PLEASE STAND BY: THE NEW SHAPE OF TELEVISION • INTELLIGENCE AND GENES

SCIENTIFIC AMERICAN

MAY 1998 $4.95

NASA astronaut describes

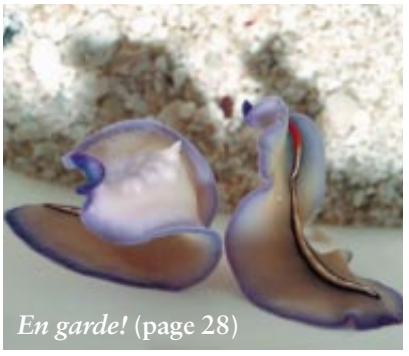
SIX MONTHS IN SPACE

Shannon Lucid peers out of Mir

SCIENTIFIC AMERICAN

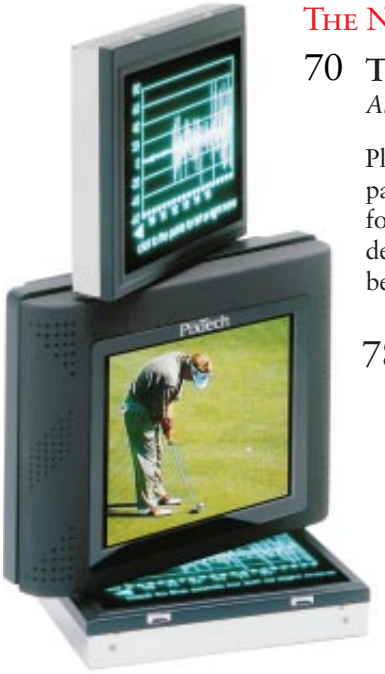
May 1998 Volume 278 Number 5

NEWS AND ANALYSIS



En garde! (page 28)

"For six months, at least once a day, and many times more often, I floated above the large observation window in the Kvant 2 module of Mir and gazed at the earth...." So astronaut Shannon Lucid begins the description of her record-breaking sojourn on board the Russian space station. Here she discusses the rigors of training, the dexterity of mind and hand required in zero-g, the need for fast-paced music and other details of life in space.



THE NEW SHAPE OF TELEVISION

About the Cover
On board the Mir space station, astronaut Shannon W. Lucid gazes out of a porthole while earthlight reflects off the glassy surface, in this artist's conception. Painting by Don Dixon.

Visit the Scientific American Web site (http://www.sciam.com) for more information on articles and other on-line features.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111.

Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97 (outside U.S. $47). Institutional price: one year $39.95 (outside U.S. $50.95). Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. Reprints available: write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.**

FROM THE EDITORS

Outsmarting Our Genes

Even the experts disagree about the best way to define and measure intelligence, but one opinion is unanimous: nobody wants less of it. When it comes to brain power, everyone wants to live in Garrison Keillor's Lake Wobegon, where all the children are above average. That is why, when the subject is the heritability of intelligence, fistfights break out so easily. Nobody wants to round out the bottom of the bell curve—especially not for reasons that seem beyond control.

"Nature versus nurture" and "biology as destiny" are long-dead issues for science. Genes and the environment in which they operate cannot be disentangled. All that genetics can do is lay out a physiological landscape in which a mind can grow. Estimates vary, but most studies say inherited factors alone can explain about half the measured differences in people's cognitive abilities.

The emphasis in that sentence should be on *measured differences*. Studies in this area look at the distribution of individual scores around some statistical mean. Saying that genetics can explain 50 percent of that *distribution* is not the same as saying that genetics can explain 50 percent of a person's *score*. Therefore, it would be wrong to say that half of intelligence is known to be genetic.

Nobody wants to round out the bottom of the bell curve.

When we judge someone's intelligence, we are usually guided by his or her particular intellectual skills: verbal fluency, a knack for solving math problems, musical aptitude and so on. For many decades, psychologists have noticed that these separate abilities tend to correlate, which has fostered the idea of an underlying global intelligence at work. Behavioral geneticists Robert Plomin and John C. DeFries, however, have gone back to look for genetic involvement in the distinct skills. They describe their results, beginning on page 62.

Their approach does not deny the possibility of genes for overall cognitive achievement. Plomin has in fact been hunting for genes associated with higher IQ (and might have just found one). But by teasing out the verbal components of intelligence, for example, investigators may more easily locate genes involved specifically in reading disability or giftedness. With that knowledge comes the possibility of intervening for the better.

If genes affect intelligence strongly, then a nurturing environment becomes only more important. And perhaps biomedical remedies based on genetic discoveries could offer everyone a helping hand up. (Some of the social consequences of that might give us pause.) By whatever means, studies of intelligence confer on us the opportunity to take that road to Lake Wobegon—if we really want to.

JOHN RENNIE, *Editor in Chief*
editors@sciam.com

SCIENTIFIC AMERICAN®
Established 1845

John Rennie, EDITOR IN CHIEF

Board of Editors
Michelle Press, MANAGING EDITOR
Philip M. Yam, NEWS EDITOR
Ricki L. Rusting, ASSOCIATE EDITOR
Timothy M. Beardsley, ASSOCIATE EDITOR
Gary Stix, ASSOCIATE EDITOR
Carol Ezzell; W. Wayt Gibbs; Alden M. Hayashi; Kristin Leutwyler; Madhusree Mukerjee; Sasha Nemecek; David A. Schneider; Glenn Zorpette
CONTRIBUTING EDITORS: Marguerite Holloway, Steve Mirsky, Paul Wallich

Art
Edward Bell, ART DIRECTOR
Jana Brenning, SENIOR ASSOCIATE ART DIRECTOR
Johnny Johnson, ASSISTANT ART DIRECTOR
Jennifer C. Christiansen, ASSISTANT ART DIRECTOR
Bryan Christie, ASSISTANT ART DIRECTOR
Bridget Gerety, PHOTOGRAPHY EDITOR
Lisa Burnett, PRODUCTION EDITOR

Copy
Maria-Christina Keller, COPY CHIEF
Molly K. Frances; Daniel C. Schlenoff; Katherine A. Wong; Stephanie J. Arthur

Administration
Rob Gaines, EDITORIAL ADMINISTRATOR
David Wildermuth

Production
Richard Sasso, ASSOCIATE PUBLISHER/ VICE PRESIDENT, PRODUCTION
William Sherman, DIRECTOR, PRODUCTION
Janet Cermak, MANUFACTURING MANAGER
Tanya DeSilva, PREPRESS MANAGER
Silvia Di Placido, QUALITY CONTROL MANAGER
Carol Hansen, COMPOSITION MANAGER
Madelyn Keyes, SYSTEMS MANAGER
Carl Cherebin, AD TRAFFIC; Norma Jones

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:
Thomas Potratz, EASTERN SALES DIRECTOR; Kevin Gentzel; Stuart M. Keating.
DETROIT, CHICAGO: 3000 Town Center, Suite 1435, Southfield, MI 48075; Edward A. Bartley, DETROIT MANAGER; Randy James.
WEST COAST: 1554 S. Sepulveda Blvd., Suite 212, Los Angeles, CA 90025; Lisa K. Carden, WEST COAST MANAGER; Debra Silver. 225 Bush St., Suite 1453, San Francisco, CA 94104
CANADA: Fenn Company, Inc. DALLAS: Griffith Group

Marketing Services
Laura Salant, MARKETING DIRECTOR
Diane Schube, PROMOTION MANAGER
Susan Spirakis, RESEARCH MANAGER
Nancy Mongelli, PROMOTION DESIGN MANAGER

International
EUROPE: Roy Edwards, INTERNATIONAL ADVERTISING DIRECTOR, London. HONG KONG: Stephen Hutton, Hutton Media Ltd., Wanchai. MIDDLE EAST: Peter Smith, Peter Smith Media and Marketing, Devon, England. BRUSSELS: Reginald Hoe, Europa S.A. SEOUL: Biscom, Inc. TOKYO: Nikkei International Ltd.

Business Administration
Joachim P. Rosler, PUBLISHER
Marie M. Beaumonte, GENERAL MANAGER
Alyson M. Lane, BUSINESS MANAGER
Constance Holmes, MANAGER, ADVERTISING ACCOUNTING AND COORDINATION

Chairman and Chief Executive Officer
John J. Hanley

Corporate Officers
Joachim P. Rosler, PRESIDENT
Robert L. Biewen, Frances Newburg, VICE PRESIDENTS
Anthony C. Degutis, CHIEF FINANCIAL OFFICER

Electronic Publishing
Martin O. K. Paul, DIRECTOR

Ancillary Products
Diane McGarvey, DIRECTOR

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

PRINTED IN U.S.A.

LETTERS TO THE EDITORS

PLACEBO EFFECT

Kudos to Walter A. Brown for a most overdue article on placebos ["The Placebo Effect," January]. Although modern medicine has no doubt revolutionized health care, the healing process is too complex to be explained by medicine alone. But I think his statement "Gone are the potions, brews and bloodlettings of antiquity" stands a small correction. Bloodletting, or rather therapeutic phlebotomy, is still in use today. Polycythemia, which is the overproduction of red blood cells, and porphyria cutanea tarda, an inability to process the porphyrin ring of old hemoglobin, are the only two diseases I am aware of that are treated with bloodletting.

WILLIAM B. CRYMES, JR.
University of South Carolina

The placebo effect has never been disparaged by skilled clinicians, who use it regularly. Pharmacological researchers, however, have long been baffled while searching for a mechanism to explain it. About 20 years ago it became clear that the analgesic effect of placebos could be nullified by administration of drugs that blocked the active sites of the endogenous opioids in the brain. This result shows that psychological factors (trust in a relationship with the physician) could increase production of a neurohumoral compound that diminished the body's exaggerated and harmful stress reaction and promoted healing. These findings not only explain the placebo effect but also point to the powerful impact of mental activity on body processes.

HENRY KAMINER
Tenafly, N.J.

WHAT'S IN A NAME?

In the 1995 version of the periodic table shown in Ruth Lewin Sime's article "Lise Meitner and the Discovery of Nuclear Fission" [January], elements 105 and 109 were (tentatively) called hahnium (Ha) and meitnerium (Mt), respectively, in honor of Otto Hahn and Lise Meitner. But last September the International Union of Pure and Applied Chemistry (IUPAC) officially assigned the name dubnium (Db) to element 105, while keeping the name meitnerium for element 109. If Hahn's treatment of Meitner (as described in Sime's article) was anything but decent, justice has been served by the action of the IUPAC. Meitner might have been overlooked by the Nobel committee, but who needs a Nobel Prize when one is immortalized in the periodic table?

Y. JACK NG
Chapel Hill, N.C.

A STELLAR ODYSSEY

Concerning the January article "The Ulysses Mission," by Edward J. Smith and Richard G. Marsden: Strange, isn't it, to honor a mythical Greek warrior by giving a space probe the Latin name *Ulysses* instead of his Greek name *Odysseus?*

HARRY ZANTOPULOS
North Canton, Ohio

RADIOACTIVE WASTE DISPOSAL

The article "Burial of Radioactive Waste under the Seabed," by Charles D. Hollister and Steven Nadis [January], briefly touches on other options for the disposal of radioactive waste, including the combination of plutonium with uranium oxide to create a mixed-oxide fuel for commercial reactors. The authors go on to state that most nuclear power plants in the U.S. would require substantial modifications before they could use mixed-oxide fuel. No basis was cited for this statement. I would like to know what changes would be required in the U.S., considering that mixed-oxide fuels are currently used in some 20 European nuclear plants of similar design. I am continually surprised by discussions that suggest that solutions to U.S. nuclear power issues are difficult or impossible when the solutions are being implemented throughout the rest of the world.

MARK BURZYNSKI
Hixson, Tenn.

Hollister and Nadis reply:

Although it is true that many European nuclear reactors routinely "burn" mixed-oxide fuels, no commercial plants in the U.S. are licensed to do so. Conventional light-water reactors, moreover, would have to be "significantly modified" to run exclusively on mixed-oxide fuels, according to a 1994 National Academy of Sciences study. These changes would involve installing more control rods and perhaps boosting their effectiveness. The study indicates that altering existing reactors would also require a "safety review and a substantial shutdown period," the costs of which have not yet been determined.

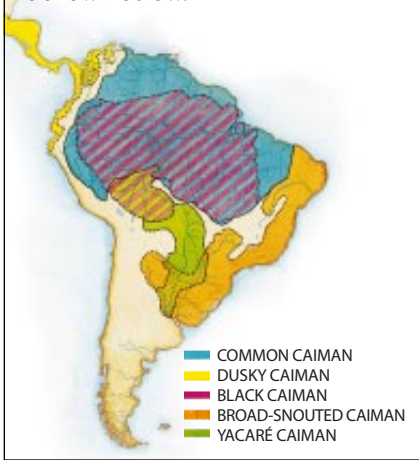
We do not consider solutions to nuclear power issues "impossible." On the contrary, our article points to a potential solution to two vexing problems—the disposal of high-level radioactive waste and of decommissioned nuclear weapons—through the interment of that material in geologic formations below the oceans.

Letters to the editors should be sent by e-mail to editors@sciam.com or by post to Scientific American, 415 Madison Ave., New York, NY 10017. Letters may be edited for length and clarity.

ERRATA

In "The Search for Blood Substitutes" [February], the chart on page 74 contains an error. The correct number of platelets in human blood is between 150,000 and 400,000 per cubic centimeter of blood.

The key for the map accompanying "The Caiman Trade" [March] was incorrect. The corrected version is shown below.



ROBERTO OSTI

50, 100 AND 150 YEARS AGO

SCIENTIFIC AMERICAN

MAY 1948

FUTURE OF THE AMAZON—"The Hylean (from the Greek *hyle,* meaning 'wood') Amazon Institute is an enterprise of breathtaking scope. Its purpose is not to gouge raw material and food out of the untamed forest. The new Institute will try a more thoughtful and subtle approach. Its strategy is to study the region's physiography, natural history and ecology (in this case, the relationship between the environment and man) and to evolve a process whereby man will learn to live harmoniously and richly in the environment instead of fighting it. To civilize the wild, rich Amazon and open it to colonization would itself be a gigantic achievement."

An ancient statue from Mexico

COSMIC ORIGIN—"The Dust Cloud Hypothesis, as it is called, suggests that planets and stars were originally formed from immense collections of submicroscopic particles floating in space. Interstellar space, formerly supposed to be empty, is now known to contain an astonishing amount of microscopic material. Jan Oort of the Netherlands, the president of the International Astronomical Union, has calculated that the total mass of this interstellar dust and gas is as great as all the material in the stars themselves, including all possible planet systems.—Fred L. Whipple."

MAY 1898

AZTEC WARRIOR—"Our illustration shows a statue of terra cotta, 5 1/4 feet tall, found by an Indian in a cavern near the city of Tezcoco. It is certain that this statue antedates the Spanish conquest. The clothing consists of a blouse (*uipilli*) with very short sleeves, a cotton girdle (*maxlatl*), leggings and sandals. The hypothesis that the statue was that of a chief or warrior is strengthened by the cotton armor, which Torquemada calls *ichcauhuitl*. This offered so efficacious a protection that the Spaniards hastened to adopt it to protect themselves against the formidable wood and obsidian saber (*maquahuitl*) of the Mexicans."

AFRICAN PLAGUE—"French physicians in Algeria have discovered a disease in Africa which, if the meager reports which have been received prove true, is as fatal as the bubonic plague now spreading in India. It first shows itself by the patient having an inordinate desire to sleep. Its symptoms resemble those manifested in laudanum poisoning. If the patient be not at once aroused, he soon falls into a stupor, which is succeeded by death. From its symptoms it has been called by the correspondents of French medical journals in Algeria 'La Maladie du Sommeil' (the disease of sleep). Two doctors of the University of Coimbe have a theory that the disease is microbic."

DRY GREASE—"Graphite as a lubricant is now recommended even by the organ of the Prussian steam boiler inspection society. However, the graphite must not only be free from all hard foreign bodies, such as quartz, but must also be in the shape of flakes, which cling to the rough surface of the metal and fill up all irregularities left in the manufacturing. Such graphite is, according to recent experiments, three times as effective as the best mineral sperm oil. It is at present only placed on the market from Ceylon and from Ticonderoga, N.Y."

DOG TAGS—"The War Department has prepared a system for identifying the men in the United States armies who may go into action. They will wear around their necks little tags of aluminum, by which they may be identified if found on the field of battle. In the last war it was often impossible to properly identify the dead soldiers, and thousands were buried in graves marked 'unidentified.' "

MAY 1848

WIRE FENCE—"This mode of fence is becoming quite common in the northern part of Illinois. We hear of many pieces of it at various places near Rock River—one of them being about two miles in length. The cost, as near as we can learn, is about 35 cents to the rod [16 1/2 feet]. It is said to be most admirable against all stock but swine. Cattle and horses particularly, after having their noses well sawed by it once, can scarcely be got near it again."

SAVING JOBS—"A mob of Journeymen brick makers was dispersed by the Baltimore police on Thursday, during an attempt to destroy some labor-saving machines introduced in certain brickyards, under the insane pretence that with the machines the owners would dispense with hands."

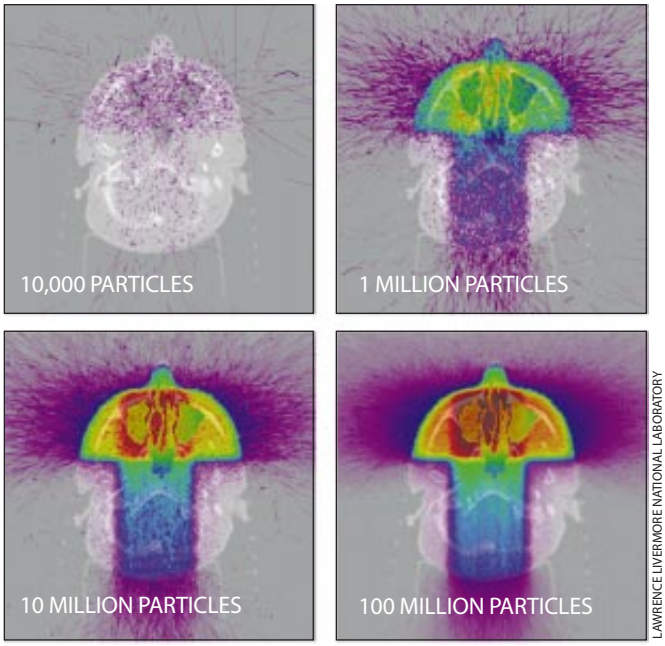
BALL OF FIRE?—"Many philosophers have firmly believed that the centre of the earth was a great fire and that the inhabitants of our globe lived, walked and slumbered on the crust of a huge furnace of which Vesuvius, Stromboli and other volcanoes were but the smoke pipes. These views of the igneous theory, as it was named, have lately been yielding to more rational ones. All the phenomena attributed to fire may be produced by electro-magnetic currents. Earthquakes and volcanic action may be the result of fluctuations in opposing electrical currents."

NEWS AND ANALYSIS

IN FOCUS

TAKING AIM AT TUMORS

Radiation is still a blunt weapon against cancer. New software may soon make it much more effective



RADIATION DOSE DELIVERED TO A BRAIN TUMOR (high-dose areas in red) ***is predicted by calculating every event that will occur to individual x-rays as they pass through a patient's skull, seen in this top view. A new computer system can simulate up to 100 million x-rays*** (purple) ***in about 30 minutes.***

A tour of Lawrence Livermore National Laboratory leaves no doubt that this bastion of fundamental physics research is still obsessed with the design and safekeeping of weapons of mass destruction. In one building, a 20-meter-long (65-feet-long) gun fires projectiles at up to 29,000 kilometers (18,000 miles) an hour to simulate the impact of a ballistic missile. On the other side of campus, 10 giant lasers zap tiny pellets with 30 trillion watts to study the genesis of nuclear fusion. This is not a place one would expect to produce a significant advance in cancer therapy.

But a small team of physicists and engineers, working for five years with funding only from Livermore itself, has taken computer algorithms once used for designing nuclear weapons and assembled them into a promising new tool for treating cancer with radiation. The Peregrine system, as it is called, "is genuinely a major step forward," praises Francis J. Mahoney, head of radiotherapy development at the National Cancer Institute. The technology should be ready for installation at cancer centers sometime next year.

The basic goal of the Peregrine project, explains Christine L. Hartmann-Siantar, its principal investigator, is to improve the precision of radiation treatment. Every year roughly 60 percent of cancer patients in the U.S.—some 750,000 people—receive such therapy, in which beams of x-rays or gamma rays are aimed at tumors in order to kill the malignant cells. About half those people reasonably hope to be cured, because their tumors are localized and vulnerable to high-energy light. Yet Radhe Mohan of the Medical College of Virginia estimates that some 120,000 of those curable patients die every year with their primary tumors still intact.

One reason the success rate is not higher, says Lynn J. Verhey, a medical physicist at the University of California at San Francisco, is that it is very difficult to predict exactly how much energy an x-ray beam will deposit into a tumor and into the

healthy tissue surrounding it. "For many years," recalls Edward I. Moses, who manages the Peregrine project, "doctors simply approximated the human body as a bag of water that reduces the energy of the x-rays exponentially with depth."

More recent computer programs, based on convolution codes, use computed tomographic (CT) scans of the area around a patient's malignancy to arrive at estimates that are more realistic. "But they still have problems wherever air meets tissue or tissue meets bone," Verhey says. Consequently, Mahoney reports, the limiting factor in radiation therapy is usually the side effects caused by inadvertent overdoses to normal tissue. "So if you can reduce that," he continues, "you should be able to increase the dose and increase the amount of cancer you kill"—and perhaps save lives.

Scientists have in fact known for decades of a way to calculate radiation doses much more accurately. Called Monte Carlo analysis, the technique tracks the life of a solitary photon as it journeys from the x-ray machine through the patient's body (or, more precisely, through a three-dimensional CT scan of it). Everything that happens to the photon—colliding with an electron in the skin, ionizing a hydrogen atom in the blood, perhaps even being absorbed by calcium in the bone—is calculated from the fundamental laws of physics and empirical measurements taken at Livermore and verified by blowing up H-bombs.

That is step one. Step two is to repeat step one for about 100 million randomly generated photons. "It's a brute-force way of solving the problem," Hartmann-Siantar admits. "As recently as 1995, a full Monte Carlo analysis for a single patient took something like 200 hours," she says. "Clearly, that would never work in a clinic, where many patients must be treated each day."

Peregrine can complete a Monte Carlo analysis of a patient's radiation treatment plan in about 30 minutes. "When people hear that, they immediately assume that we threw out some of the physics," Moses says. "That's not true: we include even rare interactions." The team did toss out many unnecessary frills in the standard Monte Carlo programs, however, such as their ability to handle moving targets and beams that change shape on the fly. "That made the problem vastly simpler," he explains.

Souped-up hardware gives the system another kick. The team designed the software to run on multiple processors simultaneously; the system currently uses 16 Intel Pentium Pro chips. That may sound expensive, but Moses estimates that Peregrine will add only 10 to 15 percent to the price of typical radiation-planning systems, which can reach $300,000.

If Peregrine makes as much of a difference in the clinic as it seems to in the lab, it may be well worth the premium. In cases where radiologists have compared Peregrine's predictions with those made by conventional codes, the doctors have found discrepancies that Mahoney describes as "fairly shocking." If the Monte Carlo analysis is correct, he notes, then "in some standard treatments in regions that are hard to plan"—for tumors in the breast, lung, spinal column, head or neck, for example—"there's a lot of missing going on."

In retrospective studies of patients with prostate and spinal cancers, for example, Peregrine has shown that standard radiotherapy deposited 5 to 10 percent less energy into the tumors than was thought. Beams that radiologists had predicted would deliver a lethal dose to an area well around one larynx tumor actually failed to kill any of the cancer.

"The most interesting case is breast cancer," Moses says. "Standard codes typically predict a low dose to [the skin on the breast]. Yet it is well known that skin burning is a real problem: sometimes women have to stop treatment because the burns are so severe. Peregrine shows four times as much dose to the skin. And where the old method predicts a very uniform dose at the chest wall, our system shows that it would in fact be pretty spotty. This might explain the fact that some women experience a hardening of their breast, called radiation fibrosis, after therapy."

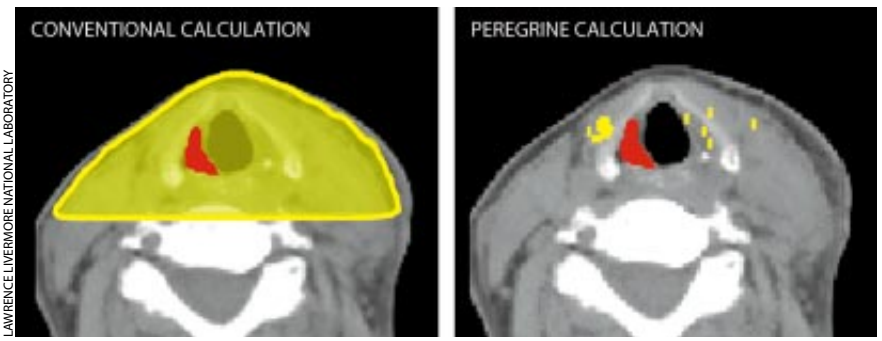
Although the Livermore scientists are confident that Peregrine will make all the difference for perhaps thousands of patients, Verhey is more cautious. "A few basic lab measurements disagree with what Peregrine says we should expect," Verhey says. "That tells me there is at least one minor problem that still has to be worked out. But it doesn't dampen my enthusiasm at all for this important planning tool."

The real test, of course, will come when the system enters the clinic. "We are writing our application for market clearance as fast as we can," Moses assures. No clinical trials will be needed to satisfy the Food and Drug Administration of Peregrine's safety. "It's much easier than getting a new drug cleared," he points out. "We simply have to prove that our calculations are at least as good as those already in use."

Livermore Laboratory has been talking to more than a dozen companies about integrating Peregrine into the leading treatment-planning systems, and the lab is planning to begin licensing negotiations later this year. If all goes smoothly, Moses says, the system could start showing up in hospitals in early 1999.

For Hartmann-Siantar, that will not be the end of this project—she and her colleagues intend to expand the program to work with electrons and protons as well as x-rays—but it will be the payoff. "When I was a graduate student, seven people in my family died of cancer within a single year," she recalls. "That's the reason I'm doing this. It touches everybody."

—W. Wayt Gibbs in Livermore, Calif.



LAWRENCE LIVERMORE NATIONAL LABORATORY

TREATMENT OF LARYNX TUMOR

(red) ***was planned using standard techniques*** **(left)** ***to expose a wide area of the neck with a curative dose of x-rays*** **(yellow).** ***Peregrine's analysis*** **(right)** ***shows that in fact the tumor did not receive enough radiation to kill any of it. The x-ray image shows a horizontal slice through the neck, near the Adam's apple*** **(top of image),** ***revealing the esophagus*** **(black)** ***and vertebrae*** **(white).**

SPACE HAZARDS

MAKING A DEEP IMPACT

Hollywood tackles the threat of near-earth objects

It's not clear just what kind of impact Asteroid 1997 XF-11 has left on the earth. On March 11 Brian G. Marsden of the Harvard-Smithsonian Center for Astrophysics reported that in 2028, an object about 1.5 kilometers (a mile) wide would pass some 50,000 kilometers (30,000 miles) from the earth—a hair's breadth in astronomical terms. In fact, researchers at the time couldn't say for certain that the asteroid would miss the planet. The next day astronomers found photographs of the object taken in 1990 and recalculated the asteroid's orbit; they figured that it would miss the earth by nearly a million kilometers, more than twice the distance to the moon. A few criticized Marsden, who tabulates observations and catalogues space bodies that might hit the earth. The fear was that people might not take the next call seriously.

Some indication of public attitudes toward the threat of near-earth objects might come soon, when Hollywood releases a film this month about such a possibility. For several months, promoters of the film trained their sights on *Scientific American, Sky and Telescope,* the Learning Channel and other media that would not be confused with *Entertainment Weekly.* That's because *Deep Impact* may represent the most lavish effort yet of Hollywood's trying to get the science right.

The Paramount Pictures–DreamWorks Pictures film, directed by Mimi Leder and co-executive-produced by Steven Spielberg, tells of a comet due to strike the earth in one year. To keep humans from suffering the same fate as the dinosaurs, the world's leaders must devise a scheme to deflect the comet—and come up with a way to save at least some people should the attempt fail. Similar disaster movies have been released (and another one, a Disney movie called *Armageddon,* is due out this summer), but the $100-million *Deep Impact* apparently differs from them in relying on half a dozen experts—including Carolyn S. and Eugene M. Shoemaker, the co-discoverers of Comet Shoemaker-Levy, which spectacularly crashed into Jupiter in 1994. (Eugene Shoemaker died in a car accident last year.)

Hollywood has been pushing to make the science more accurate, opines Warren Betts, the film's director of marketing and education of science and technology. And "I personally experienced a desire from the scientific community to come to us. NASA was so eager to work with us," Betts says of the National Aeronautics and Space Administration.

Of course, some dramatic license in a movie goes without saying. "Cometary dust is blacker than a charcoal briquette," explains Chris B. Luchini, who computationally models comets at the Jet Propulsion Laboratory in Pasadena, Calif., and was one of the film's technical advisers. But that would lead to filming black snow over a black surface, in the blackness of space—not visually appealing, so the comet dust is white. Still, Luchini found the filmmakers receptive to the science and willing to modify the script for accuracy. For instance, the original description of the comet—which is basically a dirty snowball—was incorrect. "They had the density higher than uranium," Luchini says. "A lot of details like that were flat-out wrong" but were subsequently corrected.

Perhaps the biggest stretch of realism, at least scientifically, has to do with the astronauts landing on the incoming comet to plant explosives. "A comet is not big enough to produce gravity" to land, notes Gerald D. Griffin, another adviser and a former flight director who also helped on *Apollo 13* and *Contact.*

But even a rendezvous with a comet is not practical. John L. Remo, who organized a United Nations meeting on near-earth objects (NEOs) in 1995 and is affiliated with the Harvard-Smithsonian Center, notes that a comet could move rapidly, some 50 kilometers per second, and could rotate around its axes. Matching such a complicated trajectory would be exceedingly difficult. A more reasonable approach is a detonation just off the cometary surface, which might shift the comet's motion. Simply ramming the object with a heavy-duty projectile might also work.

And given today's technology, a one-year warning isn't enough time. Experts think that 50 to 100 years may be necessary for a successful diversion (a longer lead time means a smaller nudge is required). For asteroids, that prediction time may be feasible; compared with comets, asteroids are rather stately, moving only about 20 kilometers per second, and follow predictable orbits. Comets when close to the sun emit gases to produce their characteristic tails; that outgassing affects their trajectories and makes them harder to track accurately.

A few organizations look for near-earth objects. So far they have found 108 objects that might pose a hazard—about 10 percent of the estimated total. And no concerted effort exists to develop deflection technologies. In part, it is

MYLES ARONOWITZ *Paramount Pictures and DreamWorks, L.L.C.;* PARAMOUNT PICTURES AND DREAMWORKS, L.L.C. *(inset)*

***PLANTING EXPLOSIVES ON A COMET LENDS DRAMA to the film* Deep Impact *but is not the way to divert an incoming space object* (inset).**

IN BRIEF

F's for U.S. Schools

Results from the latest and most comprehensive comparison of education in 23 nations showed that American high school seniors fall further behind their foreign counterparts than anyone thought. In tests of general mathematics, students from only two nations—Cyprus and South Africa—fared worse than U.S. 12th graders. And no country performed more poorly in tests of advanced mathematics and physics. Only those American students taking advanced placement calculus ranked higher than the average in that field.

Carbon Dioxide Crystals Up Close

At last, scientists have viewed solid carbon dioxide crystals. Because these eight-sided structures typically evaporate at temperatures higher than –134 degrees Celsius (–210 degrees Fahrenheit), they had never before been seen. But William P. Wergin and his colleagues at the U.S. Agricultural Research Service found a way to glimpse the tiny crystals—measuring some 0.13 micron—by chilling them to –196 degrees C (–320 degrees F) in a special scanning electron microscope.

AGRICULTURAL RESEARCH SERVICE-USDA

Brain Aging

For some time, scientists have known that receptors in the brain for the neurotransmitter dopamine become fewer and farther between with age. And now they have linked this depletion directly to a loss of motor skills and mental agility. Nora Volkow and her colleagues from Brookhaven National Laboratory, the State University of New York at Stony Brook and the University of Pennsylvania took positron emission tomography (PET) scans of 30 healthy volunteers, aged 24 to 86. They compared the density of dopamine receptors in sundry brain regions with each subject's results on a range of tests. Invariably, higher concentrations of dopamine receptors correlated with higher performance scores. Volkow believes it may be possible to mitigate the neurological symptoms of aging by improving the functioning of the dopamine system in the elderly.

More "In Brief" on page 26

because many proposals rely on nuclear devices, which run into international security issues, Remo notes. Such political considerations may soon change: the threat of NEOs may be on the agenda of a July 1999 U.N. conference about space (called UNISPACE III).

So what exactly are the odds of getting hit? Small objects, less than about 0.1 kilometer wide but powerful enough to level a city, slam into the earth about once a century (one such object exploded over the Tunguska Valley in Siberia in 1908). The odds that an "extinctor," an object two to five kilometers wide (about twice that of Asteroid 1997 XF-11), will strike the planet this century ranges from about one in 1,000 to one in 10,000, according to Remo. In more prosaic terms, he figures that is 10 times greater than the odds of the *Titanic*'s being sunk on its maiden voyage by an iceberg. "People should really wake up" to the threats, Remo argues. After the March asteroid scare and *Deep Impact*, perhaps they will. —*Philip Yam*

FIELD NOTES

SNOW MEN

To predict runoff, they fight bears and collect cosmic rays

March may tatter the white blanket of winter in most places, but for hydrologists Frank D. Gehrke and David M. Hart, it is the month when the snow really gets interesting. As the chief researchers overseeing California's snow surveys program, Gehrke and Hart must estimate the size of the great white lake draped over the mountains and alpine meadows that dominate the eastern flank of the state. Typically about 80 percent of the water that feeds California's inhabitants, farms and hydroelectric generators arrives in solid form and usually remains frozen until the start of the growing and air-conditioning season.

So, late each winter, the local TV camera crews strap on snowshoes and trudge out to observe Gehrke and Hart measure the snow and prognosticate on the prospects of a wet and bountiful summer. It is the West Coast version of Groundhog Day.

Snow measurement is more precise than dragging a drowsy Punxsutawney Phil from his burrow, but not by much. Despite all that hydrologists have divined about the intricacies of the delicate flakes and their life cycle, scientists have no convenient, accurate and reliable way to measure the bulk of snow covering a region. A new sensor that Gehrke and Hart are testing at their snow lab may change that by the next turn of the century. But today Gehrke, Hart and dozens of other surveyors scattered about the state have strapped on skis to measure snowfall in the same way that it has been done since the last turn of the century: by hand.

Hart skis around waist-high pines with remarkable grace, considering the 12-foot metal pipe balanced on his shoulder. At a clearing marked by two orange signs, he plunges the tube into the snow. As the tube slides in, and in, and in, I realize that those waist-high pines are in fact the tops of 15-foot-high trees, and I cinch the straps on my snowshoes a notch tighter. At last the pipe hits soil: "132 inches," Hart announces to Gehrke, who scribbles the figure into his notebook. The tube comes up, a core of snow inside it, and Hart places it on a spring scale Gehrke has strapped to his ski pole.

This is the critical measurement, because the weight of the snow reveals how much water it contains. "Depth is useful as a check to make sure we get good cores," Hart explains. But one foot of wet California snow can contain more water than four feet of dry Utah powder. Tabulating the day's data, Gehrke reports that El Niño has been more than generous. "We're measuring about 40 inches of water in this part of the Sierra Nevada, 74 percent above normal," he says. A few of the 300 survey sites in the state, he adds, are seeing more than 80 inches of water on the ground.

Snow measurements may be the best way to forecast spring runoff, but collecting them can be arduous and frustrating. Some surveys require 80-mile-long treks and climbs to altitudes of 11,450 feet. Bad weather has forced survey teams to hole up in remote cabins for a week. The state has set up about 100 pillowlike scales to weigh the snowpack automatically. But Gehrke shakes his head when I ask how they perform.

"Installing these things is a major pain," he says. "They're big—four of them connect into an 80-square-foot array—and we have to lug them in by mule. Bears like nothing better than to tear the hell out of them, and if bears

In Brief, *continued from page 24*

Dream On

Recent findings show that one popular theory about sleep may be all wrong. Because people awakened during rapid-eye-movement (REM) sleep—that phase of sleep during which the eyes dart back and forth beneath the eyelids—often remember their dreams, scientists speculated in the 1950s that REM somehow helped the brain process information collected throughout the day. But David Maurice and his colleagues at Columbia University have demonstrated that the aqueous humor—the liquid in the anterior chamber behind the cornea—does not circulate enough to deliver fresh oxygen to the cornea when sleeping eyes are still. Instead they pose that, in fact, REM serves to stir up the aqueous humor and ensure that corneal cells do not suffocate. The model could explain why periods of REM sleep become longer over the course of the night.

J. A. HOBSON *Photo Researchers, Inc.*

Evidence of Antigravity

An international collaboration of top-ranked cosmologists, called the High-z Supernova Search Team, made a surprising announcement in the journal *Science* this past February. The scientists reported that empty space may well be filled with a repulsive force of as yet unknown origins. In support of this antigravity theory, the group cites data on supernovae. The researchers maintain that the seeming brightness of these distant, exploding stars suggests that the universe is expanding at an ever increasing rate—an acceleration that the standard big bang model could not account for were there no antigravity.

Falling Cancer Rates

Americans appear to have won the latest battle in the war on cancer. A new study, "Cancer Incidence and Mortality, 1993–1995: A Report Card for the U.S.," shows that cancer rates peaked in 1992. Although incidence rates rose each year by an average of 1.2 percent between 1973 and 1990, they dropped by 0.7 percent between 1990 and 1995. So, too, whereas cancer death rates increased by 0.4 percent each year during the earlier period, they fell by 0.5 percent annually in recent years.

More "In Brief" on page 28

W. WAYT GIBBS

WEIGHING THE EFFECTS OF EL NIÑO *are David M. Hart* **(left)** *and Frank D. Gehrke* **(right).**

don't disable them, flooding often does." Even when the pillows successfully return data via satellite, Hart adds, "the figures are often off by 10 percent or more." Bridges of compacted snow often form around the edges of the pillow, relieving it of some of the mass above.

Gehrke hopes that he will soon be able to begin replacing the pillows with a new device developed at Sandia National Laboratory. The sensor uses two cosmic-radiation detectors, one above the snowpack and one beneath it. Because water (and thus snow) absorbs the gamma rays to which the sensors are tuned, the difference between the two readings can be converted directly into inches of water.

Gehrke installed two prototype machines two winters ago. "We have had to work out lots of bugs in the electronics, but now we're getting very accurate signals," he says. "If they are as bulletproof as we hope, we might eventually be able to place them in locations that are more representative of the terrain" than the meadows and ridge tops to which snow pillows and human surveyors are often limited.

Could the technology eliminate the need for researchers such as Gehrke and Hart to have to ski through whispering incense cedars in search of some pristine glade, I ask? The snow men grimace and look at each other. Naaaw.

—W. Wayt Gibbs in the Sierra Nevada

ASTRONOMY

WATER, WATER EVERYWHERE

Ice found on the moon

In 1961, during the very month that President John F. Kennedy launched the race to the moon, Kenneth Watson, Bruce C. Murray and Harrison Brown of the California Institute of Technology noted the importance of the fact that some craters in the moon's polar regions are permanently in shadow. Rather than being subjected to two weeks of blistering rays from the sun each lunar month, these sites remain eternally dark and frigid. Such "cold traps," they argued, might snare water dumped on the lunar surface by crashing comets or spewed forth by lunar volcanoes. And over the aeons, inky crater floors near the poles might accumulate substantial amounts of ice. Those deposits would be immensely valuable to people on future lunar bases, who could distill water from them or separate out the oxygen and hydrogen to use as rocket propellant. It took nearly three decades, but the latest robot probe, Lunar Prospector, has seemingly confirmed that frozen caches of water can indeed be found on the moon.

Because none of the Apollo missions visited the moon's poles, the proposal of Watson, Murray and Brown had remained untested for 30 years. The first experimental indication came when the Department of Defense and the National Aeronautics and Space Administration launched a probe called Clementine in 1994, with the intent of eventually flying it past a nearby asteroid. Before attempting that rendezvous, however, Clementine was sent into a polar orbit around the moon. (A software glitch later wrecked the asteroidal segment of the mission.)

Clementine found evidence for ice by bouncing radar signals off the lunar surface and back to antennas on the earth. Some of the signals that were returned suggested that ice might be present near the moon's south pole. Yet Clementine uncovered no indications of ice at the north pole, even though the probe flew much lower there, and the radar experiment should have been more sensitive to ice on the surface.

A 1994 report by the late Eugene M.

ANTI GRAVITY

Now You See It, Now You Don't

Picture the Beatles in a boat on a river, with or without tangerine trees and marmalade skies. They're chasing another boat. Assume that Rolling Stone Keith Richards is piloting that other boat, so its path is highly erratic. The Beatles pursue, turning their boat while continuously closing the distance to their prey. Hey, it could happen.

Switching from Beatles to beetles reveals, however, that pursuit of prey in one corner of the insect world turns out to be far less smooth. As was first reported more than 70 years ago, hungry tiger beetles, which have compound if not kaleidoscope eyes, run as fast as they can toward a prospective live meal but then come to a screeching stop. During this time-out, the beetles reorient toward their sidestepping targets. After zeroing in again, they resume running as fast as they can. They may have to do this three or four times before catching their prey or giving up.

Cole Gilbert, an entomologist at Cornell University, has seen this halting hunting technique in the woods near his lab. Finding the beetles' mystery tour toward their prey vexing, Gilbert decided to observe the stuttering stalkers under controlled conditions.

Gilbert set individuals loose after fruit flies, pursuits that he filmed and analyzed down to the millisecond. Without any direct studies of the beetles' eyes or brains, Gilbert came to a few conclusions about their sensory system and their behavior, which he published recently in the *Journal of Comparative Physiology.* Based on the angular movements of the prey, the angles between neighboring ommatidia (the units of the beetles' compound eyes) and the duration of beetle breaks, Gilbert thinks beetles, which don't see all that clearly to begin with, actually outrun their visual systems.

"Think of Elmer Fudd when he's scanning with binoculars," Gilbert explains. "So Elmer is looking for the wabbit." (Gilbert actually said "wabbit.") "And he's going pretty fast, and then there's a little blip, a slight change in light intensity in those fields. He goes past it, and then he backs up and there's Bugs chewing his carrot." The faster he pans, the smaller Bugs's blip gets. Eventually, Elmer pans so fast that too few bunny photons fall on Elmer's photoreceptors. Bugs is, in effect, invisible to Elmer's sensors. Result: no wabbit stew.

Gilbert contends that, in a similar way, the movement of the prey, combined with the beetle's own speed, results in too few prey photons making it to the beetle's photoreceptor. In effect, the beetle goes blind until it can stop and reduce the relative velocity of the prey to the point where it registers again on beetle radar.

The consideration of such biological tracking systems might help optimize devices such as the Mars Rover, Gilbert believes. "You want to move quickly to explore a large area, but if you move too fast for the optical sensors to gather enough information to form an image, the exploration is fruitless. Through knowledge of biological tracking systems, we can learn how nature has coped with this trade-off. It might allow for strategies that engineers wouldn't necessarily think of."

The intermittent sensing of the beetles might be a leaner and more efficient system than one with enough circuitry to incorporate a constant feedback of information and response. Which means that it sometimes makes sense to take the long and winding road. —*Steve Mirsky*

MICHAEL CRAWFORD

In Brief, *continued from page 26*

Cracking the Pumpkin Hack

The mysterious orange orb that sat atop Cornell University's 173-foot-tall bell tower since early October has at last come down. The object was whisked away to a nearby laboratory, where a team of professors confirmed that it was a pumpkin. Earlier on, two teams of students had unofficially demonstrated that the gourd was genuine. Both used tethered weather balloons to take samples—one with syringes, the other with a robotic drill. The pumpkin perpetrator remains at large.

FRANK DIMEO *Cornell University Photography*

Chaotic Communications

Scientists have developed chaotic encryption in electronic communications systems, and now a group from the Georgia Institute of Technology has done so in an all-optical system. Rajarshi Roy and Gregory D. VanWiggeren first encoded a message within the fluctuations of a laser's light. They then passed this signal through an erbium-doped fiber amplifier (EDFA), mixed it with a chaotic signal and transmitted it through an optical fiber. An EDFA on the receiving end generated chaotic fluctuations that were synchronized with those produced by the transmitting laser, making it possible to subtract the chaos from the combined signal.

Partible Paternity

The Bari people in Venezuela have an unusual view of paternity: when women take lovers during pregnancy, these men become so-called secondary fathers. Anthropologist Stephen Beckerman and his colleagues from Pennsylvania State University found that promiscuity during pregnancy is likely adaptive behavior for the mothers. Survival rates among children with secondary fathers was 80 percent, whereas it was a mere 64 percent among those children with only primary fathers. The reason: the Bari diet, which consists primarily of a starchy tuber called manioc, is not sufficient for children. Secondary fathers provide supplementary food—in the form of fish and meat—to their offspring. *—Kristin Leutwyler*

Shoemaker and two colleagues at the U.S. Geological Survey noted that the south pole of the moon contains "much larger" areas of permanent shadow than the north does, although just how much was hard to say. So Clementine's finding evidence for ice only in the south seemed to make some sense. But last year three radio astronomers reported that radar reflections of the type seen by Clementine could also be found for sunlit parts of the moon, casting doubt on this earlier indication of an icy southern pole. And the latest results from Lunar Prospector have completely reversed the bias that had, up to this point, placed the moon's south pole in the spotlight—or rather, out of it.

According to Alan B. Binder of the Lunar Research Institute, the leader of the Lunar Prospector science team, measurements from the spacecraft show "about twice as much water ice in the north polar regions as in the south polar regions." Actually, the relevant instrument on Lunar Prospector can only sense the presence of hydrogen. The conclusion that the hydrogen detected is from water, Binder admits, is "a leap of faith" but a logical one. The ice is apparently mixed with a great deal of rock, so that it makes up only a tiny fraction of the lunar soil. Assuming, however, that the ice-tinged soil extends a couple of meters deep, Binder and his colleagues guess that there may be anywhere from 10 million to 300 million metric tons of water in all.

