue) to produce their song and in many different sections of the thoracic ganglion to copulate.

support to the likelihood that behavior is regulated by a myriad of interacting genes, each of which handles diverse responsibilities in the body.

Learning from Experience

One thing a male can learn during courtship is not to waste time on a female who has already mated and who, consequently, will not be receptive. As Hall and Richard W. Siegel of the University of California at Los Angeles found, male flies will court virgin females tirelessly but will lose interest in mated females after about 30 minutes or an hour—when they finally become impressed by the inhibitory pheromone, or scent, emitted by mothers-to-be. Once males give up the chase, they become uninterested in all females, virgin or not, for a few hours. If there is discernible evolutionary logic to this behavior, it may be that the presence of a mated female in a group of females is a sign that most or all of them have already mated; hence, a male's efforts would be better spent elsewhere.

My explorations of the genetic underpinnings of this response began a few years ago and were undertaken with Leslie C. Griffith, who is now at Bran-

deis. We knew from the work of other investigators that an enzyme called calcium/calmodulin-dependent protein kinase II (CaMKII) can help record the effects of experience in neurons, in the process inducing molecular changes that are likely to be essential to learning. We therefore decided to see whether male fruit flies needed this protein—and thus the corresponding gene—in order to respond appropriately to mated females.

As a first step, Griffith engineered a strain of flies whose CaMKII protein could be quieted simply by increasing the body temperature of the flies. Sure enough, when the enzyme's activity was reduced even mildly, males of this strain behaved oddly. They were as avid as normal flies in their courtship of virgin females, and they lost interest in mated females after the usual hour or so, but they seemed to forget their rejection almost immediately. If they were placed with females soon after being with a mated one, they began their pursuit anew. When CaMKII was inhibited even more, the males did not learn at all: they pursued mated females unabated for hours. (Even in the world of fruit flies, it seems, some men never learn.)

Once we knew that the *CaMKII* gene, through its enzyme product, did participate in learning during courtship, we naturally began to wonder how the enzyme itself helped to record experience. All protein kinases act by phosphorylating, or adding phosphate groups to, other molecules, thereby activating or

inactivating the targets. But what was the kinase's target in neurons, and what happened after that target was phosphorylated? Such questions ultimately led us to demonstrate that a second gene expressed in neurons—*eag*—is instrumental in such learning as well.

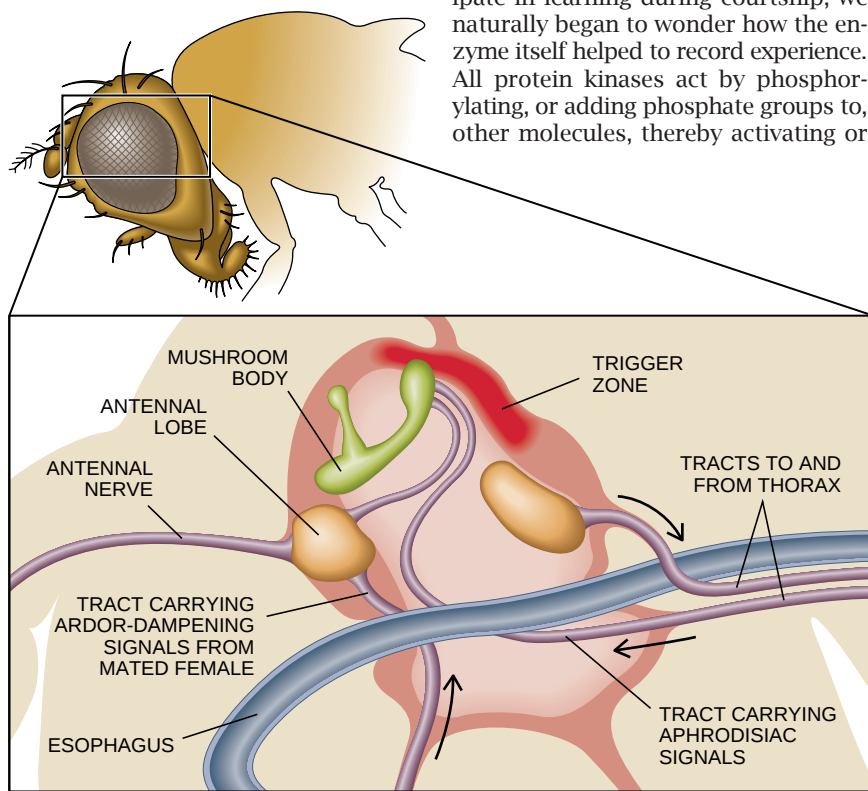
Another Learning Gene

The protein product of this gene is a component of certain membrane-spanning channels that regulate the flow of potassium ions out of neurons. Opening these channels helps to control excitability and the release of neurotransmitters, which carry messages from cell to cell. (The name derives from the fact, discovered in the 1960s, that when flies bearing mutated *eag* genes are anesthetized, their legs shake: in a sign of the times, the discoverers called the gene *ether-a-go-go*.)

On the basis of a number of clues from our own research, from that of our collaborators Jing Wang and Chun-Fang Wu of the University of Iowa and from others, Griffith and I began to think the CaMKII enzyme might participate in learning by modifying the EAG protein in potassium channels. For instance, Eric R. Kandel and his colleagues at Columbia University had shown in the marine mollusk *Aplysia* that one kind of potassium channel is modified by a kinase during a simple form of learning. Moreover, we found that mutations of the EAG protein led to essentially the same "thickheadedness" in courting males as did inhibiting the activity of CaMKII. Such discoveries implied the two proteins might operate in the same cascade of molecular interactions and that CaMKII might act on the EAG protein.

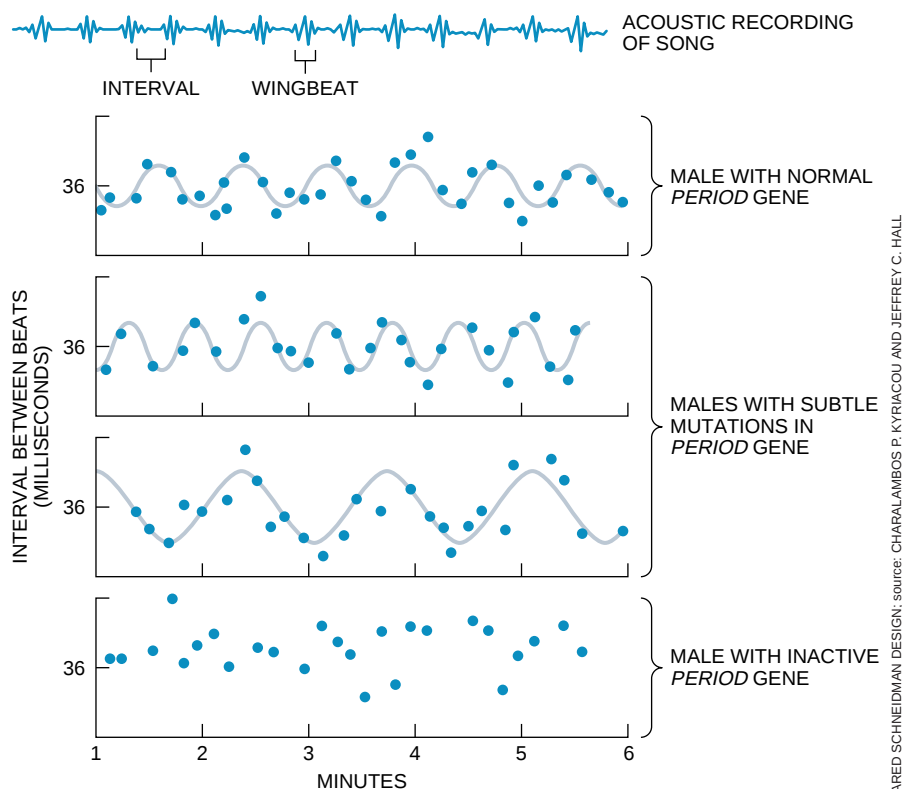
To our delight, Griffith confirmed that the enzyme could indeed modify the EAG protein, at least in the test tube. Based on these findings and on electrical recordings from synapses in mutant animals, we now speculate that males learn to give up on mated females through activation of the following sequence of molecular reactions.

First, exposure of males to antiaphrodisiac pheromones while they are courting mated females stimulates sensory systems that feed into the trigger region of the brain. As a result of this stimulation, calcium builds up locally in cells that normally promote sexual arousal during courtship. The buildup activates CaMKII, which phosphorylates the EAG protein in potassium channels carrying that protein. Such modification causes the channels to open, allowing potassium ions to flow out of the neurons, thereby quieting the cells and reducing their ability to release neurotransmit-



SEAT OF ATTRACTION to females resides in two sites in the male fruit fly's brain (shown schematically). One is the antennal lobe (gold, left); the other is the mushroom body (green), which lies close to the trigger zone responsible for initiating courtship. The importance of these regions was discovered when males engineered to contain female cells in either site began courting males as well as females.

JARED SCHNEIDMAN DESIGN



MALE FLY EXTENDS A WING (*photographic sequence at right*) before vibrating it to produce his song. In the normal song the interval between each beat (*highlighted in an acoustic recording at top*) increases gradually over about 27 seconds and then decreases equally gradually; a plot of the intervals resembles a sine curve (*top curve*). This rhythmicity has been shown to be controlled by a gene called *period*. Flies carrying a healthy *period* gene display the usual song. But those with mildly defective genes produce abnormal rhythms (*middle curves*), and those harboring an inactive gene completely lose their ability to “carry a tune” (*bottom*).



NANCI KANE, SUSAN BROUGHTON AND KAREN RAPHAEL, New York University

ters. As the cells become silent, the males lose interest in mating. Conversely, flies carrying defects in the genes for either protein presumably retain misplaced interest in mated females because the potassium channels remain closed in the critical cells, allowing the neurons to become hyperactive.

The *CaMKII* and *eag* genes turn out to be just two of several known to affect learning and memory in fruit flies. Some of the others also participate in courtship—a finding that meshes rather well with the view that behaviors arise from the interactions of vast networks of genes, most of which take part in many different aspects of an organism's biology.

Lessons for Humans?

Do the lessons from genetic studies of fruit-fly behavior bear any relevance to human beings? I think they do—within limits. There is every reason to believe that the genetic influences on behavior will be at least as complicated in people as they are in fruit flies. Hence, the notion of many, multipurpose genes

making small contributions is likely to apply. And many of the gene products that function in the brains of flies will probably turn out to be important in the human brain. Human counterparts have already been discovered for a number of genes originally identified in the fly, such as *eag*. These findings should provide insights into the molecular interactions that enable the central nervous system to produce behavior.

New technologies hold promise for detecting the contributions of individual genes to human attributes. The techniques are already being applied to a variety of complex traits, including mu-

sical ability—though more carefully than in Davenport's day. Such work, and extrapolations from animal research, can probably help pinpoint some of the genes that contribute to specific human behaviors. But any research claiming to explain human activity in purely genetic terms must be viewed with caution. Society's well-founded unwillingness to rear human subjects in perfectly controlled environments makes it virtually impossible to prove the validity of such claims.

FURTHER READING

INHIBITION OF CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE IN *DROSOPHILA* DISRUPTS BEHAVIORAL PLASTICITY. L. C. Griffith et al. in *Neuron*, Vol. 10, No. 3, pages 501-509; March 1993.
THE MATING OF A FLY. J. C. Hall in *Science*, Vol. 264, pages 1702-1714; March 25, 1994.
CALCIUM/CALMODULIN-DEPENDENT PROTEIN KINASE II AND POTASSIUM CHANNEL

SUBUNIT EAG SIMILARLY AFFECT PLASTICITY IN *DROSOPHILA*. L. C. Griffith et al. in *Proceedings of the National Academy of Sciences*, Vol. 91, No. 21, pages 10044-10048; October 11, 1994.
GENETIC FEMINIZATION OF BRAIN STRUCTURES AND CHANGED SEXUAL ORIENTATION IN MALE *DROSOPHILA MELANOGASTER*. J.-F. Ferveur et al. in *Science*, Vol. 267, pages 902-905; February 10, 1995.

The Art Historian's Computer

*Riddles posed by ancient works of art fall
to historical analyses and electronic explorations*

by Lillian Schwartz

Eight years ago my colleague Gerard Holzmann and I were testing a program that can flip, scale and juxtapose pictures. I have always had a strong interest in comparing faces and sometimes combine portraits to create different effects. I had previously scanned in many images, including Leonardo da Vinci's *Mona Lisa* and his self-portrait. Both faces are three-quarter views but turned in opposite directions. We flipped the self-portrait and scaled the images by giving them the same distance between the centers of the pupils. We then bisected the portraits and aligned the tips of the noses before juxtaposing the two halves.

As the composite image scrolled down the screen, the hairlines appeared to merge. Next, the foreheads and strong brows came into view, bearing a remarkable resemblance. Then the eyes came on, the pupils aligning at the same level. Finally, the cheekbones, noses and mouths matched precisely. I gasped. "Something wrong with my program?" Gerard asked.

Other details confirmed the similarity between the portraits. The distances between the inner corners of the eyes, one of the most individual characteristics of a face, matched to within 2 percent. The protruding brow, or supraorbital ridge—found in more than 90 percent of males—is quite distinct in Leonardo's portrait as well as in the *Mona Lisa*. (Neither of these features matches those in an earlier drawing, claimed to be a preparatory cartoon for the *Mona Lisa*. The cartoon, however, bears a resem-

blance to yet another sketch, revealed by x-ray analysis to lie underneath the *Mona Lisa*.)

Near the end of the 15th century Leonardo left Milan with his initial sketch for the *Mona Lisa* and continued to work on it for years. The painting was with him when he died in 1519. I suggest that in completing the work in the absence of his sitter, Leonardo used himself as a model and infused the portrait with his own features. He used his new shading technique, called *sfumato*, to blend and replace parts of the first model's face with his features.

With his love of riddles, Leonardo offered what may be one of the most important clues of all. It appears in the endlessly knotted cloverleaf patterns on the bodice of the *Mona Lisa*'s dress. As the art historian Roy McMullen said, "We can be quite sure that they are not there by accident." Leonardo was fascinated with knots and once adapted such a basketry design

to embellish his own name. The Italian word for the osier branches used for basketry is *vinco*. The connection between *vinci* and *vinco* reveals the identity of the model for the *Mona Lisa*—Leonardo himself.

A few years later another puzzle posed by Leonardo's works caught my attention. Scholars have long debated where one should stand to view the *Last Supper*, painted on one wall of the refectory of Santa Maria delle Grazie in Milan. Leonardo had painted tapestries in the fresco and bands of designs on the adjoining walls of the refectory. The tapestries and designs do not line up—unless one views the painting from 15 feet



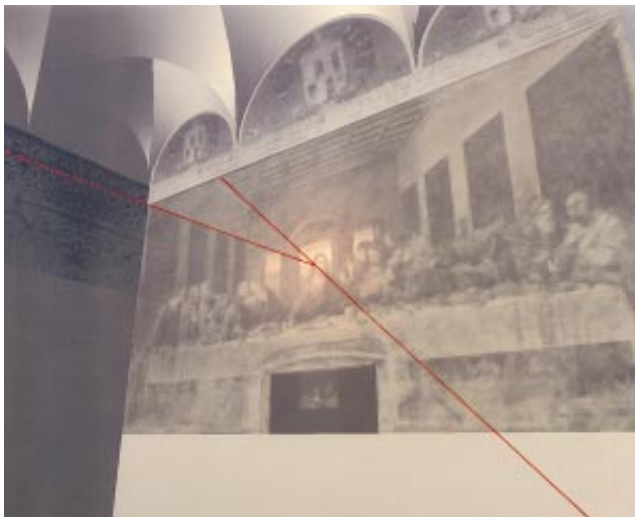
JUXTAPOSED IMAGES of the *Mona Lisa* and Leonardo da Vinci (opposite page) reveal the unity of a single face. When Leonardo's self-portrait is flipped and scaled, the lips, brows, cheekbones and shape of the head match those of the *Mona Lisa*. The similarities suggest that Leonardo, who worked on the *Mona Lisa* in the absence of his original sitter, used himself as a model and infused the painting with his own features. The original sitter may have been Isabella, Duchess of Aragon. Leonardo had made a cartoon, or preparatory drawing, of her in the same composition as the *Mona Lisa*. (The Renaissance historian Vasari, on the basis of

rumors, assumed that the model was Mona Lisa Gherardini; hence the painting's name.) The cartoon, however, does not match up with the finished painting, as the composite above shows. Despite a superficial similarity, the eyes, nose and jaws are different. The cartoon of Isabella does match a sketch that x-ray analysis has revealed to lie underneath the *Mona Lisa*. In a computerized transmutation (left) known as a morph, from "metamorphosis," a forbidding Leonardo transforms into a smiling *Mona Lisa* when age lines and beard are subtracted and the corners of his lips are turned up. The strong brows and positions of the eyes do not change.





LILLIAN SCHWARTZ © 1987 Computer Creations Corp.



LILLIAN SCHWARTZ © 1987 Computer Creations Corp.



SCALA/ART RESOURCE

LAST SUPPER (*bottom right*) was painted by Leonardo da Vinci in the refectory of Milan's Santa Maria delle Grazie. The 15th-century fresco shows Christ dining in a room that appears as an extension of the refectory (*computer re-creation, top*). Leonardo drew designs on the long walls of the room to line up with the tops of the tapestries in the fresco to create a powerful illusion—but only when viewed from a point 15 feet above the

floor. Thus, visitors do not perceive the illusion. But the computer re-creation (*bottom left*) reveals that anyone who had a view from a side wall or entered through the original door (*visible in top illustration*), which is now walled up, would perceive the supper as taking place in the refectory: the observer's eye is drawn to Christ's hand, to Christ's head, along the tops of the tapestries and finally out into the real room.

above the floor. It would seem that Leonardo intended the fresco to appear as an extension of the refectory. But the illusion does not work.

In the computer, I constructed a three-dimensional room based on the one in the fresco, in order to view it from different locations. Instead of true linear perspective, Leonardo used an accelerated perspective common to theater settings. In creating the painting, as if setting a stage, Leonardo placed Christ and the apostles up front, tilted the floor and table, designed side walls of uneven lengths and tapestries of different sizes and spacing.

There are, I found, other positions in the refectory from which the lines of the designs continue smoothly into the lines of the tapestries. To one side of the fresco and 18 feet in front of it had once been the main door to the refectory, now walled up. Anyone entering the room and getting a first glimpse of the painting would perceive the supper as taking place in the refectory, with Christ's extended hand welcoming him in. Viewing the painting from the sides, clearly what Leonardo intended, also enabled monks seated at their dinner tables, which ran along the lengths of the refectory's side walls, to feel they were partaking of their meals in the company of Christ. In the computer re-creation, I was also able to show Christ's feet (destroyed by a door later cut through the fresco but evident from an early copy), which appear suspended over the floor, the form of his figure foreboding his crucifixion.

