SPACE DISK DYNAMICS • LET THE INTERNET RUN EVERYTHING

SCIENTIFIC AMERICAN

Genetic Junk and the Secrets of Complexity

OCTOBER 2004
WWW.SCIAM.COM

CONTROLLING HURRICANES

Ways to Tame Nature's Most Destructive Storms

FIXING THE VOTE
What Electronic Ballots Can Do for Democracy

CRYSTAL CLEAR
How Cells Die (Slightly) to Get Transparent

october 2004 contents

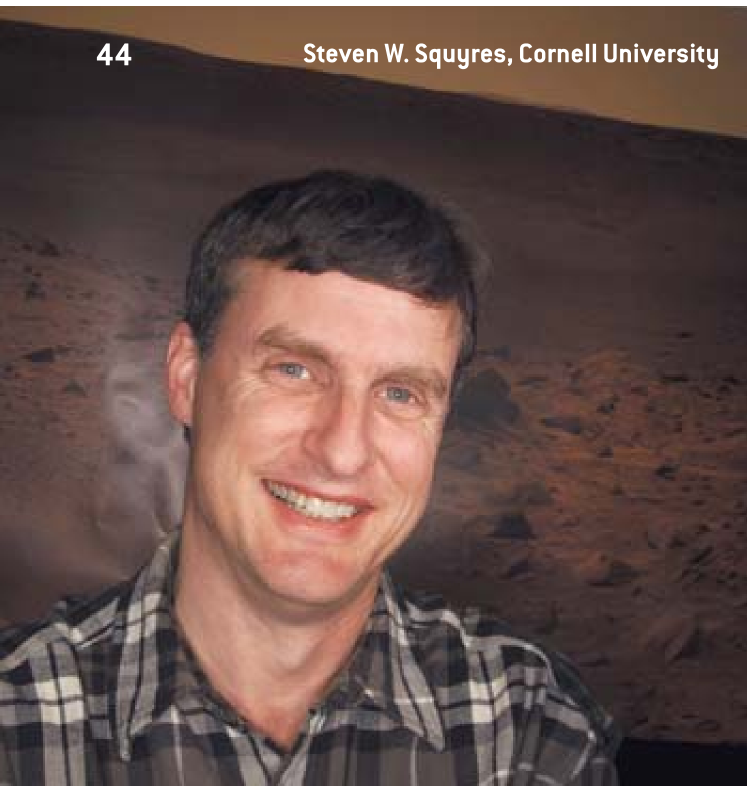
SCIENTIFIC AMERICAN Volume 291 Number 4

features

departments

26

104



44 Steven W. Squyres, Cornell University

columns

Cover image by Jean-Francois Podevin.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2004 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40012504. Canadian BN No. 127387652RT; QST No. Q1015332537. Publication Mail Agreement #40012504. Return undeliverable mail to Scientific American, P.O. Box 819, Stn Main, Markham, ON L3P 8A2. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. 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; (212) 451-8877; 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.** Printed in U.S.A.

SA Perspectives

Breathing Difficulties

A day after the August 2003 blackout shut down power plants in much of the Midwest and Northeast, scientists at the University of Maryland made a remarkable discovery. Collecting air samples from a plane flying over central Pennsylvania, the researchers found a 90 percent drop in sulfur dioxide levels and a 50 percent reduction in ozone. Visibility increased by more than 25 miles. Says Lackson Marufu, the lead author of the researchers' report: "The improvement in air quality was so great that you could not only measure it, but could actually see it as a much clearer less hazy sky."

AIR POLLUTION is an issue in the presidential election.

Improving air quality should not require shutting down the power grid. In 1970 President Richard M. Nixon signed the Clean Air Act, which ordered power companies to install pollution-control devices on all newly built plants. In 1977 Congress added the new-source review provision, which compelled electric companies to clean up older plants as well if their emissions increased because of modifications. Two decades later, though, the Environmental Protection Agency charged that many companies were skirting the law by modifying generators—say, to burn more coal—without putting in pollution controls. In 1999 the U.S. Justice Department sued seven companies for alleged violations at 17 power plants in the South and Midwest. The tough stance produced results: some companies settled the lawsuits by agreeing to reduce their emissions, and others began negotiating with the EPA.

That progress has now stalled, however. EPA officials appointed by President George W. Bush have rewritten the rules governing new-source review, restricting the agency's ability to force electric companies to install pollution controls. Although the new rules have not yet gone into effect because of legal challenges, the change in policy has already eviscerated the EPA's enforcement efforts. Last November the agency dropped 70 investigations into violations of the Clean Air Act. The loss to public health is catastrophic: according to a study by Abt Associates, a technical consulting firm, weakening the regulation will result in thousands of premature deaths every year—from chronic lung and cardiovascular diseases—that could have been prevented under the old rules. The people who would be hurt most are those who live downwind of the power plants alleged to be illegal polluters—for example, residents of the areas around Cincinnati, Pittsburgh and Charleston, W.Va., which have some of the worst levels of particulate pollution in the country.

The Bush administration has defended its changes by claiming that new-source review is costly and cumbersome. But the new rules contain loopholes that effectively release utilities of their duty to clean their oldest, dirtiest generators. For example, the Clean Air Act allows companies to make routine repairs to old power plants without being required to install pollution controls, but the new rules expand the definition of routine maintenance to include major upgrades costing up to 20 percent of a plant's value. The new rules do not even make economic sense for the nation as a whole: although the Department of Energy estimates that adding pollution controls to old generators would cost $73 billion over 20 years, the health benefits would total more than $1 trillion over the same period.

New-source review has become an important issue in the upcoming presidential election because changing the rules does not require the approval of Congress. Senator John Kerry of Massachusetts, the Democratic nominee for president, has vowed to restore the stricter regulation. We urge President Bush to do the same.

THE EDITORS editors@sciam.com

RICK BOWMER *AP Photo*

How to Contact Us

EDITORIAL

For Letters to the Editors:
Letters to the Editors
Scientific American
415 Madison Ave.
New York, NY 10017-1111
or
editors@sciam.com
Please include your name and mailing address, and cite the article and the issue in which it appeared. Letters may be edited for length and clarity. We regret that we cannot answer all correspondence.

For general inquiries:
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-754-0550
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

For new subscriptions, renewals, gifts, payments, and changes of address:
U.S. and Canada
800-333-1199
Outside North America
515-247-7631
or
www.sciam.com
or
Scientific American
Box 3187
Harlan, IA 51537

REPRINTS

To order reprints of articles:
Reprint Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8877
fax: 212-355-0408
reprints@sciam.com

PERMISSIONS

For permission to copy or reuse material from *SA*:
www.sciam.com/permissions
or
212-451-8546 for procedures
or
Permissions Department
Scientific American
415 Madison Ave.
New York, NY 10017-1111
Please allow three to six weeks for processing.

ADVERTISING

www.sciam.com has electronic contact information for sales representatives of Scientific American in all regions of the U.S. and in other countries.

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138

Los Angeles
310-234-2699
fax: 310-234-2670

San Francisco
415-403-9030
fax: 415-403-9033

Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138

Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457

Southeast
Publicitas North America, Inc.
404-262-9218
fax: 404-262-3746

Southwest
Publicitas North America, Inc.
972-386-6186
fax: 972-233-9819

Direct Response
Special Aditions
914-461-3269
fax: 914-461-3433

Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430

Canada
Derr Media Group
847-615-1921
fax: 847-735-1457

France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329

Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21

Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

India
Convergence Media
+91-22-2414-4808
fax: +91-22-2414-5594

Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139

Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662

Middle East
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320

Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +46-8-442-7059

U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922

On the Web

WWW.SCIAM.COM

UPDATED EVERY WEEKDAY

Visit **www.sciam.com/ontheweb** to find these recent additions to the site:

Early Bird Had the Brains to Fly

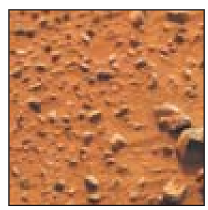
As any ostrich knows, getting off the ground requires more than just wings and feathers. A thorough examination of the earliest known bird, *Archaeopteryx*, has provided evidence of the specific neural machinery thought to be necessary for flight.

Antibody Therapy Halts Early-Stage Alzheimer's in Mice

Alzheimer's disease is characterized by two separate brain lesions: plaques and tangles. The results of a new study indicate that an experimental immune therapy that targets the former can also impact the latter, so long as the disease has not progressed to an advanced state.

Spirit's Sojourn Leaves Ancient Lake Hypothesis High and Dry

The Spirit rover traversed 637 meters in its first 90 Martian days. But to the dismay of project scientists, its trek through the Gusev crater has not revealed any evidence of the ancient lakebed that geologists thought might be there. With winter approaching, the question is, How long can the solar-powered explorer continue its quest for water before it runs out of steam?



Ask the Experts

How are past temperatures determined from an ice core?
Robert Mulvaney, a glaciologist with the British Antarctic Survey, explains.

Exclusive Online Issue: Mysteries of the Milky Way

Unraveling the secrets of our home galaxy with a decade of stunning discoveries.

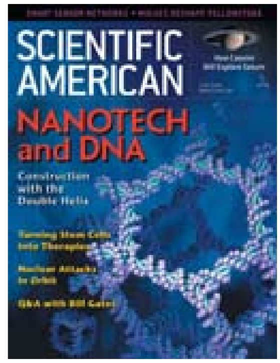
Purchase today for $5 at Scientific American Digital

www.sciamdigital.com

JOHN SIBBICK/NHM (*top*); © *SCIENCE* (*bottom*)

Letters

EDITORS@SCIAM.COM



TAKE AN ISSUE that featured an SA Perspectives, "Stem Cells: A Way Forward"; add "The Stem Cell Challenge," by Robert Lanza and Nadia Rosenthal; then fold in a Q&A with Microsoft's Bill Gates; and garnish it with "Nuclear Explosions in Orbit," by Daniel G. Dupont—and you have a recipe for a hot-and-spicy June.

The hottest topic was stem cell research. An e-mail from Bobby Winters poses a dilemma: When asked, "If your child could be helped, would you use an embryo?" he responds, "To save the life of one of my children, I would have no problem dismembering unwilling adults, but one of the purposes of law is to tame such emotional reactions. The question is not the depth of emotion nor the worth of the cure, but the value of human life."

On the other side, trauma center neurosurgeon Harold A. Wilkinson of Wellesley, Mass., writes: "I know well how valuable transplanted tissues and organs (from loved ones with fatal brain injuries) have been to their recipients. If this is so widely endorsed by the American public, and approved by the world's great religions, why is there so much angst about tissue transplantation of stem cells from embryos that no longer can live?"

STEM CELL ETHICS

We were glad to find a positive mention of the President's Council on Bioethics report "Reproduction and Responsibility" in "Stem Cells: A Way Forward" [SA Perspectives]. I would, however, correct errors in your description of it.

First, our report does not say that research on early-stage embryos should be "acceptable." Second, the council did not recommend that embryos should not be "maintained" past a certain stage but only that they should not be subject to research past that stage.

Finally, in a July 2002 report, a majority of the council's members argued that all human cloning, whether for research or for producing children, should be prohibited for now, whereas a minority of members argued that only cloning to produce children should be banned. The latest report does not change those recommendations.

Dean Clancy
Executive Director
President's Council on Bioethics

RESTRAINING ATTACHMENT THERAPY

I was astonished by the inaccuracies and gross generalizations in "Death by Theory," by Michael Shermer [Skeptic]. His representation of attachment therapy is an injustice to the therapists who practice appropriate therapy with children who suffer from attachment disorders as well as the families who benefit from this treatment.

Like all responsible professionals, members of the Association for Treatment and Training in the Attachment of Children (ATTACh) were appalled at the death of Candace Newmaker. The treatment she endured, however, in no way characterizes the type of treatment endorsed by ATTACh and practiced by its members. Therapist Connell Watkins and her associates were not members of ATTACh at the time of Candace's death. By linking ATTACh and attachment therapy to this unfortunate situation, Shermer does a disservice to your readership, cheapens Candace's death and does nothing to advance the science of child therapy.

ATTACh considers restraint a last-resort intervention to maintain the safety of an individual in danger of hurting himself or others and not a therapeutic intervention, as Shermer states in his column. In fact, ATTACh has issued a Position Statement on Coercive Therapy opposing the use of force in therapy. That statement and other information about ATTACh are available on our Web site, www.attach.org.

Todd Nichols
President, Association for Treatment and Training in the Attachment of Children

SHERMER REPLIES: Nichols's letter only confirms my general thesis that restraint is endorsed as a form of therapy by ATTACh, regardless of whether it is a first or last resort. The column did not say that Watkins or her associates were members of ATTACh. It said that Candace's mother "sought help from a therapist affiliated with" ATTACh and that Watkins was a past clinical director of the Attachment Center at Evergreen. Contrary to Nichols's objections about the depiction of ATTACh practices, the ATTACh Professional Practice Manual, adopted after Watkins's conviction, states: "Because the child's defenses against healthy relationships are so strong, therapeutic interventions may be confrontational and challenging and may involve holding, touch, or physical proximity, while never losing sight of everyone's need to feel and be safe."

"Confrontational," "challenging," "holding" and "physical proximity" create coercion. The American Psychiatric Association's published position statement warns about the use of coercive holding therapies in treating attachment disorder: "In fact, there is a strong clinical consensus that coercive therapies are contraindicated in this disorder." Finally, there is no scientific evidence for attachment therapy or its practice; it is therefore a pseudoscience.

MICROSOFT'S BRAIN TRUST

"A Confederacy of Smarts," by Gary Stix [Innovations], featuring Microsoft's research division, was both compelling and chilling, because it has made such a huge commitment to attracting the best and brightest but has consistently shown it is only interested in self-serving innovation.

MS operating systems run more than 90 percent of the world's computers. And now Microsoft Research is co-opting the globe's best minds. To do what? Promote the free flow of information and other altruistic things?

Scientific American should be first to champion changes in antitrust laws to stop the concentration of corporate power that not only has stifled scientific innovation but may actually threaten our future.

Nick Mathe
Madison, Wis.

Research laboratories such as Bell Labs, IBM Research and PARC (Palo Alto Research Center, formerly part of Xerox) were funded by the profits generated by dominant market positions, without which, they could no longer sustain the funding of the research labs—they are now a shadow of what they once were.

Microsoft makes dominant products that generate the resources necessary to create and sustain its research. Since it is inevitable that its profits will revert to a mean level, it is only a matter of time before research funds will dwindle away. In the meantime, we can all hope for meaningful discoveries and innovations, and for a new company that will fund the next-generation research laboratory.

James Wattengel
São Paulo, Brazil

BILL GATES **inspires hope and hostility.**

CHANCES WITH WOLVES

In "Lessons from the Wolf," by Jim Robbins, a revelation is attributed to William J. Ripple of Oregon State University: when he heard in 1997 that aspen trees were on the decline in Yellowstone and no one knew why, he was drawn to the park to try to solve the mystery.

This is an erroneous statement. Alan Beetle of the University of Wyoming had solved this mystery in the 1960s and long tried to call attention to the problem of elk browsing on aspen shoots.

Elks' effect on woody vegetation can be easily seen in the Lamar Valley—a teaching tool I use to show how native herbivore populations, when allowed to expand unchecked, can overgraze or overbrowse just as readily as domestic cattle.

Charles (Chuck) H. Butterfield
Range Management Program
Chadron State College

ROBBINS REPLIES: Critics have said for years that elk in Yellowstone have an adverse impact on vegetation. Even the Park Service once held this view, and rangers used to shoot elk to reduce their numbers. But the picture is more complex. Since the 1970s park scientists have pointed to climate change and perhaps other factors. Overgrazing is admittedly a big part of the problem, but is it because there are too many elk? Or, as Ripple's work suggests, is it where those elk grazed as a result of an absence of predators? Much more work remains to understand the situation thoroughly.

NUKES OVERHEAD

Although a few low Earth orbit satellites are engaged in specialized communications, "Nuclear Explosions in Orbit," by Daniel G. Dupont, trades fact for fantasy with dire predictions of how an orbital nuclear explosion could cripple the world's telecommunications infrastructure.

A great majority of the world's networks are connected with fiber-optics. Save a near-direct hit by a nuclear weapon, fiber-optics networks are far more tolerant of an electromagnetic pulse than any other existing broadband technology.

Mike Hansen
Lakewood, Ill.

ERRATA The weight of the Los Angeles class submarine USS *Miami* was incorrectly given as 395 tons in "Deep Silence," by Mark Fischetti [Working Knowledge]. It has a full displacement of 6,146 tons.

In "A Transparent Enigma," by Madhusree Mukerjee [Insights], Temple Grandin was incorrectly affiliated with the University of Illinois; she is a professor at Colorado State University.

KIM KULISH *Corbis*

50, 100 & 150 Years Ago
FROM *SCIENTIFIC AMERICAN*

Crick and DNA ▪ Water Trumps Steam ▪ Whaling and Cholera

OCTOBER 1954

ECONOMIC PSYCHOLOGY—"The notion of 'saturation' of the market is based on old-fashioned psychological assumptions which in turn rest on the analogy of biological drives: for example, if an animal is hungry, it is motivated to search for food; after it has eaten, the motive disappears. The saturation concept has resulted in dire predictions about the future of the U.S. economy [*see illustration*]. Some people point to the large proportion of U.S. families that already possesses major goods, such as refrigerators (over 80 per cent), and they argue that in the future sales will be limited largely to replacement needs. But social motives are different from biological ones. Fulfillment of one aim leads to striving for another."

STRUCTURE OF DNA—"J. D. Watson and I, working in the Medical Research Council Unit in the Cavendish Laboratory at Cambridge, were convinced that we could get somewhere near the DNA structure by building scale models based on the X-ray patterns obtained by M.H.F. Wilkins, Rosalind Franklin and their co-workers at King's College, London. The most important assumption we had to make had to do with the fact that the crystallographic repeat did not coincide with the repetition of chemical units in the chain but came at much longer intervals. A possible explanation was that all the links in the chain were the same but the X-rays were seeing every tenth link, say, from the same angle and the others from different angles. What sort of chain might produce this pattern? The answer was easy: the chain might be coiled in a helix.—F.H.C. Crick" [*Editors' note: Crick, Watson and Wilkins were awarded a Nobel Prize for this work in 1962.*]

OCTOBER 1904

POWER CHANGE—"Steam power is going out of fashion. Water power is coming in. Electrical transmission is working the change. In hundreds of villages and towns throughout the country, steam engines have been almost, if not altogether, displaced by electrically transmitted water power. In one such case, at Concord, N.H., the shops of the railway that hauls all of the coal entering the city are operated by electric motors of about 550 horse power, and the steam plant that formerly did the work is permanently out of use. Electrical energy transmitted from water power displaces steam, not so much because the former is cleaner, safer, and more conducive to good health, as because it is cheaper."

SATURATED WITH GOODS—an economist's fear, 1954

OCTOBER 1854

WINTER WHALING—"Whalemen learned from the Esquimaux that large numbers of whales resorted to certain inlets in the bays in Davis' Straits and Baffin's Bay, where they remained during the winter for shelter. This information suggested to Captain Penny the idea of fitting out parties to winter in the Polar Regions to secure as many as possible in the fall and spring, and to boil the oil out during the winter. Two vessels were accordingly prepared for the purpose, and sailed from Aberdeen on the 13th of August, 1853, manned with 33 men and 3 boys. Their efforts were entirely successful, notwithstanding the cold was 40 degrees below zero. The active duties of the men tended to preserve their health, and none felt time to hang heavily. But, singular as it may seem, that dreadful scourge, the cholera, broke out among the Esquimaux, and swept off many, while the crew escaped with slight premonitory symptoms."

CANNONBALL TRAINS—"A paper was lately read by Judge Meigs, before the American Institute Farmer's Club, at New York, in which he said: 'Such is part of the future; the railroad from point to point a mathematical line; the rails ten times stronger than any now used; the locomotives on wheels of far greater diameter, say twelve or fifteen feet; instead of one hundred miles an hour, we shall more safely travel three hundred miles an hour!' If the driving wheels were twenty-four feet in circumference (whoppers!), they would have to make 1,100 revolutions per minute. As the wheels cannot turn round without steam, the boiler would have to evaporate about a ton of water per minute. How can such a quantity of steam be evaporated in a locomotive boiler, in this space of time?"

BIOTERROR

An Uncertain Defense

HOW DO YOU TEST THAT A HUMAN EBOLA VACCINE WORKS? YOU DON'T **BY W. WAYT GIBBS**

The Ebola virus is among the deadliest on earth; in outbreaks last year in the Republic of Congo, 157 of the 178 people infected with it died of hemorrhagic fever. Because it can be exceedingly contagious in aerosol form, the Ebola virus ranks with smallpox and anthrax as one of the most worrisome potential biological weapons. Although there is no effective treatment, recent tests have shown a new vaccine able to prevent infection in monkeys. Clearly, researchers could never intentionally expose human volunteers to the lethal virus. And there are no populations at especially high risk for Ebola, as there are for HIV. So how can doctors determine whether the vaccine works in people?

That question, which applies as well to experimental vaccines for smallpox and anthrax, took on new significance on July 21, when President George W. Bush signed the Project Bioshield Act. The law authorizes the U.S. Department of Homeland Security to spend up to $5.6 billion over 10 years to increase its stockpile of antibioweapons medicines, including drugs that the U.S. Food and Drug Administration has not yet approved as safe and effective.

The secretary of health and human services can now recommend that the president order the distribution of experimental drugs to the armed forces or even to the general populace, should the secretary perceive "a significant potential for a domestic emergency, involving a heightened risk of attack." Although scientific evidence of some kind

EXTRA PRECAUTIONS are required when doing experiments on Ebola virus. Strict regulations confine all experiments to a "Biosafety Level 4" laboratory. Such labs are very expensive to operate, so only a handful exist in the world, such as this one in the U.S. Army Medical Research Institute of Infectious Disease in Fort Detrick, Md.

U.S. ARMY *USAMRIID Photo*

must suggest that the drugs will do more good than harm, human clinical trials—for decades, the only evidence that has mattered—are no longer strictly required.

The prospect of treating thousands of people with a vaccine or drug only proved to work on animals may seem risky. But the realities of the pharmaceutical market and the lethal character of many bioweapons leave few alternatives. The pharmaceutical industry has already all but abandoned work on vaccines for many of the world's major infectious diseases, such as malaria and tuberculosis, because the people most vulnerable to such illnesses tend to be least able to afford expensive medicines. There is virtually no natural market for vaccines to prevent smallpox (which has been eradicated) or anthrax (which is not contagious) or Ebola (which occurs in vicious but sporadic outbreaks).

So federal agencies are creating a market. "The Project Bioshield law specifies initial stockpiling of Ebola vaccine at about $90 million and a long-term procurement of about $260 million," Vijay B. Samant, president of Vical, observed in an August conference call with investors. Vical is a biotech firm based in San Diego; both it and Crucell, a Dutch company, have won potentially lucrative contracts to manufacture the genetically engineered ingredients in the new vaccine.

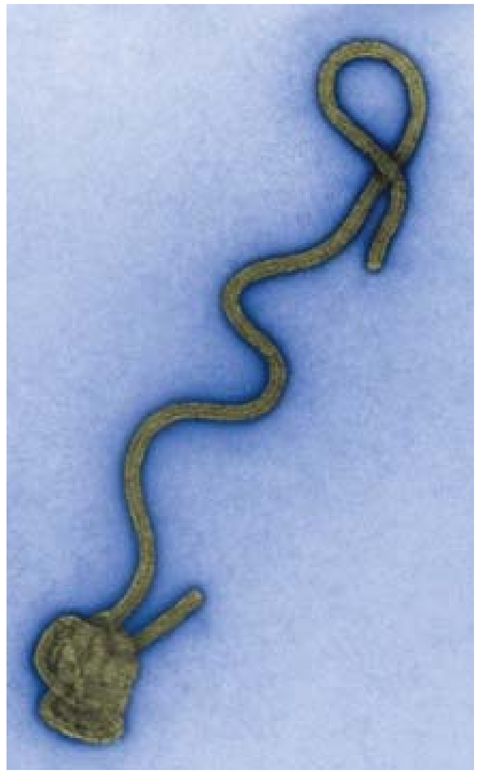
"To gain FDA approval, we will have to gather safety data on perhaps 5,000 people," says Gary J. Nabel, director of the Vaccine Research Center at the National Institutes of Health. Nabel's group designed the immunization to have two stages: a DNA primer and a viral booster shot. Both parts contain only minute fragments of Ebola virus, so the vaccine itself could not cause infection. Small human safety trials of just the primer portion, made by Vical, are now under way at the NIH. But recent experiments on macaques have shown that the booster alone led to full immunity against Ebola in less than four weeks. Jaap Goudsmit, chief scientific officer at Crucell, says it remains to be seen how long the protection will last.

That is a question typically answered by a large-scale clinical trial. But a special rule passed in 2002 allows FDA approval even without direct evidence that it works in humans, in cases where subjecting humans to clinical trials would be unethical or infeasible.

A company must first show that the vaccine works in monkeys (or another animal similar to humans). Researchers then have to figure out how the animals' immune systems respond to the optimal dose and find an equivalent dose in humans that generates a similar immune response. Project Bioshield allows the president to waive even that lowered regulation when the nation faces "heightened risk of attack."

It may be years before scientists can determine an optimal human dose for the new Ebola vaccine. That does not mean, however, that it won't soon enter the national stockpile. The Bush administration says that next year it expects to purchase 75 million doses of a new anthrax vaccine that has not yet even completed human safety trials.

Of course, Nabel says, "you can bet there would be follow-up studies if the Ebola vaccine was used in a real outbreak." He acknowledges, however, that "if the vaccine is so good that it aborts infection before the virus induces an immune response, we might not be able to tell who was exposed and who wasn't."



SINGLE VIRAL PARTICLE of Ebola is among the most lethal known. In the bloodstream, it can multiply into billions within days. But a new vaccine containing fragments of Ebola DNA prevents the virus from gaining a foothold—at least in mice and monkeys.

CAMR/A. B. DOWSETT/SCIENCE PHOTO LIBRARY

BIOWEAPONS VACCINES ON TRIAL

Vaccines against anthrax, smallpox and Ebola have entered human safety trials. Such clinical tests cannot predict whether the injections will actually protect people from infection, however. That evidence will have to come in part from animals and in part from a leap of faith. In the case of a new drug for preventing Ebola infection, the National Institutes of Health and the Dutch drug firm Crucell have given the vaccine to a small number of mice and macaques. All the immunized animals survived exposure to a high dose of the most lethal strain of Ebola virus—the same strain that the U.S.S.R. reportedly attempted to weaponize. Every unvaccinated control animal died.

SPACEFLIGHT

Readying for a Relaunch

NASA MAKES THE SPACE SHUTTLE SAFER BUT LIMITS ITS MISSIONS BY MARK ALPERT

Inside the Orbiter Processing Facility at the Kennedy Space Center, NASA technicians are readying the space shuttle *Discovery*, which is expected to be launched next March. *Discovery* will be the first shuttle to go into orbit since the loss of *Columbia* in February 2003, and the space agency has redesigned the vehicle to reduce the risk of a similar catastrophe. Yet not all the recommended fixes will be in place at the time of the first flight, limiting the kinds of missions the spacecraft can tackle.

SPACE SHUTTLE *DISCOVERY* is scheduled to go into orbit in March in the first flight since the loss of *Columbia* in February 2003. Pictured here in 1998, *Discovery* is being rebuilt to make it safer.

NASA, HO *AP Photo*

First and foremost, NASA will remove the foam insulation from the metal struts connecting the shuttle to its external fuel tank; a 1.7-pound chunk of this foam fell off during *Columbia*'s launch and punched a hole in the shuttle's wing, allowing superheated gases to flow into the craft on reentry. And in case foam peels off from another part of the external tank, NASA has put sensors in *Discovery*'s wing panels to detect debris impacts and will install a digital camera for viewing the tank after it separates from the shuttle during the ascent.

Other safety steps, however, have proved harder to implement. In its August 2003 report, the Columbia Accident Investigation Board urged NASA to devise ways to inspect the shuttle while it is in orbit and repair any serious damage that is discovered. The space agency began work on a sensor system located at the end of a 50-foot-long boom that would be attached to the shuttle's robotic arm. Equipped with a laser ranger and a television camera, the system is designed to capture three-dimensional images of the reinforced carbon-carbon panels on the shuttle's wings. But a recent report by the Return to Flight Task Group, which is monitoring NASA's progress, said the inspection effort faced "enormous challenges" because of the tight schedule for developing the sensor system. The plans for repairing damage are also incomplete: although the astronauts will be able to fill cracks in the wings and plug holes up to four inches wide, they will not be able to fix a gash as large as the six- to 10-inch breach that doomed *Columbia*.

As a stopgap, NASA is preparing alternative measures that do not require as much technology development. Before *Discovery* docks with the International Space Station, the shuttle's pilot will flip the craft so that astronauts in the station can take photographs of the heat-shield tiles on the shuttle's underside from 600 feet away. The crew members may also conduct space walks to get a closer look. If they find a hole that cannot be fixed, the *Discovery*'s crew will remain in the station until the arrival of the shuttle *Atlantis*, which could be launched on a rescue mission within 45 days. The station should have enough supplies to support the astronauts for that long, assuming there are no breakdowns in critical systems such as oxygen generation or carbon dioxide removal.

Because NASA will rely on the station as a safe haven for shuttle crews, agency director Sean O'Keefe has canceled a planned 2006 mission to service the Hubble Space Telescope, which is expected to cease oper-

HUMAN VS. ROBOT

Robots might not have the touch to fix the Hubble Space Telescope, says Jeffrey A. Hoffman, a former NASA astronaut who participated in the first Hubble servicing mission in 1993. Hoffman says installing new gyroscopes would be tricky because they must be placed near the telescope's center of mass. One strategy would be to incorporate the gyroscopes into the replacement for the Wide Field Planetary Camera (affectionately known as Wiff-Pick) that Hoffman installed a decade ago. "A robot should be able to remove the old Wiff-Pick," Hoffman says. "But when the new Wiff-Pick goes in, it has to be aligned perfectly. With your hands you can maneuver the thing, you can feel the edge. But robots don't have that force-feedback capability. They could jam the Wiff-Pick, which would not be good."

COUNTING DOWN: NASA technicians work near the front landing-wheel well of *Discovery* in preparation for the resumption of shuttle flights next year. Ceramic tiles on the craft's underside protect it from the heat of reentry.

PETER COSGROVE *AP Photo*

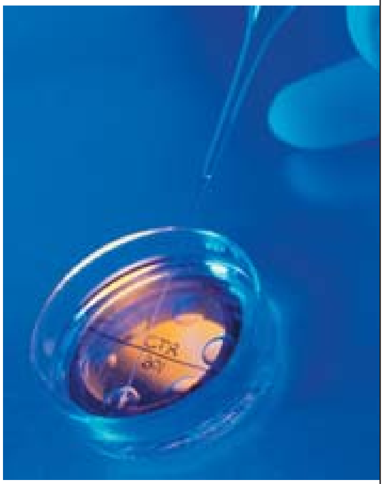
ating by 2008 if its batteries and gyroscopes are not replaced. (Hubble and the station travel in very different orbits, making it impossible for the shuttle to visit both.) Instead O'Keefe gave the go-ahead in August for a robotic mission to install new batteries, gyroscopes and scientific instruments to Hubble, at a cost estimated between $1 billion and $1.6 billion (about three times as expensive as a shuttle mission). But a robotic mission to Hubble might not prove successful, because much new technology needs to be developed, according to a July National Academy of Sciences report.

Ironically, limiting the shuttle to space station flights might actually be riskier than fixing the Hubble. Robert Zubrin, an astronautical engineer who heads the Mars Society, notes that the risk of fatal impacts with micrometeors or man-made debris is much lower during a Hubble mission because the shuttle can be oriented to minimize its vulnerability. (While docked at the station, the craft's belly is exposed.) And finishing the assembly of the station will require at least 25 shuttle flights, whereas repairing Hubble would entail only one.

BIOTECH

Transfer Troubles

CLONING SUCCESS IN ANIMALS DOESN'T EXTEND TO HUMANS **BY CATHRYN M. DELUDE**



DISH-DEFYING: Cloning in primates has been tough, in contrast to other creatures. The photograph shows cloning work with cow embryos.

MAURO FERMARIELLO/SCIENCE PHOTO LIBRARY

Mice and pigs are easy; rats are harder. Humans are harder still, but not as hard as monkeys. No one knows why some species are tougher to clone than others. But studies with mice are helping define those differences and should ultimately enable researchers to achieve greater success in obtaining human embryonic stem cells to treat diseases and replace organs.

In terms of therapeutic cloning (otherwise known as nuclear transfer), the most notable achievement so far comes from Woo Suk Hwang of Seoul National University. In February his lab reported on how it transferred the nuclei of several human donor cells into 242 eggs that had their nuclei removed. The batch produced 30 blastocysts (an early stage of embryonic development), of which only one yielded a self-renewing stem cell line.

Hwang's accomplishment highlights the enormous technical difficulty in working with human cells compared with mouse cells. The human egg is finicky: it is larger, stickier and more fragile than a mouse egg. "It is like a balloon instead of a tennis ball," says Hwang, who, instead of taking the standard approach of removing the nucleus of an egg by aspiration, gently squeezed it out of a tiny slit. He found that, compared with the situation in mice, human DNA reprogramming inside the egg takes twice as long, and embryonic stem cells divide half as fast. And unlike the case with mice, researchers must remove the inner cell mass of human embryonic stem cells from the outer (preplacental) cell layer of the blastocyst, or else the cells stop dividing, explains Kevin Eggan of the Society of Fellows at Harvard University.

Other obstacles arise from lack of experience. "We had to start from the bottom and open every door," Hwang recalls. For want of a better selection, he used a derivative of a culture media meant for growing bovine embryonic stem cells. Limited access to human embryonic stem cells as well as legal and ethical issues also slows progress, explains Douglas Melton of Harvard Medical School, who has made available 17 new embryonic stem cell lines he developed from donated embryos. The "starting material" may also make a difference, Melton suggests. "In humans, we have only worked with frozen embryos, but in mice we work with fresh ones," he says of U.S. researchers.

Other quandaries involve the optimal donor cell for nuclear transfer. The cell should be both accessible and efficient. Hwang used cumulous cells from the ovaries—not exactly easy to access and infeasible for males.

As for efficiency, research suggests that the more immature and undifferentiated a donor cell, the better. The most efficient—and the toughest to obtain—are embryonic stem cells, explains Rudolf Jaenisch, a leading expert in mouse cloning at the Whitehead Institute for Biomedical Research in Cambridge, Mass. Next come the relatively immature but rare and inaccessible somatic stem cells, followed by the more accessible but mature differentiated cells that make up tissues. At the most difficult and inefficient extreme lie the exquisitely specialized immune B and T cells and then the nondividing, highly specialized olfactory sensory neurons, which until recently were considered essentially unclonable. In February, however, Jaenisch, Eggan and Richard Axel of Columbia University successfully cloned mice from olfactory neurons, a feat hailed as a technical tour de force.

Now researchers need to find cells that are more efficient than neurons and more practical than embryonic stem cells. Eggan says experiments with mice indicate that nuclear transfer works best when an egg is ripe for fertilization and ready to divide, and it divides more readily if the transferred nucleus comes from a cell that divides frequently, like an immune cell. Embryonic stem cells, however, more readily develop a self-renewing cell line if the donor cell was quiescent, like a neuron or somatic stem cell. The trick is to find a donor cell that starts the egg dividing and that yields perpetual stem cells.

Melton believes such technical difficulties are just a sign that these are the early days of nuclear transfer. "In a few years, this will all seem easy," he predicts.

Cathryn M. Delude is a freelancer based in Andover, Mass.

A LONG WAY TO GO

The idea of cloning is straightforward: transfer the nucleus of an adult cell to an egg cell that has had its nucleus removed, then harvest the stem cells that are made as the egg starts dividing. But despite successes with other mammals, nuclear transfer (as scientists prefer to call cloning for therapeutic purposes) has proved troublesome and inefficient in humans. Questions about the quality of any harvested stem cells also persist. Researchers also face the challenge of cajoling embryonic stem cells to produce a desired tissue and inducing that tissue to work inside the body [see "The Stem Cell Challenge," by Robert Lanza and Nadia Rosenthal; SCIENTIFIC AMERICAN, June].

PHYSIOLOGY

Heartbeat Poetry

VERSE SPEAKS TO MATTERS OF THE HEART—LITERALLY **BY NICOLE GARBARINI**

Reciting the *Iliad* could have epic effects on your health. German physiologists have recently shown that such poetry can get your heart beating in time with your breaths. This synchronization may improve gas exchange in the lungs as well as the body's sensitivity and responsiveness to blood pressure changes.

Cardiovascular and respiratory responses are not normally in sync. Rhythmic fluctuations in blood pressure take place naturally in 10-second-long cycles known as Mayer waves, whereas spontaneous breathing normally occurs at a rate of approximately 15 breaths per minute.

Dirk Cysarz of the Herdecke Community Hospital and Institute of Mathematics at the University of Witten/ Herdecke wanted to explore the connection between these oscillating mechanisms, which are known to couple weakly at times. The type of poetry was a key aspect of the study. Cysarz and his colleagues specifically used Homer's *Odyssey* translated into German, which maintains the original hexametric pace of the verse—that is, six meters, or rhythmic units, per line. Hexameter is common in ancient epic poems, and both the *Odyssey* and the *Iliad* consist primarily of dactylic hexameter, which has three syllables per meter.

Several factors led the researchers to choose such verse. Cysarz says that one of his collaborators, speech therapist Petric Von Bonin, had extensive experience with this poetry form and felt it would yield the most promising results. Cysarz also cites historical accounts of Greek choruses and audiences gathering to recite more than 10,000 lines of hexameter without pausing. The verse must have produced feel-good effects, Cysarz surmises, "otherwise no one would want to listen to this poetry."

READING POETRY could be a health exercise.

KEVIN FLEMING *Corbis*

For this study, healthy subjects practiced three activities: hexameter reading, controlled breathing at six breaths per minute and spontaneous breathing. They recited while walking, breathing and lifting their arms. The researchers found increased synchronization between heart rate and breathing during the poetry readings but not during the spontaneous breathing. Controlled breathing also boosted synchronization, though not to the extent of recitation. Also, subjects found poetry reading stimulating but controlled breathing boring.