Binder does not yet know why the new results from Lunar Prospector show

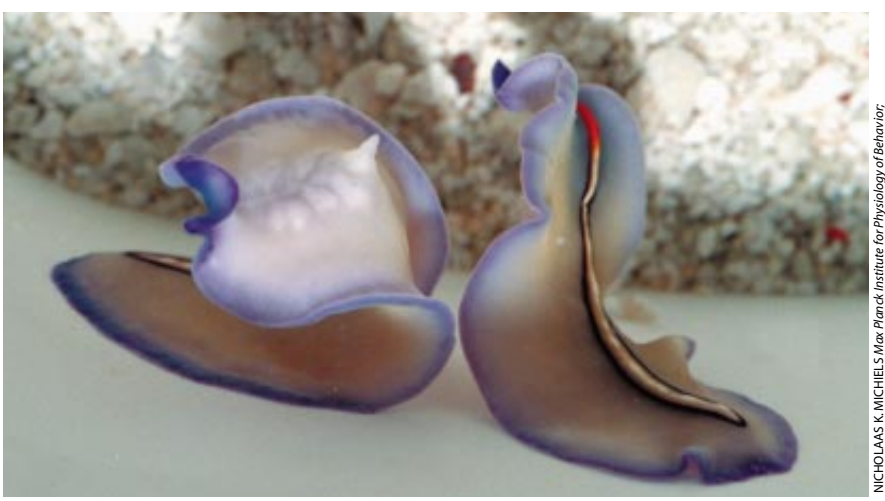
EVOLUTIONARY BIOLOGY

Dances of Worms

Some 15 to 20 meters (49 to 66 feet) below the ocean surface, in the warm waters off the coast of Queensland, Australia, an unusual mating dance takes place among hermaphrodite flatworms—and whichever wins gets to be the male, at least for the moment. In bouts referred to as penis fencing, these marine creatures will spend 20 minutes to an hour attempting to inject sperm under the skin of their mates before being injected or injured themselves.

Most other hermaphroditic animals, such as earthworms, exchange sperm during sexual trysts. But when a lone *Pseudoceros bifurcus* encounters another of its kind, it will stop, curl its body back in an intimidating backbend and display its penis. Each worm, about four to six centimeters (two inches) long, will then repeatedly strike the other until one succeeds in injecting sperm, and both will maneuver themselves to avoid being pierced.

Nicolaas K. Michiels and Leslie J. Newman published their findings in the February 12 issue of *Nature*. They explain that the duel to act as the male flatworm in a hermaphrodite pair could be understood as pure evolutionary selfishness. Because sperm are biologically cheaper to produce than eggs, males are by design able to produce more descendants than females over a lifetime, provided that they fertilize as many females as possible. So the loser leaves with the burden of fertilized eggs to care for, while the winner goes on, with a parry and a thrust, to mate again. *—Krista McKinsey*



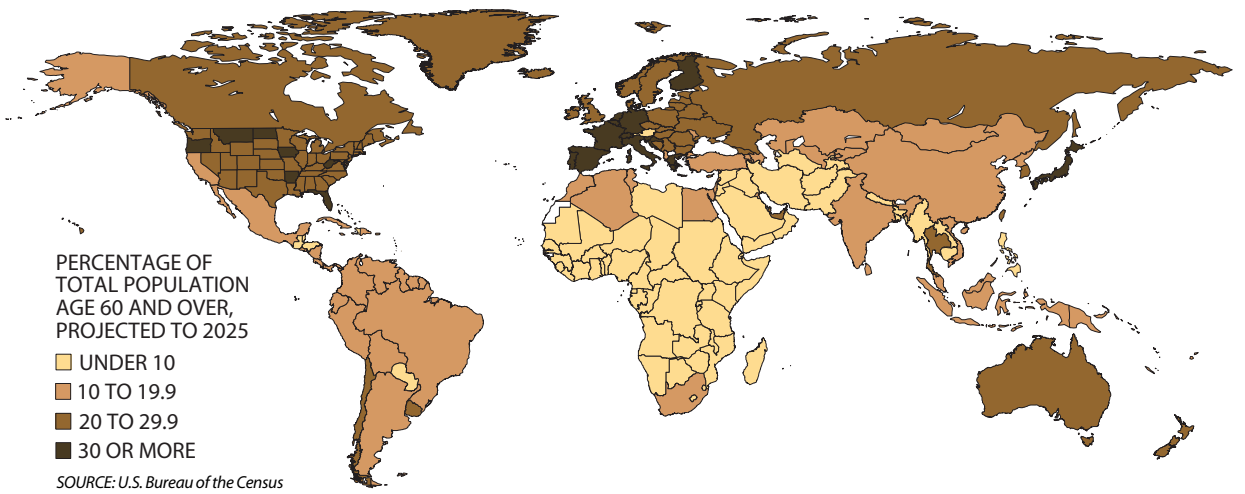
NICHOLAAS K. MICHIELS *Max Planck Institute for Physiology of Behavior;* LESLIE J. NEWMAN *University of Queensland*

***PENIS FENCING** is a flatworm mating game.*

more ice in the north than in the south. He suggests that the shadow maps previously obtained from Clementine may have been misleading. Commenting on the surprising hemispheric asymmetry, Paul D. Spudis, a geologist at the Lunar and Planetary Institute in Houston, remarks, "It's very puzzling to me." He notes that when Clementine visited the moon, the north pole was maximally lit. So the small amount of dark area seen up north at the time shows that the extent of permanent shadow there cannot be very large. Could previous assessments of the southern pole's permanently shadowed real estate have been that exaggerated? The answer unfortunately remains elusive for the moment: Lunar Prospector carries no camera, so the scientists cannot just take a quick look. *—David Schneider*

BY THE NUMBERS

The Future of the Old



To see the future of world population, look to Europe, where the birth rate is low and the number of elderly is rising dramatically. In Germany, for example, those 60 and older now account for 22 percent of the population and by 2025 will be at 35 percent, the majority of them women. Furthermore, because of continued low fertility, the total population of Germany is actually projected to decline by 8 percent between now and 2025.

Germany, together with other western European nations and Japan, is the advance guard of the historic demographic changes accompanying the rise of technological societies. If the global economy continues to raise living standards, developing countries will most likely follow the European model to eventual low fertility and large elderly populations. They are less likely to follow the American experience, which is atypical in part because of a continuing huge influx of immigrants. Immigrants are expected to be a major contributor to the 24 percent increase in U.S. population projected between now and 2025. In 2025 those 60 and older will account for 25 percent of the U.S. population, compared with 31 percent in western Europe.

Whether the world can afford adequate care for the elderly has been a matter of concern for at least a century and to this day is a potentially explosive issue in intergenerational politics. In one of the most provocative statements of recent years, Lester C. Thurow, the Massachusetts Institute of Technology economist, claimed that the demands of older Americans for social services threatens the investment on which the future of society depends. Whether such pessimism is justified will turn on imponderables such as the future trends of worker productivity, longevity and disability rates. Imaginative solutions, such as new ways for employing elderly people in the care of the disabled, can obviously make a difference. Among the harbingers on the pessimistic side, at least for the U.S., is the probability that baby boomers will increasingly overtax social services as they reach retirement age after 2010.

But there are grounds for optimism, including the recent report that disability rates of Americans 65 and older fell between 1982 and 1994. Also, there is an additional potential for reducing disability in later life by cutting rates of chronic nonlethal illnesses such as arthritis, back pain, migraines, depression and osteoporosis. In general, reducing lethal illnesses such as cancer and heart disease extends life but does not in itself lead to a healthy old age, because people are still prone to chronic nonlethal diseases. If these major killers were suppressed in tandem with chronic nonlethal disease, life without disabilities could be considerably extended.

The elimination or control of infectious disease, which is a goal of the World Health Organization, would greatly boost the disability-free lifespan of those in the developing countries, where infectious diseases are far more important as a cause of disability. Investment in basic biomedical research such as the Human Genome Project will in all likelihood help in understanding disease and hence in reducing disability rates. Possibly the biggest influences on disability rates will be the rising levels of affluence and education forecast for the next century. Educated people tend to follow healthier practices such as exercising, eating nutritious foods, abstaining from tobacco and seeking medical help—habits that are likely to reduce disability. *—Rodger Doyle (rdoyle2@aol.com)*

RODGER DOYLE

PROFILE

Rebottling the Nuclear Genie

Information warrior Thomas B. Cochran is fighting hard against U.S. reliance on nuclear weapons

Thomas B. Cochran is gazing intently at his computer screen, paging through hundreds of targets for a planned nuclear attack on Russia. The individually named and numbered strategic sites are organized by category: antiballistic-missile radars, launch-control centers, submarine docks, silo fields. Speaking quietly with a Tennessee twang, he apologizes for not yet having the exact geographical coordinates of all the missile silos—he's working on that. But he has the boundaries of the silo fields. And as a leading authority on nuclear weapons and official adviser to the Department of Energy, which oversees the nuclear stockpile, he also has a pretty clear idea which warheads to use against each target. "Information is extremely important in this business," he observes dryly.

Cochran is senior scientist and director of the nuclear program at the Natural Resources Defense Council, where for almost 20 years he has used aggressive political pressure tactics—information warfare of a sort—to reduce the U.S.'s reliance on nuclear weapons. "The people on Capitol Hill aren't going to pay any attention to you until they read about you in the *New York Times* or *Scientific American*," he says pleasantly. "So you litigate to get publicity."

The plan is to make a list of nuclear targets that matches as closely as possible the Pentagon's own highly classified list. "We think they reduced the target list last December from about 11,000 to about 2,000 in Russia and 500 elsewhere," Cochran notes. At the same time, he drops the first name of the National Security Council officer who drafted new "guidance" on targeting and points out that it explicitly allows the military to target nonnuclear weapons of mass destruction. Cochran and his associates Christopher E. Paine and Matthew G. McKinzie plan to use their homegrown target database to model the effects of different war scenarios, so they will have better information to aim new campaigns, like strategic warheads, at the DOE. "It will show the absurdity of keeping the number of weapons we keep," Cochran declares.

He also litigates to block the executive branch from deviating from statutes already on the books. His record—extraordinary by any standard—is grounded in technical analysis: Cochran's degrees are in physics and mathematics, not political science. Frank von Hippel, a nuclear weapons expert at Princeton University, says Cochran broke new ground in the 1970s, when, while working for the think tank Resources for the Future, he published a damning analysis of the Clinch River Breeder reactor project, a huge government program to develop a reactor that would create more fuel than it consumed. The government's logic "was based on several key assumptions, none of which turned out to be correct," Cochran recounts gloomily. "It was a total loser." Cochran fought the reactor on economic and environmental grounds for 12 years, until Congress canceled the scheme in 1983. "You have to be ready to stay the course," Cochran advises would-be activists.

Cochran "has extraordinary chutzpah," von Hippel remarks. "He is willing to take on what most people wouldn't bother with because they assume it's hopeless." And despite Cochran's even demeanor, he is quite capable of making "unvarnished statements," says von Hippel, who has served as a government official and been on the receiving end of some of Cochran's assessments of "idiotic" executive branch decisions.

The DOE now puts Cochran on its advisory committees, but if the aim is to soften his tough stance against the agency it does not seem to be working. The government scientist in charge of the National Ignition Facility, a giant laser-fusion device under construction at Lawrence Livermore National Laboratory, initially sought Cochran's approval for the plan. Cochran declined to give it

KATHERINE LAMBERT

***GERMANIUM GAMMA-RAY DETECTOR** was used by Thomas B. Cochran on a Soviet warship to show that such devices can identify warheads and so verify arms treaties.*

and eventually sued the DOE, challenging its decision to set up a committee at the National Academy of Sciences to advise on the project.

The committee, Cochran recalls shaking his head, was stacked with people with close economic ties to the DOE's weapons program and to Lawrence Livermore. Yet it had been asked to give a judgment on whether the machine should be built. A well-connected lawyer friend of Cochran's made short work of that arrangement in court last year, using an open-government statute. As a result, the DOE is not allowed to consider that committee's work.

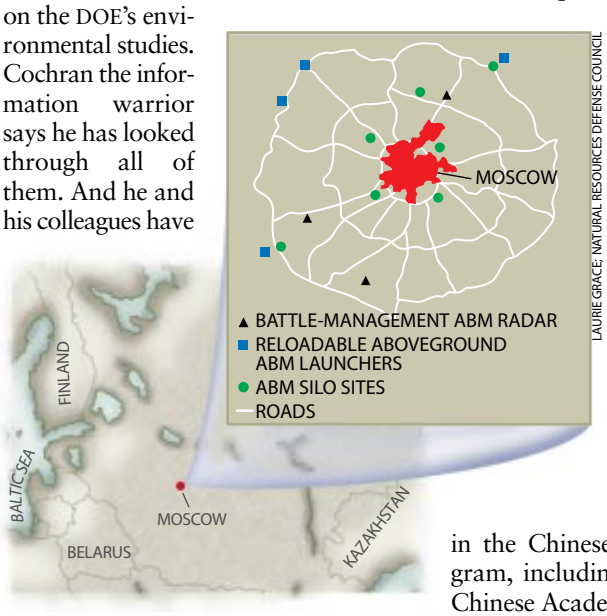
The National Academy of Sciences went into shock at the prospect that it might have to open up all its committee meetings to the public. In response to the contretemps, Congress hurriedly passed legislation that allows independent groups such as Cochran's to comment on the makeup of academy committees advising the government. The compromise also opens up to the public such committees' fact-finding sessions.

Cochran and his associates are now negotiating with DOE lawyers to settle a broader legal assault on the DOE's Stockpile Stewardship and Management Program, which the National Ignition Facility supports. The government says the program aims to ensure that nuclear weapons remain safe and reliable even though none have been tested since 1992. Cochran charges, however, that the program is actually intended to give the U.S. the capability to design new and more effective nuclear weapons. The DOE is working hard at developing supercomputers 100,000 times faster than today's machines. These devices, Cochran judges, will be able to simulate the explosion of the "physics package" in a warhead with unprecedented precision, from first principles of physics. The $4.5 billion to $5 billion being spent every year on the program is vastly more than necessary for the relatively simple job of keeping the existing stockpile safe and reliable, Cochran asserts. He says that of the other nuclear states, only France is in a position to pursue a similar course.

If Cochran is correct, the purpose of the stockpile stewardship program rests uneasily with the administration's stated policy of not developing new nuclear weapons. At least one new weapon has arrived already, Cochran points out. The recently introduced B61-11, a ground-penetrating warhead has crucial military advantages over its nonpiercing predecessor, the B61-7, Cochran states, even though the physics package is similar. Yet the DOE, as a semantic decoy, calls the B61-11 a mere modification.

The stockpile stewardship effort has ballooned because of political pressure from the weapons laboratories, Cochran charges. So he is unapologetic about dogging the program with a lawsuit brought under the National Environmental Policy Act, which requires comprehensive environmental impact statements for major projects. The same action challenges the lack of an environmental impact statement for cleanup efforts at the DOE's weapons laboratories.

Cochran does his homework. Half the floor of his office is taken up with 40 fat ring binders containing 1,600 documents on the DOE's environmental studies. Cochran the information warrior says he has looked through all of them. And he and his colleagues have other irons in the fire as well. Cochran has a particular interest in opposing commerce in weapons-usable material such as plutonium and highly enriched uranium and in maintaining scientific contacts with other nuclear powers.



ANTIBALLISTIC-MISSILE SITES around Moscow would likely be among the first targets destroyed in a nuclear attack on Russia.

In 1986 he made headlines when he took 20 tons of seismic-measuring equipment to the then Soviet Union's main nuclear test site in Kazakhstan to monitor shock waves from tests. Soviet scientists later came to monitor testing at the Nevada site. The project was startling for the degree of cooperation the Soviets offered. It got under way when von Hippel set up a meeting in which Cochran presented his plan to Evgeny P. Velikhov, a vice president of the Soviet Academy of Sciences. Velikhov, who was close to then Soviet President Mikhail S. Gorbachev, "immediately saw the political implications," Cochran tells.

The project demonstrated the feasibility of utilizing seismic monitoring to verify a low-threshold test ban. It was later taken over by the government and earned Cochran the American Physical Society's Szilard Award. The American Association for the Advancement of Science also acknowledged the project by giving the Natural Resources Defense Council its Award for Scientific Freedom and Responsibility.

In the late 1980s Cochran led another U.S. team that used radiation detectors near a live warhead on a Soviet cruiser, to prove the detectors could verify arms-control limits. On this visit he flew around the Soviet Union in the minister of defense's private plane. Cochran also escorted congressional delegations to sensitive Soviet military installations, such as the Krasnoyarsk early-warning radar in central Siberia.

Now Cochran is trying to pull off the same trick in China. He says he has developed "good relations" with a number of influential figures in the Chinese nuclear weapons program, including Hu Side, head of the Chinese Academy of Engineering Physics, and his deputy Du Xiangwan. Both are "very strong arms-control advocates," Cochran insists. An international group of physicists now meets with Chinese weapons experts every couple of years to discuss arms control and environmental policy, Cochran reports. He says the U.S. government has never asked him to acquire specific information, although he has been "informally debriefed" after some foreign visits.

Von Hippel says Cochran's early activism has served as a model for other environmental groups. Most have now realized that they must master the technical complexities of their subject to be taken seriously in policy circles. For those who would emulate his career trajectory, Cochran cites some advice he heard about propaganda: "Always tell the truth. Always make understatements. And talk to your enemies."

—Tim Beardsley in Washington, D.C.

EXPERIMENTAL SURGERY

WHEN LESS IS MORE

Trying to assess how well trimming hearts and lungs improves function

In May 1996, when a bedridden Jim Absalom of Youngstown, Ohio, had just learned that a heart transplant was his only other option, two doctors at the Cleveland Clinic made him an offer he couldn't refuse. Cardiologist Randall C. Starling and transplant surgeon Patrick M. McCarthy had recently been to Hospital Angelina Caron in Curitiba, Brazil, to see Randas Batista perform a revolutionary operation that could give Absalom's own heart a new lease on life. Afraid he would die before a donor heart became available, the 65-year-old agreed to the surgery.

Absalom's problem was congestive heart failure, a disorder in which the left ventricle of the heart cannot pump enough blood to the rest of the body because the ventricle has gotten too large and flabby. Batista pioneered the procedure in 1983, when he was working at a hospital where transplants were not done. It is based on the premise that making the ventricle smaller by taking a wedge out of its muscular wall will improve its efficiency.

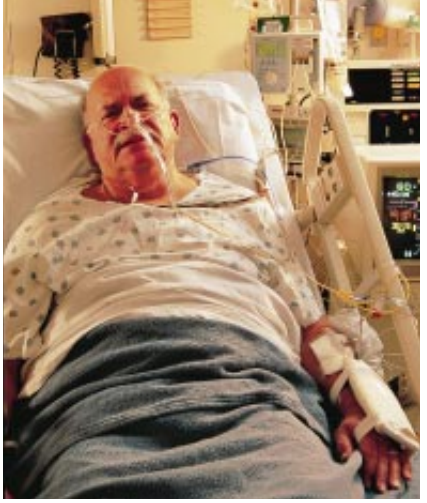
Although this approach defies the conventional medical wisdom that heart muscle should be preserved at all costs, Absalom reports that the surgery has worked well for him. "I have to be careful not to overdo it," he says, "but I now climb stairs without getting breathless or having chest pain. In fact, regular exercise is a part of my treatment. And instead of having to take a dozen different medicines, I now need only three."

To date, Starling and McCarthy have steered 59 patients through the downsizing procedure, the largest such series next to the 500-plus that Batista has done. Of those who are at least a year past the surgery, 82 percent are alive—about the same as transplant patients. They have, besides, been spared the serious side effects and high cost of antirejection drugs transplant recipients must take for life.

McCarthy and Starling are not so naive as to think Batista's surgery could help all congestive heart failure sufferers, of whom there are almost five million in the U.S. alone. The disease has become the leading cause of hospitalization for people 65 and older (400,000 new cases a year now, twice that projected by 2030). Half of those diagnosed with it die within five years, and few survive more than 10.

Rather McCarthy and Starling seek to determine which patients will really benefit from the operation and for how long. "We start with the reality that 20 percent of transplant candidates die while waiting for a donor heart," Starling says. "So we accept only patients with severe congestive failure and arrange in advance to list them for transplant should the Batista procedure fail. When an operation is new, as this one is, it makes sense to concentrate on the most straightforward cases—to best learn whether the procedure is worthwhile."

Once McCarthy and Starling have



B. SEITZ, Photo Researchers, Inc.

CARDIOVASCULAR PATIENTS could benefit from a controversial, organ-reducing procedure.

more follow-up data on their patients, larger and more formal trials of the surgery may take place. The method has been slow to catch on because Batista began without first doing systematic animal studies. That, plus lack of good patient follow-up, has made both him and his operation controversial with many in this country's medical establishment.

Also controversial—though less so—is a downsizing operation for emphysema patients. A disorder chiefly of smokers, emphysema causes the lungs to become chronically overinflated, which in turn puts a squeeze on the diaphragm and chest wall. It results in shortness of breath that, when severe, tethers the patient to bottled oxygen. The surgery, which is for such cases only, removes the lungs' most damaged parts to optimize the function of the rest and give the chest wall and diaphragm more room to move.

Though done a few times in the 1950s, the surgery was forgotten until some 40 years later, when Joel D. Cooper of the Washington University School of Medicine refined the technique by, among other things, devising a way to prevent fragile emphysemic lungs, which have the consistency of cotton candy, from springing air leaks when cut. Basically, he used membrane strips from cattle hearts to reinforce stress points.

After Cooper documented dramatic improvement in his first 20 patients, interest in the procedure revived. So much so, in fact, that claims submitted to Medicare for surgical treatment of emphysema—which had run to 200 to 300 a year—rose to more than 2,000 in 1995. But when the Health Care Financing Administration (HCFA), which runs Medicare, reviewed 700 claims for lung reduction, it found that 26 percent of patients died within 15 months. Not knowing why, the agency stopped routinely paying for the surgery in 1996.

The HCFA has since decided to join forces with the National Heart, Lung and Blood Institute (NHLBI) to get a better sense of the procedure's safety and efficacy. Some 4,700 patients will be recruited for a trial of up to five years that is now gearing up at 21 hospitals across the nation. Half the patients will be treated only nonsurgically; the rest will also have both lungs made smaller.

"Two questions are most at issue here," says Gail G. Weinmann, NHLBI project officer for the study. "One is whether patients have more to gain from surgery and maximal medical therapy than from maximal medical therapy alone. The other is that, if so, are there nonetheless some for whom it can be predicted that they will do as well or better without the surgery."

These are tricky questions and important to answer. Not only are they in the best interest of patients, they are also in the best interest of the wise use of health care dollars.

—Judith Randal in Lovettsville, Va.

RESISTANCE FIGHTING

Will natural selection outwit the king of biopesticides?

The soil bacterium known as *Bacillus thuringiensis,* or *Bt,* has remained the cornerstone of natural pest-control efforts for more than three decades. It's not hard to understand why. The bacterial toxins generally kill the bad guys (corn earworm and Colorado potato beetle, among others) and spare the good guys (humans, other mammals and beneficial insects).

But the rapidly growing acreage of corn, cotton and potatoes that are genetically engineered to produce *Bt*'s pesticidal toxins highlights fears about emerging insect resistance. Genetically engineered *Bt* crops provide exposure to the toxins throughout the growing season, leading to selection pressures that might enable only resistant pests to survive. In contrast, *Bt* sprays, which are used by home gardeners and organic and other farmers, degrade quickly, making resistance less likely.

A recent report by six prominent entomologists called on the Environmental Protection Agency to adopt more stringent requirements for *Bt* resistance management. "If *Bt* is lost in the next five years, there's a question about whether there will be a replacement or not," says Fred L. Gould, an entomologist at North Carolina State University and one of the authors.

The report for the Union of Concerned Scientists (UCS)—*Now or Never: Serious New Plans to Save a Natural Pest Control*—calls for strengthening requirements for the centerpiece of resistance management, a strategy called "refuge/high dose." Insects must be exposed to sufficient levels of *Bt* toxin from the plants to kill almost all pests. The rare resistant survivors are then apt to mate with susceptible pests bred in selected areas (refuges) that are not planted with *Bt* crops. Hybrid offspring then remain susceptible to *Bt* toxins.

The report's scientists call for expanding the size of existing refuges for cotton and corn, ensuring that these areas are close enough to the *Bt* crops to be effective, and for making establishment of refuges mandatory. (The EPA requires refuges only for *Bt* cotton and has asked seed producers to submit plans for protecting corn crops this year.)

The report notes that the need for expanded refuges—to as much as half of the planted acreage—is underlined by *Bt* cotton's ability to kill only 60 to 90 percent of cotton bollworms, a situation that became apparent during a large outbreak of the pests in the south during the summer of 1996. An EPA scientific advisory panel, some of whose members were authors of the UCS document, was scheduled to release a series of resistance-management recommendations to the agency sometime this spring.

So far no resistance to genetically engineered *Bt* crops has been documented. But several pests have shown resistance in the laboratory, and a vegetable pest, the diamondback moth, has demonstrated resistance after intensive field exposure to *Bt* sprays. "I call it the moth that roared," says Bruce E. Tabashnik of the University of Arizona about his work with diamondbacks. "It has sent a clear message that *Bt* resistance can evolve in open-field populations of a major crop pest."

Monsanto, the biggest marketer of *Bt* crops, does not foresee any need for strengthening existing protective measures nor for federal regulation of resistance-management plans. Officials cite the company's mandate that farmers should establish refuges for *Bt* corn and potatoes, even in the absence of an EPA requirement.

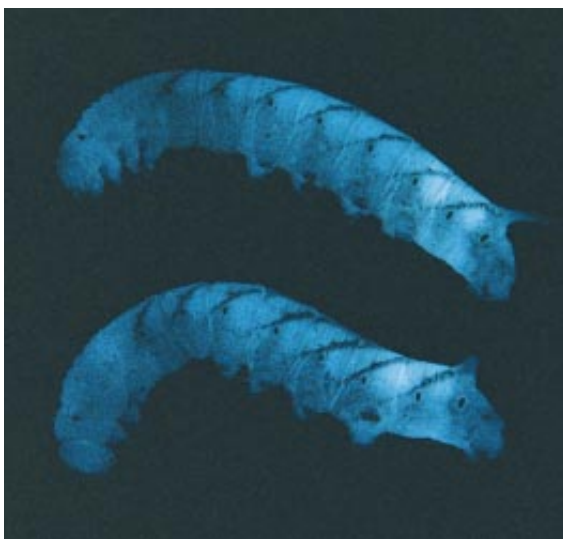
Eric Sachs, business director for YieldGard, the company's *Bt* corn product, says scientists who demand larger refuges fail to take into account the potential for improving genetically engineered crops. "Resistance is unlikely to happen within five years, and within that time frame we'll offer new technology that will further reduce the likelihood of resistance," Sachs comments.

One way for insects to develop resistance is by altering receptors in the gut where the toxins bind. Monsanto and others are working on "gene stacking": engineering of plants that express multiple toxins that bind different classes of receptors or, alternatively, combining *Bt* toxins with other proteins that disrupt the insect's life cycle.

Even these approaches may not be foolproof. Tabashnik, Gould and others have shown that some pests can evolve resistance to multiple *Bt* toxins that target different receptors and that resistance can evolve as a dominant trait. Pests can also develop new mechanisms of resistance. Brenda K. Oppert and William H. McGaughey, along with other U.S. Department of Agriculture researchers, reported in the *Journal of Biological Chemistry* in September that one pest, the Indianmeal moth, can become resistant if it lacks a key enzyme, a proteinase, that is needed to activate *Bt* toxins.

Over time, biopesticides that serve as alternatives to *Bt* crops may be needed. A study presented at the Entomological Society of America meeting last December by researchers at the University of Wisconsin–Madison discussed the cloning of genes for a toxin from a bacterium that inhabits the guts of nematodes. Just a few cells of *Photorhabdus luminescens* can kill some insect pests. A bacterial enzyme has the unusual property of making the dying insect glow blue, perhaps to scare off mammalian predators. Dow AgroSciences is now trying to insert the cloned genes into a variety of plants, which can then produce their own pesticide.

Even if *Bt* alternatives can be found, the bugs may ultimately win. Some 500 insect species have developed resistance to synthetic pesticides. Natural selection may ultimately prove a match for natural pesticides as well. —*Gary Stix*



UNIVERSITY OF WISCONSIN–MADISON

GHOULISH GLOW
emanates from insects killed by the bacterium* Photorhabdus luminescens, *which may yield new pesticides.

OPTOELECTRONICS

SEEING THE LIGHT

CMOS image sensors are poised to take on CCDs

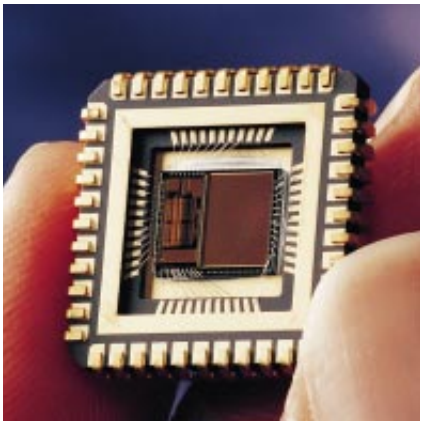
One of the success stories of the electronic age is the charge-coupled device (CCD), which shows up almost everywhere an image has to be converted to electrical signals. CCDs are the electronic, infinitely reusable "film" in digital cameras and are also critical components of countless other consumer products.

Entrenched as they are, though, CCDs are about to face a major challenge. The upstart competitor is the CMOS image sensor, which has been under development for years and is finally about to take the market by storm, thanks to heavy recent investments by the likes of Eastman Kodak, Motorola, Toshiba, Intel, Rockwell and Sarnoff.

An anticipated wave of modestly priced or novel consumer items based on CMOS sensors is expected to boost the market share of CMOS image sensors over the next few years. The sensors captured about 1.5 percent of the $678 million spent on image sensors in 1996. By 2001, they could account for at least 9 percent of a market worth $1.564 billion, according to Brian O'Rourke, an analyst who follows image sensors for Strategies Unlimited, a market research firm in Mountain View, Calif.

The advantages of CMOS sensors stem from the fact that they are fabricated using essentially the same complementary metal oxide semiconductor process as the vast majority of modern integrated circuits, such as microprocessors and dynamic random-access memories. CCDs, on the other hand, are fabricated using a variant of a largely obsolete fabrication technology called N- (for "n-channel") MOS; the process is no longer used for anything except CCDs.

Another advantage that the CMOS sensors share with all CMOS chips is very low power consumption. Perhaps most significant, the CMOS sensors can be directly integrated with other circuitry—for example, for analog-to-digital conversion or image processing—resulting in further savings in cost, power consumption and size. This feature has led some observers to predict that it is



SARNOFF CORPORATION; LEIGH PHOTOGRAPHIC GROUP

CMOS IMAGE SENSOR is integrated with other electronics.

only a matter of time before some manufacturer introduces an extraordinarily portable "camera on a chip."

Their economy and frugality with power are also expected to open up entirely new applications for CMOS image sensors. For example, the idea of putting one in a small cell phone to produce a pocket videophone is now quite feasible, according to Michael D. McCreary, director of operations for Kodak's microelectronics division. Other possible applications include ones in personal digital assistants, pagers and even toys. "I have a feeling that by 2001 they will be using these things in ways we really can't even imagine now," O'Rourke says.

With such advantages, why won't CMOS sensors replace CCDs? Because at present there is a trade-off, and it is in image quality. CMOS image sensors are typically "noisier" than CCDs, meaning that unwanted signals from various sources degrade the image signal. To get around this problem, some producers—notably, Motorola, which is allied with Kodak in a CMOS sensor venture—are tweaking their CMOS fabrication lines to be better suited to producing sensors.

The drawback to modifying a fabrication line to produce CMOS sensors, says Gary W. Hughes, head of imaging technology at Sarnoff, is that it can detract from the fundamental advantage of making sensors out of CMOS: the economy of fabricating them on standard lines. Citing Sarnoff studies, Hughes says that by fabricating chips on standard lines it will be possible to get the price of an image sensor chip down to $6 or less, paving the way for a $200 digital camera—less than half the cost of current digital cameras. —*Glenn Zorpette*

MICROMACHINES

A TONGUE FOR LOVE

A microsensor "tongue" could detect spoiled foods at the checkout register

Computers, Bill Gates is fond of pointing out, lack most of the basic senses that humans often take for granted. Some advanced machines can hear and speak—poorly—but most are blind and oblivious to touch, smell and temperature. Electronic devices may soon gain a sense of taste, however, thanks to tiny electromechanical machines invented at Pennsylvania State University by husband-and-wife engineers Vijay K. and Vasundara V. Varadan. The Varadans and their collaborators presented their designs in March at a conference in San Diego.

Although the Penn State group has built only a few working prototypes of the so-called smart tongues, the Varadans predict that within a few years the devices will be cheap and sensitive enough to find myriad uses. Stuck inside milk cartons and juice bottles, some sensors might enable checkout scanners at the grocery market to detect the growth of unwanted bacteria, they suggest. Others might be placed in giant vats in food and chemical factories to monitor the blending of ingredients. Slightly different versions could be mounted on aircraft wings to alert pilots when ice begins to form.

There is good reason to believe that this is more than daydreams and speculation. The smart tongues have at least three strong advantages over traditional chemical sensors. First is cost. They are very small, measuring just a few millimeters on a side. And they are made from simple materials—silicon, quartz, aluminum—using the same process by which the cheapest computer chips are produced. So, Vasundara Varadan says, "We should be able to make them for pennies each in quantity."

A second benefit is that they are wireless. Unlike most other micromachines, these sensors both receive power and transmit their readings by radio waves or microwaves. With no batteries and no tether to a computer, "we can place the microsensors [which are chemically inert] inside milk cartons, flush against helicopter rotors, inside pipelines—almost anywhere," Vijay Varadan asserts.

The third innovation is that unlike biological tongues and most laboratory tests, the smart tongues can detect chemical changes without performing any chemical reactions. Instead they sense shifts in the viscosity, or stickiness, of fluids, as well as changes in their electrical and acoustic properties, by purely electromechanical means.

The secret, believe it or not, is Love waves. These good vibrations are not those the Beach Boys crooned about but those discovered by British geophysicist Augustus E. H. Love almost a century ago. Love calculated that as earthquakes send up-and-down ripples shuddering through the mantle, a secondary set of side-to-side vibrations should propagate through the earth's relatively thin crust. His observation, which at last enabled geologists to measure the thickness of the planetary crust, also happens to apply to microscopically thin layers.

The Varadans' machines contain two pairs of intermeshed combs sandwiched between a thin layer of silicon dioxide and a slab of quartz. When connected to a small antenna—"which could be as simple as a one-centimeter spiral of aluminum foil," Vasundara Varadan hastens to add—each comb will tune in to any radio signal whose wavelength matches the space between its tines. At the right tone (just above 900 megahertz), the combs begin to shimmy back and forth. The Love waves start flowing.

The waves race down the length of the chip, half of which is protected by a metal coating and half of which is exposed to whatever liquid is being tested. After bouncing off a wall, the Love waves return to the combs, where their vibrations are converted back into an electrical signal that radiates out through the antenna.

The chemical measurement happens during the Love waves' round trip through the top layer of silicon. Those exposed to the fluid will be slightly slower and weaker than those that were shielded. The difference shows up in the device's radio transmission, which a scanner compares to a database of normal values.

The Varadans found that placing a second sensor just a micron higher on the same chip adds another dimension of information that boosts the sensitivity of the system remarkably. With such

a contraption, "we can measure the number of pits in orange juice," Vijay Varadan claims. Other tests have demonstrated that the sensors can sense the hint of curdling in milk left at room temperature for six hours, he reports. The engineers have shown that the same sensors, once calibrated, reliably discriminate slightly spoiled fruit juice from fresh-squeezed, ice from water, and tap water from distilled.

Although such micromachines could presumably save a fortune in an industry that routinely throws out a large fraction of food because conservative "sell by" dates have expired, the Varadans report no luck getting research funding from that quarter. Like most other micromachine projects, theirs is still paid for by the military, which is more interested in smart bombs than smart tongues. Perhaps grocers just haven't felt the Love waves yet.

—W. Wayt Gibbs in San Diego

REFRIGERATION

A COOL IDEA

Will magnetic refrigerators come to your kitchen?

The refrigerator of the future may be sitting in a laboratory in Madison, Wis. It doesn't have an ice-cube maker or a vegetable crisper. Nor does it have the standard gas-compression cooling system. What it does have are two cylinders of powdered gadolinium—a dense, gray, rare-earth metal—and a superconducting magnet. The device is the first magnetic refrigerator working at near-room temperature to produce substantial amounts of cooling power—more than 500 watts, three times the power of a large household refrigerator.

Carl B. Zimm, senior scientist at Astronautics Corporation of America, led the development of the magnetic refrigerator under a contract with the U.S. Department of Energy's Ames Laboratory. It relies on the magnetocaloric effect, the ability of ferromagnetic materials to heat up in the presence of a magnetic field and cool down when the field is removed. When a ferromagnet, such as gadolinium, is placed in a magnetic field, the magnetic moments of its atoms become aligned, making the material more ordered. But the amount of entropy in the magnet must be conserved, so the atoms vibrate more rapidly, raising the material's temperature. Conversely, when the gadolinium is taken out of the field, the material cools.

Refrigeration systems taking advantage of this effect have long been used by scientists to cool rare-earth oxides to several thousandths of a degree above absolute zero (liquid helium cools the oxides to 1.4 kelvins, and then they are demagnetized). But commercial applications have been stymied by the fact that the magnetocaloric effect is relatively weak in most ferromagnetic materials at room temperature. Gadolinium, however, is an exception. Each atom of gadolinium has seven unpaired electrons in an intermediate shell, giving the element a strong magnetic moment. What is more, the magnetocaloric effect reaches its maximum at the Curie temperature—the transition point above which a material is no longer ferromagnetic—and that point for gadolinium is 20 degrees Celsius (68 degrees Fahrenheit). In contrast, the Curie temperature for iron is 770 degrees C.

Even under ideal conditions, the magnetocaloric effect is not huge. The powerful superconducting magnet in Zimm's refrigerator produces a maximum temperature change of only 14 degrees C in the cylinders of gadolinium. But the machine uses an ingenious regeneration system to increase its cooling power. Water is pumped into one of the cylinders of gadolinium immediately after it moves out of the magnetic field. The water cools as it moves through the porous bed of demagnetized gadolinium, then flows through a heat exchanger. Next, the water passes through the cylinder of gadolinium that is inside the magnetic field. The water heats up and flows through another exchanger, providing ample refrigeration power by continually heating one exchanger and cooling the other. After a preset interval, the two cylinders of gadolinium switch places, and the flow of water is reversed. Antifreeze can be added to the water to allow the machine to cool below 0 degree C.

Zimm's refrigerator is remarkably efficient because very little energy is lost during the magnetic warming and cooling. The experimental prototype has run at 30 percent of the Carnot limit—the maximum possible efficiency for a refrigerator—which is comparable with the efficiency of most household units. Zimm believes that a larger magnetic refrigerator could operate at 70 percent of the limit, making it competitive with the best industrial-scale refrigerators. Two of his colleagues at Ames, Karl A. Gschneidner, Jr., and Vitalij K. Pecharsky, have recently discovered that a class of gadolinium alloys—mixed with silicon and germanium—exhibit an even greater magnetocaloric effect. Plans are under way to test these alloys in Zimm's machine.

The biggest obstacle to the development of magnetic refrigerators is the cost of the superconducting magnet. At least in the near future, commercial uses would be limited to large-scale operations—such as supermarket freezers and air-conditioning systems for office buildings. But if the magnetocaloric effect can be sufficiently enhanced, the device may be able to run efficiently with the weaker field generated by a permanent magnet. Then the magnetic fridge could well become a standard fixture of the 21st-century kitchen. *—Mark Alpert*

DAVID SANDELL *The Capital Times*

MAGNETIC REFRIGERATOR ***made by Carl Zimm may soon have industrial uses.***

CYBER VIEW

Bringing Down the Internet

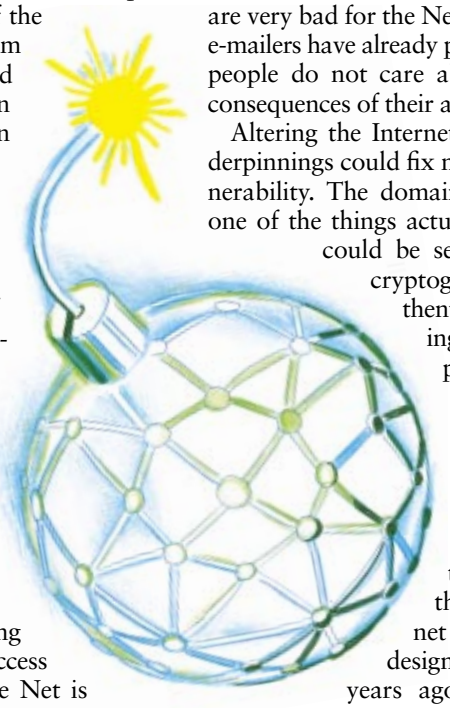
"How many bombs would it take to bring the Net down, and where would you drop them?" This kind of question can silence a roomful of Netheads, and so it did at a panel presented at the Computers, Freedom and Privacy Conference, held this past February in Austin, Tex. The group needs no reminding that, historically, the Net was built to withstand precisely that: nuclear-bomb outages. "Zero" was the consensus of Matt A. Blaze and Steve M. Bellovin, both researchers specializing in network security at AT&T and both Net heroes—Blaze for leading the technical wing of the protests against government regulation of cryptography and Bellovin for being one of the three 1979 creators of Usenet.

"Zero" is not good news, because it turns out there are far more effective ways of crippling the Internet, whether by malice or accident. In the past year several incidents, most accidental, have demonstrated the problem, all of which had to do with the routing of traffic around the Net. Routers depend on having access to accurate information to match numbered addresses to named domains, such as ".com." That information is stored in about a dozen worldwide root servers, which are top-level computers that hold the database that matches names and addresses.

Routers make easy targets. For about 10 days starting in late January, for instance, longtime Net expert Jon Postel instructed the administrators of about half those dozen root servers to get their updates from him rather than from Network Solutions, which manages the assignment of named Internet addresses. The results could have been disastrous had Postel relayed incorrect updates. Postel, who runs the Internet Assigned Numbers Authority, which hands out numbered Internet addresses, said afterward that it was a test in preparation for the revamping of the domain name system.

Less reputably, last summer Eugene Kashpureff of the renegade AlterNIC domain name service diverted traffic intended for Network Solutions's World Wide Web site to his own, using a technique called spoofing. (Don't try this at home: Kashpureff now faces sentencing for computer fraud.) And last July a large chunk of the Net was cut off when Network Solutions bungled a database update, allowing a corrupted version of the file to be propagated. "I live in fear that somebody with a malicious bent of mind will notice these accidents and say, 'Gee, I could do that, too,'" Bellovin said in recounting some of these incidents at the conference.

Underlying all these events is the same basic fact: routers believe what they're told by other routers. If a misconfigured router in a small company convinces an upstream router that it is the best path to much of the rest of the Internet, the upstream router will act on it and pass the information on to other routers. When this happened last year, it took hours for all parts of the system to stabilize, even though the misconfigured link was shut down in half an hour. "What if somebody did that deliberately from a few different points? This could be a massive denial-of-service attack on the Net—or an eavesdropping attack," Bellovin said.



This might be possible, because, according to Blaze, "powerful access to the structure of the Net is available to everyone on the Net, and second of all, every machine on the Net can be controlled, to some extent, remotely by any other machine on the Net." Your $19.95 a month gives more than just access to information; it enables you, if you are knowledgeable and malicious, to send false packets of data to the routers and reprogram them. This puts the Net in danger from everything from software bugs to human error, mechanical failure or deliberate damage. More localized Net failures in the past year have involved hardware problems such as sliced cables or unplanned electrical blackouts.

Much of the Internet's functioning has also always depended on voluntary good citizenship. The Internet protocols, which indicate how data are to be delivered, are designed so that end points on the Net—a company's network or an individual machine—behave in ways that are best for the network as a whole. For example, if a packet fails to get through, the sending machine waits longer and longer to try again, on the assumption that the problem is network congestion, rather than clog the Net by trying continuously until the data get through. "Right now all the vendors of software that run these protocols distribute well-behaved software that does this global optimization," Blaze said. "But a vendor could do well on its own benchmarks by behaving in ways that are very bad for the Net at large." Junk e-mailers have already proved that some people do not care about the global consequences of their actions.

Altering the Internet's technical underpinnings could fix much of this vulnerability. The domain name system, one of the things actually centralized, could be secured by using cryptography for authentication. According to Blaze, such protocols exist: "It's a matter of getting them deployed." Upgrading the Internet protocols would help, too. Blaze noted that secure Internet protocols were designed about 10 years ago, and there is software to run them. "If everyone in the room could convince a million of their friends to run it, the Net would be a lot safer," he said.

All of this matters because of the widespread, giddy assumption that the Net really is as invulnerable as its hype would have us believe. If it's not, we need to rethink: we can make the Net more secure by deploying better technology; we can reengineer our social structures to take into account an insecure and somewhat flaky infrastructure and incorporate ways of coping when failure inevitably happens; or we can refuse to rely on the Net for certain types of uses. In any case, we should remember that where the Net is concerned there are worse things than bombs.

—Wendy M. Grossman in Austin, Tex.

DAVID SUTER

ASTRONAUT SHANNON W. LUCID on board the Mir space station during her six-month mission.

NASA/RUSSIAN SPACE AGENCY

Six Months on Mir

As the Shuttle-Mir program draws to a close, a veteran NASA astronaut reflects on her mission on board the Russian spacecraft and the implications for the International Space Station

by Shannon W. Lucid

For six months, at least once a day, and many times more often, I floated above the large observation window in the Kvant 2 module of Mir and gazed at the earth below or into the depths of the universe. Invariably, I was struck by the majesty of the unfolding scene. But to be honest, the most amazing thing of all was that here I was, a child of the pre-Sputnik, cold war 1950s, living on a Russian space station. During my early childhood in the Texas Panhandle, I had spent a significant amount of time chasing windblown tumbleweeds across the prairie. Now I was in a vehicle that resembled a cosmic tumbleweed, working and socializing with a Russian air force officer and a Russian engineer. Just 10 years ago such a plot line would have been deemed too implausible for anything but a science-fiction novel.

In the early 1970s both the American and Russian space agencies began exploring the possibility of long-term habitation in space. After the end of the third Skylab mission in 1974, the American program focused on short-duration space shuttle flights. But the Russians continued to expand the time their cosmonauts spent in orbit, first on the Salyut space stations and later on Mir, which means "peace" in Russian. By the early 1990s, with the end of the cold war, it seemed only natural that the U.S. and Russia should cooperate in the next major step of space exploration, the construction of the International Space Station. The Russians formally joined the partnership—which also includes the European, Japanese, Canadian and Brazilian space agencies—in 1993.

The first phase of this partnership was the Shuttle-Mir program. The National Aeronautics and Space Administration planned a series of shuttle missions to send American astronauts to the Russian space station. Each astronaut would stay on Mir for about four months, performing a wide range

of peer-reviewed science experiments. The space shuttle would periodically dock with Mir to exchange crew members and deliver supplies. In addition to the science, NASA's goals were to learn how to work with the Russians, to gain experience in long-duration spaceflight and to reduce the risks involved in building the International Space Station. Astronaut Norm Thagard was the first American to live on Mir. My own arrival at the space station—eight months after the end of Thagard's mission—was the beginning of a continuous American presence in space, which has lasted for more than two years.

My involvement with the program began in 1994. At that point, I had been a NASA astronaut for 15 years and had flown on four shuttle missions. Late one Friday afternoon I received a phone call from my boss, Robert "Hoot" Gibson, then the head of NASA's astronaut office. He asked if I was interested in starting full-time Russian-language instruction with the possibility of going to Russia to train for a Mir mission. My immediate answer was yes. Hoot tempered my enthusiasm by saying I was only being assigned to study Russian. This did not necessarily mean I would be going to Russia, much less flying on Mir. But because there was a possibility that I might fly on Mir and because learning Russian requires some lead time—a major understatement if ever there was one—Hoot thought it would be prudent for me to get started.