Recently the scholar Patricia Trutty-Cooill of Western Kentucky University has drawn me back to Leonardo's work, this time to his 700 grotesques. Did he conjure them up or sketch them from life? These bizarre faces confuse many art histo-

LILLIAN SCHWARTZ (self-portrait at right) is a pioneer in computer art and computer graphics. Her work is in major art collections around the world, including the Museum of Modern Art and the Metropolitan Museum of Art in New York City and the Georges Pompidou Center in Paris. A Fellow of the World Academy of Art and Science, Schwartz wrote, with Laurens R. Schwartz, *The Computer Artist's Handbook* (W. W. Norton, 1992). For many years she has been a consultant in computer graphics at AT&T Bell Labs.

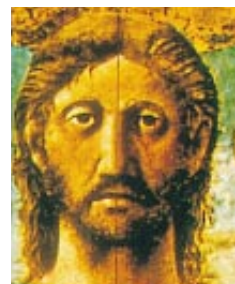


LILLIAN SCHWARTZ © 1992 Computer Creations Corp.

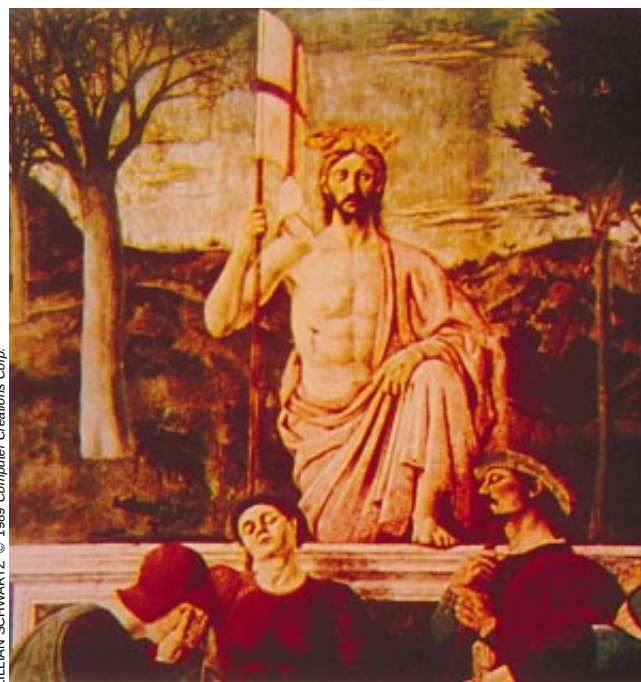
rians because they appear very far from the canon of proportions that Leonardo had established for drawing faces.

Measuring the features and their alignments, I was surprised to find that the grotesques are actually "normal." Leonardo had exercised the only leeway that his rules allowed: to exaggerate chins, noses, mouths and foreheads. As he stated, "the distance from the eyebrow to where the lip joins the chin ... to the (posterior) angle of the jaw ... to the upper limit of the ear ... form a perfect square the sides of which measure half the head." The grotesques seem so lifelike because they have "correct" proportions. Leonardo used a common "armature," as Kenneth Clark suggested.

THE RESURRECTION OF CHRIST, a fresco painted by Piero della Francesca in the 15th century, was plastered over sometime during the 17th century. Although some of the plaster has since been removed, the fresco still looks faded (*bottom left*). Electronically subtracting the white coating, the author found that the fresco was painted in warm colors (*bottom right*). When the fresco was magnified and brightened, she also discovered a symbolic tree of thorns (*right*) and what appears to be a cauliflower ear on Christ (*far right*), perhaps carried over from the model Piero used.



MUSEO DI CIVICO SANSEPOLCRO, Italy



LILLIAN SCHWARTZ © 1989 Computer Creations Corp.



Intrigued by the analytical possibilities of the computer, the late art historian Eugenio Battisti proposed that I study the color of the frescoes of another Italian artist, Piero della Francesca. Piero's *The Resurrection of Christ*, in Sansepolcro, had once been plastered over. Part of the coating was since removed, but was the restoration complete? Restorers needed to know what the original colors might have been before proceeding further.

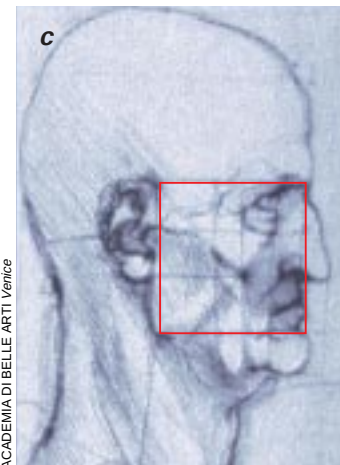
To study this fresco, I selected a high-performance system that interprets the hue, brightness and saturation of a color,

one pixel at a time. Electronically subtracting the white of the plaster coating, I found that *The Resurrection* is painted in sunset colors. Battisti pointed out, in addition, that Piero most likely took advantage of the afternoon sun, pouring in through a window high up on the west wall in the original architecture, to accent this palette. The on-screen enhancement also revealed for the first time a symbolic tree of thorns on one side of the fresco, and one curiosity—Christ has a cauliflower ear.

Not far from Italy, yet another mystery had been brewing



LEONARDO'S GROTESQUES (two are shown, *a* and *b*), bizarre as they are, have the same proportions that the painter laid down for drawing a normal face (*c*). The brow lines up with the top of the ear, the ear with the posterior corner of the jaw, the corner of the jaw with the lower edge of the lip, and the lip edge with the front of the brow, forming a square (*red line*). To draw his grotesques, Leonardo exaggerated the mouth, nose or chin. If the profiles are scaled and a feature is cut out along the prescribed lines, it fits perfectly onto any of the other faces. Thus, one grotesque acquires an eye from the normal face (*d*) and a normal mouth and chin (*e*). Or this grotesque lends his chin to a second grotesque (*f*). These pastiches also lead to lifelike faces.

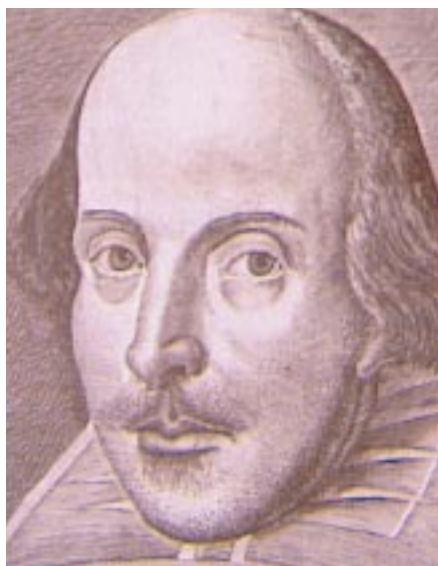


d



LILLIAN SCHWARTZ © 1992 Computer Creations Corp.

e



THE BRITISH LIBRARY

QUEEN ELIZABETH'S portrait by George Gower (a) looks, at first sight, not much like that of Shakespeare (e) by Martin Droeshout. But many features of the faces match. In these images, sections of Shakespeare's portrait have been overlaid with the queen's. In (b), the outlines of the right side of the face merge and the chins match up, but the width of the eye is diminished. As (c) shows, when a fragment on the forehead is aligned on the right, the hairline to the left matches. A second line (known as the mystery line) is etched curving up from the queen's chin. Because both jawlines were then visible, Droeshout shaded the jaw and the adjacent area; he then raised the hairline from the brow and put the queen's forehead in place. The effect is the strange bulbous forehead we see on Droeshout's drawing. In (d), the queen's forehead continues into Shakespeare's; the right jaws line up; the left eye falls into place. These and other details suggest Shakespeare's face was traced from a pattern of the queen's face.

for centuries. In 1990 I started to compare portraits of Shakespeare with those of his contemporaries, in the hope of establishing the bard's identity. Soon afterward I was invited to England to continue this work. Some scholars consider the most authentic of Shakespeare's portraits to be one that accompanies the First Folio edition of his plays, published in 1623. That engraving, by a young artist named Martin Droeshout, was denounced a century and a half later by Thomas Gainsborough, who declared that he "never saw a stupider face. It is impossible that such a mind and such a rare talent should shine with such a face." Indeed, the bulbous forehead on this portrait is anatomically absurd. The double line at the chin, others argued, represents a mask.

Digitally comparing the features of this drawing with those of the Earl of Oxford and many other notables, I found myself reaching a stalemate. Nothing seemed to match. Feeling that I had disappointed my hosts, on the last day of my stay I returned to the National Portrait Gallery. There, gazing out from a portrait of Queen Elizabeth I, I saw the very same eyes I had been scrutinizing for weeks.

A detailed comparison on the computer revealed that most of the lines in the Droeshout engraving and the queen's portrait by George Gower are the same. The eyes, noses and curvature of the faces match perfectly. But there is an odd shift in some features. I have a theory for the similarity and for the shift. In having to invent a face for Shakespeare, Droeshout

must have based his engraving on a cartoon of the queen's face. Artists frequently resort to such expediciencies. And authorized sketches of the queen, on which all her portraits were required to be based, were widely available.

Etching metal is laborious work; perhaps, having finished the left part of the face, Droeshout came back to the engraving sometime later and inadvertently (or deliberately) shifted the cartoon. The outline to the right, though following the contour of Elizabeth's face, became displaced. The head became elongated and bulbous, causing the jaw to appear too wide for the now more narrow face, an error that Droeshout tried to correct. Hence, the double jaw lines.

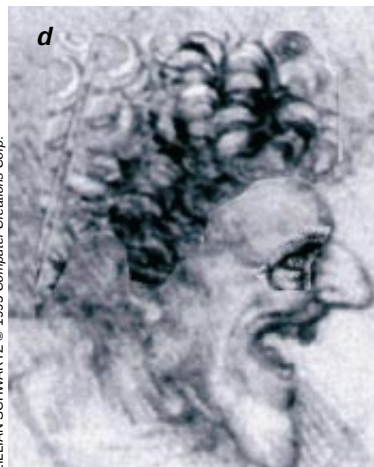
Although the face on Shakespeare's plays appears to be based on the queen's cartoon, lively debates continue as to who wrote Shakespeare's works. Even as the computer aids in solving riddles, it continues to provoke them.

f



LILLIAN SCHWARTZ © 1995 Computer Creations Corp.

d



LILLIAN SCHWARTZ © 1995 Computer Creations Corp.

e



LILLIAN SCHWARTZ © 1995 Computer Creations Corp.

A Brief History of Infinity

The infinite has always been a slippery concept. Even the commonly accepted mathematical view, developed by Georg Cantor, may not have truly placed infinity on a rigorous foundation

by A. W. Moore

For more than two millennia, mathematicians, like most people, were unsure what to make of the infinite. Several paradoxes devised by Greek and medieval thinkers had convinced them that the infinite could not be pondered with impunity. Then, in the 1870s, the German mathematician Georg Cantor unveiled transfinite mathematics, a branch of mathematics that seemingly resolved all the puzzles the infinite had posed. In his work Cantor showed that infinite numbers existed, that they came in different sizes and that they could be used to measure the extent of infinite sets. But did he really dispel all doubt about mathematical dealings with infinity? Most people now believe he did, but I shall suggest that in fact he may have reinforced that doubt.

The hostility of mathematicians toward infinity began in the fifth century B.C., when Zeno of Elea, a student of Parmenides, formulated the well-known paradox of Achilles and the tortoise [see "Resolving Zeno's Paradoxes," by William I. McLaughlin; SCIENTIFIC AMERICAN, November 1994]. In this conundrum the swift demigod challenges the slow tortoise to a race and grants her a head start. Before he can overtake her, he must reach the point at which she began, by which time she will have advanced a little. Achilles must now make up the new distance separating them, but by the time he does so, she will have advanced again. And so on, ad infinitum. It seems that Achilles can never overtake the tortoise. In like manner

Zeno argued that it is impossible to complete a racecourse. To do so, it is necessary to reach the halfway point, then the three-quarters point, then the seven-eighths point, and so on. Zeno concluded not only that motion is impossible but that we do best not to think in terms of the infinite.

The mathematician Eudoxus, similarly wary of the infinite, developed the so-called method of exhaustion to circumvent it in certain geometric contexts. Archimedes exploited that method some 100 years later to find the exact area of a circle. How did he proceed? In the box on page 114, I present not his actual derivation but a corruption of it. Part of Archimedes' own procedure was to consider the formula for the area of a polygon with n equal sides—call it P_n —inscribed inside a circle C . According to the distortion of his argument, this formula can be applied to the circle itself, which is just a polygon with infinitely many, infinitely small sides.

The perversion of Archimedes' argument has some intuitive appeal, but it would not have satisfied Archimedes. We cannot uncritically make use of the infinite as though it were just some unusually big integer. Part of what is going on here is that the larger n is, the more nearly P_n matches C . But it is also true that the larger n is, the more nearly P_n approximates a circle with a bulge—call it C^* . The key point intuitively is that C , unlike its deformed counterpart C^* , is the limit of the polygons—or what they are tending toward.

Still, it is very hard to see any way of capturing this intuition without, once again, thinking of C as an "infinigon." Archimedes provided a way. He pinpointed the crucial difference between C and C^* by proving the following point: no matter how small an area you consider, call it ε (the Greek letter epsilon),

there exists an integer n that is large enough for the area of P_n to be within ε of the area of C . The same is not true of C^* . This fact, combined with a similar result for circumscribed polygons and supplemented with a refined version of the logic contained in that argument, finally enabled Archimedes to show, without ever invoking the infinite, that the area of a circle equals πr^2 .

The Actual and Potential Infinite

Although Archimedes successfully ducked the infinite in this particular exercise, the Pythagoreans (a religious society founded by Pythagoras) happened on a case in which the infinite was truly inescapable. This find shattered their belief in two fundamental cosmological principles: Peras (the limit), which subsumed all that was good, and Apeiron (the unlimited or infinite), which encompassed all that was bad. They had insisted that the whole of creation could be understood in terms of, and indeed was ultimately constituted by, the positive integers, each of which is finite. This reduction was made possible, they maintained, by the fact that Peras was ever subjugating Apeiron.

Pythagoras had discovered, however, that the square of the hypotenuse (the longest side) of a right-angled triangle is equal to the sum of the squares of the other two sides. Given this theorem, the ratio of a square's diagonal to each side is $\sqrt{2}$ to 1, since $1^2 + 1^2 = (\sqrt{2})^2$. Were Peras impervious, this ratio should be expressible in the form p to q , where p and q are both positive integers. Yet this is impossible. Imagine two positive integers, p and q , such that the ratio of p to q , or p divided by q , is equivalent to $\sqrt{2}$. We can assume that p and q have no common factor greater than 1 (we could, if necessary,

A. W. MOORE is a tutorial fellow in philosophy at St. Hugh's College of the University of Oxford. He studied for his Ph.D. in the philosophy of language at Balliol College at Oxford. His main academic interests are in logic, metaphysics and the philosophies of Immanuel Kant and Ludwig Wittgenstein, all of which have informed his work on the infinite. He is at work on a book about the metaphysics of objectivity and subjectivity.

UNLIMITED EXPANSE is conveyed by the baked floor of California's Death Valley. To grasp the infinite, mathematicians have confronted several paradoxes, concluding that infinities come in different types and that some are bigger than others.

divide by that factor). Now, p^2 is twice q^2 . So p^2 is even, which means that p itself is even. Hence, q must be odd, otherwise 2 would be a common factor. But consider: if p is even, there must be a positive integer r that is exactly half

of p . Therefore, $(2r)^2$ equals $2q^2$, or $2r^2$ equals q^2 , which means that q^2 is even, and so q itself is also even, contrary to what was proved above.

For the Pythagoreans, this result was nothing short of catastrophic. (Accord-

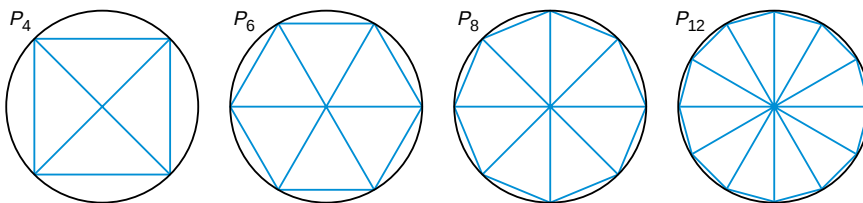
ing to legend, one of them was shipwrecked at sea for revealing the discovery to their enemies.) They had come on an "irrational" number. In doing so, they had seen the limitations of the positive integers, and they had been forced



Archimedes and the Area of a Circle

How did Archimedes use the method of exhaustion to find the area of a circle? Here is the corruption of his argument. Imagine a circle C that has a radius r . For each integer n greater than 2, we can construct a regular polygon with n sides and inscribe it inside C . This n -sided polygon—call it P_n —can be divided into n congruent triangles. Label the base of each triangle b_n and its height h_n . Then the area of each triangle is $1/2 b_n h_n$. Thus, the area of P_n as a whole is $n(1/2 b_n h_n)$, or $1/2 n b_n h_n$. But C itself is a polygon with infinitely many, infinitely small sides. In other

words, C results when we extend the original definition of P_n and allow n to be infinite. In this case, $n b_n$ is the circumference of C , which equals $2\pi r$ (which follows from the definition of π), and h_n is the radius r . So the area of C is $1/2(2\pi r r)$, or simply πr^2 .



JARED SCHNEIDMAN DESIGN

to acknowledge the presence of the infinite in their very midst. Indeed, a modern mathematician would say that $\sqrt{2}$ is a kind of “infinite object.” Not only is its decimal expansion infinite, but this expansion never adopts a recurring finite pattern.

In the fourth century B.C. Aristotle recognized a more general problem. On the one hand, we are under pressure to acknowledge the infinite. Quite apart from what we may have to say about $\sqrt{2}$, time appears to continue indefinitely, numbers seem to go on endlessly, and space, time and matter seem to be forever divisible. On the other hand, we are under pressure from various sources, including Zeno's paradoxes, to repudiate the infinite.

Aristotle's solution to this dilemma was masterful. He distinguished between two different kinds of infinity. The actual infinite is that whose infinitude exists at some point in time. In contrast, the potential infinite is that whose infinitude is spread over time. All the objections to the infinite, Aristotle insisted, are objections to the actual infinite. The potential infinite, on the other hand, is a fundamental feature of reality. It deserves recognition

in any process that can never end, including counting, the division of matter and the passage of time itself. This distinction between the two types of infinity provided a solution to Zeno's paradoxes. Traversing a region of space does not involve moving across an actual infinity of subregions, which would be impossible. But it does mean crossing a potential infinity of subregions, in the sense that there can be no end to the process of dividing the space. This conclusion, fortunately, is harmless.

Aristotle's parting of the actual and the potential infinite long stood as orthodoxy. Nevertheless, scholars usually interpreted his reference to time as a metaphor for something deeper and more abstract. Existing “in time” or existing “all at once” assumed much broader meanings. To take exception to the actual infinite was to object to the very idea that some entity could have a property that surpassed all finite measure. It was also to deny that the infinite was itself a legitimate object of study.