Cardiologist Juan Carlos Kaski of St. George's Hospital Medical School at the University of London has studied the effects of transcendental meditation on cardiac patients. He comments that other variables, such as vocal harmonics or differences in air-intake volume, may have caused the difference between the controlled breathing and hexameter recitation exercises.

Whether this recitation translates linguistically is another question. Rafael Campo, a poet and physician at Harvard Medical School, notes that hexameter in English is not one of the most appealing forms of poetry. He postulates, however, that something inherent in our physiology may have enabled this pattern of poetry to take shape.

Other studies have shown that rhythmic vocal recitations enhance cardiovascular activity. In 2001 physicians in Italy found that subjects who vocalized the *Ave Maria* in Latin or a typical yoga mantra in their original languages inherently slowed breathing to approximately six breaths per minute. Additionally, the poetry reading enhanced the baroreflex response—that is, the cardiovascular mechanism that adjusts to stabilize blood pressure. Increases in this response are a good sign for cardiac patients.

Cysarz wants to study the effects of this poetry recitation in people with hypertension. He expects that it will lower their blood pressure—and that patients will have fun while they're at it.

PHYSICS

Hawking a Theory

IS THE BLACK HOLE INFORMATION PARADOX SOLVED? BY GRAHAM P. COLLINS

Show us the math. That would be one way to sum up the physics community's response to Stephen W. Hawking's headline-making announcement this July that he had solved the black hole information paradox, a profound puzzle of quantum physics and gravity that he himself uncovered 30 years ago. For although he presented an outline of his new theory at the General Relativity 17 meeting in Dublin, he has not yet released a research paper detailing all the mathematical steps that lie behind the general reasoning. "We are waiting for the preprint," says physicist Joe Polchinski of the University of California at Santa Barbara.

The information paradox arises out of two contradictory properties of black holes. Analyzed classically, that is, without taking account of quantum mechanics, anything that falls within the hole's event horizon is lost from the universe forever. Nothing—neither light nor information—is fleet enough to emerge from the hole's intense gravitational clutches. Hawking discovered a quantum chink in this armor in 1974, when he deduced that black holes should in fact emit a random trickle of particles and radiation (now called Hawking radiation).

The problem, underlined by another Hawking paper in 1976, is the randomness, which violates a deep fundamental property of quantum physics known as unitarity. In essence, unitarity states that information should be preserved: if an encyclopedia falls into a black hole and later its mass is emitted as Hawking radiation, the radiation should encapsulate, in principle, all the information that was encoded in the encyclopedia. (Of course, the information will be as irretrievable in practice as if the encyclopedia had been incinerated.) But the randomness of Hawking radiation obliterates the information—an outcome that many physicists found dubious. "I don't believe Hawking's 1976 result, but I don't know what's wrong with the calculation either," says Harvard University string theorist Andrew Strominger.

About 10 years ago physicists discovered a way to analyze black holes using the mathematical technology of string theory. The result appears to show that unitarity, and consequently information, is preserved during black hole evaporation (the process involving strings is intrinsically unitary). Unfortunately, the string calculations do not reveal what is wrong with Hawking's 1976 calculations.

BLACK HOLES don't destroy information, Stephen W. Hawking conceded at a Dublin conference in July.

JOHN COGILL AP Photo

BETTING ON BLACK (HOLES)

At a Dublin meeting this past July, attendees witnessed a ceremonial handing over of a baseball encyclopedia from Stephen W. Hawking of the University of Cambridge to John Preskill of the California Institute of Technology. In 1997 Hawking and Preskill made a bet on the outcome of the information paradox, and Hawking's announcement represented his concession that Preskill had won the bet. A third party in the bet, Kip S. Thorne, also of Caltech, has not yet conceded to Preskill. "I want to see more details of Hawking's analysis," Thorne says. "He is preparing a manuscript on it, which I am eager to study."

Either his argument is incorrect (but physicists call it "simple and beautiful") or one of the theory's assumptions is incorrect (which would be a profound discovery about the nature of the world).

Hawking's new work uses a different technique than did his 1976 analysis and builds on recent research by Princeton University theorist Juan Maldacena. Roughly speaking, Hawking says that unitarity and information are satisfactorily preserved if you wait long enough for the black hole to completely evaporate. Only at intermediate times, when the black hole is still evaporating, is there the appearance of information loss. But the analysis does not address in detail how the information emerges or how the result might be reconciled with either the 1976 analysis or the more recent string theory computations.

And it left many physicists scratching their heads. "I was there in Dublin, did not understand the key point, and neither did the colleagues I talked to," remarks Jacob D. Bekenstein of the Hebrew University of Jerusalem. "It may require years for the experts to reach a consensus," says Kip S. Thorne of the California Institute of Technology. Some express unhappiness about the manner of the announcement before any paper had been written. Others are happy that Hawking now agrees with what string theorists have been saying for a decade, albeit for his own unique reasons. Some, such as Bekenstein, consider the information paradox to have been settled well before Hawking's announcement. "It has been a red herring for years," he states.

Others are not so sure the case is closed, even with the weight of string theory and Hawking combined. Physicist Cumrun Vafa of Harvard thinks that much more work is needed to settle the paradox definitively: "There are so many aspects to the issue of unitarity of black holes, it is hard to believe it would be settled by one single paper—no matter who the author is! The verdict is still not in."

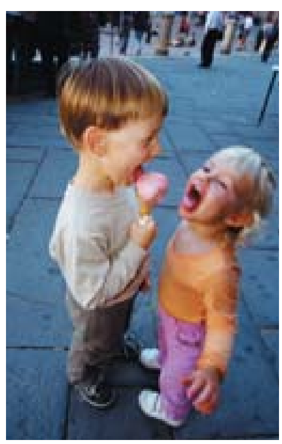
SOCIOLOGY

The Service of Siblings

SOCIALLY SPEAKING, HAVING BROTHERS AND SISTERS MAY BE BETTER THAN BEING AN ONLY CHILD BY LISA DEKEUKELAERE

Siblings may be good for something after all. Decades of research have shown that children with more siblings have no developmental advantage in social skills over only children, but a new study casts doubt on the prevailing wisdom.

Sociologists Douglas B. Downey and Dennis J. Condron of Ohio State University analyzed the social skills of 20,000 kindergarten students and discovered that children with more siblings had more interpersonal skills and better self-control. Moreover, socioeconomic status, parental age and education, race or participation in an after-school program made no difference in the disparity, although biological siblings proved to be more socially beneficial than stepsiblings. Sibling gender was also unrelated to social skills, a surprising observation because research had suggested that a brother disrupts a child's academic performance more so than a sister.


SIBLINGS may be social boosters.

JOHN AND LISA MERRILL *Corbis*

Past research has found few if any significant links between siblings and social skills; this discrepancy may have arisen because earlier studies primarily relied on self-reports from fewer than 1,000 participants.

Downey and Condron, in contrast, drew from teacher reports of behavioral traits such as frequency of disruptive behavior and ability to accept and befriend others. Such reports may offer a more general view of a child's behavior, but not without a possible cost.

Sociologist Denise Polit of Humanalysis, Inc., an independent research consultant based in Saratoga Springs, N.Y., co-authored the largest prior study on the subject. She notes that common stereotypes about only children could have inadvertently influenced a teacher's appraisal. Moreover, previous studies looked at teenagers and adults, not kindergarteners, and social skill deficits at an early age may not carry over into adulthood.

Downey, who believes that his measure of social skill level is one of the distinguishing points of his study, notes that formulating and testing an accurate indicator of social skills can be tricky. This could mean that other studies measured different variables. Heidi Riggio of Claremont McKenna College published a study in 1999 that showed no sibling-related differences in self-reports by adults of their ability to mask emotions and regulate verbal expression. She believes that her study measured different facets of social skill level than Downey and Condron's did—Riggio looked at communication ability, whereas they looked at competence in peer interaction. "The ability to communicate well and accurately does not necessarily reflect that a person is friendly, sociable or well liked," she notes.

The study's data derived from an ongoing survey that will follow the children until fifth grade, and Downey hopes to perform similar analyses on future data to test age patterns. The relevance of the work may grow because of the shrinking size of the average American family: women between the ages of 40 and 44 now have approximately 1.9 children, compared with 3.1 children as measured in 1976.

Downey asserts that these findings alone should not encourage parents to have another child because there are alternative ways to expose a child to peer groups. In addition, previous research has discovered that having fewer siblings can be beneficial. Only children show higher levels of self-esteem, motivation and academic achievement, as well as better parent-child relationships.

MAKING THE GRADES

The study by Douglas B. Downey and Dennis J. Condron of Ohio State University also found that the math and reading skill levels of children with one sibling were similar to those of only children, an unexpected pattern given popular theory on academic performance. Many sociologists believe that additional siblings dilute the resource pool of parental attention, college funds and educational equipment, in which case the achievement of only children should be markedly better than that of children with one sibling. Downey believes this model has merit, but it may require modification in light of his results.

news SCAN

BY THE NUMBERS

Energy Geopolitics

WORLD POWER COULD SHIFT IF NATURAL GAS SUPPLANTS OIL **BY RODGER DOYLE**

Natural gas will become the preeminent fuel of the 21st century, according to widespread expectation. Indeed, reserves of gas already stand at 87 percent of oil reserves in terms of energy content.

If this shift happens, the geopolitical consequences would be momentous, for it would take power away from OPEC and shift it at least partly toward Russia, which, as the map shows, has more gas reserves than any other country. The U.S. relationship with Saudi Arabia, a relatively minor producer of natural gas, would diminish, perhaps to be replaced by new ties with Iran and Qatar, the second and third leading producers of gas.

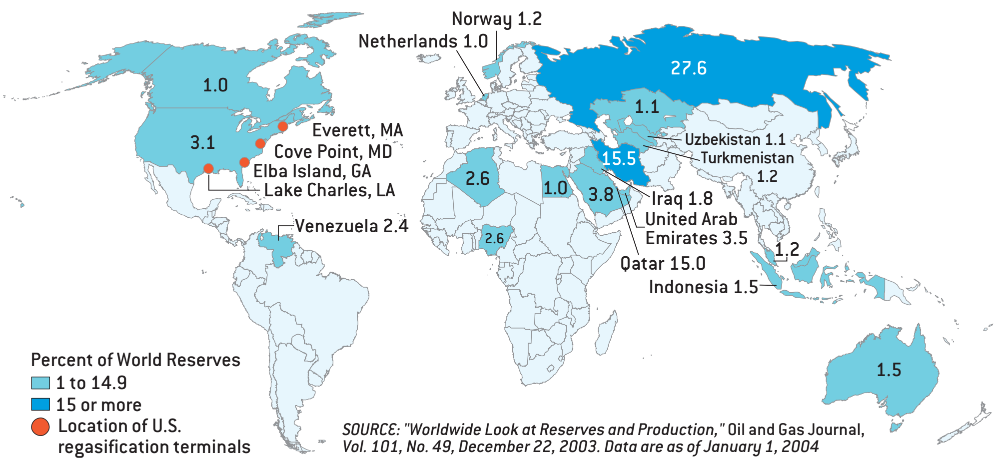
Before this transition happens, a worldwide infrastructure for natural gas, such as that now enjoyed by oil, must emerge. Natural gas is abundant, but it is located in out-of-the-way places, and the system for transporting it is comparatively undeveloped. Unlike oil, it cannot be readily stored and therefore must be transported immediately, either by pipeline or by ocean-going vessel. To go by sea, it must first be liquefied by cooling, which shrinks it to 1/600 of its original volume. The fuel is then loaded on a specially constructed vessel, regasified at its port of destination, and finally piped to where it is needed. This process is expensive: a regasification terminal costs about $4 billion, and pipeline costs at least $1 million to $5 million per mile.

The lack of a well-developed infrastructure helps to explain why the U.S. now has a growing shortage. Domestic production has been more or less flat for several years; rising demand has escalated prices, leading to higher heating bills for consumers and trouble for businesses. For instance, Louisiana's chemical industry depends on natural gas not only for power but also for raw material for its products. The industry there has closed plants and laid off thousands. In Texas, Dow Chemical shut down two plants and is now investing in countries with lower energy costs, such as Kuwait and Argentina.

The obvious solution to the U.S. shortage is to expand the number of liquefied natural gas (LNG) terminals from the current four now in operation. Investors, however, have been skittish because of the volatility of prices in recent years. Fears of an explosion, as happened in Algeria, have led to the blocking of permits for LNG terminals in California and New England. In Baja California, public outcry forced at least one company—Marathon Oil—to cancel its project. The fears of explosion are somewhat exaggerated: LNG, being liquefied, is far less flammable than gasoline. Because it takes three years to build an LNG terminal and because the permit process faces public scrutiny, the shortage of natural gas in the U.S. is unlikely to be resolved soon.

Rodger Doyle can be reached at rdoyle2@adelphia.net

TOP 20 COUNTRIES IN RESERVES OF NATURAL GAS



RODGER DOYLE

FAST FACTS: GASSING UP

Percent of U.S. natural gas consumption in 2003 for:

Industry: 31.9

Residential use: 23.3

Electric power: 22.5

Commercial activity: 14.3

Vehicle fuel: 0.1

Other: 8.0

SOURCE: Energy Information Administration, U.S. Department of Energy

FURTHER READING

The Global Liquefied Natural Gas Market: Status & Outlook. Energy Information Administration, December 2003. www.eia.doe.gov/

International Energy Outlook, 2004. Energy Information Administration, April 2004. www.eia.doe.gov

Matt Simmons Speaks Frankly with Julian Darley. Interview, May 19, 2004. www.globalpublicmedia.com/INTERVIEWS/MATT.SIMMONS

Global Distribution of Natural Gas Resources. Ronald R. Charpentier in *Encyclopedia of Energy.* Elsevier, 2004.

DATA POINTS: MAN ALIVE?

The discrepancy in mortality rates for men and women has been noted and studied since the 18th century. Today masculinity continues to be the single largest demographic risk factor for early death in developed countries, according to Daniel J. Kruger and his colleagues at the Institute for Social Research at the University of Michigan at Ann Arbor. Several factors—including physiological, behavioral, historical and cultural differences—contribute to the disparity.

Number of male deaths in the U.S. per female death by age group:

10–14: 1.50
15–19: 2.37
20–24: 2.94
25–29: 2.51
30–34: 2.06
35–39: 1.82
40–44: 1.76
75–79: 1.46

Number of male deaths per female death before age 50: 1.6

Percent of male deaths from the top four causes of excess male mortality:

Cardiovascular diseases: 26
Suicide: 9
Automobile accidents: 9
Other accidents: 10

SOURCES: Evolutionary Psychology, May 21, 2004; National Center for Health Statistics. Data are from 2000.

BIOLOGY

Whale Today, Gone to Marrow

After a whale dies, most deep-sea scavengers gorge themselves on the sunken carcass and scuttle on. Two new species of tubeworm, however, discovered by marine biologists studying a whale's bones in California's Monterey Bay, are exquisitely adapted to live solely on the bony remains. Crowned in feathery red gills, the females swell with eggs at their base and project green roots that burrow deep into the bone. Symbiotic bacteria in the roots seem to sustain the worms by digesting the bone's rich deposit of oils. Dozens of microscopic male worms cling to the inner surface of a female tube, apparently jockeying for position around the oviduct. The researchers say the worms may represent a sort of Davy Jones's dandelion, casting offspring on the ocean currents to await the next whale that goes belly up. Dig into the July 30 *Science* to extract further details. *—JR Minkel*

BONE MUNCHING: Newly discovered species of tubeworms, with feathery red gills, devour whalebone.

GEOPHYSICS

Surf's Up—Way Up

Bad weather has sunk more than 200 supertankers and container ships during the past 20 years, and researchers believe that monster 10-story-tall ocean waves are often the culprit. But until now, they didn't know whether such "rogue waves" were common on a global scale—a problem for the shipping industry, which designs ships to withstand waves only about half the size, about 15 meters. German and U.S. researchers analyzed three weeks' worth of global satellite radar data and identified 10 separate rogue waves taller than 25 meters. The team, whose results were announced by the European Space Agency on July 21, is now compiling a two-year, worldwide atlas of freak waves, which the scientists could compare with locations of sunken ships and possibly use to develop forecasting models of monster surf. They have already noticed that some rogue waves occur where multiple high-energy wave systems intersect. *—JR Minkel*

PSYCHIATRY

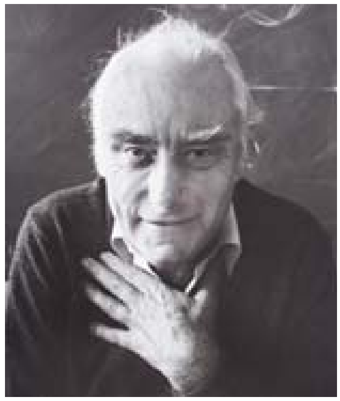
Fever in, Schizophrenia Out

Researchers had suspected that exposure to influenza during pregnancy raises the risk of offspring developing schizophrenia later in life but were dubious because the connection was based on dates of influenza outbreaks or mothers' recollections. Now a group at Columbia University and the New York State Psychiatric Institute has measured flu antibodies in 40-year-old blood samples from the mothers of schizophrenics and unaffected individuals. Those exposed to flu during the first half of pregnancy were three times more likely to develop schizophrenia, but those exposed during the second half had no increased risk, they found. (The average risk of schizophrenia is no more than 1 percent.) If the link is confirmed, further work might lead to preventive strategies against some cases of schizophrenia, says epidemiologist and senior author Ezra Susser, although he cautions that the evidence, based on only 200 offspring, is not yet solid enough for women to act on. Subtle brain damage from an inflammatory or autoimmune response to viral infection could be contributing to the disease, he speculates. See the August *Archives of General Psychiatry* for more. *—JR Minkel*

MONTEREY BAY AQUARIUM RESEARCH INSTITUTE, 2003; ILLUSTRATION BY MATT COLLINS

IN MEMORIAM

Francis Crick, 1916–2004



FRANCIS CRICK in 1979.

Science appealed to young Francis Crick the most, leading him at an early age down the road to a life in research. He feared one snag—by the time he grew up, everything would have been discovered. "I confided my fears to my mother, who reassured me, 'Don't worry, Ducky,'" Crick remembered in his 1988 memoir *What Mad Pursuit*. "'There will be plenty left for you to find out.'" The most renowned discovery of Crick's life is undoubtedly that of the double-helix structure of DNA, reported in his 1953 *Nature* paper with James Watson. This breakthrough, which earned them a Nobel Prize in 1962, sparked a revolution in biology whose impact has only just begun, from insights into the origins of species to research transforming the future of our genetic blueprint. His endeavors ranged far beyond genetics—during World War II, Crick designed magnetic and acoustic mines for the British Admiralty, and he devoted his later life to a scientific exploration of consciousness [see "The Puzzle of Conscious Experience," by Francis Crick and Christof Koch; Scientific American, September 1992]. Crick died in July of colon cancer in San Diego. *—Charles Choi*

PHYSICS

The Phantom Menace

Wormholes are theoretical tunnels in spacetime that often provide cosmic shortcuts for stories with faster-than-light travel. Now it seems that wormholes could provide fodder for end-of-the-universe tales, too. The end would come because of mysterious dark energy, thought to make up 70 percent of the universe and permeate all of space. It may exert a repulsive force that could explain why the universe's expansion is accelerating. One theorized version of dark energy is phantom energy, which grows more repulsive over time. As phantom energy builds up in a wormhole, it would force the wormhole's throat open, says theoretical physicist Pedro González-Díaz of the Institute of Mathematics and Fundamental Physics in Madrid. Eventually, the wormhole would grow to engulf the entire universe and become infinite in scope. Oddly, because wormholes allow time travel, the entrapped universe could go back to relive its past or forward into its future. The paper appears in the August 13 *Physical Review Letters*. *—Charles Choi*

WORMHOLE could eat the universe.

AQUACULTURE

Trout from Salmon

Researchers at Tokyo University of Marine Science and Technology got East Asian masu salmon (*Oncorhynchus masou*) to sire rainbow trout (*Oncorhynchus mykiss*). The scientists extracted primordial germ cells from newly hatched trout embryos and implanted them into 60 salmon embryos. Ultimately, 10 salmon incorporated the trout germ cells, which produced normal trout sperm and eggs. Growing germ cells in surrogate parents could significantly reduce the time, cost, rearing space and labor required to raise such fish, write the investigators in the August 5 *Nature*, if the parents were smaller and matured faster. For instance, male rainbow trout reach sexual maturity two years after hatching, whereas masu salmon males mature in one. *—Charles Choi*

ARNOLD NEWMAN *Getty Images* (*top*); MARK GARLICK/SCIENCE PHOTO LIBRARY (*bottom*)

BRIEF POINTS

- **Bug-free living is not healthy: fruit flies exposed to bacteria early in adulthood lasted up to 35 percent longer than those raised in sterile environments, although exposure in later life slightly decreased longevity.**

  Proceedings of the National Academy of Sciences USA online, August 16, 2004

- **The stuff of Asian myth and legend, a white elephant was spotted in Sri Lanka in mid-July. The 11-year-old female appears to be the first recorded sighting of a wild albino pachyderm.**

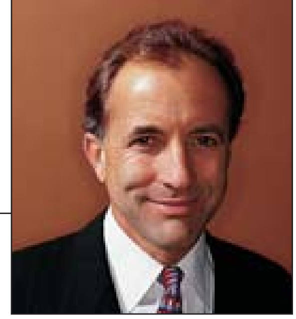
  www.wildlifetrust.org

- **Scanning tunneling microscopes can move individual atoms; now researchers have managed to use the STM to remove and replace electrons, thereby converting an atom into an ion and back again.**

  Science, July 23, 2004

- **Nicely scented and environmentally safe, cinnamon oil might protect against mosquitoes: it has proved more lethal to mosquito larvae than DEET, the most commonly used repellant.**

  Journal of Agricultural and Food Chemistry, July 14, 2004

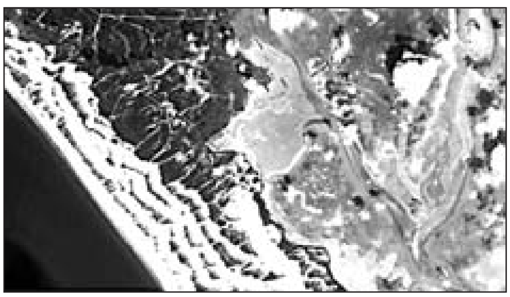


The Myth Is the Message

Yet another discovery of the lost continent of Atlantis shows why science and myth make uneasy bedfellows By MICHAEL SHERMER

Myths are stories that express meaning, morality or motivation. Whether they are true or not is irrelevant. But because we live in an age of science, we have a preoccupation with corroborating our myths.

Consider the so-called Lost Continent of Atlantis, a mythic place that has been "found" in so many places around the planet that one wouldn't think there was anywhere left to look. Think again. On June 6 the BBC released a story about satellite images locating Atlantis in, of all places, the south of Spain (http://news.bbc.co.uk/2/hi/science/nature/3766863.stm). The story quoted Rainer Kühne of the University of Dortmund in Germany as saying, "Plato wrote of an island of five stades (925 m) diameter that was surrounded by several circular structures—concentric rings—some consisting of Earth and the others of water. We have in the photos concentric rings just as Plato described."



CAN YOU SEE the concentric rings of Atlantis?

Kühne reported his findings in the online edition of the journal *Antiquity,* claiming to have identified two rectangular structures surrounded by concentric rings near the city of Cádiz, Spain. He suggests that the structures match the description in Plato's dialogue *Critias* of the silver and golden temples devoted to the Greek god Poseidon and his mortal lover Cleito and that the high mountains of Atlantis are actually those of the Sierra Morena and Sierra Nevada. "Plato also wrote that Atlantis is rich in copper and other metals," he adds. "Copper is found in abundance in the mines of the Sierra Morena."

Atlantis also has been "found" in the Mediterranean, the Canaries, the Azores, the Caribbean, Tunisia, West Africa, Sweden, Iceland and even South America. But what if there is nothing to find? What if Plato made up the story for mythic purposes? He did. Atlantis is a tale about what happens to a civilization when it becomes combative and corrupt. Plato's purpose was to warn his fellow Athenians to pull back from the precipice created by war and wealth.

In a second Plato dialogue, *Timaeus,* Critias explains that Egyptian priests told the Greek wise man Solon that his ancestors once defeated a mighty empire located just beyond the "Pillars of Hercules" (usually identified by Atlantologists as the Strait of Gibraltar), after which "there were violent earthquakes and in a single day and night all sank into the earth and the island of Atlantis in like manner disappeared into the depths of the sea." Critias describes the city as a series of circular canals lined with colorful palaces adorned in gold. Poseidon resided in a silver temple with an ivory roof, and a racecourse was built between the canals. Atlantean wealth afforded a military industrial complex of 10,000 chariots, 24,000 ships, 60,000 officers, 120,000 hoplites, 240,000 cavalry, and 600,000 archers and javelin throwers. (Your myth-detection alarm should be going off about now.) Corrupted by excessive belligerence and avarice, Zeus called forth the other gods to his home, "and when he had gathered them there he said. . . . " The sentence ends there. Plato had made his point.

The fodder for Plato's imagination came from his experiences growing up at the terminus of Greece's golden age, brought about, in part, by the costly wars against the Spartans and Carthaginians. He visited cities such as Syracuse, which featured numerous Atlantean-like temples, and Carthage, whose circular harbor was controlled from a central island. Earthquakes were common: when he was 55, one leveled the city of Helice, only 40 miles from Athens, and, most tellingly, the year before he was born an earthquake flattened a military outpost on the small island of Atalantë.

Plato wove historical fact into literary myth. As he wrote of his parables: "We may liken the false to the true for the purpose of moral instruction." The myth is the message. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* The Science of Good and Evil.

BRAD HINES (*top*); SATELLITE IMAGE BY EUROMAP, COURTESY OF WERNER WICKBOLDT (*bottom*)

Father of Spirit and Opportunity

With the success of twin rovers on the Red Planet, Steven W. Squyres and his team are showing how to conduct robotic missions—and setting the stage for human exploration By DAVID APPELL

It's 8 A.M. at NASA's Jet Propulsion Laboratory, and the Mars Opportunity rover team is gathering in a conference room for its daily science kickoff meeting. For "Martian sol 149"—the 149th day on the Red Planet since the start of the rover missions—it is assembling the minute-by-minute plan of what the rover will do: A little spectroscopy. "Ratting" a few rocks—that is, drilling them with the rover's Rock Abrasion Tool, or RAT. "I'm interested in knowing if this stuff is red or not," says Steven W. Squyres of the rocks that the rover is currently rolling over.

FOREST McMULLIN

STEVEN W. SQUYRES: MAN FOR MARS

- As a graduate student, worked with Joseph Veverka and Carl Sagan of Cornell University on the Voyager flybys of Jupiter, earning a Ph.D. in three years.
- On Martian rovers Spirit and Opportunity: "They had their own personality quirks even when they were babies."
- On going to the Red Planet: "Human exploration of Mars can't happen soon enough as far as I'm concerned."

A professor of astronomy at Cornell University, Squyres, 48, is the principal investigator for the Mars Exploration Team, which consists of 170 members. He is responsible for all the scientific activities of both the Opportunity and Spirit rovers, leading colleague John Grotzinger to liken him to a "flea on a hot griddle," with his hands in everything.

On this late June day in 2004, the Opportunity team is facing a critical decision about whether to go farther down into Endurance Crater, a 20-meter-deep hole near the landing site on the Meridiani Planum. The team members have been eyeing the crater for weeks. But first they have to figure out if Opportunity can negotiate an inclined rock step right in front of it that drops 30 centimeters along a 25-degree incline. The step is slightly above the rover's design limits, but everyone on the team wants to do it if they can.

"Welcome to the monster truck convention," jokes Opportunity mission manager Matt Wallace at a 10 A.M. traversability meeting. A test rover at the JPL Mars yard successfully went up and down a mock-up of the step; slight slippage and a minor wheelie were the only hitches. After some discussion, the engineering team decides to go ahead and send Opportunity down the step. They will drill several RAT holes on the way in case the rover cannot climb back up, Squyres says.

Departing from the way it normally does business, NASA entrusted the science package of the Mars rovers to a single individual. And unlike many earlier principal investigators, Squyres was heavily involved in the design and engineering of the rovers. In fact, he has been working on them in one form or another since 1987, when he developed the panoramic camera now in use on the rovers. In 1992 he assembled the RAT as well as spectrometers to measure thermal emissions and to detect iron and other chemicals in rocks. Then he spent the next few years writing proposals for the rover project. NASA selected Squyres and his team in 1997, and the Mars Exploration Rover (MER) project came together in its final form in July 2000 [see "The Spirit of Exploration," by George Musser; Scientific American, March].

The Spirit and Opportunity landers have proved to be the most complex robotic mission NASA has ever attempted. They have performed beautifully, operating for at least twice their projected lifetimes. Barring mechanical failure—and Spirit has lately been dragging around a bad wheel—the rovers' lives are limited by the buildup of dust on their solar arrays, which cuts the electricity being generated. (Over both rovers, the power loss is about 0.15 percent per sol, slightly less than the 0.18 percent loss per sol of Pathfinder, the rover that landed on Mars in 1997.) With Mars now entering winter, the arrays are taking longer to recharge the batteries, which will most likely result in extended sleep cycles for the vehicles. But there is hope the rovers will survive the Martian winter.


MODEL ROVER is displayed at Cornell University. Working duplicates enable maneuvers to be tested on Earth first.

The Mars rovers are about 20 light-minutes from Earth, too far for any type of real-time joystick control. So they are programmed to run one Martian day at a time between code uploads. In the afternoon Squyres sits with the computer jocks who translate the agreed-on strategy into the precise series of wheel and arm movements that will accomplish the tasks. It is tight, detailed code that calls for an in-depth understanding of the vehicles—the kind of difficult day-to-day engineering that Squyres has lived through since the rovers landed in January 2004. Although the rovers have performed beyond anyone's expectations, some small problems have arisen: a heating-element switch remains stuck open on Opportunity, and Spirit had software problems just after emerging from its landing cocoon.

But neither problem has detracted from the science that's been done. "It's fair to say that the rovers would certainly have not been as successful as they are and possibly would never even have happened if it wasn't for Steve," says Jim Bell, lead scientist for the panoramic cameras on both rovers and a close colleague of Squyres' at Cornell. To him, Squyres has set the example for those leading future missions. You not only "have to be a top-notch scientist," Bell states, "but you also have to be willing to get head over heels into the design of the instruments."

The most important discovery is the evidence for great amounts of water at Meridiani Planum. Opportunity detected, for instance, sulfate salts and hematite concretions—small, grayish, iron-containing spherules that scientists have been calling "blueberries." And as Squyres had wondered, shavings of the ratted rock in the crater were indeed brick-red, typical of hematite when it is pulverized. Sulfate and hematite are left in rocks by water, so they suggest that Opportunity is working on what was once the shoreline of a salty sea, although clues gathered so far do not indicate how long, or how long ago, liquid water covered the area. "Not only did we find evidence for a habitable environment at Meridiani," Squyres remarks, "but we've got these wonderful geologic deposits—sulfates and in particular the hematite concretions—that are very good at preserving evidence of whether there was interesting organic chemistry, whether there was life."

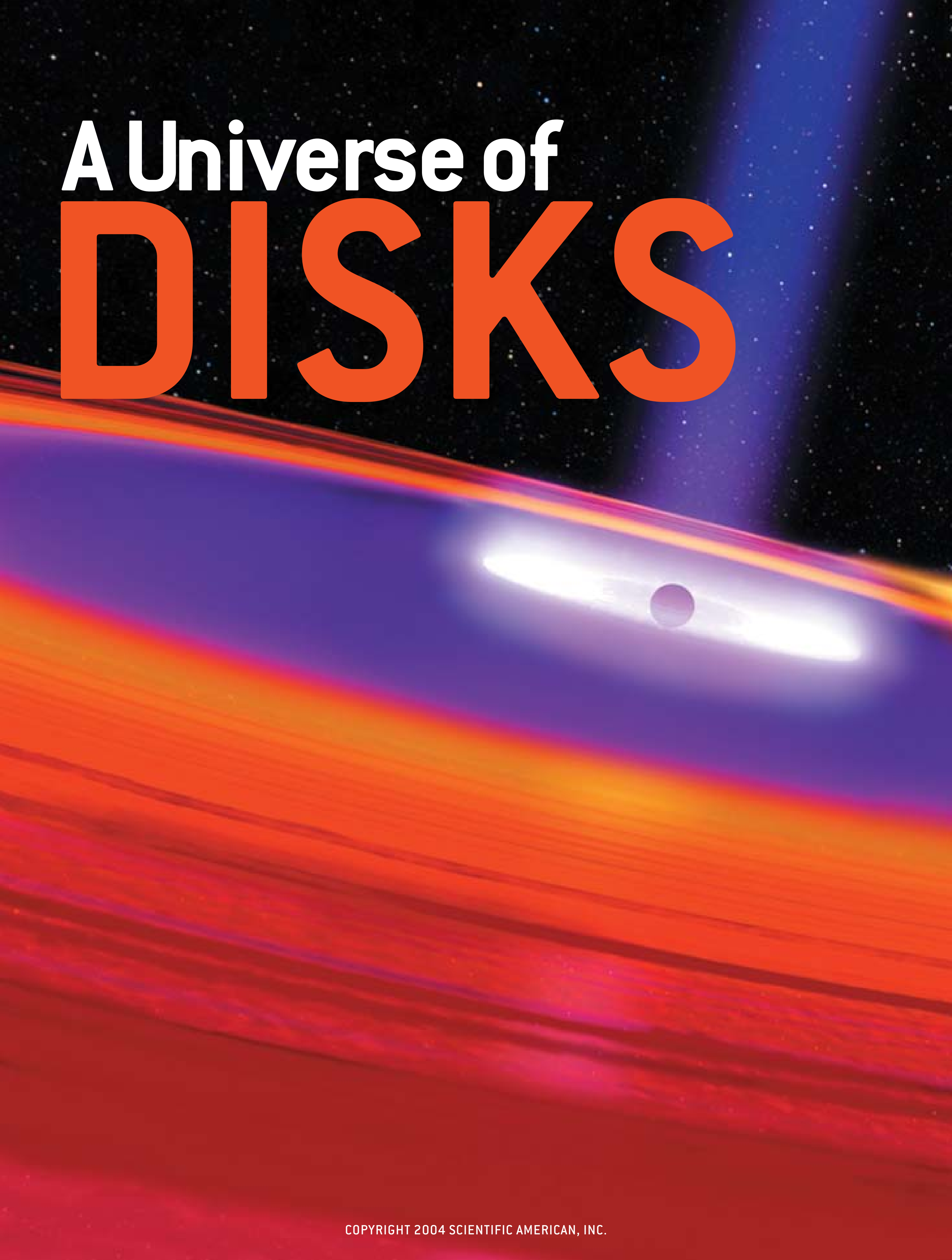
It's unlikely the rovers themselves will directly discover signs of life, though. Instead they are laying the groundwork for a sample-return mission, by robots or someday by humans. Squyres is all in favor of a manned mission to Mars. "I'm a huge fan of sending robots to Mars, obviously—that's what I do for a living. But even I believe that the best exploration, the most comprehensive, the most inspiring exploration is going to be conducted by humans."

Although some argue that NASA's many successful robotic missions prove that costly human flights are unnecessary, Squyres doesn't buy it. "I think people who would point to the successes of these two rovers as evidence that you don't need to send humans to Mars are missing the point entirely. I view our rovers not as competitors to humans but as precursors. And they're precursors in the sense of telling us more about the Martian surface and what it's like, what it takes to walk across it, build on it, launch from it, that kind of thing." A human base on Meridiani Planum, now being explored by Squyres and his MER team, would be a good place to start. SA

David Appell is based in Newmarket, N.H.

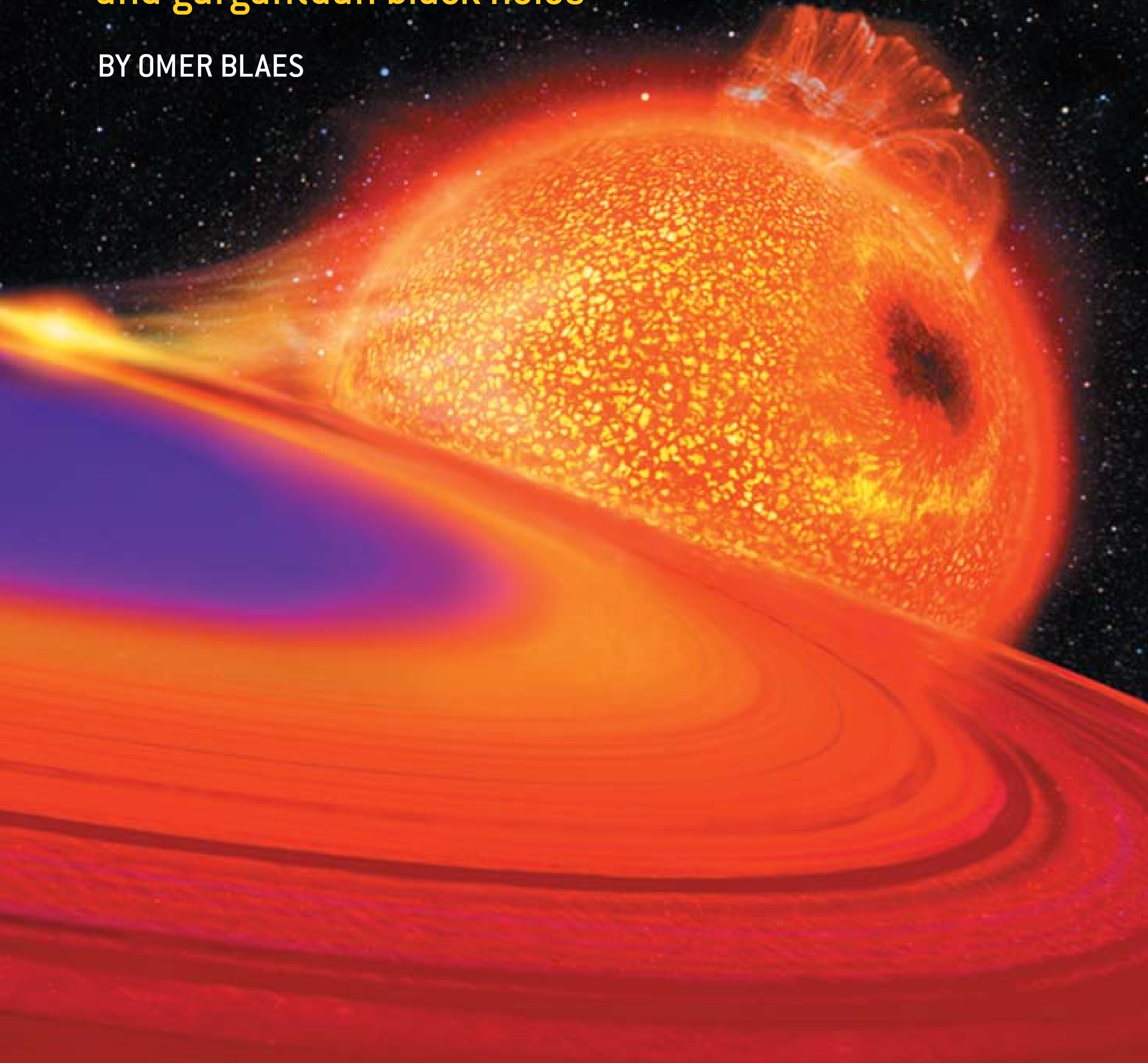
FOREST McMULLIN

A Universe of DISKS



New research reveals the dynamics of the spinning disks of gas that surround young stars and gargantuan black holes

BY OMER BLAES



ACCRETION DISK surrounds a black hole in this artist's rendering of an x-ray binary system. The powerful gravity of the black hole pulls gas off the companion star, a red giant. The disk of gas emits copious amounts of x-ray radiation as it spirals inward toward the black hole. The disk also generates jets of particles that shoot from the innermost region.