I hung up the phone and for a few brief moments stared reality in the face. The mission on which I might fly was less than a year and a half away. In that time I would have to learn a new language, not only to communicate with my crewmates in orbit but to train in Russia for the mission. I would have to learn the systems and operations for Mir and Soyuz, the spacecraft that transports Russian crews to and from the space station. Because I would be traveling to and from Mir on the space shuttle, I needed to maintain my familiarity with the American spacecraft. As if that were not enough, I would also have to master the series of experiments I would be conducting while in orbit.

It is fair at this point to ask, "Why?" Why would I wish to live and work on Mir? And from a broader perspective, why are so many countries joining together to build a new space station? Certainly one reason is scientific research. Gravity influences all experiments done on the earth except for investigations conducted in drop towers or on airplanes in parabolic flight. But on a space station, scientists can conduct long-term investigations in an environment where gravity is almost nonexistent—the microgravity environment. And the experience gained by maintaining a continuous human presence in space may help determine what is needed to support manned flights to other planets.

From a personal standpoint, I viewed the Mir mission as a perfect opportunity to combine two of my passions: flying airplanes and working in laboratories. I received my private pilot's license when I was 20 years old and have been flying ever since. And before I became an astronaut, I was a biochemist, earning my Ph.D. from the University of Oklahoma in 1973. For a scientist who loves flying, what could be more exciting than working in a laboratory that hurtles around the earth at 17,000 miles (27,000 kilometers) per hour?

After three months of intensive language study, I got the go-ahead to start my training at Star City, the cosmonaut training center outside Moscow. My stay there began in January 1995, in the depths of a Russian winter. Every morning I woke at five o'clock to begin studying. As I walked to class I was always aware that one misstep on the ice might result in a broken leg, ending my dreams of a flight on Mir. I spent most of my day in classrooms listening to Mir and Soyuz system lectures—all in Russian, of course. In the evenings I continued to study the language and struggled with workbooks written in technical Russian. At midnight I finally fell exhausted into bed.

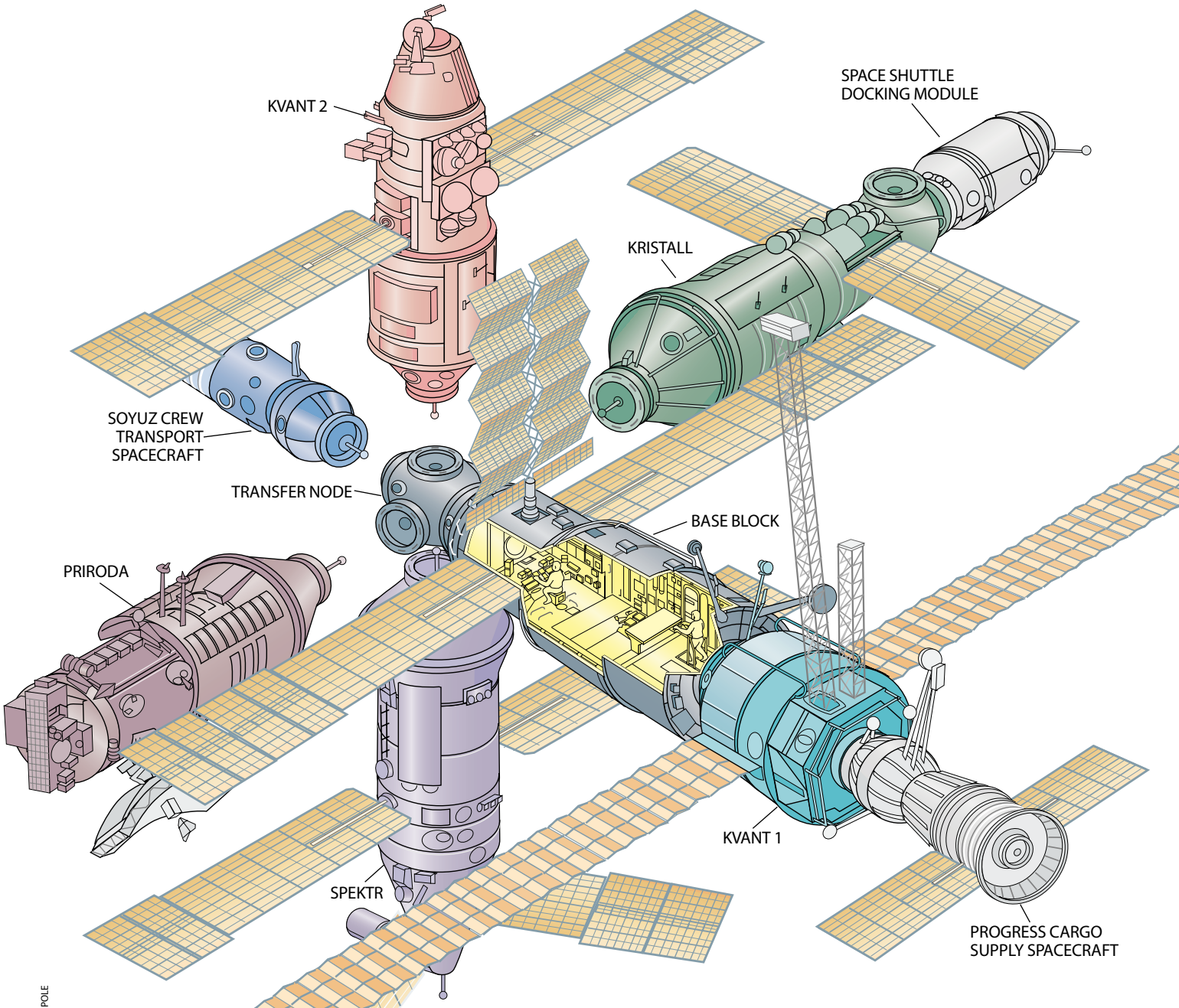
NASA/RUSSIAN SPACE AGENCY

FLOATING FREELY in the Base Block of Mir, Lucid posed for a patriotic snapshot taken by one of her cosmonaut crewmates on July 4, 1996.

I worked harder during that year than at any other time in my life. Going to graduate school while raising toddlers was child's play in comparison. (Fortunately, my three children were grown by this point, and my husband was able to visit me in Russia.) At last, in February 1996, after I had passed all the required medical and technical exams, the Russian spaceflight commission certified me as a Mir crew member. I traveled to Baikonur, Kazakhstan, to watch the launch of the Soyuz carrying my crewmates—Commander Yuri Onufriyenko, a Russian air force officer, and flight engineer Yuri Usachev, a Russian civilian—to Mir. Then I headed back to the U.S. for three weeks of training with the crew of shuttle mission STS-76. On March 22, 1996, we lifted off from the Kennedy Space Center on the shuttle *Atlantis*. Three days later the shuttle docked with Mir, and I officially joined the space station crew for what was planned to be a four-and-a-half-month stay.

Living in Microgravity

My first days on Mir were spent getting to know Onufriyenko and Usachev—we spoke exclusively in Russian—and the layout of the space station. Mir has a modular design and was built in stages. The first part, the Base Block, was launched in February 1986. Attached to one end of the Base Block is Kvant 1, launched in 1987, and at the other end is Mir's transfer node, which serves the same function as a hallway in a house. Instead of being a long corridor with doors, though, the transfer node is a ball with six hatches. Kvant 2 (1989), Kristall (1990) and Spektr (1995) are each docked to a hatch. During my stay on Mir, the Russians launched Priroda, the final module of the space station, and attached it to the transfer node. Priroda contained the laboratory where I conducted most of my experiments. I stored my personal belongings in Spektr and slept there every night. My commute to work was very short—in a matter of seconds I could float from one module to the other.



IAN WORPOLE

EXPLODED VIEW of Mir shows how the space station's modules fit together. Inside the Base Block is the command post, one of the treadmills and the table where the crew eat their meals.

The two cosmonauts slept in cubicles in the Base Block. Most mornings the wake-up alarm went off at eight o'clock (Mir runs on Moscow time, as does the Russian mission control in Korolev). In about 20 minutes we were dressed and ready to start the day. The first thing we usually did was put on our headsets to talk to mission control. Unlike the space shuttle, which transmits messages via a pair of communications satellites, Mir is not in constant contact with the ground. The cosmonauts can talk to mission control only when the space station passes over one of the communications ground sites in Russia. These "comm passes" occurred once an orbit—about every 90 minutes—and generally lasted about 10 minutes. Commander Onufriyenko wanted each of us to be "on comm" every time it was available, in case the ground needed to talk to us. This routine worked out well because it gave us short breaks throughout the day. We gathered in the Base Block and socialized a bit before and after talking with mission control.

After the first comm pass of the day, we ate breakfast. One of the most pleasant aspects of being part of the Mir crew was that we ate all our meals together, floating around a table in the Base Block. Preflight, I had assumed that the repetitive nature of the menu would dampen my appetite, but to my surprise I was hungry for every meal. We ate both Russian and American dehydrated food that we reconstituted with hot water. We experimented with mixing the various packages to create new tastes, and we each had favorite mixtures that we recommended to the others. For breakfast I liked to have a bag of Russian soup—usually borscht or vegetable—and a bag of fruit juice. For lunch or supper I liked

the Russian meat-and-potato casseroles. The Russians loved the packets of American mayonnaise, which they added to nearly everything they ate.

Our work schedule was detailed in a daily timeline that the Russians called the Form 24. The cosmonauts typically spent most of their day maintaining Mir's systems, while I conducted experiments for NASA. We had to exercise every day to prevent our muscles from atrophying in the weightless environment. Usually, we all exercised just before lunch. There are two treadmills on Mir—one in the Base Block and the other in the Kristall module—and a bicycle ergometer is stored under a floor panel in the Base Block. We followed three exercise protocols developed by Russian physiologists; we did a different one each day, then repeated the cycle. Each protocol took about 45 minutes and alternated periods of treadmill running with exercises that involved pulling against bungee cords to simulate the gravitational forces we were no longer feeling. Toward the end of my stay on Mir I felt that I needed to be working harder, so after I finished my exercises I ran additional kilometers on the treadmill.

I'll be honest: the daily exercise was what I disliked most about living on Mir. First, it was just downright hard. I had to put on a harness and then connect it with bungee cords to the treadmill. Working against the bungees allowed me to stand flat on the device. With a little practice, I learned to run. Second, it was boring. The treadmill was so noisy you could not carry on a conversation. To keep my mind occupied, I listened to my Walkman while running, but soon I realized I'd made a huge preflight mistake. I had packed very few tapes with a fast beat. Luckily, there was a large collection of music tapes on Mir. During my six-month stay, I worked through most of them.

When we had finished exercising, we usually enjoyed a long lunch, then returned to our work. Many times in the late afternoon we had a short tea break, and in the late evening we shared supper. By this point we had usually finished all the assignments on the Form 24, but there were still many housekeeping chores that needed to be done: collecting the trash, organizing the food supply, sponging up the water that had condensed on cool surfaces. Clutter was a problem on Mir. After we had unloaded new supplies from the unmanned Progress spacecraft that docked with the space station once every few months, we could put human wastes and trash into the empty vehicles, which would burn up on reentry into the atmosphere. But there was usually no room left on Progress for the many pieces of scientific equipment that were no longer in use.

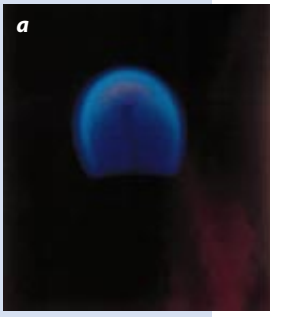
After supper, mission control would send us the Form 24 for the next day on the teleprinter. If there was time, we had tea and a small treat—

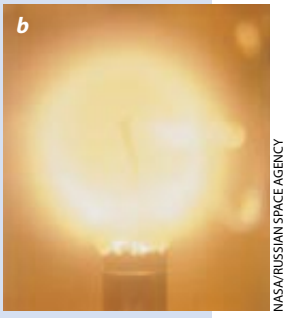
Fires in Free Fall

by Howard D. Ross

On the earth, flames are shaped by gravity. The heated gases rise, creating a buoyant airflow that feeds oxygen to the fire. But on Mir or on the space shuttle, there is no buoyant flow. Fires must get their oxygen from diffusion—the slow drift of molecules through the air—or from air currents generated by ventilation fans. The NASA Lewis Research Center has overseen experiments to determine how flames spread in a microgravity environment. Our research has shed light on the physics of combustion and helped NASA to better identify the risks of fires on spacecraft.

One of the first experiments on Mir was designed to answer a simple but long-standing question: Could a candle burn in microgravity? Using a NASA-built chamber that isolated the experiment from the space station's atmosphere, astronaut Shannon Lucid lit candles of various sizes by touching a loop of wire—heated by an electrical current—to each wick. Immediately after ignition each candle flame was hemispherical, with a bright yellow core. Because less oxygen was reaching the flame, it consumed candle wax nearly five times more slowly than it would on the earth. But the lack of an upward airflow increased the conduction of heat downward into the candle. Within two minutes the entire candle melted. The fire didn't go out, though; surface tension kept the swirling ball of liquid wax anchored to the wick and the candleholder.





NASA/RUSSIAN SPACE AGENCY

A candle that would burn for about 10 minutes on the earth stayed lit for up to 45 minutes on Mir. Four to 10 minutes after ignition the flames turned blue, becoming too cold to produce yellow, luminous soot (*a*). The flames were so weak that the video cameras on board the space station lacked the sensitivity to observe them, even when the lights in the combustion chamber were switched off. Lucid, however, peeked inside and used still photography and on-the-spot sketches to record the results. The biggest surprise came after the fire went out. When Lucid switched on the chamber lights, she observed a large spherical cloud surrounding the candle tip (*b*). We believe this cloud contained wax droplets condensing in the colder air. This phenomenon provided a valuable lesson for fighting fires on spacecraft: even after a fire is extinguished, flammable material can continue to issue from the source.

Astronauts on Mir and the space shuttle also studied the burning of plastic and cellulose materials, measuring how quickly the flames spread under different airflows. In a quiescent environment, where the air is still, most materials burn more slowly in microgravity than they would on the earth. But under a low-speed airflow—two to eight inches (five to 20 centimeters) per second—certain materials become more flammable. For example, a sheet of paper under a low-speed airflow in microgravity will burn 20 percent faster than paper under the same conditions on the earth. This result provided another lesson in spacecraft safety: the first line of defense against fires is to still the air by turning off the ventilation systems.

These investigations were only part of a much broader program. Other experiments conducted on the space shuttle focused on gas-jet fires, burning fuel droplets, hydrogen-air mixtures and the soot produced by fires in microgravity. The International Space Station will include a fully equipped combustion chamber, offering opportunities for further study.

HOWARD D. ROSS is a senior researcher at the Microgravity Science Division of the NASA Lewis Research Center in Cleveland, Ohio.

cookies or candy—before the last comm pass of the day, which usually occurred between 10 and 11 at night. Then we said good night to one another and went to our separate sleeping areas. I floated into Spektr, unrolled my sleeping bag and tethered it to a handrail. I usually spent some time reading and typing letters to home on my computer (we used a ham radio packet system to send the messages to the ground controllers, who sent them to my family by e-mail). At midnight I turned out the light and floated into my sleeping bag. I always slept soundly until the alarm went off the next morning.

Quail Eggs and Dwarf Wheat

Our routine on Mir rarely changed, but the days were not monotonous. I was living every scientist's dream. I had my own lab and worked independently for much of the day. Before one experiment became dull, it was time to start another, with new equipment and in a new scientific field. I discussed my work at least once a day with Bill Gerstenmaier, the NASA flight director, or Gaylen Johnson, the NASA flight surgeon, both at Russian mission control. They coordinated my activities with the principal investigators—the American and Canadian scientists who had proposed and designed the experiments. Many times when we started a new experiment, Gerstenmaier arranged for the principal investigators to be listening to our radio conversations, so they would be ready to answer any questions I might have. And this was in the middle of the night back in the U.S.!

My role in each experiment was to do the onboard procedures. Then the data and samples were returned to the earth on the space shuttle and sent to the principal investigators for analysis and publication. I believe my experience on Mir clearly shows the value of performing research on manned space stations. During some of the experiments, I was able to observe subtle phenomena that a video or still camera would miss. Because I was familiar with the science in each experiment, I could sometimes examine the results on the spot and modify the procedures as needed. Also, if there was a malfunction in the scientific equipment, I or one of my crew-

WORST PART of life on Mir: daily treadmill running (*top left*). The space station's main controls are in the Base Block command post (*top right*). Cosmonaut Yuri Usachev won a gelatin dessert for finding Lu-

cid's lost sneaker (*below*). Lucid used interlocking plastic bags to conduct an experiment with quail eggs (*bottom left*) and a specially built "glove box" for flame and fluid investigations (*bottom right*).

mates could usually fix it. Only one of the 28 experiments scheduled for my mission failed to yield results because of a breakdown in the equipment.

I started my work on Mir with a biology experiment examining the development of embryos in fertilized Japanese quail eggs. The eggs were brought to Mir on the same shuttle flight that I took, then transferred to an incubator on the space station. Over the next 16 days I removed the 30 eggs one by one from the incubator and placed them in a 4 percent paraformaldehyde solution to fix the developing embryos for later analysis. Then I stored the samples at ambient temperature.

This description makes it sound like a simple experiment, but it required creative engineering to accomplish the procedure in a microgravity environment. NASA and Russian safety rules called for three layers of containment for the fixative solution; if a drop escaped, it could float into a crew member's eye and cause severe burns. Engineers at the NASA Ames Research Center designed a system of interlocking clear bags for inserting the eggs into the fixative and cracking them open. In addition, the entire experiment was enclosed in a larger bag with gloves attached to its surface, which allowed me to reach inside the bag without opening it.

Investigators at Ames and several universities analyzed the quail embryos at the end of my mission to see if they differed from embryos that had developed in an incubator on the ground. Remarkably, the abnormality rate among the Mir embryos was 13 percent—more than four times higher than the rate for the control embryos. The investigators believe two factors may have increased the abnormality rate: the slightly higher temperature in the Mir incubator and the much higher radiation levels on the space station. Other experiments determined that the average radiation exposure on Mir is the equivalent of getting eight chest x-rays a day. NASA scientists believe, however, that an astronaut would have to spend at least several years in orbit to raise appreciably his or her risk of developing cancer.

PHOTOGRAPHS COURTESY OF NASA/RUSSIAN SPACE AGENCY

I was also involved in a long-running experiment to grow wheat in a greenhouse on the Kristall module. American and Russian scientists wanted to learn how wheat seeds would grow and mature in a microgravity environment. The experiment had an important potential application: growing plants could provide oxygen and food for long-term spaceflight. Scientists focused on the dwarf variety of wheat because of its short growing season. I planted the seeds in a bed of zeolite, an absorbent granular material. A computer program controlled the amount of light and moisture the plants received. Every day we photographed the wheat stalks and monitored their growth.

At selected times, we harvested a few plants and preserved them in a fixative solution for later analysis on the ground. One evening, after the plants had been growing for about 40 days, I noticed seed heads on the tips of the stalks. I shouted excitedly to my crewmates, who floated by to take a look. John Blaha, the American astronaut who succeeded me on Mir, harvested the mature plants a few months later and brought more than 300 seed heads back to the earth. But scientists at Utah State University discovered that all the seed heads were empty. The investigators speculate that low levels of ethylene in the space station's atmosphere may have interfered with the pollination of the wheat. In subsequent research on Mir, astronaut Michael Foale planted a variety of rapeseed that successfully pollinated.

The microgravity environment on the space station also provided an excellent platform for experiments in fluid physics and materials science. Scientists sought to further improve the environment by minimizing vibrations. Mir vibrates slightly as it orbits the earth, and although the shaking is imperceptible to humans, it can have an effect on sensitive experiments. The movements of the crew and airflows on the station can also cause vibrations. To protect experiments from these disturbances, we placed them on the Microgravity Isolation Mount, a device built by the Canadian Space Agency. The top half of the isolation mount floats free, held in place solely by electromagnetic fields.

After running an extensive check of the mount, I used it to isolate a metallurgical experiment. I placed metal samples in a specially designed furnace, which heated them to a molten state. Different liquid metals were allowed to diffuse in small tubes, then slowly cooled. The principal investigators wanted to determine how molten metals would diffuse without the influence of convection. (In a microgravity environment, warmer liquids and gases do not rise, and colder ones do not sink.) After analyzing the results, they learned that the diffusion rate is much slower than on the earth. During the procedure, one of the brackets in the furnace was bent out of alignment, threatening the completion of the experiment. But flight engineer Usachev simply removed the bracket, put it on a workbench and pounded it straight with a hammer. Needless to say, this kind of repair would have been impossible if the experiment had taken place on an unmanned spacecraft.

Many of the experiments provided useful data for the engineers designing the International Space Station. The results from our investigations in fluid physics are helping the space station's planners build better ventilation and life-support systems. And our research on how flames propagate in microgravity may lead to improved procedures for fighting fires on the station [see "Fires in Free Fall," by Howard D. Ross, on page 49].

Safety in Space

Throughout my mission I also performed a series of earth observations. Many scientists had asked NASA to photograph parts of the planet under varying seasonal and lighting conditions. Oceanographers, geologists and climatologists would incorporate the photographs into their research. I usually took the pictures from the Kvant 2 observation window with a handheld Hasselblad camera. I discovered that during a long spaceflight, as opposed to a quick space shuttle jaunt, I could see the flow of seasons across the face of the globe. When I arrived on Mir at the end of March, the higher latitudes of the Northern Hemisphere were covered with ice and snow. Within a few weeks, though, I could see huge cracks in the lakes as the ice started to break up. Seemingly overnight, the Northern Hemisphere glowed green with spring.

We also documented some unusual events on the earth's

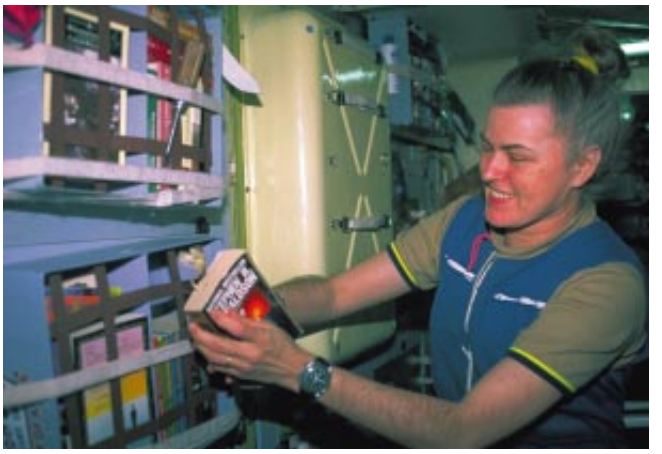
PHOTOGRAPHS COURTESY OF NASA/RUSSIAN SPACE AGENCY

SENSITIVE EXPERIMENTS were placed on the Microgravity Isolation Mount (*far left*) to protect them from vibrations. Lucid recorded her observations on a laptop computer (*above left*). Crew members injected each other with vaccines to test their immune systems (*above*) and collected blood samples (*above right*). A Russian device tested Lucid's cardiovascular system by creating a vacuum around the lower half of her body (*right*).

surface. One day as we passed over Mongolia we saw giant plumes of smoke, as though the entire country were on fire. The amount of smoke so amazed us that we told the ground controllers about it. Days later they informed us that news of huge forest fires was just starting to filter out of Mongolia.

For long-duration manned spaceflight, the most important consideration is not the technology of the spacecraft but the composition of the crew. The main reason for the success of our Mir mission was the fact that Commander Onufriyenko, flight engineer Usachev and I were so compatible. It would have been very easy for language, gender or culture to divide us, but this did not happen. My Russian crewmates always made sure that I was included in their conversations. Whenever practical, we worked on projects together. We did not spend time criticizing one another—if a mistake was made, it was understood, corrected and then forgotten. Most important, we laughed together a lot.

The competence of my crewmates was one of the reasons I always felt safe on Mir. When I began my mission, the space station had been in orbit for 10 years, twice as long as it had been designed to operate. Onufriyenko and Usachev had to spend most of their time maintaining the station, replacing parts as they failed and monitoring the systems critical to life support. I soon discovered that my crewmates could fix just about anything. Many spare parts are stored on Mir, and more are brought up as needed on the Progress spacecraft. Unlike the space shuttle, Mir cannot return to the earth for repairs, so the rotating crews of cosmonauts are trained to keep the station functioning.



PHOTOGRAPHS COURTESY OF NASA/RUSSIAN SPACE AGENCY

SPACE TOYS were gifts from Lucid's children, delivered by Progress (*top left*). Lucid read 50 books during her six months on Mir, keeping some in a zero-g bookcase (*above*). Lucid, Usachev and Commander Yuri Onufriyenko posed for a group shot in the Priroda module (*left*). They have stayed friends since their mission.

Furthermore, the crews on Mir have ample time to respond to most malfunctions. A hardware failure on the space shuttle demands immediate attention because the shuttle is the crew's only way to return to the earth. If a piece of vital equipment breaks down, the astronauts have to repair the damage quickly or end the mission early, which has happened on a few occasions. But Mir has a lifeboat: at least one Soyuz spacecraft is always attached to the space station. If a hardware failure occurs on Mir, it does not threaten the crew's safe return home. As long as the space station remains habitable, the crew members can analyze what happened, talk to mission control and then correct the malfunction or work around the problem.

Only two situations would force the Mir crew to take immediate action: a fire inside the space station or a rapid depressurization. Both events occurred on Mir in 1997, after I left the station [see "Learning from Disaster," on the opposite page]. In each case, the crew members were able to contain the damage quickly.

My mission on the space station was supposed to end in August 1996, but my ride home—shuttle mission STS-79—was delayed for six weeks while NASA engineers studied abnormal burn patterns on the solid-fuel boosters from a previous shuttle flight. When I heard about the delay, my first thought was, "Oh, no, not another month and a half of treadmill running!" Because of the delay, I was still on Mir when a new Russian crew arrived on the Soyuz spacecraft to relieve Onufriyenko and Usachev. By the time I finally came back on the shuttle *Atlantis* on September 26, 1996, I had logged 188 days in space—an American record that still stands.

This June, astronaut Andrew Thomas—the last of the seven NASA astronauts who have lived on Mir over the past three years—is scheduled to return to the earth, ending the Shuttle-Mir program. Based on my own experience, I believe there are several lessons that should be applied to the operation of the International Space Station. First, the station crew must be chosen carefully. Even if the space station has the latest in futuristic technology, if the crew members do not enjoy working together, the flight will be a miserable experience. Second, NASA must recognize that a long-duration flight is as different from a shuttle flight as a marathon is from a 100-yard dash. On a typical two-week shuttle flight, NASA ground controllers assign every moment of the crew's time to some task. But the crew on a long-duration flight must be treated more like scientists in a laboratory on the earth. They must have some control over their daily schedules.

Similarly, when a crew trains for a science mission on the space shuttle, the members practice every procedure until it can be done without even having to think about it. Training for a mission on the International Space Station needs to be different. When a crew member starts a new experiment on a long-duration flight, it might be up to six months after he or she trained for the procedure. The astronaut will need to spend some time reviewing the experiment. Therefore, their training should be skill-based. Crew members should learn the skills they will need during their missions rather than practice every specific procedure. Also, crew members on a long-duration flight need to be active partners in the scientific investigations they perform. Experiments should be designed such that the astronaut knows the science involved and can make judgment calls on how to proceed. An intellectually engaged crew member is a happy crew member.

When I reflect on my six months on Mir, I have no shortage of memories. But there is one that captures the legacy of the Shuttle-Mir program. One evening Onufriyenko, Usachev and I were floating around the table after supper. We were drinking tea, eating cookies and talking. The cosmonauts were

Learning from Disaster

The year 1997 was a bad one for Mir. On February 23 the solid-fuel oxygen generator in the Kvant 1 module began to burn uncontrollably, creating a blowtorchlike flame that reached a temperature of 900 degrees Fahrenheit (500 degrees Celsius). The Mir crew—which at that time included NASA astronaut Jerry Linenger—scrambled for oxygen masks and fire extinguishers. The distilled water foam in the extinguishers failed to douse the blaze, but it kept the walls of the space station cool until the fire burned itself out 14 minutes later. The fire caused no injuries, but the crew had to wear surgical masks for several days afterward because of the fear of smoke contamination.

An even worse mishap occurred four months later. On June 25, while the crew was testing a new system for docking the Progress supply ship to Mir, the space station's commander misjudged the approach of the vessel. The Progress crashed into an array of solar-energy panels (*right*) and punctured the hull of the Spektr module. As air leaked out of the space station, astronaut Michael Foale—who had replaced Linenger the month before—raced to the Soyuz escape vessel and started preparing it for departure. Luckily, the puncture was small, and the crew was able to seal off Spektr and repressurize the rest of the station.

In the wake of the accidents, leaders in the U.S. Congress called for a thorough review of the Shuttle-Mir program. NASA commissioned two panels to study the safety of sending astronauts to Mir and decided in September to go ahead with the final two missions. But critics of the U.S. space agency have raised concerns about the next stage of the cooperative program, the construction of the International Space Station, which is scheduled to begin this summer. Many of the same types of equipment used on Mir—including the solid-fuel oxygen generators—will also be used in the Russian modules of the new space station.

SOLAR ARRAY damaged by the Progress collision

NASA/RUSSIAN SPACE AGENCY

To reduce the risks of another fire or depressurization, NASA has worked with the Russians to determine the causes of the accidents. Unfortunately, the origin of the fire remains a mystery. The solid-fuel generator burns lithium perchlorate to produce oxygen. Mir crews have frequently used these devices as a backup to the station's primary oxygen generator. Some NASA engineers believe the solid-fuel generator might have been damaged by multiple misfirings of its ignition system. Russian and American officials are now considering whether to modify that system and add more shielding to the device.

The causes of the Progress collision are clearer. The test of the new docking system was poorly planned. Vasili Tsibliyev, Mir's commander at the time, was supposed to gauge the distance between Progress and the space station by monitoring a video image of Mir taken by a camera mounted on the supply vessel. But the angle of approach was such that the earth appeared behind Mir on the video monitor, making it more difficult to see the space station. Also, Mir's radar had been turned off for the test, which occurred while the space station was out of communications contact with the ground.

Officials at NASA were not aware of these deficiencies beforehand. To prevent a similar collision on the International Space Station—which will have three docking ports for Russian and European spacecraft and two for the space shuttle—U.S. officials insisted on having more scrutiny of Russian docking tests. Says Frank Culbertson, director of the Shuttle-Mir program: "Our safety efforts have definitely improved. We'll have a more experienced crew and ground team to deal with future incidents."

—Mark Alpert, staff writer

very curious about my childhood in Texas and Oklahoma. Onufriyenko talked about the Ukrainian village where he grew up, and Usachev reminisced about his own Russian village. After a while we realized we had all grown up with the same fear: an atomic war between our two countries.

I had spent my grade school years living in terror of the Soviet Union. We practiced bomb drills in our classes, all of us crouching under our desks, never questioning why. Similarly, Onufriyenko and Usachev had grown up with the knowledge that U.S. bombers or missiles might zero in on their villages. After talking about our childhoods some more, we marveled at what an unlikely scenario had unfolded. Here we were, from countries that were sworn enemies a few years earlier, living together on a space station in harmony and peace. And, incidentally, having a great time. SA

The Author

SHANNON W. LUCID is an astronaut at the National Aeronautics and Space Administration Johnson Space Center in Houston, Tex. She has participated in five spaceflights, including her mission on Mir, logging a total of 223 days in orbit. She is currently the astronaut representative to the Shuttle-Mir program. She is still an active-duty astronaut and hopes to be assigned to another NASA spaceflight.

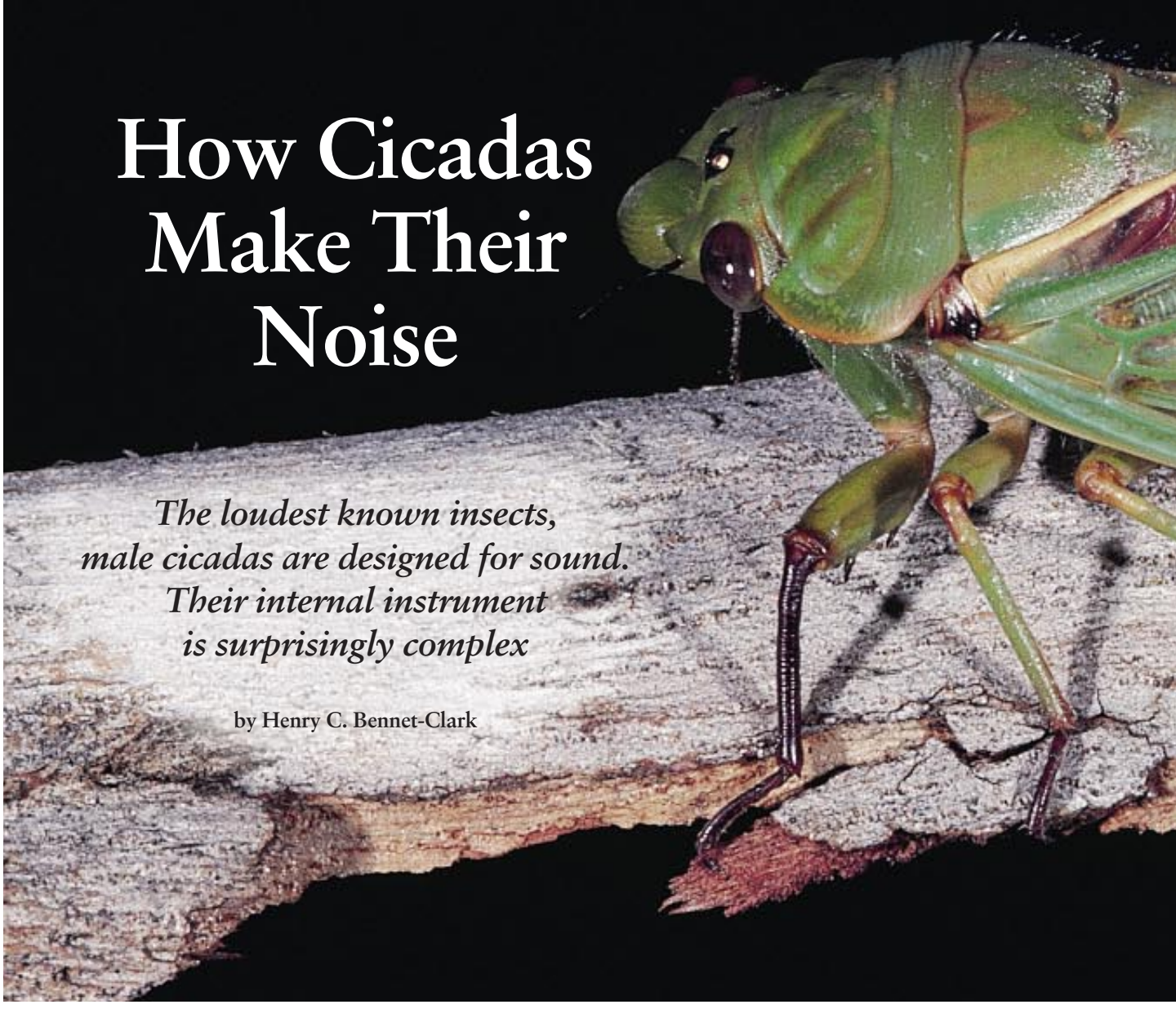
Further Reading

DIARY OF A COSMONAUT: 211 DAYS IN SPACE. Valentin Lebedev. PhytoResource Research, 1988; and Bantam Books, 1990.

Information on the Shuttle-Mir program is available at http://shuttle-mir.nasa.gov on the World Wide Web.

Information on the International Space Station is available at http://station.nasa.gov on the World Wide Web.

How Cicadas Make Their Noise



The loudest known insects, male cicadas are designed for sound. Their internal instrument is surprisingly complex

by Henry C. Bennet-Clark

In many places the world over, the early dusk of late spring can be quite a noisy time. Every year some of the several thousand species of cicada emerge from underground, and the males begin to sing their raucous, almost deafening, song. Among these virtuosos of the insect realm, the males of one species of Australian cicada (*Cyclochila australasiae*) are distinguished for having the loudest insect call measured so far.

Transmitting at 100 decibels within a one-meter range at a frequency of 4.3 kilohertz, this cicada has a cry whose volume and intensity resemble that of a personal alarm going off. Except, of course, that the male may not be singing alone but rather with a chorus often of tens or even hundreds of others. So the effect is similar to that of a roomful of simultaneously ringing alarms.

How these relatively small creatures—they are only 60 millimeters (2.3 inches) long overall—are capable of making such a rasping ruckus has remained a mystery until recently. Despite the attention these insects receive when they come out—in the U.S. the appearance of the periodic cicadas is a media event—their bioacoustics have not been well studied. In 1954 J.W.S. Pringle published a seminal paper on cicada sound production, which was followed by a period of relative silence on this front. During the past seven years, David Young of the University of Melbourne and I have focused our investigations on the acoustics and mechanics of these insects. With the use of tiny probe microphones that I developed, we have been able to describe aspects of the marvelous sound system of *C. australasiae*.

A Curious Volume

Cicadas are plant-sucking bugs, related most closely to aphids and leafhoppers. They are large—and noisy—for insects that feast on sap: the smallest are 10 millimeters (just under half an inch) long, the largest 100 millimeters (four inches). Female cicadas lay their eggs in the stems of plants or in trees. Later, the newly hatched young drop to the ground and burrow in search of plant roots to tap. They remain underground throughout their larval life, which can last many years; the 13- and 17-year life cycles of the American periodic cicadas are among the longest. When the larvae emerge, they molt into winged adults and then live for only a few weeks. During this time, the males sing their hearts out to at-

C. RENTZ *Bruce Coleman Inc.*

AUSTRALIAN CICADA (*Cyclochila australasiae*) is one of the more than 100 species found "down under." The male has the loudest call of any insect so far recorded.

tract mates. (Cicadas can also produce other calls: a protest song and even a quiet courting rasp.)

Why the insects' calling song is so loud is not entirely understood. One possibility is that the sound may saturate the ears of predators, making it hard to locate the cicada precisely. Or it may be that larger cicadas have a larger home range and need to call louder to attract a mate. We do know that the females are not hard of hearing: they hear just fine with thresholds of 30 to 40 decibels. And it is quite likely that female cicadas can discriminate between males on the basis of song quantity and quality, just as female crickets do.

Surround Sound

The key to understanding how the males make their powerful noise lies in isolating the different aspects of the sound-production mechanism. The primary equipment is a pair of domed tymbals, or drumlike structures, that sit on either side of the abdomen. Each of these elastic, resonant organs has a row of four convex ribs that runs longitudinally up and down its surface. These ribs are flexibly connected to one another as well as to a large oval plate at the rear of the tymbal.

Attached to each oval tymbal plate, in turn, is a large, fast muscle. The contraction of these muscles distorts the domed tymbals, creating a pulse of sound. Because the two tymbal muscles contract alternately at 120 hertz apiece, the song produced by the tymbals has a 240-hertz modulation. (A hertz is a unit of frequency that measures one cycle per second; accordingly, one kilohertz indicates a frequency of 1,000 cycles per second.)

Because they contract so rapidly, the muscles produce enough energy in the first three milliseconds of each contraction to cause two or three of the curved ribs on each tymbal to buckle in sequence. This motion causes two or three stepwise inward movements of the oval plate. The stored elastic energy released during these stepwise movements produces a brief, loud click as the individual ribs buckle. This series of clicks merges to form a train of vibrations at the cicada's song frequency of 4.3 kilohertz.

A click from the rib of a tymbal produces high sound pres-

sures—as great as 158 decibels—within the cicada's abdomen. (This pressure is roughly equivalent to that generated by a grenade exploding one meter away.) In most species of cicada, this region is filled largely by an air sac, and in this Australian cicada the air sac is 1.8 milliliters in volume, occupying 70 percent of the insect's abdomen. The abdomen contains a pair of thin eardrums that extend the full width of the ventral surface and serve as acoustic windows, connecting the air sac to the outside world.

The high-pressure sound pulses produced by the tymbals and ribs create a sympathetic resonance in the abdominal air sac. Because the eardrums are larger than the tymbals, they serve as an effective means of radiating the sound beyond the body—causing it to be about 20 times louder than if the sound were to emanate simply from the tiny tymbals. Thus, the insect manages to broadcast over a far greater range.

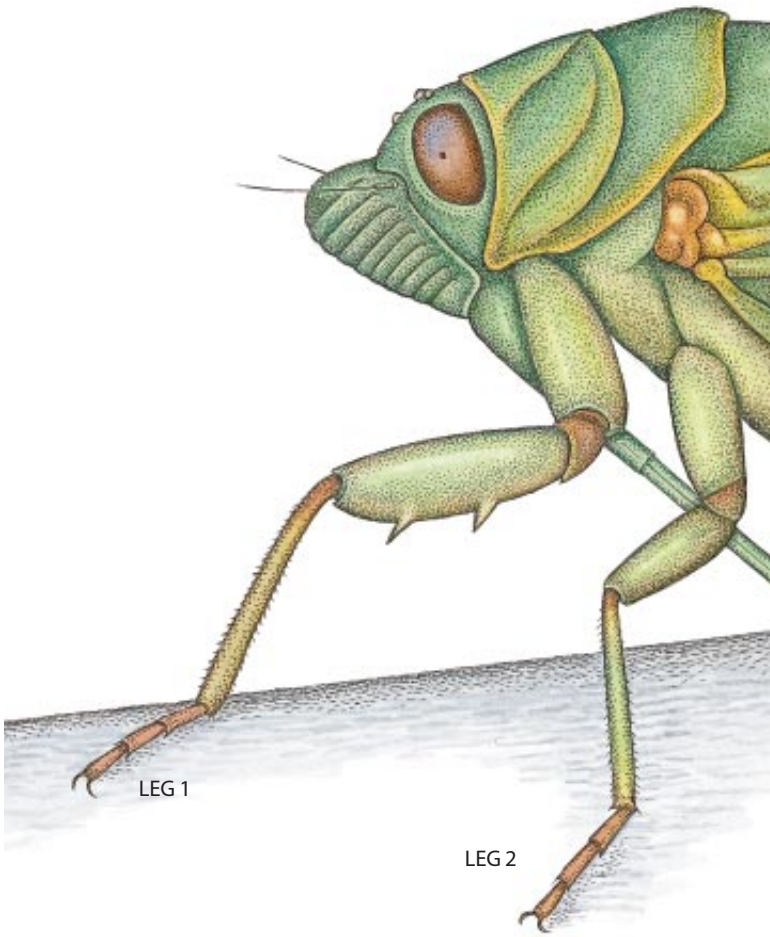
Strangely, this 158-decibel sound does not blast the male's ears to bits. Although sound is radiated via the eardrums, the sensory part of the male ear is in a separate capsule, connected to the eardrum by way of a small canal. This separation may protect the males, preventing them from going deaf—but we do not know for sure.

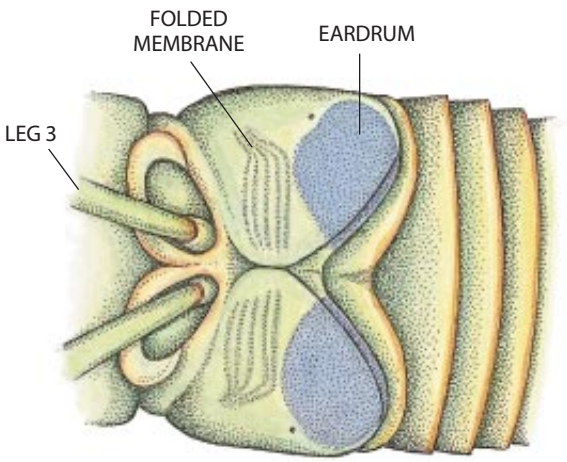
Orchestrating the Song

The abdomen and eardrums do more, though, than simply throw the sound into the world. They serve to maintain the quality of the all-important mating song. By extending the abdomen and by opening the opercula that cover the eardrums, the cicada can tune the abdominal resonator to the 4.3-kilohertz frequency of the sound pulses produced by the tymbals. The abdominal structures can therefore increase the purity or the loudness of the song. They act as what is called in musical terminology a resonant acoustic load to the tymbals—that is, they correct and balance the sound system.

A similar mechanism seems to occur in many other species of cicada, and related sound-production systems have been described in other loud insects as well. Yet to the best of our knowledge, the males of this Australian cicada species remain the noisiest insects around—as the residents of Melbourne and Sydney know only too well. SA

The Anatomy of a Cicada's Song





UNDERSIDE VIEW

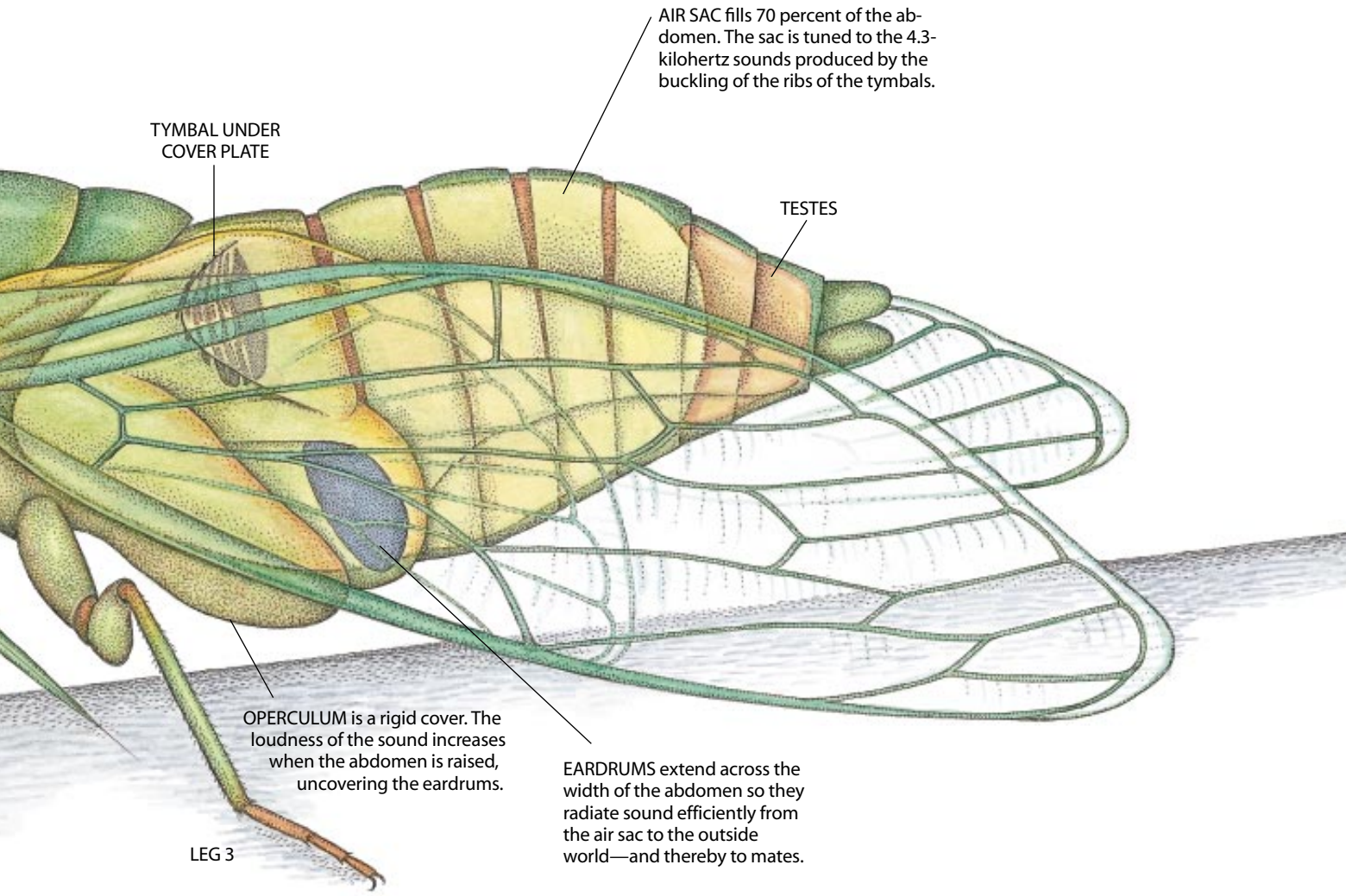
The Author

HENRY C. BENNET-CLARK is Reader in invertebrate zoology at the University of Oxford. He studies the bioacoustics and biomechanics of insects, looking, for example, at how insects jump and at the mechanical properties of insect skeletal materials. Bennet-Clark's work on fruit flies led him to develop a microphone able to record the tiny creatures' equally tiny love songs. For the past seven years, he has collaborated with David Young, a Reader in the department of zoology at the University of Melbourne, who has worked on the biology of cicadas for 20 years.