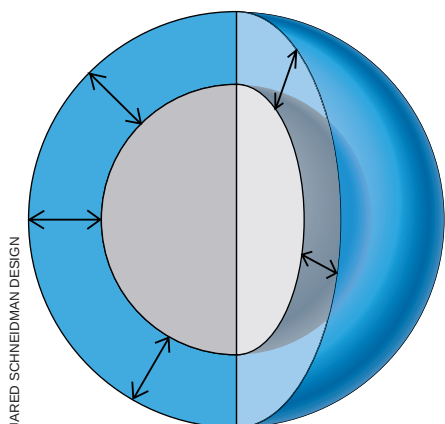
Some 2,000 years later the infinite, both actual and potential, exercised mathematicians once more as they developed the calculus. Early work on the calculus, ushered in by Isaac Newton

and Gottfried Wilhelm Leibniz in the late 17th century, fell far short of Greek standards of rigor. Indeed, mathematicians had made extensive, uncritical use of infinitesimals, items taken to be too small for measure. Sometimes these quantities were considered equal to zero. For example, when they were added to another number, the value of the original number remained the same. At other times, they were taken to be different from zero and used in division. Guillaume François Antoine de l'Hôpital wrote: “A curve may be regarded as the totality of an infinity of straight segments, each infinitely small: or ... as a polygon with an infinite number of sides.” Only in the 19th century did French mathematician Augustin-Louis Cauchy and German mathematician Karl Weierstrass resuscitate the method of exhaustion and give the calculus a secure foundation.

The Infinite and Equinumerosity

As a result of Cauchy's and Weierstrass's work, most mathematicians felt less threatened by Zeno's paradoxes. Of more concern by then was a family of paradoxes born in the Middle Ages dealing with equinumerosity. These puzzles derive from the principle that if it is possible to pair off all the members of one set with all those of another, the two sets must have equally many members. For example, in a nonpolygamous society there must be just as many husbands as wives. This principle looks incontestable. Applied to infinite sets, however, it seems to flout a basic notion first articulated by Euclid: the whole is always greater than any of its parts. For instance, it is possible to pair off all the positive integers with those that are even: 1 with 2, 2 with 4, 3 with 6 and so on—despite the fact that positive integers also include odd numbers.

The medievals proffered many simi-



JARED SCHNEIDMAN DESIGN

1	2	3	4	...	n	...
↑	↑	↑	↑		↑	
↓	↓	↓	↓		↓	
1	4	9	16	...	n ²	...

SETS ARE THE SAME SIZE if all their members can be paired with one another. But this principle seems to be violated in infinite sets. All the squared integers can be matched with every single positive integer (above), even though the set of squares seems smaller. Similarly, all the points on the smaller sphere can be paired off with those on the larger one (left).

lar examples, some of which were geometric. In the 13th century the Scottish mathematician John Duns Scotus puzzled over the case of two concentric circles: all the points on the shorter circumference of the smaller circle can be paired off with all the points on the longer circumference of the bigger circle. The same result applies to two spheres [see bottom illustration on opposite page]. Some 350 years later Galileo discussed a variation of the pairing example of the even integers, based instead on squared integers. Particularly striking is the fact that as increasingly larger segments of the sequence of positive integers are considered, the proportion of these integers that are squares tends toward zero. Nevertheless, the pairing still proceeds indefinitely.

It is certainly tempting, in view of these difficulties, to eschew infinite sets entirely. More generally, it is tempting to deny, as did Aristotle, that infinitely many things can be gathered together all at once. Eventually, though, Cantor challenged the Aristotelian view. In work of great brilliance he took the paradoxes in his stride and formulated a coherent, systematic and precise theory of the actual infinite, ready for any skeptical gaze. Cantor accepted the "pairing off" principle and its converse, namely, that no two sets are equinumerous unless their members can be paired off. Accordingly, he accepted that there are just as many even positive integers as there are positive integers altogether (and likewise in the other paradoxical cases).

Let us for the sake of argument, and contemporary mathematical convention for that matter, follow suit. If this principle means that the whole is no greater than its parts, so be it. We can in fact use this idea to define the infinite, at least in its application to sets: a set is infinite if it is no bigger than one of its parts. More precisely, a set is infinite if it has as many members as does one of its proper subsets.

What remains an open question, once things have been clarified in this way, is whether all infinite sets are equinumerous. Much of the impact of Cantor's work came in his demonstration that they are not. There are different infinite sizes. This proposition results from what is known as Cantor's theorem: no set, and in particular no infinite set, has as many members as it has subsets. In other words, no set is as big as the set of its subsets. Why not? Because if a set were, it would be possible to pair off all its members with all its subsets. Some members would then be paired off with subsets that contained them, others not. So what of the set of those mem-

SUBSETS	POSITIVE INTEGERS					
	1	2	3	4	5	...
EVEN INTEGERS	NO	YES	NO	YES	NO	
ODD INTEGERS	YES	NO	YES	NO	YES	
PRIMES	NO	YES	YES	NO	YES	
SQUARES	YES	NO	NO	YES	NO	
MULTIPLES OF 4	NO	NO	NO	YES	NO	
...						
DIAGONALIZED SET	YES	YES	NO	NO	YES	...

JARED SCHNEIDMAN DESIGN

CANTOR'S THEOREM—that no set has as many members as it has subsets—is proved by diagonalization, which creates an extra subset. Each subset of the set of positive integers is represented as a series of yeses and noes. A yes indicates that the integer belongs to the subset; a no that it does not. Replacing each yes with a no, and vice versa, down the diagonal (shaded area) creates another subset.

bers that were not included in the set with which they had been paired? No member could be paired off with this subset without contradiction.

The argument can be recast in a diagram [see illustration above]. For convenience, I will focus on the set of positive integers. I can represent any subset of the set of positive integers by an infinite sequence of yeses and noes, registering whether successive positive integers do or do not belong to the set. For example, the set of even integers can be represented by the sequence <no, yes, no, yes, no...>, corresponding to 1, 2, 3, 4, 5, and so forth. We can do the same for the set of odd integers <yes, no, yes, no, yes...>, the set of prime numbers <no, yes, yes, no, yes...> and the set of squares <yes, no, no, yes, no...>. Generally, then, any assignment of different subsets to individual positive integers (such as the purely arbitrary example illustrated) can be represented as an infinite square of yeses and noes.

To show that at least one subset is nowhere on this list of subsets, we make a new subset by moving down the "square's diagonal," replacing each yes with a no, and vice versa. In the case illustrated, we write <yes, yes, no, no...>. What results represents the subset in question. For by construction it differs from the first subset listed with respect to whether 1 belongs to it, from the second with respect to whether 2 belongs to it, from the third with respect to whether 3 belongs to it, and so on. There is a pleasant historical quirk here: just as studying a diagonal had led the Pythagoreans to acknowledge an infinitude beyond the grasp of the positive integers, the same was true in a different way in Cantor's case.

Cantor later devised infinite cardi-

nals—numbers that can be used to measure the size of infinite sets. He invented a kind of arithmetic for them as well. Having defined his terms, he explored what happens when one infinite cardinal is added to another, when it is multiplied by another, when it is raised to a power, and so forth. His work showed mathematical craftsmanship of the highest caliber. But even in his own terms, difficulties remained. The continuum problem is perhaps the best known of these troubles. The set of positive integers, we have seen, is smaller than the set of sets of positive integers. But how much smaller? Specifically, are there any sets of intermediate size?

Cantor's Continuum Hypothesis

Cantor's own hypothesis, his famous "continuum hypothesis," was that there are not. But he never successfully proved this idea, nor did he disprove it. Subsequent work has shown that the situation is far graver than he had imagined. Using all the accepted methods of modern mathematics, the issue cannot be settled. This problem raises philosophical questions about the determinacy of Cantor's conception. Asking whether the continuum hypothesis is true may be like asking whether Hamlet was left-handed. It may be that not enough is known to form an answer. If so, then we should rethink how well Cantor's work tames the actual infinite.

Of even more significance are questions surrounding the set of all sets. Given Cantor's theorem, this collection must be smaller than the set of sets of sets. But wait! Sets of sets are themselves sets, so it follows that the set of sets must be smaller than one of its own proper subsets. That, however, is

Diagonalization and Gödel's Theorem

The diagonalization used in establishing Cantor's theorem also lies at the heart of Austrian mathematician Kurt Gödel's celebrated 1931 theorem. Seeing how offers a particularly perspicuous view of Gödel's result.

Gödel's theorem deals with formal systems of arithmetic. By arithmetic I mean the theory of positive integers and the basic operations that apply to them, such as addition and multiplication. The theorem states that no single system of laws (axioms and rules) can be strong enough to prove all true statements of arithmetic without at the same time being so strong that it "proves" false ones, too. Equivalently, there is no single algorithm for distinguishing true arithmetical statements from false ones. Two definitions and two lemmas, or propositions, are needed to prove Gödel's theorem. Proof of the lemmas is not possible within these confines, although each is fairly plausible.

Definition 1: A set of positive integers is arithmetically definable if it

can be defined using standard arithmetical terminology. Examples are the set of squares, the set of primes and the set of positive integers less than, say, 821.

Definition 2: A set of positive integers is decidable if there is an algorithm for determining whether any given positive integer belongs to the set. The same three sets above serve as examples.

Lemma 1: There is an algorithmic way of pairing off positive integers with arithmetically definable sets.

Lemma 2: Every decidable set is arithmetically definable.

Given lemma 1, diagonalization yields a set of positive integers that is not arithmetically definable. Call this set D . Now suppose, contrary to Gödel's theorem, there is an algorithm for distinguishing between true arithmetical statements and false ones. Then D , by virtue of its construction, is decidable. But given lemma 2, this proposition contradicts the fact that D is not arithmetically definable. So Gödel's theorem must hold after all. Q.E.D.



Kurt Gödel, 1906–78

impossible. The whole can be the same size as the part, but it cannot be smaller. How did Cantor escape this trap? With wonderful pertinacity, he denied that there is any such thing as the set of sets. His reason lay in the following picture of what sets are like. There are things that are not sets, then there are sets of all these things, then there are sets of all those things, and so on, without end. Each set belongs to some further set, but there never comes a set to which every set belongs.

Cantor's reasoning might seem somewhat ad hoc. But an argument of the sort is required, as revealed by Bertrand Russell's memorable paradox, discovered in 1901. This paradox concerns the set of all sets that do not belong to themselves. Call this set R . The set of mice, for example, is a member of R ; it does not belong to itself because it is a set, not a mouse. Russell's paradox turns on whether R can belong to itself. If it does, by definition it does not belong to R . If it does not, it satisfies the condition for membership to R and so does belong to it. In any view of sets, there is something dubious about R . In Cantor's view, according to which no set belongs to itself, R , if it existed, would be the set of all sets. This argument makes Cantor's picture, and the rejection of R that goes with it, appear more reasonable.

But is the picture not strikingly Aristotelian? Notice the temporal metaphor. Sets are depicted as coming into being

"after" their members—in such a way that there are always more to come. Their collective infinitude, as opposed to the infinitude of any one of them, is potential, not actual. Moreover, is it not this collective infinitude that has best claim to the title? People do ordinarily define the infinite as that which is endless, unlimited, unsurveyable and immeasurable. Few would admit that the technical definition of an infinite set expresses their intuitive understanding of the concept. But given Cantor's picture, endlessness, unlimitedness, unsurveyability and immeasurability more properly apply to the entire hierarchy than to any of the particular sets within it.

In some ways, then, Cantor showed that the set of positive integers, for example, is "really" finite and that what is "really" infinite is something way beyond that. (He himself was not averse to talking in these terms.) Ironically, his work seems to have lent considerable substance to the Aristotelian orthodoxy that "real" infinitude can never be actual.

Some scholars have objected to my suggestion that, in Cantor's conception, the set of positive integers is "really" finite. They complain that this assertion is at variance not only with standard mathematical terminology but also, contrary to what I seem to be suggesting, with what most people would say.

Well, certainly most people would say the set of positive integers is "really" infinite. But, then again, most people are unaware of Cantor's results. They

would also deny that one infinite set can be bigger than another. My point is not about what most people would say but rather about how they understand their terms—and how that understanding is best able, for any given purpose, to absorb the shock of Cantor's results. Nothing here is forced on us. We could say some infinite sets are bigger than others. We could say the set of positive integers is only finite. We could hold back from saying either and deny that the set of positive integers exists.

If the task at hand is to articulate certain standard mathematical results, I would not advocate using anything other than standard mathematical terminology. But I would urge mathematicians and other scientists to use more caution than usual when assessing how Cantor's results bear on traditional conceptions of infinity. The truly infinite, it seems, remains well beyond our grasp.

FURTHER READING

- INFINITY AND THE MIND: THE SCIENCE AND PHILOSOPHY OF THE INFINITE. Rudy Rucker. Harvester, 1982.
- TO INFINITY AND BEYOND: A CULTURAL HISTORY OF THE INFINITE. Eli Maor. Birkhauser, 1986.
- THE INFINITE. A. W. Moore. Routledge, 1990.
- INFINITY. Edited by A. W. Moore. Dartmouth, 1993.
- UNDERSTANDING THE INFINITE. Shaughan Lavine. Harvard University Press, 1994.

The Tapestry of Power in a Mesopotamian City

Mashkan-shapir was for a brief time one of the most important cities in the civilized world. Its remains challenge traditional notions of power distribution in early urban society

by Elizabeth C. Stone and Paul Zimansky

Remains of the world's first cities are the most noteworthy features of the landscape in southern Iraq, and for nearly two centuries archaeologists have probed them and puzzled over their artifacts. Built up over the course of five and a half millennia of intermittent occupation, these tells—mounds of building rubble and associated artifacts—can be as large as a mile in diameter; some rise more than 100 feet above the plain. Babylon, Ur, Uruk, Nippur and Kish have yielded abundant evidence of the material culture of Mesopotamian society. Thanks

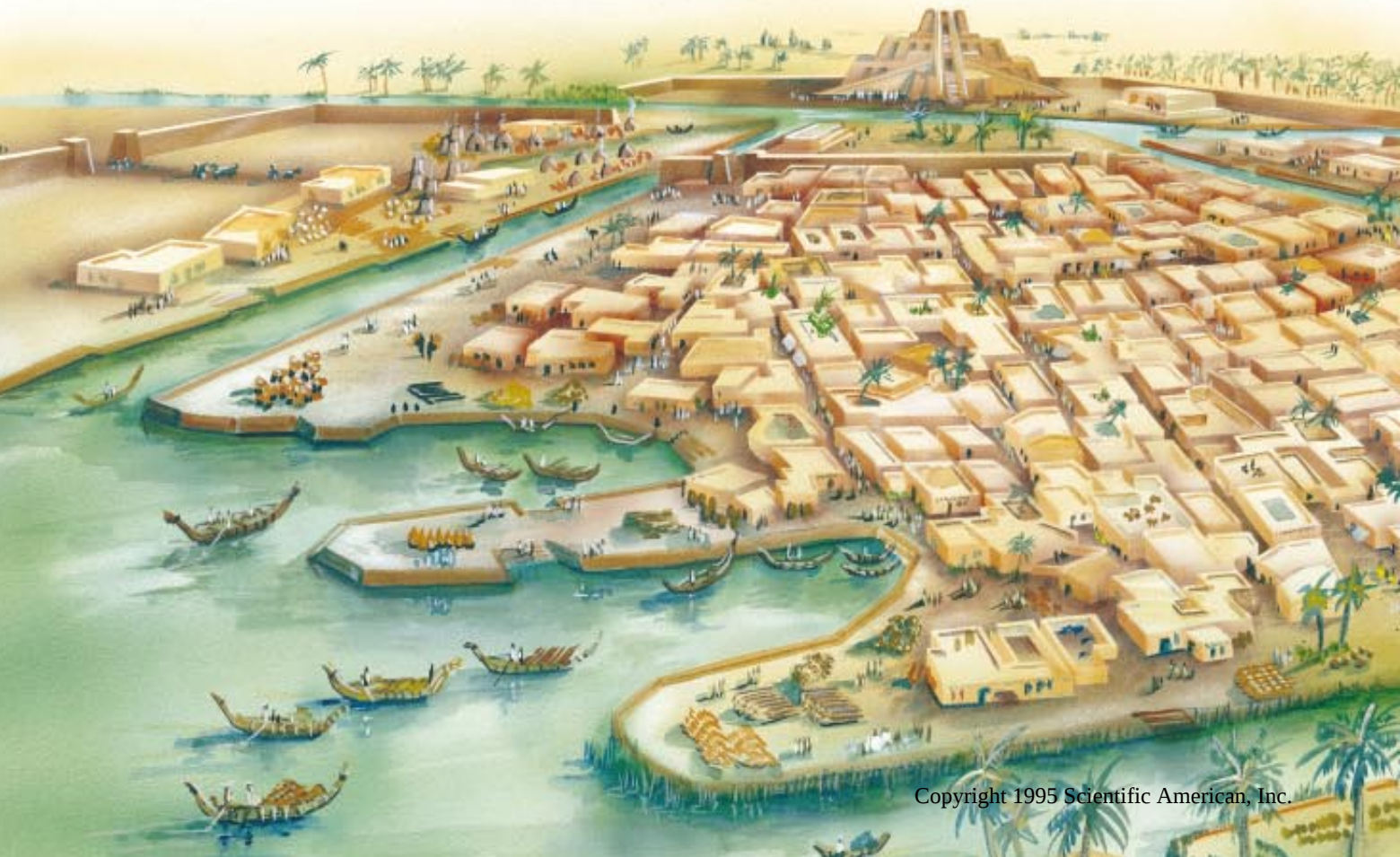
to their citizens' relatively imperishable writing medium—clay tablets—they have also provided detailed textual testimony on political, intellectual, religious and social institutions.

Nevertheless, the physical and social organization of these most ancient cities is still poorly understood, for a variety of reasons. Paradoxically, the very richness of evidence has led to ignorance. The tells are so massive that even the best-financed field parties can excavate only tiny fractions of each city. More important, the arrangements of buildings that archaeologists uncover

generally do not represent a city that actually existed at one particular time. The ancient inhabitants built on earlier structures in some cases and swept them away or modified them in others. One might imagine a similar problem facing archaeologists trying to understand London a few thousand years hence: they would be confronted by the mixed remains of modern skyscrapers, Victorian buildings, Norman castles and even a Roman garrison; reconstructing the city as it looked during any given period would be almost impossible.

Urban sociologists have long known

JANA BRENNING



that the plans of contemporary cities reflect patterns of social organization. Our own survey of non-Mesopotamian early cities shows that similar conclusions can be drawn about early urban sites. Where power is highly focused and based on coercion, centers of administration, religion, manufacture and trade cluster together, surrounded by residences of the elite. In contrast, societies in which diverse groups share control and in which decision making takes place at various levels of the social hierarchy show little or no evidence of such concentration. The intimate ties between elites and the rest of the population in these decentralized cities are mirrored by a mixture of rich and poor houses in each residential district.

Where are humankind's first cities to be placed in this spectrum? Archaeologists have tended to emphasize centralization, but a close look at their reasoning, combined with our recent findings at a site called Mashkan-shapir, indicate that this view needs revision. Early excavations in Mesopotamia focused on seats of wealth and power—palaces and temples—and led researchers to take a similarly narrow view in reconstructing the society that built them. Yet concentration on the physical remains of high status obscured the fact that Mesopotamian texts do not identify clearly differentiated social classes. Instead they record the importance of general assemblies in decision making.