Look up at the sky on a clear night

and try to find some of the planets that are visible to the naked eye—Mercury, Venus, Mars, Jupiter or Saturn. If you locate three or more, you will see that they all appear to line up within a fairly narrow band that forms a great circle around the sky. This band includes the ecliptic, the path of the sun's apparent movement through the constellations of the zodiac over the course of a year. And if you now focus on the fuzzy trail called the Milky Way, you will notice that it traces a different great circle across the sky.

These observed geometric facts are not accidental. The planets of our solar system, including Earth, revolve around the sun in the same direction and (apart from Pluto) in nearly the same plane. This arrangement is strong evidence that the planets formed from a pancakelike disk of material (mostly gas and dust) that orbited the early sun. Similarly, the appearance of the Milky Way—which is a hazy agglomeration of the light from many billions of stars—shows that our galaxy is also disklike in shape. Because our solar system is situated within this disk, our galaxy appears to encircle us.

Structures in the shape of disks are common in the universe on a vast variety of scales. Saturn's rings are a graceful local example, but not the only one; all the giant planets in our solar system have rings. Disks have also been observed around many young stars; astronomers often call them protoplanetary disks because they appear to be similar to the one that must have formed our own solar system. In some binary star systems, gas escapes from one star and is captured by the gravity of the other to form a disk. Inside the disk, the gas slowly works its way down to the stellar surface in a tight, spiraling motion like a whirlpool. Such structures, called accretion disks, are also thought to exist around supermassive black holes (which can weigh as much as a billion suns) at the centers of galaxies. The largest disks are spiral galaxies such as our own Milky Way, which stretches more than 100,000 light-years across.

Given the ubiquity of disks in the universe, understanding how they work is an important problem in astrophysics. Astronomers believe that accretion disks around supermassive black holes may have influenced the way galaxies formed and evolved. And exploring the dynamics of accretion disks around young stars may shed some light on the early history of our own solar system. Thanks to new theoretical insights and modern computer simulations, scientists have recently discovered an explanation for the roiling turbulence of accretion disks that makes them powerful energy sources. But other phenomena, such as the jets of particles that often stream from the disks, remain a mystery. Researchers still have much to study in the billions of whirling disks that populate our universe.

Overview/*Accretion Disks*

- Because disks of gas are so ubiquitous in the universe, appearing around newly born stars, in binary star systems and at the centers of galaxies, astrophysicists are keenly interested in learning their dynamics.
- To explain the radiation from the disks, scientists have long assumed they must be turbulent. Physicists believe that a phenomenon called magnetorotational instability is causing the turbulence.
- Researchers are now exploring how this phenomenon works in different kinds of accretion disks.

The Celestial Merry-Go-Round

A DISK'S ROTATION holds it up against gravity. Picture yourself on a merry-go-round that is spinning dangerously fast. If you do not hold on tightly to one of the painted horses, you will be flung in a straight line tangent to the merry-go-round's circle. The tension in your arm provides exactly the force needed to cause your body to move in a circle and therefore to stay on the ride. The rotation of the merry-go-round prevents you from moving inward; you would need to make a tremendous effort to pull yourself toward the center. In the same way, the rotation of the material in a disk prevents it from collapsing inward under the force of gravity.

Rotating objects are endowed with angular momentum, a quantity that is proportional to the object's rotation rate and the distribution of mass around the axis of rotation. (The farther the mass is from the axis, the greater the object's angular momentum.) Angular momentum is central to our understanding of the behavior of rotating systems because, like energy, it is conserved: it can be neither created nor destroyed. A twirling ice skater, for example, can spin faster by pulling in her arms. Because her angular momentum must remain constant, the movement of mass closer to the rotation axis of her body is offset by an increase in her spin rate.

The conservation of angular momentum explains why disks are so prevalent in the universe. Consider a cloud of gas that is contracting inward under the attractive power of its own gravity. Almost everything in the universe rotates at some level, so suppose this cloud has some angular momentum. As it contracts, the principle of angular momentum conservation forces it to rotate faster [*see box on opposite page*]. Material in the equatorial region of the cloud—that is, in the plane perpendicular to the rotation axis—moves inward more and more slow-

DON DIXON (*preceding pages*)

ly as rotation starts to balance the pull of gravity. Material along the rotation axis falls vertically toward the equatorial plane much faster. The resulting object is a rotationally supported disk.

Scientists believe this process explains how protoplanetary disks form around young stars and perhaps how gas disks coalesce around black holes at the centers of galaxies. Whether an entire galaxy becomes a disk is a timing issue: spiral galaxies emerge from gas that becomes rotationally supported before patches of the gas contract into stars. If stars are born in the gas before the galactic cloud becomes rotationally supported, the stars will maintain their individual orbits around the galactic center, creating an elliptical galaxy. In general, galaxies do not form in isolation, and galactic collisions and mergers complicate matters considerably. At least some elliptical galaxies, as well as the bulges and halos of spiral galaxies, may have arisen from such collisions.

Accretion disks also form in binary star systems when one of the stars (for example, a compact, dense white dwarf) gravitationally pulls gas off its companion (usually a larger, less compact star). This gas has considerable angular momentum from the original orbital motion of the two stars around their common center of mass, so it typically cannot fall directly inward toward the white dwarf. Instead the gas ends up forming a disk around the dwarf.

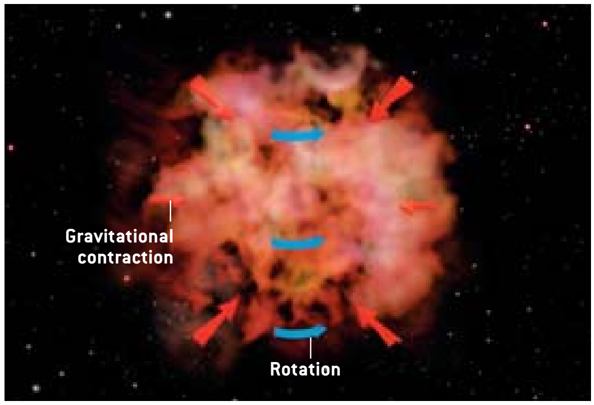
Just as Mercury has a much shorter year than Earth—a mere 88 days—the material in the inner parts of a disk invariably takes less time to complete one orbit than does material in the outer parts. This gradient in orbital periods causes shear: bits of material at slightly different distances from the center of the disk slide past one another [*see box on page 54*]. If some form of friction is present in the disk material, it tries to slow down the more rapidly orbiting inner regions and speed up the more slowly orbiting outer regions. Angular momentum is therefore transported from the inner to the outer regions of the disk. As a consequence, material in the inner regions loses rotational support against gravity and falls inward. The overall result is a gradual spiraling of matter toward the central star or black hole.

As material spirals down to the innermost orbit of an accretion disk, it must give up gravitational potential energy. Some of the potential energy goes into giving the material the faster orbital speed it gains as it falls inward; the rest is dissipated into heat or other forms of energy by the friction itself. Thus, the material in the disk can become very hot, emitting copious amounts of visible, ultraviolet and x-ray radiation. The energy release can make accretion disks formidable power sources.

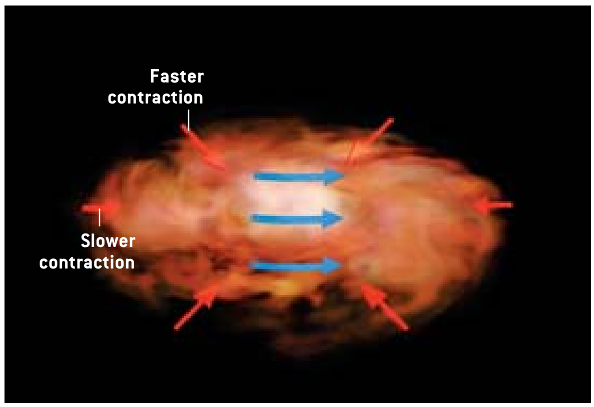
This phenomenon is what first alerted astronomers to the existence of black holes. Black holes themselves cannot emit light, but the accretion disks around them can. (This general statement ignores the theorized Hawking radiation, an emission that would be undetectable for all but the smallest black holes and that has not yet been observed anywhere in the universe.) According to Einstein's general theory of relativity, the energy released by an accretion disk around a black hole should be equivalent to roughly 10 percent of the material's rest-mass energy

BIRTH OF A DISK

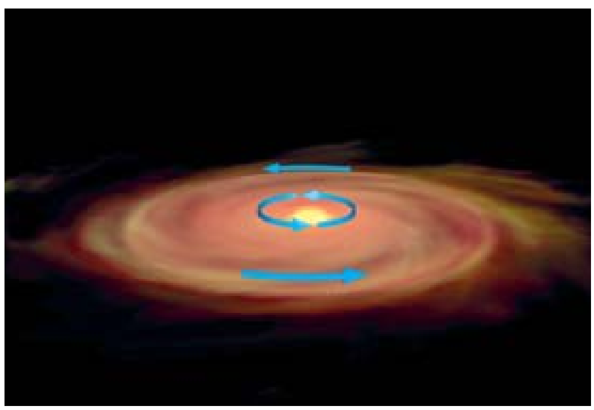
The principle of angular momentum conservation explains why disks are so common in the universe. Angular momentum is proportional to an object's rotation rate and the distribution of mass around its axis of rotation. Because angular momentum must be conserved, a rotating object that is also contracting must spin faster as it shrinks.



A cloud of interstellar gas is rotating slowly around its axis and contracting because of the attractive pull of its own gravity. As the cloud collapses, it rotates faster.



The gas in the cloud's equatorial plane moves inward more slowly because its rotation starts to balance the gravity. Gas above and below the plane falls inward much faster.



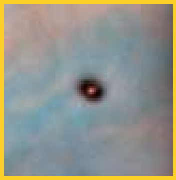
Over time, all the material in the cloud falls into the equatorial plane, where the gas becomes rotationally supported—its motion holds it up against gravity.

ALFRED T. KAMAJIAN

A Gallery of Disks and Jets

Astronomers have observed disks across the universe—around young stars in nebulas in our own galaxy and at the centers of galaxies millions of light-years away. Many of the disks emit long jets of particles in a process that is still not well understood.

Protoplanetary Disk

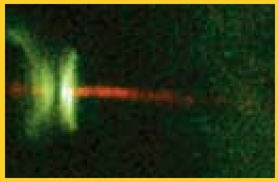


In the Orion nebula, about 1,500 light-years from Earth, a protoplanetary disk surrounds a star that is only one million years old. The disk is about 40 billion kilometers across (three times the size of our solar system) and is composed of 99 percent gas and 1 percent dust. As the disk evolves, it may form a planetary system like our own.

Spiral Galaxy

NGC 7331, a spiral galaxy about 50 million light-years from Earth, is a disk just like our own Milky Way galaxy. Data from the Spitzer Space Telescope, a new observatory that looks at infrared radiation, indicate the presence of a supermassive black hole in the galaxy's core.

Jet from a Nascent Star



HH-30, a newborn star about 450 light-years from Earth, is embedded in a protoplanetary disk (*viewed edge-on at left*). Two jets of gas stream in opposite directions from the center of the disk, moving as fast as 960,000 kilometers per hour. The star's magnetic field may be channeling the gas.

Jet from an Active Galaxy

The active nucleus of M87, a giant elliptical galaxy about 50 million light-years from Earth, is emitting a jet of high-speed electrons that stretches 6,500 light-years from the galaxy's core. An accretion disk spinning around a supermassive black hole is putting most of its power into the jet.

(which is equal to its mass times the speed of light squared). This amount is spectacularly high, more than 10 times as great as the energy that would be released if the material underwent thermonuclear reactions, such as occur in stars or hydrogen bombs. And yet this prediction agrees with observations of the radiation from quasars, highly luminous objects that are believed to be powered by accretion disks around supermassive black holes in the centers of early galaxies. When one calculates the total energy radiated over time by all the quasars in a given region of space, it turns out to be about 10 percent of the mass of all the supermassive black holes currently observed in an equivalent region times the speed of light squared.

Turbulence in Space

BUT WHAT IS THE NATURE of the friction inside the accretion disks that touches off this enormous energy release? One possibility is that the particles that make up the material in the disk undergo collisions in which they exchange small amounts of energy and angular momentum. This mechanism operates in Saturn's rings: as the pebbles, rocks and boulders that make up the rings collide, their energy is lost as heat, and angular momentum is transferred outward. Ordinary fluids act in much the same way; in fact, Saturn's rings can be thought of as a viscous fluid in which the colliding molecules are actually rocks! The collisions give the rings a tendency to spread radially, but Saturn's moons act as reservoirs of angular momentum that keep the rings confined.

Unfortunately, this simple process cannot explain the activity of many other types of accretion disks. In the accretion disks in binary systems or at the centers of galaxies, particle collisions would produce an inflow of mass that is too small by many orders of magnitude to produce the brilliant luminosity of these disks. Another possibility is that large-scale spiral waves in the disk, similar to the spiral arms observed in galaxies, hasten the inflow of matter. Just as sound waves transport energy through the air, spiral waves can transport both energy and angular momentum outward and facilitate the accretion of material inward. And astronomers, in fact, have seen evidence of spiral-wave patterns in accretion disks in some binary systems. But the spiral waves in these systems do not appear to be large enough to produce the rate of matter inflow needed to explain the observed radiation from the disks.

Many astrophysicists believe, however, that the most widespread mechanism for friction inside accretion disks is turbulence, which would accelerate the inflow of matter by generating violent, large-scale collisions. When water flows in a pipe, the viscosity of the liquid causes the flow speed to be highest at the center of the pipe and lowest near the pipe's inner surface. If the water is forced to move faster, this velocity shear becomes larger and eventually destabilizes the flow, making it turbulent and chaotic. Because accretion disks also contain flows with very high rates of shear, scientists proposed in the 1970s that the disks must also be highly turbulent. But when researchers tried to demonstrate this phenomenon using the basic equations of

THE AUTHOR

OMER BLAES has long been intrigued by the dynamics of accretion disks. A professor of physics at the University of California, Santa Barbara, he earned his Ph.D. in 1985 from the International School for Advanced Studies in Trieste, Italy. He then did postdoctoral research at the California Institute of Technology and the Canadian Institute for Theoretical Astrophysics in Toronto. Blaes is a theorist who works in the area of high-energy astrophysics; in addition to accretion disks, he is particularly interested in the physics of compact objects such as black holes, neutron stars and white dwarfs.

M. J. MCCAUGHREAN *MPIA*, C. R. O'DELL *Rice University*, NASA, ESA (*protoplanetary disk*); C. BURROWS *STScI*, J. HESTER *Arizona State University*, J. MORSE *STScI*, NASA (*HH-30*); M. REGAN ET AL. *STScI*, JPL, CALTECH, NASA (*NGC 7331*); STSCI, NASA (*M87*) fluid flow and computer simulations, they found no indication that turbulence would develop in an accretion disk.

The reason for this negative result is still controversial. It could be that the computer simulations are somehow faulty, but it is also possible that the analogy with pipe flow is simply incorrect and that rotating systems like accretion disks are intrinsically different. Investigators have conducted laboratory experiments to search for turbulence in flows that resemble accretion disks, but again the results have been debated. Although turbulence is sometimes seen in these experiments, it may arise from effects that would not be present in a real accretion disk.

Nevertheless, astrophysicists have persisted in their belief that accretion disks are indeed turbulent. Proceeding under this assumption, most researchers have adopted a crude mathematical guess for the effects of disk turbulence that was introduced in 1973 by Soviet physicists Nikolai Shakura and Rashid Sunyaev. By cutting the Gordian knot in this way, astrophysicists have been able to build theoretical models of accretion disks to compare with observations of actual disks. This research program has achieved several successes over the years. For example, some accretion disks in binary star systems occasionally undergo large, temporary increases in luminosity. (Dwarf novae, explosions of light from the accretion disk around a white dwarf in a binary system, are examples of this phenomenon.) Scientists have convincingly demonstrated that these transient increases are triggered by an instability in the disk that causes material to flow rapidly inward.

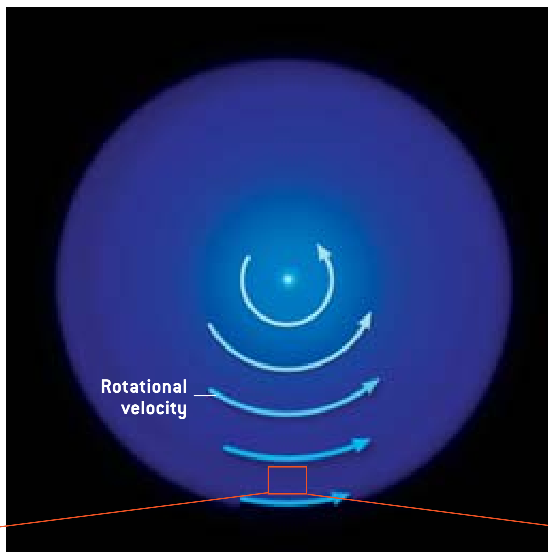
Despite these accomplishments, however, the Shakura-Sunyaev approach has really just concealed our ignorance. Discrepancies between model predictions and observational data might arise simply because the widely accepted guess for disk turbulence is wrong. In addition, the turbulence might have observable consequences besides providing angular momentum transport in the disk, but researchers are unable to predict these consequences without understanding the process behind them.

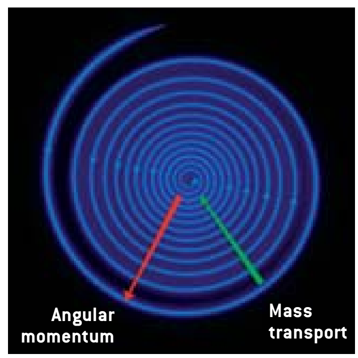
An Astronomical Racetrack

FORTUNATELY, A SPECTACULAR breakthrough in the turbulence problem came in 1991. Steven Balbus and John Hawley of the University of Virginia realized that if the material in the accretion disk was highly electrically conductive and magnetized, even if only weakly, then the magnetic field would pro-

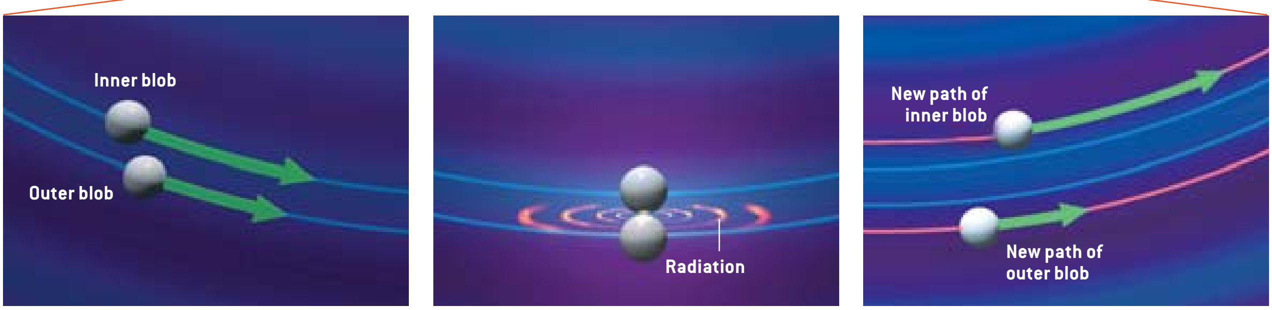
HOW A DISK SPIRALS AND RADIATES

Material in the inner parts of an accretion disk takes less time to complete an orbit than does material in the outer parts (*right*). Bits of material that are closer to the center of the disk slide past material that is slightly farther out. In an accretion disk around a star or black hole, large-scale blobs of gas collide violently in a turbulent flow (*bottom*). This process transports angular momentum outward, causing some of the gas to lose rotational support and spiral inward (*far right*). And because the collisions make the material very hot, the disk radiates large amounts of visible, ultraviolet and x-ray radiation.





The result of all these collisions is that angular momentum is transferred to the outer reaches of the disk while the gas whirls inward to the central star or black hole.



Two blobs of gas in slightly different orbits collide with each other because the inner blob is moving a bit faster than the outer one.

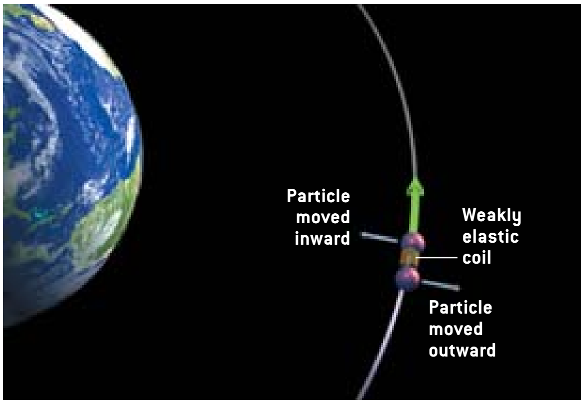
The collision transfers energy and angular momentum from the inner to the outer blob. The heated gas generates radiation.

Deprived of energy, the inner blob falls to a closer orbit and gains speed. The outer blob is flung to a farther orbit, slowing it down.

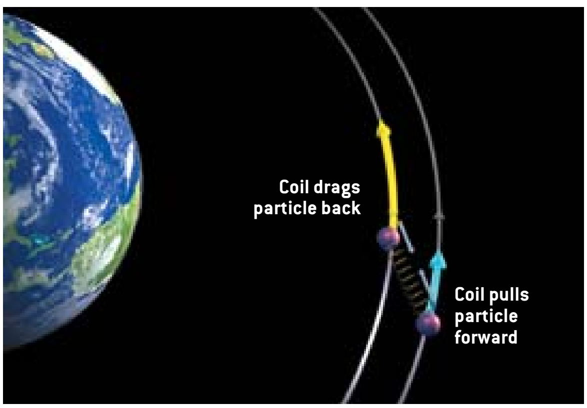
ALFRED T. KAMAJIAN

A DISK INSTABILITY

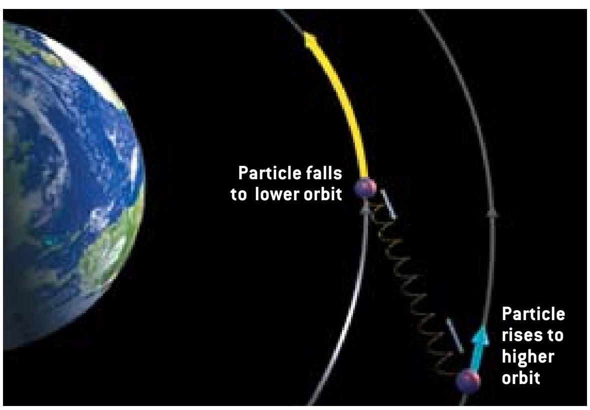
Astrophysicists believe that turbulence develops in accretion disks because the charged particles in the disk are linked by magnetic field lines, which try to bring the particles together if they move apart. But if the magnetic field is weak, it can actually drive the particles farther apart. This phenomenon, called magnetorotational instability, is best described by an analogy of two particles connected by a weakly elastic coil.



Say the two particles are orbiting Earth. A random collision pushes one particle a bit closer to Earth (hence speeding it up) and the other a bit farther away (slowing it down).



The elastic coil tries to bring the particles back together. The tension in the coil drags the faster, inner particle back and pulls the slower, outer particle forward.



Because the inner particle loses energy, it falls lower and gains speed. The outer particle rises to a higher orbit and slows down. This instability leads to turbulent flow.

duce a fluid instability in the disk. The instability would invariably cause a turbulent flow that would transport angular momentum outward and dissipate gravitational binding energy. This effect, which was called the magnetorotational instability (MRI), is now believed to play a central role in the way many accretion disks operate.

Magnetic field lines in a highly conductive medium must move with the medium's underlying flow. Where the material goes, so goes the field. Magnetic field lines also exert forces on the medium. In particular, just like elastic strings, the field lines exert tension forces when they are bent or curved.

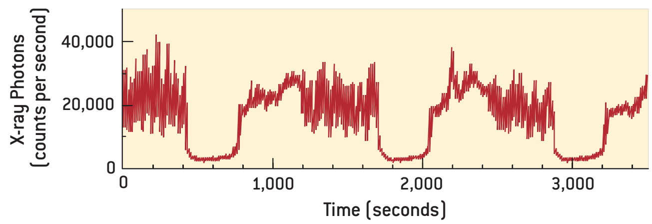
To understand the effect of magnetic field lines, imagine two particles in orbit about Earth that are tethered together by an elastic coil [*see box at left*]. If the particles are momentarily pulled apart (with one particle moved a bit closer to Earth and the other a bit farther away), most people would assume that the resulting tension in the coil would cause them to snap back to their original configuration. If the tension is sufficiently weak, however, its presence can actually drive the particles farther apart. The particle moved inward must speed up in its orbit to conserve angular momentum, and the particle perturbed outward must slow down for the same reason. The stretched coil acts to slow the faster, inner particle and speed the slower, outer particle. Deprived of some of its kinetic energy, the inner particle falls farther inward (and paradoxically speeds up in its orbit), while the velocity boost flings the outer particle farther out (where it moves even more slowly). In effect, the coil is transferring angular momentum from the inner to the outer particle. In an accretion disk composed of charged particles, magnetic field lines work in exactly the same way.

It is easy to see how this instability would lead to turbulent flow. Consider another analogy: a circular racetrack with the cars in the inner lanes moving faster than the cars in the outer lanes. Suppose someone hooked chains between cars in different lanes. The cars in the inner lanes would lose angular momentum as they are dragged backward, while the cars in the outer lanes would gain angular momentum as they are flung forward. The result would be considerable chaos and mayhem, just like the turbulence that develops in an accretion disk.

The discovery of the MRI has revolutionized our understanding of accretion disks. The situation is rather similar to that prevailing in the early 20th century when astronomers first realized that the primary energy source for stars was nuclear fusion reactions occurring in the stellar core. Now astrophysicists have deduced the mechanism that powers even greater energy sources such as quasars and active galactic nuclei (highly energetic galactic cores that are also thought to be fueled by matter falling into supermassive black holes). Researchers are currently exploring how MRI turbulence works in different physical situations and how that might explain the observed deviations in behavior among various types of accretion disks.

For example, some scientists are interested in whether and how MRI turbulence acts in protoplanetary disks. Such disks form a much cooler environment than those around white dwarfs, neutron stars and black holes because of the much

ALFRED T. KAMAJIAN



ODD PATTERNS appear in the x-ray radiation from accretion disks such as GRS 1915+105, which surrounds a black hole in a binary system about 40,000 light-years from Earth. Physicists do not know what causes these oscillations.

ELIZA JEWETT, ADAPTED FROM FIGURE 4D BY BELLONI ET AL. IN A&A, VOL. 355, 271–290; 2000

smaller gravity of the central star. As a result, the disks are composed largely of electrically neutral dust and gas rather than ionized plasma. Whether magnetic fields can affect the flow of such material is far from clear.

My group and others are trying to figure out how MRI works in hot, opaque accretion disks around black holes. The turbulence in these disks can be effectively supersonic, forming and reforming shock waves of charged particles just as supersonic aircraft produce sonic booms. Because these motions can kick photons to high energies, and because the photons can move more easily through the relatively transparent regions between shock waves, MRI turbulence can produce characteristic patterns of radiation that astronomers should be able to observe from black hole systems.

Oscillations and Jets

GIVEN THAT MANY accretion disks are thought to contain very turbulent flows, it is hardly surprising that observations show a high degree of variability in their output of radiation. The variations are usually random and chaotic, but there is occasionally order within the chaos. Intriguing, inexplicable patterns in the light output occur over and over again [*see illustration above*], and oscillations with reasonably well defined frequencies are sometimes evident. The Rossi X-ray Timing Explorer, a satellite that can measure rapid changes in x-ray brightness, has significantly aided the study of oscillations in accretion disks around neutron stars and stellar-mass black holes (those with a mass four to 15 times as great as the sun's).

Astrophysicists do not know what causes these variability patterns or the oscillation frequencies. An exciting possibility, proposed by Robert Wagoner of Stanford University and others, is that the oscillations reflect discrete modes of vibration of the disk, very much like the harmonics of a violin string. And just as the notes produced by a violin string can reveal the string's tension and mass, the observed frequencies of an accretion disk might be able to tell us about the disk's structure and the spacetime around the neutron star or black hole.

Although much of the gravitational binding energy released by the spiraling material in accretion disks ends up in the form of radiation, sometimes the energy also drives winds and jets of particles from the disk [*see box on page 52*]. Astronomers are intently exploring how such outflows are generated and what determines the partitioning of accretion power into radiative and kinetic luminosity. In all likelihood, different types of disks have different mechanisms for expelling particles. In some cases, the outflows probably wield a controlling influence on the accretion disk, because they carry not only mass and energy outward but perhaps significant amounts of angular momentum as well.

One possible driving mechanism for some types of outflows is pressure from the photons that are produced by the accretion disk. Even though photons have zero rest mass, they still carry momentum. When photons scatter off material, they exchange momentum with the particles they hit and thereby exert a force on them. (This is the principle behind solar sails.) Ultraviolet photons radiating from young massive stars are known to drive particle winds outward by scattering off the atoms and ions surrounding the star. In the same way, ultraviolet photons from accretion disks around white dwarfs and in active galactic nuclei or quasars may also accelerate winds from the disk.

Some systems, such as young stars and certain classes of active galactic nuclei, produce very fast, narrow jets of particles extending up to several light-years in the case of young stars and to more than several million light-years for active galactic nuclei. The fact that these jets remain collimated in a narrow beam over such great distances suggests that magnetic fields may be involved. (Astronomers have also inferred the presence of such fields in active galactic nuclei from their effects on the polarization of radio waves.) Because the accretion disk itself is believed to be magnetized, the rotation of the disk can twist the magnetic field lines into a helix. Tension in the field lines that spiral around a jet of particles can help confine it. Back in the 1980s, Roger Blandford and David Payne of the California Institute of Technology suggested that the rotation of the disk may also help fling material outward along the field lines, providing the initial acceleration and mass loading for the jet. Unfortunately, we do not yet know how to relate the inward accretion flow in the disk, with its complex MRI turbulence, to the apparently more ordered field structure in a jet outflow. But the rapid progress we are making in studying the magnetic fields in accretion disks may help us crack these kinds of problems.

Astrophysicists have spent decades trying to figure out how accretion disks work, and now we believe we have a basic understanding of these systems. As we investigate how magnetic turbulence operates in different environments, we hope to someday comprehend the remarkable variety of phenomena these spinning disks exhibit. It was an accretion disk that gave birth to our solar system, so unraveling the dynamics of these fascinating objects may ultimately help explain how we came to be. SA

MORE TO EXPLORE

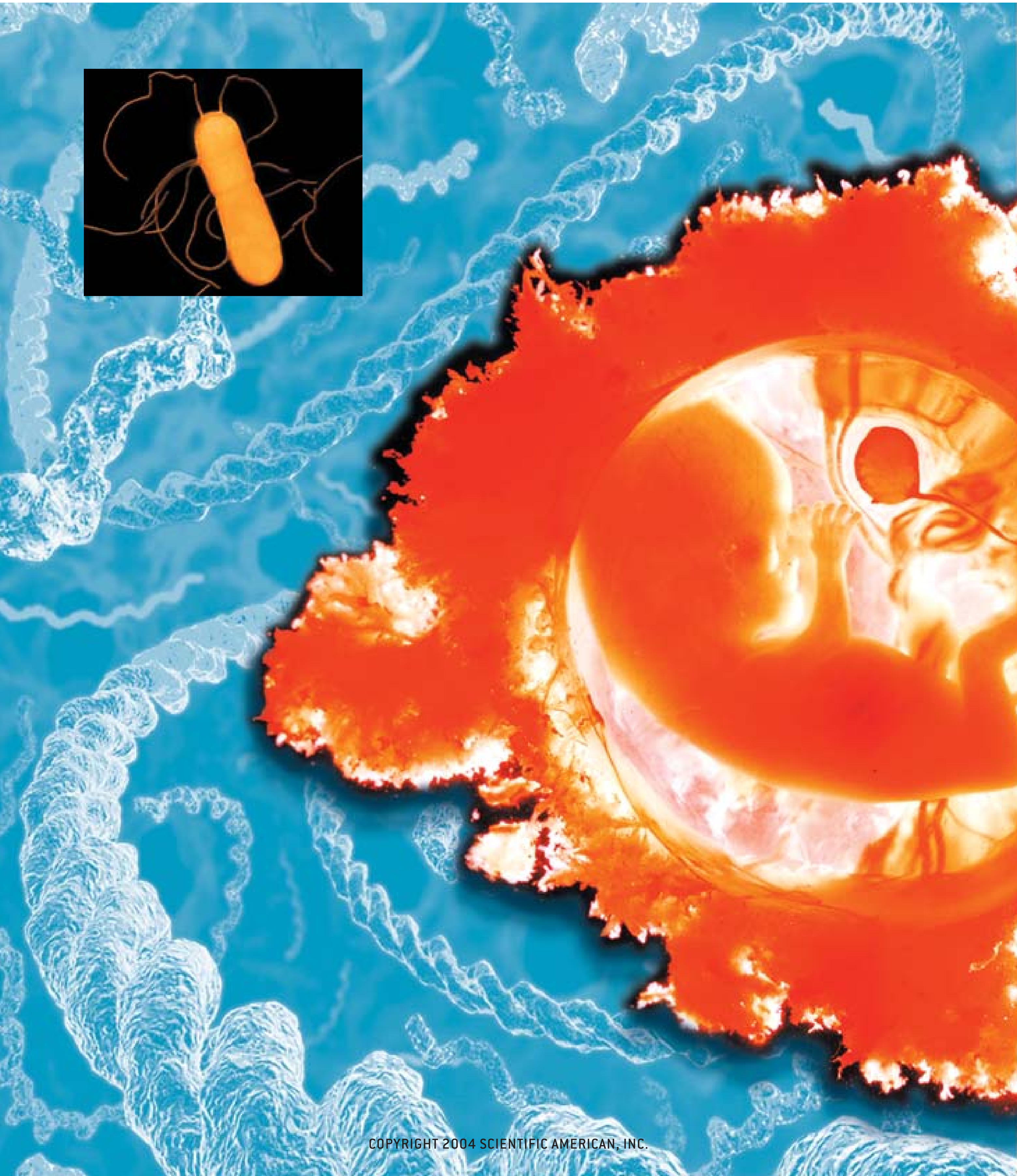
Gravity's Fatal Attraction: Black Holes in the Universe. M. Begelman and M. J. Rees. W. H. Freeman and Company, 1998.

Accretion Power in Astrophysics. Third edition. Juhan Frank, Andrew King and Derek Raine. Cambridge University Press, 2002.

Accretion Processes in Star Formation. Lee Hartmann. Cambridge University Press, 2004.

More information about accretion disks, black holes and other astrophysical objects can be found at **http://imagine.gsfc.nasa.gov/**

THE HIDDEN GENETIC



PROGRAM of COMPLEX ORGANISMS

By John S. Mattick

Biologists assumed that proteins alone regulate the genes of humans and other complex organisms. But an overlooked regulatory system based on RNA may hold the keys to development and evolution

BACTERIA AND HUMANS differ greatly in their structural and developmental complexity, but biologists have long assumed that all organisms used the same genetic mechanisms. Yet new work hints that complexity arises from an additional program hidden in "junk" DNA.

Assumptions can be dangerous, especially in science. They usually start as the most plausible or comfortable interpretation of the available facts. But when their truth cannot be immediately tested and their flaws are not obvious, assumptions often graduate to articles of faith, and new observations are forced to fit them. Eventually, if the volume of troublesome information becomes unsustainable, the orthodoxy must collapse.

We may be witnessing such a turning point in our understanding of genetic information. The central dogma of molecular biology for the past half a century and more has stated that genetic information encoded in DNA is transcribed as intermediary molecules of RNA, which are in turn translated into the amino acid sequences that make up proteins. The prevailing assumption, embodied in the credo "one gene, one protein," has been that genes are generally synonymous with proteins. A corollary has been that proteins, in addition to their structural and enzymatic roles in cells, must be the primary agents for regulating the expression, or activation, of genes.