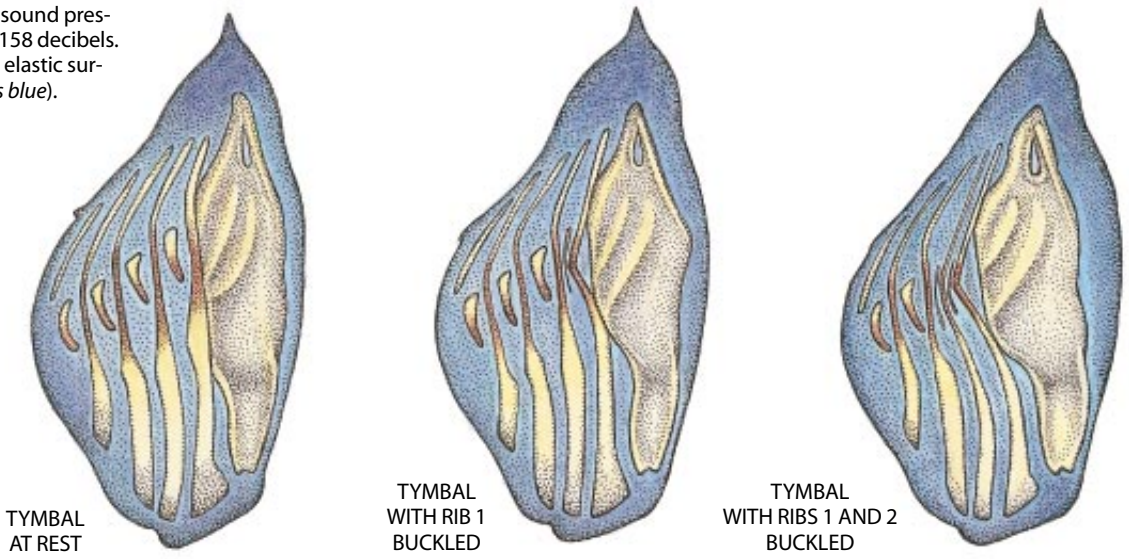
Further Reading

A Model of the Mechanism of Sound Production in Cicadas. H. C. Bennet-Clark and D. Young in *Journal of Experimental Biology,* Vol. 173, pages 123–153; 1992.

The Role of the Tymbal in Cicada Sound Production. D. Young and H. C. Bennet-Clark in *Journal of Experimental Biology,* Vol. 198, pages 1001–1019; 1995.



RIBS are buckled in sequence by the strong tymbal muscle. Each buckling creates sound pressures as great as 158 decibels. Tymbal ribs have elastic surrounds (*shown as blue*).



PATRICIA J. WYNNE

The Genetics of Cognitive Abilities and Disabilities

Investigations of specific cognitive skills can help clarify how genes shape the components of intellect

by Robert Plomin and John C. DeFries

People differ greatly in all aspects of what is casually known as intelligence. The differences are apparent not only in school, from kindergarten to college, but also in the most ordinary circumstances: in the words people use and comprehend, in their differing abilities to read a map or follow directions, or in their capacities for remembering telephone numbers or figuring change. The variations in these specific skills are so common that they are often taken for granted. Yet what makes people so different?

It would be reasonable to think that the environment is the source of differences in cognitive skills—that we are what we learn. It is clear, for example, that human beings are not born with a full vocabulary; they have to learn words. Hence, learning must be the mechanism by which differences in vocabulary arise among individuals. And differences in experience—say, in the extent to which parents model and encourage vocabulary skills or in the quality of language training provided by schools—must be responsible for individual differences in learning.

Earlier in this century psychology was in fact dominated by environmental explanations for variance in cognitive abilities. More recently, however, most psychologists have begun to embrace a more balanced view: one in which nature and nurture interact in cognitive development. During the past few decades, studies in genetics have pointed to a substantial role for heredity in molding the components of intellect, and researchers have even begun to track down the genes involved in cognitive function. These findings do not refute the notion that environmental factors shape the learning process. Instead they suggest that differences in people's genes affect how easily they learn.

Just how much do genes and envi-

ronment matter for specific cognitive abilities such as vocabulary? That is the question we have set out to answer. Our tool of study is quantitative genetics, a statistical approach that explores the causes of variations in traits among individuals. Studies comparing the performance of twins and adopted children on certain tests of cognitive skills, for example, can assess the relative contributions of nature and nurture.

In reviewing several decades of such studies and conducting our own, we have begun to clarify the relations among specialized aspects of intellect, such as verbal and spatial reasoning, as well as the relations between normal cognitive function and disabilities, such as dyslexia. With the help of molecular genetics, we and other investigators have also begun to identify the genes that affect these specific abilities and disabilities. Eventually, we believe, knowledge of these genes will help reveal the biochemical mechanisms involved in human intelligence. And with the insight gained from genetics, researchers may someday develop environmental interventions that will lessen or prevent the effects of cognitive disorders.

Some people find the idea of a genetic role in intelligence alarming or, at the very least, confusing. It is important to understand from the outset, then, what exactly geneticists mean when they talk about genetic influence. The term typically used is "heritability": a statistical measure of the genetic contribution to differences among individuals.

Verbal and Spatial Abilities

Heritability tells us what proportion of individual differences in a population—known as variance—can be ascribed to genes. If we say, for example, that a trait is 50 percent heritable, we are in effect saying that half of the variance in that trait is linked to heredity. Heritability, then, is a way of explaining what makes people different, not what constitutes a given individual's intelligence. In general, however, if heritability for a trait is high, the influence of genes on the trait in individuals would be strong as well.

Attempts to estimate the heritability of specific cognitive abilities began with family studies. Analyses of similarities between parents and their children and between siblings have shown that cognitive abilities run in families. Results of the largest family study done on specific cognitive abilities, which was conducted in Hawaii in the 1970s, helped to quantify this resemblance.

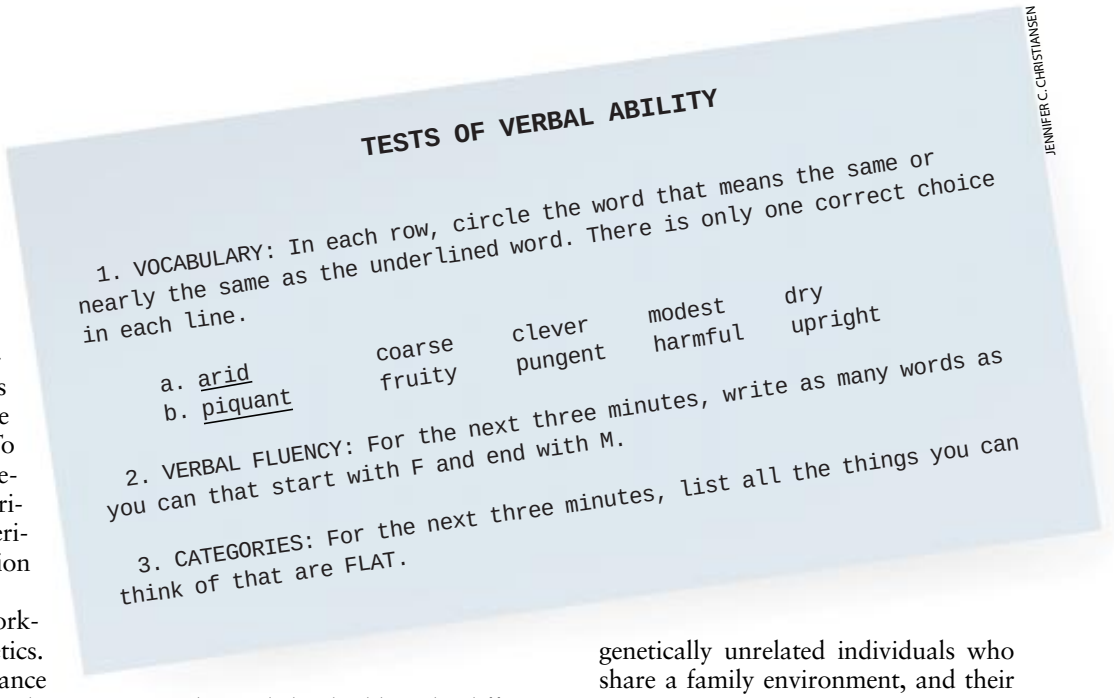
The Hawaii Family Study of Cognition was a collaborative project between researchers at the University of Colorado at Boulder and the University of Hawaii and involved more than 1,000 families and sibling pairs. The study determined correlations (a statistical measure of resemblance) between relatives on tests of verbal and spatial ability. A correlation of 1.0 would mean that the scores of family members were identical; a correlation of zero would indicate that the scores were no more similar than those of two people picked at random. Because children on average share half their genes with each parent and with siblings, the highest correlation in test scores that could be expected on genetic grounds alone would be 0.5.

The Hawaii study showed that fami-

TWINS ARE COMMON RESEARCH SUBJECTS in studies of specific cognitive abilities. The identical (*opposite page*) and fraternal (*below*) pairs depicted here are participants in the authors' research. They are performing a task in a test of spatial ability, trying to reconstruct a block model with their own toy building blocks. On such tests, which are given to each child individually, the scores of identical twins (who have all the same genes) are more similar than the scores of fraternal twins (who share about half their genes)—a sign that genetic inheritance exerts an influence on spatial ability.

DAVID KAMPFNER *Gamma Liaison*

ly members are in fact more alike than unrelated individuals on measures of specific cognitive skills. The actual correlations for both verbal and spatial tests were, on average, about 0.25. These correlations alone, however, do not disclose whether cognitive abilities run in families because of genetics or because of environmental effects. To explore this distinction, geneticists rely on two "experiments": twinning (an experiment of nature) and adoption (a social experiment).



JENNIFER C. CHRISTIANSEN

Twin studies are the workhorse of behavioral genetics. They compare the resemblance of identical twins, who have the same genetic makeup, with the resemblance of fraternal twins, who share only about half their genes. If cognitive abilities are influenced by genes, identical twins ought to be more alike than fraternal twins on tests of cognitive skills. From correlations found in these kinds of studies, investigators can estimate the extent to which genes account for variances in the general population. Indeed, a rough estimate of heritability can be made by doubling the difference between identical-twin and fraternal-twin correlations.

Adoption provides the most direct way to disentangle nature and nurture in family resemblance, by creating pairs of genetically related individuals who do not share a common family environment. Correlations among these pairs enable investigators to estimate the contribution of genetics to family resemblance. Adoption also produces pairs of genetically unrelated individuals who share a family environment, and their correlations make it possible to estimate the contribution of shared environment to resemblance.

Twin studies of specific cognitive abilities over three decades and in four countries have yielded remarkably consistent results [*see illustration on page 66*]. Correlations for identical twins greatly exceed those for fraternal twins on tests of both verbal and spatial abilities in children, adolescents and adults. Results of the first twin study in the elder-

How Do Cognitive Abilities Relate to General Intelligence?

by Karen Wright

Since the dawn of psychology, experts have disagreed about the fundamental nature of intelligence. Some have claimed that intelligence is an inherent faculty prescribed by heredity, whereas others have emphasized the effects of education and upbringing. Some have portrayed intelligence as a global quality that permeates all facets of cognition; others believe the intellect consists of discrete, specialized abilities—such as artistic talent or a flair for mathematics—that share no common principle.

In the past few decades, genetic studies have convinced most psychologists that heredity exerts considerable influence on intelligence. In fact, research suggests that as much as half of the variation in intelligence among individuals may be attributed to genetic factors.

And most psychologists have also come to accept a global conceptualization of intelligence. Termed general cognitive ability, or "g," this global quality is reflected in the apparent overlap among specific cognitive skills. As Robert Plomin and John C. DeFries point out, people who do well on tests of one type of cognitive skill also tend to do well on tests of other cognitive abilities. Indeed, this intercorrelation has provided the rationale for IQ (intelligence quotient) tests, which yield a single score from combined assessments of specific cognitive skills.

Because specific and general cognitive abilities are related in this manner, it is not surprising that many of the findings regarding specific abilities echo what is already known about general ability. The heritabilities found in studies of specific cognitive abilities, for example, are comparable with the heritability determined for g. The developmental trend described by the authors—in which genetic influence on specific cognitive abilities seems to increase throughout childhood, reaching adult levels by the mid-teens—is also familiar to researchers of general cognitive ability.

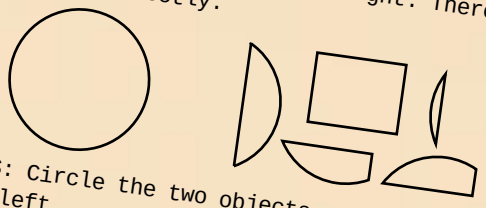
And because measures of g are derived from intercorrelations of verbal and spatial abilities, a gene that is linked with both those traits is almost guaranteed to have some role in general cognitive ability as well—and vice versa. This month in the journal *Psychological Science,* Plomin and various collaborators report the discovery of the first gene associated with general cognitive ability. Although the finding should further understanding of the nature of cognition, it is also likely to reignite debate. Indeed, intelligence research may be one realm where understanding does little to quell disagreement.

KAREN WRIGHT is a freelance writer living in New Hampshire.

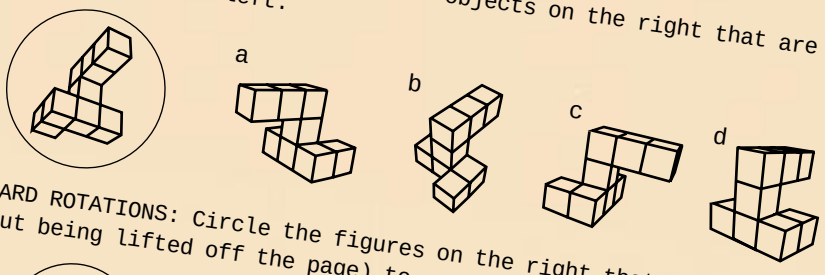
TESTS OF SPECIFIC ABILITIES administered to adolescents and adults include tasks resembling the ones listed here. The tests gauge each cognitive ability in several ways, and multiple tests are combined to provide a reliable measure of each skill. (Answers appear on page 69.)

TESTS OF SPATIAL ABILITY

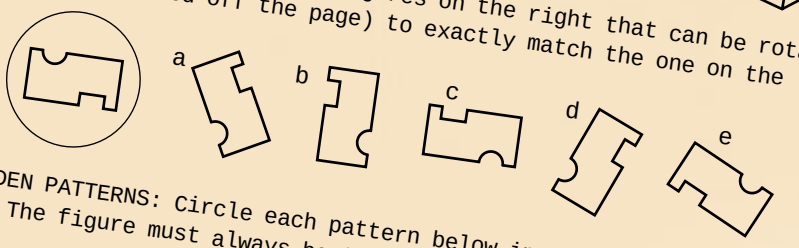
1. IMAGINARY CUTTING: Draw a line or lines showing where the figure on the left should be cut to form the pieces on the right. There may be more than one way to draw the lines correctly.



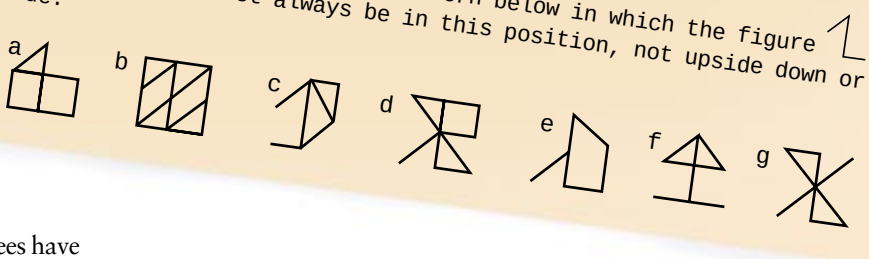
2. MENTAL ROTATIONS: Circle the two objects on the right that are the same as the object on the left.



3. CARD ROTATIONS: Circle the figures on the right that can be rotated (without being lifted off the page) to exactly match the one on the left.



4. HIDDEN PATTERNS: Circle each pattern below in which the figure appears. The figure must always be in this position, not upside down or on its side.



JENNIFER C. CHRISTIANSEN

ly—reported last year by Gerald E. McClearn and his colleagues at Pennsylvania State University and by Stig Berg and his associates at the Institute for Gerontology in Jönköping, Sweden—show that the resemblances between identical and fraternal twins persist even into old age. Although gerontologists have assumed that genetic differences become less important as experiences accumulate over a lifetime, research on cognitive abilities has so far demonstrated otherwise. Calculations based on the combined findings in these studies imply that in the general population, genetics accounts for about 60 percent of the variance in verbal ability and about 50 percent of the variance in spatial ability.

Investigations involving adoptees have yielded similar results. Two recent studies of twins reared apart—one by Thomas J. Bouchard, Jr., Matthew McGue and their colleagues at the University of Minnesota, the other an international collaboration headed by Nancy L. Pedersen at the Karolinska Institute in Stockholm—have implied heritabilities of about 50 percent for both verbal and spatial abilities.

In our own Colorado Adoption Project, which we launched in 1975, we have used the power of adoption studies to further characterize the roles of genes and environment, to assess developmental trends in cognitive abilities and to explore the extent to which specific cognitive skills are related to one another. The ongoing project compares the correlations between more than 200 adopted children and their birth and adoptive parents with the correlations for a control group of children raised by their biological parents [*see illustration on page 67*].

These data provide some surprising insights. By middle childhood, for example, birth mothers and their children who were adopted by others are just as similar as control parents and their children on measures of both verbal and spatial ability. In contrast, the scores of adopted children do not resemble those of their adoptive parents at all. These results join a growing body of evidence suggesting that the common family environment generally does not contribute to similarities in family members. Rather family resemblance on such measures seems to be controlled almost entirely by genetics, and environmental factors often end up making family members different, not the same.

The Colorado data also reveal an interesting developmental trend. It appears that genetic influence increases during childhood, so that by the mid-teens, heritability reaches a level comparable with that seen in adults. In correlations of verbal ability, for example, resemblance between birth parents and their children who were adopted by others increases from about 0.1 at age three to about 0.3 at age 16. A similar pattern is evident in tests of spatial ability. Some genetically driven transformation in cognitive function seems to take place in the early school years, around age seven. The results indicate that by the time people reach age 16, genetic factors account for 50 percent of the variance for verbal ability and 40 percent for spatial ability—numbers not unlike those derived from twin studies of specific cognitive abilities.

The Colorado Adoption Project and other investigations have also helped

clarify the differences and similarities among cognitive abilities. Current cognitive neuroscience assumes a modular model of intelligence, in which different cognitive processes are isolated anatomically in discrete modules in the brain. The modular model implies that specific cognitive abilities are also genetically distinct—that genetic effects on verbal ability, say, should not overlap substantially with genetic effects on spatial ability.

Psychologists, however, have long recognized that most specialized cognitive skills, including verbal and spatial abilities, intercorrelate moderately. That is, people who perform well on one type of test also tend to do well on other types. Correlations between verbal and spatial abilities, for example, are usually about 0.5. Such intercorrelation implies a potential genetic link.

From Abilities to Achievement

Genetic studies of specific cognitive abilities also fail to support the modular model. Instead it seems that genes are responsible for most of the overlap between cognitive skills. Analysis of the Colorado project data, for example, indicates that genetics governs 70 percent of the correlation between verbal and spatial ability. Similar results have been found in twin studies in childhood, young adulthood and middle age. Thus, there is a good chance that when genes associated with a particular cognitive ability are identified, the same genes will be associated with other cognitive abilities.

Research into school achievement has hinted that the genes associated with cognitive abilities may also be relevant to academic performance. Studies of more than 2,000 pairs of high school–age twins were done in the 1970s by John C. Loehlin of the University of Texas at Austin and Robert C. Nichols, then at the National Merit Scholarship Corporation in Evanston, Ill. In these studies the scores of identical twins were consistently and substantially more similar than those of fraternal twins on all four domains of the National Merit Scholarship Qualifying Test: English usage, mathematics, social studies and natural sciences. These results suggest that genetic factors account for about 40 percent of the variation on such achievement tests.

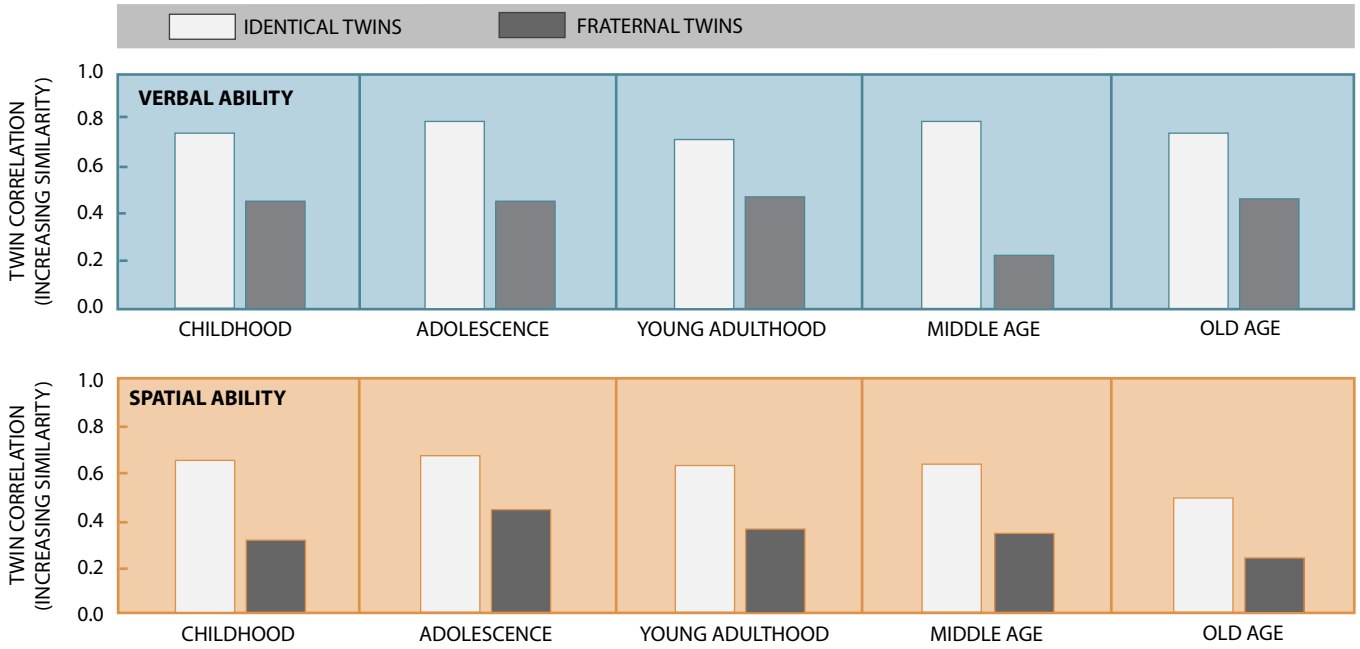
Genetic influence on school achievement has also been found in twin studies of elementary school–age children as well as in our work with the Colorado Adoption Project. It appears that genes may have almost as much effect on school achievement as they do on cognitive abilities. These results are surprising in and of themselves, as educators have long believed that achievement is more a product of effort than of ability. Even more interesting, then, is the finding from twin studies and our adoption project that genetic effects overlap between different categories of achievement and that these overlapping genes are probably the very same genetic fac-

What Heritability Means

The implications of heritability data are commonly misunderstood. As the main text indicates, heritability is a statistical measure, expressed as a percentage, describing the extent to which genetic factors contribute to variations on a given trait among the members of a population.

The fact that genes influence a trait does not mean, however, that "biology is destiny." Indeed, genetics research has helped confirm the significance of environmental factors, which generally account for as much variance in human behavior as genes do. If intelligence is 50 percent heritable, then environmental factors must be just as important as genes in generating differences among people.

TWIN STUDIES have examined correlations in verbal (*top*) and in spatial (*bottom*) skills of identical twins and of fraternal twins. When the results of the separate studies are put side by side, they demonstrate a substantial genetic influence on specific cognitive abilities from childhood to old age; for all age groups, the scores of identical twins are more alike than those of fraternal twins. These data seem to counter the long-standing notion that the influence of genes wanes with time.



Moreover, even when genetic factors have an especially powerful effect, as in some kinds of mental retardation, environmental interventions can often fully or partly overcome the genetic "determinants." For example, the devastating effects of phenylketonuria, a genetic disease that can cause mental retardation, can often be nullified by dietary intervention.

Finally, the degree of heritability for a given trait is not set in stone. The relative influence of genes and environment can change. If, for instance, environmental factors were made almost identical for all the members of a hypothetical population, any differences in cognitive ability in that population would then have to be attributed to genetics, and heritability would be closer to 100 percent than to 50 percent. Heritability describes what is, rather than what can (or should) be. *—R.P. and J.C.D.*

tors that can influence cognitive abilities.

This evidence supports a decidedly nonmodular view of intelligence as a pervasive or global quality of the mind and underscores the relevance of cognitive abilities in real-world performance. It also implies that genes for cognitive abilities are likely to be genes involved in school achievement, and vice versa.

Given the evidence for genetic influence on cognitive abilities and achievement, one might suppose that cognitive disabilities and poor academic achievement must also show genetic influence. But even if genes are involved in cognitive disorders, they may not be the same genes that influence normal cognitive function. The example of mental retardation illustrates this point. Mild mental retardation runs in families, but severe retardation does not. Instead severe mental retardation is caused by genetic and environmental factors—novel mutations, birth complications and head injuries, to name a few—that do not come into play in the normal range of intelligence.

Researchers need to assess, rather than assume, genetic links between the normal and the abnormal, between the traits that are part of a continuum and true disorders of human cognition. Yet genetic studies of verbal and spatial disabilities have been few and far between.

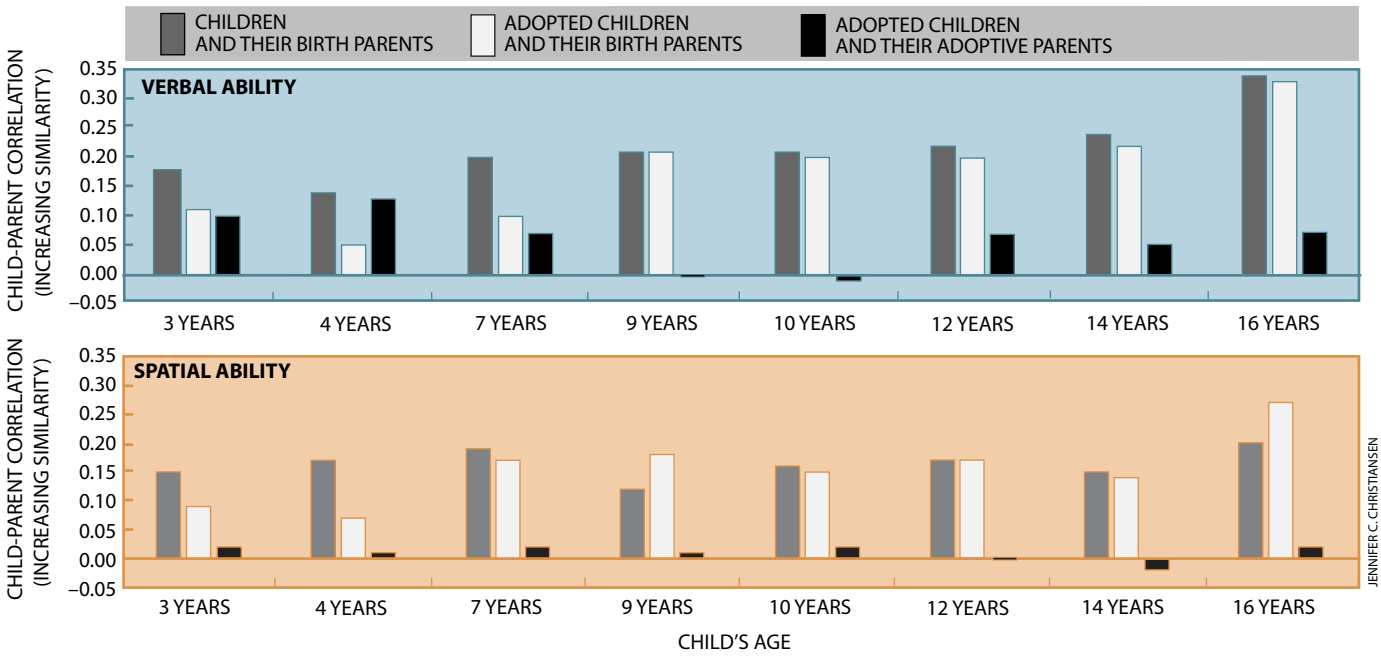
Genetics and Disability

Most such research has focused on reading disability, which afflicts 80 percent of children diagnosed with a learning disorder. Children with reading disability, also known as dyslexia, read slowly, show poor comprehension and have trouble reading aloud [see "Dyslexia," by Sally E. Shaywitz, Scientific American, November 1996]. Studies by one of us (DeFries) have shown that reading disability runs in families and that genetic factors do indeed contribute to the resemblance among family members. The identical twin of a person diagnosed with reading disability, for example, has a 68 percent risk of being similarly diagnosed, whereas a fraternal twin has only a 38 percent chance.

Is this genetic effect related in any way to the genes associated with normal variation in reading ability? That question presents some methodological challenges. The concept of a cognitive disorder is inherently problematic, because it treats disability qualitatively—you either have it or you don't—rather than describing the degree of disability in a quantitative fashion. This focus creates an analytical gap between disorders and traits that are dimensional (varying along a continuum), which are by definition quantitative.

During the past decade, a new genetic technique has been developed that bridges the gap between dimensions and disorders by collecting quantitative information about the relatives of subjects diagnosed qualitatively with a dis-

COLORADO ADOPTION PROJECT, which followed subjects over time, finds that for both verbal (*top*) and spatial (*bottom*) abilities, adopted children come to resemble their birth parents (*white bars*) as much as children raised by their birth parents do (*gray bars*). In contrast, adopted children do not end up resembling their adoptive parents (*black bars*). The results imply that most of the family resemblance in cognitive skills is caused by genetic factors, not environment.

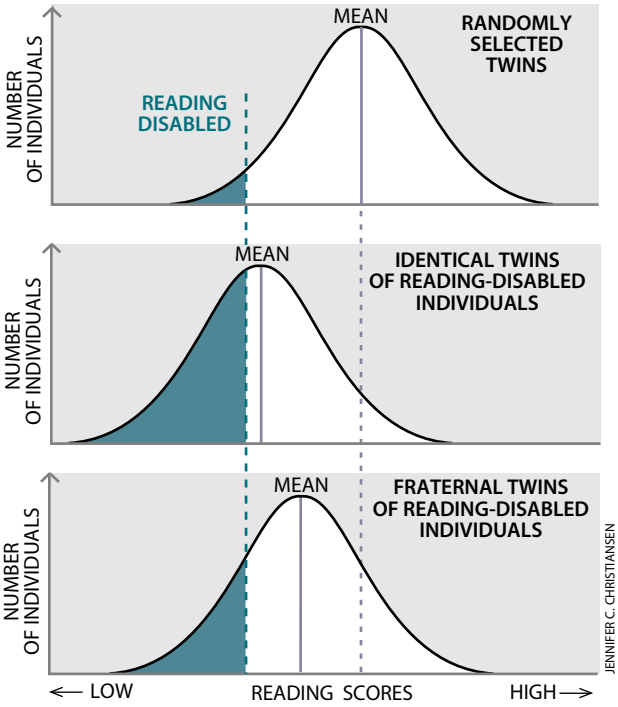


ability. The method is called DF extremes analysis, after its creators, DeFries and David W. Fulker, a colleague at the University of Colorado's Institute for Behavioral Genetics.

For reading disability, the analysis works by testing the identical and fraternal twins of reading-disabled subjects on quantitative measures of reading, rather than looking for a shared diagnosis of dyslexia [*see illustration at right*]. If reading disability is influenced by genes that also affect variation within the normal range of reading performance, then the reading scores of the identical twins of dyslexic children should be closer to those of the reading-disabled group than the scores of fraternal twins are. (A single gene can exert different effects if it occurs in more than one form in a population, so that two people may inherit somewhat different versions. The genes controlling eye color and height are examples of such variable genes.)

It turns out that, as a group, identical twins of reading-disabled subjects do perform almost as poorly as dyslexic subjects on these quantitative tests, whereas fraternal twins do much better than the reading-disabled group (though still significantly worse than the rest of the population). Hence, the genes involved in reading disability may in fact be the same as those that contribute to the quantitative dimension of reading ability measured in this study. DF extremes analysis of these data further suggests that about half the difference in reading scores between dyslexics and the general population can be explained by genetics.

For reading disability, then, there could well be a genetic link between the normal and the abnormal, even though such links may not be found universally for other disabilities. It is possible that reading disability represents the extreme end of a continuum of reading ability, rather than a distinct disorder—that dyslexia might be quantitatively rather than qualitatively different from the normal range of reading ability. All this suggests that if a gene is found for reading disability, the same gene is likely to be associated with the normal range of variation in reading ability. The definitive test will come when a specific gene is identified that is associated with either reading ability or disability. In fact, we and other investigators are already very close to finding such a gene.



READING SCORES of twins suggest a possible genetic link between normal and abnormal reading skills. In a group of randomly selected members of twin pairs (*top*), a small fraction of children were reading disabled (*blue*). Identical (*middle*) and fraternal (*bottom*) twins of the reading-disabled children scored lower than the randomly selected group, with the identical twins performing worse than the fraternal ones. Genetic factors, then, are involved in reading disability. The same genes that influence reading disability may underlie differences in normal reading ability.

The Hunt for Genes

Until now, we have confined our discussion to quantitative genetics, a discipline that measures the heritability of traits without regard to the kind and number of genes involved. For information about the genes themselves, researchers must turn to molecular genetics—and increasingly, they do. If scientists can identify the genes involved in behavior and characterize the proteins that the genes code for, new interventions for disabilities become possible.

Research in mice and fruit flies has succeeded in identifying single genes related to learning and spatial perception, and investigations of naturally occurring variations in human populations have found mutations in single genes that result in general mental retardation. These include the genes for phenylketonuria and fragile X syndrome, both causes of mental retardation. Single-gene defects that are associated with Duchenne's muscular dystrophy, Lesch-Nyhan syndrome, neurofibromatosis type 1 and Williams syndrome may also be linked to the specific cognitive disabilities seen in these disorders [see "Williams Syndrome and the Brain," by Howard M. Lenhoff, Paul P. Wang, Frank Greenberg and Ursula Bellugi; Scientific American, December 1997].

In fact, more than 100 single-gene mutations are known to impair cognitive development. Normal cognitive functioning, on the other hand, is almost certainly orchestrated by many subtly acting genes working together, rather than by single genes operating in isolation. These collaborative genes are thought to affect cognition in a probabilistic rather than a deterministic manner and are called quantitative trait loci, or QTLs. The name, which applies to genes involved in a complex dimension such as cognition, emphasizes the quantitative nature of certain physical and behavioral traits. QTLs have already been identified for diseases such as diabetes, obesity and hypertension as well as for behavioral problems involving drug sensitivity and dependence.

But finding QTLs is much more difficult than identifying the single-gene mutations responsible for some cognitive disorders. Fulker addressed this problem by developing a method, similar to DF extremes analysis, in which certain known variations in DNA are correlated with sibling differences in quantitative traits. Because genetic effects are easier to detect at the extremes of a dimension, the method works best when at least one member of each sibling pair is known to be extreme for a trait. Investigators affiliated with the Colorado Learning Disabilities Research Center at the University of Colorado first used this technique, called QTL linkage, to try to locate a QTL for reading disability—and succeeded. The discovery was reported in 1994 by collaborators at Boulder, the University of Denver and Boys Town National Research Hospital in Omaha.

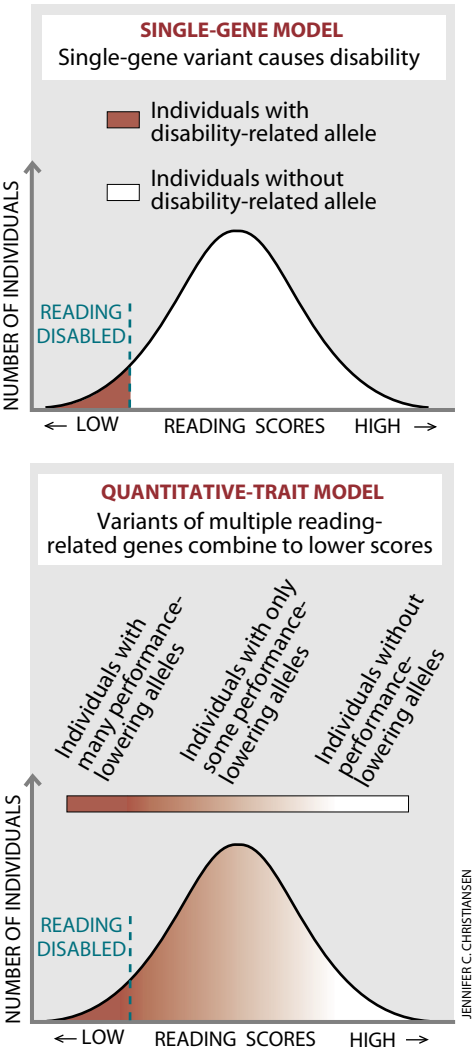
Like many techniques in molecular genetics, QTL linkage works by identifying differences in DNA markers: stretches of DNA that are known to occupy particular sites on chromosomes and that can vary somewhat from person to person. The different versions of a marker, like the different versions of a gene, are called alleles. Because people have two copies of all chromosomes (except for the gender-determining X and Y chromosomes in males), they have two alleles for any given DNA marker. Hence, siblings can share one, two or no alleles of a marker. In other words, for each marker, siblings can either be like identical twins (sharing both alleles), like fraternal twins (sharing half their alleles) or like adoptive siblings (sharing no alleles).

The investigators who found the QTL for reading disability identified a reading-disabled member of a twin pair and then obtained reading scores for the other twin—the "co-twin." If the reading scores of the co-twins were worse when they shared alleles of a particular marker with their reading-disabled twins, then that marker was likely to lie near a QTL for reading disability in the same chromosomal region. The researchers found such a marker on the short arm of chromosome 6 in two independent samples, one of fraternal twins and one of non-twin siblings. The findings have since been replicated by others.

It is important to note that whereas these studies have helped point to the location of a gene (or genes) implicated in reading disability, the gene (or genes) has not yet been characterized. This distinction gives a sense of where the genetics of cognition stand today: poised on the brink of a new level of discovery. The identification of genes that influence specific cognitive abilities will revolutionize researchers' understanding of the mind. Indeed, molecular genetics will have far-ranging consequences for the study of all human behavior. Researchers will soon be able to investigate the genetic connections between different traits and between behaviors and biological mechanisms. They will be able to better track the developmental course of genetic effects and to define more precisely the interactions between genes and the environment.

The discovery of genes for disorders and disabilities will also help clinicians design more effective therapies and to identify people at risk long before the appearance of symptoms. In fact, this scenario is already being enacted with an allele called Apo-E4, which is associated with dementia and cognitive decline in the elderly. Of course, new knowledge of specific genes could turn up new problems as well: among them, prejudicial labeling and discrimination. And genetics research always raises fears that DNA markers will be used by parents prenatally to select "designer babies."

We cannot emphasize too much that genetic effects do not imply genetic determinism, nor do they constrain environmental interventions. Although some readers may find our views to be controversial, we believe the benefits of identifying genes for cognitive dimensions and disorders will far outweigh the potential abuses. SA



TWO MODELS illustrate how genetics may affect reading disability. In the classic view (*top*), a single variant, or allele, of a gene is able to cause the disorder; everyone who has that allele becomes reading disabled (*graph*). But evidence points to a different model (*bottom*), in which a single allele cannot produce the disability on its own. Instead variants of multiple genes each act subtly but can combine to lower scores and increase the risk of disability.

TEST ANSWERS VERBAL: 1a. dry; 1b. pungent SPATIAL: 1.  2. b, c; 3. a, c, d; 4. a, b, f

The Authors

ROBERT PLOMIN and JOHN C. DEFRIES have collaborated for more than 20 years. Plomin, who worked with DeFries at the University of Colorado at Boulder from 1974 to 1986, is now at the Institute of Psychiatry in London. There he is research professor of behavioral genetics and deputy director of the Social, Genetic and Developmental Psychiatry Research Center. DeFries directs the University of Colorado's Institute for Behavioral Genetics and the university's Colorado Learning Disabilities Research Center. The ongoing Colorado Adoption Project, launched by the authors in 1975, has so far produced three books and more than 100 research papers. Plomin and DeFries are also the lead authors of the textbook *Behavioral Genetics*, now in its third edition.

Further Reading

NATURE, NURTURE AND PSYCHOLOGY. Edited by Robert Plomin and Gerald E. McClearn. American Psychological Association, Washington, D.C., 1993.

GENETICS OF SPECIFIC READING DISABILITY. J. C. DeFries and Maricela Alarcón in *Mental Retardation and Developmental Disabilities Research Reviews*, Vol. 2, pages 39–47; 1996.

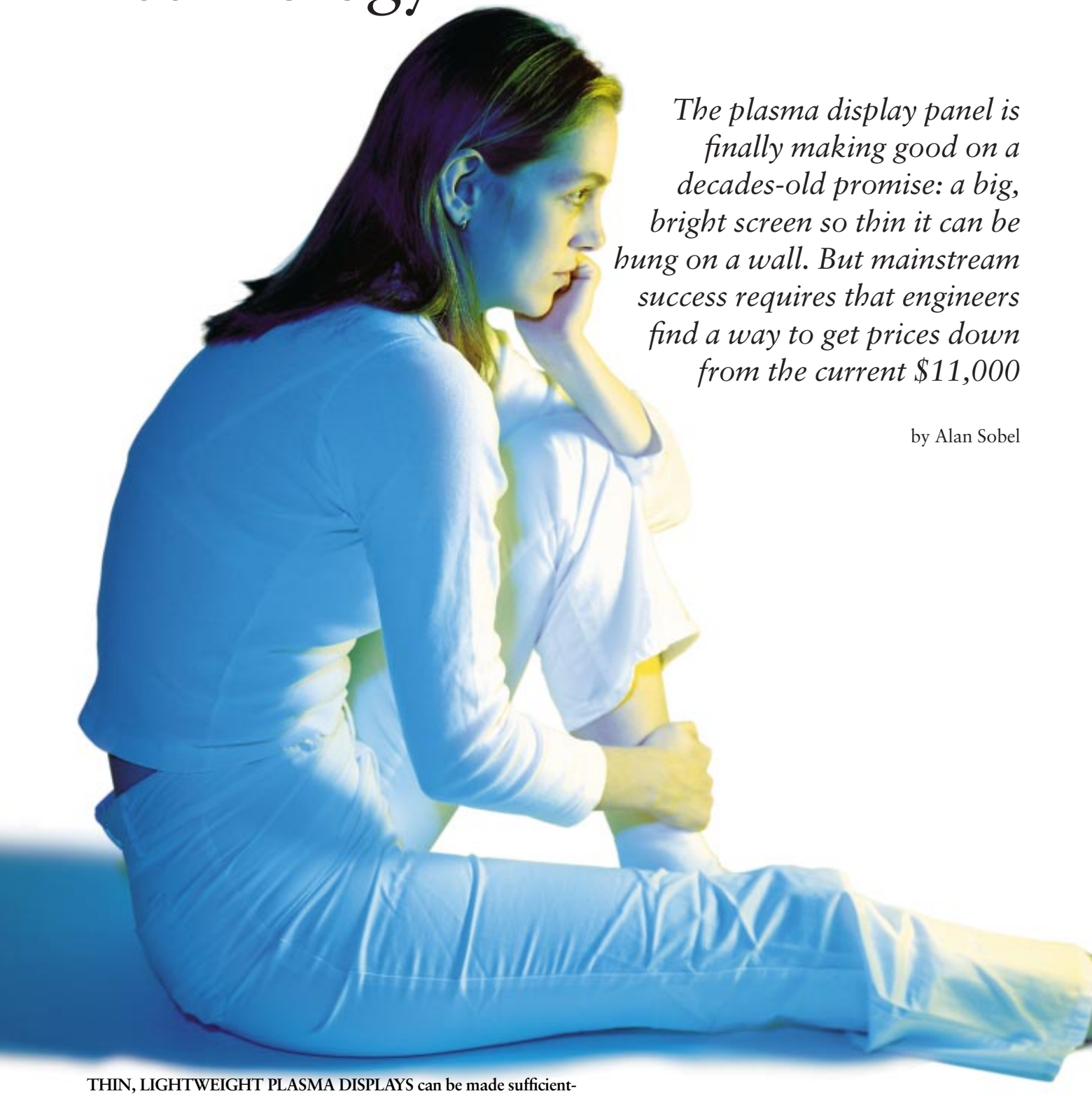
BEHAVIORAL GENETICS. Third edition. Robert Plomin, John C. DeFries, Gerald E. McClearn and Michael Rutter. W. H. Freeman, 1997.

SUSCEPTIBILITY LOCI FOR DISTINCT COMPONENTS OF DEVELOPMENTAL DYSLEXIA ON CHROMOSOMES 6 AND 15. E. L. Grigorenko, F. B. Wood, M. S. Meyer, L. A. Hart, W. C. Speed, A. Schuster and D. L. Pauls in *American Journal of Human Genetics*, Vol. 60, pages 27–39; 1997.

Television's Bright New Technology

The plasma display panel is finally making good on a decades-old promise: a big, bright screen so thin it can be hung on a wall. But mainstream success requires that engineers find a way to get prices down from the current $11,000

by Alan Sobel



THIN, LIGHTWEIGHT PLASMA DISPLAYS can be made sufficiently large to show enough of the detail available with high-definition television to absorb the viewer in the way a big-screen movie does.

The television set is probably the most successful electronic product of all time. Essentially all homes in developed countries have at least one, and in many countries there are considerably more TV sets than telephones.