There may also be a more subtle bias at work. Historians recognize that industrialism and capitalism have so transformed the world that there are no modern analogues for ancient cities. Rather than considering a wide range of potential urban organizations, however, some scholars have perhaps too readily posited a unified model for a "preindustrial city" based on a few, well-studied (and centralized) examples. In devising this model they have rarely looked further afield than ancient Greece and sometimes no further than medieval Europe.

As a result, researchers have in effect taken for granted that cities in Mesopotamia were shaped by the same forces as were later European ones, among them a stable agricultural base and a fixed value for any given plot of land. In fact, the economic base in this region was anything but geographically stable—as indicated by the importance of nomadic herding. Even cultivated land was impermanent: annual floods, high evaporation rates and rapid poisonous salinization of land under cultivation led to a constantly shifting mosaic of rich irrigated fields and orchards, deserts and marshes, in which wealth or power had little to do with permanent control of a particular parcel. Detailed descriptions of many preindustrial urban civilizations—in West Africa, the Islamic Middle East and the New World at the time of the conquistadors—show considerable variability in organization;

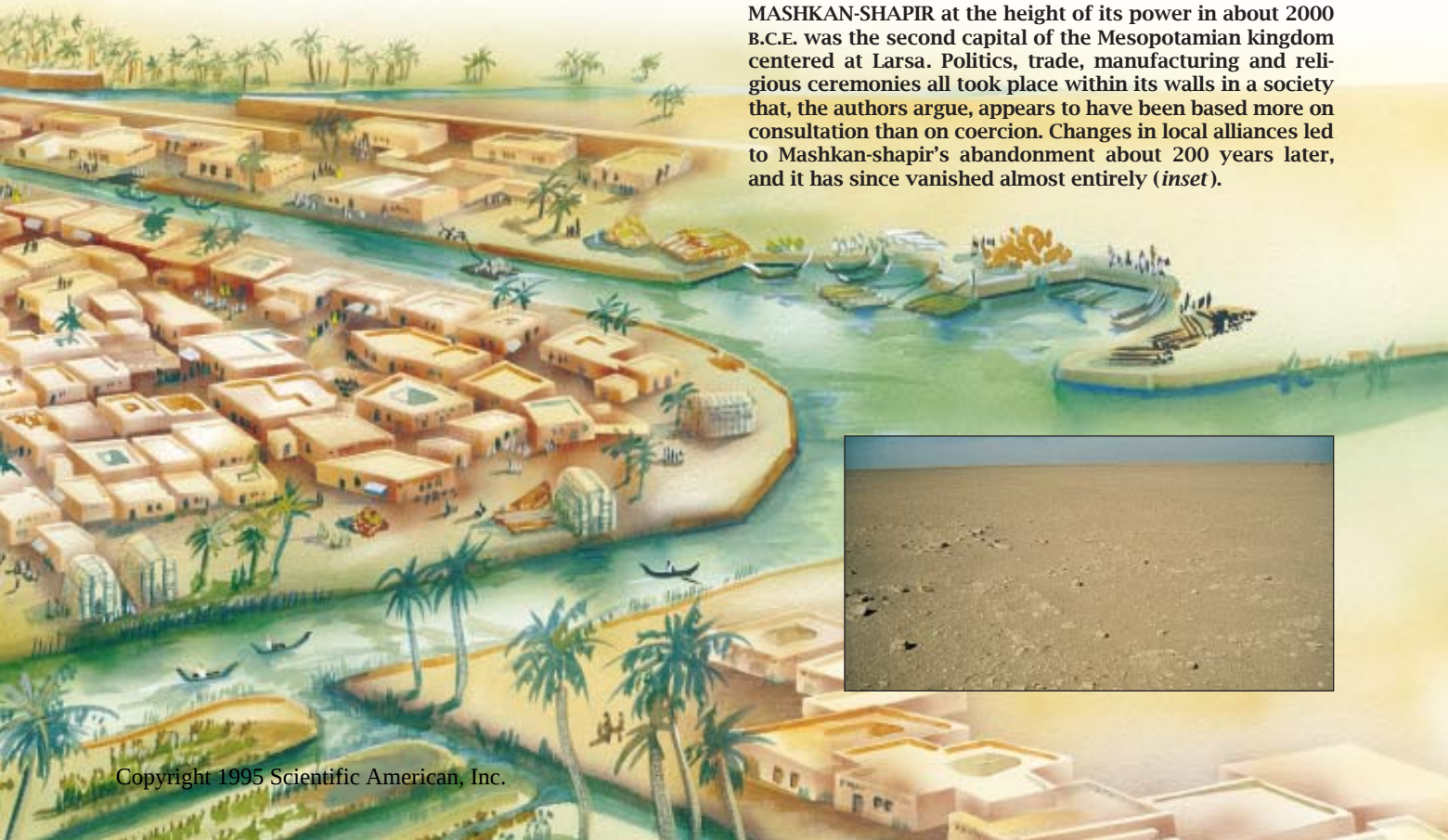
ELIZABETH C. STONE and PAUL ZIMANSKY have collaborated for more than 20 years on the study of ancient cities in the Middle East. Stone, who received her Ph.D. in 1979 from the University of Chicago, is professor of anthropology at the State University of New York at Stony Brook. Zimansky received his Ph.D. in 1980 from the University of Chicago and is now associate professor of archaeology at Boston University.

they also suggest a link between the permanence of agricultural land and the degree of social and political centralization. There is thus no reason to assume a priori that Mesopotamian cities were especially centralized.

An Untouched Site

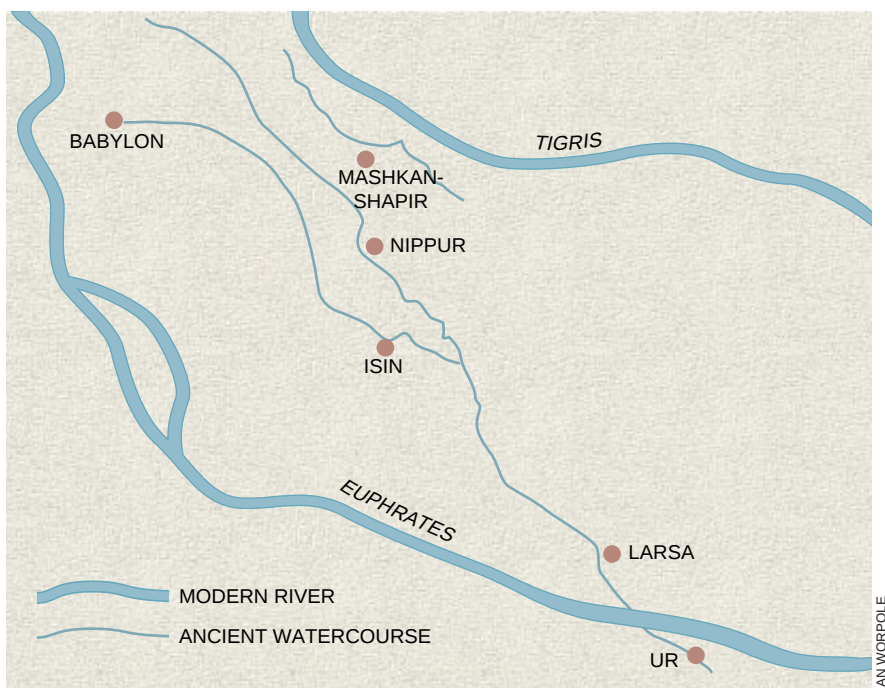
The archaeological project that led to the discovery of Mashkan-shapir came about when we decided to attack the question of urban organization by seeking out a site that had been occupied only during a single period. The ruins of such a short-lived city would provide a snapshot of urban layout; by analyzing that physical organization, we would be able to draw some conclusions about whether it arose through coercion by priests and kings or by consensus among diverse segments of society. We were looking for a site that was of urban scale but otherwise quite

MASHKAN-SHAPIR at the height of its power in about 2000 B.C.E. was the second capital of the Mesopotamian kingdom centered at Larsa. Politics, trade, manufacturing and religious ceremonies all took place within its walls in a society that, the authors argue, appears to have been based more on consultation than on coercion. Changes in local alliances led to Mashkan-shapir's abandonment about 200 years later, and it has since vanished almost entirely (inset).





ANCIENT MESOPOTAMIA covered the region between the Tigris and Euphrates rivers in what is now Iraq (*map, left*). Mashkan-shapir (*center*) rose to prominence because of its strategic position near the Tigris; it controlled trade in many goods that came down the waterway from the north. A satellite photograph (*right*) shows the location of the city (*green*), its canals and the Tigris at the time Mashkan-shapir flourished.



different from the great cities that had been investigated before—occupied for a short time and left relatively undisturbed since its abandonment.

Looking through data collected by other archaeologists on the overall distribution of ancient settlements in Iraq, we chose a site that Robert McC. Adams, then at the University of Chicago, had found in the mid-1970s. He had given it a number but no name.

Our initial visit in January 1987 took place under less than ideal circumstances—it coincided with the first day of Iran's "final offensive" against Iraq during the Iran-Iraq war. Nevertheless, we could see that the site was littered with traces of occupation. Walls, pottery, graves, even ancient canals were all clearly visible across an area more than half a mile in diameter. Most of the shards dated to the first quarter of the second millennium B.C.E. Both the quantity and quality of the remains made it clear that the only recent visitors to this place, Bedouin and their camels, had left it largely undisturbed.

The site stayed nameless for another two years while we made arrangements to return. Then, in 1989, shortly after we began our survey, we chanced on a chunk of baked clay near the remains of a gate in the city wall. The chunk bore a cuneiform inscription. In short order we brushed clean 150 similar fragments, which together turned out to carry multiple copies of an inscription commemorating the wall's construction. On the third piece we examined were three clearly legible signs of the four that are used to write "Mashkan-shapir": the

name of what was once, albeit briefly, one of the most important cities in the world.

Mashkan-shapir first appears in the historical record as a small sheep-rearing village on the fringe of the Mesopotamian heartland in the latter part of the third millennium B.C.E. It probably would have remained obscure had it not been for the political intricacies of the early second millennium. Shortly before 2000 B.C.E., an empire centered at Ur, which had controlled the entire alluvial plain, collapsed. For the next two centuries, several cities vied for hegemony, foremost among them Isin and Larsa. Larsa was probably more powerful, but Isin was farther upstream on the Euphrates River and so could impede its rival's access to essential commodities, such as wood, metal and stone, coming down the river from the northwest. Larsa countered by moving to control the eastern part of the valley and to secure access to the Tigris River. As Larsa's northern outpost, Mashkan-shapir quickly grew to urban size, eventually serving as a second capital for the kingdom.

The city retained its role after Isin was defeated, because the rise of Babylon under Hammurabi's rule in the early 18th century B.C.E. created similar competition for control of access to goods. Soon enough, however, Hammurabi's conquests, which reunited most of southern Mesopotamia, made Mashkan-shapir's strategic position irrelevant. The city was abandoned around 1720 B.C.E., and the watercourses that had sustained it fell into disrepair. Mash-

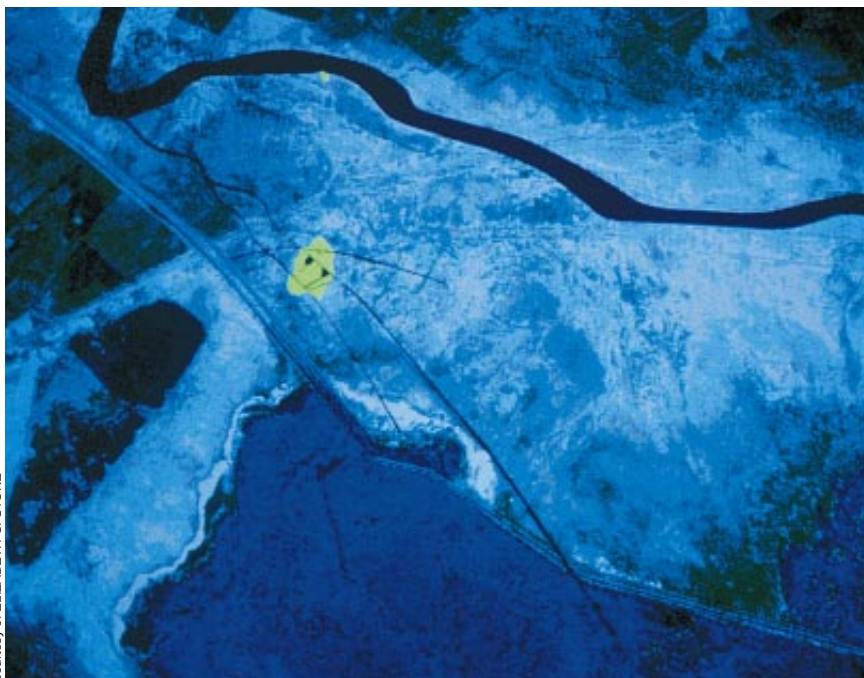
kan-shapir disappeared into the desert.

There is ample historical testimony as to what kind of a place Mashkan-shapir was during its brief flowering. As a second capital to Larsa, it was a political city where much diplomatic activity was conducted. It was also a gateway to the trade route up the Tigris. Finally, it housed a primary sanctuary to Nergal, god of death, among the most powerful deities in the Mesopotamian pantheon. Because administration, religion and trade were the major activities in other large Mesopotamian cities, the organization of Mashkan-shapir has strong implications for that of other sites.

Difficulties in the Field

Opportunities for work at the site have been constrained both by the time required to raise funds for extended fieldwork and by the political vicissitudes of the region. We undertook the initial exploration in two three-week campaigns in 1987 and 1989. From January through May 1990 we were able to work continuously at the site thanks to the support of the National Science Foundation, the National Endowment for the Humanities, the National Geographic Society and the American Schools of Oriental Research. We conducted a thorough surface survey and aerial reconnaissance to map the main features of the site, supplemented by modest excavations to determine the relation between surface traces and subsurface remains. We also made use of satellite imagery to understand local geology.

Shortly after we finished the first



phase of the project and returned to the U.S., Iraq's invasion of Kuwait brought an end to archaeological work by foreigners. We hope that someday we can return and apply the insight we now have to make excavations at locations that will be most informative about the operation of this ancient city.

Survey by Kite and Foot

Mapping Mashkan-shapir was not easy. The city's remains hardly form a tell in the traditional sense. The site has been seriously eroded by wind, which has reduced the latest buildings to foundations and left heavier artifacts exposed on the surface. Only in a few places do the contours of the site rise more than two meters above the plain.

As a result, larger architectural patterns, such as the path of the city wall, were difficult to make out from ground level. Indeed, Mashkan-shapir's wall is visible only from the air for much of its circumference, and in some places it has vanished entirely. Aerial reconnaissance was essential to our work. We took advantage of the site's strong winds to loft a camera-bearing kite. The elevation of the kite changed with wind velocity, and so the area included in each image varied considerably. We did our best to make up for these irregularities by taking many pictures and having a great deal of redundancy in our coverage.

The 1,600 aerial photographs we took would have been of little use in mapping were it not for software designed to facilitate analysis of satellite images and to create maps for city planners

and geographers. We had marked the corners of the squares in our survey grid, each measuring 50 meters on a side, with crosses that would be visible in the photographs, thus making it possible to orient the images and compensate for geometric distortions caused by the swinging of the camera. Digitized versions of the images were corrected, recomposed at a uniform scale and assembled in a mosaic detailed enough to identify the position of each individual brick on the site's surface.

We combined aerial reconnaissance with a pedestrian survey based on the same 50-meter grid. A member of our team walked over each grid square in a pattern designed to ensure that nothing was missed, marking key features and artifacts with surveyor's flags. We indicated scattered fragments of bricks, potsherds, copper or ceramic slag, kiln fragments and bitumen on our map in terms of concentrations rather than individual pieces because there were far too many to count. From sampling in limited areas we estimate that at least 30 million pieces of pottery larger than a fingernail lay on the surface.

During the course of the survey, we also mapped graves, platforms of baked brick and mud brick, edges of canals and traces of the city wall. And we located more than 1,200 individual artifacts: tools, weapons, jewelry, pieces of statuary, plaques, figurines, small models and whole pots. The variable distribution of artifacts bore testimony to the complexity of this short-lived city.

For a broader view of the geography around Mashkan-shapir, we turned to

an image taken by a French SPOT remote-sensing satellite in May 1988. The image reveals an ancient bed of the Tigris near the site, which explains why the city was built more than 20 miles from the nearest modern watercourse. The picture also shows the outlines of a series of canals originating in the river and bisecting the city.

A City of Canals

From these findings we were able to discern a great deal about life in Mashkan-shapir. Like all Mesopotamian cities of its era, it was surrounded by a mud-brick wall interrupted by a number of gates. We identified three gates. Two were for road traffic; they were located near major canals presumably because then, as now, roads ran along the waterways that served as vital lines of communication. The pylons of the third gate are on either side of a canal; perhaps it served to regulate either the flow of water or of water traffic.

Surprisingly, the city wall was not always built close to the edge of dense settlement. An area between the settlement and the wall remained empty, except for six isolated buildings that seem to have been storehouses. Many Mesopotamian texts suggest that commodity exchange took place near city gates, and so this space may have served as some kind of market. Another apparently unoccupied area, near a canal, may have been a garden. We know that some Mesopotamian cities incorporated such gardens because a map of Nippur, scratched on a tablet dated a few centuries later, shows a large garden planted on one corner of the city.

The city proper was laid out in five districts, separated by canals. A quadrilateral region in the middle was surrounded by four larger areas to the north, south, east and west [*see illustration on next page*]. The biggest districts, to the north and east, were themselves divided by canals. Broad harbors occupied two of the canal junctions; they must have been centers for commerce.

Many of the city's buildings were lined up along the waterways, making the canals an integral part of the urban fabric. Mashkan-shapir is not the only Mesopotamian city known to have been arranged in this manner, but the extent of canals elsewhere has been obscured by the great height of the deposits that have filled them in. Streets, which were also covered by debris at other sites, are just barely visible at Mashkan-shapir. Some follow the lines of canals; others cut across districts. On one such street, the baked-brick traces of a bridge (or perhaps two quays) can be seen

where the street crosses a canal. Excavations at other Mesopotamian cities have shown that the main streets were supplemented by a network of alleys that gave access to individual houses.

Form Follows Function

The aerial photographs reveal additional demarcations. One wall surrounded a segment of the western district, and another wall cut across the southern part of the central one. Both are similar to an internal wall at Ur, which marked the sacred space surrounding the city's primary temples.