This conclusion derived from studies primarily on bacteria such as *Escherichia coli* and other prokaryotes (simple one-celled organisms lacking a nucleus). And indeed, it is still essentially correct for prokaryotes. Their DNA consists almost entirely of genes encoding proteins, separated by flanking sequences that regulate the expression of the adjacent genes. (A few genes that encode RNAs with regulatory jobs are also present, but they make up only a tiny fraction of most prokaryotes' genetic ensembles, or genomes.)

Researchers have also long assumed that proteins similarly represent and control all the genetic information in animals, plants and fungi—the multicellular organisms classified as eukaryotes (having cells that contain nuclei). Pioneering biologist Jacques Monod summarized the universality of the central dogma as "What was true for *E. coli* would be true for the elephant."

Monod was only partly right. A growing library of results reveals that the central dogma is woefully incomplete for describing the molecular biology of eukaryotes. Proteins do play a role in the regulation of eukaryotic gene expression, yet a hidden, parallel regulatory system consisting of RNA that acts directly on DNA, RNAs and proteins is also at work. This overlooked RNA-signaling network may be what allows humans, for example, to achieve structural complexity far beyond anything seen in the unicellular world.

Some molecular biologists are skeptical or even antagonistic toward these unorthodox ideas. But the theory may answer some long-standing riddles of development and evolution and holds great implications for gene-based medicine and pharmaceuticals. Moreover, the recent discovery of this system affords insights that could revolutionize designs for complex programmed systems of all kinds, cybernetic as well as biological.

RNAs AND PROTEINS may communicate regulatory information IN PARALLEL.

The Ubiquitous Junk

A DISCOVERY in 1977 presaged that something might be wrong with the established view of genomic programming. Phillip A. Sharp of the Massachusetts Institute of Technology and Richard J. Roberts of New England Biolabs, Inc., and their respective colleagues independently showed that the genes of eukaryotes are not contiguous blocks of protein-coding sequences. Rather they are mosaics of "exons" (DNA sequences that encode fragments of proteins) interspersed with often vast tracts of intervening sequences, or "introns," that do not code for protein. In the nucleus, a gene is first copied in its totality as a primary RNA transcript; then a process called splicing removes the intronic RNAs and reconstitutes a continuous coding sequence—messenger RNA, or mRNA—for translation as protein in the cytoplasm. The excised intronic RNA, serving no apparent purpose, has been presumed to be degraded and recycled.

But if introns do not code for protein, then why are they ubiquitous among eukaryotes yet absent in prokaryotes? Although introns constitute 95 percent or more of the average protein-coding gene in humans, most molecular biologists have considered them to be evolutionary leftovers, or junk. Introns were rationalized as ancient remnants of a time before cellular life evolved, when fragments of protein-coding information crudely assembled into the first genes. Perhaps introns had survived in complex organisms because they had an incidental usefulness—for example, making it easier to reshuffle segments of proteins into useful new combinations during evolution. Similarly, biologists have assumed that the absence of introns from prokaryotes was a consequence of intense competitive pressures in the microbial environment: evolution had pruned away the introns as deadweight.

One observation that made it easier to dismiss introns—and other seemingly useless "intergenic" DNA that sat between genes—as junk was that the amount of DNA in a genome does not correlate well with the organism's complexity. Some amphibians, for example, have more than five times as much DNA as mammals do, and astonishingly, some amoebae have 1,000 times more. For decades, researchers assumed that the underlying number of protein-coding genes in these organisms correlated much better with complexity but that the relationship was lost against the variable background clutter of introns and other junk sequences.

But investigators have since sequenced the genomes of diverse species, and it has become abundantly clear that the correlation between numbers of conventional genes and complexity truly is poor. The simple nematode worm *Caenorhabditis elegans* (made up of only about 1,000 cells) has about 19,000 protein-coding genes, almost 50 percent more than insects (13,500) and nearly as many as humans (around 25,000). Conversely, the relation between the amount of nonprotein-coding DNA sequences and organism complexity is more consistent.

Put simply, the conundrum is this: less than 1.5 percent of the human genome encodes proteins, but most of it is transcribed into RNA. Either the human genome (and that of other complex organisms) is replete with useless transcription, or these nonprotein-coding RNAs fulfill some unexpected function.

This line of argument and considerable other experimental evidence suggest that many genes in complex organisms—perhaps even the majority of genes in mammals—do not encode protein but instead give rise to RNAs with direct regulatory functions [see "The Hidden Genome," by W. Wayt Gibbs, SCIENTIFIC AMERICAN, November and December 2003]. These RNAs may be transmitting a level of information that is crucial, par-

Overview/*Revising Genetic Dogma*

- A perplexingly large portion of the DNA of complex organisms (eukaryotes) seems irrelevant to the production of proteins. For years, molecular biologists have assumed this extra material was evolutionary "junk."
- New evidence suggests, however, that this junk DNA may encode RNA molecules that perform a variety of regulatory functions. The genetic mechanisms of eukaryotes may therefore be radically different from those of simple cells (prokaryotes).
- This new theory could explain why the structural and developmental complexity of organisms does not parallel their numbers of protein-coding genes. It also carries important implications for future pharmaceutical and medical research.

JEFF JOHNSON (*DNA background*); MOREDUN ANIMAL HEALTH LTD/SCIENCE PHOTO LIBRARY (*bacterium*); LENNART NILSSON (*embryo*) (*preceding pages*) ticularly to development, and that plays a pivotal role in evolution.

From Parasites to Parallel Controls

THE CLUE to understanding this point may lie in a new interpretation of introns. Contrary to early assumptions that introns generally date back to the dawn of life, evidence amassed more recently indicates that these sequences invaded the genes of higher organisms late in evolution. Most likely, they derived from a type of self-splicing mobile genetic element similar to what are now called group II introns. These elements are parasitic bits of DNA that have the peculiar ability to insert themselves into host genomes and to splice themselves out when expressed as RNA.

Group II introns are found only occasionally in bacteria, and it is easy to see why. Because bacteria lack a nucleus, transcription and translation occur together: RNA is translated into protein almost as fast as it is transcribed from DNA. There is no time for intronic RNA to splice itself out of the protein coding RNA in which it sits, so an intron would in most cases disable the gene it inhabits, with harmful consequences for the host bacterium. In eukaryotes, transcription occurs in the nucleus and translation in the cytoplasm, a separation that opens a window of opportunity for the intron RNA to excise itself. Introns can thus be more easily tolerated in eukaryotes.

Of course, as long as introns needed to splice themselves in and out of genomes, their sequences could not have deviated much from that of group II introns. But a further leap in intron evolution may have accompanied the evolution in eukaryotes of the structure called the spliceosome. This is a complex of small catalytic RNAs and many proteins; its job is to snip intron RNA out of messenger RNA precursors efficiently.

By freeing introns from the need to splice themselves, the spliceosome would in effect have encouraged introns to proliferate, mutate and evolve. Any random mutation in an intron that proved beneficial to the host organism would have been retained by natural selection. Intronic RNAs would therefore be evolving independently and in parallel with proteins. In short, the entry of introns into eukaryotes may have initiated an explosive new round of molecular evolution, based on RNA rather than protein. Instead of being junky molecular relics, introns could have progressively acquired genetic functions mediated by RNA.

If this hypothesis is true, its meaning may be profound. Eukaryotes (especially the more complex ones) may have developed a genetic operating system and regulatory networks that are far more sophisticated than those of prokaryotes: RNAs and proteins could communicate regulatory information in parallel. Such an arrangement would resemble the advanced information-processing systems supporting network controls in computers and the brain.

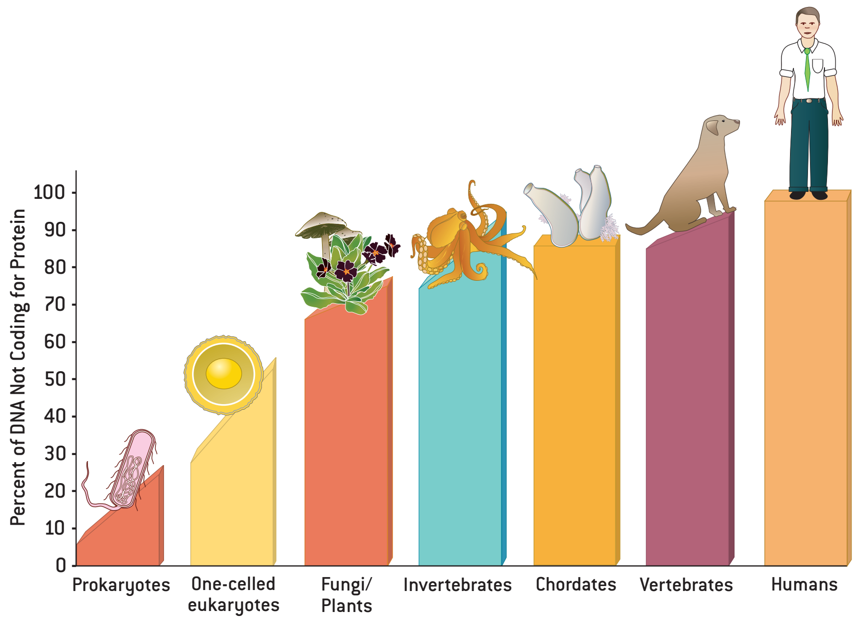
Functional jobs in cells routinely belong to proteins because they have great chemical and structural diversity. Yet RNA has an advantage over proteins for transmitting information and regulating activities involving the genome itself: RNAs can encode short, sequence-specific signals as a kind of bit string or zip code. These embedded codes can direct RNA molecules precisely to receptive targets in other RNAs and DNA. The RNA-RNA and RNA-DNA interactions could in turn create structures that recruit proteins to convert the signals to actions.

The bit string of addressing information in the RNA gives this system the power of tremendous precision, just as the binary bit strings used by digital computers do. It is not too much of a stretch to say that this RNA regulatory system would be largely digital in nature.

We may have totally misunderstood THE NATURE OF THE GENOMIC PROGRAMMING.

The evidence for a widespread RNA-based regulatory system is strong, albeit still patchy. If such a system exists, one would expect that many genes might have evolved solely to express RNA signals as higher-order regulators in the network. That appears to be the case: thousands of RNAs that never get translated into protein (noncoding RNAs) have been identified in recent analyses of transcription in mammals. At least half and possibly more than three quarters of all RNA transcripts fit this category.

One would also expect that many of



NONPROTEIN-CODING SEQUENCES make up only a small fraction of the DNA of prokaryotes. Among eukaryotes, as their complexity increases, generally so, too, does the proportion of their DNA that does not code for protein. The noncoding sequences have been considered junk, but perhaps it actually helps to explain organisms' complexity.

LUCY READING; SOURCE: R. J. TAFT AND J. S. MATTICK

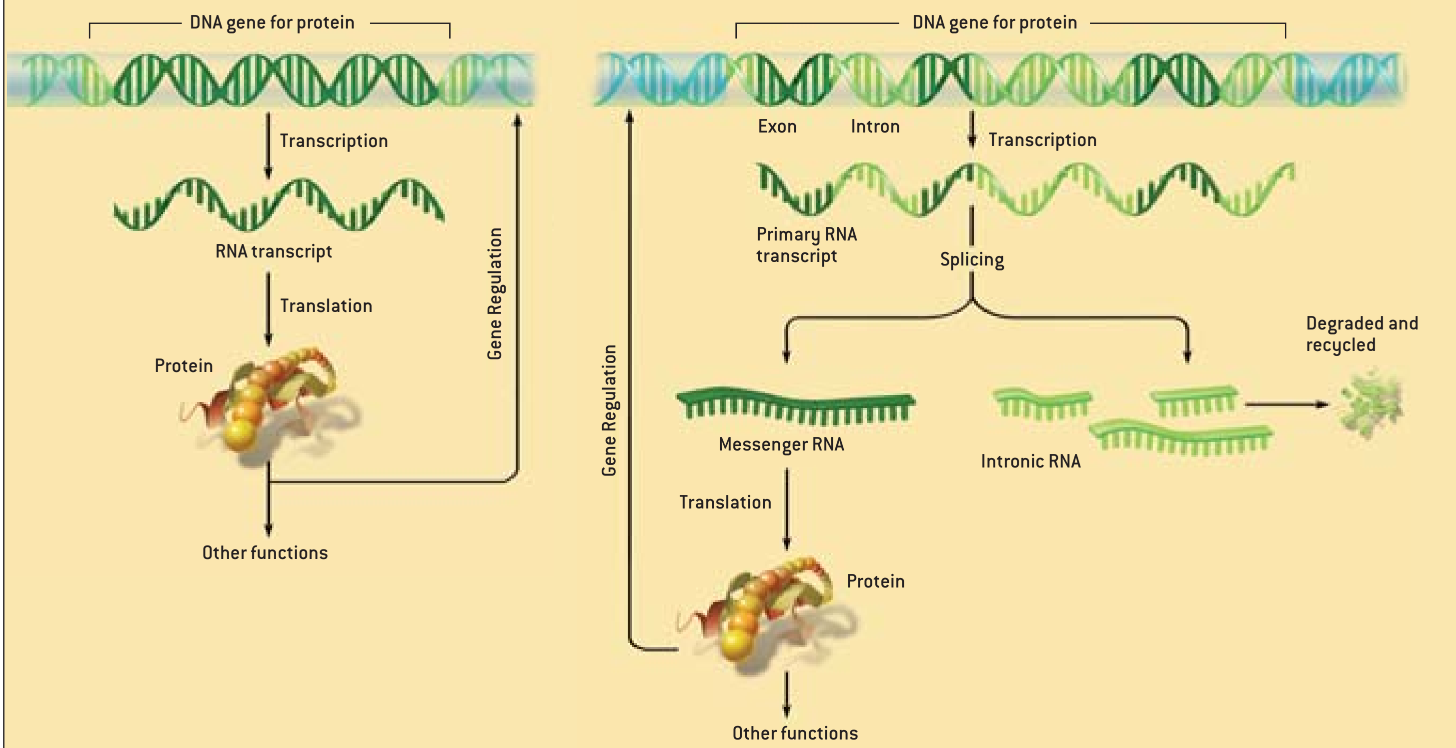
AN EVOLVING VIEW OF GENE ACTIVITY

GENE ACTIVITY IN PROKARYOTES

Prokaryotes (bacteria and other simple cells) have DNA that consists almost entirely of protein-coding genes. When those genes are active, they give rise to RNA transcripts that are immediately translated into proteins, which in turn regulate genetic activity and provide other functions.

TRADITIONAL VIEW OF GENE ACTIVITY IN EUKARYOTES

In the DNA of eukaryotes (complex organisms), individual genes comprise "exon" sequences that code for segments of protein separated by noncoding "intron" sequences. When a gene is active, it is entirely transcribed as RNA, but then the intronic RNA is spliced out and the exonic RNA is assembled as messenger RNA. The cell translates the messenger RNA into protein while breaking down and recycling the intronic RNA, which serves no purpose.



these RNAs might be processed into smaller signals capable of addressing targets in the network. Hundreds of "microRNAs" derived from introns and larger nonprotein-coding RNA transcripts have in fact already been identified in plants, animals and fungi. Many of them control the timing of processes that occur during development, such as stem cell maintenance, cell proliferation, and apoptosis (the so-called programmed cell death that remodels tissues). Many more such small RNAs surely await discovery.

These RNA signals, by finding targets on other RNAs, DNA and proteins, could influence a cell's genetic program in many ways. For example, they could inform various genes that a particular protein-coding sequence has been transcribed, and that feedback could trigger a host of parallel adjustments. More important, however, the RNA signals could serve as a powerful feed-forward program embedded in the genetic material that controls the trajectories of gene expression. If so, they could explain some of the deep mysteries surrounding cell differentiation and organism development.

Regulating Development

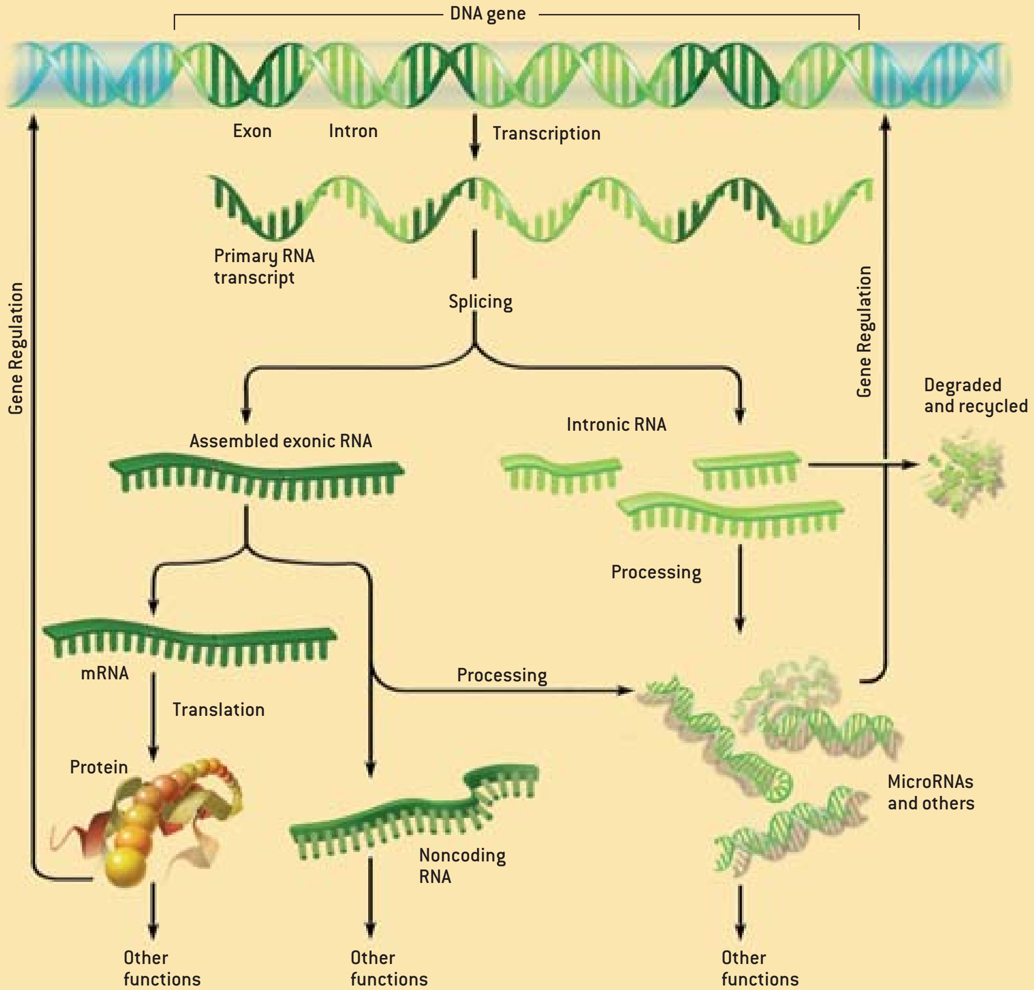
CONSIDER WHAT HAPPENS during human embryonic development: a single fertilized cell progresses to become a precisely structured, beautifully sculptured organism of an estimated 100 trillion cells with distinct positions and functions. The pattern of gene expression that makes this transformation possible relies heavily on two phenomena: modification of chromatin and alternative splicing.

Chromatin is the material that makes up chromosomes; it consists of DNA complexed with proteins. Within cells, small chemical tags (such as methyl and acetyl groups) can attach to segments of the DNA and to the chromatin proteins and thereby determine whether the genes in the associated DNA will be accessible for transcription or will stay silent. Recent results indicate that RNA signaling directs the tagging of the chromatin and thus gene expression. Indeed, a number of complex chromosomal processes, such as mitosis (cell division) and meiosis (the formation of sperm and egg precursors), as well as a range of complex genetic phenomena appear to depend on biochemical pathways that affect RNA processing.

Alternative splicing generates divergent repertoires of RNAs and proteins in the cells of a body's different tissues, all of which share a common set of genes. Most protein-coding transcripts are alterna-

TERESE WINSLOW

NEW VIEW OF GENE ACTIVITY IN EUKARYOTES
Some of the intronic RNA and even some of the assembled exonic RNA may play a direct regulatory role by interacting with the DNA, other RNA molecules or proteins. By modifying protein production at various levels, these noncoding RNAs may superimpose additional genetic instructions on a cell.



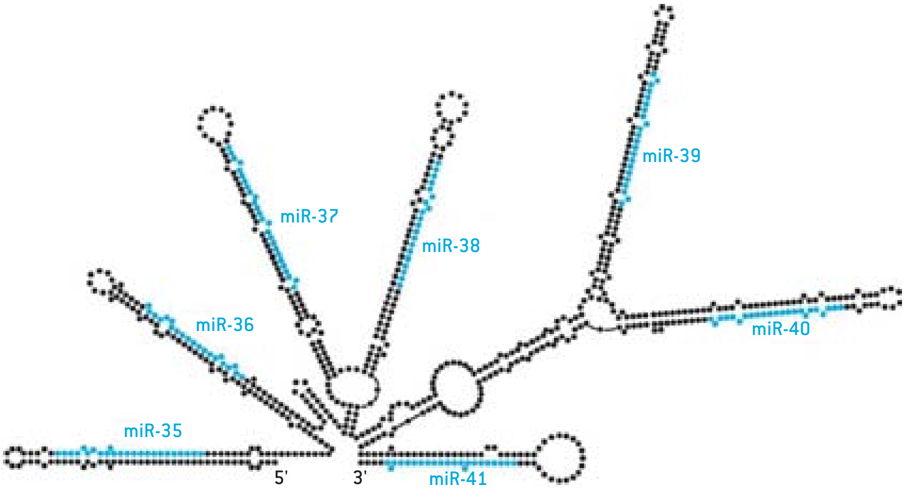
tively spliced in mammals. When intron RNA is spliced out of a gene's transcript, the protein-coding RNA regions may be assembled in more than one way to yield more than one type of protein. The phenomenon is of fundamental importance to animal and plant development, but no one yet understands how cells specify which form of a protein they will make. Few protein factors that control the alternative splicing of specific genes have been found. Consequently, researchers have usually supposed that subtle combinations of general factors activate or repress alternative splicing in different contexts. But no strong evidence has backed up that presumption.

A more likely and mechanistically appealing possibility, however, is that RNAs regulate the process directly. In principle, these molecules could exert exquisitely flexible control by tagging or grabbing particular sequences in primary gene transcripts and steering how the spliceosome joins the pieces. In keeping with that idea, DNA sequences at the intron-exon junctions where alternative splicing occurs are often resistant to change during evolution. Also, a number of laboratories have demonstrated that artificial antisense RNAs designed to bind to such sites can modify splicing patterns in cultured cells, as well as in whole animals. It is perfectly plausible that this phenomenon occurs naturally in vivo, too, but has just not yet been detected.

Controlling Complexity

SUCH CONSIDERATIONS lead naturally to a more general consideration of what type of information, and how much of it, might be required to program the development of complex organisms. The creation of complex objects, whether houses or horses, demands two kinds of specifications: one for the components and one for the system that guides their assembly. (To build a house, one must specify the needed bricks, boards and beams, but one must also have an architectural plan to show how they fit together.) In biology, unlike engineering, both types of information are encoded within one program, the DNA.

The component molecules that make up different organisms (both at the individual and the species levels) are fundamentally alike: around 99 percent of the proteins in humans have recognizable equivalents in mice, and vice versa; many of those proteins are also conserved in other animals, and those involved in basic cellular processes are conserved in all



PROPOSED PRECURSOR molecule for microRNAs is a primary RNA transcript that may produce multiple small RNAs (*blue*). The structure of the precursor might guide the excision of these small RNA signals.

eukaryotes. Thus, the differences in animals' forms surely arise more fundamentally from differences in the architectural information.

Protein-coding genes obviously specify the components of organisms, but where does the architectural information reside? Biologists have widely assumed that the instructions for assembling complex organisms are somehow embedded in the diverse combinations of regulatory factors within cells—that is, in the permutations of regulatory proteins interacting with one another and with the DNA and RNA. Yet, as Daniel C. Dennett of Tufts University has observed, although such combinatorics can generate almost endless possibilities, the vast majority will be chaotic and meaningless—which is problematic for biology. Throughout their evolution and development, organisms must navigate precise developmental pathways that are sensible and competitive, or else they die. Generating complexity is easy; controlling it is not. The latter requires an enormous amount of regulatory information.

Both intuitive and mathematical considerations suggest that the amount of regulation must increase as a nonlinear (usually quadratic) function of the number of genes. So, as the system becomes more complex, an increasing proportion of it must be devoted to regulation. This nonlinear relation between regulation and function appears to be a feature of all integrally organized systems. Therefore, all such systems have an intrinsic complexity limit imposed by the accelerating growth of their control architecture, until or unless the regulatory mechanism changes fundamentally.

Generating COMPLEXITY is easy; CONTROLLING IT IS NOT.

In agreement with this prediction, the number of protein regulators in prokaryotes has been found to increase quadratically with genome size. Moreover, extrapolation indicates that the point at which the number of new regulators is predicted to exceed the number of new functional genes is close to the observed upper limit of bacterial genome sizes.

Throughout evolution, therefore, the complexity of prokaryotes may have been limited by genetic regulatory overhead, rather than by environmental or biochemical factors as has been commonly assumed. This conclusion is also consistent with the fact that life on earth consisted solely of microorganisms for most of its history. Combinatorics of protein interactions could not, by themselves, lift that complexity ceiling.

Eukaryotes must have found a solution to this problem. Logic and the available evidence suggest that the rise of multicellular organisms over the past billion years was a consequence of the transition to a new control architecture based largely on endogenous digital RNA signals. It would certainly help explain the phenomenon of the Cambrian explosion about 525 million years ago, when invertebrate animals of jaw-dropping diversity evolved, seemingly abruptly, from much simpler life. Indeed, these results suggest a general rule with relevance beyond biology: organized complexity is a function of regulatory information—and, in virtually all systems, as observed by Marie E. Csete, now at Emory University School of Medicine, and John C. Doyle of the California Institute of Technology, explosions in complexity occur as a result of advanced controls and embedded networking.

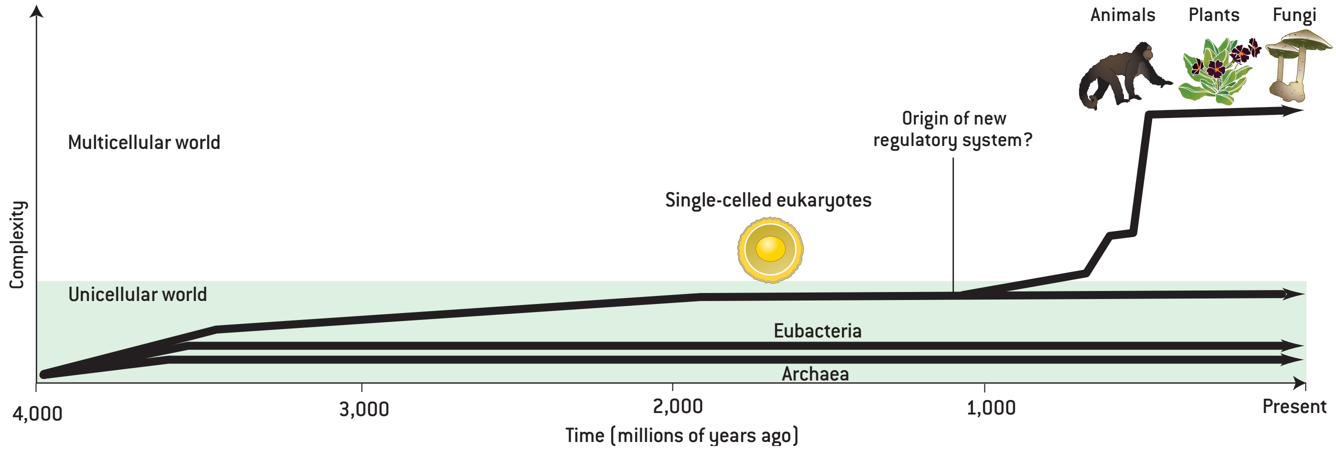
The implications of this rule are staggering. We may have totally misunderstood the nature of the genomic programming and the basis of variations in traits among individuals and species. The rule implies that the greater portion of the genomes in complex organisms is not junk at all—rather it is functional and subject to evolutionary selection.

The most recent surprise is that vertebrate genomes contain thousands of noncoding sequences that have persisted virtually unaltered for many millions of years. These sequences are much more highly conserved than those coding for proteins, which was totally unexpected. The mechanism that has frozen these sequences is unknown, but their extreme constancy suggests that they are involved

THE AUTHOR

JOHN S. MATTICK, born and raised in Sydney, today is a professor of molecular biology at the University of Queensland and director of the Institute for Molecular Bioscience. Formerly he was also foundation director of the Australian Genome Research Facility. His accomplished career includes the development of Australia's first genetically engineered vaccine. In 2001 Mattick was appointed an Officer in the Order of Australia, and in 2003 he was awarded the Australian government's Centenary Medal. Married with three sons, he enjoys walking and body surfing as time permits.

ALICE CHEN; FROM NELSON C. LAU ET AL. IN *SCIENCE*, VOL. 294; OCTOBER 26, 2001



UNICELLULAR LIFE, primarily prokaryotes, ruled the earth for billions of years. When multicellular life appeared, however, its complexity rose with dizzying speed. The evolution of an additional genetic regulatory system might explain both the jump to multicellularity and the rapid diversification into complexity.

in complex networks essential to our biology. Thus, rather than the genomes of humans and other complex organisms being viewed as oases of protein-coding sequences in a desert of junk, they might better be seen as islands of protein-component information in a sea of regulatory information, most of which is conveyed by RNA.

The existence of an extensive RNA-based regulatory system also has ramifications for pharmacology, drug development and genetic screening. Traditional genetic diseases such as cystic fibrosis and thalassemia are caused by catastrophic component damage: one of the individual's proteins simply doesn't work. Yet many, if not most, of the genetic variations determining susceptibility to most diseases and underpinning our individual idiosyncrasies probably lie in the noncoding regulatory architecture of our genome that controls growth and development. (Noncoding RNAs have already been linked with several conditions, including B cell lymphoma, lung cancer, prostate cancer, autism and schizophrenia.)

Such defects will not be easy to identify by molecular genetic epidemiology, nor will they necessarily be easy to correct. But understanding this regulatory system may ultimately be critical to understanding our physical and psychological individuality, as well as trait variation in plants and animals. It may also be the prelude to sophisticated strategies for medical intervention to optimize health and for truly advanced genetic engineering in other species.

Aside from introns, the other great source of presumed genomic junk—accounting for about 40 percent of the human genome—comprises transposons and other repetitive elements. These sequences are widely regarded as molecular parasites that, like introns, colonized our genomes in waves at different times in evolutionary history. Like all immigrants, they may have been unwelcome at first, but once established in the community they and their descendants progressively became part of its dynamic—changing, contributing and evolving with it.

Good, albeit patchy, evidence suggests that transposons contribute to the evolution and genomic regulation of higher organisms and may play a key role in epigenetic inheritance (the modification of genetic traits). Moreover, this past July Erev Y. Levanon of Compugen and colleagues elsewhere announced an exciting discovery involving a process called A-to-I (adenosine-to-inosine) editing, in which an RNA sequence changes at a very specific site. They demonstrated that A-to-I editing of RNA transcripts is two orders of magnitude more widespread in humans than was previously thought and overwhelmingly occurs in repeat sequences called Alu elements that reside in noncoding RNA sequences. A-to-I editing is particularly active in the brain, and aberrant editing has been associated with a range of abnormal behaviors, including epilepsy and depression.

Although RNA editing occurs to some extent in all animals, Alu elements are unique to primates. An intriguing possibility is that the colonization of the primate lineage by Alu elements made it possible for a new level of complexity to arise in RNA processing and allowed the programming for neural circuitry to become more dynamic and flexible. That versatility may have in turn laid the foundation for the emergence of memory and higher-order cognition in the human species.

Finally, understanding the operation of the expanded and highly sophisticated regulatory architecture in the genomes of complex organisms may shed light on the challenges of designing systems capable of self-reproduction and self-programming—that is, true artificial life and artificial intelligence. What was dismissed as junk because it was not understood may well turn out to hold the secrets to human complexity and a guide to the programming of complex systems in general. SA

MORE TO EXPLORE

Darwin's Dangerous Idea. Daniel C. Dennett. Simon and Schuster, 1995.

Challenging the Dogma: The Hidden Layer of Non-Protein-Coding RNAs In Complex Organisms. John S. Mattick in *BioEssays*, Vol. 25, No. 10, pages 930–939; October 2003.

The Unseen Genome: Gems among the Junk. W. Wayt Gibbs in *Scientific American*, Vol. 289, No. 5, pages 46–53; November 2003.

Noncoding RNAs: Molecular Biology and Molecular Medicine. Edited by J. Barciszewski and V. A. Erdmann. Landes Bioscience/Eurekah.com, Georgetown, Tex., 2003.

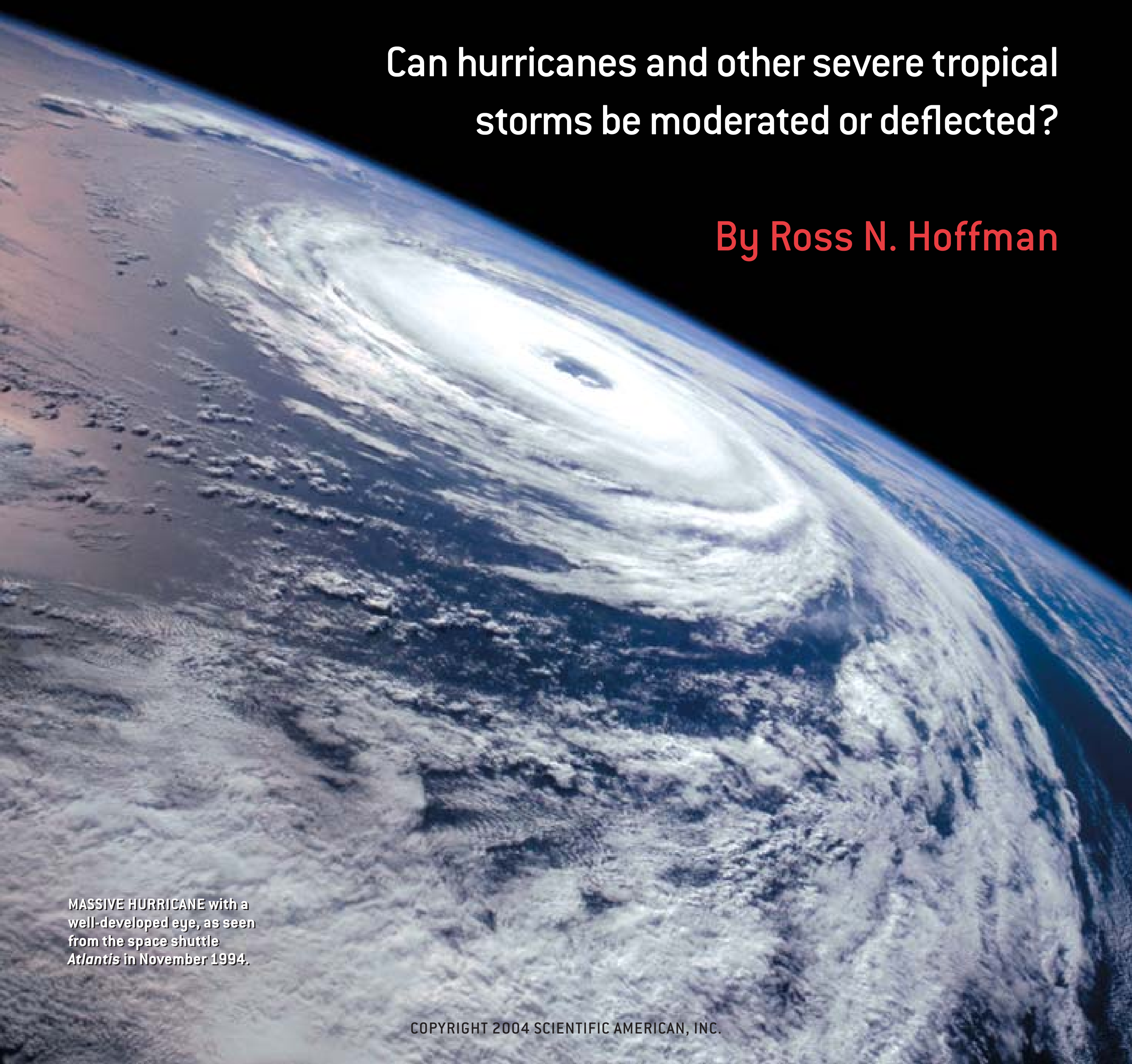
More information and lists of publications will be found at the author's Web site (under construction) at **http://imb.uq.edu.au/groups/mattick**

LUCY READING

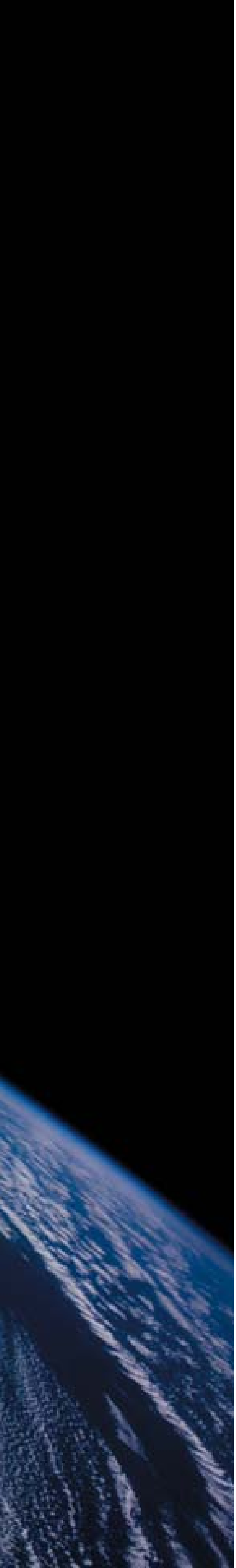
Controlling Hurricanes

Can hurricanes and other severe tropical storms be moderated or deflected?

By Ross N. Hoffman



MASSIVE HURRICANE with a well-developed eye, as seen from the space shuttle *Atlantis* in November 1994.