This phenomenal success notwithstanding, as an appliance the television set leaves a great deal to be desired. It is boxy, heavy and somewhat delicate. Perhaps most disappointing is that it is not really commercially feasible to make a conventional, picture-tube-based set with a screen size exceeding about 1,000 millimeters (40 inches), measured diagonally. (In this article, millimeters are used for screen measurements, following the convention in engineering.) Besides being extremely expensive, such a set would weigh hundreds of kilograms and would be difficult to get through a standard, 76-centimeter-wide (30-inch-wide) residential door. Larger-screen televisions today generally use one of several projection technologies, all of which suffer to some extent from limited brightness or viewing angles in comparison with picture tubes.

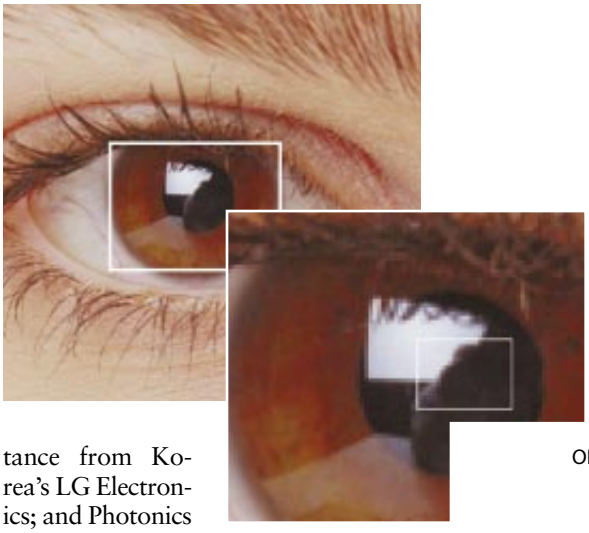
With the imminent advent of high-definition television (HDTV), this inability to produce large, bright, full-color screens is about to become unacceptable. The conventional wisdom is that HDTV—the much higher resolution television technology that is already available in Japan and will soon reach the Americas and Europe—has very little effect on the viewer unless it is displayed on a screen with a diagonal measurement of, at the very least, 1,000 millimeters (or, better yet, 1,250, 1,500 or more). Only on a very large screen can all the detail be appreciated, absorbing the viewer in the way a big-screen movie does.

To prepare for this anticipated demand for thin, large screens, high-tech concerns in Japan, the U.S. and Europe are working on dozens of different flat-panel-display technologies, most of which were invented years if not decades ago. Of this group, the only technology that is close to commercial viability is the plasma display panel, or PDP, also known as the gas-discharge display. As of this writing, engineers at NEC have reportedly demonstrated a 1,270-millimeter (50-inch) screen in the 16:9 width-to-height ratio of HDTV. Fujitsu, Mitsubishi, Toshiba and Plasmaco (a U.S.-based company now owned by Matsushita) have all demonstrated panels up to 1,056 millimeters (42 inches) in diagonal, in both the 4:3 aspect ratio of current TV and HDTV's 16:9.

CHRIS BURKE *Quesada/Burke Photography;* PLASMA SCREEN COURTESY OF QFTV, INC.

As of mid-March, approximately $11,000 would buy from Fujitsu a 1,056-millimeter monitor, which differed from a television set only in its lack of a tuner and speakers. Sales were said to be steady, if not frequent, in a few commercial markets. The New York Stock Exchange, for example, has installed on its trading floor about 1,000 Fujitsu panels to display stock prices, and these panels are beginning to appear at trade shows, where a big picture can impress an audience. Costs for the displays will surely plummet as more production equipment comes online. Indeed, prices must fall to less than $1,000 if the plasma display is to become a mass-market product.

The stakes are high: a predicted global market for perhaps hundreds of millions of screens worth many billions of dollars. Accordingly, the Japanese companies mentioned above have collectively invested more than $1 billion in new or expanded production facilities and have also invested in some U.S. firms working in this field. Other companies that have sizable plasma display efforts but that have not yet produced 1,000-millimeter screens include the Netherlands-based giant Philips; Thomson Tubes Electroniques in France, which has concentrated on smaller, high-resolution panels; and a number of Korean, Taiwanese and Japanese firms (other than the ones mentioned above). In the U.S., only three companies are developing the displays: Plasmaco; ElectroPlasma, which is getting substantial assis-



PLASMA PICTURE ELEMENT, or "pixel," consists of three tiny cells, each coated inside with a chemical phosphor (*below*). The phosphors emit light of a primary color when the gas in the cell is ionized. A cell is initially ionized by applying a voltage pulse between the address electrode and the sustain electrodes; the continuously applied voltage pulses between the sustain electrodes continue to ionize the gas, so each cell emits light until it is turned off (by another, carefully timed pulse between address and sustain electrodes). The ionized gas, known as a plasma, emits ultraviolet radiation, which stimulates the phosphors to emit the colored light (*bottom left*). (The bus electrode, adjacent to the sustain electrode, delivers the higher current level needed at the beginning of each pulse discharge.) As in an ordinary video or computer monitor, different relative brightness levels among the three primary colors in each of many thousands of pixels enable the display to create essentially any color and render any image (*left*).

tance from Korea's LG Electronics; and Photonics Systems, which is still independent.

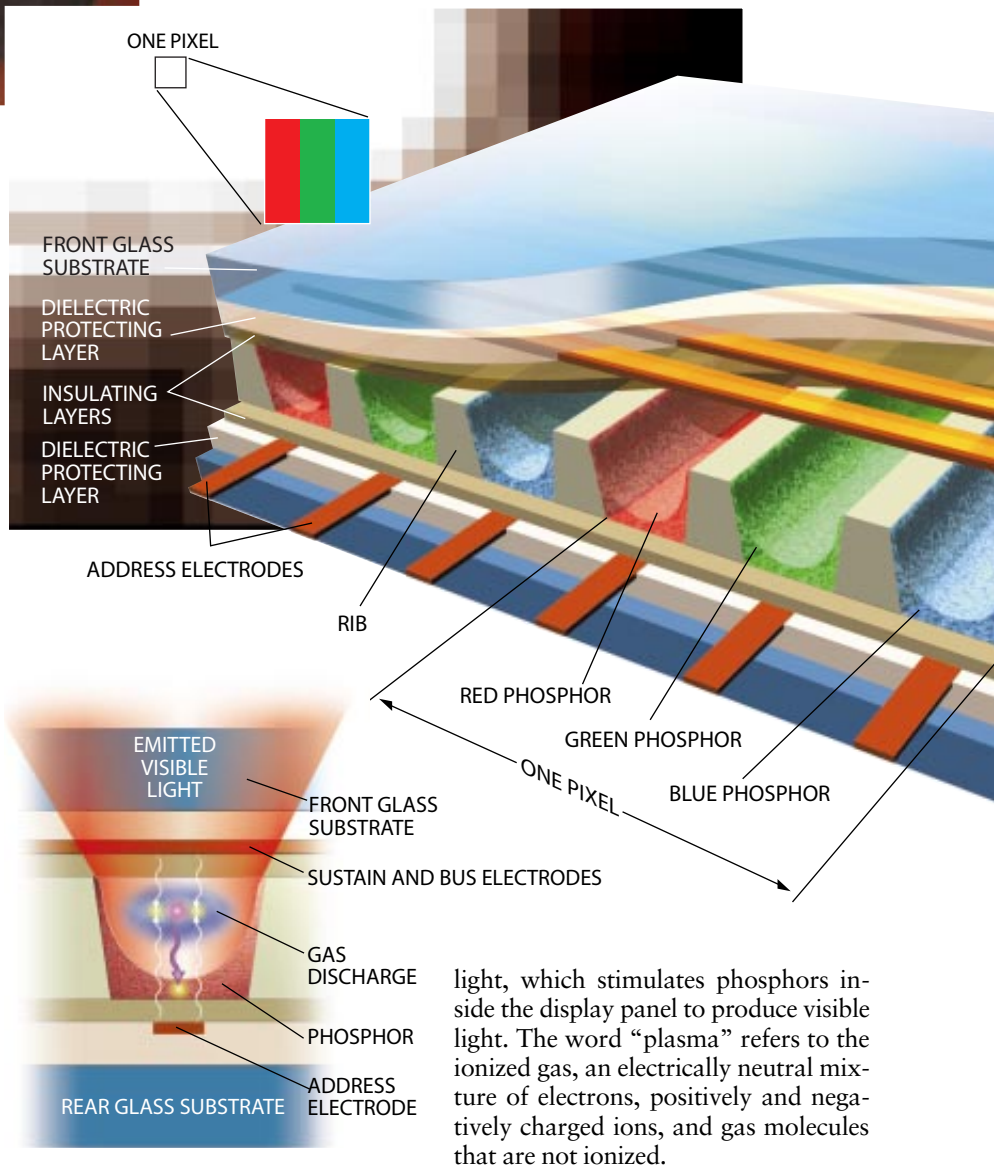
Besides making HDTV feasible, the plasma display will also fulfill a goal that has existed almost since the first days of television, when engineers envisioned a thin, light TV screen that would hang on the wall like a framed picture. Over the years, many different flat-panel displays were invented, and a few, notably liquid-crystal displays, have achieved commercial success in laptop computers, aircraft cockpits and various electronic instruments. Other than the plasma display, however, no flat-panel technology can currently be produced in the sizes necessary for HDTV. Moreover, most of the other types are not well suited to displaying full-motion video, although small, full-color liquid-crystal displays are now often used in portable televisions and video equipment.

How They Work

Flat-panel displays are superior in several important respects to television picture tubes, which are more properly known as cathode-ray tubes, and the projection systems used in today's big-screen TVs [*see box on pages 74 and 75*]. In addition to the obvious weight and thickness advantages, the resolution and luminance of a flat-panel display are uniform over its entire surface; cathode-ray tubes and projectors tend to lose both resolution and luminance at their corners and edges. Flat panels are generally better-looking, more efficient and less power-hungry than projection displays of the same size but are less efficient than cathode-ray tubes.

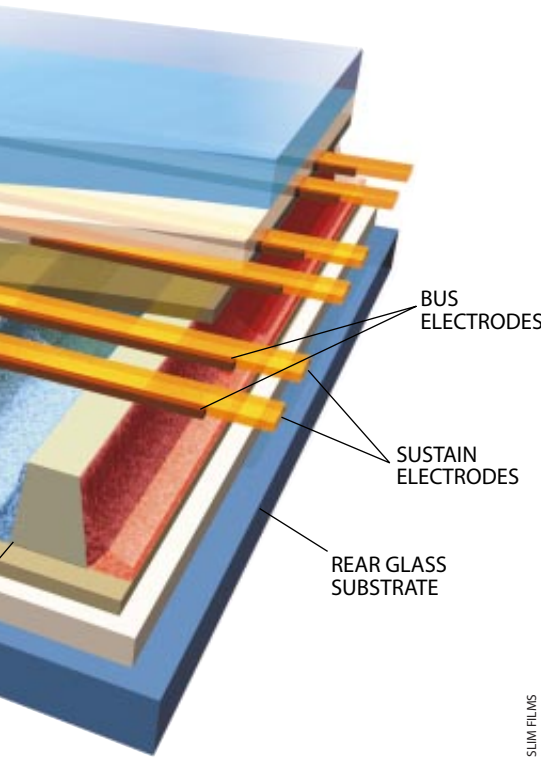
The technology that is on the verge of fulfilling the decades-old goal of a flat-screen television has much in common with the ordinary fluorescent lamp. Like a fluorescent tube, a plasma panel works by passing an electric current through a gas, knocking electrons from the atoms or molecules—ionizing them, in other words. In a plasma display, the gas is typically a blend of helium and xenon. The ionized gas emits ultraviolet light, which stimulates phosphors inside the display panel to produce visible light. The word "plasma" refers to the ionized gas, an electrically neutral mixture of electrons, positively and negatively charged ions, and gas molecules that are not ionized.

This light-emitting phenomenon occurs in miniature throughout the panel. Essentially the screen is made up of tiny contiguous pixels (picture elements), each containing three cells. Each cell comprises an electrode pair or triad as well as chemicals, known as phosphors, painted on the walls, the back and sometimes the front of the cell. Current



flowing between the electrodes ionizes the gas, which emits the ultraviolet light that stimulates the phosphor in that cell. Each cell contains phosphor that emits one primary color—red, green or blue; the combination can produce almost any color.

This basic description glosses over a few aspects that made the development of plasma screens so challenging. One is the way in which the gas is ionized. If two electrodes are immersed in a gas and a voltage is applied between them, nothing will happen until there is at least one charged particle—an electron or ion—available in the gas to conduct



SLIM FILMS

electric current between the electrodes. As this charged particle, which is known as a priming particle, moves through the gas under the force of the applied voltage, it collides with other neutral molecules or atoms and ionizes them. These charged particles in turn ionize other atoms, and the current increases. The ultraviolet (and visible and infrared) light occurs when the positive ions of the gas capture electrons from the plasma, emitting photons.

Two of the major challenges in designing the cells in a display involve supplying priming particles and limiting the current that flows between electrodes. For a priming particle, an itinerant cosmic ray might do the trick, ionizing a molecule—if we could wait until one happened to pass between the electrodes. We cannot, of course—the cells in a display device must be turned on and off many times a second. So display devices incorporate an auxiliary discharge to provide priming particles, which make it possible for a cell to turn on with little delay.

The cells in a display must also be fabricated in such a manner that the current flowing between the electrodes does not keep increasing, which it naturally will if not checked. Without this limit on current, the gas between the electrodes would become fully ionized, breaking down its resistance and permitting an arc discharge to occur. Such an arc would destroy the device.

There are two types of plasma displays, alternating current and direct current. The alternating-current model is the dominant type. In this display, current is inherently limited by the design of the cells. The two electrodes of each cell are separated from the surrounding gas by a thin layer of insulating material. This thin film behaves like a capacitor, allowing only alternating current to pass through it. Furthermore, the capacitor limits the flow of current, so that the discharge cannot increase to the arc level.

Organizing the Signals

Just as a bank is more than a collection of money, a display is more than an array of light-emitting elements. Thus, understanding how a plasma display works requires a knowledge not only of how a cell works but also of how signals are changed into images on the display. Like those of any flat-panel display, the cells of a plasma panel are arranged in matrix fashion. The electrical leads to the electrodes are arranged in rows and columns, with a light-emitting device—a cell, in other words—at each intersection. Drive circuits are required for every row and column in the display, to apply the voltages necessary both to light and to turn off a cell.

Television displays were the first electronic image displays, and all displays now inherit the TV image process, which was designed for cathode-ray tubes. In a cathode-ray tube, an electron beam sweeps across the screen in what is known as raster fashion, starting at the top left, moving across the screen, down one line, across the screen again, and so on. A complete picture comprises a frame, and TV frames are repeated 30 times per second (25 times in much of Europe and many other places). Most computer displays are "refreshed" at 60 or more frames per second, even though the image content may not have changed.

With their matrix arrangement of cells, lack of a single, primary set of electron beams and slower response speeds than cathode-ray tubes, flat-panel displays cannot use the raster technique. In these displays, the signal must be arranged and applied in such a way that it addresses an entire row of cells at once, then those in the row below it, and so on until a complete frame has been created.

In an alternating-current PDP, a sequence of alternating-polarity pulses of voltage just too low to excite a discharge is continuously applied to the entire panel; this is called the "sustain" voltage. A cell is turned on by applying a higher voltage to the corresponding pair of row and column electrodes. This higher voltage breaks down the gas, allowing current to flow and charge to build up on the capacitor in the cell. The accumulated charge produces a voltage opposed to the driving voltage, so the discharge terminates and the cell stops emitting light. On the next, opposite-polarity pulse, however, that stored voltage adds to the applied sustain voltage, and the cell fires again and continues to fire as long as this sustaining waveform is applied. Each gas-discharge pulse is only a microsecond or so long and produces only a small amount of light. Yet there may be as many as 50,000 pulses per second, so the total amount of light is substantial.

A cell is turned off by applying a pulse that reduces the stored charge on the capacitors. Differences in brightness ("gray scale") are achieved by controlling the fraction of time that each cell is on. The main drawback of the scheme, and a problem common to all flat-panel displays, is the huge number of drive circuits required around the periphery of the display. A typical computer display, with 480 rows and 640 columns of pixels, with each pixel containing three cells, requires $480 + (640 \times 3) = 2{,}400$ drive circuits, independent of the physical size of the display. In comparison, a cathode-ray tube needs only two deflection circuits, one each for horizontal and vertical movement of the electron beam, and three modulating

Big, Bright and Organic

by Paul E. Burrows, Stephen R. Forrest and Mark E. Thompson

Of the dozens, if not hundreds, of flat-panel-display technologies under development, organic electroluminescent devices may be the most promising. With active layers only about a thousandth of the thickness of a human hair, these displays were first demonstrated in their present form just over 10 years ago. Already the best of them are many times brighter per unit area than a conventional television picture tube and function—at less than 10 volts—for several tens of thousands of hours. Significantly, organic displays are fast enough to handle full-motion video, making them suitable for use not only as computer monitors but also as televisions.

The most impressive features of these displays, however, are unusual ones made possible by intrinsic characteristics of their organic constituents. For example, weak intermolecular bonding in these materials permits the active layers to be deposited on thin, flexible plastic substrates. The resulting displays could be rolled up or made to conform to surfaces of various contours. Both the active layers and their plastic substrates are so light that the weight of a computer monitor could in principle be reduced to a few ounces.

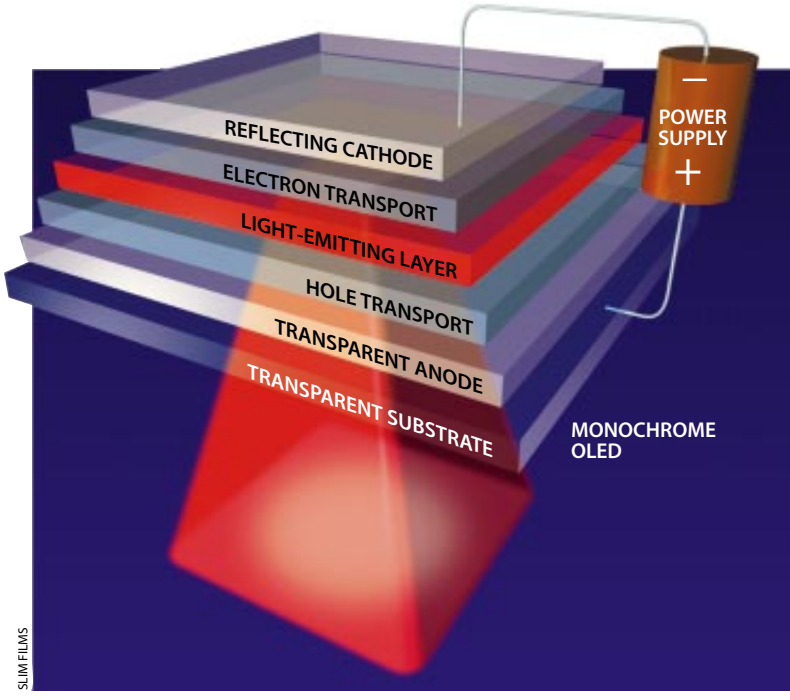
The ability of organic semiconductors to form high-quality thin films on practically any flat substrate is potentially transformative, because it promises to allow engineers to grow organic electroluminescent devices over large areas at very low cost—a critical requirement for large-area flat-panel displays. Conventional semiconductors, on the other hand, must be grown on costly substrates, known as wafers, chosen to have a crystal structure similar to the film. Someday it may very well be possible to create wall-size organic displays, although so far the largest full-color displays, all of which are experimental, measure 125 millimeters (five inches) diagonally.

Such potential has spurred multimillion-dollar commercial research programs at Eastman Kodak, Motorola and many other companies in the U.S., Japan and Europe. Small independent firms, such as Universal Display Corporation in the U.S. and Cambridge Display Technology in the U.K., have been created solely to turn the innovation, known as an organic light-emitting device (OLED), into a commercial product.

In contrast to conventional semiconductors like silicon, organic semiconductors are composed mainly of carbon, hydrogen, nitrogen and oxygen, each molecule containing dozens or hundreds of atoms. These organic molecules emit photons when they fall back from an electronically excited state to the ground, or relaxed, state. Key organic semiconductors include Alq_3, α-NPD and a red-light-emitting molecule called DCM, which is also used in organic-based lasers and whose full formula is 4-(dicyanomethylene)-2-methyl-6-(4-dimethylaminostyryl)-4H-pyran.

Light, Naturally

Typically, an OLED consists of three layers of organic semiconductor material—the electron-transport, light-emitting and hole-transport layers—that are sandwiched between a cathode and an anode [*see illustration below*]. Negative charge carriers (electrons) and positive charge carriers (called holes; a hole is an absence of an electron) are injected from the cathode and anode, respectively, and transported to the light-emitting region under the influence of an electrical field. In the organic molecules of the light-emitting region, an electron and a hole associate with one another, forming an entity called an exciton. This electrically neutral entity can jump from molecule to molecule, typically decaying in a few billionths of a second to give light of a particular energy and, hence, color. Pick the right organic mole-



circuits, one for each color. The flat panel has a price; even with the economy of integrated circuits, the sheer number of circuits required is a major contributor to the relatively high cost of flat-panel displays.

In the other type of plasma panel, the direct-current gas-discharge display, the applied voltages are also pulses. Unlike those in the alternating-current type, however, they are always of the same polarity. Researchers have lavished a great deal of creativity on this form, but at the moment there is one dominant large-screen type. It was developed mainly under the 30-year sponsorship of NHK Laboratories. (NHK is the Japanese government broadcasting system.)

NHK has shown 1,056-millimeter panels with good luminance and contrast. In comparison with alternating-current panels, however, they have been heavy and inefficient, generating excessive heat along with the light. In addition, their complicated panel geometry is expensive to produce.

Another serious problem is life span. The material used on the negative electrode (the cathode) tends to sputter—atoms are actually knocked out of the cathode and redeposited elsewhere in the cell. The problem is worse than in alternating-current panels because the sputtered material is a metal, not an insulator. It is opaque, causing a loss of light, and is conductive, possibly pro-

cules for the light-emitting layer, and red, green or blue light can be produced. Either the anode or the cathode must be transparent for the light to escape to the viewer.

Alternatively, both electrodes can be made transparent by constructing them from the semiconductor indium tin oxide. These clear OLEDs open up intriguing possibilities—imagine, for instance, your office window doubling as a large display screen.

A second class of organic compounds used in OLEDs are polymers. Similar to conventional plastics, the polymeric materials consist of long chains of a molecular unit repeated as many as thousands of times in the molecule. Polymer OLED displays made by this method can be almost as bright and efficient as their vacuum-deposited, small molecule-based cousins. To date, however, the operating lifetime of vacuum-deposited OLEDs exceeds that of the polymeric devices by an order of magnitude, caused in part by the exceptional material purity obtainable with small organic molecules.

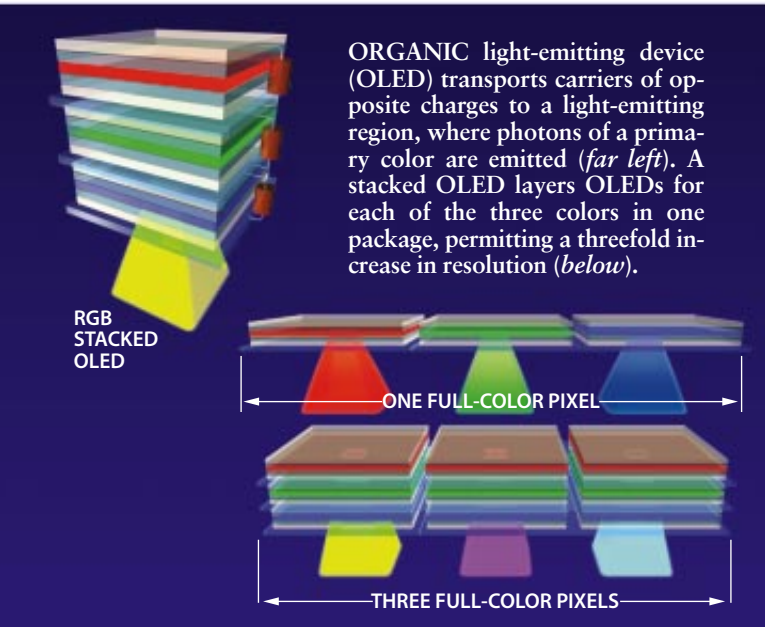
Commercial Applications

The most efficient and longest-lived vacuum-deposited OLED displays emit green light, whereas red- and blue-emitting materials lag behind in both respects. The best devices currently deliver a luminous efficiency of up to 18 lumens per watt, which is five to 10 times the figure for liquid-crystal displays, but only about a quarter as much as for television picture tubes, which are extraordinarily efficient. After encapsulation to keep out atmospheric water vapor, OLEDs have a continuous operating lifetime of tens of thousands of hours (to 50 percent initial brightness), which is comparable with that of a picture tube. The displays' operating temperature, however, must be kept below 100 degrees Celsius. The fast pace of OLED research suggests that before long, better red- and blue-emitting materials will be available, and the temperature problem will be solved.

In the meantime, monochrome OLEDs emitting only green light are already good enough for commercial applications. Pioneer Corporation, the Tokyo-based high-fidelity and electronics giant, has recently introduced a 2.5-by-10-centimeter monochrome OLED flat panel for eventual use in automobiles to display navigational information.

Monochrome OLEDs are just the beginning. The unique properties of organic semiconductors allow for new device architectures unobtainable with conventional technologies. For example, we have recently shown that three transparent OLEDs, one each for red, blue and green, can be built on top of one another. Each of these three devices emits its own color through the transparent organic layers, the contacts and the substrate, allowing the entire device area to emit any combination of the three colors, as shown in the illustration at the left. This stacked OLED may replace the conventional architecture for fabricating full-color displays, in which red, green and blue pixels (picture elements) are adjacent to one another in a single layer in the display. The drawbacks of the conventional arrangement are that the need for at least three pixels at each point in the display limits its resolution and that for close-up viewing (such as for a head-mounted display), extremely small pixels may be required to maintain the illusion of full color.

A stacked OLED display avoids this problem because each pixel can be any desired color—effectively tripling the resolution of the display, while also cutting down on the amount of dark space between the pixels. As a bonus, because the three colors are vertically stacked on top of one another, the arrangement offers full color at any resolution or viewing distance, because there are no side-by-side pixels for the eye to resolve.



ORGANIC light-emitting device (OLED) transports carriers of opposite charges to a light-emitting region, where photons of a primary color are emitted (*far left*). A stacked OLED layers OLEDs for each of the three colors in one package, permitting a threefold increase in resolution (*below*).

PAUL E. BURROWS, STEPHEN R. FORREST and MARK E. THOMPSON all began working on organic displays at Princeton University. Burrows and Forrest are still at the university, Burrows as a research scholar at the Center for Photonics and Optoelectronic Materials and Forrest as professor and chair of the department of electrical engineering. Thompson is now an associate professor in the department of chemistry at the University of Southern California.

ducing electrical short circuits in the cell.

Another interesting direct-current display has been developed by a struggling group in St. Petersburg, Russia. It uses a simpler structure. Its most interesting feature is that it is built in tiles, each tile about 20 centimeters square, containing 64 × 64 pixels. The seams at the edges of each tile are very thin, so the tiles can be put together with the seam lines almost invisible. Each tile is driven independently; thus, the luminance of a display is independent of its size, unlike the NHK display, in which the luminance is inversely proportional to the number of rows. This characteristic leads to the possibility of very large displays.

Problems, Problems

The most fundamental problem of all matrix displays is that many drive circuits are required—one for each row and each column, plus the control circuitry needed to activate the electrodes at the right times. In plasma displays the problem is exacerbated by the fact that the voltages needed are high (for microcircuitry, that is)—about 100 volts. Furthermore, the circuitry must be able to accommodate the brief but large pulse current at the beginning of each discharge cycle. As a result, the cost of drive circuitry for plasma displays is higher than for many competing display technologies, although this cost

The Competition

In the push for big pictures, plasma display panels (PDPs) are leading the race at the moment. But a multitude of competitors is getting better all the time. Some of the most important of these include the following:


PANASONIC CONSUMER ELECTRONICS

Projectors. Projection displays are not hang-on-the-wall displays, but they are a means of making big pictures and thus are real competition. The projector image can come either from three hard-working monochrome cathode-ray tubes, typically 180 to 230 millimeters (seven to nine inches) in diameter, or from a high-intensity light source that sends its light through one or more so-called light valves—essentially electronic slides. Bulky in comparison to plasma displays, projectors are also hard-put to meet the requirements of high-definition television (HDTV).

Cathode-ray tubes. Big, heavy and power-hungry, the cathode-ray tube is also a mature technology that is good and getting better. Its main drawback is limited size: although it is possible to make models bigger than about 900 millimeters (35 inches) in diagonal, they are extremely impractical for most applications. And the consensus is that HDTV requires pictures larger than 1,000 millimeters.

Liquid-crystal displays (LCDs). The notebook computer has created a huge demand for LCDs, and the display industry has responded by producing displays of improving resolution, viewing angle, color rendition and efficiency, at steadily decreasing cost. This more desirable active-matrix type incorporates at each cell an electronic switch (a thin-film transistor); this configuration adds to the cost but greatly improves performance over passive displays. Larger displays are being worked on—Sharp has demonstrated a 1,015-millimeter (roughly 40-inch) diagonal color screen—but the cost goes up dramatically with display area. Viewing angles are less limited than in the past, but LCDs continue to have difficulty with moving images because the material is inherently slow to switch.

An interesting hybrid is the plasma-addressed LCD, invented at Tektronix in the U.S. and now being pursued by Sony, Sharp and Philips. Row selection is effected by narrow channels filled with gas, which is ionized row by row; modulation of the individual cells is effected by controlling the voltage on the column electrodes, as in more conventional LCDs. The plasma-filled channels are much less expensive than thin-film transistors, and this approach is particularly interesting for large displays. Panels measuring 1,065 millimeters have been demonstrated at trade shows. Luminance is excellent; unfortunately, reproduction of fast-moving images is subject to the usual LCD problems, and the viewing angle has been more limited than that of some advanced LCDs.

Light-emitting diodes (LEDs). These are semiconductor devices in which electrons and electron voids, called holes, are injected into a region where they can combine, emitting light (and a lot of heat). Now that blue LEDs are available, as a result of a long and expensive struggle, full-color displays have been built.


STANLEY

The displays are based on many thousands of pixels, each consisting of a red, a green and a blue LED. Such displays can be yards across and bright enough to see in broad daylight. Their

will certainly come down as production volumes increase.

Another lingering problem with plasma displays is that ion bombardment tends to destroy the phosphors over time, reducing their light output; this phenomenon has limited life to 10,000 hours or so in the past, but now lifetimes of 30,000 hours (to half the initial luminance) are claimed by some manufacturers. Most life-lengthening strategies involve keeping the discharge away from the phosphor.

The efficiency and luminance of alternating-current plasma displays could also stand some improvement. Efficiencies are now about one lumen per watt and are in fact getting better. Luminances are moving well past 100 candelas per square meter (29 foot-lamberts). At least one manufacturer is already claiming 300 candelas per square meter—close to the 350 candelas per square meter that is generally considered the minimum requirement for television sets intended for sale to consumers. (In general, television displays must be brighter than computer displays because they are usually at greater distances from the viewer than computer monitors are and must look good even when viewed in bright room light.)


FUJITSU MICROELECTRONICS

PLASMA DISPLAY measures 1,056 millimeters (42 inches) diagonally and costs approximately $11,000. A mere 150 millimeters (six inches) thick, the monitor weighs 40 kilograms (87 pounds) and can display 16.7 million colors.

Although the luminance of both alternating- and direct-current plasma display panels is already high enough for many applications, it is not yet as high as that of cathode-ray tubes. The easiest way of increasing the luminance is to boost the power; unfortunately, this tactic pushes up the cost of the drive circuitry and the power supplies. Hence, there is a continued demand for greater efficiency. Developers are now investigating two possibilities: getting more useful photons

high cost, however, has limited them to use in public squares, shopping centers and the like.

It has not been possible to build large matrices of LEDs in single-crystal compound semiconductors. But a new development—organic LEDs—may well change this picture [*see box on pages 74 and 75*].

PLANAR

Electroluminescence. This technology makes use of a thin phosphor film sandwiched between insulator layers. A high voltage applied across the sandwich generates electrons at the interface between insulator and phosphor, and these electrons produce light in much the same way that electrons produce light in the phosphors of a cathode-ray tube. Thin-film electroluminescent displays are rugged, with wide viewing angles and excellent life. At the moment, though, they are expensive, small and have a limited color gamut. Nevertheless, they have found modest success in a few niche markets.

DOUG PECK *PixTech*

A new development, electroluminescence from organic polymers, may have a significant impact on this technology [*see box on pages 74 and 75*].

Field-emission displays. More than 20 years ago workers at the Stanford Research Institute showed that modest voltages applied to very sharp points of molybdenum or some other refractory metal produced an electrical field high enough to pull electrons out of the metal at room temperature. They realized that these so-called cold-cathode electron emitters could be the basis of a cathode-ray-tubelike device with one or more cathodes for each pixel.

The hangup at the time was making the cathodes. Since then, advances in semiconductor fabrication technology and materials have made these devices much more interesting. One company, PixTech in France, is now in commercial production. There is a host of other development efforts, including major ones at Candescent in California and at Motorola in Arizona.

Canon, the Japanese company, has announced a type of cathode that could transform this display type into a more marketable product. Rather than sculpting single-crystal silicon into microscopic points, the Canon surface-conduction-emission cathode pushes electrons into the vacuum from a narrow gap between two electrodes printed on a glass substrate. If this device meets early expectations, cathodes of almost any size could be produced by ink-jet printing or similar inexpensive means.

Although many other problems remain to be solved, the field-emission display could be a substantial competitor to the LCD for small displays and has perhaps a better chance than the LCD of seriously challenging the PDP in the large-screen area.

Vacuum fluorescent displays. Based on another old idea, this display is a form of low-voltage, flat cathode-ray tube. Electrons are emitted from wire cathodes at low temperature and directed to phosphors capable of efficient light emission at low voltage. The pixels to be illuminated are selected by metal-mesh grids between the cathodes and the phosphor. Vacuum fluorescent displays are often seen as clock readouts, especially in automobiles, on kitchen appliances and on many videocassette recorders, but attempts to move this technology to large image displays have been unsuccessful so far. —*A.S.*

FUTABA CORPORATION

from the discharge, and using those ultraviolet photons more effectively in the cells to produce visible light.

Invention and the Mass Market

Like any form of commercial technology, a successful display requires both the invention of a device and the invention and improvement of manufacturing methods. The two are inextricably intertwined; an extremely clever device that is too expensive to be produced at a competitive price cannot become a mass-market item. In practice, this has meant that clever ideas, unless pursued by companies with extraordinarily deep pockets (often assisted by government grants for research and development), get shelved.

It has been a particular strength of the big Japanese companies that they have had the will and the wherewithal to invest in manufacturing technology over long periods. They have also benefited from an engineering and labor force that has painstakingly improved manufacturing technologies. These are major reasons why the centers of display production are mostly in the Far East.

Display work is exciting because there are so many novel ideas. New techniques keep appearing, even as the older techniques get better. It is also exciting because some form of display device is a key component of most systems. Whether it washes clothes or controls a satellite launch, a piece of equipment must interact with humans, making a good display an essential feature. The industry, as well as the displays it produces, is well worth watching. SA

The Author

ALAN SOBEL, a consultant based in Evanston, Ill., has been doing research and development on electronic displays for more than 35 years. He is editor of the *Journal of the Society for Information Display* and contributes to the society's monthly *Information Display* magazine. He has degrees in electrical engineering from Columbia University and a Ph.D. in physics from the Polytechnic Institute of Brooklyn.

Further Reading

FLAT-PANEL DISPLAYS AND CRTS. Edited by Lawrence E. Tannas, Jr. Van Nostrand Reinhold, 1985.

ELECTRO-OPTICAL DISPLAYS. Edited by Mohammad A. Karim. Marcel Dekker, 1992.

DISPLAY SYSTEMS: DESIGN AND APPLICATIONS. Edited by Lindsay W. MacDonald and Anthony C. Lowe. Wiley SID series in display technology. Published in association with the Society for Information Display. John Wiley & Sons, 1997.

Digital Television: Here at Last

After a long and contentious process, a digital standard in the U.S. has finally emerged. It will soon replace today's antiquated television system

by Jae S. Lim

Within a few months television in North America will undergo a change as fundamental and sweeping as the advent of color. In November broadcasters in large metropolitan areas will begin digital broadcasts, which promise much sharper picture and sound than the current system provides. Called digital television, or DTV for short, the new system also has many features that are absent from conventional broadcasting, such as auxiliary channels for data and easy connection to computers and telecommunications networks.

The changeover from the current system, established in the 1940s and 1950s by the National Television System Committee (NTSC), to the new digital form has been a slow, often contentious process. There were many years of competition and cooperation among major organizations. Officials at the Federal Communications Commission (FCC), broadcasters, television manufacturers and academics tried to come up with a digital standard that would not render existing TVs immediately obsolete. Some of the details, especially those regarding the introduction of computer services to the television realm, still need to be worked out.

Meanwhile the international situation remains unclear. So far Canada and the Republic of Korea have committed themselves to the new U.S. standard. Most of Asia, Europe and Latin America, however, are now evaluating that standard, as well as other possibilities.

EDWIN H. REMSBERG *Gamma Liaison*

But most of the dust has settled, and advances in communications, signal processing and very large scale integration technologies have enabled a major overhaul of the television system rooted in half-century-old technology. The new system will operate in channels mainly in the ultrahigh-frequency (UHF) band, spanning 470 to 890 megahertz (channels 14 to 83). The old and new systems will coexist until 2006, when broadcasters are expected to cease using NTSC signals on both the very high frequency (VHF) band, between 54 and 216 megahertz (channels 2 to 13), and the UHF band. The FCC will then reallocate those channels to digital TV or to other services, such as wireless communications.

Historically, the FCC has exercised authority over technical standards for terrestrial broadcast media. The FCC's influence, however, does extend to cable and satellite operators, because they may wish to adopt terrestrial broadcast standards in order to avoid confusion, added expense and technical complexities for themselves and for their customers. In this scenario, though, the complication is that some cable and satellite operators have already begun broadcasting digital signals—but not in high-definition format. At some point, this situation will have to be resolved if broadcast, cable and satellite operators are all to coexist in technical harmony.

Driving television to the new system was the desire for a better picture, which actually began before the digital era. In the late 1960s NHK, Japan's government-sponsored TV broadcaster, made the first foray into high-definition television, or HDTV. Together with Japan's electronics manufacturers, NHK developed an analog system called MUSE, for multiple sub-nyquist encoding. The encoding was a scheme that delivered five times more information to produce a sharper image. The problem was that it required five times as much broadcasting airspace as well. The NTSC system delivers audio and video signals within a six-megahertz-wide channel, but MUSE would need some 30 megahertz of channel space.

There was not enough room to accommodate such a scheme. In the early days of NTSC, plenty of bandwidth existed. The NTSC method, by today's standards, is highly inefficient in the way it uses bandwidth. In video the intensities among neighboring picture elements (pixels) are quite often very similar or at least dependent on those immediately near them. Because NTSC transmits entire scenes without exploiting this dependence, much of the redundant information is transmitted again. This type of inefficient use of the spectrum generates interference among the different NTSC signals. As the number of TV stations increased, interference became a serious problem.

The solution was to leave some channels, known as taboo channels, unused. Typically, in a highly populated area in the U.S., only one of two VHF channels and only one of six UHF channels are assigned. Mobile radio and other telecommunications services also placed demands on the available bandwidth. The end result was that there was simply no room in the U.S. spectrum for bandwidth-hungry HDTV systems such as MUSE.

At the request of the broadcast organizations, the FCC created the Advisory Committee on Advanced Television Service (ACATS) in September 1987. The ACATS was chartered to advise the FCC on matters related to the standardization of advanced television service in the U.S., including the establishment of a technical standard. In 1988 the ACATS asked industries, universities and research laboratories to propose advanced television standards.

Dueling Approaches

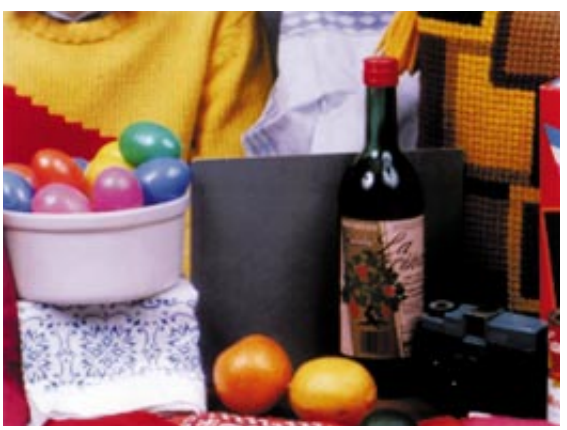
While the ACATS screened the proposals and prepared testing laboratories for formal technical evaluation, the FCC made a key decision. In March 1990 it selected the simulcast approach for advanced television service, rather than the receiver-compatible approach.

In the latter approach, current TV sets would be able to obtain an HDTV signal and generate a viewable picture. The NTSC relied on such a method when it introduced color, so that existing black-and-white sets would not become obsolete. The receiver-compatible approach worked because color information did not require a large amount of bandwidth. A small portion of the six-megahertz-wide channel could contain the color information without seriously affecting the black-and-white picture.

An HDTV signal, however, requires much more information than a color signal. So the receiver-compatibility requirement would demand an augmentation channel to carry the additional information HDTV requires—basically, another six-megahertz-wide channel.

BROADCASTING PREPARATIONS for digital television (DTV), such as those undertaken here by WRC-TV (*opposite page*), an NBC affiliate in Washington, D.C., include digital television cameras, a wider set and higher-quality furnishings. DTV can offer a sharp, high-definition image (HDTV) and greater aspect ratio, or width-to-height proportion, than the current system, as shown in the simulated comparison (*below*).

EXISTING TV



DIGITAL TV

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Avoiding a Hard Decision

Even though the new standard adopted by the Federal Communications Commission (FCC) exceeds the initial requirements of a digital television system and will serve very well for years to come, its ultimate degree of success depends in part on certain decisions the FCC left to the marketplace. The standard is flexible: it permits not only the six transmission formats originally proposed by the Grand Alliance but also a large number of additional formats, in part because of a last-minute compromise reached among industries with differing preferences. To understand the potential consequences of this decision, it is helpful to review the basic differences between interlaced scanning and its modern replacement, progressive scanning.

The interlaced scanning format is the basis for the National Television System Committee (NTSC) standard. The format delivers pictures that are snapshots of a scene recorded a specific number of times per second. In interlaced scanning, a single snapshot consists of only odd lines; it is followed by a snapshot consisting of only even lines, and the sequence then repeats. Each snapshot—odd and even—is called a field.

Although only snapshots of a scene are shown, the human visual system perceives continuous motion, as long as there are enough snapshots. In the NTSC system, the snapshots are flashed at a rate of 60 fields per second. This rate is sufficiently high for the viewer to perceive continuous motion and to avoid notice of display flickering. The total number of active lines is approximately 480, and each line contains approximately 420 pixels, or picture elements. (The number of lines represents the vertical spatial resolution in the picture, and the number of pixels per line represents the horizontal spatial resolution.)

An alternative to interlaced scanning is progressive scanning, in which all the lines for each snapshot are scanned. The snapshot in progressive scanning is called a frame. This type of scanning, however, proved impractical for the NTSC method; because of bandwidth limitations, it could be done only at 30 frames per second, which would cause images to flicker badly. In principle, a receiver could avoid flicker if each progressively scanned frame is displayed twice. But that would require a frame memory, a technology that did not exist when NTSC was standardized. Hence, interlace took hold of the television manufacturing and broadcasting industries. Subsequently, much HDTV video equipment, including cameras, has been developed for interlaced scanning.

Unfortunately, interlace introduces certain video artifacts, such as interline flicker. Though usually not very bothersome for entertainment material, they can be quite objectionable in the display of text, computer graphics or other material that has sharp lines. Consider a sharp horizontal line that is in the odd field but not in the even field. Even though the overall scan rate is 60 hertz, the scan rate for the sharp horizontal line is slower, at 30 hertz. As a result, the line flickers. Partly for this reason, almost all computer monitors use progressive scanning. It is possible to deinterlace a signal so that it can be viewed on a progressive-scan monitor, but the conversion requires complex signal processing to achieve high performance. And the image is still not as good as it would be if the transmission were originally in a progressive-scan format. Not surprisingly, the computer industry advocates progressive scanning, which would help it gain a greater foothold in the broadcast industry and the general home market. Other advocates include the movie industry, which desires a better display for film.

The disagreement over transmission formats generated considerable heat and delayed digital TV in the U.S. The Grand Alliance HDTV system proposed five progressive-scan formats and one interlaced-scan format. To accommodate standard definition, the recommendation by the Advisory Committee on Advanced Television Service (ACATS) specified 12 additional standard formats that consisted of both progressive and interlaced formats. But the FCC in December 1996 adopted a compromise that was reached between the broadcast and computer industries. This hastily reached agreement removed most of the restrictions on transmission formats; it not only allowed all the 18 formats that included interlaced scanning but also allowed many additional progressive and interlaced formats. In effect, this decision left the choice of transmission formats to free-market forces.

In fact, there is very good justification for permitting multiple transmission formats. A TV system must accommodate many video input sources: video cameras, film, magnetic and optical media, synthetic imagery and the like, all of which have

SLIM FILMS

This approach poses several major problems. It requires an NTSC channel as a basis to transmit an HDTV signal, which means that the highly spectrum inefficient NTSC system cannot be converted to a modern, efficient technical system. In addition, the introduction of HDTV would permanently require a new channel for each existing NTSC channel.

For these reasons, the FCC adopted the simulcast approach. Unlike the receiver-compatible case, in which the HDTV signal is obtained from the NTSC signal and the additional information is in the augmentation channel, a simulcast HDTV signal is broadcast in its own six-megahertz-wide channel, independently of the NTSC signal. In this way, a modern transmission system could be crafted for the entire HDTV signal.

The drawback, of course, is that today's television sets would not be able to receive an HDTV signal. To ensure that they do not become obsolete immediately, the FCC decided to give one new channel for HDTV service to each of the U.S.'s 1,500 stations that requested it. During the transition period, both services would coexist. Initially the FCC required the same program be broadcast simultaneously or within a short period from each other on both channels, hence the name simulcast. (This requirement was later eliminated.) Once enough of the country was using HDTV, NTSC

their own kind of format. In the NTSC system, which uses only a single transmission format, the various input sources are converted to one format and then transmitted. For instance, the format of film is 24 frames per second; it is converted to 60 fields per second with interlaced scanning (that is, the NTSC format) and then transmitted. One transmission format, however, is an inefficient way of using the available spectrum.

To understand why, consider a movie broadcast. Each new channel can handle 60 frames per second (at 720 lines of progressive scanning). Film is shot at 24 frames per second, so only about half the bandwidth needs to be used (the empty part could contain an additional HDTV movie, computer data or other programming). Forcing a single format means that the movie would have to be transmitted at 60 frames per second—in other words, each frame would be sent two or three times, and there would be less room for other data. The spectrum therefore becomes occupied with repetitive data.

Permitting multiple transmission formats also allows the use of different formats for different applications. For example, broadcasters of sports events can use a format that features high resolution. On the other hand, a news program can choose a format with somewhat less resolution and use the resulting savings in channel capacity to transmit an additional program.

In permitting the new digital television standard to have many different transmission formats, the FCC allows for a video transmission format that is identical or very similar to the format of the original source. From the viewpoint of spectrum efficiency, allowing all possible video formats may be ideal.