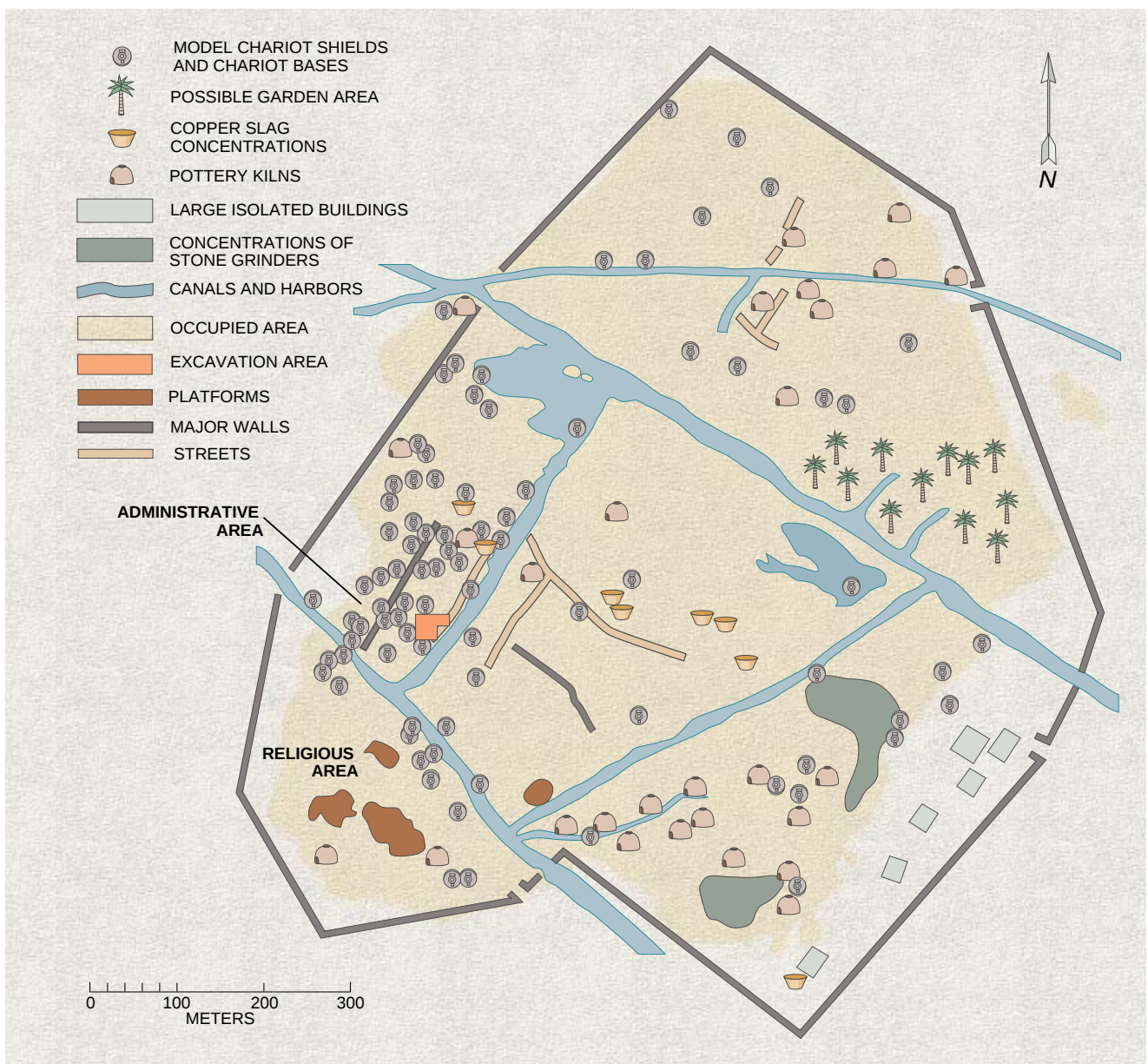
How were the political, religious, eco-

nomics and social functions of Mashkan-shapir laid out within the physical structure defined by streets, canals and internal walls? From the ruins, we can tell where various activities took place.

The primary temple at Mashkan-shapir—that of Nergal, god of death—would have been the visual focal point of the city. Raised up on a platform or ziggurat, it could be seen for miles, and, much like a medieval cathedral, it was a symbol of power. The remains of the baked-brick and mud-brick platforms that seem to have supported the most important sanctuaries lie in the southern region, which was cut off by a canal. The religious character of these plat-

forms was made clear by the discovery of 70 fragments of life-size terra-cotta statues of humans, lions, dogs and horses. Statues of lions frequently adorned the entrances of even minor temples in Mashkan-shapir's time, but archaeologists have found the more complex human and animal statues only at major cities such as Isin.

Across the canal in the central district was another area with religious overtones, identifiable by another platform—the only one so far discovered outside the religious quarter. This region contains traces of numerous burials and a concentration of such grave goods as jewelry and weapons. The area



URBAN AREA about a kilometer square was separated into five regions by canals. Remnants of religious and administrative activities were found at the periphery. Workshops for the making of stone, copper and ceramic objects were scat-

tered rather than being concentrated in a single district; coppersmiths plied their trade along the main street of the central district. This relatively even distribution implies a city in which different classes lived for the most part side by side.

is also cut off from the rest of the settlement by a wall. Most graves at other Mesopotamian sites (and many at Mashkan-shapir) appear in domestic areas, and so this cemetery suggests that one segment of the society—perhaps that associated with the religious or administrative center—had a separate burial ground, probably with an attached temple.

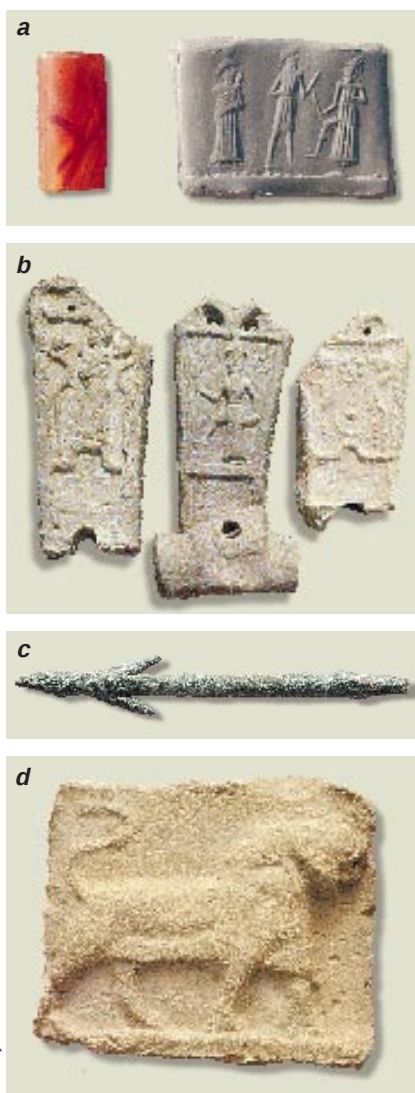
A third distinctive region within the city was the walled-off enclosure in the west, which we believe was an administrative center. It contained a series of very regularly built structures, quite unlike the more haphazard construction of private houses. Although the structures clearly do not constitute a palace such as the one at Mari, which dates to the same era, they may have performed some of the same administrative functions. Like palaces elsewhere, this enclave at Mashkan-shapir is at the periphery of the city. Furthermore, during the 1990 season we excavated numerous unbaked clay sealings from the buildings in this enclosure. These pieces of clay, bearing the impression of carved stone seals, were attached to ropes that closed doors or were embedded in the material that closed storage jars, much as a wax seal might be placed on a letter. They have no place in domestic contexts.

The enclosure contained an oddity as well—a concentration of model chariots decorated with representations of the major gods of the city, Nergal and Shamash, the sun god. It is difficult to say for certain what function these small, two-wheeled vehicles served. Their prevalence in an official space rules out the possibility that they were toys. One possible interpretation is that they served to represent these gods in such legal issues as oath taking.

No Focus of Power

Our survey has identified no other areas of specialized architecture at Mashkan-shapir. Everywhere else the artifacts may be characterized as domestic—figurines, small tools and weapons, commonplace jewelry (such as shell rings) and the traces of houses and burials.

The major roads and canals probably divided this large urban space into discrete residential neighborhoods, but if so they were not segregated by wealth or status. Stone bowls and metal objects (made of imported material and representing significant investments of labor) were scattered quite evenly. Had they been more prevalent in one area than another, an argument could be made for elite residential districts. Cyl-



courtesy of ELIZABETH C. STONE

inder seals—ancient badges of office and items of considerable value in themselves—were also evenly dispersed.

Similarly, we found that manufacturing took place throughout the city. There appears to have been a slight concentration of “smokestack” industry, such as copper smelting and pottery making on the southeastern, leeward side of the city, but artisans appear to have undertaken their work in houses surrounded by other residences. No single section of the site could be called a manufacturing district. Coppersmiths—their workshops marked by concentrations of copper slag—plied their trade along the main street of the central district, for example, but we found most decorative stones and their associated grinders in the southeast. There were two centers of ceramic production (marked by ceramic slag and kiln fragments), one in the north and one in the east. Both were clustered around the smaller canals that ran through these areas. In short, the spatial arrangement

of manufacturing yields the same mostly decentralized picture as the arrangement of houses and artifacts.

Our survey of Mashkan-shapir does not appear to support a highly centralized model for Mesopotamian cities—or for their social organization. It is true that we identified clear foci of religion and administration. But they were sequestered in the southern part of the site. Indeed, they were separated from each other and from the rest of the city by major canals. Moreover, these potential power centers were far from the regions where commerce took place—the harbors and the known city gates. The production of goods seems to have been in the hands of artisans who lived within broader residential neighborhoods that housed both commoners and members of the elite.

The overall organization of Mashkan-shapir suggests that textual sources have not misled us about the broad involvement of Mesopotamian city dwellers in shaping their local power relationships. This conclusion, in turn, may reflect on earlier social structures: if citizens lived in a relatively uncentralized system during the Old Babylonian period, when Mashkan-shapir was thriving, it appears highly unlikely that local authority was more firmly in the hands of a small elite in earlier periods. As a result, the grounds for seeking the origins of civilization in processes of conquest and coercion seem, at the very least, far from compelling.

FURTHER READING

- CULTURAL ATLAS OF MESOPOTAMIA AND THE ANCIENT NEAR EAST. Michael Roaf. Facts on File, 1990.
- THE TELL ABU DUWARI PROJECT, IRAQ, 1987. Elizabeth C. Stone in *Journal of Field Archaeology*, Vol. 17, No. 2, pages 141–162; Summer 1990.
- EARLY MESOPOTAMIA: SOCIETY AND ECONOMY AT THE DAWN OF HISTORY. J. Nicholas Postgate. Routledge, 1992.
- THE TELL ABU DUWARI PROJECT, 1988–1990. Elizabeth C. Stone and Paul Zimansky in *Journal of Field Archaeology*, Vol. 21, No. 4, pages 437–455; Winter 1994.

The Price of Prevention

by Kristin Leutwyler, *staff writer*

During the past century, preventive medicine has revolutionized how most Americans live and die. Immunizations now protect against diseases that only a few generations ago crippled or killed scores of children every year. Public health initiatives mandating cleaner air and food, improved water and sewage systems, and safer home and work environments have shielded the population from infection and injury. Education programs hymn the merits of good nutrition, proper hygiene and regular exercise.

As a direct consequence, the leading causes of death in the U.S. are now those conditions that most often come with age and wear—heart disease, chronic breathing disorders and cancer. And even those pose less threat than they once did, thanks to a host of diagnostic tests, medications and surgical procedures. Some experts now estimate that modern medicine can forestall 70 percent of all illnesses.

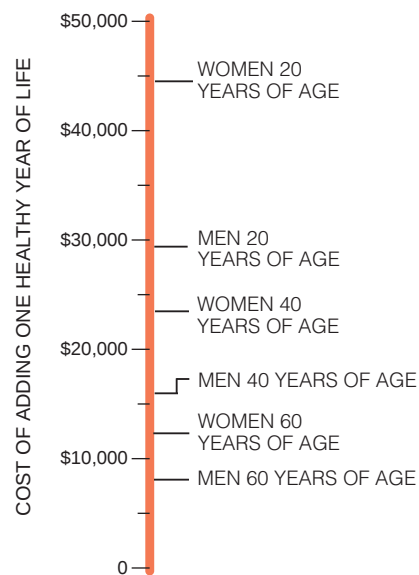
Not surprisingly, then, when lawmakers began seriously discussing health care reform two years ago, most pushed for the inclusion of preventive medical services in insurance benefits packages. They reasoned that once paid for, preventive medicine would reach more people and thereby improve the nation's well-being. Many further asserted that the increased use of prevention would help cut medical spending. The logic was simple: if fewer people got sick, fewer would require medical care, and society would pay fewer bills. Here lay a terrific bargain—greater health for less money.

But in fact, a burgeoning collection of studies shows that disease prevention is rarely a steal. In a recent review of the available data, the Office of Technology Assessment (OTA) reported that of all the preventive services they evaluated, only three paid for themselves in the end: prenatal care for poor women, tests in newborns for some congenital disorders (such as phenylketonuria and hypothyroidism) and most childhood immunizations. In contrast, unless limited to high-risk individuals, screening for cancer costs more than does therapy. Moreover, only when restricted to select groups do adult immunizations and procedures to detect sexually transmitted diseases prove cost-saving. And screening for high blood pressure generally costs more than does treating heart attack and stroke victims.

Indeed, preventing disease often is more expensive than treating it. In absolute terms, even commonplace illnesses have a fairly low incidence in the general population. Vaccinations and screening tests must be given to many people to avert even one case of some sicknesses. Providing a \$10 test to 10,000 people, then, entails the same expense as administering a \$100,000 treatment to one person. Further, preventive measures are not perfect. Some small percentage of any group receiving an immunization or screening test will develop the disease anyway. And guarded against any one ailment, these people systematically become more likely to succumb to another. "Even if we were to be successful with the prevention of all diseases," notes Faith T. Fitzgerald of the University of California at Davis, "we would bear a population of older people undergoing biological decay, who could live longer and cost more while they were dying."

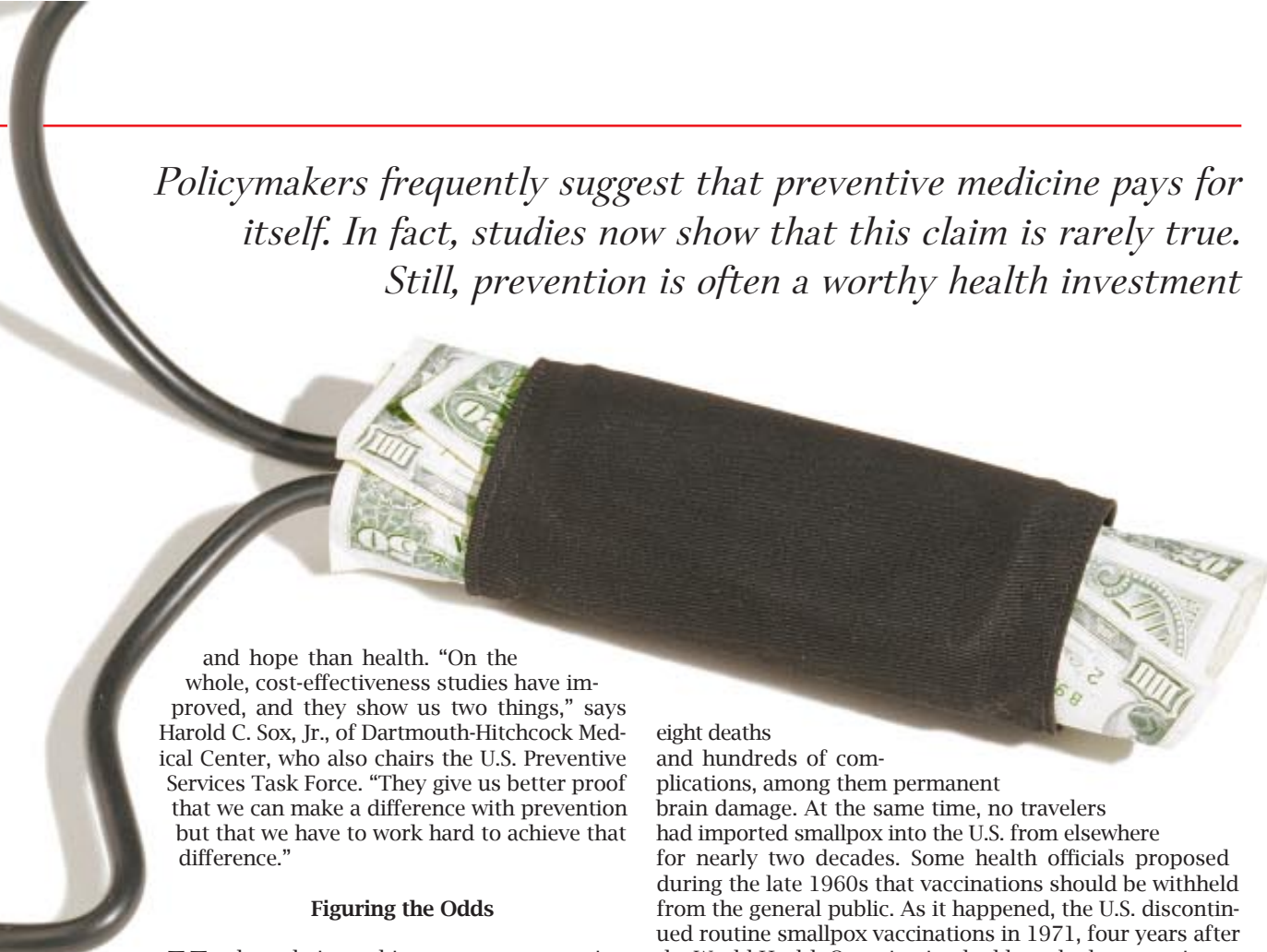
The fact that we all eventually die is no reason not to sidestep pain and suffering whenever possible, whatever the price. If the goals of health care reform are to be met, however, policymakers cannot afford to take gambles. The nation already lacks the resources needed to provide care for its sick and dying. And investing in prevention indiscriminately will further deplete those funds. Analyses that stack the effectiveness of care up against its cost are proving useful for weeding out those preventive services that buy more hype

BLOOD PRESSURE SCREENING would seem to be prevention that pays for itself. The test is cheap, and treating high blood pressure helps many people avoid heart attacks, strokes and other expensive conditions. But in fact, the cost-effectiveness of hypertension screening varies considerably. It costs far less to add a healthy year of life by screening 60-year-olds than by screening younger individuals. So, too, monitoring any age group buys less health for the females than for the males, who are more often hypertensive.



DAN WAGNER

Policymakers frequently suggest that preventive medicine pays for itself. In fact, studies now show that this claim is rarely true. Still, prevention is often a worthy health investment



and hope than health. "On the whole, cost-effectiveness studies have improved, and they show us two things," says Harold C. Sox, Jr., of Dartmouth-Hitchcock Medical Center, who also chairs the U.S. Preventive Services Task Force. "They give us better proof that we can make a difference with prevention but that we have to work hard to achieve that difference."