COURTESY OF NASA/CORBIS

Every year huge rotating storms packing winds greater than 74 miles per hour sweep across tropical seas and onto shorelines—often devastating large swaths of territory. When these roiling tempests—called hurricanes in the Atlantic and the eastern Pacific oceans, typhoons in the western Pacific and cyclones in the Indian Ocean—strike heavily populated areas, they can kill thousands and cause billions of dollars of property damage. And nothing, absolutely nothing, stands in their way.

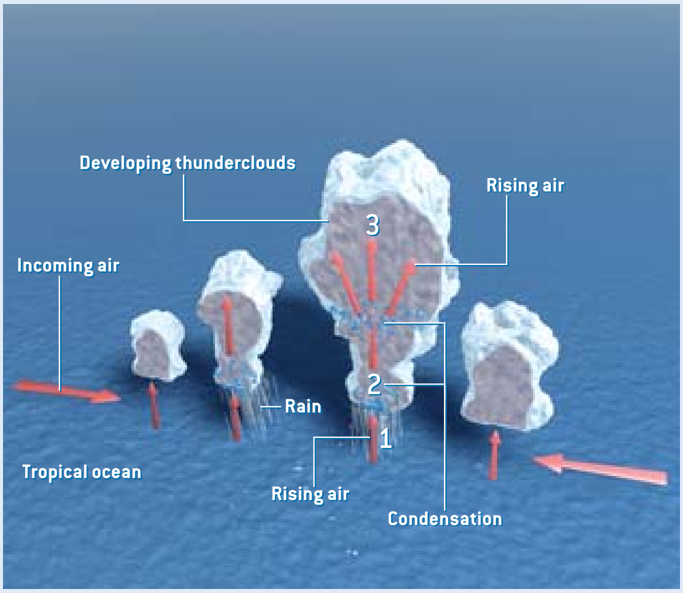
But must these fearful forces of nature be forever beyond our control? My research colleagues and I think not. Our team is investigating how we might learn to nudge hurricanes onto more benign paths or otherwise defuse them. Although this bold goal probably lies decades in the future, we think our results show that it is not too early to study the possibilities.

To even consider controlling hurricanes, researchers will need to be able to predict a storm's course extremely accurately, to identify the physical changes (such as alterations in air temperature) that would influence its behavior, and to find ways to effect those changes. This work is in its infancy, but successful computer simulations of hurricanes carried out during the past few years suggest that modification could one day be feasible. What is more, it turns out the very thing that makes forecasting any weather difficult—the atmosphere's extreme sensitivity to small stimuli—may well be the key to achieving the control we seek. Our first attempt at influencing the course of a simulated hurricane by making minor changes to the storm's initial state, for example, proved remarkably successful, and the subsequent results have continued to look favorable, too.

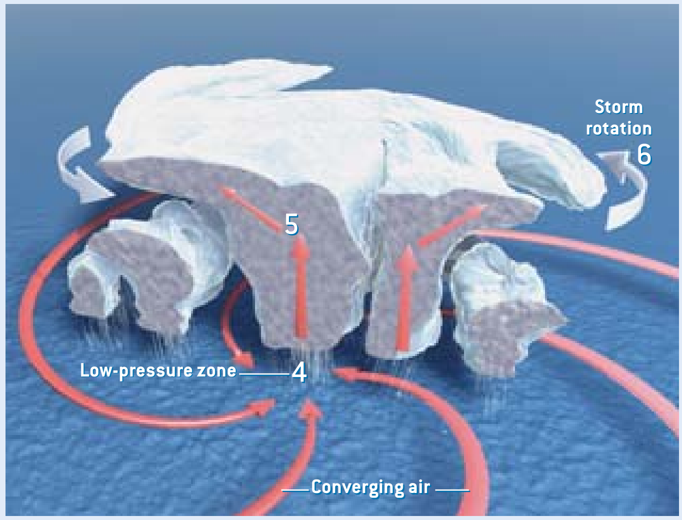
To see why hurricanes and other severe tropical storms may be susceptible to human intervention, one must understand their nature and origins [*see box on next two pages*]. Hurricanes grow as clusters of thunderstorms over the tropical oceans. Low-latitude seas continuously provide heat and moisture to the atmosphere, producing warm, humid air above the sea surface. When this air rises, the water vapor in it condenses to form clouds and precipitation. Condensation releases heat—the solar heat it took to evaporate the water at the ocean surface. This so-called latent heat of condensation makes the air more buoy-

ANATOMY OF A HURRICANE

Some scientists believe that they may be able to weaken or move hurricanes onto less dangerous tracks by altering the initial physical conditions (the air temperature or humidity, for example) in the center of the storm or even in the surrounding areas. To succeed, they need to make accurate and detailed forecasts of hurricanes. Here are the outlines of how these powerful storms arise.



Hurricanes start to form when tropical oceans release heat and water into the atmosphere, producing large amounts of warm, humid air above the surface (*1*). Warm air rises, and as it does so, the water vapor in it condenses to form clouds and rain (*2*). This condensation produces heat, causing air in the developing thunderclouds to climb still farther (*3*).



The release of heat above the tropical seas creates a surface low-pressure zone, where additional warm, moist air from the outer perimeter converges (*4*). This continuous movement into the burgeoning thunderstorm shifts huge amounts of heat, air and water skyward (*5*). This upward transfer and release of heat further enhance the convergence of surrounding air toward the growing storm center, which starts to circulate under the influence of the earth's rotation (*6*). The process continues apace, strengthening and organizing the storm.

DAVID FIERSTEIN

ant, causing it to ascend still higher in a self-reinforcing feedback process. Eventually, the tropical depression begins to organize and strengthen, forming the familiar eye—the calm central hub around which a hurricane spins. On reaching land, the hurricane's sustaining source of warm water is cut off, which leads to the storm's rapid weakening.

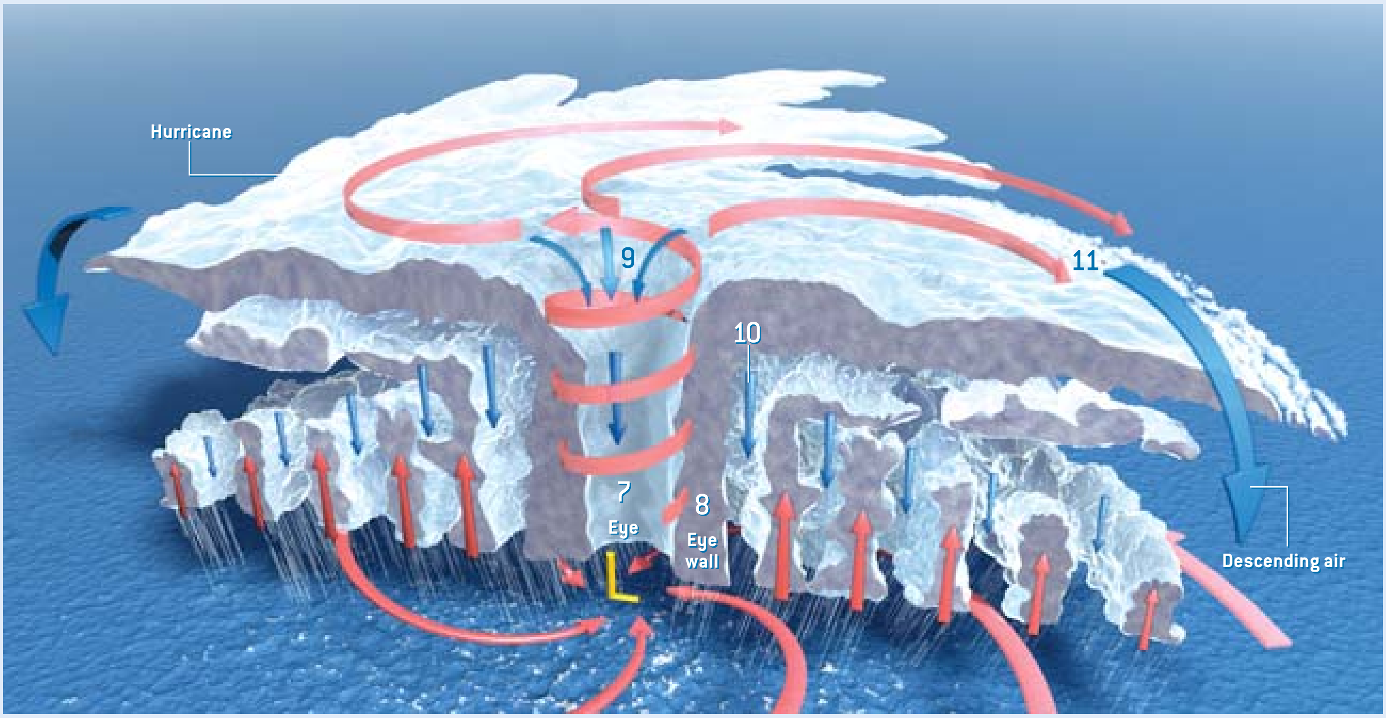
Dreams of Control

BECAUSE A HURRICANE draws much of its energy from heat released when water vapor over the ocean condenses into clouds and rain, the first researchers to dream of taming these unruly giants focused on trying to alter the condensation process using cloud-seeding techniques—then the only practical way to try to affect weather. In the early 1960s a U.S. government-appointed scientific advisory panel named Project Stormfury performed a series of courageous (or perhaps foolhardy) experiments to determine whether that approach might work.

Project Stormfury aimed to slow the development of a hurricane by augmenting precipitation in the first rain band outside the eye wall—the ring of clouds and high winds that encircle the eye [see "Experiments in Hurricane Modification," by R. H. Simpson and Joanne S. Malkus; Scientific American, December 1964]. They attempted to accomplish this goal by seeding the clouds there with silver iodide particles dispersed by aircraft, which would serve as nuclei for the formation of ice from water vapor that had been supercooled after rising to the highest, coldest reaches of the storm. If all went as envisioned, the clouds would grow more quickly, consuming the supplies of warm, moist air near the ocean surface, thus replacing the old eye wall. This process

Overview/*Taming Hurricanes*

- Meteorological researchers are simulating past hurricanes using sophisticated weather-forecasting models that closely reproduce the complex internal processes crucial to the development and evolution of severe tropical storms.
- The work confirms that these massive, chaotic systems are susceptible to minor changes in their initial conditions—for instance, the air temperature and humidity near the center of the storm and in the surrounding regions.
- Using complex mathematical optimization techniques, the researchers are learning what modifications to a hurricane could weaken its winds or divert it from populated areas.
- If these theoretical studies are ultimately successful, they should point the way toward practical methods for intervening in the life cycle of hurricanes to protect life and property.



As the storm intensifies, an eye—a calm, low-pressure hub—typically forms (7). The eye is encircled by a ring of clouds and high winds called the eye wall (8). The storm has become a hurricane. At the same time, the rising air, now heated and having lost much of its moisture, can rise no further, because the stratosphere acts like a lid above the hurricane. Some of this dry air falls into the eye (9) and between the cloud bands (10), while the remainder spirals away from the storm center and descends (11). Meanwhile large-scale air currents nearby steer the hurricane along its path.

would then expand the radius of the eye, lessening the hurricane's intensity in a manner akin to a spinning skater who extends her arms to slow down.

The Stormfury results were ambiguous at best. Meteorologists today do not expect this particular application of cloud seeding to be effective in hurricanes because, contrary to the early beliefs, the storms contain little supercooled water vapor.

Chaotic Weather

OUR CURRENT STUDIES grew out of an intuition I had 30 years ago when I was a graduate student learning about chaos theory. A chaotic system is one that appears to behave randomly but is, in fact, governed by rules. It is also highly sensitive to initial conditions, so that seemingly insignificant, arbitrary inputs can have profound effects that lead quickly to unpredictable consequences. In the case of hurricanes, small changes in such features as the ocean's temperature, the location of the large-scale wind currents (which drive the storms' movements), or even the shape of the rain clouds spinning around the eye can strongly influence a hurricane's potential path and power.

The atmosphere's great sensitivity to tiny influences—and the rapid compounding of small errors in weather-forecasting models—is what makes long-range forecasting (more than five days in advance) so difficult. But this sensitivity also made me wonder whether slight, purposely applied inputs to a hurricane might generate powerful effects that could influence the storms, whether by steering them away from population centers or by reducing their wind speeds.

I was not able to pursue those ideas back then, but in the past decade computer simulation and remote-sensing technologies have advanced enough to renew my interest in large-scale weather control. With funding support from the NASA Institute for Advanced Concepts, my co-workers and I at Atmospheric and Environmental Research (AER), an R&D consulting firm, are employing detailed computer models of hurricanes to try to identify the kinds of actions that might eventually be attempted in the real world. In particular, we use weather-forecasting technology to simulate the behavior of past hurricanes and then test the effects of various interventions by observing changes in the modeled storms.

Modeling Chaos

EVEN TODAY'S BEST weather prediction computer models leave much to be desired when it comes to forecasting, but with effort they can be useful for modeling these storms. The models depend on numerical methods that simulate a storm's complex development process by computing the estimated atmospheric conditions in brief, successive time steps. Numerical weather prediction calculations are based on the assumption that within the atmosphere there can be no creation or destruction of mass, energy, momentum and moisture. In a fluid system like a hurricane, these conserved quantities are carried along with the storm's flow. Near the boundaries or margins of the system, however, things get more complicated. At the sea surface, for example, our simulations account for the atmosphere gaining or losing the four basic conserved quantities.

Modelers define the atmospheric state as a complete specification of the measurable physical variables, including pressure, temperature, relative humidity, and wind speed and direction. These quantities correspond to the conserved physical properties on which the computer simulations are based. In most weather models these observable variables are defined on a three-dimensional grid representation of the atmosphere, so one can plot a map of each property for each elevation. Modelers call the collection of values of all the variables at all the grid points the model state.

To generate a forecast, a numerical weather prediction model repeatedly advances the model state from one instant through a small time step (a few seconds to a few minutes depending on the scales of motion resolved by the model). It calculates the effects during each time step of winds carrying along the various atmospheric properties and of the processes of evaporation, rainfall, surface friction, infrared cooling and solar heating that occur in the area of interest.

Unfortunately, meteorological forecasts are imperfect. In the first place, the beginning model state is always incomplete and inexact. Initial states for hurricanes are particularly difficult to define because direct observations are few and difficult to make. Yet we do know from satellite cloud images that hurricanes have complex and detailed structures. Although these cloud images are potentially useful, we need to know much more. Second, even with a perfect initial state, computer models of severe tropical storms are themselves prone to error. The atmosphere, for example, is modeled only at a grid of points. Features smaller than the grid length, the distance between two neighboring grid points, cannot be handled correctly. Without very high resolution, a hurricane's structure near the eye wall—its most important feature—is smoothed out and the details are unclear. In addition, the models, just like the atmosphere they simulate, behave in a chaotic fashion, and inaccuracies from both these error sources grow rapidly as the forecast computations proceed.

Despite its limitations, this technology is still valuable for our purposes. We have modified for our experiments a highly effective forecast initialization system called four-dimensional variational data assimilation (4DVAR). The fourth dimension to which the name refers is time. Researchers at the European Center for Medium-Range Weather Forecasts, one of the world's premier meteorological centers, use this sophisticated technique to predict the weather every day. To make best use of all the observations collected by satellites, ships, buoys and airborne sensors before the forecast begins, 4DVAR combines these measurements with an educated first guess of the initial atmospheric state—a process called data assimilation. This first guess is usually a six-hour forecast valid at the time of the original observations. Note that 4DVAR accounts for each observation just when it was taken rather than grouping them across a time interval of several hours. The result of merging the observational data and the first guess is then used to initiate the subsequent six-hour forecast.

In theory, data assimilation produces an optimal approximation of the weather in which the fit of the model's representation to the observations is balanced against its fit to the first guess. Although the statistical theory for this problem is clear, the assumptions and information needed for its proper application are only approximate. As a result, practical data assimilation is part art and part science.

Specifically, 4DVAR finds the atmospheric state that satisfies the model equations and that is also close to both the first guess and the real-world observations. It accomplishes this difficult task by back-adjusting the original model state at the start of the six-hour interval according to the difference between observations and model simulation made during that period. In particular, 4DVAR employs these differences to calculate the model's sensitivity—how small changes in each of the parameters would affect the degree to which the simulation fit the observations. This computation, using the so-called adjoint model, runs backward in time over the six-hour interval. An optimization program then chooses the best adjustments to make to the original model state to achieve a simulation that most closely matches the progress of the actual hurricane during the six-hour period.

Because this adjustment is made using an approximation of the model equations, the entire process—the simulation, the comparisons, the adjoint model and the optimization—must be repeated again and again to fine-tune the results. When the process is complete, the conditions of the simulation at the end of the

The altered version of Hurricane Iniki veered off, so that Kauai escaped the storm's most damaging winds.

THE AUTHOR

ROSS N. HOFFMAN is a principal scientist and vice president for research and development at Atmospheric and Environmental Research (AER) in Lexington, Mass. His primary areas of interest cover objective analysis and data assimilation methods, atmospheric dynamics, climate theory and atmospheric radiation. He has been a member of several NASA science teams and was a member of the National Research Council Committee on the Status and Future Directions in U.S. Weather Modification Research and Operations. Hoffman would like to thank the NASA Institute for Advanced Concepts for supporting his work as well as his AER colleagues, particularly John Henderson, for their efforts in this research.

six-hour period then supplies the first guess for the next six-hour interval.

After simulating a hurricane that occurred in the past, we can then change one or more of its characteristics at any given time and examine the effects of these perturbations. It turns out that most such alterations simply die out. Only interventions with special characteristics—a particular pattern or structure that induces self-reinforcement—will develop sufficiently to have a major effect on a storm. To get an idea of what this means, think of a pair of tuning forks, one vibrating, the other stationary. If the forks are tuned to different frequencies, the second fork does not move, despite being struck repeatedly by sound waves emitted by the first. But if the devices share the same frequency, the second fork will respond in a resonant manner and vibrate sympathetically. In an analogous fashion, our challenge is to find just the right stimuli—changes to the hurricane—that will yield a robust response that leads to the desired results.

Calming the Tempest

TO EXPLORE WHETHER the sensitivity of the atmospheric system could be exploited to modify atmospheric phenomena as powerful as hurricanes, our research group at AER conducted computer simulation experiments for two hurricanes that occurred in 1992. When Hurricane Iniki passed over the Hawaiian island of Kauai in September of that year, several people died, property damage was enormous and entire forests were leveled. Hurricane Andrew, which struck Florida just south of Miami the month before, left the region devastated.

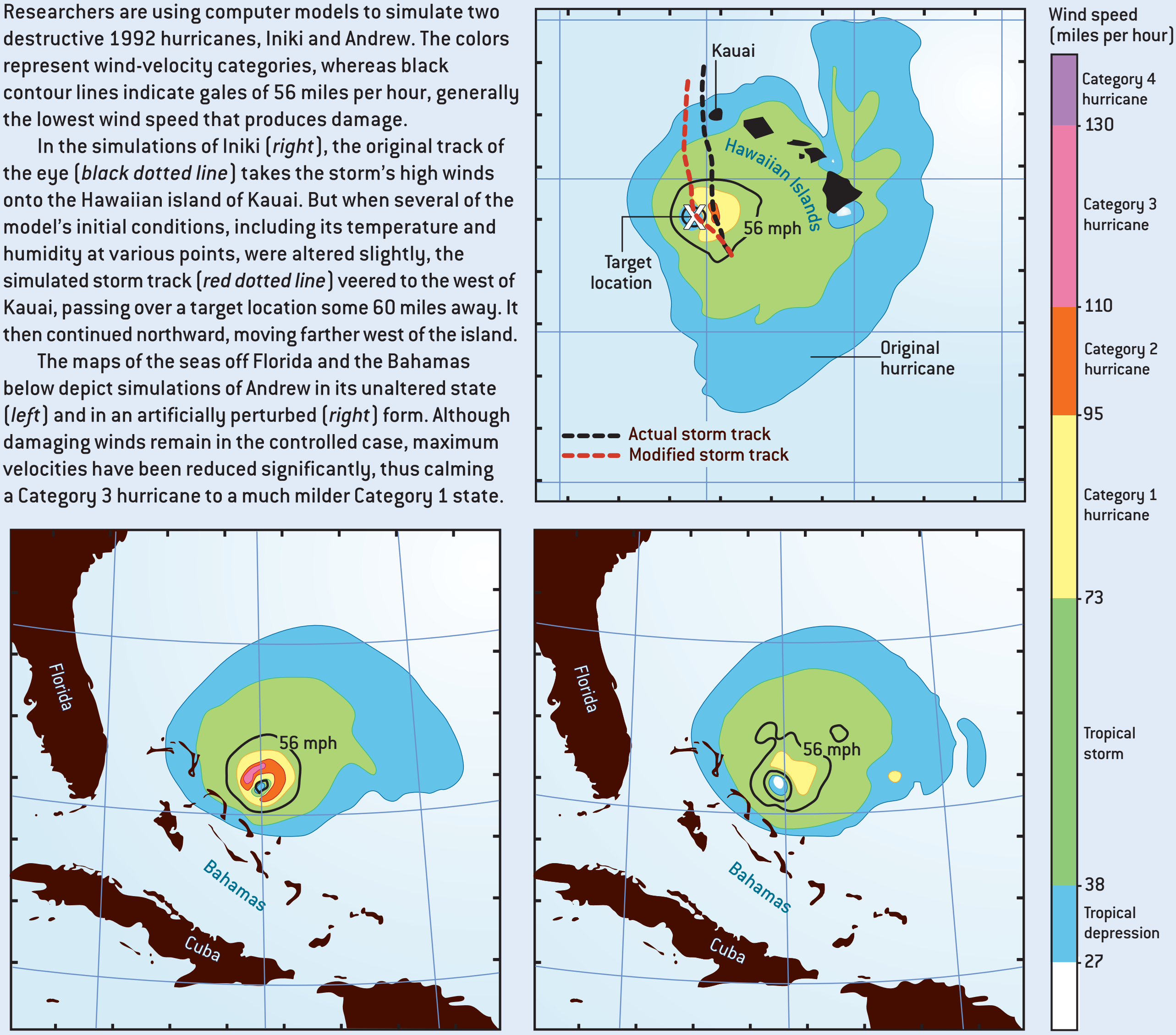
Surprisingly, given the imperfections of existing forecasting technologies, our first simulation experiment was an im-

CONTROL OF SIMULATED HURRICANES

Researchers are using computer models to simulate two destructive 1992 hurricanes, Iniki and Andrew. The colors represent wind-velocity categories, whereas black contour lines indicate gales of 56 miles per hour, generally the lowest wind speed that produces damage.

In the simulations of Iniki (*right*), the original track of the eye (*black dotted line*) takes the storm's high winds onto the Hawaiian island of Kauai. But when several of the model's initial conditions, including its temperature and humidity at various points, were altered slightly, the simulated storm track (*red dotted line*) veered to the west of Kauai, passing over a target location some 60 miles away. It then continued northward, moving farther west of the island.

The maps of the seas off Florida and the Bahamas below depict simulations of Andrew in its unaltered state (*left*) and in an artificially perturbed (*right*) form. Although damaging winds remain in the controlled case, maximum velocities have been reduced significantly, thus calming a Category 3 hurricane to a much milder Category 1 state.



LUCY READING

mediate success. To alter the path of Iniki, we first chose where we wanted the storm to end up after six hours—about 60 miles west of the expected track. Then we used this target to create artificial observations and fed these into 4DVAR. We set the computer to calculate the smallest change to the initial set of the hurricane's key defining properties that would yield a track leading to the target location. In this early experiment we permitted any kind of possible artificial alteration to the storm system to take place.

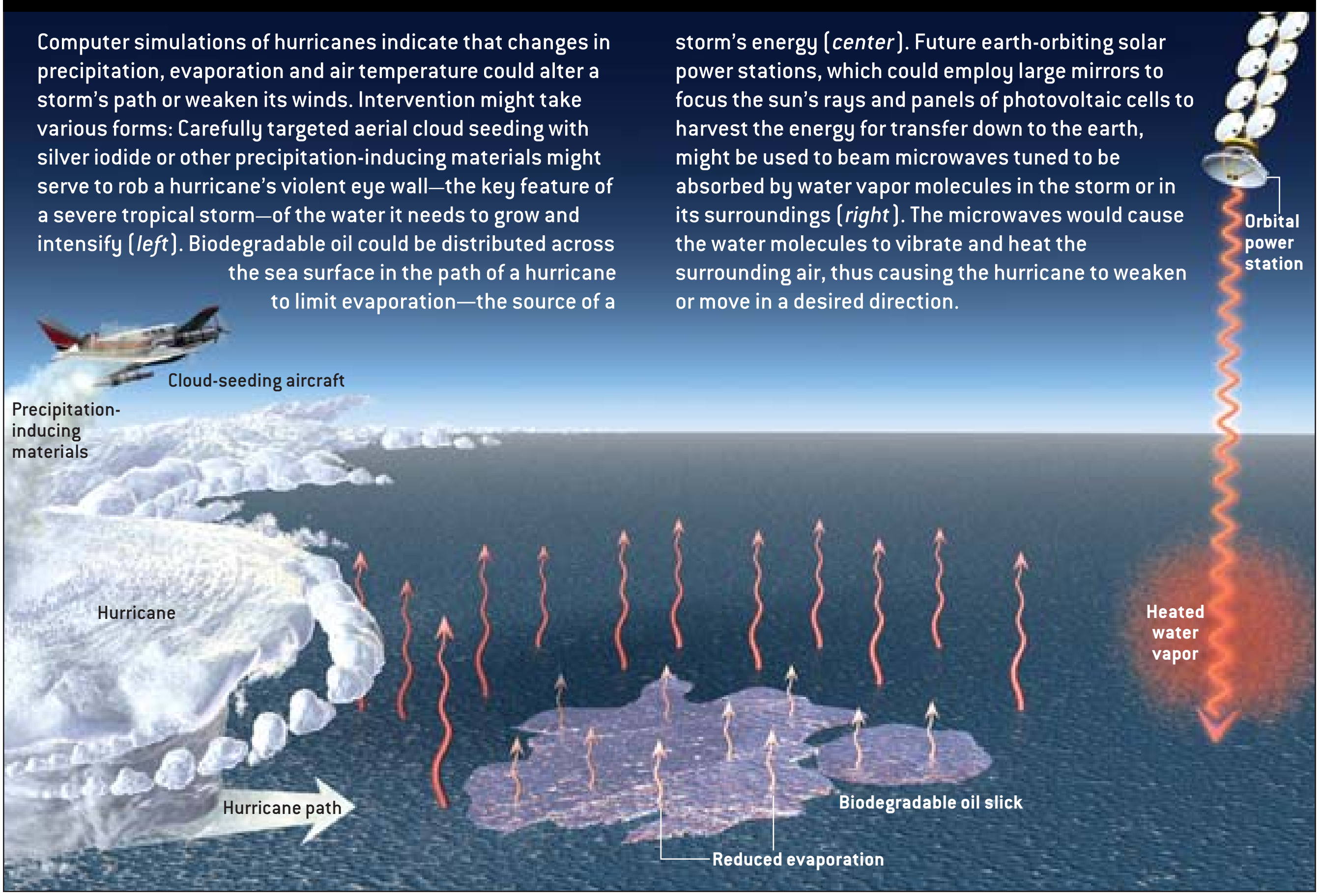
The most significant modifications proved to be in the starting temperatures and winds. Typical temperature adjustments across the grid were mere tenths of a degree, but the most notable change—an increase of nearly two degrees Celsius—occurred in the lowest model layer west of the storm center. The calculations yielded wind-speed alterations of two or three miles per hour. In a few locations, though, the velocities changed by as much as 20 mph because of minor redirections of the winds near the storm's center.

Although the original and altered versions of Hurricane Iniki looked nearly identical in structure, the changes in the key variables were large enough that the latter veered off to the west for the first six hours of the simulation and then traveled due north, so that Kauai escaped the storm's most damaging winds. The relatively small, artificial alterations to the storm's initial conditions had propagated through the complex set of nonlinear equations that simulated the storm to result in the desired relocation after six hours. This run gave us confidence that we were on the right path to determining the changes needed to modify real hurricanes. For the subsequent hurricane simulation trials, our team used higher grid resolutions to model the hurricane and set 4DVAR to the goal of minimizing property damage.

In one experiment using the modified code, we calculated the temperature increments needed to limit the surface wind damage caused by Hurricane Andrew as it hit the Florida coast. Our goal was to keep the initial temperature perturbation to a minimum (to make it as easy to accomplish as possible in real life) and to curtail the most destructive winds over the last two hours of the first six-hour interval. In this trial, 4DVAR determined that the best way to limit wind damage would be to make the greatest modifications to the beginning temperature near the storm's eye. Here the simulation produced changes as large as two or three degrees C at a few locations. Smaller temperature alterations (less than 0.5 degree C) extended out 500 to 600 miles from the eye. These perturbations feature a wavelike pattern of alternating rings of heating and cooling centered on the hurricane. Although only temperature had been changed at the start, all key variables were

HURRICANE INTERVENTION

Computer simulations of hurricanes indicate that changes in precipitation, evaporation and air temperature could alter a storm's path or weaken its winds. Intervention might take various forms: Carefully targeted aerial cloud seeding with silver iodide or other precipitation-inducing materials might serve to rob a hurricane's violent eye wall—the key feature of a severe tropical storm—of the water it needs to grow and intensify (*left*). Biodegradable oil could be distributed across the sea surface in the path of a hurricane to limit evaporation—the source of a storm's energy (*center*). Future earth-orbiting solar power stations, which could employ large mirrors to focus the sun's rays and panels of photovoltaic cells to harvest the energy for transfer down to the earth, might be used to beam microwaves tuned to be absorbed by water vapor molecules in the storm or in its surroundings (*right*). The microwaves would cause the water molecules to vibrate and heat the surrounding air, thus causing the hurricane to weaken or move in a desired direction.



DAVID FIERSTEIN

soon affected. In the case of the original simulated hurricane, damaging winds (greater than about 56 mph) covered populated areas in South Florida by the end of six hours, but in the altered model run, they did not do so.

As a test of the robustness of these results, we applied the same perturbation to a more sophisticated, higher-resolution version of the model. We obtained very similar results, which show that our experiments are reasonably insensitive to our particular choice of model configuration. After six hours, however, damaging winds reappeared in the altered simulation, so additional interventions would have been required to keep South Florida safe. Indeed, it looks as if a series of planned disturbances would be required to control a hurricane for any length of time.

Who Can Stop the Rain?

IF IT IS TRUE, as our results suggest, that small changes in the temperature in and around a hurricane can shift its path in a predictable direction or slow its winds, the question becomes, How can such perturbations be achieved? No one, of course, can alter the temperature throughout something as large as a hurricane instantaneously. It might be possible, however, to heat the air around a hurricane and thus adjust the temperature over time.

Our team plans to conduct experiments in which we will calculate the precise pattern and strength of atmospheric heating needed to moderate hurricane intensity or alter its track. Undoubtedly, the energy required to do so would be huge, but an array of earth-orbiting solar power stations could eventually be used to supply sufficient energy. These power-generating satellites might use giant mirrors to focus sunlight on solar cells and then beam the collected energy down to microwave receivers on the ground. Current designs for space solar power stations would radiate microwaves at frequencies that pass through the atmosphere without heating it, so as to not waste energy. For weather control, however, tuning the microwave downlink to frequencies better absorbed by water vapor could heat different levels in the atmosphere as desired. Because raindrops strongly absorb microwaves, parts of the hurricane inside and beneath rain clouds would be shielded and so could not be heated in this way.

In our previous experiments, 4DVAR determined large temperature changes just where microwave heating could not work, so we ran an experiment in which we forced the temperature in the center of the hurricane to remain constant during our calculation of the optimal perturbations. The final results resembled those of the original, but to compensate for making no initial temperature changes in the storm center the remaining temperature changes had to be larger. Notably, temperature changes developed rapidly near the storm center during the simulation.

Another potential method to modify severe tropical storms would be to directly limit the availability of energy by coating the ocean surface with a thin film of a biodegradable oil that slows evaporation. Hurricanes might also be influenced by introducing gradual modifications days in advance of their approach and thousands of miles away from their eventual targets. By altering air pressure, these efforts might stimulate changes in the large-scale wind patterns at the jet-stream level, which can have major effects on a hurricane's intensity and track. Further, it is possible that relatively minor alterations to our normal activities—such as directing aircraft flight plans to precisely position contrails and thus increase cloud cover or varying crop irrigation practices to enhance or decrease evaporation—might generate the appropriate starting alterations.

What if Control Works?

IF METEOROLOGICAL control does turn out to work at some point in the future, it would raise serious political problems. What if intervention causes a hurricane to damage another country's territory? And, although the use of weather modification as a weapon was banned by a United Nations Convention in the late 1970s, some countries might be tempted.

Before those kinds of concerns arise, however, our methods would need to be proved on atmospheric phenomena other than hurricanes. In fact, we believe our techniques should first be tried out in an effort to enhance rainfall. This approach could then serve as a test bed for our concepts in a relatively small region that could be instrumented densely with sensors. For such reduced size scales, perturbations could be introduced from aircraft or from the ground. If our understanding of cloud physics, computer simulation of clouds and data assimilation techniques advance as quickly as we hope, these modest trials could be instituted in perhaps 10 to 20 years. With success there, larger-scale weather control using space-based heating may become a reasonable goal that nations around the globe could agree to pursue. SA

Small changes can strongly influence a hurricane's potential path and power.

MORE TO EXPLORE

The Rise and Fall of Weather Modification: Changes in American Attitudes toward Technology, Nature, and Society. Chunglin Kwa in *Changing the Atmosphere: Expert Knowledge and Environmental Governance.* Edited by Clark A. Miller and Paul N. Edwards. MIT Press, 2001.

Controlling the Global Weather. Ross N. Hoffman in *Bulletin of the American Meteorological Society,* Vol. 83, No. 2, pages 241–248; February 2002. Available at **http://ams.allenpress.com**

Critical Issues in Weather Modification Research. Michael Garstang et al., National Research Council of the National Academies of Sciences. National Academies Press, Washington, D.C., 2003. Available at **www.nap.edu/books/0309090539/html**

NOAA's Hurricane Research Division: **www.aoml.noaa.gov/hrd/tcfaq**

Ross N. Hoffman's technical presentations on weather modification can be found at **www.niac.usra.edu/studies/**

The Internet of Things

The principles that gave rise to the Internet are now leading to a new kind of network of everyday devices, an "Internet-0"

By Neil Gershenfeld, Raffi Krikorian and Danny Cohen

In Barcelona about a century ago, Antoni Gaudí pioneered a fluid building style that seamlessly integrated visual and structural design. The expressive curves of his buildings were not just ornamental facades but also integral parts of the load-bearing structure. Unfortunately, a similar unification has yet to happen for the electronic infrastructure in a building. Switches, sockets and thermostats are grafted on as afterthoughts to the architecture, with functions fixed by buried wiring. Appliances and computers arrive as after-the-fact intrusions. Almost nothing talks to anything else, as evidenced by the number of devices in a typical house or office with differing opinions as to the time of day.

These inconveniences have surprisingly broad implications for construction economics, energy efficiency, architectural expression and, ultimately, quality of life. In the U.S., building buildings is a $1-trillion industry. Of that, billions are spent annually on drawing wiring diagrams, then following, fixing and revising them. Over the years, countless "smart home" projects have sought to find new applications for intelligent building infrastructure—neglecting the enormous existing demand for facilities that can be programmed by their occupants rather than requiring contractors to fix their functionality in advance.

Any effort to meet that demand, though, will be doomed if a lightbulb requires a skilled network engineer to install it and the services of a corporate IT department to manage it. The challenge of improving connectivity requires neither gigabit speeds nor gigabyte storage but rather the opposite: dramatic reductions in the cost and complexity of network installation and configuration.

Over the years, a bewildering variety of standards have been developed to interconnect household devices, including X10, LonWorks, CEBus, BACnet, ZigBee, Bluetooth, IrDA and HomePlug. The situation is analogous to that in the 1960s when the Arpanet, the Internet's predecessor, was developed. There were multiple types of computers and networks then, requiring special-purpose hardware to bridge these islands of incompatibility.

The solution to building a global network out of heterogeneous local networks, called internetworking, was found in two big ideas. The first was packet switching: data are chopped up into packets that can be routed independently as needed and then recombined. This technique marked a break from the traditional approach, used in telephone networks, of dedicating a static circuit to each connection. The second idea was the "end-to-end" principle: the behavior of the network should be determined by what is connected to it rather than by its internal construction, a concept embodied in the Internet Protocol (IP). Gradually the Internet expanded to handle applications ranging from remote computer access to e-commerce to interactive video. Each of these services introduced new types of data for packets to carry, but engineers did not need to change the network's hardware or software to implement them.

These principles have carried the Internet through three decades of growth spanning seven orders of magnitude in both performance and size—from the Arpanet's 64 sites to today's 200 million registered hosts. They represent timeless insights into good system design, and, crucially, they contain no specific performance requirements. With great effort and discipline, technology-dependent parameters were kept out of the specifications so that hardware could evolve without requiring a revision of the Internet's basic architecture.

These same ideas can now solve the problem of connecting

EVEN SOMETHING AS SIMPLE as a lightbulb could be connected directly to the Internet, if suitably equipped with cheap circuitry that sends signals along the electrical wiring.

heterogeneous devices rather than heterogeneous networks. Extending the Internet all the way down to an individual lightbulb requires recognizing the similarities, and differences, between a bulb and the mainframe computers for which the Internet was originally developed.

Smart Spaces

WE ENCOUNTERED the opportunity, and demand, for incorporating the Internet into a physical infrastructure through a series of installations we did with colleagues around the world. One, at a demonstration of future technologies for the White House/Smithsonian millennium events, was a smart bathroom shelf that detected pill bottles. It could remind people when to take their medication, let the pharmacist know when a refill was needed and help a doctor supervise care. Such a system could assist with health care compliance, which represents one of the greatest social and economic costs associated with aging.