PROGRESSIVE

INTERLINE FLICKER occurs in interlaced scanning when fine lines (such as the red ones boxing "WATCH.") fall on individual scan lines. When the odd lines are scanned (*top left*), image portions that fall on the even lines are not displayed; the reverse happens in the subsequent (even-line) scan (*bottom left*). As a result, the eye detects a flickering. In progressive scanning (*right*), all lines are scanned in each instance, so there is no interline flicker.

But from the viewpoint of the receiver—that is, the consumer's TV set—too many formats can be costly. A monitor typically can display images in only one format, so the different formats received must be converted to one display format. Allowing for too many formats can complicate the conversion. In any case, most of the benefits derived from multiple formats can be obtained by carefully selecting a small subset of formats.

All new displays in the not too distant future are likely to be progressive because of the format's overall superior performance. But the decision by the FCC to allow both progressive and interlaced formats in the transmission path gives broadcasters a choice. If broadcasters use progressive scanning in the transmission, both interlaced- and progressive-scan displays can be accommodated with little additional cost to the receivers. That is because the progressive-to-interlace conversion can be performed well with relatively simple digital processing. All digital receivers are expected to have such capabilities. But if broadcasters use interlaced scanning, a progressive-scan receiver has to convert the interlaced-scan format. Such a deinterlacing operation demands complex signal processing to achieve good performance, raising the cost of progressive-scan displays, which are currently more expensive than interlaced ones.

Broadcasters have some incentive to send interlaced signals: cameras and other production equipment that utilize interlaced scanning are more readily available now. In addition, substantial amounts of video material exist that have been captured in the interlaced format.

By allowing interlaced formats in the transmission path, the FCC in essence requires deinterlacers in all progressive-scan receivers (unless every single major broadcaster shuns interlaced formats). That is far more burdensome to the general public than requiring deinterlacers at the transmitters, of which there are far fewer. In addition, deinterlacers perform much better at the transmitter than at the receiver, because the transmitter has access to the uncoded original video.

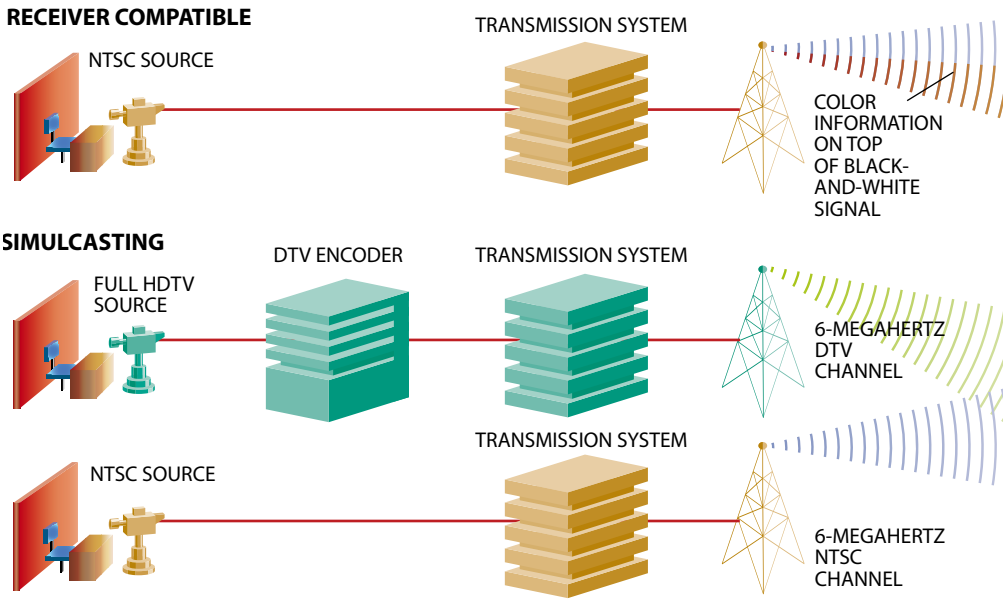
One important function of a standard is to make restrictions that serve the public interest. Limited to a specific video-compression approach, for example, digital TV receivers need to implement only the decoding method for the video-compression approach adopted by the FCC. In the case of the transmission formats, however, the FCC avoided making a difficult decision. The decision now will be made by short-term market forces. That may leave the broadcast industry and the general public tied to interlaced formats and their associated equipment much longer than if the FCC had made a hard decision that narrowed the options to the long-term, preferred progressive-scan formats. —*J.S.L.*

service would be discontinued. The spectrum it previously occupied would be used for additional HDTV channels or for other services.

There are a number of advantages to the simulcast approach. It permits the design of a new, spectrum-efficient HDTV signal that requires significantly less power and that interferes much less with other signals, including the NTSC signal. This design enables the use of the taboo channels. (Without the taboo channels, the FCC cannot give an additional channel to each existing NTSC broadcaster for HDTV service.) In addition, simulcast ultimately eliminates the spectrum-inefficient NTSC channels following the transition period. Furthermore, removing the NTSC signals that have strong interference characteristics allows the more efficient use of other channels. The 1990 FCC ruling was a key decision in the process to standardize HDTV in the U.S.

Soon thereafter many groups began proposing HDTV systems based on the simulcast approach. Ultimately, the FCC's ACATS approved five proposals for formal evaluation. One advocated an analog system; the other four were completely digital. Laboratory tests at the Advanced Television Testing Center in Alexandria, Va., evaluated the five systems. The Advanced Television Evaluation Laboratory in Ottawa, Canada, subjectively judged picture quality.

TRANSMISSION APPROACHES to DTV that were debated early on focused on how televisions should receive both the new and the existing NTSC (National Television System Committee) signals. The receiver-compatible approach was used when color was introduced; the color information could fit in a six-megahertz-wide channel. Because a high-definition and an NTSC signal both cannot fit into one channel, the Federal Communications Commission adopted the simulcast approach, which calls for broadcasters to send out both types of signals.



In February 1993 a special ACATS panel reviewed the test results and made a recommendation. It concluded that the four digital systems substantially outperformed the analog system. The panel also found that each of the four excelled in different aspects. Therefore, it could not recommend one in particular. Instead the panel urged that each digital system be retested after improvements were made by the proponents.

The ACATS accepted the panel's recommendation. But as an alternative, the ACATS also encouraged the four proponents to combine the best elements of the different systems and submit one single system for evaluation.

Toward a Single Standard

The four digital proponents indeed decided to work on a single scheme. In May 1993 they formed a consortium called the Grand Alliance, which comprised seven organizations: AT&T, Zenith, the David Sarnoff Research Center in Princeton, N.J., Chicago-based General Instrument Corporation, the Massachusetts Institute of Technology (for which I was the representative), Philips Electronics North America and France-based Thomson Consumer Electronics. Between 1993 and 1994 the Grand Alliance chose the best technical elements from the four systems and made further improvements on them. A technical standard based on the Grand Alliance HDTV prototype was documented by the Advanced Television Systems Committee, an industry consortium.

To deliver all the information needed for a high-definition image within a six-megahertz-wide channel (which handles about 20 million bits per second), data compression is needed (an uncompressed high-definition image needs about a billion bits per second). The Grand Alliance proposal relies on a scheme called MPEG-2 (the acronym stands for Moving Pictures Experts Group). A key in MPEG compression is that rather than sending whole images, as the current NTSC system does, it sends only the changes in the images. For instance, in a newscast, MPEG would not repeatedly send images of the desk or the background, which remain stationary; instead it would transmit mostly the movements the anchors made. As a result, many fewer data are needed to update an image (of course, every so often, the entire image would be transmitted).

The compressed video, audio and any other data are first multiplexed—that is, combined into one sequence of bits. The bit stream is modulated (imposed on an electromagnetic wave). The modulated signal is then transmitted over the air for terrestrial broadcasting. An antenna would pick up the signal and send it to a receiver, which would demodulate the signal to generate a bit stream. The bit stream is demultiplexed to produce compressed data, which are then decompressed.

In November 1995 the ACATS recommended the standard documented by the Advanced Television Systems Committee to the FCC, which accepted it except for one detail. The standard restricted broadcasters to 18 video resolution formats—that is, ways to transmit images. The FCC eased this restriction in December 1996 [*see box on page 80*].

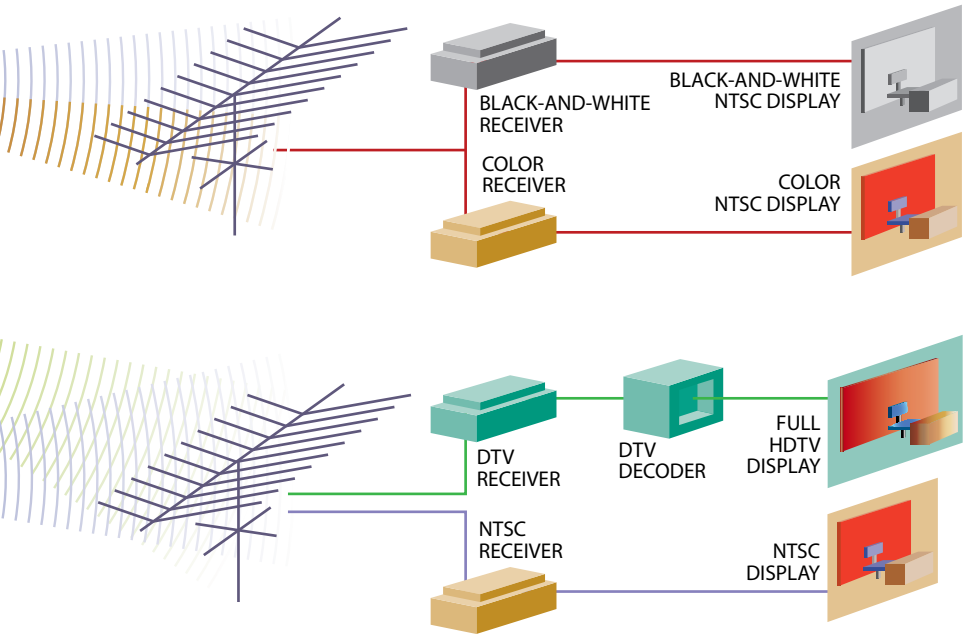
In early 1997 the FCC made additional rulings to support the new technical standard, such as the allocation of channels. Starting this fall, major broadcasters are scheduled to transmit digital TV signals in large metropolitan areas, including Boston, Los Angeles and Washington, D.C.

A digital television system based on the standard will be quite flexible. It can deliver spectacular high-definition video and multichannel surround sound within a six-megahertz bandwidth. Or it can send several programs of standard-definition TV, resolution of which is comparable to that of today's programs. Because of the flexibility, the acronym describing the system changed from HDTV, or high-definition television, to the broader term DTV, for digital television. Moreover, the new standard can reach a larger area with much lower transmitter power than the NTSC system can.

The standard can also incorporate future improvements in technology. For example, MPEG-2 specifies only the syntax of the coded bit streams and the decoding process. This leeway allows some flexibility in the encoding process, which can be upgraded without changing the standard.

Getting the Digital Picture

As we have seen, today's televisions cannot receive the new signal and display a viewable picture. There are two solutions. One, of course, is to obtain a new set: digital TV receivers were demonstrated at the Consumer Electronics Show held in Las Vegas this past January. They will be commercially



available in the fall. Initially a large-screen digital television receiver with HDTV display capability will be expensive—more than $5,000. As more receivers are sold, however, the price will decrease rapidly.

An alternative approach is to attach a "set-top" box to an NTSC receiver. The box, which will cost a few hundred dollars, decodes the digital television signal and converts it to an analog NTSC signal. Although a viewer will not experience HDTV resolution, the video quality will be better than that of the same program broadcast through the analog NTSC channel.

One improvement that will be apparent even without a digital set will be the reception quality. Unlike the analog NTSC signal, which suffers from channel degradations such as multipath effects ("ghosts") and random noise ("snow"), the digital video will be perfect within a certain coverage area, or else there will be no picture at all. This all-or-nothing case mirrors that for digital music, in which a player cannot read dirty or damaged compact discs.

The high spatial resolution of a digital TV system will also allow for increased realism, by way of a bigger screen. In the NTSC system, the recommended viewing distance is approximately seven times the picture height, to avoid the visibility of a line structure on the screen. Hence, for a display screen two thirds of a meter high (a 40-inch-diagonal set), the recommended viewing distance is 4.25 meters. This distance makes it difficult to place a large-screen television receiver in many homes. And because of the distance, the viewing angle is approximately 10 degrees.

For a digital TV with high definition, the recommended viewing distance is typically three times the picture height. So for the set described above, this distance would be 1.8 meters, which is practical for many homes. And with a wider screen, this distance affords a larger, more realistic field of view of about 30 degrees.

Another plus of the new standard is the increased aspect ratio, or the relation between width to height of the image. An NTSC television receiver displays images that have an aspect ratio of 4:3, which mirrored movies that were made when the NTSC system was first developed. Since then, movies have become much wider. To reflect this change, a digital TV system has a larger aspect ratio, such as 16:9.

A digital television system can also deliver CD-quality, multichannel sound, which enhances the visual experience. Multiple audio channels can produce the effect of surround sound employed in movie theaters. They can also be used for transmitting different languages in the same video program.

The new digital system will transform television in another, perhaps more profound way. Traditionally, the TV has been a stand-alone device whose primary purpose was entertainment. But in addition to a better picture and multiple programs, the digital system will be able to accommodate data from telecommunications services, delivering such information as stock quotes or e-mail. In this way, the display can be used as a videophone, a newspaper service or a computer monitor.

The promise of this integration has been the impetus for the computer industry's entrance into the domain previously dominated by broadcasters. Much still needs to be worked out between the two camps, but the convergence seems inevitable. Almost assuredly, in the near future, the television will be regarded as the home center for entertainment, telecommunications and information. In the end, it may be this integration that will have a far greater effect on us than the better picture and sound that the digital system will deliver. SA

The Author

JAE S. LIM is professor in the department of electrical engineering and computer science at the Massachusetts Institute of Technology, which he joined in 1978 as an assistant professor. He is also director of M.I.T.'s Advanced Telecommunications Research Program. His research interests include digital signal processing and its applications to image, video, audio and speech processing. During the past 10 years, he has participated in the Federal Communications Commission's work on standardizing advanced television. He represented M.I.T. in the design of the M.I.T./G.I. system (one of the four finalist systems) and in the design of the Grand Alliance system that led to the FCC's adoption of the U.S. Digital Television Standard in December 1996. The opinions expressed here are those of the author only. Lim thanks David Staelin of M.I.T. for his valuable comments and suggestions.

Further Reading

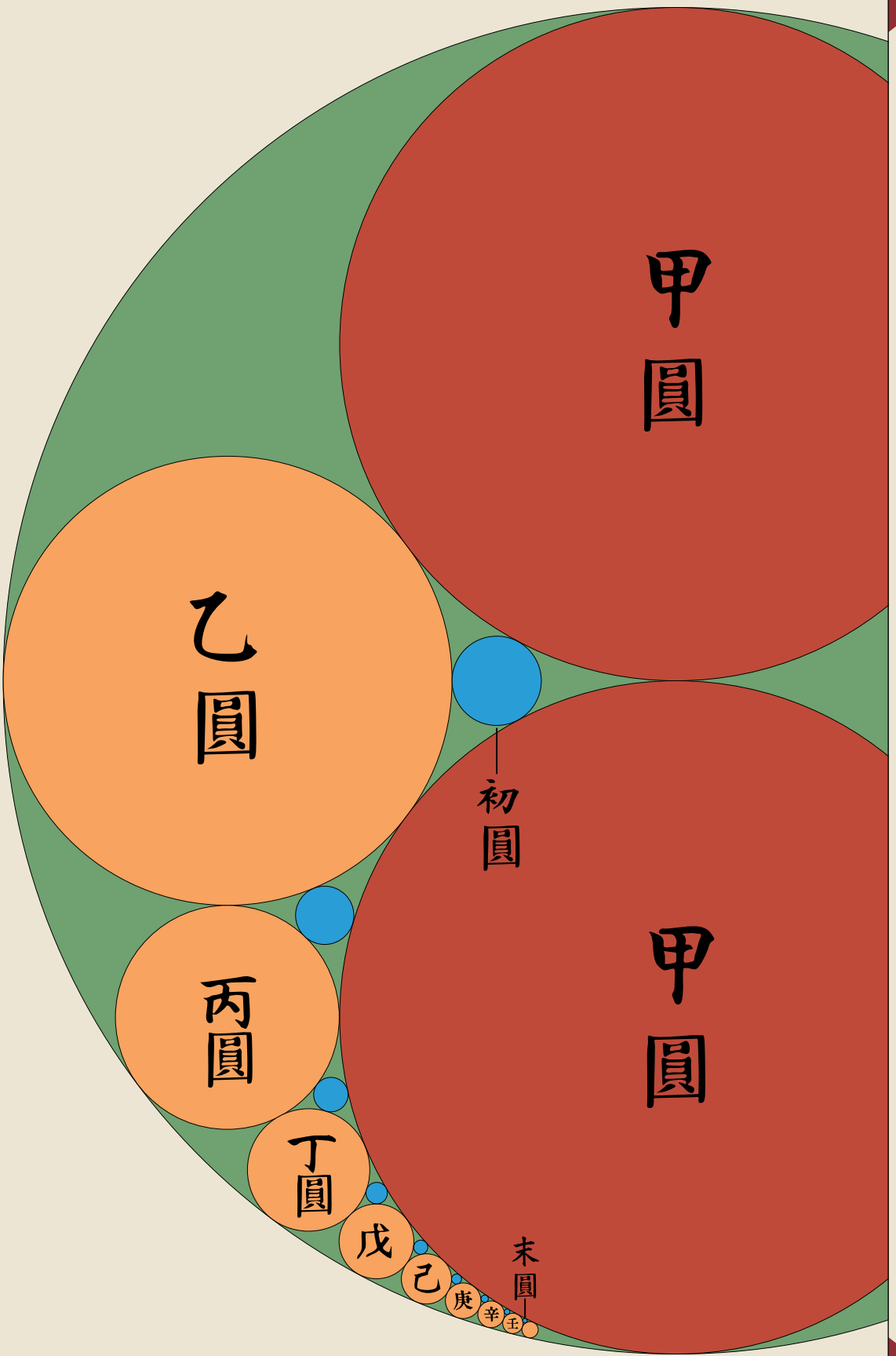
TWO-DIMENSIONAL SIGNAL AND IMAGE PROCESSING. Jae S. Lim. Prentice-Hall, 1990.

DIGITAL PICTURES: REPRESENTATION, COMPRESSION AND STANDARDS. Second edition. A. N. Netravali and B. G. Haskell. Plenum Press, 1995.

HDTV AND THE NEW DIGITAL TELEVISION. Special report in *IEEE Spectrum,* Vol. 32, No. 4, pages 34–79; April 1995.

ATSC DIGITAL TELEVISION STANDARD. Advanced Television Systems Committee; September 16, 1995.

所懸干東都本所枡島妙見堂者一事



Japanese Temple Geometry

During Japan's period of national seclusion (1639–1854), native mathematics thrived, as evidenced in sangaku*—wooden tablets engraved with geometry problems hung under the roofs of shrines and temples*

by Tony Rothman, with the cooperation of Hidetoshi Fukagawa

Of the world's countless customs and traditions, perhaps none is as elegant, nor as beautiful, as the tradition of *sangaku,* Japanese temple geometry. From 1639 to 1854, Japan lived in strict, self-imposed isolation from the West. Access to all forms of occidental culture was suppressed, and the influx of Western scientific ideas was effectively curtailed. During this period of seclusion, a kind of native mathematics flourished.

Devotees of math, evidently samurai, merchants and farmers, would solve a wide variety of geometry problems, inscribe their efforts in delicately colored wooden tablets and hang the works under the roofs of religious buildings. These *sangaku,* a word that literally means mathematical tablet, may have been acts of homage—a thanks to a guiding spirit—or they may have been brazen challenges to other worshipers: Solve this one if you can!

For the most part, *sangaku* deal with ordinary Euclidean geometry. But the problems are strikingly different from those found in a typical high school geometry course. Circles and ellipses play a far more prominent role than in Western problems: circles within ellipses, ellipses within circles. Some of the exercises are quite simple and could be solved by first-year students. Others are nearly impossible, and modern geometers invariably tackle them with advanced methods, including calculus and affine transformations. Although most of the problems would be classified today as recreational or educational mathematics, a few predate known Western results, such as the Malfatti theorem, the Casey theorem and the Soddy hexlet theorem. One problem reproduces the Descartes circle theorem. Many of the tablets are exceptionally beautiful and can be regarded as works of art.

Pleasing the *Kami*

It is natural to wonder who created the *sangaku* and when, but it is easier to ask such questions than to answer them. The custom of hanging tablets at shrines was established in Japan centuries before *sangaku* came into existence. Shintoism, Japan's native religion, is populated by "eight hundred myriads of gods," the *kami.* Because the *kami,* it was said, love horses, those worshipers who could not present a living horse as an offering to the shrine might instead give a likeness drawn on wood. As a result, many tablets dating from the 15th century and earlier depict horses.

Of the *sangaku* themselves, the oldest surviving tablet has been found in Tochigi Prefecture and dates from 1683. Another tablet, from Kyoto, is dated 1686, and a third is from 1691. The 19th-century travel diary of the mathematician Kazu Yamaguchi refers to an even earlier tablet—now lost—dated 1668. So historians guess that the custom first arose in the second half of the 17th century. In 1789 the first collection containing typical *sangaku* problems was published. Other collections followed throughout the 18th and 19th centuries. These books were either handwritten or printed with wooden blocks and are remarkably beautiful. Today more than 880 tablets survive, with references to hundreds of others in the various collections. From a survey of the extant *sangaku,* the tablets seem to have been distributed fairly uniformly throughout Japan, in both rural and urban districts, with about twice as many found in Shinto shrines as in Buddhist temples.

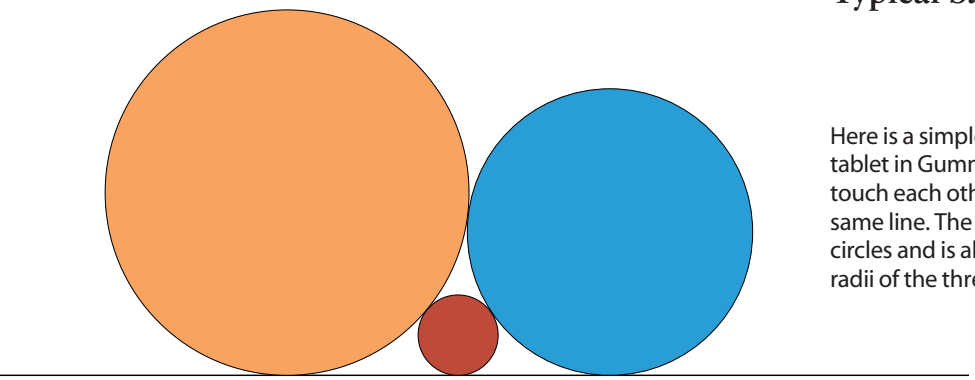
Most of the surviving *sangaku* contain more than one theorem and are frequently brightly colored. The proof of the theorem is usually not given, only the result. Other information typically includes the name of the presenter and the date. Not all the problems deal solely with geometry. Some ask for the volumes of various solids and thus require calculus. (This point raises the interesting question of what techniques the practitioners brought into play; some speculations will be offered in the following discussion.) Other tablets contain Diophantine problems—that is, algebraic equations requiring solutions in integers.

In modern times the *sangaku* have been largely forgotten but for a few devotees of traditional Japanese mathematics. Among them is Hidetoshi Fukagawa, a high school teacher in Aichi Prefecture, roughly halfway between Tokyo and Osaka. About 30 years ago Fukagawa decided to study the history of Japanese mathematics in hopes of finding better ways to teach his courses. A mention of the math tablets in an old library book greatly astonished him, for he had never heard of such a thing. Since then, Fukagawa, who holds a Ph.D. in mathematics, has traveled widely in Japan to study the tablets and has amassed a collection of books dealing not only

BRYAN CHRISTIE, AFTER TOSHIHISA IWASAKI

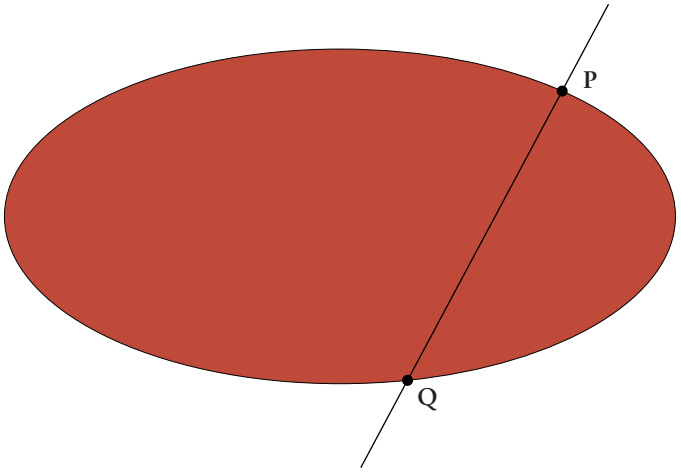
***SANGAKU* PROBLEMS typically involve multitudes of circles within circles or of spheres within other figures. This problem is from a *sangaku,* or mathematical wooden tablet, dated 1788 in Tokyo Prefecture. It asks for the radius of the *n*th largest blue circle in terms of *r,* the radius of the green circle. Note that the red circles are identical, each with radius *r*/2. (Hint: The radius of the fifth blue circle is *r*/95.) The original solution to this problem deploys the Japanese equivalent of the Descartes circle theorem.** **(The answer can be found on page 91.)**

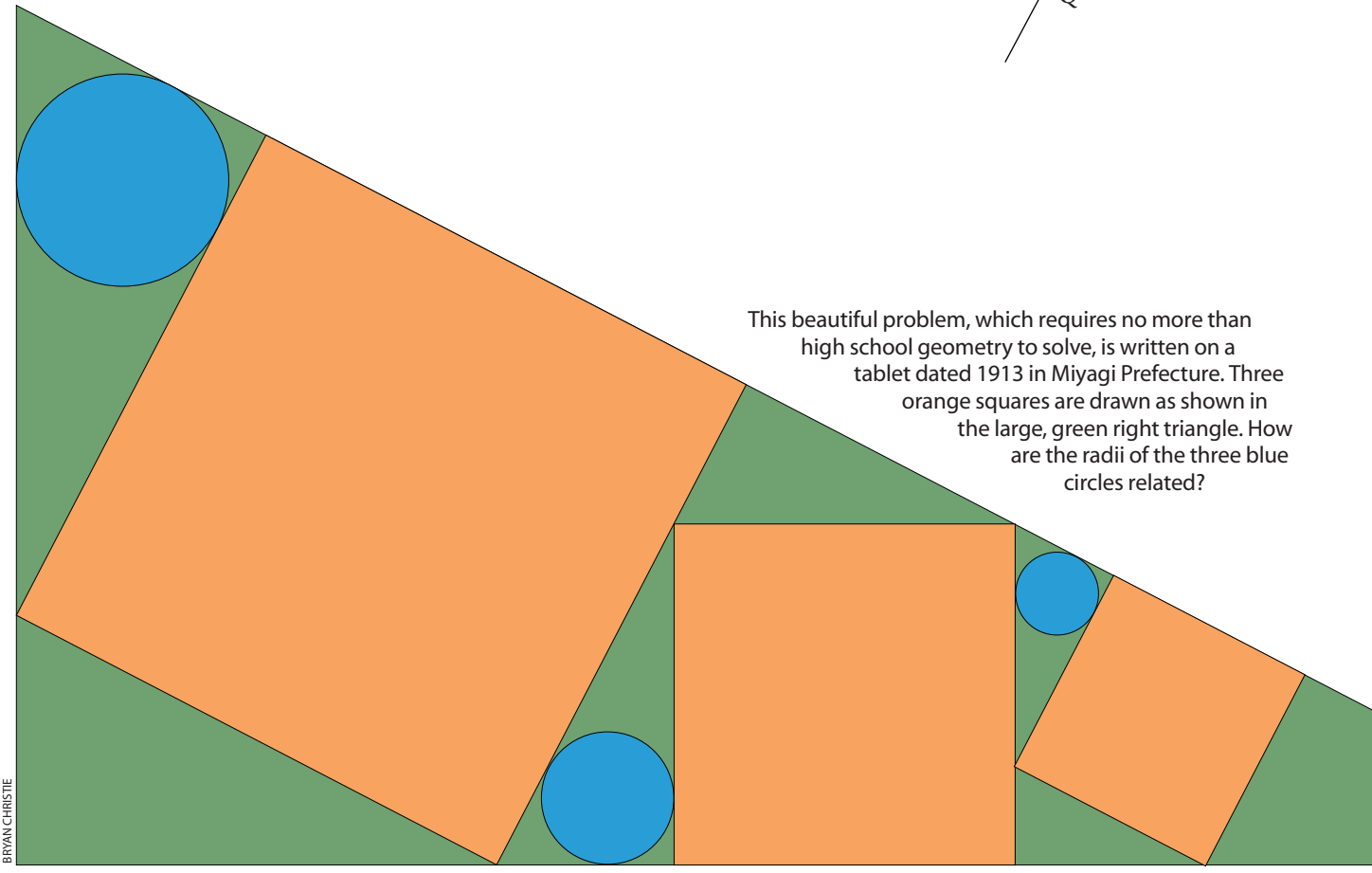
Typical *Sangaku* Problems*



Here is a simple problem that has survived on an 1824 tablet in Gumma Prefecture. The orange and blue circles touch each other at one point and are tangent to the same line. The small red circle touches both of the larger circles and is also tangent to the same line. How are the radii of the three circles related?

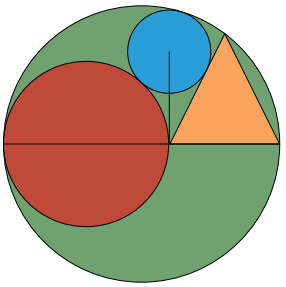
This striking problem was written in 1912 on a tablet extant in Miyagi Prefecture; the date of the problem itself is unknown. At a point P on an ellipse, draw the normal PQ such that it intersects the other side. Find the least value of PQ. At first glance, the problem appears to be trivial: the minimum PQ is the minor axis of the ellipse. Indeed, this is the solution if $b < a \leq \sqrt{2}b$, where a and b are the major and minor axes, respectively; however, the tablet does not give this solution but another, if $2b^2 < a^2$.



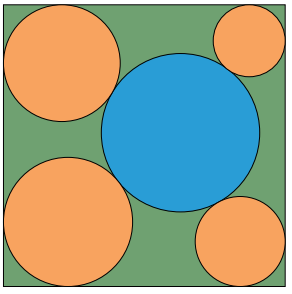


This beautiful problem, which requires no more than high school geometry to solve, is written on a tablet dated 1913 in Miyagi Prefecture. Three orange squares are drawn as shown in the large, green right triangle. How are the radii of the three blue circles related?

BRYAN CHRISTIE



In this problem, from an 1803 *sangaku* found in Gumma Prefecture, the base of an isosceles triangle sits on a diameter of the large green circle. This diameter also bisects the red circle, which is inscribed so that it just touches the inside of the green circle and one vertex of the triangle, as shown. The blue circle is inscribed so that it touches the outsides of both the red circle and the triangle, as well as the inside of the green circle. A line segment connects the center of the blue circle and the intersection point between the red circle and the triangle. Show that this line segment is perpendicular to the drawn diameter of the green circle.



This problem comes from an 1874 tablet in Gumma Prefecture. A large blue circle lies within a square. Four smaller orange circles, each with a different radius, touch the blue circle as well as the adjacent sides of the square. What is the relation between the radii of the four small circles and the length of the side of the square? (Hint: The problem can be solved by applying the Casey theorem, which describes the relation between four circles that are tangent to a fifth circle or to a straight line.)

*Answers are on page 91.

with *sangaku* but with the general field of traditional Japanese mathematics.

To carry out his research, Fukagawa had to teach himself *Kambun,* an archaic form of Japanese that is closely related to Chinese. *Kambun* is the Japanese equivalent of Latin; during the Edo period (1603–1867), scientific works were written in this language, and only a few people in modern Japan are able to read it fluently. As new tablets have been discovered, Fukagawa has been called in to decipher them. In 1989 Fukagawa, along with Daniel Pedoe, published the first collection of *sangaku* in English. Most of the geometry problems accompanying this article were drawn from that collection.

Wasan versus *Yosan*

Although the origin of the *sangaku* cannot be pinpointed, it can be localized. There is a word in Japanese, *wasan,* that is used to refer to native Japanese mathematics. *Wasan* is meant to stand in opposition to *yosan,* or Western mathematics. To understand how *wasan* came into existence—and with it the unusual *sangaku* problems—one must first appreciate the peculiar history of Japanese mathematics.

Of the earliest times, very little is definitely known about mathematics in Japan, except that a system of exponential notation, similar to that employed by Archimedes in the *Sand Reckoner,* had been developed. More concrete information dates only from the mid-sixth century A.D., when Buddhism—and, with it, Chinese mathematics—made its way to Japan. Judging from the works that were taught at official schools at the start of the eighth century, historians infer that Japan had imported the great Chinese classics on arithmetic, algebra and geometry.

According to tradition, the earliest of these is the *Chou-pei Suan-ching,* which contains an example of the Pythagorean theorem and the diagram commonly used to prove it. This part of the tome is at least as old as the sixth century B.C.

A more advanced state of knowledge is represented in the *Chiu-chang Suan-shu,* considered the most influential of Chinese books on mathematics. The *Chiu-chang* describes methods for finding the areas of triangles, quadrilaterals, circles and other figures. It also contains simple word problems of the type that torment many high school students today: "If five oxen and two sheep cost eight taels of gold, and two oxen and eight sheep cost eight taels, what is the price of each animal?" The dates of the *Chiu-chang* are also uncertain, but most of it was probably composed by the third century B.C. If this information is correct, the *Chiu-chang* contains perhaps the first known mention of negative numbers and an early statement of the quadratic equation. (According to some historians, the ancient Egyptians had begun studying quadratic equations centuries before, prior to 2000 B.C.)

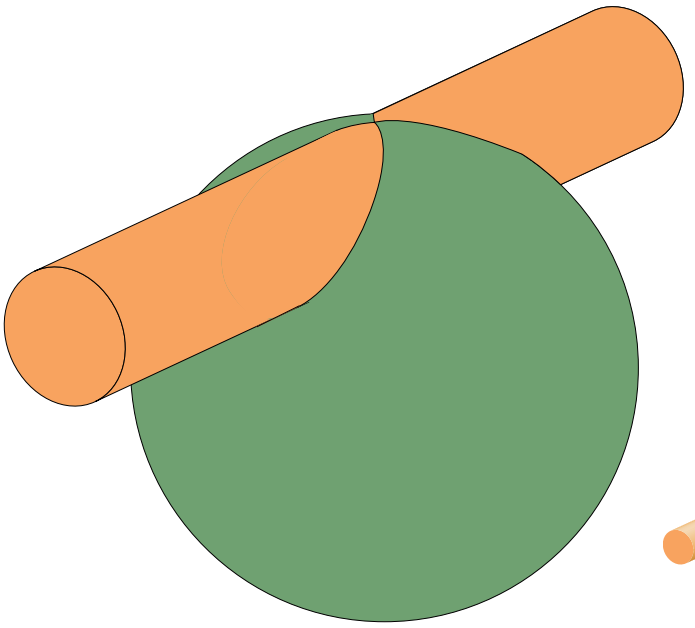
Despite the influx of Chinese learning, mathematics did not then take root in Japan. Instead the country entered a dark age, roughly contemporaneous with that of Western Europe. In the West, church and monastery became the centers of learning; in Japan, Buddhist temples served the same function, although mathematics does not seem to have played much of a role. By some accounts, during the Ashikaga shogunate (1338–1573) there could hardly be found in all Japan a person versed in the art of division.

It is not until the opening of the 17th century that definite historical records exist of any Japanese mathematicians. The first of these is Kambei Mori, who prospered around the year 1600. Although only one of Mori's works—a booklet—survives, he is known to have been instrumental in developing arithmetical calculations on the *soroban,* or Japanese abacus, and in popularizing it throughout the country.

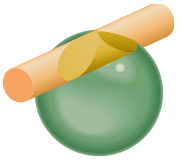
The oldest substantial Japanese work on mathematics actually extant belongs to Mori's pupil Koyu Yoshida (1598–1672). The book, entitled *Jinko-ki* (literally, "small and large numbers"), was published in 1627 and also concerns operations on the *soroban. Jinko-ki* was so influential that the name of the work often was synonymous with arithmetic. Because of the book's influence, computation—as opposed to logic—became the most important concept in traditional Japanese mathematics. To the extent that it makes sense to credit anyone with the founding of *wasan,* that honor probably goes to Mori and Yoshida.

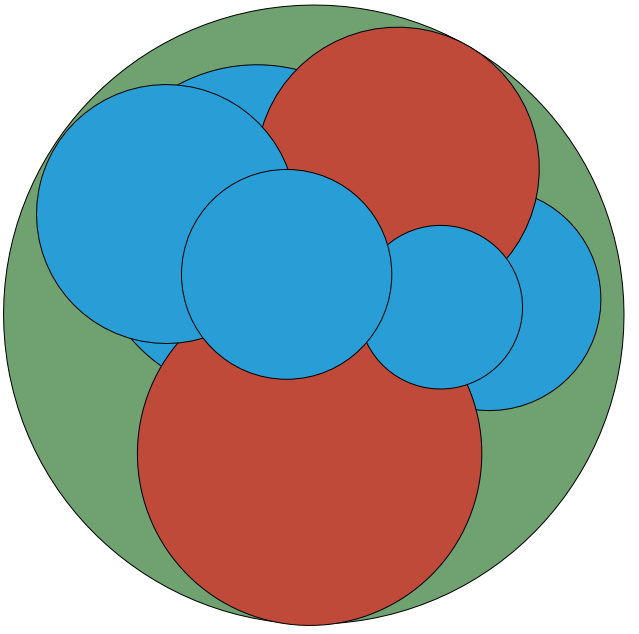
A Brilliant Flowering

Wasan, though, was created not so much by a few individuals but by something much larger. In 1639 the ruling Tokugawa shogunate (during the Edo period), to strengthen its power and diminish challenges to its reign, de-

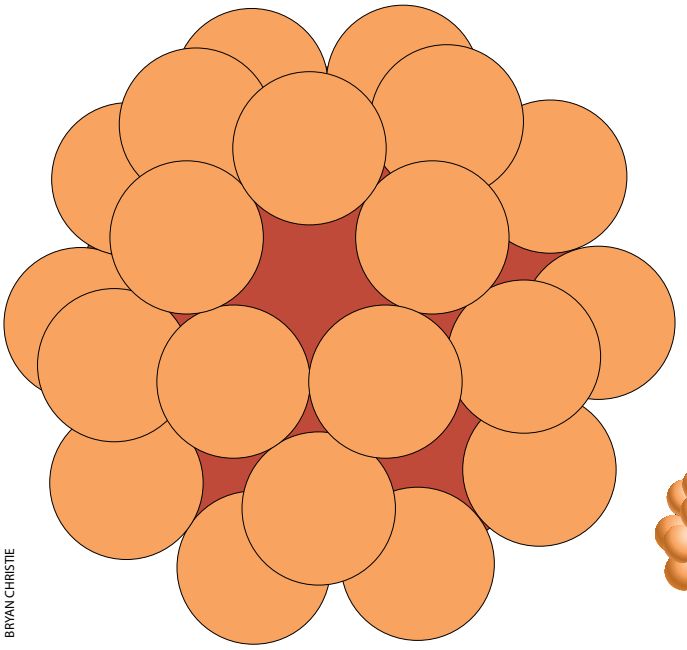


From a *sangaku* dated 1825, this problem was probably solved by using the *enri*, or the Japanese circle principle. A cylinder intersects a sphere so that the outside of the cylinder is tangent to the inside of the sphere. What is the surface area of the part of the cylinder contained inside the sphere? (The inset shows a three-dimensional view of the problem.)

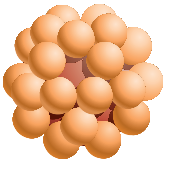




This problem is from an 1822 tablet in Kanagawa Prefecture. It predates by more than a century a theorem of Frederick Soddy, the famous British chemist who, along with Ernest Rutherford, discovered transmutation of the elements. Two red spheres touch each other and also touch the inside of the large green sphere. A loop of smaller, different-size blue spheres circle the "neck" between the red spheres. Each blue sphere in the "necklace" touches its nearest neighbors, and they all touch both the red spheres and the green sphere. How many blue spheres must there be? Also, how are the radii of the blue spheres related? (The inset shows a three-dimensional view of the problem.)



Hidetoshi Fukagawa was so fascinated with this problem, which dates from 1798, that he built a wooden model of it. Let a large sphere be surrounded by 30 small, identical spheres, each of which touches its four small-sphere neighbors as well as the large sphere. How is the radius of the large sphere related to that of the small spheres? (The inset shows a three-dimensional view of the problem.)



BRYAN CHRISTIE

*Answers are on page 91.

creed the official closing of Japan. During this time of *sakoku,* or national seclusion, the government banned foreign books and travel, persecuted Christians and forbade Portuguese and Spanish ships from coming ashore. Many of these strictures would remain for more than two centuries, until Commodore Matthew C. Perry, backed by a fleet of U.S. warships, forced the end of *sakoku* in 1854.

Yet the isolationist policy was not entirely negative. Indeed, during the late 17th century, Japanese art and culture flowered so brilliantly that those years go by the name of *Genroku,* for "renaissance." In that era, *haiku* developed into a fine art form; No and Kabuki theater reached the pinnacle of their development; *ukiyo-e,* or "floating world" pictures, originated; and tea ceremonies and flower arranging reached new heights. Neither was mathematics left behind, for *Genroku* was also the age of Kowa Seki.

By popular accounts, Seki (1642–1708) was Japan's Isaac Newton or Gottfried Wilhelm Leibniz, although this reputation is difficult to substantiate. If the numbers of manuscripts attributed to him are correct, then most of his work has been lost. Still, there is no question that Seki left many disciples who were influential in the further development of Japanese mathematics.

The first—and incontestable—achievement of Seki was his theory of determinants, which is more powerful than that of Leibniz and which antedates the German mathematician's work by at least a decade. Another accomplishment, more relevant to temple geometry but of debatable origin, is the development of methods for solving high-degree equations. (Much traditional Japanese mathematics from that era involves equations to hundreds of degrees; one such equation is of the 1,458th degree.) Yet a third accomplishment sometimes attributed to Seki, and one that might also bear on *sangaku,* is the development of the *enri,* or circle principle.

The *enri* was quite similar to the method of exhaustion developed by Eudoxus and Archimedes in ancient Greece for computing the area of circles. The main difference was that Eudoxus and Archimedes used *n*-sided polygons to approximate the circle, whereas the *enri* divided the circle into *n* rectangles. Thus, the limiting procedure was somewhat different. Nevertheless, the *enri* represented a crude form of integral calculus that was later extended to other figures, including spheres and ellipses. A type of differential calculus was also developed around the same period. It is conceivable that the *enri* and similar techniques were brought to bear on *sangaku.* Today's mathematicians would use modern calculus to solve these problems.

Spheres within Ellipsoids

During Seki's lifetime, the first books employing the *enri* were published, and the first *sangaku* evidently made their appearance. The dates are almost certainly not coincidental; the followers of Yoshida and Seki must have influenced the development of *wasan,* and, in turn, *wasan* may have influenced them. Fukagawa believes that Seki encountered *sangaku* on his way to the shogunate castle, where he was officially employed as court mathematician, and that the tablets pushed him to further researches. A legend? Perhaps. But by the next century, books were being published that contained typical native Japanese problems: circles within triangles, spheres within pyramids, ellipsoids surrounding spheres. The problems found in these books do not differ in any important way from those found on the tablets, and it is difficult to avoid the conclusion that the peculiar flavor of all *wasan* problems—including the *sangaku*—is a direct result of the policy of national seclusion.

But the question immediately arises: Was Japan's isolation complete? It is certain that apart from the Dutch who were allowed to remain in Nagasaki Harbor on Kyushu, the southernmost island, all Western traders were banned. Equally clear is that the Japanese themselves were severely restricted. The mere act of traveling abroad was considered high treason, punishable by death. It appears safe to assume that if the isolation was not complete, then it was most nearly so, and any foreign influence on Japanese mathematics would have been minimal.

The situation began to change in the 19th century, when the *wasan* gradually became supplanted by *yosan,* a process that produced hybrid manuscripts written in *Kambun* with Western mathematical notations. And, after the opening of Japan by Commodore Perry and the subsequent collapse of the Tokugawa shogunate in 1867, the new government abandoned the study of native mathematics in favor of *yosan.* Some practitioners, however, continued to hang tablets well into the 20th century. A few *sangaku* even date from the current decade. But almost all the problems from this century are plagiarisms.

The final and most intriguing question is, Who produced the *sangaku?* Were the theorems so beautifully drawn on wooden tablets the works of professional mathematicians or amateurs? The evidence is meager.

Only a handful of *sangaku* are mentioned in the standard *A History of Japanese Mathematics,* by David E. Smith and Yoshio Mikami. They cite the 1789 collection *Shimpeki Sampo,* or *Mathematical Problems Suspended before the Temple,* which was published by Kagen Fujita, a professional mathematician. Smith and Mikami mention a tablet on which the following was appended after the solution: "Feudal district of Kakegawa in Enshu Province, third month of 1795, Sonobei Keichi Miyajima, pupil of Sadasuke Fujita of the School of Seki." Mikami, in his *Development of Mathematics in China and Japan,* mentions the "Gion Temple Problem," which was suspended at the Gion Temple in Kyoto by Enkyu Tsuda, pupil of Enri Nishimura. Furthermore, the tablets were written in the specialized language of *Kambun,* signifying the mark of an educated class of practitioners.

From such scraps of information, it is tempting to conclude that the tablets were the work primarily of professional mathematicians and their students. Yet there are reasons to believe otherwise.

Many of the problems are elementary and can be solved in a few lines; they are not the kind of work a professional mathematician would publish. Fukagawa has found a tablet from Mie Prefecture inscribed with the name of a merchant. Others have names of women and children—12 to 14 years of age. Most, according to Fukagawa, were created by the members of the highly educated samurai class. A few were probably done by farmers; Fukagawa recalls how about 10 years ago he visited the former cottage of mathematician Sen Sakuma (1819–1896), who taught *wasan* to the farmers in nearby villages in Fukushima Prefecture. Sakuma had about 2,000 students.

Such instruction recalls the Edo period itself, when there were no colleges or universities in Japan. During that time, teaching was carried out at private schools or temples, where ordinary people would go to study reading, writ-

Works of Art

Sangaku can be found in many Shinto shrines and Buddhist temples throughout Japan (*near right*). The tablets are traditionally hung below the eaves of the religious buildings, in a centuries-old practice of worshipers presenting wooden tablets as acts of homage to their guiding spirits (*center right*). The *sangaku* contain mathematical problems that almost always deal with geometry (*far right*). Many of the tablets are delicately colored (*bottom left*); some have been engraved in gold (*bottom right*). —*T.R.*

ing and the abacus. Because laypeople are more often drawn to problems of geometry than of algebra, it would not be surprising if the tablets were painted with such artistic care specifically to attract nonmathematicians.