Figuring the Odds

Hard work is nothing new to prevention proponents. It is often exceedingly difficult to determine how well any one measure works or how much it is really needed. In the realm of therapeutics, a patient generally requires immediate care, and his or her recovery gives some indication of whether a doctor's treatment succeeded. Prevention, on the other hand, applies to large numbers of ostensibly healthy people. No preventive action comes free from possible side effects. Hence, a physician's obligation to guarantee greater benefit than harm, or *primum non nocere*, becomes more stringent. For this reason, the medical community has come to hold prevention to a higher standard of proof than therapeutics. The increasing pressure to reduce medical spending has only upped the ante. "You had better not do something that is really expensive and has potential harm as well as benefit unless you're pretty darn certain it will make a difference," Sox says.

That difference is usually fairly clear for primary prevention, which encompasses counseling and immunizations. Such measures are generally inexpensive, and so whatever health they buy is a relative bargain. Indeed, workplace programs that encourage employees to exercise, eat properly and quit smoking, among other things, have been found to reduce a company's medical bills on average by some 20 percent, notes James F. Fries of Stanford University and the Health Project Consortium, a group of experts that annually reviews health-promotion efforts.

Similarly, the value of immunization is usually high. It comes into question only when the incidence of a disease drops so low that the odds of infection rival the risk that serious side effects will result from the vaccination. In the 1950s, for example, routine vaccinations had rid the U.S. of smallpox. A national survey published in 1967 estimated that the 14 million vaccinations given every year caused seven or

eight deaths and hundreds of complications, among them permanent brain damage. At the same time, no travelers had imported smallpox into the U.S. from elsewhere for nearly two decades. Some health officials proposed during the late 1960s that vaccinations should be withheld from the general public. As it happened, the U.S. discontinued routine smallpox vaccinations in 1971, four years after the World Health Organization had launched a campaign to eradicate the disease altogether. The WHO initiative succeeded in part because the vaccinations effectively lowered the risk of infection in the population.

"Primary prevention directly changes the probability that someone will develop a disease," notes David M. Eddy of Kaiser Permanente Southern California. "In contrast, most screening tests, often called secondary prevention, have no effect on the chances that someone will get a disease." The worth of such a test therefore depends on countless factors, including the false positive and false negative rates, the populations tested and the natural course of the disease. Also, unlike most primary preventive services, many screening tests must be repeated because certain diseases can arise at any time. "Intermittent screening is far and away the most complicated of all medical technology to analyze," Eddy notes. "At what age do you start screening, how often do you screen and when do you stop screening? The answers to all these questions depend on individual risk factors."

The current stew over cholesterol testing and treatment stems from such quandaries. "Most people whose cholesterol I lower will not benefit, because chances are they would not have suffered a heart attack anyway," says Robert H. Brook of the University of California at Los Angeles and Rand. Several clinical randomized trials have demonstrated that treatment lessens the incidence of heart disease among middle-aged men, but to date, no studies have considered whether anyone younger or older stands to gain. Moreover, some medical researchers speculate that the drugs used to treat high cholesterol can do harm. "Until that concern gets cleared up, it would be best to avoid using cholesterol-lowering drugs in people who are at little risk of developing heart disease in the first place," Sox says.

In addition, the OTA report noted that even for high-risk

Prenatal Care—A Questionable Bargain?

Public health experts justify large investments in prenatal care on the grounds that such prevention produces savings. They reason that doctor visits and screening tests during pregnancy reduce the number of low-weight births and thereby cut the costs of postnatal care. As evidence, they cite an array of analyses that estimate for every dollar spent on prenatal care anywhere from \$1.70 to \$3.38 is saved. Indeed, the Office of Technology Assessment concluded that prenatal care for poor women ultimately paid for itself.

Nevertheless, Frederick A. Connell of the University of Washington and Jane Huntington of the Group Health Cooperative argue that the cost-savings claims for prenatal care are not convincing. "Although there are many good arguments why expectant mothers should receive prenatal care," Huntington says, "the cost-savings argument is not one of them." Connell and Huntington targeted three areas in which the available cost-savings analyses of prenatal care had most often erred:



Prenatal ultrasound scanning

■ JUDGING THE IMPACT OF PRENATAL CARE

For ethical reasons, researchers could not randomly assign women to control groups receiving no prenatal care. In most studies the control groups therefore consisted of self-selected women whose health habits might have differed from those who sought prenatal care. Consequently, factors other than lack of prenatal care may have caused some babies to be born at low weights. In fact, recent analyses have not been able to demonstrate that prenatal care lowers the incidence of low birth weight.

■ ESTIMATING THE COSTS OF PRENATAL CARE

The price of prenatal care in the studies Connell and Huntington considered ranged from \$380 to \$1,042 per pregnancy. These studies assumed that routine prenatal care effectively prevented low-weight births. But more comprehensive monitoring and social support services, which undoubtedly cost more, may often be necessary to achieve the goal of preventing low-weight births, especially among low-income women. Moreover, these women often need assistance simply to see a doctor in the first place. None of the analyses took this expense into account.

■ CALCULATING THE SAVINGS

The estimated amounts saved by prenatal care in the selected studies ranged from \$347 to \$13,616 per pregnancy. But Connell and Huntington argue that one cannot predict savings based simply on the costs of caring for an "average" low-birth-weight baby. Very small infants are extremely expensive, whereas moderately low-weight babies may require only a few extra days in the hospital after delivery. Although the greatest savings can be achieved by avoiding the smallest birth weights, these occurrences are in fact the hardest to prevent.

groups, cholesterol-lowering programs appear to have no significant impact on overall mortality. In several studies the successfully treated men showed higher rates of death from violence, accidents, trauma, suicide and cancer than did untreated men. Perhaps they lived longer than they would have without treatment and therefore encountered unforeseen health hazards. Or maybe those men in the study who ate fatty foods were also more likely to smoke, drink or drive recklessly. Currently it is impossible to know whether a man who spends money and time to lower his cholesterol level has done anything to prolong his life, although he may be less likely to die from heart disease.

The preventive benefits of screening for prostate cancer are even less well

documented: the evidence collected so far does not prove that there are any. Yet many more men now seek screening than did a decade ago, when prostate-specific antigen (PSA) tests first became available. The test detects elevated PSA levels in the blood, which indicate that tumors may be present in the prostate. But many other conditions also raise PSA levels, and so the false positive rate is high. Thus, although the measure is quick and easy, it is not terribly reliable. "I have patients now, highly educated men, who, by God, want PSA screening tests," Fitzgerald says. "I tell them that the false positive risk is higher than the risk that they have prostate cancer, but they say they think the test is going to make them feel better."

In the short run, maybe a negative re-

sult would have that effect. But suppose it is positive—what then? Even if the result is correct, many of the microscopic tumors that the test can detect may never pose any problems. And starting treatment may not stop those tumors that would have progressed anyway. "There is no good evidence that the early detection of prostate cancer prolongs a man's life," Sox says. If anything, the sooner a man is diagnosed, the longer he might worry about his fate. Furthermore, the early detection of any cancer triggers a series of follow-up measures, from biopsies and prophylactic surgery to drug and radiation therapies—none of which come cheap.

"If you are going to send millions of healthy, happy people marching off for screening tests, taking time off from

work or play, looking for parking spaces and being subjected to the chance of false positive test results, you ought to have reason to believe they will come out ahead," Eddy adds.

Hedging Bets

Sometimes even mounds of data give little proof of some preventive measure's impact. Consider breast cancer, the second leading cause of cancer deaths among American women. It consumes more health care dollars than any other cancer—\$6.5 billion in 1990 alone. The results from several major trials consistently credit mammography with averting 30 percent of breast cancer deaths among women aged 50 to 70. Yet most also demonstrate that younger women gain very little. Even if a woman younger than 50 years discovers she has breast cancer through screening, the odds that she might die from the disease remain more or less the same.

"More or less" is the hitch. The relative risk of mortality has in fact decreased slightly in those investigations that have followed 40-something women a little longer. Nevertheless, there is no way to be certain that this apparent drop amounts to anything more than chance. A group at the University of California at San Francisco recently combined the numbers from several major randomized controlled trials and found that mortality among screened women decreased only after 10 to 12 years.

Perhaps women in their forties who are followed for 10 to 12 years are then simply women in their fifties and sixties, who do benefit from mammography, explains Karla M. Kerlikowske of U.C.S.F. At that age, their breasts contain more fat and less dense breast tissue, so mammogram machines can image small tumors hidden in them more easily. Perhaps had only the best technicians and radiologists—armed with the latest equipment—participated in the trials, they would have detected more breast cancers in younger women. Perhaps no matter what screening techniques were used, early detection would do little to lower mortality rates among women who acquire breast cancer in their forties because their tumors readily metastasize to other sites. The point is that the evidence to justify screening women younger than 50 for breast cancer using mammography is weak at best.

Making such distinctions is critical in developing economically rational health care policies. Sometimes the difference between screening liberally or selectively adds up to small change; other times not. In 1990 the OTA concluded that

	FOR ROUTINE USE		FOR MORE SELECTIVE USE	
	Decision	Criteria	Decision	Criteria
INTRAOCULAR PRESSURE	Yes	Over age 65	Not considered	—
STOOL FOR OCCULT BLOOD	No recommendation (C)	Over age 50	Yes (C)	Over age 40 if first-degree relative has colon cancer; history of inflammatory bowel disease or endometrial, breast or ovarian cancer
SIGMOIDOSCOPY	No recommendation (C)	Over age 50, every 3 to 5 years	Yes (C)	
URINALYSIS	Yes (C)	Over age 60	Yes (C)	Diabetes
SERUM CHOLESTEROL	Yes	Over age 18, every 5 years	Yes	More frequently if at risk for coronary artery disease
MAMMOGRAPHY	Yes (A,B)	Age 50 to 75 or beyond, yearly	Yes	Age 35 to 40 if premenopausal first-degree relative has breast cancer
CERVICAL CYTOLOGIC SCREENING	Yes	Age 20 to 65 or beyond, every 1 to 3 years	Yes	Yearly if at high risk; if over age 65 and not tested in previous 10 years
FASTING PLASMA GLUCOSE	No	Over age 50	Yes	Familial diabetes; history of gestational diabetes; obesity
THYROID TESTING	No	—	Yes	Women over age 60
HIV SEROLOGIC SCREENING	No (C)	—	Yes (B)	Over age 18 if at high risk; transfusion recipients, 1978–85
SYPHILIS SEROLOGIC SCREENING	No	—	Yes (B)	High-risk sexual behavior
RESTING ELECTROCARDIOGRAPHY	No	—	Yes	Over age 40 if at risk for coronary artery disease or starting an exercise program
EXERCISE STRESS TEST	No	—	Yes	
TUBERCULIN TESTING	No	—	Yes	If exposed; if HIV infection present; if taking high-dose steroids
BONE MINERAL ANALYSIS	No	—	Yes	Early and perimenopausal women; slender women; women considering estrogen but only because of osteoporosis
COLONOSCOPY	No	—	Yes	If more than one first-degree relative has colon cancer; history of ulcerative colitis for 10 years or more or adenomatous polyps or familial polyposis
CHEST RADIOGRAPHY FOR LUNG	No	—	No	Cigarette smokers
HEMATOCRIT	No	—	No recommendation	Low socioeconomic status; institutionalized elderly persons
PROSTATE-SPECIFIC ANTIGEN	No (C)	Over age 50	Not considered	Over age 40 if at increased risk

LISA BURNETT

RECOMMENDATIONS from the U.S. Preventive Services Task Force are summarized in the chart shown. Tests that the task force did not evaluate are marked not considered; tests on which it could reach no conclusion are labeled no recommendation. Some recommendations also have letter grades, indicating the task force's assessment of the strength of the evidence on which it was basing its decision.

saving one year of life among women aged 65 and older by screening them for cervical cancer becomes seven times more expensive if the tests are done every year rather than every three. In the case of breast cancer, screening women in their forties costs six times more to save a year of life than does screening women in their fifties, according to calculations done recently at Rand.

"Organizations responsible for paying for these tests must understandably have a very different perspective from advocacy groups that bear no responsibility for costs. The American Cancer Society can glibly recommend that women in their forties have regular mammograms because it has no fixed budget and doesn't have to worry about other worthwhile activities that will not be possible if resources are pulled away," Eddy states. "The cost-effectiveness of a screening test changes dramatically depending on how you define it—by age, frequency, the risks involved and many other things. The trick is to find from innumerable options those that are most cost-effective."

The Rand study set about designing a benefit package for the detection and early treatment of breast cancer based on a meta-analysis of the cost-effectiveness of mammography, primary surgery, adjuvant chemotherapy and follow-up care. Costs were summed and

"...we already ration care in this country on the basis of access, insurance and knowledge. I'd rather try to do it on the basis of scientific evidence."

comparisons made for a hypothetical health care organization of 500,000 people, in which 360 new breast cancer cases would crop up annually. The team concluded that it made most sense to cover mammograms only for women between the ages of 50 and 69.

Like most other researchers, the Rand group reported that mammography had no statistically significant effect on mortality from breast cancer among women younger than 50. Similarly, mammography barely changed overall mortality among women older than 69. The Rand authors also noted that chemotherapy reduced overall mortality among women younger than 50 by some 25 percent. Hence, they concluded that more younger lives could be saved by treating breast

cancer with chemotherapy after it had manifested than by trying to catch it beforehand using regular mammography. "As chemotherapy improves, the utility of mammography decreases," Brook says.

Granted, most women would rather head off breast cancer before chemotherapy is necessary. But most health organizations would rather pay for something that keeps the most women alive. "When you're designing a benefits package and you have limited resources, you have to make these kinds of decisions," Brook comments. "If a woman is 45 and wants a mammogram and to pay for it herself, that's fine. I have no problems with how people want to spend their money. At the same time, we already ration care in this country on the basis of access, insurance and knowledge. I'd rather try to do it on the basis of scientific evidence."

Cashing Out

To be certain, quantifying the outcomes of medical practice is not yet an exact science. Cost-effectiveness analysis, which compares the outcome of some procedure—typically the number of years of life it might save—to its price may be the least mature branch of the field. "Cost-effectiveness studies are subject to criticism for two reasons," Sox says. "First, many use different

Inconsistent Analyses of Health Costs

Determining the cost-effectiveness of a medical service—the number of years of life it saves and at what price—is exceedingly complex. Researchers currently tackle the problem by a variety of methods. Some studies involve randomized, controlled trials, whereas others are based on case findings alone. Some analysts tally only the years of life saved through a particular procedure; others also consider the quality of those years and devalue any that are marred by side effects.

Finding a way to carry the costs and savings associated with prevention on an accounting ledger is also a challenge. "If you don't spend some amount of money on health today, you could invest it, and you would have more at some point in the future," explains Louise B. Russell of Rutgers University, "so we take that into account." Yet no one discount rate has been adopted by all analysts. Moreover, future health benefits are not always discounted at the same rate. "If you do not discount future health benefits and costs equally, you get

strange results," Russell notes. "For example, it will appear as though it is always better to wait to spend health care dollars."

Furthermore, many analyses overlook certain costs. To illustrate, Russell describes a 1968 survey of measles vaccinations in her book, *Is Prevention Better*

COST-EFFECTIVENESS OF MEASLES VACCINATION	MILLIONS OF DOLLARS		
	WITHOUT VACCINATION	WITH VACCINATION	DIFFERENCE
DIRECT MEDICAL EXPENSES			
Administration of vaccine	-----	108.3	-108.3
Treatment for measles cases	174.3	96.8	77.5
Treatment for vaccine side effects	-----	NA	NA
SUBTOTAL	174.3	205.1	-30.8
NET EXPENDITURES FOR INSTITUTIONAL CARE			
For persons mentally retarded by measles	499.7	298.8	200.8
For persons mentally retarded by vaccine side effects	-----	NA	NA
TOTAL	674	503.9	170

LAURIE GRACE

methodologies—that's fixable. But also, we know how to figure the medical care costs of adding an extra year of life through screening. Yet we have not figured out how to place a dollar value on that benefit." For comparing interventions aimed at the same goal—say, reducing breast cancer mortality—this does not pose too much of a problem: because the desired outcome is held constant, investing in one option over another makes sense, provided it returns higher gains.

Trying to comparison shop between very different medical services, on the other hand, proves treacherous. Is it better, for example, to spend \$2 million on prenatal care instead of bone marrow transplants, given that the unborn infants will on average have a much better chance for survival than the leukemia patients? Most people would find such a proposition morally repugnant. Yet health officials in Oregon made a similar choice eight years ago. An 11-member commission ranked the cost-effectiveness of 695 conditions and treatments. The state legislature then allocated enough funds to cover the top 565 items on the list. Bone marrow transplants were cut.

Such comparisons are problematic in part because not everyone agrees on how much years of life are worth, Sox says. "Most people consider paying \$100,000 to save one additional year of

life too much but consider \$25,000 per year of life saved reasonable," he notes. "The difficulty is that we have no logical, rational basis for setting those boundaries." Moreover, different people draw different boundaries. An older person suffering from some chronic condition might not value another year as much as would a young, healthy person.

To address these highly contentious issues, the Office of Disease Prevention and Health Promotion has established a panel to standardize the design and use of cost-effectiveness analysis. "One of the main goals is to be able to look across a lot of different interventions, some of which have their primary effects on the quality of life rather than the length of life," says Louise B. Russell of Rutgers University, co-chair of the panel. "Units such as quality-adjusted life years (QALYs)—instead of just years of life saved—combine those into one summary measure."

In essence, this unit discounts the value of an added year to reflect any undesired side effects. An extra year of perfect health would be worth one full QALY. But one plagued by arthritic pain might be valued at 80 percent of a QALY, and a year on a ventilator, only 67 percent. To determine the right numbers, several researchers have surveyed the public about how much different conditions detract from the quality of their life. But the concept needs work. "How you capture quality isn't standardized at the moment," comments Marthe R. Gold of the Office of Disease Prevention and Health Promotion. "Cost-effectiveness is still enough of a developing science that at best it can serve as a decision-making tool—but not as the justification for the decision."

Financial issues aside, the crux of the debate is about the burden of proof. "The traditional school of thought [on prevention] was, when in doubt, do it," Eddy remarks. "The burden had always been to show that something had no benefit." When medical recommendations involve only individual doctors and patients—as they normally have for the past two millennia or more—it is generally worth taking the chance that some precaution might help one person. Massive screening programs, though, necessarily demand a different take. "If you screen women in their forties for breast cancer, you are affecting the lives of some 18 million healthy people," Eddy says.

Such large prescriptions are forcing physicians to serve a dualistic role, Fitzgerald comments. "They must care for individual patients who are sick and at



DAN WAGNER

the same time look out for the commonweal. Yet those two interests are diametrically opposed." Indeed, the sick—especially those who might have avoided their condition—are increasingly viewed as a burden on society. In this sense, the expansion of preventive care has revised our cultural definitions of health. "Our paradigm now is that good health is normal," she remarks, "but that simply isn't true from all the historical and empirical evidence available."

Perhaps the success of preventive medicine has raised society's expectations too high. Fitzgerald points out that the influenza pandemic starting in 1917 killed more people in the world in 16 months than HIV has killed since its identification. At that time, epidemics—similar to cases of alcoholism, lung cancer or obesity—were sad but regular occurrences. Now that medicine enables us sometimes to avoid such conditions and a growing list of other horrors, too, Americans are less content to accept that the list is not complete.