Another installation, at New York City's Museum of Modern Art in 1999, used the furniture in a gallery to guide visitors through information about the exhibits. The idea was to avoid intruding on the visual and social space of the museum with conventional computer interfaces. At the opening, a museum benefactor proclaimed, "This is great, because I hate computers, and there are no computers here!"—not realizing that the furniture contained 17 Internet-connected embedded computers communicating with hundreds of sensor microcomputers.

Then came a building in the "Media House" demonstration in 2001 in, fittingly enough, Barcelona. The structural supports of this installation carried not just weight but also electricity and data. Lights and switches contained computers that allowed them to interact with one another and with other computers over the network. The associations between lights and switches were established on the fly.

Overview/*Internet-Zero*

- Giving everyday objects the ability to connect to a data network would have a range of benefits: making it easier for homeowners to configure their lights and switches, reducing the cost and complexity of building construction, assisting with home health care. Many alternative standards currently compete to do just that—a situation reminiscent of the early days of the Internet, when computers and networks came in multiple incompatible types.
- To eliminate this technological Tower of Babel, the data protocol that is at the heart of the Internet can be adopted to represent information in whatever form it takes: pulsed electrically, flashed optically, clicked acoustically, broadcast electromagnetically or printed mechanically.
- Using this "Internet-0" encoding, the original idea of linking computer networks into a seamless whole—the "Inter" in "Internet"—can be extended to networks of all types of devices, a concept known as interdevice internetworking.

At an opening event for the Media House, one of the leaders of the high-speed Internet-2 project was on hand. He kept asking how fast data could be sent through the building. Someone reminded him that lightbulbs do not need to watch movies at broadband speeds and joked that the network of everyday devices was part of an "Internet-Zero," not Internet-2. The name stuck.

The IP processors that were developed to support these demonstrations were themselves not a research project, but the recurring interest in them led to the launch of the Internet-0 (I0) project. Those devices embodied seven principles that together extend the original notion of internetworking to interdevice internetworking.

The Sevenfold Way

FIRST, EACH I0 DEVICE uses IP. In contrast, the many competing approaches for connecting devices introduce alternative standards. If a computer wants to communicate with one of these devices, the message must first be translated from IP into another protocol—a task that requires a special interface. Designers took this approach out of a belief that IP would be too demanding to implement in simple devices. But that need not be so. The code to run IP can be squeezed into a few kilobytes and run on a one-dollar microcontroller. The IP information adds about 100 bits to each message, which typically has a negligible impact on the response time and power requirements. In return for this modest overhead, the network avoids the cost of configuring and maintaining complex interfaces.

Second, the software is simplified by implementing the communications protocols jointly rather than separately. In a conventional computer, the tasks associated with networking are rigidly segregated. Low-level code handles the actual signals, such as the generation of electrical pulses sent over an Ethernet cable or through a telephone modem. That code sends its output to another layer of software that encodes and decodes the data. Above that are layers that supervise the sending and receiving of packets, assemble and disassemble the packets, and interpret standards for the content of the packets. Finally, the data reach the application, such as a Web browser.

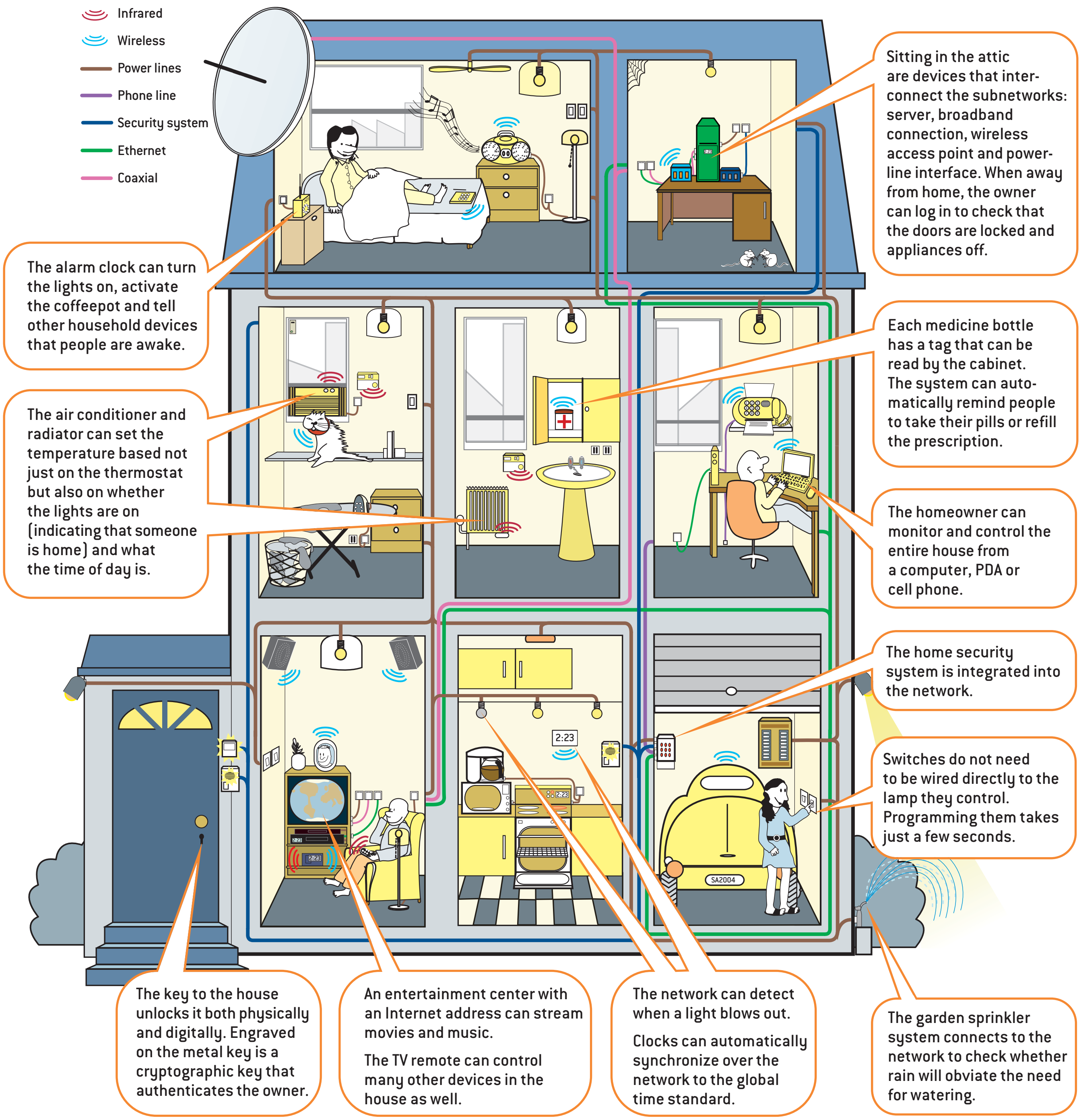
Each of these layers is implemented separately as a kind of software version of a human bureaucracy; much of the code is dedicated to interlayer message passing. The layers are useful abstractions for developing the standards, so that one of them can be changed without needing to modify the rest, but that generality does not need to be preserved when they are actually executed. In an I0 device, the software implementation takes advantage of knowledge of the application.

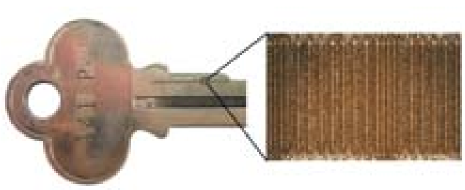
Third, two I0 devices do not require the existence of a third one in order to operate. Most computers on the Internet are either clients (such as Web browsers) or servers; the clients are useless without the servers. But each of the I0 lights and switches stores the data and routines it needs rather than relying on a central server, which would reduce reliability and increase costs. Although servers could enhance the system's value—say, by turning all the lights on or off at a particular time of the day—they are not necessary for it to function.

SLIM FILMS (*preceding page*)

One Network to Connect Them All

Internet-0 allows myriad devices to intercommunicate and interoperate: pill bottles can order refills from the pharmacy; light switches and thermostats can talk to lightbulbs and heaters; people can check on their homes from their offices. Existing technologies already allow many of these functions, but Internet-0 provides a single consistent standard. It can handle information sent through the AC power line, over a wireless connection or even engraved on a metal key, and it seamlessly integrates with the local and global computer networks. Devices can be configured by interacting with them rather than by typing on computers.





Internet-0 packets can be engraved onto a key (*left*) or printed as a bar code (*below*). In both cases, the vertical bars represent pulses and, once converted into electrical signals, can be put onto the network without further translation. The key engraving is a cryptographic key that allows the holder to reconfigure devices. The bar code contains "IO" preceded by the address of the destination device.

IP header — UDP header — "IO"

MANU PRAKESH (*key*); LUCY READING (*house*)

Whereabouts

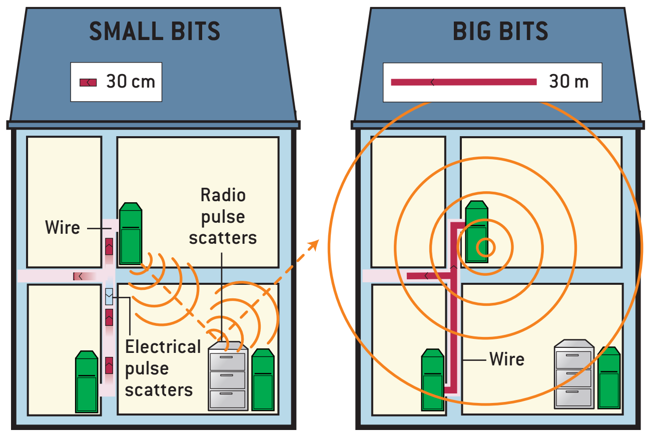
FOURTH, EACH I0 DEVICE is responsible for keeping track of its own identity. A networked computer has five different names: a hardware Media Access Control (MAC) for the physical address on the local network (such as "00:08:74:AC:05:0C"), an IP address on the global network ("18.7.22.83"), a network name ("www.mit.edu"), a functional name ("the third server from the left") and the name of a cryptographic key to communicate with it securely. Naming is one of the primary functions of servers. I0 devices must be able to manage those functions for themselves when a server is not present, as well as accept the answers from a server when one is.

The most common kind of hardware address is centrally managed by assigning manufacturers blocks of addresses to burn into their products. But that kind of coordination would not be feasible for every light and switch produced on the planet. Instead a device can simply choose a random string for its address. The probability of two devices picking the same 128-bit number, for example, is just one in 10^{38}. Users can assign functional and hardware names by interacting with a device—for instance, by pressing programming buttons on a light and switch that cause them to broadcast their addresses in succession and hence establish a control relation or by carrying the network address along with a cryptographic key between the devices to make the connection secure.

Fifth, I0 uses bits that are bigger than the network. Bits have a physical size—they are nothing more than electrical, radio or light pulses. The time it takes to send such a pulse, multiplied by the speed at which it travels (which is typically about the speed of light), is the size. It used to be that bits were bigger than the network over which they were sent. Now they are much smaller—30 centimeters long for a data rate of a gigabit per second. If the network is bigger than that, problems arise at the interfaces between pieces of the network. Slight mismatches in the transmission properties at the interfaces generate spurious signals. Moreover, two computers that start transmitting simultaneously may not discover the collision until after they have sent many bits. That is why high-speed networks require special cables, active hubs, agile transceivers and skilled installers. But at a megabit per second, which is roughly the speed of a typical home cable modem or DSL connection (certainly enough for a lightbulb), a bit is 300 meters long—so large that it can span an entire building network. It no longer matters what interfaces the network has.

When Slower Is Better

When data are transmitted slowly, the size of the corresponding electrical or radio pulses is large, greatly simplifying the functioning of a computer network. Whereas small pulses reverberate off interfaces such as metal cabinets (in the case of radio signals) or wire junctions (for electrical signals), large pulses literally fill every inch of the air or wiring within a house. Therefore, they do not require hubs and other special equipment.



On the Telegraph Road

SIXTH, THE USE of big bits allows the data that make up a packet to be represented in the same way no matter what physical medium conveys them. When bits are small, their physical representation, or modulation, must be customized to each communications channel. The modulation used by a dial-up modem is very different from that used by a cable modem, because a twisted-pair phone line and a coaxial cable differ in the amplitude, frequency and phase of the signals they can carry. But when the bits are big, the detailed propagation characteristics of each channel make no difference.

Morse Code exploits this principle. It can be tapped on a telegraph, flashed from ship to ship or banged out on a pipe. These very different channels carry the same data at the same rate using the same encoding scheme. No translation is needed. The information is conveyed simply by the arrival time of a change in the physical medium—a voltage jump on the telegraph line or the sudden onset of sound. The precise tone or amplitude does not matter.

I0 is similar except that it uses "0" and "1" impulses instead of dots and dashes. Like today's modems, an I0 device sends a packet in a string of eight-bit bytes, each framed by start and stop bits. A 0 bit is represented by an impulse followed by a

THE AUTHORS

NEIL GERSHENFELD, RAFFI KRIKORIAN and *DANNY COHEN* are researchers that seem to thrive by defying traditional disciplinary boundaries. Gershenfeld directs the Center for Bits and Atoms at the Massachusetts Institute of Technology, which gets support from the National Science Foundation. He studies the relation between physical form and logical function in everything from quantum computers to automobile safety systems, from a computerized cello he developed for Yo-Yo Ma to tools used by rural Indian villagers. Krikorian is an M.I.T. graduate student who has led the development of hardware and software for Internet-0. His earlier work in academia and industry spanned very small embedded IP stacks to very large distributed computation engines. The Internet-0 project grew out of their collaboration with Cohen, a Distinguished Engineer at Sun and one of the fathers of the Internet. He pioneered real-time interactive applications over the Arpanet, which helped to lead to the development of the Internet Protocol (IP), and started the MOSIS IC fabrication service.

LUCY READING

pause; a 1 bit is a pause followed by an impulse; a start or stop bit is a pair of impulses. Such a scheme, called Manchester encoding, simplifies distinguishing between a valid 0 or 1 and an interfering or absent signal. In addition, the spacing between the pulses in the start bit allows the receiver to measure the transmission rate; the rate does not need to be fixed in advance. If extra noise immunity is needed, the sender and receiver can agree to use a procedure to vary the time between bytes (as is done in ultrawideband radios), which would help them separate the signal from noise while retaining backward compatibility with simpler I0 devices that use just the framing pulses.

As long as the bits are sent slowly enough to be bigger than the network, the encoding can be the same across all types of physical media. The pulses could be sent through a wire, coupled onto a power line, clicked by a speaker, printed onto a page or engraved on a key. Each of these media would pass a different part of the pulse: a power line filters out high frequencies and a radio antenna the low ones. All that is required is for some of the frequencies in the impulse to get through. (The detailed frequency response would be useful, however, if the I0 device needed to probe its physical environment.)

This representation extends the end-to-end principle of the Internet to modulation. When a computer transmits packets using IP, it does not need to know anything about the networks that will carry the packet. Similarly, when a device uses I0 impulses, it does not need to know about the media that will carry the signal.

Less Is More

THE SEVENTH AND FINAL attribute of I0 is the use of open standards. The desirability of open standards should not need saying, but it does. Many of the competing standards for connecting devices are proprietary. The recurring lesson of the computer industry has been that proprietary businesses should be built on top of, rather than in conflict with, open standards.

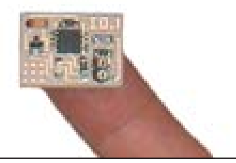
As an example of I0 in action, return now to the bathroom shelf for managing medication. Our demonstration project used radio-frequency identification (RFID) tags in the pill bottles—tiny disposable chips that are powered by the radio signals that interrogate them. We had to configure the tag reader to know what to do with the data it received. The same is true for the RFID systems now being deployed across consumer and military supply chains: an army of consultants and contractors is needed to configure all the RFID readers.

The process would be much easier using I0. The tag would encode an IP packet—one might call it an IPID tag—and the reader would merely have to relay the packet to the network. The packet might contain the addresses of the pharmacist and doctor rather than a prearranged code that the reader would have to translate into English. The information would be carried by the pill bottle rather than programmed into the reader.

If these features are such a good thing, then why were they not put into practice sooner? The problem is that communications engineers have had a longstanding bias that bandwidth is scarce and hence must be used efficiently, dating back to the days when it was. The developers of the original Ethernet were faulted at the time because it did not reach the fundamental communications limits set by quantum mechanics. That was true but irrelevant. Ethernet succeeded because of its relative simplicity.

Today networks are indeed approaching quantum limits, sacrificing simplicity for ever more spectacular performance gains. I0 reverses this trend. It is a technological case of less being more; speed is sacrificed for interoperability.

Data could be sent through a wire, CLICKED by a speaker, PRINTED onto a page or engraved on a key—all using the same Internet-0 encoding.



INTERNET-0 PROCESSOR

I0 is aimed at the scaling limits imposed by network complexity rather than raw performance. It is not intended to replace the existing Internet; it provides a compatible layer below it. An I0 device depends on the current routers, gateways and name servers to carry packets between I0 subnetworks. Over time, however, the distinction between I0 and the rest of the Net may recede. The protocols that run in the Internet's servers, such as the ones used to direct IP packets to their destinations, are described as algorithms—a set of instructions for finding the best path for a packet to take. But the protocols can also be understood as optimizations—a way to make best use of the available communications resources given their constraints. Recent research has shown how to solve such constrained optimizations using distributed systems rather than central processors. Thus, I0 nodes might one day be configured to solve global network management problems through their local behavior so that the higher-level Internet architecture can emerge from their interactions.

If so, the ultimate destiny of Internet-0 is not just turning on the lights. An I0 network will be indistinguishable from the computers that it connects; it really will be the computer. By allowing devices for communications, computation, storage, sensing and display to exchange information in exactly the same representation—around the corner or around the world—the components of a system can be dynamically assembled based on the needs of a problem rather than fixed by the boundaries of a box. SA

MORE TO EXPLORE

How the Internet Came to Be. Vinton Cerf in *The Online User's Encyclopedia.* Edited by Bernard Aboba. Addison-Wesley, 1993. Available at **www.internetvalley.com/archives/mirrors/cerf-how-inet.txt**

When Things Start to Think. Neil Gershenfeld. Henry Holt, 1999.

Other publications are available at **cba.mit.edu/projects/I0**

DYING

TO SEE

Studies of the lens of the eye not only could reveal ways to prevent cataracts but also might illuminate the biology of Alzheimer's, Parkinson's and other diseases in which cells commit suicide

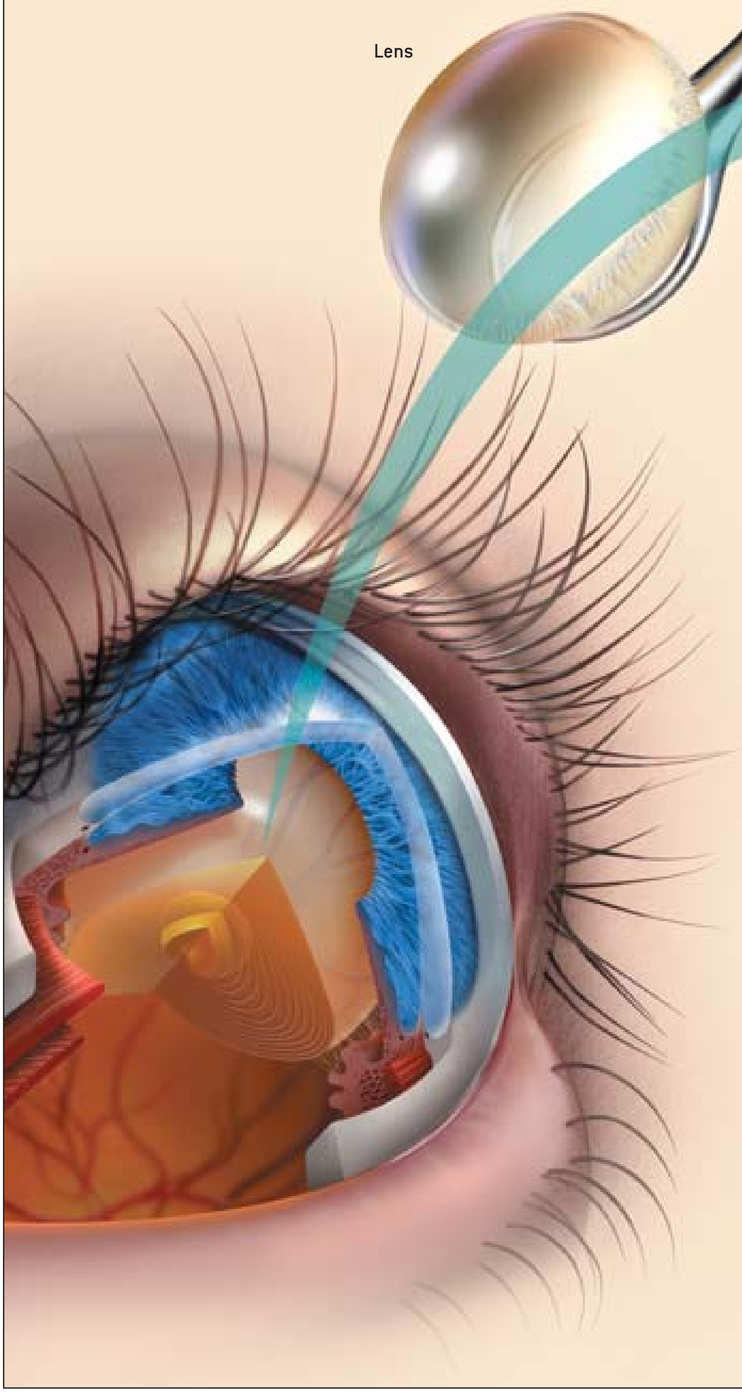
By Ralf Dahm

RENEE LYNN *Corbis* (*impalas and lionesses*); JANA BRENNING (*photocomposition*)

AGING AND DAMAGE to the eye's lens can cause cataracts and yellowing that ruin vision, which may put the viewer at risk.

The lens of the eye is the only transparent tissue in the human body. In the past few years, scientists have determined that this transparency—critical for focusing light—stems in large part from the unique ability of the lens to activate a self-destruct program in its cells that aborts just before completion, leaving empty but sustainable cells that transmit visible rays.

The eye lens is transparent both because of its architecture and because of its unusual developmental program. The cells of the fully formed lens fit together in a regular arrangement that limits the scattering of light (*diagram at right* and *micrographs at far right*). And those cells become free of light-obstructing material during development (*bottom right*) by initiating a suicide program that dissolves their innards but halts before the cells actually die.



A better understanding of how lens cells become and remain transparent should suggest ways to prevent lens-clouding cataracts. More than half of all Americans older than 65 develop these sight-blocking occlusions. The only recourse is to surgically remove the person's lens and insert an artificial implant, and even then, complications requiring a second operation occur in a large proportion of patients. Given that cataracts affect primarily older people, for whom any kind of surgery is worrisome, a method to slow, stop or reverse cataracts would be a great aid indeed.

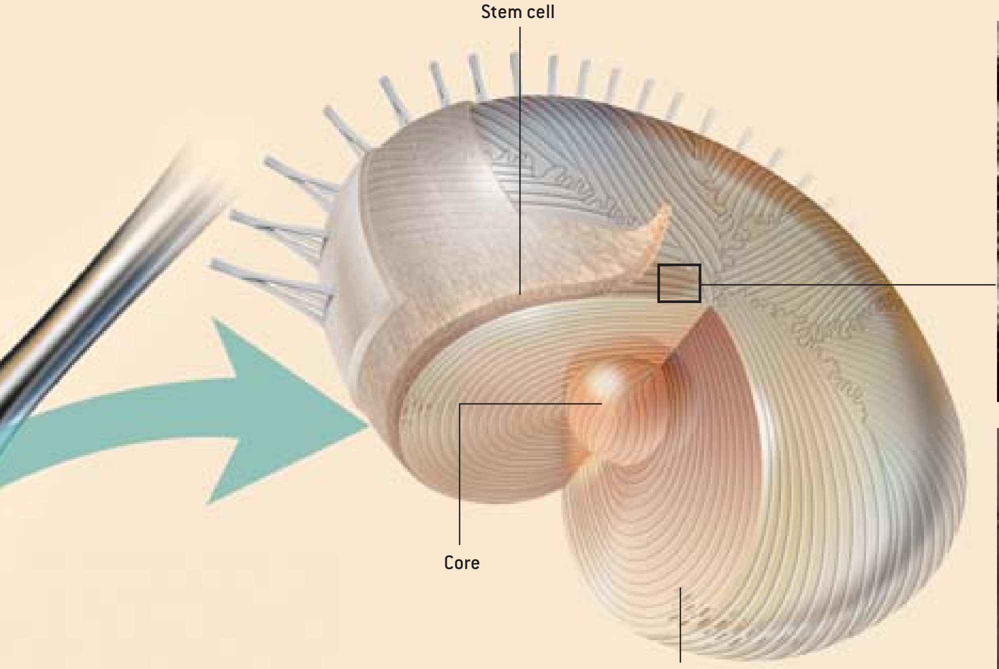
Beyond protecting vision, improved knowledge of how the lens tightly controls cell suicide could reveal ways to treat debilitating conditions characterized by excessive or inappropriate cell death, chief among them Parkinson's disease, Alzheimer's disease and chronic infections such as AIDS.

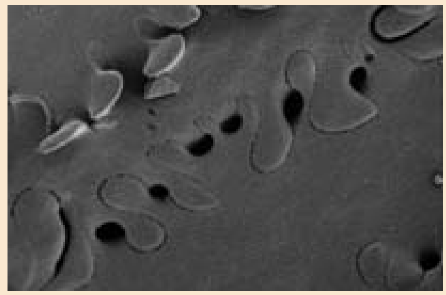
Barely Alive

THE EYE'S LENS is a biological marvel, being at once dense, flexible and clear. If it bore the slightest obscurities, our visual world would be a fun house of warped and blurred images and glare. And if the lens had any hint of color, it would absorb light, preventing us from seeing certain shades.

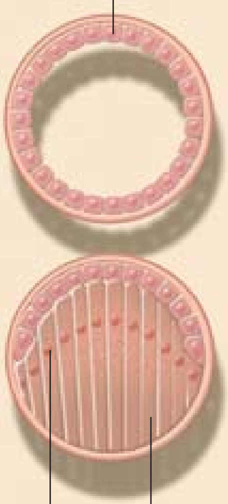
Many animals possess translucent parts, such as insect wings, but truly transparent tissue in nature is rare and difficult to achieve. In humans the cornea is clear, but it is more a thin, gelatinous layer of proteins and sugars than true cellular tissue. The lens comprises about 1,000 layers of perfectly clear, living cells. Other than vision, the only significant exploitation of transparency

THE LENS: KILLING ITSELF FOR CLARITY





Layers of lens cells align in parallel (*top*), so that light passes perpendicularly through them, as in this bovine lens. Within a layer (*bottom*), adjacent cells interlock like jigsaw puzzle pieces to prevent gaps from forming when the lens changes shape during focusing; the layering and interlocking of cells enables light to pass across cell boundaries without scattering.

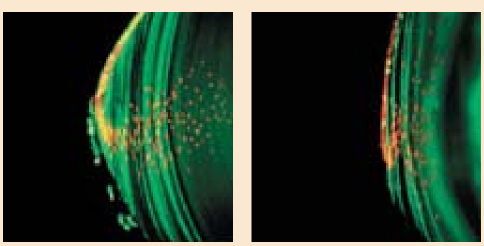


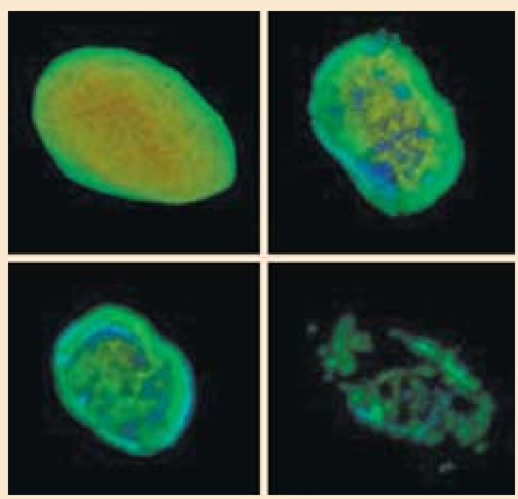
THE DEVELOPING LENS

Lens development begins in early embryos when undifferentiated (stem) cells lining a spherical vesicle (*left, top*) differentiate into lens cells that reach across the cavity (*left, bottom*). After this core forms, more stem cells differentiate into cells that elongate around the outside rim, adding layers in an onionlike manner (*above*). Initially these cells have a nucleus, mitochondria, endoplasmic reticulum and other typical organelles. But as they are encapsulated by newer cells, they degrade their organelles, leaving nothing but an outer membrane and a thick solution of special proteins called crystallins. This barely living material has a uniform index of refraction, so it does not scatter light.

Aspects of the process can be seen in a developing lens (*below, left*) and in a nearly fully developed lens (*below, right*) of a mouse. New cells stretch down across the equatorial region and effectively move inward as even newer cells cover them. Cell nuclei (*red*) traveling down and in persist for a time but dissolve as they are buried.

The nucleus of a developing lens cell dissolves over several days (*right*), with the nuclear envelope and DNA inside breaking down in tandem.





KEITH KASNOT (*drawings*); RALF DAHM (*top SEM, layers*); ALAN R. PRESCOTT *University of Dundee, Scotland* (*bottom SEM, jigsaw pieces*); FROM "DEVELOPMENT OF A MACROMOLECULAR DIFFUSION PATHWAY IN THE LENS," BY V. I. SHESTOPALOV AND S. BASSNETT, IN *JOURNAL OF CELL SCIENCE*, VOL. 15; 2003 (*bottom left two images, nucleated cells*); RALF DAHM (*bottom right four images, degrading nuclei*) in the natural world occurs among certain ocean and freshwater creatures, which use the trait to blend into the open water and hide from predators. Yet almost all these animals, such as jellyfish, qualify only as "very translucent," not totally see-through.

Transparency is unusual because cells have organelles—internal structures such as the nucleus (which stores DNA), the energy-producing mitochondria, and the Golgi apparatus and endoplasmic reticulum, which are important in the synthesis of proteins and lipids. Each structure has its own refractive index, and when a light ray crosses an area where the index changes, the light scatters, creating a degree of opaqueness.

In addition, some cells absorb certain wavelengths of light, resulting in color. The heme of the hemoglobin in blood cells gives them their characteristic red hue. Because organs and muscles have a blood supply, they appear primarily in shades of red, too. Furthermore, many cells, especially those in hair and skin, are populated with melanins—pigment molecules that come in colors ranging from red to black.

The lens has no melanins and no blood supply. Yet that alone is not enough for transparency. Cartilage has no melanins or blood supply and is colorless, but it is at best translucent. That is because in virtually all tissues, cells or fibers are oriented at various angles, creating different refractive indices that scatter light as it passes through. The lens is composed of only one cell type, and the cells are precisely aligned.

Given that lens cells have no blood supply, no connective or nervous tissue, and no organelles, can they even be considered alive? The answer depends on how "life" is defined. Lots of small animals without a blood supply are happily populating the planet. Human cartilage receives no blood, but any biologist would consider it living. If life means a cell has a metabolism, then lens cells are alive—albeit barely. Although they have no mitochondria to produce energy, certain nutrients and other molecules diffuse into the lens's outermost cells and slowly pass inward, cell to cell.

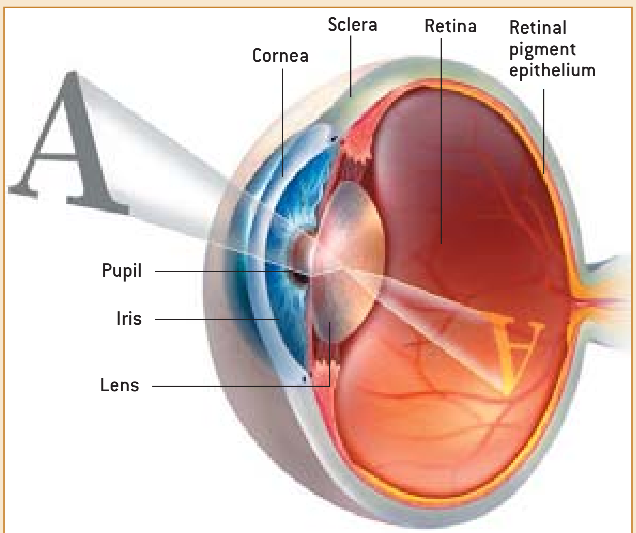
Young lens cells do have organelles when they first form from stem cells in a fetus, but the organelles are destroyed during early development. (The same occurs for new cells that are periodically laid down during adulthood.) What remains is a cytoplasm consisting of an unusually thick solution of special proteins called crystallins. Although the lens is often described as a crystal, it does not qualify in the chemical sense—where the geometric position of ions or molecules with respect to one another is systematically repeated. The lens is a "biological crystal"—that is, it has a very regular arrangement of cells. Each cell contains large molecules—crystallin proteins—that form complexes with paracrystalline arrangements. This construction makes the cytoplasm optically homogeneous; the refractive index does not change inside the cell or from one cell to another.

Brown Eyes Blue

Knowing how the eye focuses light (*diagram*) explains not only how you see but why your eyes may be brown, hazel or blue—or red in a photograph. The iris blocks incoming light, leaving a neat hole—the pupil—through which light rays strike the lens and are focused onto the retina. The rays that hit the iris are scattered back. The shorter the light's wavelength the greater the scattering, so blue light is scattered more than red, giving the iris a "natural" blue color. (The same principle causes the sky and sea to appear blue.) Yet the iris also contains melanin—pigment molecules that absorb various wavelengths. A lot of melanin will absorb much of the light, making the iris appear dark brown. Less melanin leads to lighter browns and greens, and very little melanin allows blue to dominate.

The pupil appears black because a melanin-rich layer of cells just behind the retina—the retinal pigment epithelium—absorbs all light that the retina has not. This absorption prevents light from randomly scattering back to the retina's photoreceptors, which would blur vision. (The black lining of a camera serves the same purpose.) Because no light is emitted back through the pupil, it appears black.

Albinos cannot synthesize melanin; their retinal pigment epithelium does not absorb much light and thus causes poor vision and near-blindness in bright light. As light scatters back toward the pupil and iris, it illuminates blood vessels, making them appear pink or red. A similar effect can occur during flash photography of any person: the flash is so bright that the epithelium cannot absorb all the rays, and backscatter creates "red eye" in the photograph. *—R.D.*



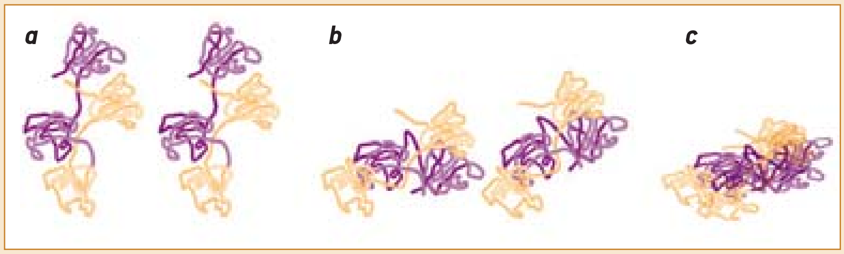
KEITH KASNOT

Seeing through a Glass, Dimly

CLARITY, OF COURSE, comes at a cost. Although lens cells survive the controlled suicide of organelles, this degradation has drastic implications. Without nuclei, the genetic programs for synthesizing new parts are gone. Mature lens cells cannot regenerate or repair themselves, as cells in other tissues do.

The ability to replace damaged parts is a prime advantage of biological systems. The molecules that compose human cells typically have half-lives lasting a few minutes to several days. Within six months or so, 90 percent of the molecules that make

How Cataracts Form



Cataracts in the lens blur the vision or blind millions of people every year. Lens cells contain a thick solution of large proteins called crystallins (*a*) in an ordered arrangement. Researchers are not yet sure why, but as crystallins accumulate damage, such as from ultraviolet light, oxidation or dehydration, they collapse into misfolded fibers (*b*). Then the misfolded proteins can aggregate into a tangled mass (*c*). The clumped mass blocks or distorts incoming light, creating a cloudy spot in a person's field of view (*photograph*). Elevated levels of misfolded proteins have been found in the brains of people with Alzheimer's or Parkinson's disease, prompting scientists to look deeper for common clues. *—R.D.*

ALICE CHEN (*drawing*); SCIENCE PHOTO LIBRARY (*photograph*)

up our bodies are replaced by new ones. Lens cells, however, must function for a lifetime—a spectacular span.

This lack of repair mechanism makes the cells vulnerable to certain stresses. For example, severe dehydration can cause crystallin proteins to precipitate, prompting their cells to crumble into a clump—a cataract. This speck disrupts the otherwise uniform index of refraction, creating a cloudy spot in a person's field of vision. Just a few weeks of extreme dehydration can initiate cataract formation.

Even in the absence of such conditions, the inability to repair means that over the long term, small insults accumulate. Regular exposure to highly reactive molecules such as oxygen free radicals, or to ultraviolet radiation, or to years of elevated blood sugar from diabetes eventually leads to cataracts in many people—and to many cataract operations.

References to the removal of clouded lenses date back as early as 1800 B.C. to the Babylonian Code of Hammurabi. Ancient Egyptian texts and medieval European and Islamic writings describe detaching the lens from the ciliary muscle and pushing it down into the vitreous humor—the thick fluid in the back of the eye. Although this procedure removed the veil from the light path, it left no lens to focus rays. Patients could see only blurred images, as if their eyes were open underwater.

The application of special spectacles in the 17th and 18th centuries finally compensated for the lost focusing power. Today's artificial lenses eliminate any need for glasses. Doctors perform more than one million cataract operations annually in the U.S. alone. Fortunately, the procedure now has a success rate of nearly 100 percent and takes no more than 45 minutes. Still, approximately one third of patients return with aftercataracts, caused by undifferentiated cells—stem cells—that are inadvertently left behind during surgery. These cells start proliferating, but in contrast to their behavior during embryonic development, they form a disorganized mass that obscures vision and has to be surgically removed. In those developing countries that lack surgical resources, cataracts account for half of all cases of blindness. In India alone, cataracts blind an estimated 3.8 million people every year.