The best answer, then, to the question of who created temple geometry seems to be: everybody. On learning of the *sangaku*, Fukagawa came to understand that, in those days, many of the Japanese loved and enjoyed math, as well as poetry and other art forms.

It is pleasant to realize that some *sangaku* were the works of ordinary mathematics devotees, carried away by the beauty of geometry. Perhaps a village teacher, after spending the day with students, or a samurai warrior, after sharpening his sword, would retire to his study, light an oil lamp and lose the world to an intricate problem involving spheres and ellipsoids. Perhaps he would spend days working on it in peaceful contemplation. After finally arriving at a solution, he might allow himself a short rest to savor the result of his hard labor. Convinced the proof was a worthy offering to his guiding spirits, he would have the theorem inscribed in wood, hang it in his local temple and begin to consider the next challenge. Visitors would notice the colorful tablet and admire its beauty. Many people would leave wondering how the author arrived at such a miraculous solution. Some might decide to give the problem a try or to study geometry so that the attempt could be made. A few might leave asking, "What if the problem were changed just so...."

Something for us all to consider. SA

The Author

TONY ROTHMAN received his Ph.D. in 1981 from the Center for Relativity at the University of Texas at Austin. He did postdoctoral work at Oxford, Moscow and Cape Town, and he has taught at Harvard University. Rothman has published six books, most recently *Instant Physics*. His next book is *Doubt and Certainty,* with E.C.G. Sudarshan, to be published this fall by Helix Books/Addison-Wesley. He has also recently written a novel about nuclear fusion. SCIENTIFIC AMERICAN wishes to acknowledge the help of Hidetoshi Fukagawa in preparing this manuscript. Fukagawa received a Ph.D. in mathematics from the Bulgarian Academy of Science. He is a high school teacher in Aichi Prefecture, Japan.

Further Reading

A HISTORY OF JAPANESE MATHEMATICS. David E. Smith and Yoshio Mikami. Open Court Publishing Company, Chicago, 1914. (Also available on microfilm.)

THE DEVELOPMENT OF MATHEMATICS IN CHINA AND JAPAN. Second edition (reprint). Yoshio Mikami. Chelsea Publishing Company, New York, 1974.

JAPANESE TEMPLE GEOMETRY PROBLEMS. H. Fukagawa and D. Pedoe. Charles Babbage Research Foundation, Winnipeg, Canada, 1989.

TRADITIONAL JAPANESE MATHEMATICS PROBLEMS FROM THE 18TH AND 19TH CENTURIES. H. Fukagawa and D. Sokolowsky. Science Culture Technology Publishing, Singapore (in press).

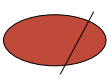
PHOTOGRAPHS BY HIROSHI UMEOKA

Answers to *Sangaku* Problems

Unfortunately, because of space limitations the complete solutions to the problems could not be given here. Additional details can be found at http://www.sciam.com on the Scientific American *World Wide Web site.*

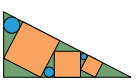
Answer: $r/[(2n-1)^2+14]$. The original solution to this problem applies the Japanese version of the Descartes circle theorem several times. The answer given here was obtained by using the inversion method, which was unknown to the Japanese mathematicians of that era.

Answer: $1/\sqrt{r_3} = 1/\sqrt{r_1} + 1/\sqrt{r_2}$, where r_1, r_2 and r_3 are the radii of the orange, blue and red circles, respectively. The problem can be solved by applying the Pythagorean theorem.



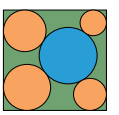
Answer: $PQ = \dfrac{\sqrt{27}\,a^2b^2}{(a^2+b^2)^{3/2}}$

The problem can be solved by using analytic geometry to derive an equation for PQ and then taking the first derivative of the equation and setting it to zero to obtain the minimum value for PQ. It is not known whether the original authors resorted to calculus to solve this problem.



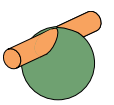
Answer: $r_2^2 = r_1 r_3$, where r_1, r_2 and r_3 are the radii of the large, medium and small blue circles, respectively. (In other words, r_2 is the geometric mean of r_1 and r_3.) The problem can be solved by first realizing that all the interior green triangles formed by the orange squares are similar. The original solution then looks at how the three squares are related.

Answer: In the original solution to this problem, the author draws a line segment that goes through the center of the blue circle and is perpendicular to the drawn diameter of the green circle. The author assumes that this line segment is different from the line segment described in the statement of the problem on page 87. Thus, the two line segments should intersect the drawn diameter at different locations. The author then shows that the distance between those locations must necessarily be equal to zero—that is, that the two line segments are identical, thereby proving the perpendicularity.



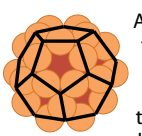
Answer: If a is the length of the square's side, and r_1, r_2, r_3 and r_4 are the radii of the upper right, upper left, lower left and lower right orange circles, respectively, then

$$a = \frac{2(r_1 r_3 - r_2 r_4) + \sqrt{2(r_1 - r_2)(r_1 - r_4)(r_3 - r_2)(r_3 - r_4)}}{r_1 - r_2 + r_3 - r_4}$$



Answer: $16t\sqrt{t(r-t)}$, where r and t are the radii of the sphere and cylinder, respectively.

Answer: Six spheres. The Soddy hexlet theorem states that there must be six and only six blue spheres (thus the word "hexlet"). Interestingly, the theorem is true regardless of the position of the first blue sphere around the neck. Another intriguing result is that the radii of the different blue spheres in the "necklace" (t_1 through t_6) are related by $1/t_1 + 1/t_4 = 1/t_2 + 1/t_5 = 1/t_3 + 1/t_6$.



Answer: $R = \sqrt{5}r$, where R and r are the radii of the large and small spheres, respectively. The problem can be solved by realizing that the center of each small sphere lies on the midpoint of the edge of a regular dodecahedron, a 12-sided solid with pentagonal faces.

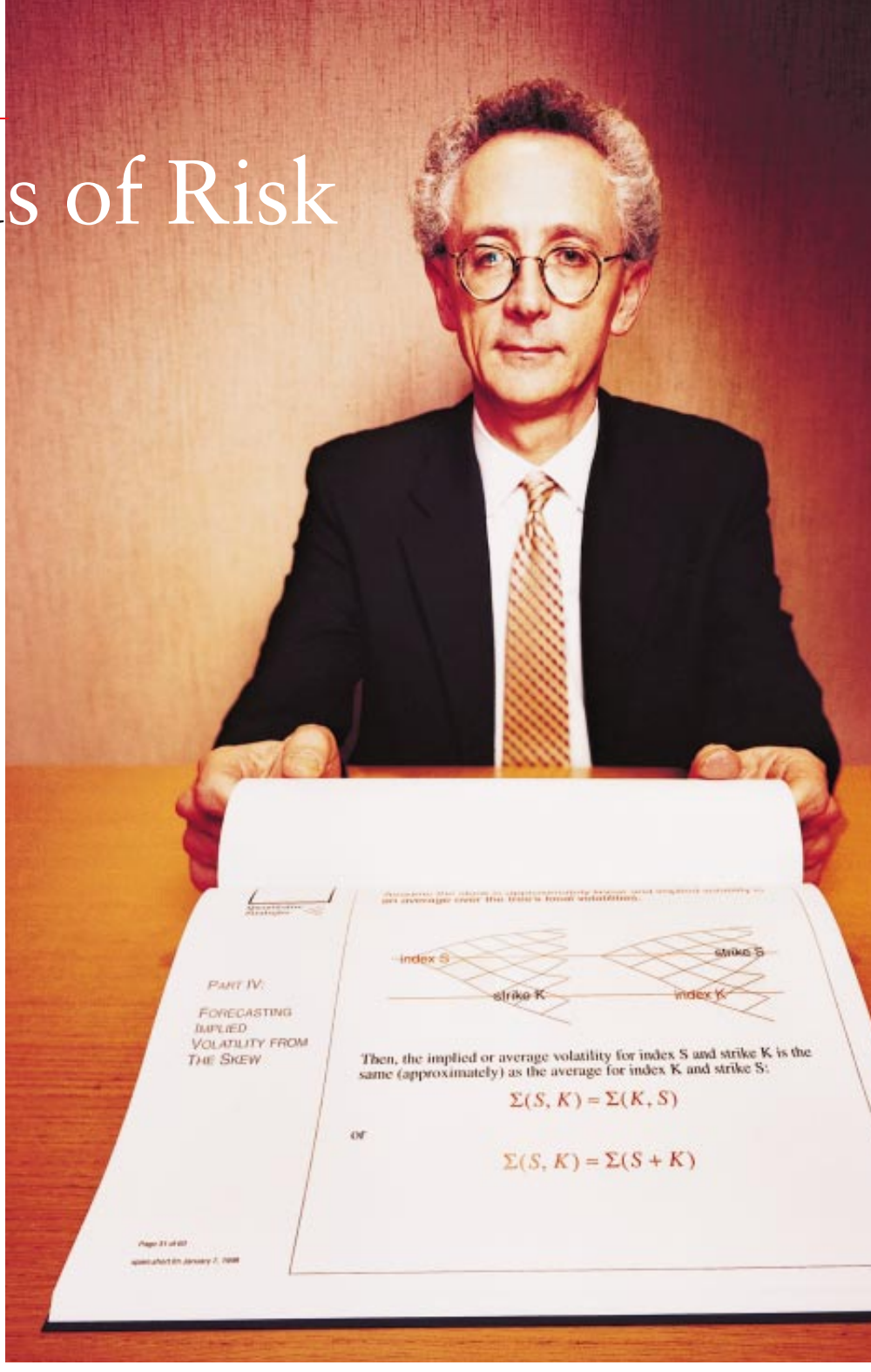
BRYAN CHRISTIE

A Calculus of Risk

by Gary Stix, *staff writer*

Months before El Niño–driven storms battered the Pacific Coast of the U.S., the financial world was making its own preparations for aberrant weather. Beginning last year, an investor could buy or sell a contract whose value depended entirely on fluctuations in temperature or accumulations of rain, hail or snow. These weather derivatives might pay out, for example, if the amount of rainfall at the Los Angeles airport ranged between 17 and 27 inches from October through April. They are a means for an insurer to help provide for future claims by policyholders or a farmer to protect against crop losses. Or the contracts might allow a heating oil supplier to cope with a cash shortfall from a warmer than expected winter by purchasing a heating degree-day floor—a contract that would compensate the company if the temperature failed to fall below 65 degrees as often as expected. "We're big fans of El Niño because it's brought us a lot of business," comments Andrew Freeman, a managing director of Worldwide Weather Trading, a New York City–based firm that writes contracts on rain, snow and temperature.

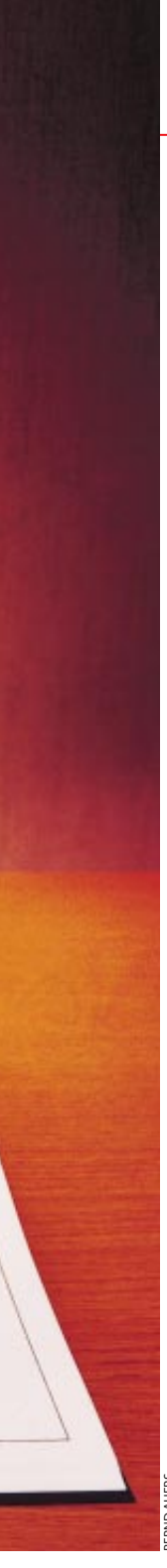
Weather derivatives mark an example of the growing reach of a discipline called financial engineering. This bailiwick of high-speed computing and the intricate mathematical modeling of mathematicians, physicists and economists can help mitigate the vagaries of running a global business. It entails the custom packaging of securities to provide price insurance against a drop in either the yen or the thermometer. The uncertainties of a market crash or the next monsoon can be priced, divided into marketable chunks and sold to someone who is willing to bear that risk—in exchange for a fee or a future stream of payments. "The technology will effectively allow you to completely manage the risks of an entire organization," says Robert A. Jarrow, a professor of finance at Cornell University.



PHYSICIST-TURNED-QUANT Emanuel Derman of Goldman Sachs holds a diagram of a "tree" model he helped to create to show the volatility of stock index prices.

The engineering of financial instruments has emerged in response to turbulence during recent decades in ever more interconnected world markets: a result of floating exchange rates, oil crises, interest-rate shocks and stock-market collapses. The creative unleashing of new products continues with increasingly sophisticated forms of securities and derivatives—options, futures and other contracts derived from an underlying asset, financial index, interest or currency exchange rate. New derivatives will help electric utilities protect against price and capacity swings in newly deregulated markets. Credit derivatives let banks pass off to other parties the risk of default on a loan. Securities that would help a business cope with the year 2000

Financial engineering can lessen exposure to the perils of running a multibillion-dollar business or a small household. But mathematical models used by this discipline may present a new set of hazards



BERND AUERS

bug have even been contemplated.

This ferment of activity takes place against a tainted background. The billions of dollars in losses that have accumulated through debacles experienced by the likes of Procter & Gamble, Gibson Greetings and Barings Bank have given derivatives the public image of speculative risk enhancers, not new types of insurance. Concerns have also focused on the integrity of the mathematical modeling techniques that make derivatives trading possible.

Despite the tarnish, financial engineering received a valentine of sorts in October. The Nobel Prize for economics (known formally as the Bank of Sweden Prize in Economic Sciences) went to Myron S. Scholes and Robert C. Merton, two of the creators of the options-pricing model that has helped fuel the explosion of activity in the derivatives markets.

Options represent the right (but not the obligation) to buy or sell stock or some other asset at a given price on or before a certain date. Another major class of derivatives, called forwards and futures, obligates the buyer to purchase an asset at a set price and time. Swaps, yet another type of derivative, allow companies to exchange cash flows—floating-interest-rate for fixed-rate payments, for instance. Financial engineering uses these building blocks to create custom instruments that might provide a retiree with a guaranteed minimum return on an investment or allow a utility to fill its future power demands through contractual arrangements instead of constructing a new plant.

Creating complicated financial instruments requires accurate pricing methods for the derivatives that make up their constituent parts. It is relatively easy to establish the price of a futures contract. When the cost of wheat rises, the price of the futures contract on the commodity increases by the same relative amount. Thus, the relationship is linear. For options, there is no such simple link between the derivative and the underlying asset. For this reason, the work of Scholes, Merton and their deceased colleague Fischer Black has assumed an importance that prompted one economist to describe their endeavors as "the most successful theory not only in finance but in all of economics."

Einstein and Options

The proper valuation of options had perplexed economists for most of this century. Beginning in 1900 with his groundbreaking essay "The Theory of Speculation," Louis Bachelier described a means to price options. Remarkably, one component of the formula that he conceived for this purpose anticipated a model that Albert Einstein later used in his theory of Brownian motion, the random movement of particles through fluids. Bachelier's formula, however, contained financially unrealistic assumptions, such as the existence of negative values for stock prices.

Other academic thinkers, including Nobelist Paul Samuelson, tried to attack the problem. They foundered in the difficult endeavor of calculating a risk premium: a discount from the option price to compensate for the investor's aversion to risk and the uncertain movement of the stock in the market.

The insight shared by Black, Scholes and Merton was that an estimate of a risk premium was not needed, because it is contained in the quoted stock price, a critical input in the option formula. The market causes the price of a riskier stock to trade further below its expected future value than a more staid equity, and that difference serves as a discount for inherent riskiness.

Black and Scholes, with Merton's help, came up with their option-pricing formula by constructing a hypothetical portfolio in which a change of price in a stock was canceled by an offsetting change in the value of options on the stock—a strategy called hedging. Here is a simplified example: A put option would give the owner the right to sell a share of a stock in three months if the stock price is at or below $100. The value of the option might increase by 50 cents when the stock goes down $1 (because the condition under which the option can be used has grown more likely) and decrease by 50 cents when the stock goes up by $1.

To hedge against risks in changes in share price, the investor can buy two options for every share he or she owns; the profit then will counter the loss. Hedging creates a risk-free portfolio, one whose return is the same as that of a treasury bill. As the share price changes over time, the investor must alter the composition of the portfolio—the ratio of the number of shares of stocks to the number of options—to ensure that the holdings remain without risk.

The Black-Scholes formula, in fact, is elicited from a partial differential equation demonstrating that the fair price for an option is the one that would bring a risk-free return within such a hedging portfolio. Variations on the hedging strategy outlined by Black, Scholes and Merton have proved invaluable to financial-center banks and a range of other institutions that can use them to protect portfolios against market vagaries—ensuring against a steep decline in stocks, for instance.

The basic options-pricing methodology can also be extended to create other instruments, some of which bear bizarre names like "cliquets" or "shouts." These colorful financial creatures provide the flexibility to shape the payoffs from the option to a customer's particular risk profile, placing a floor, ceiling or averaging function on interest or exchange rates, for example.

With the right option, investors can bet or hedge on any kind of uncertainty, from the volatility (up-and-down movement) of the market to the odds of catastrophic weather. An exporter can buy a "look-back" currency option to receive the most favorable dollar-yen

exchange rate during a six-month period, rather than being exposed to a sudden change in rates on the date of the contract's expiration.

In the early 1970s Black and Scholes's original paper had difficulty finding a publisher. When it did reach the *Journal of Political Economy* in 1973, its impact on the financial markets was immediate. Within months, their formula was being programmed into calculators. Wall Street loved it, because a trader could solve the equation easily just by punching in a few variables, including stock price, interest rate on treasury bills and the option's expiration date. The only variable that was not readily obtainable was that for "market volatility"—the standard deviation of stock prices from their mean values. This number, however, could be estimated from the ups and downs of past prices. Similarly, if the current option price was known in the markets, a trader could enter that number into a workstation and "back out" a number for volatility, which can be used to judge whether an option is overpriced or underpriced relative to the current price of the stock in the market.

Investors who buy options are basically purchasing volatility—either to speculate on or to protect against market turbulence. The more ups and downs in the market, the more the option is worth. An investor who speculates with a call—an option to buy a stock—can lose only the cost of purchase, called a premium, if the stock fails to reach the price at which the buyer can exercise the right to purchase it. In contrast, if the stock shoots above the exercise price, the potential for profit is unlimited. Similarly, the investor who hedges with options also anticipates rough times ahead and so may buy protection against a drop in the market.

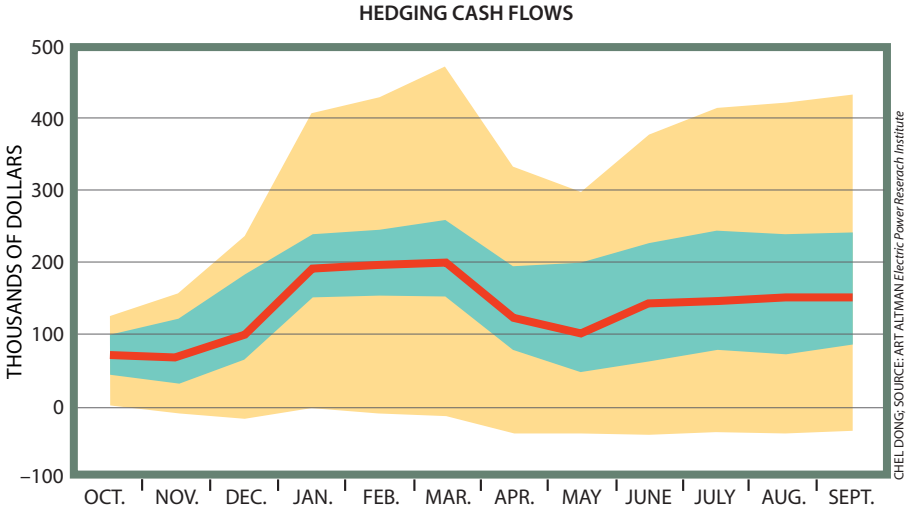
Physicists on Wall Street

Although it can be reduced to operations on a pocket calculator, the mathematics behind the Black-Scholes equation is stochastic calculus, a descendant from the work of Bachelier and Einstein. These equations were by no means the standard fare in most business administration programs. Enter the Wall Street rocket scientists: the former physicists, mathematicians, computer scientists and econometricians who now play an important role at the Wall Street financial behemoths.

Moving from synchrotrons to trading rooms does not always result in such a seamless transition. "Whenever you hire a physicist, you're always hoping that he or she doesn't think of markets as if they were governed by immutable physical laws," notes Charles Smithson, a managing director at CIBC World Markets, an investment bank. "Uranium 238 always decays to uranium 234. But a physicist must remember that markets can go up as well as go down."

Recently some universities have opened "quant schools," programs that educate M.B.A. or other master's students in the higher applied mathematics of finance, the subtleties of Ito's lemma and other cornerstones of stochastic calculus. Or else they may train physicists, engineers and mathematicians before moving on to Wall Street. "Market pressures are directing physicists to get more education to try to understand the motivation and intuition underlying financial problems," says Andrew W. Lo, who heads the track in financial engineering at the Massachusetts Institute of Technology's Sloan School of Management.

As part of their studies, financial engineers in training learn about the progression of mathematical modeling beyond the original work of Black, Scholes and Merton. The basic Black-Scholes formula made unrealistic assumptions about how the market operates. It takes a fixed interest rate as an input, but of



HEDGING by buying and selling contracts called forwards allows a utility's electricity-generating plant to reduce uncertainties in cash flows—revenues minus expenses (*red line*). The variability of cash flows—represented as 80 percent confidence intervals—narrows considerably with forwards (*green area*) as compared with operating without the contracts (*yellow area*). Forwards limit how much cash flows may fall but also restrict potential gains.

BERND AUERS

"QUANT SCHOOL" at the Massachusetts Institute of Technology's Sloan School of Management teaches students the nuances of financial engineering in a classroom built as a trading room.

course interest rates change, and that influences the value of an option—particularly an option on a bond. The formula also assumes that changes in the growth rate of stock prices fall into a normal statistical distribution, a bell curve in which events cluster around the mean. Thus, it fails to take into account extraordinary events such as the 1929 or 1987 stock market crashes. Black, Scholes and Merton—and legions of quants—have spent the ensuing years refining many of the original ideas.

Emanuel Derman, head of the quantitative strategies group at Goldman Sachs, is a physicist-turned-quant whose job over the past 13 years has been to tackle the imperfections of the Black-Scholes equation. Derman, a native of Cape Town, South Africa, received his doctorate from Columbia University in 1973 for a thesis on the weak interaction among subatomic particles. He went on to postdoctoral positions, including study of neutrino scattering at the University of Pennsylvania and charmed quark production at the University of Oxford's department of theoretical physics. In the late 1970s Derman decided to leave academia: "Physics is lonely work. It's a real meritocracy. In physics, you sometimes feel like you're either [Richard] Feynman or you're nobody. I liked physics, but maybe I wasn't as good as I might have been."

So in 1980 he went to Bell Laboratories in New Jersey, where he worked on a computer language tailored for finance. In 1985 Goldman Sachs hired him to develop methods of modeling interest rates. He has worked there since, except for a year spent at Salomon Brothers. At Goldman, he met the recently recruited Fischer Black, and the two began working with another colleague, William W. Toy, on a method of valuing bond options. Derman remembers Black as a bluntly truthful man with punctilious writing habits who wore a Casio Data Bank watch. "Black was less powerful mathematically than he was intuitively," Derman says. "But he always had an idea of what the right answer was."

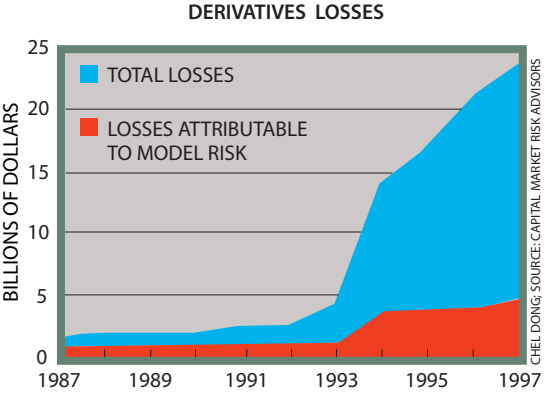
Physics Versus Finance

Much of Derman's recent work on the expected volatility of stock prices continues to refine the original 1973 paper. The Black-Scholes equation was to finance what Newtonian mechanics was to physics, Derman asserts. "Black-Scholes is sort of the foundation on which the field rests. Nobody knows what to do next except extend it." But the field, he fears, may never succeed in producing its own Einstein—or some unified financial theory of everything. Finance differs from physics in that no mathematical model can capture the multitude of ever mutating economic factors that cause major market perturbations—the recent Asian collapse, for instance. "In physics, you're playing against God; in finance, you're playing against people," Derman declares.

Outside the domain of Wall Street, the parallels between physical concepts and finance are sometimes taken more literally by academics. Kirill Ilinski of the University of Birmingham in England has used Feynman's theory of quantum electrodynamics to model market dynamics, while employing these concepts to rederive the Black-Scholes equation. Ilinski replaces an electromagnetic field, which controls the interaction of charged particles, with a so-called arbitrage field that can describe changes in option and stock prices. (Trading that brings the value of the stock and the option portfolio into line is called arbitrage.)

Ilinski's theory shows how quantum electrodynamics can model Black, Scholes and Merton's hedging strategy, in which market dynamics dictate that any gain in a stock will be offset by the decline in value of the option, thereby yielding a risk-free return. Ilinski equates

INCORRECT VALUATIONS for derivatives can cause losses, as expressed in the concept of "model risk." The cumulative dollar figure for model risk from 1987 to 1997 comprises about 20 percent of all publicly disclosed losses.



REPRINTED FROM *THE FINANCIAL POST*

DERIVATIVES DEBACLES have created their own distinctive brand of humor.

it with the absorption of "virtual particles," or photons, that damp the interacting forces between two electrons. He goes on to show how his arbitrage field model elucidates opportunities for profit that were not envisaged by the original Black-Scholes equation.

Ilinski is a member of the nascent field of econophysics, which held its first conference last July in Budapest. Nevertheless, literal parallelism between physics and finance has gained few adherents. "It doesn't meet the very simple rule of demarcation between science and hogwash," notes Nassim Taleb, a veteran derivatives trader and a senior adviser to Paribas, the French investment bank. Ilinski recognizes the controversial nature of his labors. "Some people accept my work, and some people say I'm mad. So there's a discrepancy of opinion," he says wryly.

Whether invoking Richard Feynman or Fischer Black, the use of mathematical models to value and hedge securities is an exercise in estimation. The term "model risk" describes how different models can produce widely varying prices for a derivative and how these prices create large losses when they differ from the ones at which a financial instrument can be bought or sold in the market.

Model risk comes in many forms. A model's complexity can lead to erroneous valuations for derivatives. So can inaccurate assumptions underlying the model—failing to take into account the volatility of interest rates during an exchange-rate crisis, for instance. Many models do not cope well with sudden alterations in the relation among market variables, such as a change in the normal trading range between the U.S. dollar and the Indonesian rupiah. "The model or the way you're using it just doesn't capture what's going on anymore," says Tanya Styblo Beder, a principal in Capital Market Risk Advisors, a New York City firm that evaluates the integrity of models. "Things change. It's as if you're driving down a very steep mountain road, and you thought you were gliding on a bicycle, and you find you're in a tractor-trailer with no brakes."

Custom-tailored products of financial engineering are not traded on public exchanges and so rely on valuations produced by models, sometimes making it difficult to compare the models' pricing to other instruments in the marketplace. When it comes time to sell, the market may offer a price that differs significantly from a model's estimate. In some cases, a trader might capitalize on supposed mispricings in another trader's model to sell an overvalued option, a practice known as model arbitrage.

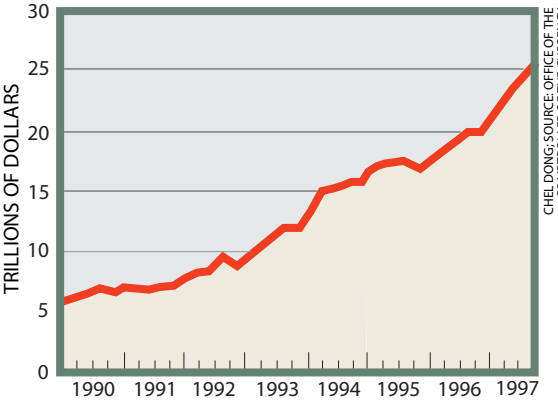
"There's a danger of accepting models without carefully questioning them," says Joseph A. Langsam, a former mathematician who develops and tests models for fixed-income securities at Morgan Stanley. Morgan Stanley and other firms adopt various means of testing, such as determining how well their models value derivatives for which there is a known price.

Problems related to modeling have accounted for about 20 percent of the $23.77 billion in derivatives losses that have occurred during the past decade, according to Capital Market Risk Advisors. Last year, however, model risk comprised nearly 40 percent of the $2.65 billion in money lost. The tally for 1997 included National Westminster Bank, with $123 million in losses, and Union Bank of Switzerland, with a $240-million hit.

A conference in February sponsored by *Derivatives Strategy,* an industry trade magazine, held a roundtable discussion called "First Kill All the Models." Some of the participants questioned whether the most sophisticated mathematical models can match traders' skill and gut intuition about market dynamics. "As models become more complicated, people will use them, and they're dangerous in that regard, because they'll use them in ways that are deleterious to their economic health," said Stanley R. Jonas, who heads the derivatives trading department for Societé Generale/FIMAT in New York City. An unpublished study by Jens Carsten Jackwerth of the London Business School and Mark E. Rubinstein of the University of California at Berkeley has shown that traders' own rules of thumb about inferring future stock index volatility did better than many of the major modeling methods.

One modeler at the session—Derman of Goldman Sachs—defended his craft. "To paraphrase Mao in the sixties: Let 1,000 models bloom," he proclaimed. He compared models to gedanken (thought) experiments, which are unempirical but which help physicists contemplate the world more clearly: "Einstein would think about what it was

NOTIONAL AMOUNT OF DERIVATIVES HELD BY U.S. COMMERCIAL BANKS



CHEL DONG; SOURCE: OFFICE OF THE COMPTROLLER OF THE CURRENCY

GROWTH in derivatives usage by insured U.S. commercial banks continues. The notional value represents the face value of the underlying asset, index or rate from which options, forwards, futures and swaps are derived.

like to sit on the edge of a wave moving at the speed of light and what he would see. And I think we're doing something like that. We are sort of investigating imaginary worlds and trying to get some value out of them and see which one best approximates our own." Derman acknowledged that every model is imperfect: "You need to think about how to account for the mismatch between models and the real world."

Financial Hydrogen Bombs

The image of derivatives has been sullied by much publicized financial debacles, which include the bankruptcies of Barings Bank and Orange County, California, and huge losses by Procter & Gamble and Gibson Greetings. Investment banker Felix Rohatyn has been quoted as warning about the perils of twentysomething computer whizzes concocting "financial hydrogen bombs." Some businesses and local governments have excluded derivatives from their portfolios altogether; fears have even emerged about a meltdown of the financial system.

The creators of these newfangled instruments place the losses in broader perspective. The notional, or face, value of all stocks, bonds, currencies and other assets on which options, futures, forwards and swap contracts are derived totaled $56 trillion in 1995, according to the Bank for International Settlements. The market value of the outstanding derivatives contracts themselves represents only a few percentage points of the overall figure but an amount that may still total a few trillion dollars. In contrast, known derivatives losses between 1987 and 1997 totaled only $23.8 billion. More mundane investments can also hurt investors. When interest rates shot up in 1994, the treasury bond markets lost $230 billion.

Derivatives make the news because, like an airplane crash, their losses can prove sudden and dramatic. The contracts can involve enormous leverage. A derivatives investor may put up only a fraction of the value of an underlying asset, such as a stock or a bond. A small percentage change in the value of the asset can produce a large percentage gain or loss in the value of the derivative.

To manage the risks of owning derivatives and other securities, financial houses take refuge in yet other mathematical models. Much of this work is rooted in portfolio theory, a statistical measurement and optimization methodology for which Harry M. Markowitz received the Nobel Prize in 1990. Markowitz elucidated how investors could minimize risk for a given level of return by diversifying into a range of assets that do not all perform the same way as the market changes.

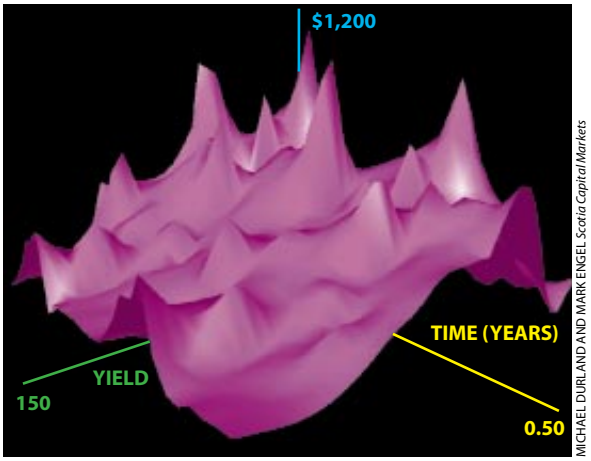
One hand-me-down from Markowitz is called value at risk. It sets forth a set of techniques that elicits a single worst-case number for investment losses. Value at risk calculates the probability of the maximum losses for every existing portfolio, from currency to derivatives. It then elicits a value at risk for the company's overall financial exposure: the worst hit that can be expected within the next 30 days with a given statistical confidence interval might amount to $85 million. An analysis of the portfolios shows where risks are concentrated.

Philippe Jorion, a professor of finance at the University of California at Irvine, has performed a case study that shows how value-at-risk measures could raise warning flags to even unsophisticated investors. Members of the school boards in Orange County that invested in the county fund that lost $1.7 billion might have reacted differently if they knew that there existed a 5 percent chance of a billion-dollar-plus loss.

Like other modeling techniques, value at risk has bred skepticism about how well it predicts ups and downs in the real world. The most widely used measurement techniques rely heavily on historical market data that fail to capture the magnitude of rare but extreme events. "If you take the last year's worth of data, you may see a portfolio vary by only 10 percent. Then, if you move a month ahead, things may change by 100 percent," comments Ron S. Dembo, president of Algorithmics, a Toronto-based risk-management software company. Algorithmics and other firms go beyond the simplest value-at-risk methods by providing banks with software that can "stress-test" a portfolio by simulating the ramifications of large market swings.

One modeling technique may beget another, and debates over their intrinsic worth will surely continue. But the ability to put a price on uncertainty, the essence of financial engineering, has already proved worthwhile in other business settings as well as in government policymaking and domestic finance. Options theory can aid in steering capital investments. A conventional investment analysis might suggest that it is better for a utility to budget for a large coal-fired plant that can provide capacity for 10 to 15 years of growth. But that approach would sacrifice the alternative of building a series of small oil-fired generators, a better choice if demand grows more slowly than expected. Option-pricing techniques can place a value on the flexibility provided by the slow-growth path.

The Black-Scholes model has also been used to quantify the benefits that accrue to a developing nation from providing workers with a general education rather than targeted training in specific skills. It reveals that the value of being able to change labor skills quickly as the economy shifts can exceed the extra cost of supplying a broad-based education. Option pricing can even be used to assess the flexibility of choosing an "out-of-plan" physician for managed health care. "The implications for this aren't just in the direct financial markets but in being able to use this technology for how we organize nonfinancial firms and how people organize their financial lives in general," says Nobelist Merton. Placing a value on the vagaries of the future may help realize the vision of another Nobel laureate: Kenneth J. Arrow of Stanford University imagined a security for every condition in the world—and any risk, from bankruptcy to a rained-out picnic, could be shifted to someone else. SA



MICHAEL DURLAND AND MARK ENGEL *Scotia Capital Markets*

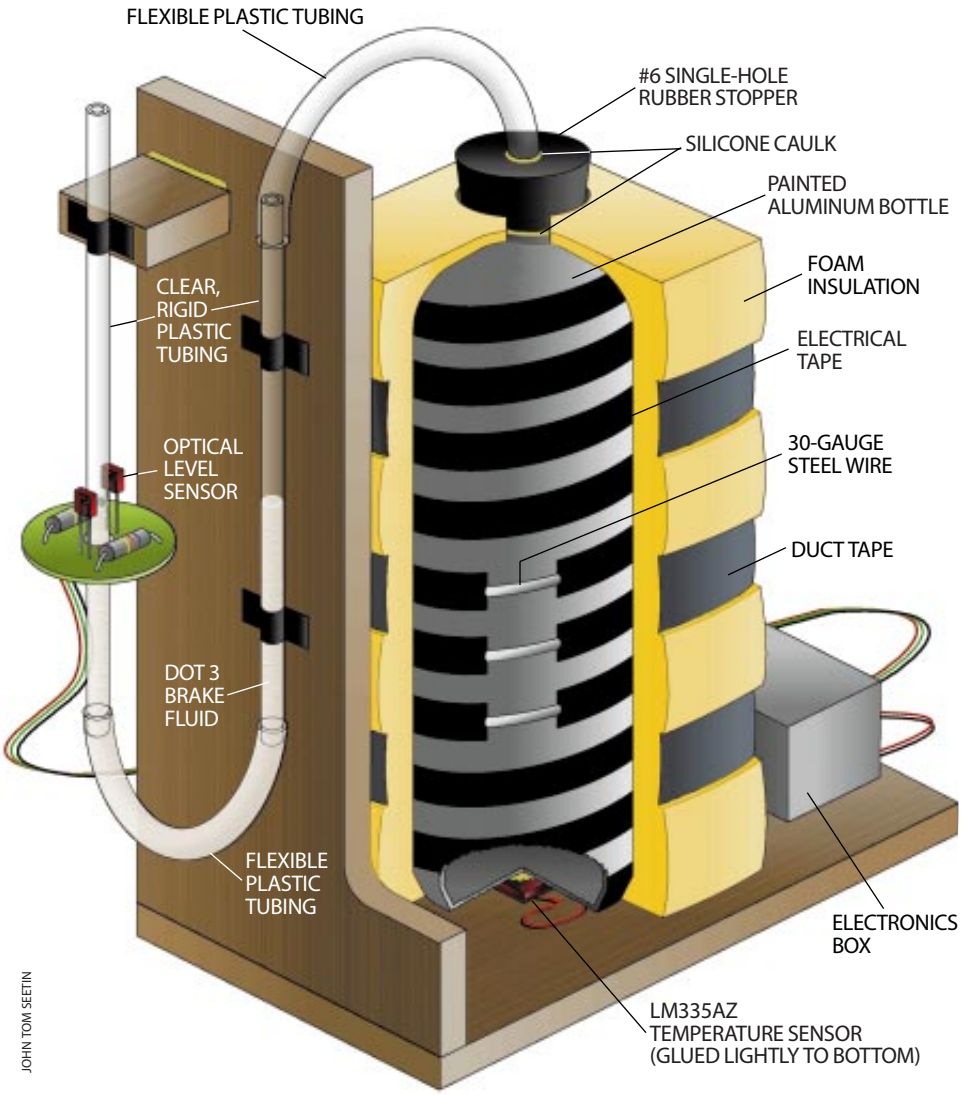
"STRESS TESTING" of interest-rate derivatives at the Bank of Nova Scotia shows how market fluctuations affect the characteristics of a portfolio.

by Shawn Carlson

Sensing Subtle Tsunamis

As I write this column, I am enjoying my regular Sunday morning bagel at a small café overlooking the ocean. It's raining, and storms at sea have whipped up unusually large waves. These ribbons of energy march thousands of miles in lockstep, finally breaking on Pacific beaches with spectacular effect. Watching these watery monsters roll up onto the nearby sand reminds me of another kind of wave passing by. The world's greatest ocean is the atmosphere, and it, too, contains extraordinarily powerful waves. Like their oceangoing counterparts, atmospheric waves are normally generated by energetic storms. But they can also arise and spread, like ripples on a quiet pond, when a meteor or volcanic explosion violently shocks the air.

Yet even the largest atmospheric tsunamis are quite difficult to detect. The pressure excursions that betray their passage are typically just a few millibars (thousandths of one atmosphere), and these tiny undulations often take tens of minutes, or even longer, to go by. Instruments that can monitor such subtle signals are called microbarographs, and professional units can cost thousands of dollars. But, thanks to Paul Neher, a gifted amateur scientist from Las Cruces, N.M., anyone can now observe these ephemeral waves for about $50.



JOHN TOM SEETIN

MICROBAROGRAPH
is easily constructed from an aluminum bottle, steel wire and plastic tubing.

Neher uses a manometer (in this case, a U-shaped tube and sensor) [*see illustration on this page*] to balance the barometric pressure against the air pressure trapped inside an aluminum bottle that is kept a bit warmer than the surrounding air. Because the pressure of an isolated gas varies with temperature, any shift in external pressure can be matched by changing the temperature of the air inside the bottle. Neher's instrument senses external pressure fluctuations by monitoring the height of a liquid in the manometer. A rise in external pressure presses the liquid down into the manometer and triggers a heating coil that warms the bottle. If the pressure drops, the level of the liquid rises, and the heater is kept off, thus allowing the bottle to cool a little to compensate. Monitoring the temperature changes with the circuit shown on the opposite page reveals minuscule shifts in air pressure.

For the aluminum container, Neher completely drained a tire inflator bottle of 32-ounce (about one-liter) size and then replaced the crimped-on cap with a single-hole rubber stopper. He fashioned his manometer from a piece of glass tubing, which he bent after heating with a propane torch. But you can link two clear, rigid plastic tubes with a short, flexible plastic hose. You'll need to fill this assembly about one-third full with a liquid that has a low viscosity and does not evaporate. Neher uses clear brake fluid in his device.

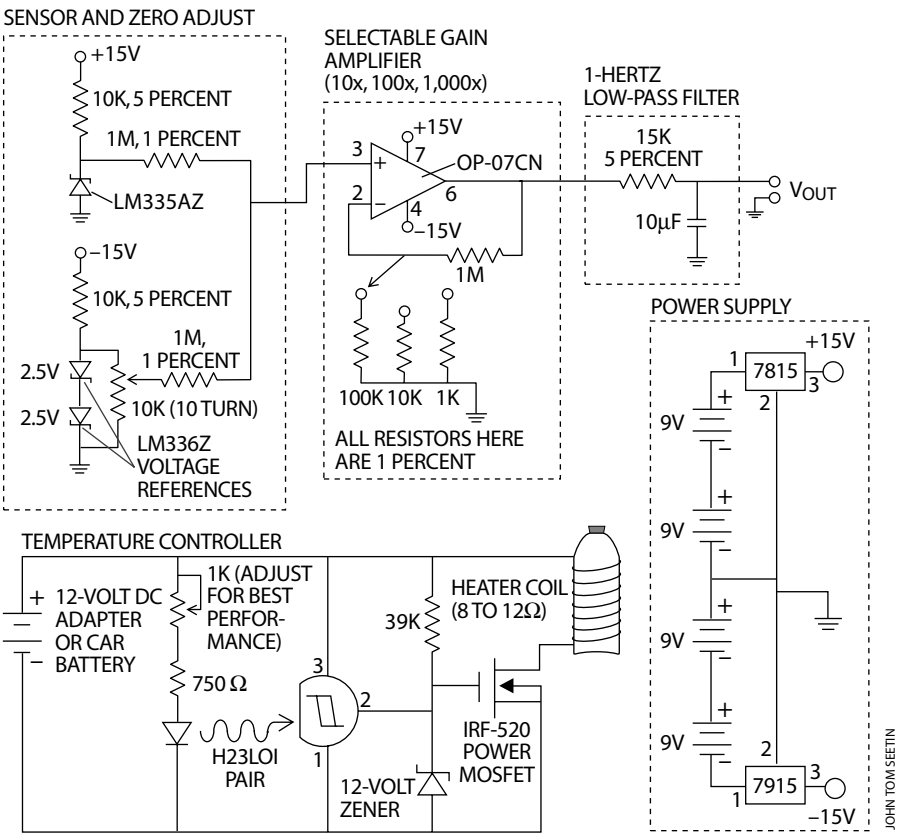
The instrument senses changes in the fluid level by using the transparent liquid to focus the light from an infrared light-emitting diode (LED) onto a phototransistor. When the liquid drops below the set point, the defocused light becomes too diffuse to detect. This change causes the circuit attached to the phototransistor to send an electric current through the heater. Neher employed

"beading wire," which he obtained from a craft store, for the heating filament. Hobbyists use this 30-gauge (1/4-millimeter) steel wire to make necklaces, but it has a resistance of about one ohm per foot (about three ohms per meter), which is ideal for this application.

After electrically insulating the bottle with a layer of enamel paint, wrap 10 evenly spaced turns along the length of the bottle and secure them with electrical tape. Surround the wrapped bottle with a few inches (about 10 centimeters) of an insulating material such as foam rubber or a spray insulation (which you can purchase at a hardware store). When operating, the circuit heats the bottle slightly every 10 seconds or so, replacing the tiny amount of heat that leaches through the insulation, thereby keeping the level of the fluid stable. The LM335AZ chip is a sensitive solid-state thermometer that varies its output voltage by 10 millivolts for each degree Celsius change in temperature. This marvel can measure temperatures to about 0.01 degree C, which corresponds to an ultimate pressure resolution approaching 20 microbars—that's a scant 20 millionths of one atmosphere.

To begin your observations, disconnect the tube that links the manometer to the aluminum bottle, then heat the bottle about 10 degrees C (18 degrees Fahrenheit) above room temperature by blocking the light from reaching the phototransistor. Reconnect the tube and allow the circuit to stabilize the bottle at this elevated temperature. It's easy to see if your instrument is working: just lift it. It should register the 100-microbar drop in pressure that results when you raise it about one yard (one meter).

You will want to calibrate the microbarograph over a larger range than you can easily generate by shifting its height. The solution is quite simple: move the LED-phototransistor pair up or down a bit on the manometer column. The circuit will then adjust the temperature inside the bottle to raise or lower the liquid between the LED and phototransistor to match. This manipulation causes the fluid level to become uneven, with the weight of the unbalanced liquid being supported by the pressure difference between the atmosphere and the air inside the bottle. The specific gravity of DOT 3 brake fluid, which is widely available in the U.S., is 1.05. For this value, each inch (centimeter) difference in the level of the liquid corresponds to a pressure difference of 2.62 millibars (1.03 millibars).



ELECTRONIC CIRCUITS FOR THIS PROJECT
can be built with components from Mouser Electronics (817-483-6848).

You can read the output of the Neher microbarograph with a digital voltmeter or, better yet, record the data continuously using your computer and an analog-to-digital converter. These versatile devices, once too pricey for amateur budgets, are now quite affordable. Both Macintosh and PC aficionados should check out the Serial Box Interface, which is available for $99 from Vernier Software (8565 Southwest Beaverton Highway, Portland, OR 97225; www.vernier.com, 503-297-5317). PC users can also consider buying a similar unit for $100 from Radio Shack (part number 11910486). Either unit will turn your home computer into a sophisticated data-collection station. Then with your own sensitive microbarograph, you will be able to detect the endless train of subtle atmospheric tsunamis that is constantly rolling by. SA

For more information about this and other amateur science projects, check out the Forum section of the Society for Amateur Scientists's Web site at http://web2.thesphere.com/SAS/. You may also write the society at 4735 Clairemont Square, Suite 179, San Diego, CA 92117, or call them at (619) 239-8807.

Cautionary note: We urge readers building the centrifuge described in the January Amateur Scientist column not to neglect to use the protective housing included in that project. One reader ran his newly built centrifuge at high speed without this housing only to discover that the rotor was imbalanced enough to shatter, perilously sending pieces of plastic in all directions (see http://web2.thesphere.com/SAS/WebX.cgi).

by Ian Stewart

Cementing Relationships

The prestigious journal *Nature* manages to combine high-powered research papers with eclectic science writing. One of its regular columns is Art and Science, whose name pretty much speaks for itself. In the December 11, 1997, issue, art historian Martin Kemp describes the remarkable landscapes of a London artist named Jonathan Callan. Unlike conventional landscape art, Callan's works are sculptures, not paintings. And his landscapes are unlike anything seen on earth or indeed on any known world. They are three-dimensional forms created by pouring cement onto a perforated board.