"Our cultural definition of disease has become one in which it is the result of some personal lapse or the lapse of some external agency," Fitzgerald remarks. That belief, however, is unrealistic. No matter how appealing the idea, prevention alone can eliminate neither medical inflation nor mortality. It does hold great potential to limit suffering and pain, and for that reason, worthwhile services should be equally available to all citizens. But prolonging life comes at a price. The test for society will be how high a price it is willing to pay.

Than Cure? [see table]. The investigation had concluded that routine vaccinations consumed \$31 million during the preceding five years. At the same time, however, the nation had saved \$201 million because fewer children were left brain-damaged by the disease.

The outlay for treating the vaccine's side effects was not included, however—an oversight that very likely sold short the total cost of vaccinations. The study also probably overestimated the costs associated with the institutionalization of sick children by failing to subtract many expenses—such as food and clothing—that the children would have incurred living at home. All told, the savings from measles vaccinations were no doubt somewhat less than \$170 million, Russell concludes.

Until standard methods for calculating cost-effectiveness ratios are developed, it will remain extremely difficult for health care providers to compare disparate studies of the same procedure or like analyses of very different interventions.



THE AMATEUR SCIENTIST conducted by Ricardo Chiv'inglese

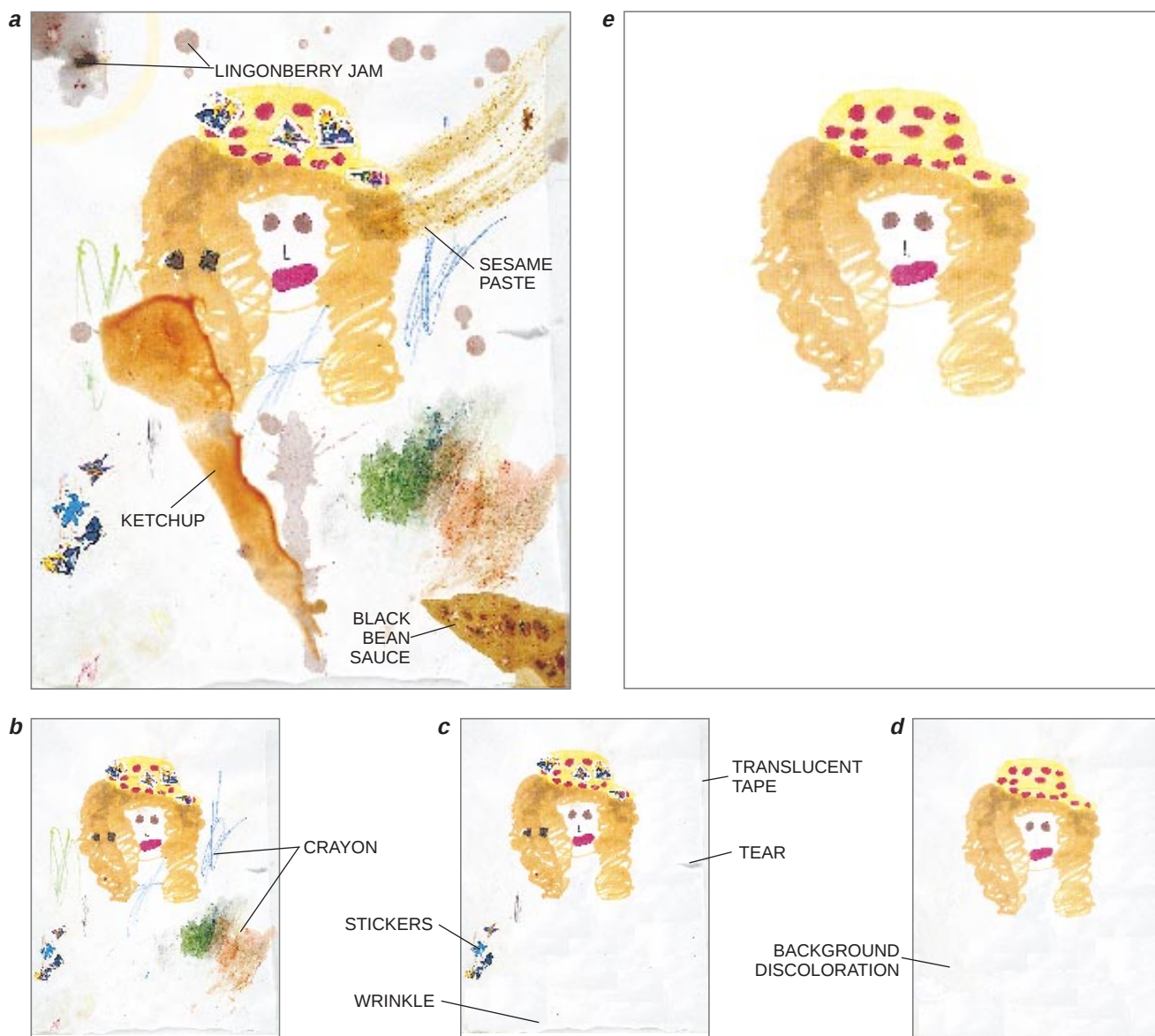
Computerized Restoration of Juvenile Art

Although computers have found their way into home and laboratory during the past three decades, they have only recently entered the studio. Increases in computing speed and greater sophistication of software have helped artists to restore as well as to analyze works of art [see "The Art Historian's Computer," by Lillian Schwartz, page 106]. Although nonprofessionals

may not have access to the complex multispectral image-analysis equipment used by specialists, the falling cost of personal computers and image-manipulation software is bringing electronic art-restoration techniques within the reach of many amateurs.

This trend is a welcome one because most households with young members possess many original works of art but

do not have the facilities for protecting them from physical decay. Indeed, typical display sites expose paintings and drawings to diverse dangers. A mixed-media work left hanging on the refrigerator, for example, may be exposed to excessive sunlight, wide swings in temperature and humidity, grease vapors, food spatters or even sibling vandalism [see *illustration below*]. Furthermore, few domestic artworks are created on acid-free, archival paper; this shortcoming renders them especially vulnerable



ELECTRONIC ART RESTORATION starts with removal of stains, rips, vandalism and other damage (a-c). After the correction of background color shifts (still visible in d), the resulting image (e) is essentially pristine.

JAMIE S. FEIGENBAUM

to a wide range of environmental insults.

Removing damage requires a range of image-manipulation techniques. First, correct the image background: yellowing and uniform deposits of dust, dirt and grease can be removed by shifting color and brightness so that the original tone returns. (You may want to obtain a fresh sample of the paper or other medium for comparison.) Wax-based pigments tend to be lightfast, but watercolors and markers fade, so their tints will need to be corrected as well. Each point in the digitized painting is represented in terms of the primary colors cyan, magenta, yellow and black (or red, green and blue), and so it is simple to calculate these values for the background colors, define a mathematical function that will map each color to its corrected value and then instruct the image-manipulation program to apply this function.

Most of the rest of the repair work—erasing stains and signs of physical damage—is simply a matter of replacing the lost colors in

the affected patches with their counterparts from surrounding areas. Fortunately, most young artists do not employ a complex palette, so the choice will be limited. (Certain artists working with wax-based pigments, however, may have utilized all 64 colors.) If the damage spans the boundary between two colors, you may have to re-create the demarcation by hand; many image-manipulation programs offer “splines” and other tools for generating smooth lines. In cases where the loss is particularly extensive, restorers may also have to re-create tonal gradations or patterns.

Sculptures and installation pieces, which are also highly vulnerable, may benefit from three-dimensional rendering programs that can recover the original form. For example, humidity changes may cause parts of a pasta bas-relief to detach from their backing. Careful measurement of the glue marks, com-

bined with finite-element models of the appropriate macaroni shapes, will permit a virtual reconstruction. (The full capabilities of such programs, however, are beyond the current discussion.)

Extensive restoration requires taking into account the artist's original intent, which may be elicited by direct consultation (“Is that a tree?”) or deduced by comparison with similar works in the corpus. At this stage, you may be able to remedy not only adverse environmental impacts but also shortcomings in the available media as well: blunt-tipped markers, crayons or pencils unable to produce the fine lines that the artist intended or perhaps the absence of a particular color from the palette.

Ultimately, the reworking of such art may reveal images that a young artist may have envisioned but was unable to put into form. Obstacles to complete realization of the artistic impulse can include a lack not only of tools but also of hand-eye coordination and of technical training. All these difficulties can be remedied by the electronic restorer.

In the bottom illustration, for example, the artist did not have access to a range of earth tones and grays in making the original drawing. Because she would have used them had they been available, the work was corrected to compensate. I recognized that the bold, sweeping strokes of green were obviously meant to evoke the swaying of tall grass in an open field. To bring out this intention, I made a series of additional refinements: modifying the grass to add finer lines and more subtle shading than a felt-tip marker could produce, removing *pentimenti* (lines she had crossed out and replaced elsewhere) and finally altering the proportions and the composition to achieve a more realistic effect. Executing these corrections on the figure produced the scene the artist would have painted had she the proper tools, motor coordination and artistic experience [see top illustration in pair at left].

Considerable controversy attends such restorations. Nevertheless, I believe careful study of the artist's corpus, combined with analysis of the school in whose tradition she is working—be it Impressionist, Expressionist, Hudson River, Pre-Raphaelite, Neo-Primitive or Elementary—may yield both aesthetic dividends and a deeper understanding of the many influences that shape juvenile art.

RICARDO CHIAV'INGLESE is an adjunct lecturer in the Interpretive Technology Program at the Freon Institute in New York City.

HIDDEN BEAUTY can be brought out in juvenile artworks by alterations that replace the original marker, crayon or brushstrokes (bottom) with what the young artist almost certainly would have intended (top).



ANDREW WYETH, "Christina's World," The Museum of Modern Art



JAMIE S. FEIGENBAUM



BOOK REVIEWS by Daniel L. Schacter

Memory Wars

MAKING MONSTERS: FALSE MEMORIES, PSYCHOTHERAPY, AND SEXUAL Hysteria, by Richard Ofshe and Ethan Watters. Charles Scribner's Sons, 1994 (\$23). **THE MYTH OF REPRESSED MEMORY: FALSE MEMORIES AND ALLEGATIONS OF SEXUAL ABUSE**, by Elizabeth F. Loftus and Katherine Ketcham. St. Martin's Press, 1994 (\$22.95). **VICTIMS OF MEMORY: INCEST ACCUSATIONS AND SHATTERED LIVES**, by Mark Pendergrast. Upper Access Books, 1995 (\$24.95).

A debate is raging—in courtrooms, in journals and in the popular press—about the validity of recovered memories of long-past events. Because these memories often involve sexual abuse and other horrible experiences, the dispute has stimulated broad interest in what is known about how the mind records events. Fascination with the workings of memory is nothing new, of course. Aristotle offered a wide-ranging discourse on memory more than 2,000 years ago, and numerous philosophers since have pondered the mind's capacity to travel backward in time. But, as recent events reaffirm, science has long had a hard time grappling with that remarkable ability.

Serious research into the nature of memory did not begin until 1885, when the German psychologist Hermann Ebbinghaus applied scientific method to the analysis of memory. The past decade has been particularly exciting, as scientists from a number of disciplines have begun to develop a thorough understanding of memory. The new analyses range from detailed models of how experiences are recorded to broader theories about the brain systems involved in various forms of memory.

Those of us studying human memory feel optimistic that we are finally at the threshold of understanding some of the deepest enigmas of the mind. Yet this exploration has also developed a dark side. During the past half a dozen years, there has been an explosion of cases in which adult men and women—most frequently, young women undergoing psychotherapy—have seemingly remembered childhood sexual abuse that they had forgotten for years or even decades. Those memories include everything from single episodes of inappropriate touching to years of rape

and torture; parents or other close family members are typically recalled as the perpetrators. Patients frequently become overwhelmingly convinced of the reality of their recovered memories; the accused often deny the memories with equal fervor.

Are these memories accurate recollections of terrible traumas or phantoms of events that never happened? Have therapists developed effective new memory-retrieval techniques, or have they employed misguided procedures that actually help to create the memories? And are the patients who recover memories of sexual abuse being empowered to speak out, or are they being diverted from the problems that brought them to therapy in the first place?

These questions have sparked passionate, sometimes acrimonious disputes. Memory researchers have been astonished to find their world of experiments, theories and laboratory paradigms intertwined with high-profile issues of incest, new-age psychotherapy and even satanic cults.

The books discussed in this review delve into the recovered-memories debate from diverse perspectives. Richard Ofshe (*Making Monsters*, co-authored with writer Ethan Watters) is a social psychologist who has studied and pub-

lished about social influences and cults. Elizabeth Loftus (*The Myth of Repressed Memory*, co-authored with writer Katherine Ketcham) is a memory researcher who has made numerous pioneering contributions to the cognitive study of memory distortion and suggestibility. And Mark Pendergrast (*Victims of Memory*) is a writer and journalist who is himself an accused parent. Despite their disparate backgrounds, the authors share a critical view of what they refer to as recovered-memory therapy.

Ofshe and Watters adopt the most confrontational approach. The scathing tone of *Making Monsters* surfaces in the first paragraph of its preface: "Our goal is to prove beyond doubt that devastating mistakes are being made within certain therapy settings.... This work is intended as an exposé of a pseudoscientific enterprise that is damaging the lives of people in need." Ofshe and Watters lay the blame squarely at the feet of incompetent, even morally reprehensible therapists. "If, for no defensible reason," they write, "some therapists are causing the same emotional and psychological trauma as an actual rape or sexual assault, then they, like those who physically victimize people, deserve moral condemnation."

These are serious charges. In support



LORIE NOVAK, "Fragments," 1987. Jayne Baum Gallery (all photographs from collection of D. Schacter and S. McGlynn)

of their case, Ofshe and Watters begin with a cursory review of experiments showing that suggestive influences can alter a memory and that people can have vivid recollections of events that never happened.

No laboratory studies have ever attempted to demonstrate the possibility of implanting false memories of sexual trauma. Indeed, it seems safe to assume that such studies never will, because it would be unethical for a researcher to attempt to do so. Ofshe and Watters consider perhaps the closest analogue, an experiment published in 1991 by the late Canadian researcher Nicholas Spanos in which he hypnotized his subjects, "regressed" them to "past lives" and suggested to some that they could have been abused in a past life. Spanos found that those who received this suggestion were later more likely to "remember" being abused than those who had not.

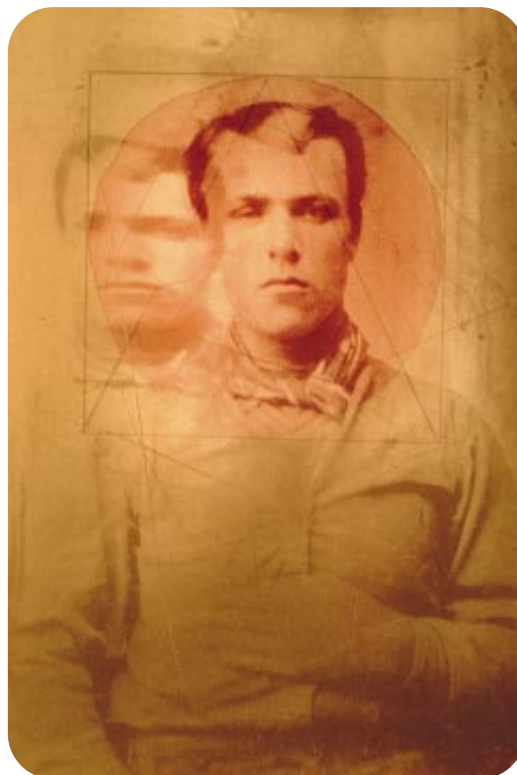
The authors of *Making Monsters* also examine the controversial memory-retrieval techniques (including hypnosis and visualization) advocated by some recovery therapists. Ofshe and Watters discuss the mounting evidence that hypnosis offers a potent method for inducing compelling but inaccurate pseudomemories. They also effectively criticize therapy techniques that involve visualizing or imagining abusive incidents as a first step toward remembering them. Ofshe and Watters argue that a therapist who believes in the reality of forgotten abuse can help validate imagined experiences as bona fide memories.

Therapists sometimes infer that forgotten abuse has occurred based on symptoms that advocates of recovered memory consider telltale signs of abuse—attributes ranging from low self-esteem and depression to avoidance of mirrors and the desire to change one's name. As Ofshe and Watters assert, there is little evidence to link such behaviors to a history of actual abuse. Moreover, many of the alleged signs of abuse are observed in people who were almost certainly not abused.

Ofshe and Watters also draw on wrenching case studies of patients who have recovered memories in therapy, some of whom later retracted those memories. We will probably never know what truly happened in these situations, so their usefulness as scientific data is questionable. Ofshe and Watters examine the most extreme cases, in which patients recount memories of extended, horrific abuse only after coming under

the influence of a therapist who uses suggestive techniques to hunt for repressed memories. Such recovered memories, the authors judge, are most likely spurious.

They make this point most convincingly when discussing recovered memories of satanic ritual abuse, which they characterize as "the Achilles' heel of the recovered memory movement." Many therapists have reported on patients who have clearly recalled savage acts carried out by satanic cults: rapes, murders, cannibalization of fetuses and re-



CHERYL CALLERI, "Fugitive Memory III," 1993 Ruth Bachofner Gallery

lated atrocities. Yet in most instances, no memories of ritual brutality existed prior to therapy, and no one has produced hard evidence of such acts. Ofshe and Watters note that investigations by the Federal Bureau of Investigation of more than 300 cases have failed to turn up any proof.

The lack of empirical support does not necessarily mean that no satanic cults exist or that no ritual abuse has ever occurred. But Ofshe and Watters's central claim—that recovered memories of ritualistic horrors are very likely to have been created during therapy—is convincing. They also link therapist-induced pseudomemories with the recent explosion in diagnoses of multiple-personality disorder, which was once thought to be exceedingly rare. The authors claim that such personalities are often fabricated in therapy and discuss

a case in which a therapist supposedly discovered that a patient was suffering from multiple personalities and repressed memories of satanic-ritual abuse. Once the patient ended therapy, she retracted her memories and abandoned her personalities.

In an attack on the underpinnings of recovered-memory therapy, Ofshe and Watters attempt to discredit the notion that there is a special mechanism, which they label "robust repression," that could cause someone to forget completely about years of repeated sexual trauma. The authors distinguish robust repression from the weaker mental repression (which some psychologists refer to as suppression) that occurs when people consciously avoid thinking about unpleasant experiences. The idea that people sometimes deliberately put aside painful recollections is not controversial. Such conscious avoidance could decrease the likelihood that someone would later remember the unpleasant experience, because it would not benefit from the postevent rehearsal—thinking and talking about the past—that ordinarily strengthens memories. But the mere lack of rehearsal probably cannot produce profound amnesia for traumatic events that happen repeatedly for years. A more potent mechanism is required; this is where the concept of robust repression comes in.