In addition to becoming vulnerable to cataracts, the aging lens tends to yellow. Proteins that absorb blue and green light slowly accumulate, blocking these rays from reaching the retina and thereby giving the lens a yellow or brownish appearance. Only reds, yellows and browns pass through, altering a person's view of the world [*see box on next page*].

Controlled Suicide

IN RECENT YEARS, scientists have done much more than marvel at the lens's qualities and fret over its age-related decline. They are finding that the process by which the lens systematically destroys its organelles may offer a marvelous opportunity to solve some of humankind's most frustrating illnesses.

Like all cells, lens cells that arise from stem cells during early fetal development contain organelles. But as they differentiate, they demolish their organelles—and the rubble that remains—to become transparent. This may not seem problematic at first, but consider what happens when other cells encounter so much as a little damage to their DNA: they embark on an irreversible process called apoptosis, or pro-

THE AUTHOR

RALF DAHM is a project manager at the Max Planck Institute for Developmental Biology in Tübingen, Germany. He has a Ph.D. in biochemistry from the University of Dundee in Scotland and oversees a pan-European project to exploit zebra fish as a model for researching human development and diseases. He is co-editor of *Zebrafish: A Practical Approach* (Oxford University Press, 2002) and is author of a popular German-language science book about human embryonic development, stem cells and cloning. Dahm is also fascinated by how eye diseases change the way artists see and therefore render the world.

PAINTING THROUGH OLD EYES

CLAUDE MONET painted the Japanese bridge in his Giverny garden near Paris in 1899 (*left*). The same scene, which he attempted to capture again between 1918 and 1924, shows that cataracts had blurred his vision and that the yellowing of his lenses had impaired his perception of blues and greens, leaving him in a world filled with murky reds and browns.

French impressionist Claude Monet (1840–1926) reached the grand old age of 86. But advancing years seriously affected his eyesight. Cataracts clouded his vision, and the yellowing of the lenses in his eyes altered his color perception. His work over the final two decades of his life offers a vivid depiction of how these common impairments skew human sight.

The yellowing started first. Gradually, proteins that absorb the "cold colors" of violet, blue and, later, green accumulated in his lenses, blocking these light rays from reaching the retina. Red and yellow light still passed through, rendering Monet's world in increasingly warm tones.

Cataracts then clouded his vision, forcing him to perceive his surroundings as if he were looking through frosted glass. Over time he had trouble discerning shapes, normal daylight became blinding, and in the late stages he could only differentiate between light and dark.

Monet first noticed that his eyes were changing during a trip to Venice in 1908. The 68-year-old painter had difficulty selecting his colors. In 1912 Monet's doctor diagnosed a cataract in each eye and recommended surgery, but the artist was afraid; in his time any operation was fraught with problems, and removing a cataract frequently ended an artist's career.

From that point on, however, Monet's works show fewer details. Yellows, reds and browns predominate. When he examined his later pictures, he was often seized by a towering rage and a desire to destroy them. In early 1922 he wrote that he was no longer able to create anything of beauty.

Later that year Monet's right eye could only detect light and the direction from which it came; his left eye could only see about 10 percent of what is considered normal. In January 1923, at the age of 83, he finally had cataract surgery to his right eye, but he complained that the glasses he had to wear thereafter made colors appear peculiar.

In 1925 he finally found suitable spectacles and was delighted. He wrote that he could see well again and would work hard. Alas, he died a year later. *—R.D.*

grammed cell death. Destructive proteins released inside a cell chop up its DNA and key proteins, and the mitochondria shut down, depriving the cell of its energy source. The tattered cell breaks apart and dissolves. Ordinarily, damaged cells commit suicide to make room for new healthy cells—otherwise an organ with an accumulating number of damaged cells would not be able to function. In some cases, damaged cells kill themselves so they do not start proliferating and turn cancerous. Lens cells destroy the nucleus and every other organelle yet halt the process just before demolition is complete, leaving an intact outer membrane, an inner cytoskeleton of proteins and a thick crystallin plasma [*see box on pages 84 and 85*].

The ability to halt cellular suicide has come as quite a surprise. The scientific community had always viewed apoptosis as an unstoppable process. Yet some unknown mechanism in the lens controls the death machinery so it destroys only certain cell components while leaving others intact. Several years ago I, along with other lens specialists, began to suspect that a deliberate braking mechanism was involved. We showed that specific compartments of differentiating cells—the nucleus or mitochondria, say—succumb to the same destruction that occurs during the full apoptosis of mature cells. But other compartments such as the cytoskeleton are unaffected. The implication is that lens cells actually use the death machinery not to destroy themselves but to choreograph the differentiation process.

The next leap in thinking came quickly: a mechanism that could control apoptosis could alter the progression of diseases characterized by excessive cellular suicide, such as neurodegenerative disorders. To harness this power, researchers must find the signals—or blockers—that stop total destruction. Similarly, discovering what triggers lens cells to degrade their organelles could suggest new ways to induce cancer cells to commit suicide.

Pieces of evidence are accumulating. One theory advanced by Steven Bassnett of Washington University to explain the onset of apoptosis holds that during development, as new lens cells are formed around existing ones—like new layers around an onion core—the older internal cells become further removed from the surface, and the amount of oxygen that reaches them decreases. If the concentration drops below a threshold, the integrity of the mitochondria, which rely on an oxygen supply for energy production, might be compromised. Sensing this problem, the cell triggers the release of proapoptotic factors. This theory seems plausible in part because damaged mitochondria are known to initiate apoptosis in mature human cells. The death machinery is always there, ready to go. If the cell senses serious damage, it can release the block on the death machinery, and all hell breaks loose.

At the same time, Bassnett has proposed another potential cause of apoptosis: the lactic acid produced during the breakdown of glucose that occurs in differentiated lens cells. Mature cells in the lens's center lack mitochondria and produce energy by turning glucose into lactic acid. The acid forms a concentration gradient, along with a gradient in pH. Either gradient could start apoptosis.

Other triggers have attracted attention as well. In studies of lens cells in culture, Michael Wride, now at Cardiff University in Wales, and Esmond Sanders of the University of Alberta in Canada showed that tumor necrosis factor appears to promote the degradation of lens nuclei. Tumor necrosis factor is a messenger protein, or cytokine, that can act as a potent inducer of apoptosis in healthy cells and certain tumor cells. No one knows how this cytokine would work naturally in the lens, however.

Klaus van Leyen of Massachusetts General Hospital and his colleagues have uncovered clues to the molecules that respond to cell-death triggers. They found, for instance, that the enzyme 15-lipoxygenase can embed itself in the membranes of lens cell organelles and create holes in them. The holes allow proteases (enzymes that destroy proteins) to enter and destroy the organelles. Exactly what elicits the 15-lipoxygenase activity at the right time during lens cell differentiation remains unclear.

Research by this author and others has recently provided possible insights into the braking mechanism. In human, rat and mouse lenses, my colleagues and I found that a protein called galectin-3, which can bind to other molecules, is produced in lens cells that still have their organelles, but its synthesis is reduced when the organelles start to degrade. This activity pattern could control the apoptosis process, but we have no idea what triggers the shutoff of galectin-3. We began to look at galectin-3 because it is known to be involved in various biological functions related to cell proliferation, apoptosis and differentiation in other tissues.

Most recently, Sogo Nishimoto of Osaka University in Japan has identified a DNAse (an enzyme that cleaves DNA) that is essential for the degradation of DNA in lens cells. When this particular DNAse is missing in laboratory mice, they are born with cataracts; also, the apoptotic breakdown of the nuclei during the differentiation of lens cells does not seem to occur, whereas apoptosis appears to occur normally in all other cells. (Children can be born with cataracts if organelles are not degraded during fetal development, possibly as a result of a viral infection, such as rubella, in the mother.)

Of course, it is conceivable that rather than actively halting apoptosis in midstream, lens cells forestall death because some components simply are resistant to the molecules that effect self-destruction. For instance, proteins occurring only in the lens might be "invisible" to the killer enzymes that degrade the cytoskeletons of other cells. Alternatively, some evidence suggests that crystallins might form a protective barrier around certain proteins, preventing the enzymes from reaching those targets.

Swimming Zebras

AS WORK ADVANCES, a small fish could offer promising clues. The zebra fish is a terrific creature in which to study embryonic development. Its embryos have very few cells and are quite translucent early on, so experts can observe the formation of internal organs. Most organs develop incredibly fast—just 48 hours after eggs are laid. During day three, the fish hatch and start swimming around. Yet because zebra fish are vertebrates, the genetic control of their development is remarkably similar to that in humans.

Various groups have undertaken large-scale searches for mutant zebra fish, among them the laboratory of Nobel laureate Christiane Nüsslein-Volhard at the Max Planck Institute. Among the mutants found were ones possessing lenses with intact organelles and others with lens cells that died completely. Some mutants had cataracts much like those in humans.

The labs are now looking to see if these mutants can provide new information about what starts and stops apoptosis. If so, the insights could advance medical research into ways to beat cell-death diseases. In the meantime, these studies should greatly improve our understanding of how and why cataracts form, which could lead to ways to slow their growth or prevent them altogether. That possibility alone keeps us focused. SA

MORE TO EXPLORE

Nuclear Degeneration in the Developing Lens and Its Regulation by TNFalpha. Michael A. Wride and Esmond J. Sanders in *Experimental Eye Research,* Vol. 66, No.3, pages 371–383; 1998.

Lens Organelle Degradation. Steven Bassnett in *Experimental Eye Research,* Vol. 74, No. 1, pages 1–6; 2002.

Developmental Aspects of Galectin-3 Expression in the Lens. R. Dahm, S. Bramke, J. Dawczynski, R. H. Nagaraj and M. Kasper in *Histochemistry and Cell Biology,* Vol. 119, No. 3, pages 219–226; 2003.

Nuclear Cataract Caused by a Lack of DNA Degradation in the Mouse Eye Lens. S. Nishimoto et al. in *Nature,* Vol. 424, pages 1071–1074; 2003.

FIXING THE VOTE

ELECTRONIC VOTING MACHINES PROMISE TO MAKE ELECTIONS MORE ACCURATE THAN EVER BEFORE, BUT ONLY IF CERTAIN PROBLEMS—WITH THE MACHINES AND THE WIDER ELECTORAL PROCESS—ARE RECTIFIED

By Ted Selker

COURTESY OF SEQUOIA VOTING SYSTEMS

Voting may seem like a simple activity—cast ballots, then count them. Complexity arises, however, because voters must be registered and votes must be recorded in secrecy, transferred securely and counted accurately. We vote rarely, so the procedure never becomes a well-practiced routine. One race between two candidates is easy. Half a dozen races, each between several candidates, and ballot measures besides—that's harder. This complex process is so vital to our democracy that problems with it are as noteworthy as engineering faults in a nuclear power plant.

Votes can be lost at every stage of the process. The infamous 2000 U.S. presidential election dramatized some very basic, yet systemic, flaws concerning who got to vote and how the votes were counted. An estimated four million to six million ballots were not counted or were prevented from being cast at all—well over 2 percent of the 150 million registered voters. This is a shockingly large number considering that the decision of which candidate would assume the most powerful office in the world came to rest on 537 ballots in Florida.

VOTING MACHINE—here, Sequoia Voting Systems's AVC Edge—is fairly typical of direct record electronic (DRE) voting machines on the market. Voters enter their votes via a touch-screen interface.

Three simple problems were to blame for these losses. The first, which made up the largest contribution, was from registration database errors that prevented 1.5 million to three million votes; this problem was exemplified by 80,000 names taken off the Florida lists because of a poorly designed computer algorithm. Second, a further 1.5 million to two million votes were uncountable because of equipment glitches, mostly bad ballot design. For example, the butterfly ballot of Palm Beach County confused many into voting for an unintended candidate and also contributed to another appalling outcome: 19,235 people, or 4 percent of voters, selected more than one presidential candidate. Equipment problems such as clogged punch holes resulted in an additional 682 dimpled ballots that were not counted there. Finally, according to the U.S. Census Bureau, about one million registered voters reported that polling-place difficulties such as long lines prevented them from casting a vote.

Thus, registration and polling-place troubles accounted for about two thirds of the documentable lost votes in 2000. The remaining one third were technology-related, most notably ballot design and mechanical failures. In the aftermath of the 2000 election, officials across the country, at both the federal and local levels, have scrambled to abandon old approaches, such as lever machines and punch cards, in favor of newer methods. Many are turning to electronic voting machines. Although these machines offer many advantages, we must make sure that these new systems simplify the election process, reduce errors and eliminate fraud.

Some countries have introduced electronic systems with great success. Brazil started testing electronic voting machines in the mid-1990s and since 2000 has been using one type of machine across its vast pool of 106 million voters. It has multiple organizations responsible for different aspects of voting equipment development as part of the safeguards. It also introduced the machines in carefully controlled stages—with 40,000 voters in 1996 (7 percent of whom failed to record their votes electronically) and 150,000 in 1998 (2 percent failure). Improvements based on those experiments reduced the failure rate to an estimated 0.2 percent in 2000.

Voting Technology

VOTING SYSTEMS have a long history of advancing with technology. In ancient Greece, Egypt and Rome, marks were made for candidates on pieces of discarded pottery called ostraca. Paper superseded pottery in the hand-counted paper ballot, which is still used by 1.3 percent of U.S. voters. Other modern technologies are lever machines, punch cards and mark-sense ballots (where each candidate's name is next to an empty oval or other shape that must be marked correctly to indicate the selection, and a scanner counts the votes automatically). The table on pages 94 and 95 summarizes the benefits and drawbacks of each of these methods and suggests ways to improve them. A lengthier discussion of nonelectronic systems is at www.sciam.com/ontheweb.

Electronic voting machines have been around for 135 years—Thomas Edison patented one in 1869. Elections started testing electronic voting machines in the 1970s, when displaying and recording a ballot directly into a computer file became economical. At first, many were mixed-media machines, using paper to present the selections and buttons to record the votes. Officials had to carefully align the paper with the buttons and indicator lights. Electronic voting machines that use such paper overlays are still on the market. More modern direct record electronic (DRE) voting machines present the ballot and feedback information on an electronic display, which may be combined with audio.

Such machines have many advantages: they can stop a voter from choosing too many candidates (called overvoting), and they can warn if no candidate is picked on a race (undervoting). For instance, when Georgia changed over to DREs in 2002, residuals (the total of overvotes and undervotes combined) were reduced from among the worst in the nation at 3.2 percent on the top race in 2000 to 0.9 percent in 2002. So-called ballotless voting allows the machines to eliminate tampering with physical ballots during handling or counting. (Lever machines, dating back to 1892, share many of those features.)

Yet the birthing of DRE voting equipment in the U.S. has not been easy. The voting machine industry is fragmented, with numerous companies pursuing a variety of products and without a mature body of industry-wide standards in place. Deciding what is a good voting machine is still being discussed by various advocacy organizations and groups such as the IEEE Project 1583 on voting equipment standards. Allegations of voting companies using money to influence testing and purchasing of equipment are not uncommon.

Complicating matters, local jurisdictions across the country have different rules and approaches to testing and using voting equipment. Some counties, such as Los Angeles, are sophisticated enough that they commission voting machines built to their own specifications. Many other municipalities know so little about voting that they employ voting companies to run the election and report the results.

Polling-place practices add further hazards of insecurity and potential malfunctions. I recall walking into the central election warehouse (where the voting machines are stored and the precinct vote tallies are combined) in Broward County, Florida, when it was being used for a recount in December 2002. The building's loading dock was opened to the outdoors for ventilation. The control center for tallying all the votes was a small computer room; the door to that room was ajar and no log was kept of personnel entering and leaving.

Beyond external issues, DRE machines themselves have had technological shortcomings that have slowed their adoption. Voters have found their displays confusing or challenging to use. Software bugs and difficulties in setting up DREs have also presented problems. During the 2002 Broward County recount, I was allowed to try out machines from Electronic Systems and Services (ESS), one of the country's major election machine makers. The ESS machines had an excessive undervote because the "move to next race" button was too close to the "deposit my ballot" button. An audio ballot was so poorly designed it took about 45 minutes to vote.

On machines made by the company Sequoia, people who chose a straight party vote and then tried to select that party's presidential candidate were unaware that they were *de*selecting their presidential choice. A massive 10 percent undervote was registered in one county using Sequoia machines in New Mexico.

Examining the insides of new voting machines still reveals

Overview/*Electronic Voting*

- Following the infamous 2000 presidential election, electoral officials around the country have scrambled to upgrade their voting technology with newer systems, such as direct record electronic voting machines (DREs).
- A state or county that is considering buying DREs should hire experts to test the machines thoroughly for bugs, malicious software and security holes and to assess the quality of the user interface.
- Election officials and polling-place workers should be well versed in the operation of their machines and should follow practices that do not compromise the security of the vote.
- In addition to these technology-related issues, the voter registration process and polling-place practices in general must be improved to prevent massive losses of votes.

AUDIT TRAILS

An audit trail printed on paper or recorded on tape or CD would enable an independent recount of votes made on an electronic voting machine.

VERIVOTE PRINTER UPGRADE to Sequoia Voting Systems's AVC Edge voting machine produces a paper copy of the votes made on it and displays it behind a window. Before leaving the voting booth, the voter can verify her vote by inspecting the paper record, which is retained by the machine for use in recounts

DON FOLEY (*illustration*); COURTESY OF SEQUOIA VOTING SYSTEMS (*photograph*)

many physical security faults. For example, some machines have a lifetime electronic odometer that is supposed to read every vote that the machine makes. But the odometer is connected to the rest of the machine by a cable that a corrupt poll worker could unplug to circumvent it without breaking a seal.

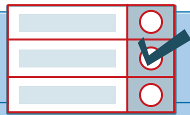
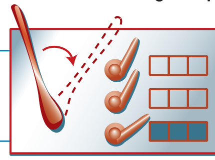
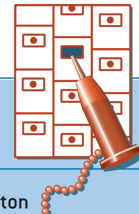
Source code for voting machines made by different companies, like most commercial software, is a trade secret. Election machine companies allow buyers to show the source code to experts under confidential terms. Unfortunately, the local election officials might not know how to find a qualified expert. And when they find one, will the voting companies be required to listen? For instance, in 1997 Iowa was considering a voting machine made by Global Election Systems, which was later bought out by Diebold. Computer scientist Douglas W. Jones of the University of Iowa pointed out security issues, and the state bought Sequoia machines instead. In February 2003 Diebold left its software on unsecured servers, and DRE critics posted Diebold's code on the Internet for everyone to see. The problems that Jones saw six years earlier had not been fixed. Any person with physical access to the machines and a moderate amount of computer knowledge could have hacked into them to produce any outcome desired.

The best computer security available depends on sophisticated encryption and carefully designed protocols. Yet to know the system has not been compromised requires testing. DRE machines have not received the constant testing that they require. Security of today's voting machines is wholly dependent on election workers and the procedures that they follow.

Because virtually all tallies, no matter what voting method is used, are now stored and transmitted in some electronic form, computer fraud is possible with all voting systems. The advent of DRE machines potentially allows such tampering to go

EXISTING VOTING TECHNOLOGIES

Improving or optimizing an existing technology may be a better choice for many counties than hasty adoption of a new system—introduction of a new technology is often accompanied by an increase in errors.

TECHNOLOGY	Hand-counted paper ballots	Lever machines	Punch cards
COMMENTS	■ Used by 1.3 percent of U.S.	■ First used in 1892 in Lockport, N.Y.	■ First used in 1964 in Fulton and De Kalb counties, Georgia
ADVANTAGES	■ Simple ■ Lowest residual* rate	■ Overvotes are impossible ■ Guarantees secrecy of vote	■ Removes human errors of tallying ■ Compact machines
DISADVANTAGES	■ Recounts differ from original count by twice as much as machine-counted votes do ■ Persistent allegations of votes being altered, added, lost, and so on	■ Bulky, massive machines ■ Defective odometers common ■ Misreading of odometers ■ Voting falloff on lower races (for Senate, state office, for example)	■ Hard to punch holes correctly ■ Often punch wrong hole ■ Ballot design troubles ■ Card readers jam frequently ■ Ballot easy to spoil
WAYS TO IMPROVE	■ Count by mechanical scanner ■ Treat paper with light, heat or coating material to make vote indelible	■ Check and service before each election ■ Monitor odometers with video cameras ■ Improve labeling of groups of levers forming a race ■ Adjustable height of machines	■ Optical way to check ballot while in booth might help

ELIZA JEWETT

unchecked from the point at which the voter attempts to cast a ballot. Schemes for altering ballots have always existed, but a computerized attack could have widespread effects were it waged on a large jurisdiction that uses one kind of software on one type of machine. Using a single system allows large jurisdictions to get organized and improve their results but must be accompanied by stringent controls.

The successful reduction of residuals across all of Georgia, mentioned earlier, is a case in point. Thorough tests on the DREs at Kenisaw State University found many problems, which were resolved before the machines were put into use. This rigorous testing and careful introduction of the machines were central to the state's success.

Electronic Fraud

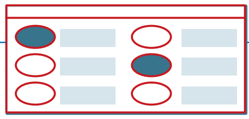
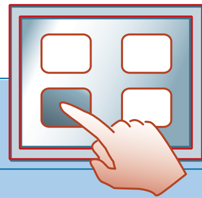
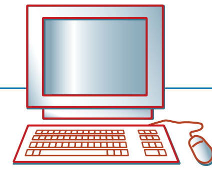
HOW CAN WE FIND all the dangers created by bad software and prevent or correct them before they compromise an election? Reading source code exposes its quality and its use of security approaches and can reveal bugs. But the only completely reliable way to test software is by running it through all the possible situations that it might be faced with.

In 1983 Ken Thompson, on receipt of the Association for Computing Machinery's Turing Award (the most prestigious award in computer science), gave a lecture entitled "Reflections on Trusting Trust." In it he showed the possibility of hazards such as "Easter eggs"—pieces of code that are not visible to a reader of the program. In a voting machine, such code would do nothing until election day, when it would change how votes were recorded. Such code could be loaded into a voting machine in many ways: in the voting software itself, in the tools that assemble the software (compiler, linker and loader), or in the tools the program depends on (database, operating system scheduler, memory management and graphical-user-interface controller).

Tests must therefore be conducted to catch Easter eggs and bugs that occur only on election day. Many electronic voting machines have clocks in them that can be set forward to the day of the election to perform a test. But these clocks could be manipulated by officials to rerun an election and create bogus voting records, so a safer voting machine would not allow its clock to be set in the field. Such machines would need to be tested for Easter egg fraud on election day. In November 2003 in California a random selection of each electronic voting system was taken aside on the day of election, and careful parallel elections were conducted to show that the machines were completely accurate at recording votes. These tests demonstrated that the voting machines were working correctly.

To prepare for a fraud-free voting day requires that every effort be made to create voting machines that do not harbor malicious code. The computer science research community is constantly debating the question of how to make provably secure software. Computer security experts have devised many approaches to keep computers reliable enough for other purposes, such as financial transactions. Financial software transfers billions of dollars every day, is extensively tested and holds up well under concerted attacks. The same security techniques can be applied to voting machines. Some researchers believe that the security precautions of "open source" (making the programs available for anyone to examine) and encryption techniques can help but not completely guard against Easter eggs.

Guarding votes against being compromised has always re-

Mark-sense ballots 	Electronic machines 	Internet voting, phone messaging, interactive TV 
■ First used in 1962 in California	■ First used in 1976	■ Internet voting first used in 2000 primary in Phoenix, Ariz.
■ With in-precinct scanning, has lowest residuals of any mechanical method ■ Easier than punching holes ■ Voter can read candidates right on ballot	■ Overvotes are impossible ■ No human errors of tallying ■ Easy for people with physical disabilities to use ■ Good feedback	■ Vote from home ■ People with physical disabilities can use their own special-needs setup ■ No human errors in tallying
■ Ballot readers are slower, harder to calibrate and more prone to jamming than card readers ■ Bulky ballot ■ Ballot easy to spoil	■ User interface often poor ■ Concerns about malicious software ■ Concerns about computer obsolescence	■ Concerns about malicious software, network problems and hackers
■ Use an in-precinct scanner to catch problems and give the voter a second chance to vote ■ Use DRE to mark ballot ■ "Fill in the shape" version better than "connect the arrow" version	■ Test ballots ■ Consider closed systems ■ Test system, including on day of election	■ Use special Web browser ■ System on a CD ■ New approaches to security needed, such as multiple software agents

*Residuals are ballots with votes for too many (overvote) or too few (undervote) candidates.

quired multiple human agents watching each other for mistakes or malice. The best future schemes might include computer agents that check one another and create internal audits to validate every step of the voting process. The Secure Architecture for Voting Electronically (SAVE) at the Massachusetts Institute of Technology is a demonstration research project to explore such an approach. SAVE works by having several programs carry out the same tasks, but while using such different methods that each program would have to be breached separately to compromise the final result. The system knows to call foul when too many modules disagree.

Audit Trails

SOME CRITICS INSIST that the best way to ameliorate such attacks is by providing a separate human-readable paper ballot. This widely promoted scheme is the voter-verified paper ballot (VVPB) suggested by Rebecca Mercuri, then at Bryn Mawr College. The voting machine prints out a receipt, and the voter can look at it after voting and assure himself that at least the paper records his intention. The receipt remains behind a clear screen so no one can tamper with it during its inspection, and it is retained by the machine. If a dispute about the electronic count arises, a recount can be conducted using the printed receipts. (It is not a good idea for the voter to have a copy, because such receipts could encourage the selling of votes.)

Although the VVPB looks quite appealing at first glance, a deeper inspection exposes some serious flaws. First, it is complicated for the voter. Elections in this country often have many races. Validating all the selections on a separate paper after the ballot has been filled out is not a simple task. Experience shows that even when confronted with a printout that tells voters in which race they have made a mistake, few are willing to go back and correct it. Anything that takes a voter's attention away from the immediate act of casting a ballot will reduce the chances of the person voting successfully. Every extra button, every extra step, every extra decision is a source of lost votes.

The scheme is also complicated for the officials. If a voter claims fraud, what is the official to do? The voter claims she voted for Jane, but both the DRE screen and the receipt show a vote for John. Should they close the polling station? On top of this, the officials are not legally allowed to see an individual voter's ballot.

VVPB addresses only a small part of the fraud problem. The paper trails themselves could be made part of a scheme for defrauding an election if a hacker tampers with the printing software. The paper can be manipulated in all the usual ways after the election.

A better option would allow people to verify their selections

THE AUTHOR

TED SELKER is the Massachusetts Institute of Technology director of the California Institute of Technology/M.I.T. voting project, which evaluates the impact of technology on the election process. A large part of his research in voting concerns inventing and testing new technology. Examples include new approaches to user interfaces and ballot design and secure electronic architectures. Selker's Context Aware Computing group at the M.I.T. Media Laboratory strives to create a world in which people's desires and intentions guide computers to help them. This work is developing environments that use sensors and artificial intelligence to form keyboardless computer scenarios.

In the Courts and in the News

In recent months, electronic voting machines have been in the news a lot, as groups file legal actions both for and against use of the machines and new problems with elections are uncovered. *—Graham P. Collins, staff editor*

March—In a case brought by the American Association of Disabled Persons, a federal judge in Florida orders Duval County to have at least one machine that allows the visually impaired to vote without assistance at 20 percent of its polling places. Duval County appeals, and in April the judge stays his own ruling.

April—In Maryland, local politicians and activists from the Campaign for Verifiable Voting file suit against the Maryland Board of Elections to block the use of the state's 16,000 direct record electronic (DRE) voting machines, which do not have printers to produce paper receipts as required by state law. The move follows reports of glitches in the March 2 primary election; some voters who demanded paper ballots were given them but later learned their votes were invalidated.

April—Citing security and reliability concerns and following problems in the March 2 primary election, California's secretary of state bans the use, in the November 2004 election, of more than 14,000 DREs made by Diebold, Inc. He also conditionally decertifies 28,000 other DREs, pending steps to upgrade their security. (Some counties have their systems recertified in June.) Three counties file suit to block his order. A group of disabled voters also sues to undo the order. In addition, the California secretary of state recommends that the state's attorney general look into possible civil and criminal charges against Diebold because of what he calls "fraudulent actions by Diebold." A report accuses the company of breaking state election law by installing uncertified software on DREs in four counties and then lying about those machines.

May—In Florida, Representative Robert Wexler sues to block the use of Election Systems and Services voting technology in Broward and Miami-Dade counties.

June—The League of Women Voters, which in 2003 endorsed paperless electronic voting, drops that support. Instead it adopts a resolution to favor "secure, accurate, recountable and accessible" systems such as those with printed receipts.

June—The head of the Election Assistance Commission calls for tougher security measures for electronic voting by the November election.

July—Advocacy groups in Florida ask a Tallahassee judge to step in before the August 31 primary election and override Governor Jeb Bush's decision not to allow manual recounts in the 15 counties that have touch-screen voting machines. Also in Florida, audit records of the 2002 governor's primary and general election are reported permanently lost because of computer failures. After a few days the records are rediscovered on a disk in an adjoining room.

September—Nevada, in a primary election, will be the first to use DREs that print paper receipts statewide.

with recorded audio feedback. An audio transcript on tape or a CD has an integrity that is harder to compromise than a collection of paper receipts. Most current electronic voting machines can be set up to speak the choices to the voter while he looks at the visual interface. The tape can be read by a computer or listened to by people. Because misreads of paper are a major difficulty with all counting machines today, the tape can be better verified than paper receipts. An audio receipt is also preferable to a paper receipt because it is hard to change or erase the audio verifications without such alterations being noticed (think about the 18-minute gap on the Watergate tapes). Also, a small number of cassette tapes or CDs are easier to store and transport than thousands of paper receipts.

Other proposals for voter verification include recording the video image of the DRE and showing the ballot as it has been received by the central counting databases while the voter is in the booth. The advantage of these techniques is that they are passive—they do not require additional actions on the part of the voter.

Here is how voting might go using a well-designed audio record. Imagine you are voting on a computer. You like Abby Roosevelt, Independent. You press the touch-screen button for your choice. The name is highlighted, and the vote button on one side is replaced with an unvote button on the other side. The tab on the screen for this race shows that a selection has been made. The earphones you are wearing tell you that you have voted for "Ben Jefferson" (and these words are recorded on a backup tape).

Wait a minute! "Ben Jefferson"? You realize that you must have pressed the wrong button by mistake. You study the screen and see a prominent "cancel vote" button. You press it. "Vote for Ben Jefferson for president canceled," the computer intones onto a tape and into your ears. The screen returns to its prevote state, and this time you press more carefully and are rewarded with "Vote cast for Abby Roosevelt, Independent, for president." You go on to the Senate race.

The features just described are designed to give feedback in ways you are most adept at understanding. People are good at noticing labels moving, tabs changing, and contrast and texture changes. We have trouble doing things accurately without such feedback. The audio verification comes right at a time when the user is performing the action. Perceptual tasks (seeing movement and hearing the audio) are easier to perform than cognitive ones (reading a paper receipt and remembering all the candidates one intended to vote for). A tape or CD recording is a permanent, independent transcript of your vote.

DIEBOLD ELECTION SYSTEMS'S AccuVote TSX, another typical modern electronic voting machine, was decertified in California.

COURTESY OF DIEBOLD ELECTION SYSTEMS

These features are all implementable now as ballot improvements on current voting machines. Extra work would be needed to allow sight- or hearing-impaired people to verify multiple records of their ballot as well.

Some researchers are studying alternatives to DREs, in the form of Internet voting or voting using familiar devices such as the phone. Since May 2002, England has been experimenting with a number of systems intended to increase turnout. These methods include mailing in optically readable paper ballots (absentee voting), using a standard phone call and the phone's keypad, using the instant-messaging facilities on cell phones and using interactive TV that is available in English homes. Swindon Borough, for example, included more than 100,000 voters in an experiment using the Internet and telephones. A 10-digit PIN was hand-delivered to voters' homes. This PIN was used in conjunction with a password the voters had been sent separately to authorize them to vote. No fraud was detected or reported. But the effort only improved turnout by 3 percentage points (from 28 to 31 percent).

In contrast, introducing the option of absentee voting increased voter turnout by 15 percentage points—but with a downside: large-scale vote buying was reported in Manchester and Bradford. (Being able to prove whom you have voted for, such as by showing the ballot you are mailing in, enables vote buying.)

What Must Be Done

THE UNIVERSAL ADOPTION of perfect voting machines will not be happening anytime soon. But quite independent of the specific machines used, much can and should be done simply to ensure that votes are collected and accurately counted in the U.S. We must be adamant about the following improvements:

1. We must simplify the registration system. The largest loss of votes in 2000 occurred because errors in registration databases prevented people from voting. Registration databases must be properly checked to make sure they include all eligible people who want to be registered. We must develop national standards and technology to ensure that people can register reliably but that they do not register and vote in multiple places.

2. Local election officials must understand the operation of their equipment and test its performance thoroughly when it is delivered and before each election. DREs should be tested on election day, using dummy precincts.

3. Local election officials must teach their workers using simple procedures to run the equipment and other processes. Ballot making, marking, collecting and counting all must be carefully set up to avoid error and fraud. Many voting officials inadvertently use procedures that compromise accuracy, security and integrity of ballots by, for example, turning off precinct scanning machines that check for overvotes and inspecting and "correcting" ballots.

4. Each step in the voting process must be resistant to tampering. Collecting, counting and storing of ballots must be done with documentation of who touches everything and with clear procedures for what to do with the materials at each stage. Multiple people must oversee all critical processes.

5. Each task in the voting process must be clear and accessible, have helpful feedback and allow a person to validate it. Perceptual, cognitive, motor and social capabilities of people must be taken into account when designing machines and ballots. Ballot designs should pass usability and countability tests before being shown for final approval to the parties invested in the election. Voters must be able to understand how to make their selections, and votes must be easy to count in mass quantities.

6. The government should invest in research to develop and test secure voting technology, including DREs and Internet voting. Rushing to adopt present-day voting machines is not the best use of funds in the long term.

7. Standards of ethics must be set and enforced for all poll workers and also for voting companies regarding investments in them and donations by them or their executives.

Only when these requirements are met will we have a truly secure and accurate voting system, no matter what underlying technology is used. SA

MORE TO EXPLORE

Misvotes, Undervotes and Overvotes: The 2000 Presidential Election in Florida. Alan Agresti and Bret Presnell in *Statistical Science,* Vol. 17, No. 4, pages 436–440; 2002. Available at **web.stat.ufl.edu/~presnell/Tech-Reps/election2000.pdf**

A Better Ballot Box? Rebecca Mercuri in *IEEE Spectrum,* Vol. 39, No. 10, pages 46–50; October 2002. Available at **www.spectrum.ieee.org/WEBONLY/publicfeature/oct02/evot.html**

Security Vulnerabilities and Problems with VVPT. Ted Selker and Jon Goler. April 2004. Available at **www.vote.caltech.edu/Reports/vtp_wp13.pdf**

The Caltech/M.I.T. Voting Technology Project is at **www.vote.caltech.edu;** the project's July 2004 report with recommendations for the 2004 presidential election is at **www.vote.caltech.edu/Reports/EAC.pdf**

The U.S. Election Assistance Commission Web site is at **www.eac.gov**

Hitting the Genetic OFF Switch

A host of start-ups is speeding development of a new class of drugs that block the action of RNA **By Gary Stix**

In 1996 *Worth* magazine proclaimed that Isis Pharmaceuticals could become the next Microsoft, a prediction that turned out to be a particularly egregious example of hyperbole run amok. To be sure, Isis remains a leader in the gene-blocking technology called antisense. But the road to successful treatments for cancer and other diseases has been littered with disappointments.

During the past few years, a new gene-silencing technology has emerged that may be poised to fulfill the promise that was once trumpeted for antisense. "I've been writing in grants for 25 years that during the next five years I'm going to test this process or that process to see if I can do gene inactivation studies in mammalian cells in culture. And I did them, and they were so awkward and so complicated that you just couldn't apply them generally," says Phillip A. Sharp, director of the McGovern Institute for Brain Research at the Massachusetts Institute of Technology. "Lo and behold, all of the time right there in front of me was a process that I could have used."

Sharp, a co-winner of the 1993 Nobel Prize in Physiology or Medicine, was referring to a series of relatively recent discoveries that cells have a mechanism, dubbed RNA interference (RNAi), which blocks gene expression. It prevents RNA transcripts of genes from giving rise to the proteins those genes encode. This natural method of gene silencing comes into play, for example, when viruses try to commandeer a cell's protein-making machinery to produce viral proteins.

A milestone arrived in 1998, when Andrew Z. Fire, now at the Stanford University School of Medicine, and Craig C. Mello of the University of Massachusetts Medical School identified in worms double-stranded RNAs that acted as the switch to turn off genes in RNAi. And in 2001 Thomas Tuschl, now at the Rockefeller University, found that an abbreviated version of double-stranded RNAs—short interfering RNAs (siRNAs)—could shut off genes in mammalian cells. The number of research papers on RNAi has mushroomed from a dozen-plus in 1998 to multiple hundreds last year. Even if the promise for therapeutics never materializes, it is quite likely that some of the seminal discoveries will garner Nobel Prizes. "This has touched everything we do in biological science, from plants to man," Sharp notes. [See "Censors of the Genome," by Nelson C. Lau and David P. Bartel; Scientific American, August 2003.]

The excitement about siRNAs as drugs relates to how they differ in critical ways from antisense therapeutics. At first glance, siRNAs seem very similar to antisense. An antisense drug consists of an artificially synthesized chain of nucleotides, or genetic building blocks, that binds to a messenger RNA containing a complementary sequence. This binding blocks gene expression. An siRNA also silences genes—and it even uses a complementary RNA, or antisense, strand to do so. Once inside a cell, an siRNA attaches to an aggregate of proteins called an RNA-induced silencing complex (RISC), which retains only the antisense strand. The siRNA-bearing RISC then binds to the targeted messenger RNA and degrades it or prevents it from functioning [*see box on page 100*].

Unlike the antisense drugs that have been under development for the past 15 years, siRNAs do not disrupt only a single messenger RNA. They act as catalysts, doing the same job over and over, one explanation for their apparent potency. "They are 100- to 1,000-fold more effective than antisense," says Judy Lieberman, a senior investigator at the CBR Institute for Biomedical Research in Boston and one of the first researchers to show the therapeutic potential of the technique in animals.