Kemp, a professor in the University of Oxford's art history department, notes a relationship between Callan's sculptures and recent work in the field of complexity theory. Certain general principles seem to govern Callan's fantastic landscapes—for instance, the highest peaks of cement occur in the regions farthest from the holes. In a letter to the editor in a later issue of *Nature* (January 29, 1998), Adrian Webster, an astronomer at the Royal Observatory in Edinburgh, points out that the curious geometry of Callan's landscapes can be understood using a more classical branch of mathematics, the theory of Voronoi cells. He also explains how Voronoi cells illustrate one of the major recent discoveries of astronomy, the foamlike distribution of matter in the universe. If ever there was an example of the unity of mathematics, art and science, this has to be it.

Since the very first cave paintings, artists have relied on processes from physics and chemistry to create their masterpieces. In ancient Greece and Rome, sculptors had to understand how stones fractured and how molten bronze flowed into a cast. Renaissance painters studied the properties of pigments. The traditional artist's technique has been to control these physical processes, using them to shape sculptures and paintings in desired ways. Callan is one of a smaller band of modern artists who relinquish that control. They allow the physical and chemical processes of their media to determine the main features of their artworks.

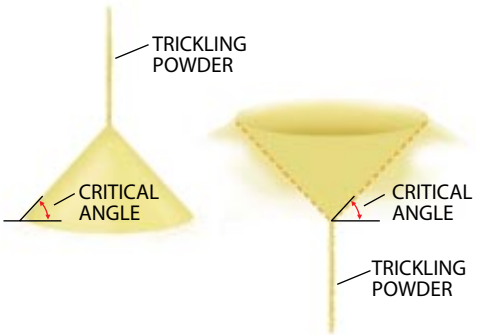
Callan begins each sculpture by drilling a random pattern of holes into a horizontal board. He then sieves cement powder evenly over the surface. Some cement falls through the holes, and some piles up in the areas between them. The sculpture hardens by absorbing moisture from the air. The final result resembles a moonscape, with jagged peaks surrounding steep craters. Callan compares the sculpture to an earthly mountain range: "A geography that seems both eminently natural and highly artificial—the Alps brand new."

A more apt comparison might liken Callan's artworks to a collection of sandpiles. Civil engineers have long been familiar with how granular materials—such as sand, soil or cement powder—pile up. The simplest and most important feature is the existence of a "critical angle." Depending on the nature of the granular material, there is a steepest slope that it can sustain without collapsing. This slope runs at a constant angle with the ground—the critical angle. If you keep piling sand higher and higher—say, by pouring it in a thin stream—the slope of the sandpile will steepen until it reaches the critical angle. Any extra sand will then trickle down the pile, causing either a tiny avalanche or a big one, to restore the constant slope. The resulting steady-state shape, in this simplest model, is a cone whose sides slope at exactly the critical angle.

Complexity theorists study the process by which the slope attains this shape and the nature of the avalanches, big or small, that accompany its growth. Danish physicist Per Bak coined the term "self-organized criticality" for such processes, and he has suggested that they model many important features of the natural world, especially evolution (where the avalanches involve not grains of sand but entire species, and the piles are in an imaginary space of potential organisms).

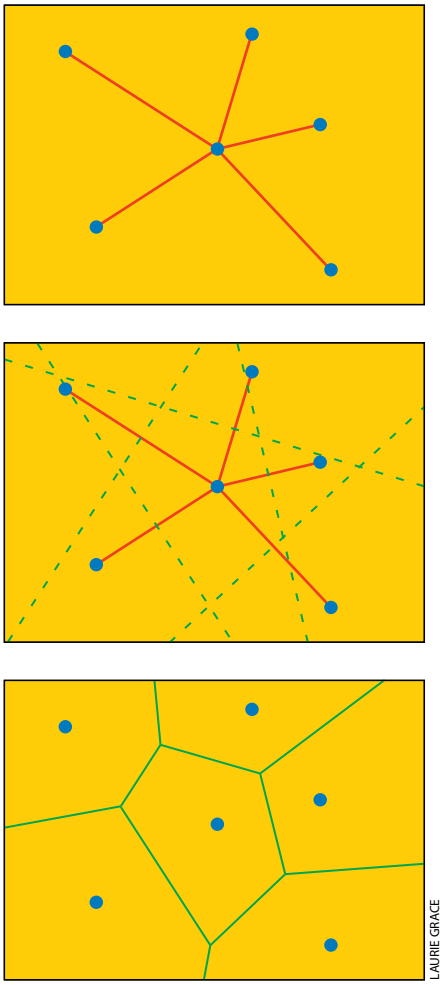
In Callan's artworks, the structure of

COURTESY OF HALES GALLERY



LAURIE GRACE

JAGGED CRATERS *pock the surface of one of Jonathan Callan's cement sculptures* (far left). *The sides of each crater slope at the same angle as the sides of a conical pile of cement powder* (left).



LAURIE GRACE

***VORONOI CELLS** can be sketched by drawing the perpendicular bisectors of the lines between the holes.*

cement powder around each hole is an inversion of the civil engineer's conical sandpile. Consider a horizontal board with just one hole. Away from the hole, cement rises in every direction at the critical angle, creating a conical depression whose tip points downward and rests at the center of the hole [*see illustration on page 100*]. These inverted cones are the craters that form Callan's striking landscapes.

But what of the geometry when there are several holes? The key point now is that any cascading cement powder that trickles through the board will fall out through the hole that is nearest to its initial point of impact. It is therefore possible to predict where the boundaries between the conical craters will occur. Divide the board into regions surrounding the holes, in such a manner that each region consists of those points that are closer to the chosen hole than they are to any other hole. The region is the hole's "sphere of influence," so to speak, except that it is not a sphere but a polygon. Provided the board is horizontal, the boundaries between these regions are directly underneath the common boundaries of adjacent craters.

A way to sketch one of these regions is to choose any hole and draw lines from its center to the centers of all the other holes [*see illustration above*]. Cut each line in half and from that point draw another line at right angles to it (that is, draw the perpendicular bisector). The result will be a crisscrossing network of bisectors. Find the smallest convex region that is bounded by segments of this network and contains the chosen hole. This region is known as a Voronoi cell. Each hole is surrounded by a unique Voronoi cell, and all the cells together tile the plane. Georgii F. Voronoi (1868–1908) was a Russian mathematician who worked on number theory and multidimensional tilings. Voro-

FEEDBACK

The column on the Double Bubble Conjecture [January] has given many readers more toil and trouble than they deserve. Several pointed out that one of the bubbles in the accompanying photographs did not have the shape of a minimal-surface catenoid, as stated. A true catenoid of this type (*below*) is stubbier and has a more tightly pinched waist.

Why the difference between the ideal shape and the photograph? In fact, the shape of a bubble that forms between two parallel rings depends on how far apart the rings are, compared with their diameters. If the separation increases beyond a critical value—as happened apparently in the photograph—the bubble elongates and then collapses, forming two disconnected disks, each spanning one ring. —*I.S.*

BRYAN CHRISTIE

noi cells go by a few other names—Dirichlet domains and Wigner-Seitz cells, for example—because they have been rediscovered in many contexts.

Callan's craters, then, rise in inverted cones at the same critical angle and meet above the edges of the Voronoi cells defined by his system of holes. One comfortable consequence of this geometry is that when two slopes meet, they do so at the same height above the board—there is no sharp discontinuity. Another feature, less obvious, can also be deduced: the shape of the ridge where one crater merges into its neighbor. In the abstract, what we have are two inverted cones rising at identical angles. They meet above the perpendicular bisector of the line that joins their vertices—above the Voronoi boundary. Consequently, their intersection lies in the vertical plane through that boundary line. What curve do you get if you cut a cone with a vertical plane? The ancient Greeks knew the answer: a hyperbola. This fact helps to explain the rather jagged nature of Callan's landscapes.

What of the connection with astronomy? Instead of holes in a plane, imagine points in three-dimensional space. In the plane, the perpendicular bisector of a pair of points is a line, but in space it is a plane. Draw these bisecting planes for the lines between a given point and all the other points. Let the Voronoi cell of the given point be the smallest convex region that surrounds it and is bounded by parts of these planes. Now the Voronoi cell is a polyhedron. Astronomers have recently discovered that the large-scale distribution of matter in the universe resembles a network of such polyhedra. Most galactic clusters seem to be located on the boundaries of neighboring Voronoi cells. This pattern has been called the Voronoi foam model of the universe because it looks somewhat like a giant bubble bath.

There is an analogy—imperfect, but still illuminating—with the distribution of cement powder in Callan's landscapes. In his sculptures, the cement piles up highest along the Voronoi boundaries. The analogous property in three-dimensional space would be that as the universe expands, matter concentrates along the same boundaries. So this one simple idea encapsulates some arresting art, elegant mathematics and deep physics about the structure of the universe. SA

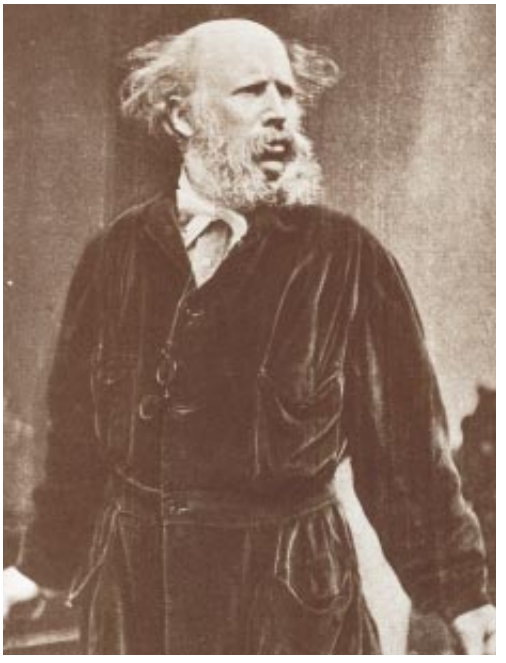
ON THE ORIGIN OF BODY LANGUAGE

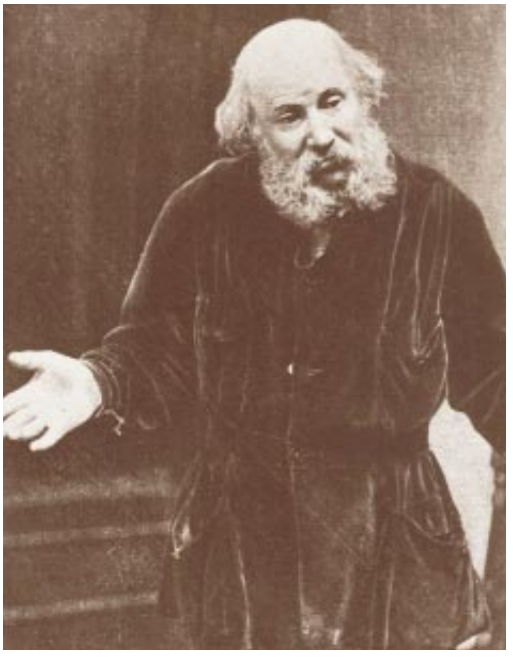
Review by Mark Ridley

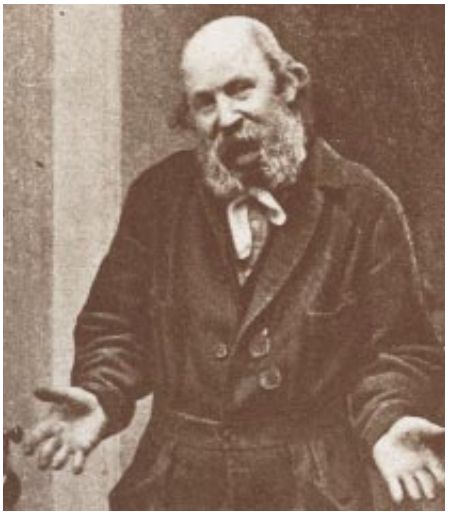
The Expression of the Emotions in Man and Animals
BY CHARLES DARWIN
A new edition, with introduction, afterword and commentaries by Paul Ekman
Oxford University Press, New York, 1998 ($30)

The emotions are a hot topic of the 1990s. They are the most conspicuous difference between the software of our brains and that of computers, creating a puzzle for artificial-intelligence theorists of what a computer would gain if it had emotions as well as a rational calculating ability. The emotional centers of the brain can be almost missing in some brain-damaged patients, creating a puzzle for neuropsychiatrists of what is wrong with these superficially normal people. And social relationships are lubricated, glued together and dissolved by the emotions, creating a puzzle for many of us of how to deploy them efficiently. Hence the stream of popular scientific books, including Antonio R. Damasio's *Descartes' Error,* Daniel Goleman's *Emotional Intelligence* and Joseph LeDoux's *The Emotional Brain,* as well as less scientific features in the glossy magazines and opinions in the talk shows. What does Charles Darwin's 1872 classic have to offer here?

One answer is that emotions are expressed not verbally but in body (and particularly facial) language. You can tell someone is angry, for example, from—in Darwin's words—"the body being held erect," "the brows being heavily contracted" and "the firmly compressed mouth, the distended nostrils, and flashing eyes." This much was known before Darwin, although his description is more complete and accurate than any of his predecessors. Indeed, it has not been replaced since. The editor of this new edition, Paul Ekman, is professor of psychology at the University of California at San Francisco and an expert on emotional expression. Ekman interpolates occasional comments within the text, usually to explain how Dar-





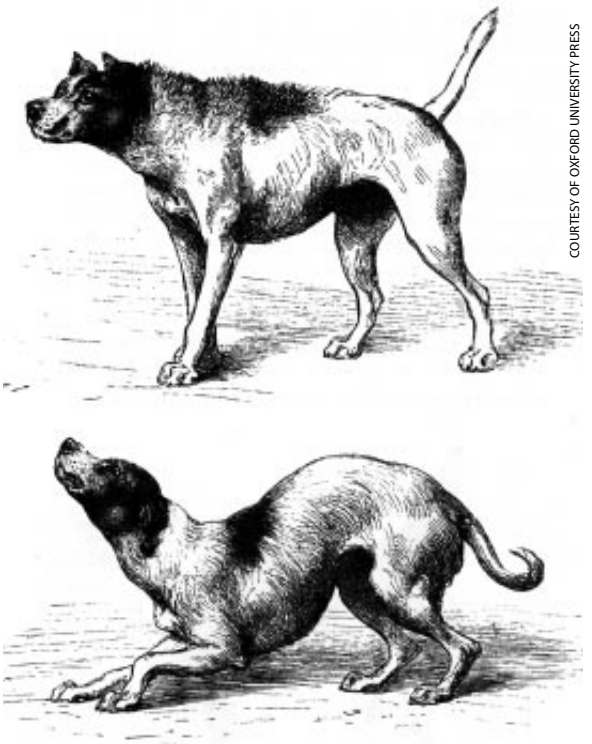


COURTESY OF OXFORD UNIVERSITY PRESS

ANGER AND HELPLESSNESS exemplify two of the principles that Charles Darwin theorized govern emotional expression. In the difficult task of obtaining images to illustrate these principles, Darwin was assisted by the photographer Oscar Rejlander, who went so far as to act out various emotions himself (bearded man). *He even agreed to trim his normally bushy mustache so that his expression would be more clearly revealed.*

win's ideas have held up after later work. Darwin's descriptions, except for some details of facial muscular contractions, get high marks.

Darwin, however, excelled as a theorizer, and *The Expression of the Emotions* is much more than a body-language dictionary. His main interest was the relation between each emotion and the form of its expression. Why do we express anger by an erect posture and direct stare, rather than by, for example, shrugging our shoulders? His answer is contained in three principles governing emotional expression. Some emotions are expressed in a way that is posturally appropriate for the related behavior. If you are angry, for instance, then you may be about to hit someone, and the form of the expression is a postural precursor of doing so: before hitting someone, it makes sense to look at them with some concentration, to tense your muscles, to raise your arm. Darwin called these expressions "serviceable associated habits," and they are the easiest to understand.



COURTESY OF OXFORD UNIVERSITY PRESS

DOGS, LIKE HUMANS, have their own antithetical expressions of hostility and humility.

His second principle, antithesis, is a more remarkable insight. Some emotions appear to be expressed by a posture opposite to that of the opposite emotion. As an example, Darwin used a famous pair of pictures of a dog. One shows a hostile, threatening dog, expressing its emotion by means of the serviceable associated habits. The other picture is of a subordinate, "humble and affectionate" dog, and its posture—relaxed muscles, body, tail and ears all lowered—is the opposite. Darwin interpreted shrugging the shoulders in the same way. It expresses, he says, helplessness or impotence: "It accompanies speeches such as, 'It was not my fault'; 'It is impossible for me to grant this favour'; 'He must follow his own course, I cannot stop him.' Shrugging the shoulders likewise expresses patience, or the absence of any intention to resist." In a set of four photographs, he shows how shoulder shrugging is posturally antithetical to someone who is in control or command.

Darwin's third principle has never found many supporters (not even Darwin himself). He called it the direct action of the nervous system and used it to explain, for instance, trembling as an expression of fear. Saying that trembling is caused by the direct action of the nervous system may not be all that enlightening, but I don't know that anyone has yet thought up anything better.

One problem that readers have had with Darwin's book is that he does not seem to have any compelling theory of why expressions should be antithetically expressed. It is easy to see where the advantage is in the posture of a serviceable associated habit, but what advantage is there in an antithetical posture? Darwin is uncharacteristically incurious, although one possibility—communication—should have been obvious. If you are feeling subordinate, the last thing you want is to be mistaken for someone who is feeling dominant—and risk getting mixed up in fights you prefer to avoid. The way to minimize the chance of that mistake is to make yourself look as unlike a dominant individual as possible. Antithesis therefore can make sense. But Darwin ignored communication. In Ekman's words, "He avoided any discussion of how expressions communicate information until the very end of the book. In the last chapter he wrote one paragraph acknowledging that expressions have communicative value, but did not mention that this was relevant to their evolution."

The omission is a mystery and has never been satisfactorily explained. Ekman does have some interesting thoughts about it, inspired by historian Richard Burkhardt, and they take us back to the origin of Darwin's interest in the emotions. In 1840 the 31-year-old Darwin had only recently formulated his theory of evolution. He was reading widely in relevant, and less relevant, material and picked up Charles Bell's *Anatomy and Philosophy of Expression*. Bell supposed that the frowning muscle (or corrugator) was unique to humans, which would suggest a separate creation of the human species. Darwin knew this was a mistake. (Ekman tells how Darwin wrote in the margin of his copy of Bell, next to a passage about the corrugator, "I have seen well developed in monkey.... I suspect he never dissected monkey.")

The general idea that emotional expression is unique to humans was older still. Social and political theorists of the 18th century, including Adam Smith, supposed that the Creator had installed blushing in humans to check antisocial behavior and make social life possible. (Blushing, by the way, is the subject of the most fascinating chapter in Darwin's book, although it can be awkward to share the fascination with others.) A connection between human emotional expression and social life was a standard, if minor, creationist theme in Darwin's time. Darwin told Alfred Russel Wallace that he began his emotional project to "upset Sir C. Bell" and to demonstrate the continuity of emotional expression between humans and many other species. This purpose can be detected in the book. But as the project grew it took on new dimensions, and when the sixtysomething Darwin finally wrote his work up (he always had a leisurely attitude to publication), it was structured around his three principles of

Early Scientific Photography

COURTESY OF OXFORD UNIVERSITY PRESS

One of the first scientific books to use photographs, Darwin's *The Expression of the Emotions* included these disturbing pictures made around 1860 by physician Guillaume-Benjamin Duchenne and photographer Adrien Tournachon. Duchenne used electric current to stimulate his subjects to produce a variety of facial expressions. Here the induced contraction of certain neck muscles produces an expression of intense fear. By applying a constant flow of electricity, he was able to hold the expressions long enough to accommodate the lengthy exposure times needed by early photographic technology. His subject suffered from a neuromuscular abnormality, which made it possible to activate individual groups of his facial muscles without causing involuntary response among others.

emotional expression and not the question of evolutionary continuity. Nevertheless, he may have avoided discussing communication because it was a creationist hot button that he did not want to push. He could make his argument subtly more persuasive by avoiding it.

Ekman's edition is no mere reprint plus introduction. The text itself is not a reprint, because Ekman has collated the previous editions and Darwin's manuscripts and corrected some errors. He has also added a particularly good afterword, in which he describes the 20th-century debate about whether emotional expressions are a human universal. Darwin had argued they are. In the middle of this century, the opposite view—promoted by Margaret Mead, among others—prevailed. Ekman himself then reinstated Darwin's view. Ekman globe-trotted, including a visit to a region of New Guinea where the people had practically no contact with Westerners, and showed people photographs of human faces and asked them what emotions they illustrated. The answers everywhere were much the same. Mead seems to have made some tart but insubstantial criticisms of Ekman's work, and the whole story is interesting to read for its sociology as well as its science.

The edition also has six new appendices and an extensive apparatus of commentaries. Special attention has been paid to the illustrations. *The Expression of the Emotions* was one of the first scientific books to use photographs, and certain problems arose in the publication. Some photographs that Darwin discussed in the text were omitted; some were inconveniently placed; some were reversed (a nontrivial matter, as Ekman points out, because we now have evidence for asymmetric facial expression in some emotions, with stronger expression on the left, perhaps related to the emotional dominance of the right brain hemisphere). All these problems have been fixed, as well as they can be. I compared the pictures with my copy of the first edition; most of them are clearer in Ekman's, with a few exceptions.

Bibliophiles tend toward a gloomy view of the progress of human civilization. The most important indicator—book-production quality—seems to show such remorseless decline. Print size shrinks (they will assure you), margins narrow, bindings become frail, and paper grows gray and thin. Ekman's edition, however, is well printed, well bound and is the best production of Darwin's book ever published.

The Expression of the Emotions in Man and Animals is one of Darwin's most readable works. It is alive with anecdotes, literary quotations and his own observations of his friends and children. Artificial-intelligence nerds, neuropsychiatric white-coats and magazine psychobabblers all have some way to go in understanding the emotions, and there will be no better inspiration for them (and the rest of us) than the ideas of one of the master intellects of all time, in this smart new edition.

MARK RIDLEY works in the department of zoology at the University of Oxford. He is the author of the standard college text Evolution *(Blackwell Science, 1996).*

IS VICTORY IN VIEW?

Review by Robert A. Weinberg

Curing Cancer: Solving One of the Greatest Medical Mysteries of Our Time
BY MICHAEL WALDHOLZ
Simon & Schuster, New York, 1997 ($24)

As one of the large horde of researchers trying to figure out the molecular origins of cancer, I am often asked when my colleagues and I are going to come up with "the cure." The questioners usually show polite interest in our research into cancer's roots, but, in the end, they are really interested only in the bottom line: When will the War on Cancer be won? And why hasn't it been won already?

The American public has been handsomely supporting research into cancer's origins for a quarter of a century; the time has arrived for some return on investment. The arcane descriptions of cancer genes and cell biology that my friends and I provide offer little comfort to someone whose mother is dying of a breast or colon carcinoma.

This understandable fixation on improving cancer treatment must have motivated the title of Michael Waldholz's new book *Curing Cancer: Solving One of the Greatest Medical Mysteries of Our Time*. Surely the book's contents didn't inspire his title, for it deals only tangentially with curing cancer. Waldholz, an editor at the *Wall Street Journal,* focuses instead on another aspect of current cancer research—the search for the genes that cause inborn susceptibility to various types of tumors. The connection between these familial cancer genes and new cancer cures is tenuous at best.

During the 1980s, the sea changes in understanding cancer came from uncovering genes that are inherited in pristine form from parents, only to suffer damage during a person's lifetime in one or another of their tissues. The resulting "sporadic" cancers, which comprise 95 percent of the tumors seen in the clinic, are determined by diet, tobacco use, radiation or plain bad luck.

The residual 5 percent of tumors are preordained from birth, being dictated by inheritance of a mutant gene from a parent. The most notorious of these genes, and one featured prominently in *Curing Cancer,* is *BRCA1*. Inheritance of mutant forms of this gene dictates a lifetime risk of breast or ovarian cancer between 50 and 80 percent, depending on the authority consulted and the latest crop of research findings.

Much of the book's drama comes from descriptions of behind-the-scenes races to isolate (clone) familial cancer genes like *BRCA1*. Still more derives from accounts of the human repercussions of the diagnostic tests spawned by the cloned genes. Some members of cancer-ridden families are informed of impending disaster; others learn they have escaped a genetic curse that has doomed many of their close relatives.

These forays into genetic prophesying are only a beginning. In 10 years we will possess a much longer catalogue of familial cancer genes. Already, automated analysis of the DNA prepared from a drop of blood can determine inborn risk for contracting many forms of cancer. Such genetic diagnosis might seem like a big step forward, but the reality is otherwise. In many cases, little can be done to deal with a diagnosed inherited cancer gene except to wait for disaster to strike. On occasion, drastic measures, such as the bilateral mastectomies undertaken by some women with diagnosed *BRCA1* mutations, may represent the most practical response.

The major game will continue to involve the sporadic cancers, where inheritance seems to play a minor role in determining cancer onset. We now know of a number of mutated genes, damaged in cancerous tissue during an individual's lifetime, that drive the formation of these tumors. The flood of genetic sequence information pouring out of the massive Human Genome Project will facilitate the discovery of many more. Each of these genes (oncogenes, for example) makes a protein, and many of these represent tempting targets for those intent on developing new therapeutic drugs.

After decades of studying the generative mechanisms of cancer, returns on investment will be forthcoming over the next decade in the form of new treatments. This book spends little time on the current attempts to develop therapies, which, one hopes, will lead to radically new ways to cure this disease. Drug development is usually a plodding, rather unexciting process, holding little human interest, unlike the breathless races to clone the genes described in the book and the heart-wrenching stories of afflicted families.

Now that we know how cellular suicide is controlled, we can learn how to turn it on selectively.

Peeking over the horizon, what kinds of new drugs and cancer cures will we see? Prophecy is usually foolhardy in basic research fields such as this one. Still, some tantalizing clues are in sight. One recent significant development is the discovery of how most cancer cells die. Their death is triggered by a suicide program, called apoptosis, that is built into most and perhaps all cells in the human body. For the past four decades, oncologists have unknowingly been triggering this suicide program in cancer cells through the use of chemotherapeutic drugs and x-rays. Now that we know how this cellular suicide program is controlled, we can learn how to turn it on selectively in cancer cells, leaving normal cells untouched.

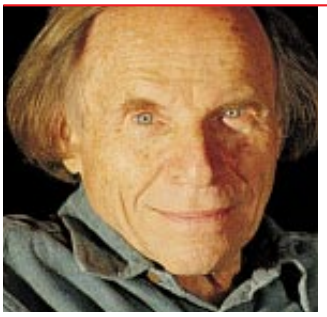
Here are three specific strategies that have attracted the attention of those who spend their time looking for new cancer cures. Cancer cells seem to be especially dependent on a constant flow of life-sustaining ("trophic") signals. Some pharmaceutical companies are considering the development of drugs to block these signals, thereby causing cancer cells to enter into apoptosis.

Growing tumor masses are dependent on a blood supply to provide them with nutrients and to evacuate wastes. The circulatory system within the tumor mass is generated by the active recruitment of blood vessels (angiogenesis) from nearby normal tissues. Completed research has now demonstrated the existence of agents that can block tumor angiogenesis, thereby starving the tumor without any discernible negative effects on normal tissues.

And finally, the recent, much heralded discovery of the telomerase gene provides the clue to how tumor cells proliferate without limit, in contrast to the limited replicative ability of normal cells. The telomerase enzyme, which makes "immortalized" cell growth possible, is a tempting target for drug intervention.

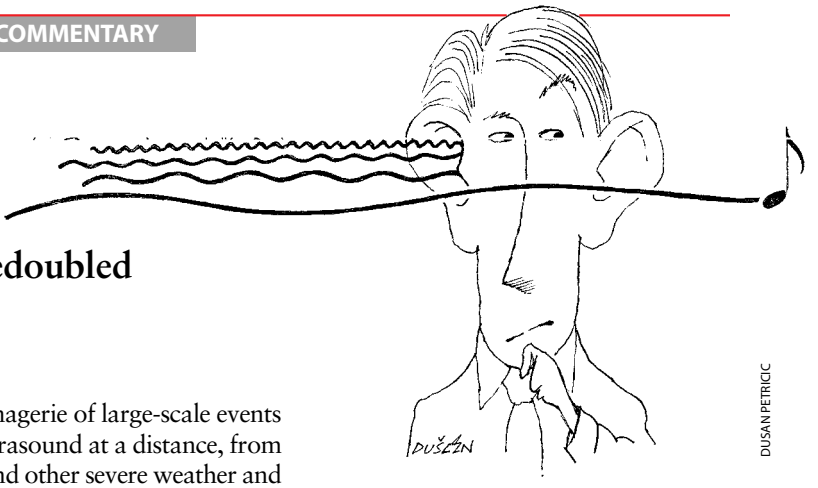
This book is engrossing, affording a view of how science is done and how its outcomes affect real lives. The task of teasing out individual cancer genes from the 80,000 and more present in humans is daunting. Waldholz shows us how it is accomplished. Never easy, it depends on wide-ranging intellects and expertise in fields as diverse as genealogy, tumor pathology and high-tech DNA analysis. For those interested in the human side of science, Waldholz offers rich fare—up-close vignettes of several of the leaders in contemporary cancer research and how drive, ambition and ample brain power have propelled their research and our understanding of this complex disease. But the real "curing cancer" book is still to be written.

ROBERT A. WEINBERG is at the Whitehead Institute for Biomedical Research and biology department of the Massachusetts Institute of Technology.



WONDERS

by Philip Morrison



DUSAN PETRICIC

Double Bass Redoubled

The deepest tones of the church organ were sensations less of ear than of chest to one young choirgirl who sat near the tall pipes. A decade ago biologist Katharine B. Payne, no longer that choirgirl but an investigator of international standing, vividly reexperienced her memorable sensation during a casual visit to the elephant house at a zoo. She and her colleagues have established how elephant families in the wild regularly communicate using sounds mainly inaudible to humans. Their purposeful trumpeting can arrange friendly rendezvous with groups five or 10 miles distant across forest or savanna, at frequencies of 35 hertz (or cycles per second) down to 12 hertz.

The familiar term "ultrasonics" is applied to sound pitched too high for our ears to hear, though not too high for dogs or bats, and for the extremely high frequency submillimeter waves much used for medical close-up imaging. "Infrasonics" is the label for the other end of the acoustic spectrum, far-ranging sounds too low in pitch to be audible. Energy loss in sound transport is the result of internal diffusion that wipes away the contrast between compressed crests and rarefied troughs as any pressure wave advances. The longer the wavelength of the sound, the farther it can go.

Ocean storms can be detected well inland by their infrasound. Poseidon's low-frequency airborne clamor is known as microbaroms, in analogy to microseisms, similar elastic waves that traverse solid rock to the seismographs. Both arise from the unceasing beat of far-off, storm-stirred waves against the shore. Microbaroms peak at frequencies near 0.2 or 0.3 hertz. The infrasonic domain now under most study begins an octave lower than that, about eight octaves below the deepest note reached by the big double bass. (Should we call infrasonics the re-re-re-re-re-re-re-redoubled bass?)

A whole menagerie of large-scale events generates infrasound at a distance, from hurricanes and other severe weather and avalanches to some auroral discharges. Volcanic eruptions, meteoric fireballs, big rocket launches and large man-made explosions, chemical and atomic, are currently the most studied of infrasound sources.

Detection begins with a quiet countryside site, a good capacitance microphone and a modest amplifier designed to work well at frequencies from a few tens of millihertz up to about 0.1 hertz. A noise filter against local wind noise is realized by a dozen simple porous hoses that extend from the microphone into yards-long spokes. By combining hose outputs at the hub, the burbles are smoothed out. The infrasound energy density flowing undetected from a volcano far overseas matches roughly the energy flow from an ordinary whisper at your side, both easily within reach of modern audio techniques. Savvy amateurs have built and operated such infrasound stations.

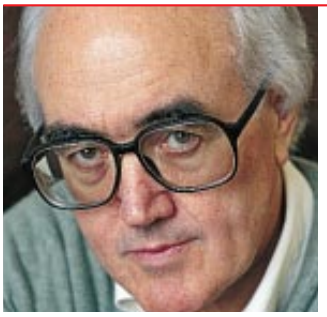
A whole menagerie of events generates infrasound at a distance, from hurricanes and avalanches to some auroral discharges.

To find the incoming wave direction, emplace a stereo array of three or four microphones spaced a few hundred yards apart. Computer-correlate all the onsets of any well-marked signal, and the best fit gives relative arrival times at the mikes as well as the wave speed. Geometry supplies directions within a few degrees. The signing of the Comprehensive Test Ban Treaty by many nations is now stimulating development of an infrasound-detection network planned to link some 60 to 80 arrays worldwide, as a smaller American network was operated for a good many cold war years. Reasonably good localization of an open-air explosion anywhere with energy release as much as one kiloton of TNT is now a practical goal, actively pursued in the U.S. by the national laboratories of Los Alamos and Sandia.

Some of the ways of atmospheric infrasound come clear in one concrete example. A recent thick volume—*Fire and Mud,* from the University of Washington Press—collects the wonderfully diverse studies, local and global, that describe the great eruption of Mount Pinatubo in June 1991. One team of infrasound experts reports what it observed in central Japan, about 1,500 miles northeast of that Philippine volcano. The researchers show the long wave trains in curve after wiggly curve. The time of major eruption was clear: after a transit delay of about three hours, infrasound pressure waves all at once grew severalfold, lasting for hours as the multiple explosions continued at the crater, pulsing minute after minute, often near the energy of a megaton of TNT.

The sound came out of the southwest, not far off the expected great circle. But within about 30 hours more, a second long-lasting signal arrived, much weaker and now from the northeast. What was that? Certainly it was the infrasound from Pinatubo that had started out for Japan the long way around. One final signal came again from out of the southwest but then delayed about 35 hours. That fits the original burst that had crossed over Japan bound for the north, to return from the south af-

Continued on page 111



CONNECTIONS

by James Burke

Cheers

DUSAN PETRICIC

A barman the other night opened my tonic water bottle with a flourish, and the tinkle of metal reminded me of William Painter, the man who devised the Crown Seal Company bottle cap. (And then blew his Hall of Fame chances by advising one of his salesmen, name of Gillette, to invent a similar use-and-throw-away gizmo, the safety razor.)

Anyway, after many earlier attempts at effecting closure—including wax, glass balls and a combo of wire and cork—metal-capped effervescence was first made popularly available by Jacob Schweppes at the 1851 Crystal Palace Exhibition in London. At which he sold vast quantities of soft drinks, thus realizing the long-dead dream of Joseph Priestley, who had invented soda water decades earlier. Getting fizz into water (drinking same, it was thought, would cure yellow fever) was one of Priestley's more successful industrial-chemistry efforts, all of which were inspired by the modern education he had received at one of the great Dissenter academies. These had originally been set up in late 17th-century England by Protestants who wouldn't accept the return of a monarchy after the failure of Oliver Cromwell's Puritan Commonwealth. These guys opened their own schools because those who had refused to sign an oath of allegiance to the throne were barred from going to university, becoming members of Parliament, preaching or trading in the major cities, and taking jobs in the army. So went their options for advancement.

The up-to-date curriculum of the Dissenter academies (with unheard-of subjects such as science and modern languages) had been generally influenced by the ideas of a Czech freethinker and educationalist named Amos Komensky. This Bohemian theologian had arrived in England in 1641 and made such an impression on the Puritans with his two books on pedagogy—*The Great Didactic* and *The School of Infancy*—that he was showered with job offers. One he turned down was the presidency of some little-known New England college called Harvard: Komensky preferred to concentrate on the further development of his philosophical views. Including thoughts about reality and how it was made up of irreducibly small elements, an idea said to have fired up a certain German math freak named Gottfried Leibniz to conceive of fundamental entities he termed "monads." And then, in 1675, to invent the calculus you needed to measure infinitesimally small matters. (Though if you're English, you believe Newton did it all himself, and first.)

Ross was unaware that the wandering magnetic spot was already on its way elsewhere.

The other, and larger, thing Leibniz was crazy about was libraries. His job as bookkeeper to the elector of Hannover was pretty much of a sinecure, given him so he could write (though he never finished it) a history of the ducal family. While book hunting in Paris, Leibniz was apparently much taken by a text on how to put together your own library, with instructions on how to catalogue, choose titles, dust books and treat the library staff. This bookworm's delight had been written in 1627 by Gabriel Naudé, who had put his money where his mouth was by collecting and organizing a gigantic library of 40,000 volumes for his boss Cardinal Mazarin, first minister of France, who then built a place to house it all and opened it to the public. Naudé's book also caught the eye of an English aristocrat and scholar, John Evelyn, who came across it during a tour of Europe a few years later. Evelyn eventually translated Naudé's book and gave a copy to a friend, who used it to organize the pile of material he had amassed (he chose books by their size rather than subject matter, except in the case of erotica) as part of his great plan to write a history of the English navy.

It was only thanks to this buy-it-by-the-yard bibliophile, Samuel Pepys, that by the late 1600s there was anything naval to write about. As secretary of the admiralty, Pepys had made sure Britannia was able to rule the waves, by introducing standardized ordnance, regular shipbuilding programs, official rates of pay and promotion, disciplinary codes and a new breed of naval captain who knew bow from stern. About the only reformatory matter at which Pepys was a signal failure was, in fact, signaling. Which at the time wasn't up to much. The contemporary limitations on what you could say with nine limp flags on a windless day is best illustrated by the fact that even in a gale, if the admiral were inviting you to lunch, the flagship was obliged to hang up a tablecloth.

By 1794 things were better, but from the point of view of the British government, the real problem was that the navy was still not getting the message fast enough. Especially between London and the fleet headquarters in Portsmouth. So that same year, when a French prisoner of war was discovered to be carrying a copy of instructions for a radically new semaphore communications system (recently invented by Frenchman Claude Chappe and already in use by Britain's deadly foe, Napoleon), suggestions for an improved version were instantly forwarded to the authorities by Reverend John Gamble, the military chaplain who had retrieved the information. Gamble's

semaphore used a wood frame with five shutters that could open and close in coded patterns that could be seen by telescope some distance away. A chain of stations to relay these patterns would be able to get messages from Portsmouth to London in mere minutes. Unfortunately, a further minor improvement was proposed by another clergyman, who also happened to be the fourth son of an earl. So Gamble's idea bit the dust, and he went back to buying foreign patents. One of which was for a French food-preserving process in tins.

By 1818 yet another attempt to search the polar regions of Canada for the Northwest Passage set sail with provisions of the new canned food (the expedition also carried 40 umbrellas as presents for Eskimos). Ultimately, the enterprise failed in its quest, but the experience whetted the appetite of James Clark Ross, onboard nephew of the expedition leader, and in 1829 he headed his own venture: to find the magnetic north pole in the same approximate area. On the morning of June 1, 1831, when Ross hung a magnetic needle on a fine thread of New Zealand flax, the needle's dip of 89′ 39″ was close enough to vertical to convince Ross that the pole was underfoot. The location—at 70°3′17″N 96° 46′43″W—was marked with a cairn of stones, a flag was flown, and the magnetic north pole was claimed in the name of Great Britain and King William IV. (Ross was unaware that even as he spoke, the wandering magnetic spot was already on its way elsewhere.)

As is often customary with explorers, Ross gave names to some of the various desolate places in which he got stuck. On this occasion these appellations included the Boothia Peninsula, the Gulf of Boothia and Felix Harbor. You'll gather from this that somebody called Felix Booth was worth commemorating. In fact, had it not been for Booth's generous gift of £20,000, Ross's entire voyage might not have been possible, the magnetic north pole might not have been (temporarily) British, and I might not have managed to make this tale end the same way all these columns do—back where it began.

In that bar, remember? Where the other thing the barman was pouring into my glass, besides tonic, was what made Felix Booth rich enough to finance polar expeditions. Booth's Gin. SA

Wonders, *continued from page 109*

ter one full passage around the world.

Infrasound sound is low loss, to be sure. But that doesn't bring it around the world, because sound, like light, travels in a straight line in a uniform medium. We don't see over the horizon. How can sound circumnavigate at all? It travels on high within an atmospheric duct or channel, bending to the earth's sphere as do the strata of the atmosphere itself. The speed of sound depends on gas temperature, on the weight of the gas molecules and on winds that move the medium. All three can be reasonably invoked to model infrasound propagation. Just as light from above bounces off the sun-heated highway to appear as a bubbling mirror suspended in the air, infrasound rays can be trapped between layers of the upper atmosphere.

The average sound speeds observed in the three signals fit well the proposal that Pinatubo sound went far above the horizon: it bent 70 miles high by refraction at the hot, thin layer of the high thermosphere, then bounced again and again between that height and the earth, so that some reached Japan. A fit can be found for the observed signal speeds, because seasonally shifting winds of the right strength and direction are expected at high altitudes along the global path. The long-delayed signals represent a mix of what would be carried in two well-known, if oversimplified, atmospheric channels aloft. Pinatubo seismic records made in the distant, quiet Black Forest of Germany suggest energy feedback among infrasound, certain seismic waves and vibrations of the dusty gas of the volcano plume itself.

Pressure waves from fast-moving meteorites high aloft often span the globe at still lower frequencies. The physics of such sound, called acoustic-gravity waves, are rather different. These waves are five to 10 miles long and at a few millihertz respond to the decline of atmospheric buoyancy in the thin heights. Their speed varies with differing directions from the vertical. They are hard to model as simple bouncing rays because they act more like particular modes of vibration in the richly resonant waveguide of air that enfolds the solid earth. It won't be too long until these sounds from afar can help us complete the reliable world census of incoming large meteorites that so far we lack. SA

POLYMERASE CHAIN REACTION

by Elizabeth A. Dragon
Roche Molecular Systems

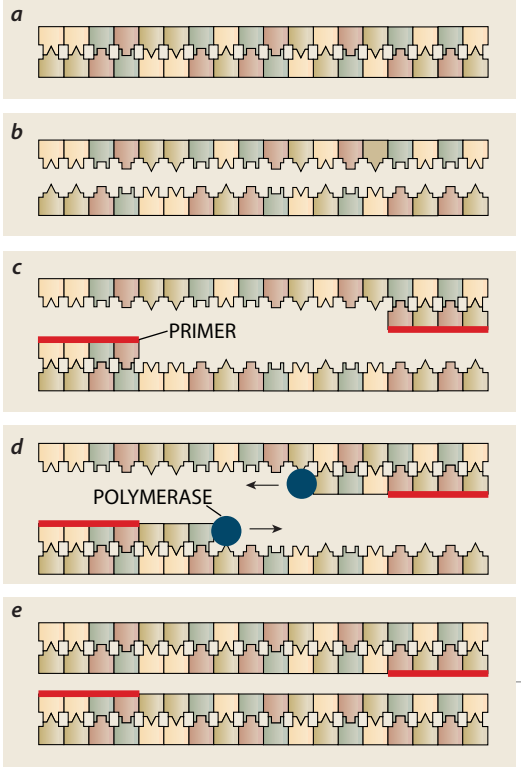
STEVEN SENNE *AP Photo*

CRIME SCENE CLUES, such as a single hair or a drop of blood, can be analyzed using PCR.

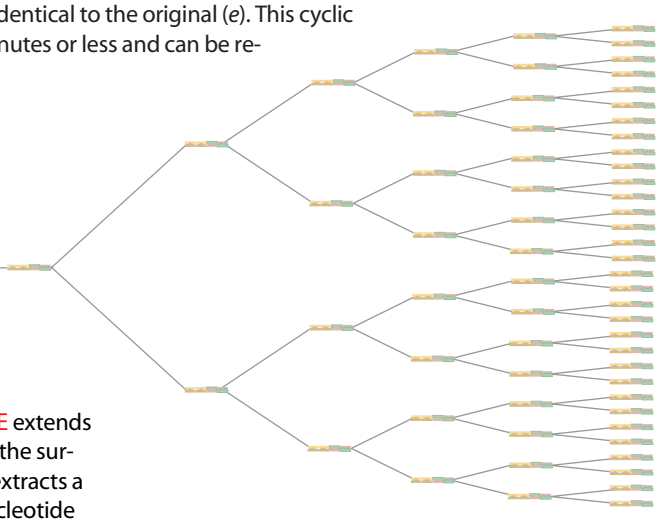
Polymerase chain reaction (PCR) is a technique that mimics nature's way of replicating DNA. First described in 1985, PCR has been adopted as an essential research tool because it can take a minute sample of genetic material and duplicate enough of it for study. PCR has been used to identify the remains of Desert Storm casualties, to analyze prehistoric DNA, to diagnose diseases and to help make identifications in police investigations.

DNA is most often found as a double-stranded molecule, twisted as a helix, in which each strand complements the other. PCR starts with the DNA sample, which is put in a reaction tube along with primers (short, synthetic pieces of single-stranded DNA that exactly match and flank the stretch of DNA to be amplified), deoxynucleotide triphosphates (dNTPs, the building blocks of DNA), buffers and a heat-resistant enzyme (polymerase). Heating the mixture separates the "template" strands of DNA. Then, at varying temperatures, the rest of the components in the mixture spontaneously organize themselves, building a new complementary strand for each original.

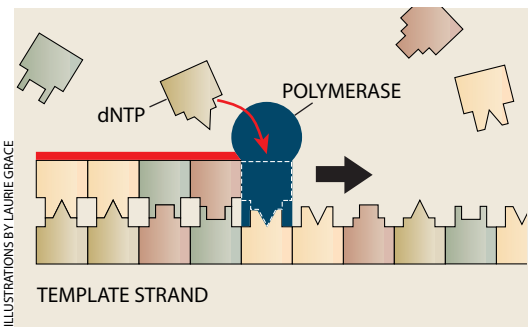
At the end of each cycle the DNA count has doubled. If you start with one DNA molecule, at the end of 30 cycles (only a few hours later) there will be about a billion copies. Thus, if you are looking for a single gene among thousands, the game changes from "searching for a needle in a haystack" to "making a haystack of needles."



DUPLICATING DNA begins with a double-stranded stretch of DNA to be amplified, or copied (*a*). In a solution heated to 95 degrees Celsius (203 degrees Fahrenheit), hydrogen bonds between the strands break, leaving two single strands (*b*). When the mixture is cooled to between 50 and 65 degrees C, specially manufactured DNA primers bind complementarily to each strand at points flanking the region to be copied (*c*). At 72 degrees C, polymerase enzymes extend the bound primers in one direction, using the original DNA as a template (*d*). The products are two new double strands of DNA, both identical to the original (*e*). This cyclic reaction takes only minutes or less and can be repeated indefinitely.



EXPONENTIAL GROWTH of the DNA target occurs because the products of each cycle become the templates for the next cycle.



ILLUSTRATIONS BY LAURIE GRACE

POLYMERASE ENZYME extends a bound primer. From the surrounding medium, it extracts a free-floating deoxynucleotide triphosphate (dNTP) that will complement the next unpaired position in the template strand of DNA. The enzyme then joins the dNTP to the end of the primer and moves on to the next position.