Ofshe and Watters effectively criticize the concepts of robust repression advanced by several advocates of recovered memory. The authors refer to a review article by David Holmes of the University of Kansas that concludes that there is no good experimental evidence for repression. The laboratory studies considered by Holmes necessarily use relatively sterile manipulations, however. It is difficult to convince even a laboratory researcher such as myself that these studies bear more than a remote relation to real emotional traumas.

There is also some information notably missing from *Making Monsters*. Ofshe and Watters do not discuss the extensive literature on psychogenic amnesia, whereby traumatic events temporarily blot out certain memories, ranging from single episodes to an individual's entire personal past. In addition, because the authors focus on extreme cases that involve forgetting years of ongoing trauma, they say little about the possibility that a person could forget and recover memories of a single abusive episode or a few such episodes. They touch on this important question

only in an appendix. Moreover, Ofshe and Watters do not seriously address forgotten memories of abuse that may be recovered outside of therapy.

Ofshe and Watters apply their conclusions broadly, indicting the entire "recovered memory movement" of faulty reasoning and hazardous practices. The judgment is justified when applied specifically to the material in the book. The authors' grand generalization is unconvincing, however, because they never define exactly who is part of the recovered-memory movement, nor do they provide much evidence that their conclusions deserve such wide application. Even if some recovery therapists have engaged in dangerous practices, it is possible that genuine cases of forgetting and recalling sexual abuse do exist.

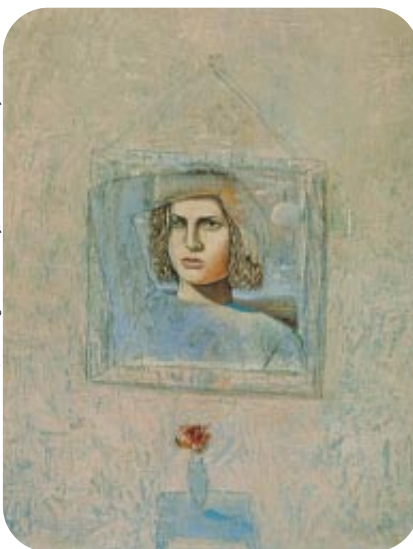
Loftus and Ketcham's book, in contrast, is entirely devoid of the contemptuous tone adopted by Ofshe and Watters. *The Myth of Repressed Memory* is written in Loftus's first-person voice and frequently takes on an autobiographical quality as she relates her experiences as an expert witness, memory researcher and participant in recovered-memory debates. Her outspoken views on the malleability of memory have made Loftus something of a lightning rod in these disputes. At a memory symposium held in Boston last year, a group of incest survivors picketed her appearance. Yet in *The Myth of Repressed Memory*, Loftus comes across as genuinely interested in trying to understand the other side of the issue.

In a revealing interlude Loftus recounts a meeting with Ellen Bass, co-author of the controversial "bible" of the recovery movement, *The Courage to Heal*. This book has been disparaged by virtually every critic of recovered memory for its sweeping and unsubstantiated claims (such as an extraordinary admonition to those who have no memory of abuse: "If you think you were abused and your life shows the symptoms, then you were.") Nevertheless, Loftus's account of her conversation with Bass reflects a good-faith attempt on both sides to consider the perspective of the other. Although the exchange ends at something of an impasse, it illustrates the kind of dialogue that is needed to resolve the controversies surrounding recovered memories.

Loftus also uses her encounter with Bass to make a crucial point. Loftus states that she does not dispute the validity of abuse memories that have never been lost, nor does she rule out the possibility that people can forget and later reclaim some memories of abuse. She is concerned primarily with the lack

of scientific evidence that extensive, severe sexual trauma can be pushed into the unconscious through a special mechanism of memory repression. Clinical and laboratory research indicate that emotionally traumatic experiences tend to be well remembered.

Loftus does briefly consider the phenomenon of psychogenic amnesia, describing a case that my colleagues and I reported some years ago, which involved a young man who temporarily forgot almost all his personal past after a traumatic experience. She correctly points out several differences between this kind of amnesia and the kind of forgetting implicated in cases of recovered memory—differences that limit the ex-



CANDACE WALTERS, "Looking for Yesterday," 1992. Clarke Gallery

tent to which one can serve as a model for another. In a disappointing omission, however, Loftus fails to discuss the studies specifically concerning loss of memory of sexual abuse, even though she lists several in her bibliography and has published such a study herself.

The Myth of Repressed Memories is most effective when it delves into substantive topics in depth. For example, Loftus gives a detailed critique of the memory-retrieval techniques advocated in a popular book by recovery therapist Renee Fredrickson. These techniques include hypnosis and visualization, as well as methods for contacting "body memories" (memories that are alleged to have been stored in body tissue) and procedures such as a "quick list," in which patients jot down whatever comes to mind when pondering possible abuse without attempting to assess the accuracy of the retrieved thoughts. Loftus notes that there is no scientific documentation of the efficacy of these techniques but good reason to believe that they pose a danger because

they encourage patients to blur the line between imagination and memory. Indeed, I was so surprised that a therapist would advocate such techniques that I checked the original source to determine whether Loftus had portrayed Fredrickson's approach fairly; she had.

I would have welcomed more such thorough treatments of the key issues in the recovered-memories debate. Although the autobiographical material in *The Myth of Repressed Memories* and Loftus's moving descriptions of families shattered by recovered memories make for gripping reading, they tend to crowd out the kind of rigorous analyses of central disputes that are so sorely needed in this minefield.

Personal experience lies at the very core of the third book, Mark Pendergrast's *Victims of Memory*. Pendergrast attempts to integrate an insider's account of the recovered-memories controversy with a scholarly analysis of it. He begins the task with two daunting strikes against him: he has been accused of unspecified abuse by his two daughters—charges he denies—and he is a journalist who has no credentials in psychology or psychiatry. He is therefore readily perceived as someone who has an ax to grind and little else to offer. In an impressive display of scholarship and sheer determination, Pendergrast has surmounted these obstacles to write a comprehensive treatment of the recovered-memories controversy.

Victims of Memory covers much of the same territory as do the other two books and offers many of the same arguments, but Pendergrast offers a broader portrayal of the social and cultural contexts of the recovered-memories phenomenon. His treatment is also distinguished by some welcome historical perspective. For instance, he describes the evolving role of therapeutic suggestion in the genesis of multiple-personality disorder, and he exhibits throughout a flair for digging out relevant quotes from pioneering psychologists and psychiatrists.

Pendergrast demonstrates a laudable ability to lay out all sides of the argument. He analyzes evidence for and against repression, carefully acknowledging the limitations of laboratory research but offering a thoughtful examination of studies concerning the forgetting of sexual abuse. He also considers several cases in which memories of abuse resurfaced without therapy. Citing the literature on psychogenic amnesia, Pendergrast concedes that people can sometimes forget traumatic experiences. But he rightly points out that evidence from such cases must be treated cautiously because in some instances

SCIENTIFIC AMERICAN

**COMING
IN THE
MAY
ISSUE...**

THE EVOLUTIONARY DYNAMICS OF HIV INFECTION

**Martin Nowak
and Andrew McMichael**
University of Oxford

DID BOHR "SHARE" NUCLEAR SECRETS?

**Kurt Gottfried
and Hans A. Bethe**
Cornell University
Ronald Z. Sagdeev
*University of Maryland,
College Park*

TRENDS IN ARCHAEOLOGICAL PRESERVATION

Marguerite Holloway
Staff writer

**ALSO
IN MAY...**

The Global Tobacco Epidemic

Binary Neutron Stars

The Ocean's Salt Fingers

Silicon Microstrip Detectors

Niels Bohr and the Drawing

**ON SALE
APRIL 27**

amnesia may be feigned deliberately.

Much of *Victims of Memory* is devoted to interviews with incest survivors, therapists, retractors and accused parents. Although it is difficult to draw general conclusions from a small and selected sample, Pendergrast allows the reader to see the issues from a variety of perspectives. Despite the author's personal situation—his daughters' allegations originated in therapy—Pendergrast renders a sympathetic portrayal of recovery therapists as well-intentioned but misinformed players in a drama that has veered out of control.

In the end, however, Pendergrast's critique of recovered-memory therapy is no less damning than that of Ofshe and Watters. He extends his discussion of suggestion-induced false recollections to include bizarre but fascinating cases in which people "remember" being abducted by aliens, which demonstrate the power of hypnosis to induce intense but inaccurate memories.

Pendergrast also links the way that therapists interpret patients' symptoms with contemporary research on implicit memory—that is, nonconscious effects of experience on subsequent behavior and cognition. Some therapists have cited this research (which is a major focus of my own work) as justification for interpreting their patients' fears, dislikes or attractions as unconscious "memories" of abuse. Although such an interpretation could possibly be valid in some cases, inferring the existence of implicit memories is a complex process, so alternative explanations need to be scrupulously considered. As Pendergrast notes, the fact that implicit memory has been established in controlled experiments "does not mean that a woman who hates bananas is necessarily reacting subconsciously to a memory of her father's erect penis, as many trauma therapists believe."

Pendergrast occasionally misses his targets. In a section entitled "Scientists as True Believers," he critiques an article co-authored by the eminent neurobiologist Eric R. Kandel that considered possible neurobiological bases for repressed memories. Kandel acknowledged, however, that false memories can be created and is hardly a "true believer" in recovered-memory therapy. Fortunately, Pendergrast is rarely so sloppy. The book concludes with a moving letter to his daughters (he no longer knows where they live or what their names are); the reader cannot help but hope that reconciliation is still possible.

Where does all this leave us in attempting to make sense of an important and painful issue? Perhaps the key point is that the standard depiction of

this debate—proponents of recovered memory versus advocates of false memory, winner take all—is simplistic and needlessly divisive. Understanding the current situation requires distinguishing among several intertwined questions.

First, there is the question of whether false recollections can originate in therapy. Extensive laboratory research indicates that suggestion and other factors can lead to memory distortion. A startling number of patients "recover" memories of satanic-ritual abuse despite an absence of evidence for such abuse. Hypnotically based therapy has helped induce recollections of exceedingly improbable events (such as past lives and alien abductions). And a growing number of people have retracted their recovered memories. Taken together, these considerations lead inexorably to the conclusion that some recovery therapists have helped create—probably un-

*The standard depiction
of the recovered-memories
debate is simplistic and
needlessly divisive.*

wittingly—pseudomemories of sexual abuse that never occurred. The phenomenon may be too widespread to attribute to a few bad-apple therapists, but it is unjustified to indict the entire field of psychotherapy for the excesses of some practitioners.

At the same time, I cannot emphasize too strongly that the foregoing conclusions do not imply that all recovered memories are inaccurate. A proper evaluation of recovered memory phenomena will require further inquiry into the evidence for robust repression, which currently lacks credible scientific support. Studies of psychogenic amnesia, though inconclusive, indicate that traumatic experiences can sometimes lead to extensive forgetting; researchers need to analyze more fully the conditions that produce memory loss.

Whether or not robust repression exists, there is the related question of whether individual incidents of abuse can be forgotten—and it seems clear that they can. Such forgetting could result from ordinary processes of memory decay or interference and might be exacerbated by conscious avoidance of the trauma and consequent lack of rehearsal. Another question is whether such forgotten incidents can later be recalled accurately. Although there is little well-documented evidence from clinical research for recovery of accu-

rate memories of abuse, several credible cases have been reported.

Further clarification of the recovered-memory controversy will require systematic study of memory processes. Some investigators have claimed that traumatic memory operates in a fundamentally different manner than does nontraumatic memory, yet there is little experimental support for this assertion. Basic psychological research is just beginning to uncover the memory mechanisms that underlie intense false recollections. There are solid indications that a phenomenon known as source amnesia (in which a person forgets the source or context in which a memory originated) renders people vulnerable to memory distortions. When people cannot remember the source of a memory, they are apt to confuse whether it reflects an actual event, a fantasy or something that was said or suggested. The role of source amnesia in therapeutically induced false recollections remains to be explored.

The stakes here extend far beyond improved understanding of the mind. Research and effective communication are needed to minimize the possibility that people who were not abused will come to the psychologically devastating conclusion that they were. It is also imperative to avoid false accusations that can fracture lives and shatter families. And a better awareness of the workings of memory will bolster the credibility of the memories reported by survivors of sexual abuse.

One unacceptable outcome of the present situation is that the memories of genuine incest survivors may be called into question. The only way to avoid this travesty is to encourage all participants to adhere to rigorous standards of scientific inference and logic. Neglect of science has contributed to the present difficulties. In an interview with Ellen Bass, Ofshe and Watters asked the co-author of *The Courage to Heal* if she could cite any scientific support for her ideas. She responded candidly: "Look, if we waited for scientific knowledge to catch up, we could just forget the whole thing. My ideas are not based on any scientific theories."

The events of the past several years suggest that the price of not waiting for scientific knowledge may be disastrously high. It is imperative that all involved in this debate work hard to ensure that the standards of science, not rhetoric or pseudoscience, constitute the framework for future discussion.

DANIEL L. SCHACTER is professor of psychology at Harvard University.



Planting Trees

to give themselves

A Future

Together, American children, and children throughout the world, are planting trees to save our environment.

If you'd like to help them, please contact:



1(800)643-0001

THE NEW BOOK from the author of GÖDEL, ESCHER, BACH



"No reader will fail to be impressed by Hofstadter's intimate knowledge of the field and his ingenious contributions."

—Martin Gardner



\$30 at bookstores or toll-free 800-331-3761

BasicBooks

A Division of HarperCollins Publishers
Also available from HarperCollinsCanada Ltd

SCIENTIFIC AMERICAN CORRESPONDENCE

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

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

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

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

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

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

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

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



ESSAY by Jeremy Bernstein

The Poor Person's Guide to *The Bell Curve*

I have, of course, never read *The Bell Curve*. I doubt that anyone has actually read it except possibly its authors, Richard J. Herrnstein and Charles Murray, and I am not entirely sure about them either. No matter. The important thing is that I have read *about The Bell Curve*. To have read it would have made the writing of this guide quite impossible. It would have confused matters irreparably. But now to the task at hand.

I begin with a simple, almost self-evident principle that for the sake of clarity I call Bernstein's First Law.

Bernstein's First Law: All tests measure something.

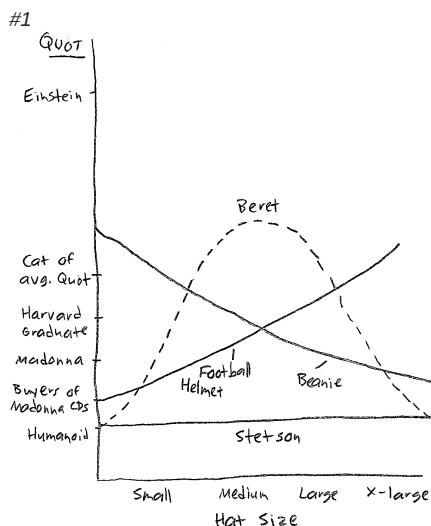
A case in point: I was once tested on my ability to crawl under a limbo stick after having drunk a jigger of Jamaican rum. I found that I could do it if the stick was set at my actual height.

Conclusion: IQ tests measure *something*. The problem is what to call it.

There has been a terrible fuss because the thing the tests measure has been called Intelligence Quota. Therefore, having a small one qualifies a person as a moron—something that is regarded as undesirable in some parts of our society. I propose to solve this problem by renaming it Quot. No stigma is attached to having a large or small Quot, any more than one is stigmatized by having a large or small telephone number.

The real question is how to pronounce Quot. Here I can be helpful. The word comes from James Joyce's *Finnegans Wake*. Recall the sentence, "Three quots from Mister Motz." Herman Motz was the owner of a liquor store in Zurich where Joyce shopped. Joyce had trouble with his *rrr*'s—hence his pronunciation of "quart." This problem much amused Motz, whose little joke seeped into *Finnegans Wake*. In short, "Quots" rhymes with "Motz."

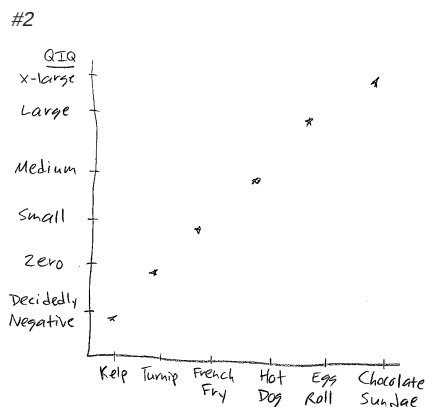
Does Quot size correlate to any other recognizable genetic feature? This is an important matter. It is so important that I give the evidence for a correlation between Quots and hats in my first graph, graph #1. I believe this diagram, which I have painstakingly sketched by hand, is about as accurate as the ones in *The Bell Curve*.



Note that beret wearers have almost perfect bell-curved Quots. That is because most of them are French.

I next turn to the delicate question, "Should you be told the size of your Quot?" The simple answer is no. If you have a large Quot and are so informed, it will give you a swelled head. And nobody wants to be confronted with the heartbreak of a small Quot. When I was about 11 years old, I was told that my Quot was substantially smaller than my sister's. This information did me no good. I blamed my parents. I asked them what *their* Quots were, and they told me that it was none of my business.

If you find that you or your loved ones have small Quots, is there anything you can do about it? Yes! That is the really good news I bring you. En-



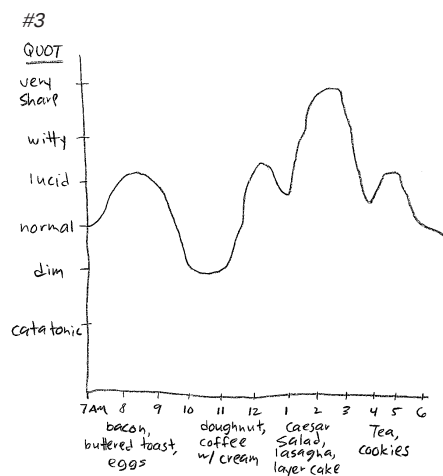
larging your Quot is all a matter of proper nutrition. This is spelled out in the second graph, graph #2. The term "GIQ" stands for Quot Improvement Quota. For the sake of economy, I have measured GIQs in the same units I used to measure hat size.

This graph is a treasure trove of information. For example, it explains why patrons of health food stores often seem a little slow. A good hot pastrami sandwich would do wonders.

Finally, I would like to present a third graph, graph #3, showing the result of a recent scientific experiment in which I was myself the subject. It takes advantage of the latest high-tech Quot meters. These devices, which can be inconspicuously disguised as Walkmans, earmuffs or cellular telephones, constantly monitor your Quot. The results speak for themselves.

It is clear from the graph that the period around 10:30 A.M. is a very dicey one. Forewarned is forearmed.

Readers of this guide have no need



to buy *The Bell Curve*. I have saved you \$30! My advice is to redirect this money toward a good Quot-enhancing dinner. Don't forget the three basic food groups: vanilla, chocolate and butter pecan.

JEREMY BERNSTEIN wears many hats, one of which is professor of physics at the Stevens Institute of Technology in Hoboken, N.J. The inspiration for this essay came while he was munching on a hoagie.