Already almost 100 companies are

involved in RNAi; nearly half supply the chemicals and technology needed to perform experiments, and the others are biotechnology or pharmaceutical companies doing commercial research with RNAi, according to Kewal K. Jain, chief executive officer of Jain PharmaBiotech, a Basel, Switzerland, market research firm. "All of this has happened within the last two or three years," Jain says.

A small fraction of these companies have dedicated themselves to producing therapies using siRNAs. As soon as Tuschl's paper documenting siRNAs in mammalian cells was published, the venture-capital community sprang into action. "It was worth it to make a bet realizing in vivo efficacy was not guaranteed," says Christoph H. Westphal, one of the founders of Alnylam in Cambridge, Mass., and a general partner with Polaris Venture Partners. Many of the early innovators in RNAi technology, including Tuschl, Sharp and David P. Bartel of M.I.T., among others, got together to form Alnylam Pharmaceuticals in 2002. Sharp, a founder of biotech giant Biogen, brought together this banner group after conversations with more established companies failed to generate sufficient interest.

Paul R. Schimmel, a professor of molecular biology and chemistry at the Scripps Research Institute, and the founder of several biotechnology companies before this one, insisted on the name Alnylam, an Arabic word meaning "string of pearls" that is also the designation for the middle star of Orion's belt. Schimmel made the case, over the protests of others, that the name—pronounced "al-NIGH-lam"—was difficult to pronounce but impossible to forget. Barry Greene, the company's chief operating officer, furnishes a simpler explanation: "The URL was open," he joked at a recent investors' conference.

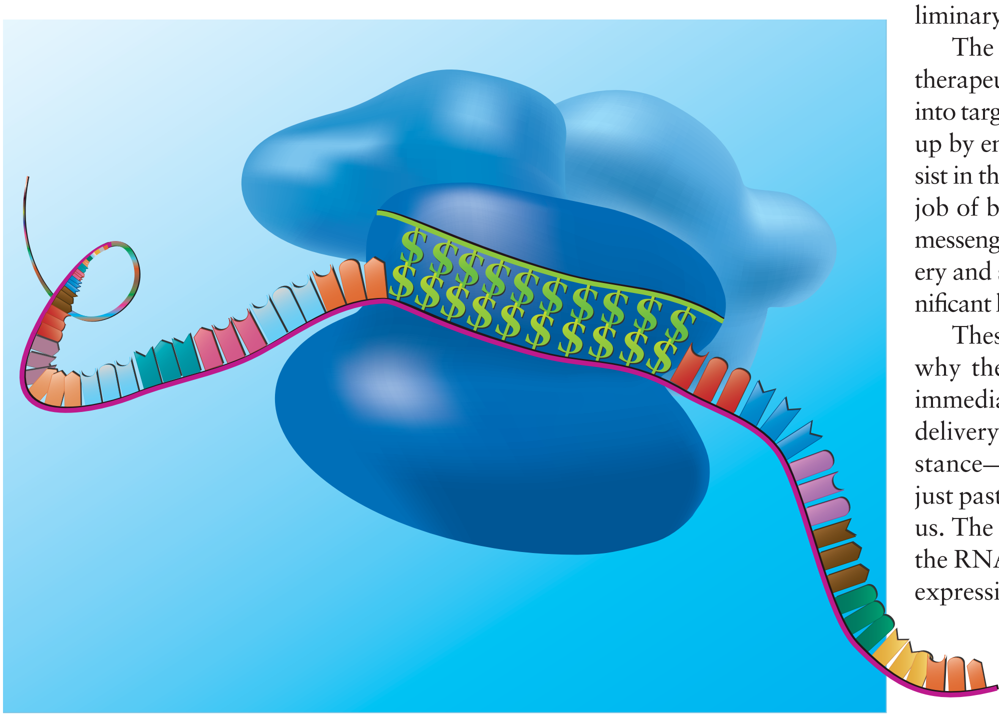
The founders constituted an all-star scientific advisory team, and some also filled slots on the board of directors. But when John M. Maraganore, the company's first permanent chief executive, a transplant from Millennium Pharmaceuticals, hired the people who actually would be entrusted with the task of making siRNA drugs, he did not at first seek out postdoctoral students of these research heavyweights. "Five people were focused on intellectual property and one on science," Maraganore says. A bulletproof patent estate, as much as scientific smarts, would determine who would thrive or falter as drug development and clinical trials got under way. "We were very focused and running very hard," he remembers. "We recognized that if we weren't first, others would grab it from us."

The fledgling Alnylam even bought the German firm Ribopharma to get a hold of a key patent. The stir created by RNAi—tagged by *Science* magazine as "breakthrough of the year" in 2002—helped to bring in venture money. The total take thus far has reached about $85 million, including a somewhat disappointing initial public offering in the spring, and provides enough to keep Alnylam going for another two years, until the first drug makes it through the preliminary safety phase of clinical trials.

The success or failure of RNAi as a therapeutic will hinge on getting the drug into target cells without its being chopped up by enzymes. The drug must then persist in the cell long enough to carry out its job of binding to and inhibiting specific messenger RNAs. The challenge of delivery and stabilization has also posed a significant hurdle for the success of antisense.

These difficulties serve as one reason why the newly formed Alnylam team immediately discarded one approach to delivery: using a vector—a virus, for instance—to ferry a stretch of DNA, not just past the cell wall, but into the nucleus. The gene would then go on to make the RNA that would interfere with gene expression. "In my mind, nothing about RNAi solves the problem of gene therapy," Maraganore notes, referring to the downsides of using viruses to deliver the drug to the right location and the unwanted side effects that they sometimes provoke. Consequently, short interfering RNAs are synthesized in the laboratory from a soup of nucleotides until they form double-stranded molecules that have 21 nucleotide pairs. Some other companies, such as Benitec in Australia, are still pursuing a gene therapy approach [*see table on page 101*].

Key expertise and intellectual property to accomplish this task came from an unlikely source. Alnylam had hired away



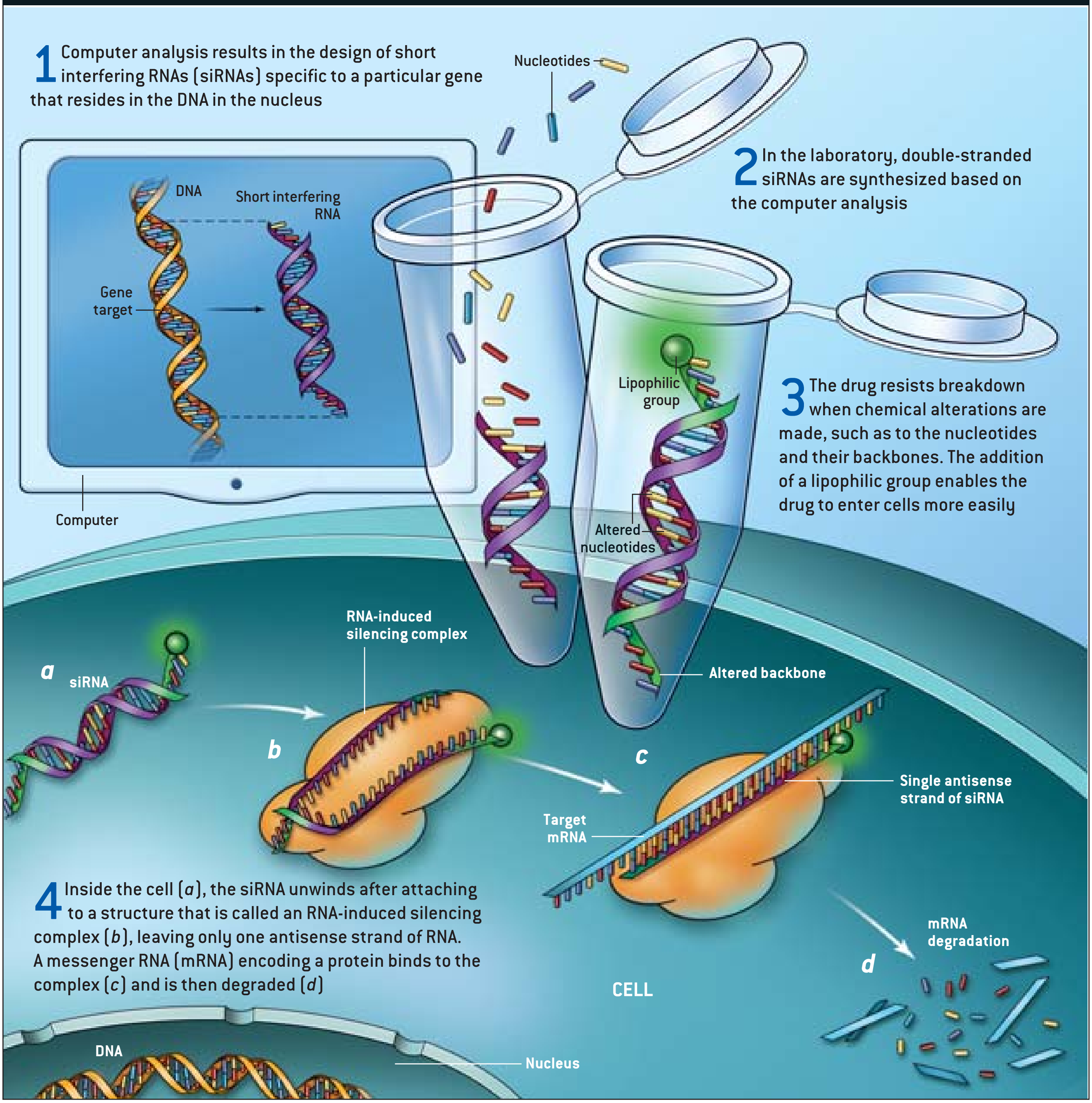
BINDING A MESSENGER RNA (*long strand*) to complementary RNA and an aggregate of proteins (*blob*) could potentially become a lucrative new approach to drug development.

ELIZA JEWETT

from Isis an executive, Muthiah Manoharan, to become vice president of drug discovery. Maraganore called Isis president Stanley T. Crooke last summer and reassured him that Alnylam still wished to be on good terms with the antisense manufacturer. A few months later a dialogue between the two companies resulted in an agreement in which Alnylam would pay $5 million to license Isis's extensive patent portfolio of chemical techniques for delivering and stabilizing RNA. "We will be able to take advantage of the 10-plus years of development of chemistry used in antisense," Maraganore says. In turn, Isis invested $10 million in Alnylam, giving it a 5 percent equity stake in the company and a stream of royalties and fees once siRNA products hit the marketplace. It will also get the rights to make some siRNA drugs.

The development trajectory for siRNA recapitulates the path that antisense has taken. The only antisense drug approved to date is Isis's Vitravene, intended to treat an eye disease once prevalent in AIDS patients. The drug is injected directly into the eye, concentrating the compound at its target while impeding it from producing adverse side effects in

MAKING A GENETIC SILENCER



TAMI TOLPA

BIOTECHS DEVELOPING DRUGS BASED ON RNAi

COMPANY	THERAPIES
Acuity Pharmaceuticals Philadelphia	Age-related macular degeneration and diabetic retinopathy
Alnylam Pharmaceuticals Cambridge, Mass.	Age-related macular degeneration, Parkinson's disease and, over the longer term, cancer, metabolic and autoimmune diseases
atugen Berlin	Cancers and metabolic diseases, for systemic applications, and ocular and skin diseases, for topical delivery
Benitec Queensland, Australia	Hepatitis C and, over the longer term, cancer, autoimmune and viral diseases using a gene therapy approach
CytRx Los Angeles	Amyotrophic lateral sclerosis, cytomegalovirus, obesity and type 2 diabetes
Intradigm Rockville, Md.	Cancer
Nucleonics Horsham, Pa.	Hepatitis B and C and, over the long run, cancer, inflammatory and viral diseases using a gene therapy approach
Sirna Therapeutics Boulder, Colo.	Macular degeneration, hepatitis C and, later, cancer, metabolic, inflammatory, dermatological and central nervous system diseases

other parts of the body. But the market for the drug virtually disappeared as other treatments for AIDS became available and prevented most cases of the cytomegalovirus retinitis infection.

Short RNAs may eventually be delivered to the bloodstream to treat systemic diseases. But, reprising the antisense experience, the first petition to begin a clinical trial was filed in August by Acuity Pharmaceuticals, a Philadelphia company that will attempt to treat age-related macular degeneration by intraocular injection. Other companies, including Alnylam, will follow suit with their own drug trials for macular degeneration. One of these filings will bring Alnylam head to head with the other drug developer that stands a chance of becoming a leader in this emerging market niche.

A Boulder, Colo., company called Sirna Therapeutics was expected to submit an application to the FDA for a macular degeneration drug in early September, perhaps half a year or more before Alnylam. Unlike Alnylam, Sirna is no start-up. It is a reincarnation of another firm, Ribozyme Pharmaceuticals, which for a decade staked its fate on a different type of RNA-related drug. Ribozymes are RNAs acting as catalytic enzymes that, in principle, can cut up messenger RNA and prevent a protein from being produced. But, as with antisense, the potency of ribozymes came into question. A drug to combat hepatitis C caused a monkey to go blind, probably because of the massive doses injected. And another drug did not seem to slow growth of tumors in patients with advanced breast cancer. At the time, Howard W. Robin, a new chief executive, who had managed development of drugs like Betaseron for multiple sclerosis at Berlex Laboratories, was faced with a decision about whether to close shop. The company had only $2 million in cash left and risked being delisted from the Nasdaq.

The darkest days coincided with Tuschl's publication about RNAi in mammalian cells. Instead of turning out the lights, Ribozyme Pharmaceuticals became Sirna Therapeutics. Sirna rejiggered the chemical techniques it had used to deliver and stabilize ribozymes and adapted them to siRNAs. Robin claims that single doses of its siRNAs have remained active inside cells of live animals for up to 22 days. The revamping succeeded in attracting $72 million in new investment during an 18-month period. Besides macular degeneration, the company has programs in hepatitis, oncology and Huntington's disease, among others. "It's not often that you take the skills and intellectual property from a technology that's not working very well and transfer it to the hottest area of biology," Robin says.

Sirna has filed for 90 patents that Robin believes cover the most attractive drug prospects. Patent fights may loom as the technology gets nearer the marketplace. "If you look at our competitors, we believe almost everything they're doing violates our patents," Robin proclaims. But Maraganore begs to differ: "We have a bit of a toll road for anyone doing therapeutics." Any dispute is not likely to emerge until siRNA drugs are much closer to approval. In the meantime, investigators will be closely watching whether siRNAs produce unwanted immune responses or shut down genes they are supposed to leave intact.

For the time being, optimism about RNAi reigns. "If you do with RNAi in man what you do in cell culture, you have the most unbelievably powerful technology for making pharmaceuticals," Maraganore says. "It's the dream of medicine to do selective and efficient gene silencing," Robin asserts. But Isis's Crooke, tempered by the failure of trials for a few of his company's antisense drugs, has a slightly different perspective: "Any time you think something is magic, you're going to get in trouble. [RNAi] is a complicated system with lots of interesting nuances that mechanistically should lead to some unexpected effects as well as those you desire."

RNAi has become a preeminent research tool in a remarkably short time. But its potential as a genetically based pharmaceutical will not become clear for several years, when the first clinical trials prove whether a simple injection is capable of shutting down the effects of a disease-causing gene. SA

MORE TO EXPLORE

The RNAi Revolution. Carl D. Novina and Phillip A. Sharp in *Nature,* Vol. 430, pages 161–164; July 8, 2004.

The Silent Revolution: RNA Interference as Basic Biology, Research Tool and Therapeutic. Derek M. Dykxhoorn and Judy Lieberman in *Annual Review of Medicine,* Vol. 56 (in press).

EARTHQUAKE PROTECTION

Shock Absorbed

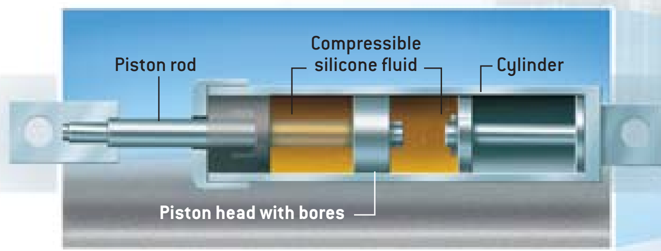
Earthquakes kill thousands of people and cause billions of dollars in damage every year. Reinforced concrete and special trusses have toughened large buildings, but mechanisms that actually reduce the shake from a quake are still relatively new.

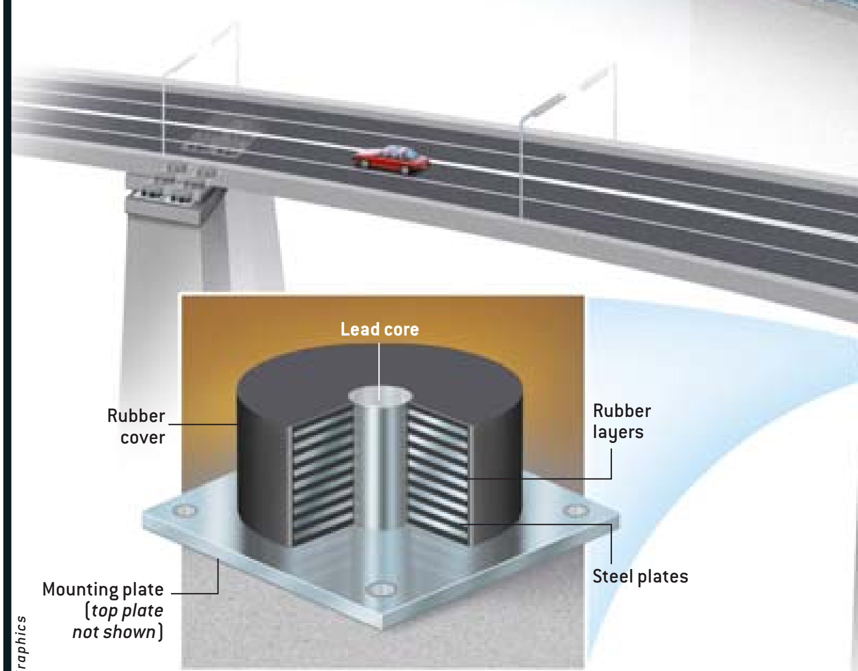
Building codes require structures to provide "life safety"—that is, they should not collapse, so people can evacuate. But the real challenge is economic. "We could design a building with conventional techniques that would survive the largest earthquake without damage, but it would be so expensive no one would build it," says André Filiatrault, deputy director of the Multidisciplinary Center for Earthquake Engineering Research at the State University of New York at Buffalo. More and more, building owners are considering the incremental cost to minimize damage to the structure, to its mechanical systems, to the contents inside and, ultimately, to allow for immediate reoccupancy. "In a hospital, for example," Filiatrault says, "the structure accounts for only 10 percent of the total cost. The other 90 percent is equipment, and designing for life safety won't spare it."

The leading techniques "try to absorb a lot of the ground motion energy so the building doesn't have to," says James Malley, senior principal at Degenkolb Engineers in San Francisco. That often means inserting heavy-duty fixtures between the building and its foundation, such as base isolators that act as rubber mats, viscous fluid dampers that operate like shock absorbers, or slide bearings that allow the building to sway instead of snap. Much of the hardware is adapted from military gear for hardening missile silos, ship decks and submarines against bombs and missiles, and it is being rolled out by contractors looking for civilian work, notes Douglas P. Taylor, CEO of Taylor Devices in North Tonawanda, N.Y.

A few researchers are examining novel technologies, such as actuators that would pull on tendonlike beams to counteract ground motion or electrorheological fluids in a foundation that would turn from liquid to gel to filter out shock waves. In the meantime, installations at new sites and retrofits to old ones are booming. Even in California, most buildings are not yet outfitted. —*Mark Fischetti*

VISCOUS FLUID DAMPERS act like a car's shock absorbers, neutralizing ground movement and minimizing displacement between floors so they do not tear apart. A piston head with bores pushes through a silicone oil, dissipating the quake's mechanical energy as heat.





KENT SNODGRASS *Precision Graphics*

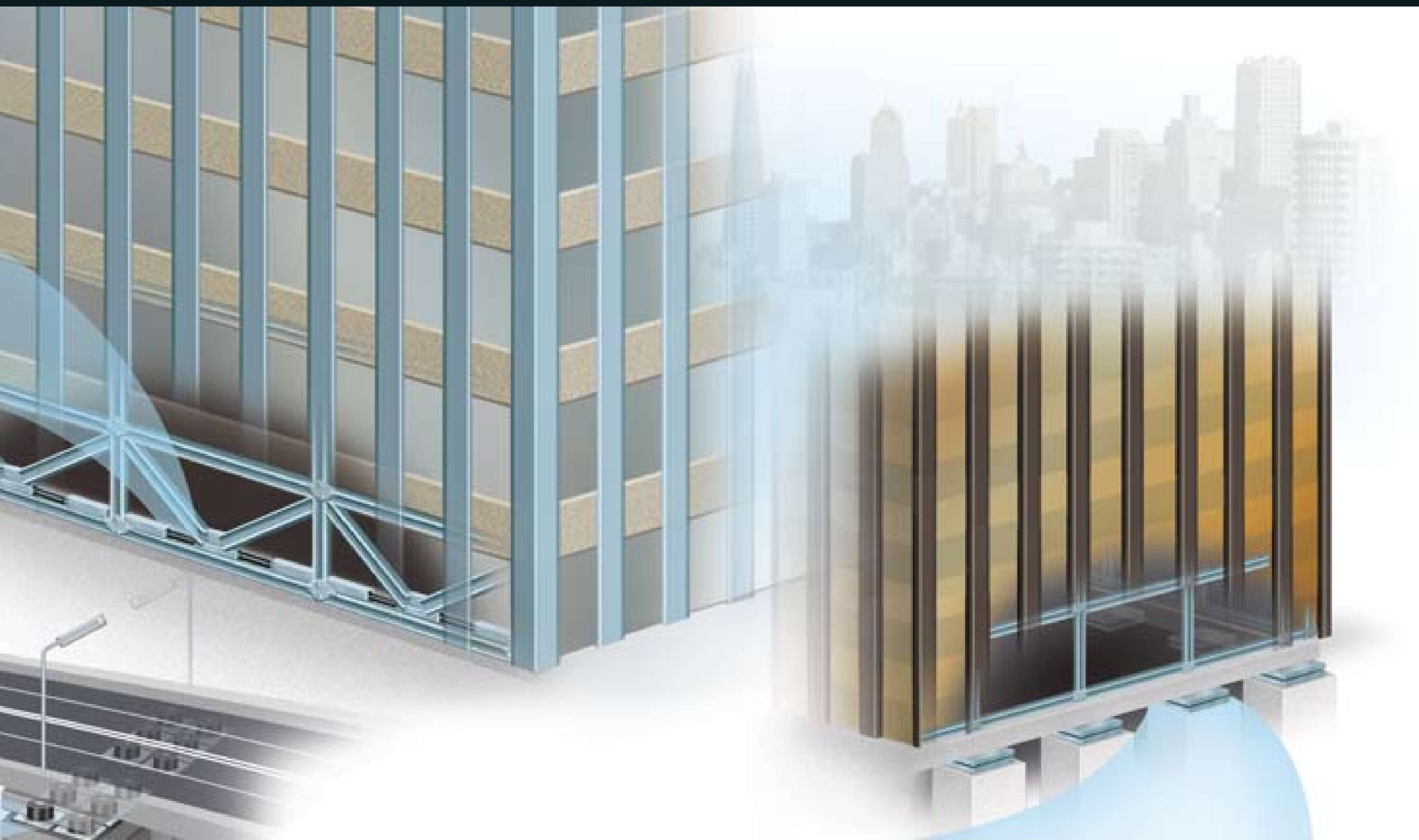
BASE ISOLATORS underneath a bridge pillar, roadway or building column lessen earthquake shear and therefore damage. Rubber layers displace sideways to absorb lateral motion and spring back to return the structure to its original position. Steel layers bonded to the rubber create stiffness that prevents vertical motion. The lead core stops the structure from shifting in the wind.

DID YOU KNOW . . .

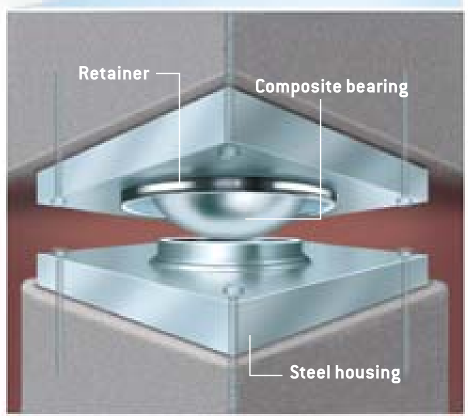
➤ SHARE A SHAKE: In 1999 the National Science Foundation launched the Network for Earthquake Engineering Simulation to research how best to protect structures against earthquakes. The NSF will spend more than $100 million over 15 years at almost 20 university research centers. The centers, which are building gigantic shake tables and wave tanks, will run experiments at one another's facilities over high-speed networks that should be fully operational this fall.

➤ IT TAKES A DISASTER: Work on structural isolation techniques has accelerated sharply because of two disasters. The January 1994 Northridge earthquake near Los Angeles took 60 lives and caused $20 billion in damage. The January 1995 quake in Kobe, Japan, killed more than 5,000 people and crippled more than 50,000 buildings. History repeats: the 1971 San Fernando Valley quake in California prompted codes requiring steel-reinforced concrete in buildings.

➤ ILL WIND: Even light skyscrapers can withstand strong winds. Still, they may sway at a frequency that can make occupants feel seasick. Builders can cancel the movement by installing huge roof tanks with water that swishes to counter the motion (and is available to quench fires). Alternatively, slide bearings [*see illustration*] can support the roof, allowing it to rock back and forth. This technique can also isolate the heavy mass during a quake, lessening stress on vertical columns.



SLIDE BEARING allows a building to slowly glide back and forth like a pendulum as the ground shimmies. Made of a high-strength, low-friction, self-lubricating composite, the bearing's semispherical shape distributes the building's weight and the quake's shear forces. The bearing's shape determines how far and fast the building can glide.



Send topic ideas to workingknowledge@sciam.com

Gadget Envy

ALL-IN-ONE CELL PHONES CAN DO JUST ABOUT ANYTHING

BY MARK ALPERT

I do not own a cell phone. For most of the 1990s I was convinced that wireless phones were nothing more than a passing fad. In more recent years I've held a similar disdain for PDAs, digital cameras and MP3 players. My philosophy was: Who needs all this junk? In my crotchety, old-fashioned view, technological innovation reached its pinnacle in the 1970s with the introduction of the Technics SL-1300 turntable and the Canon A-1 camera. Now those were great gadgets.

But my attitude changed last year when my wife acquired both a cell phone and a BlackBerry for sending and receiving e-mail wirelessly. (She works in the political arena, where BlackBerries seem to be particularly popular.) She became expert at punching out messages on the BlackBerry's keyboard with her thumbs. At odd moments during the day, when the kids were relatively quiet, I would find her clutching the device in a sort of trance, her eyes focused on its drab green screen. After a while I started to get jealous. So, in a classic case of marital one-upmanship, I decided to obtain an even more impressive communications system.

THROW AWAY your old gadgets: a single all-in-one cell phone can take the place of your laptop, personal organizer, digital camera and portable music player.

JOHN FRASER

For handheld devices, the current trend is convergence. Manufacturers are offering cell phones that can send messages, browse the Web, take pictures and play music. A dizzying array of all-in-one gadgets is already on the market, and I had some trouble selecting a product to evaluate. In the end I picked the Treo 600 Smartphone made by PalmOne, the Milpitas, Calif., company that introduced the Palm Pilot PDA eight years ago. I made this choice largely on the recommendation of the *New York Times,* which called the Treo 600 "the favored wireless phone among Silicon Valley's digerati." (In case you were wondering, "digerati" is the computer-world equivalent of "literati." An awful word, isn't it?)

PalmOne sent me a Treo 600 to review, with wireless service provided by T-Mobile. Weighing in at six ounces, the Treo is a bit larger than an ordinary cell phone, but it still fits comfortably in one hand. It has a color screen measuring almost two inches on each side and a QWERTY keyboard with tiny buttons about an eighth of an inch wide. More and more cell phone makers are putting keyboards on their devices to make it easier to send e-mail. (Composing messages on a conventional numeric keypad requires an absurd amount of tapping.) But the Treo's keyboard is too small to be really useful; adult thumbs are so much larger than the keys, you frequently press the wrong buttons. Luckily, you can use the device's stylus to dial

TREO 600 SMARTPHONE can take snapshots and wirelessly send them to your friends.

phone numbers on the screen and to tap out messages on a displayed keyboard.

One of the chief selling points of the Treo is that it includes all the personal organization tools of the PalmOne PDAs: the calendar, phone book, to-do list and so on. But the built-in software is not as capable as I would have liked. For example, the Treo allows you to view Microsoft Word and Excel documents that are e-mailed to the device, but you can't edit the files or create new ones; to do so, you have to download additional software (at a cost of $30 or more). Considering the Treo's relatively steep price—$450 and up, depending on which cellular service you use—I expected the basic package to include more applications. In contrast, T-Mobile offers a roughly comparable Pocket PC phone that comes with a full range of programs (including versions of Word and Excel) for only $400. And the Pocket PC's screen is larger and has better resolution than the Treo's.

Perhaps the best feature of the Treo is its Web browser. Unlike earlier generations of Internet-connected gadgets, the Treo can view virtually any Web site, not just those specially formatted for handheld devices. The Treo's browser reshapes the Web page into a narrow column so that all the content can appear on the small screen, albeit with a lot of scrolling. I had great fun with the Treo during my

morning commute, reading the online edition of the *New York Post* as I rode the crosstown bus. The device even showed some practical value: while boating in the Peconic Bay off Long Island, I used the Treo to get a weather report.

The only major disappointment was the browser's speed, which is limited by the laggardly data-transmission systems employed by the cellular networks. Unfortunately, the Treo 600 cannot connect to the Internet by Wi-Fi, the wireless system that transmits data at much faster rates from network "hot spots." If you want a PDA phone with Wi-Fi access, you might want to consider the Motorola MPx or Hewlett-Packard's iPAQ h6315, both of which are expected this fall.

The Treo also contains a digital camera that takes pictures through a minuscule lens on the back of the device. Like other camera-equipped cell phones, the Treo can transmit the photographs wirelessly as attachments to messages. Although the quality of the images is pretty lousy, this spontaneous sharing of snapshots can definitely be entertaining. While walking down the street, I took a picture of my foot and e-mailed it to my

COURTESY OF PALMONE

wife along with a snippet of electronic music that the Treo identified, rather vaguely, as "Loud Aria." (The tune is actually from the Queen of the Night aria in Mozart's *Magic Flute*.) I titled the message "Song about My Foot." Sadly, my wife wasn't amused.

I decided to take a different tack and try to impress her with the incredible usefulness of the Treo. I showed her how I had linked a photo of one of my colleagues to his telephone number, so that his picture popped up on the Treo's screen whenever he called me.

"Isn't that great?" I said. "I'm going to do the same thing for everyone in my phone book. I'll always know who's calling even if I don't recognize the voice."

My wife was skeptical. "You don't need a picture for that. The person's name will be on the screen."

"Well, let's say it's someone I don't know very well," I countered. "Like a person I've seen only once at a business meeting. I'll probably remember his face better than his name."

"So the next time you're in a meeting, you're going to photograph everyone there?"

I pointed at the Treo's miniature lens. "I can do it surreptitiously."

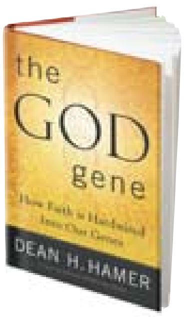
My wife shook her head. "That's the most ridiculous thing I've ever heard."

She was just jealous, I thought. After all, she was a mere BlackBerry user, whereas I was a member of the digerati! To console myself, I downloaded a game called Marbles Squared to my Treo. It was good for playing on the subway, where I couldn't connect to the Web. But one morning on the downtown local, I looked up from my screen and saw that the person sitting next to me was playing the same game on his Treo. He was a pimply teenager with peach fuzz on his upper lip. So much for my delusions of technological grandeur.

Bottom line: the Treo 600 is fun, and it might even enhance your productivity at work, but don't expect it to enhance your marriage. SA

Faith-Boosting Genes

A SEARCH FOR THE GENETIC BASIS OF SPIRITUALITY **BY CARL ZIMMER**



THE GOD GENE: HOW FAITH IS HARDWIRED INTO OUR GENES
by Dean H. Hamer
Doubleday, New York, 2004 ($24.95)

By page 77 of *The God Gene,* Dean H. Hamer has already disowned the title of his own book. He recalls describing to a colleague his discovery of a link between spirituality and a specific gene he calls "the God gene." His colleague raised her eyebrows. "Do you mean there's just one?" she asked.

"I deserved her skepticism," Hamer writes. "What I meant to say, of course, was 'a' God gene, not 'the' God gene."

Of course. Why, the reader wonders, didn't Hamer call his book *A God Gene*? That might not have been as catchy, but at least it wouldn't have left him contradicting himself.

Whatever you want to call it, this is a frustrating book. The role that genes play in religion is a fascinating question that's ripe for the asking. Psychologists, neurologists and even evolutionary biologists have offered insights about how spiritual behaviors and beliefs emerge from the brain. It is reasonable to ask, as Hamer does, whether certain genes play a significant role in faith. But he is a long way from providing an answer.

Hamer, a geneticist at the National Cancer Institute, wound up on his quest for the God gene by a roundabout route. Initially he and his colleagues set out to find genes that may make people prone to cigarette addiction. They studied hundreds of pairs of siblings, comparing how strongly their shared heredity influenced different aspects of their personality. In addition to having their subjects fill out psychological questionnaires, the researchers also took samples of DNA from some of them. Hamer then realized that this database might let him investigate the genetics of spirituality.

He embarked on this new search by looking at the results of certain survey questions that measured a personality trait known as self-transcendence, originally identified by Washington University psychiatrist Robert Cloninger. Cloninger found that spiritual people tend to share a set of characteristics, such as feeling connected to the world and a willingness to accept things that cannot be objectively demonstrated. Analyzing the cigarette study, Hamer confirmed what earlier studies had found: heredity is partly responsible for whether a person is self-transcendent or not. He then looked at the DNA samples of some of his subjects, hoping to find variants of genes that tended to turn up in self-transcendent people.

His search led him to a gene known as VMAT2. Two different versions of this gene exist, differing only at a single position. People with one version of the gene tend to score a little higher on self-transcendence tests. Although the influence is small, it is, Hamer claims, consistent. About half the people in the study had at least one copy of the self-transcendence-boosting version of VMAT2, which Hamer dubs the God gene.

Is the God gene real? The only evidence we have to go on at the moment is what Hamer presents in his book. He and his colleagues are still preparing to submit their results to a scientific journal. It would be nice to know whether these results can withstand the rigors of peer review. It would be nicer still to know whether any other scientists can replicate them. The field of behavioral genetics is littered with failed links between particular genes and personality traits. These alleged associations at first seemed very strong. But as other researchers tried to replicate them, they faded away into statistical noise. In 1993, for example, a scientist reported a genetic link to male homosexuality in a region of the X chromosome. The report brought a huge media

GENETIC FEEL for the divine?

CREATION OF ADAM, BY MICHELANGELO; ERICH LESSING *Art Resource, NY* fanfare, but other scientists who tried to replicate the study failed. The scientist's name was Dean Hamer.

To be fair, it should be pointed out that Hamer offers a lot of details about his study in *The God Gene,* along with many caveats about how hard it is to establish an association between genes and behavior. But given the fate of Hamer's so-called gay gene, it is strange to see him so impatient to trumpet the discovery of his God gene. He is even eager to present an intricate hypothesis about how the God gene produces self-transcendence. The gene, it is well known, makes membrane-covered containers that neurons use to deliver neurotransmitters to one another. Hamer proposes that the God gene changes the level of these neurotransmitters so as to alter a person's mood, consciousness and, ultimately, self-transcendence. He goes so far as to say that the God gene is, along with other faith-boosting genes, a product of natural selection. Self-transcendence makes people more optimistic, which makes them healthier and likely to have more kids.

These speculations take up the bulk of *The God Gene,* but in support Hamer only offers up bits and pieces of research done by other scientists, along with little sketches of spiritual people he has met. It appears that he has not bothered to think of a way to test these ideas himself. He did not, for example, try to rule out the possibility that natural selection has not favored self-transcendence, but some other function of VMAT2. (Among other things, the gene protects the brain from neurotoxins.) Nor does Hamer rule out the possibility that the God gene offers no evolutionary benefit at all. Sometimes genes that seem to be common thanks to natural selection turn out to have been spread merely by random genetic drift.

Rather than address these important questions, Hamer simply declares that any hypothesis about the evolution of human behavior must be purely speculative. But this is simply not true. If Hamer wanted, he could have measured the strength of natural selection that has acted on VMAT2 in the past. And if he did find signs of selection, he could have estimated how long ago it took place. Other scientists have been measuring natural selection this way for several years now and publishing their results in major journals.

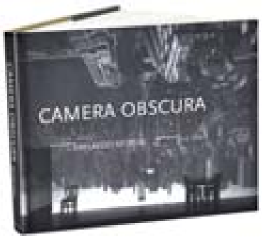
The God Gene might have been a fascinating, enlightening book if Hamer had written it 10 years from now—after his link between VMAT2 and self-transcendence had been confirmed by others and after he had seriously tested its importance to our species. Instead the book we have today would be better titled: *A Gene That Accounts for Less Than One Percent of the Variance Found in Scores on Psychological Questionnaires Designed to Measure a Factor Called Self-Transcendence, Which Can Signify Everything from Belonging to the Green Party to Believing in ESP, According to One Unpublished, Unreplicated Study.* SA

Carl Zimmer's books include Soul Made Flesh *and* Evolution: The Triumph of an Idea.

THE EDITORS **RECOMMEND**

CAMERA OBSCURA

Photographs by Abelardo Morell. Introduction by Luc Sante. Bulfinch Press, 2004 ($60)

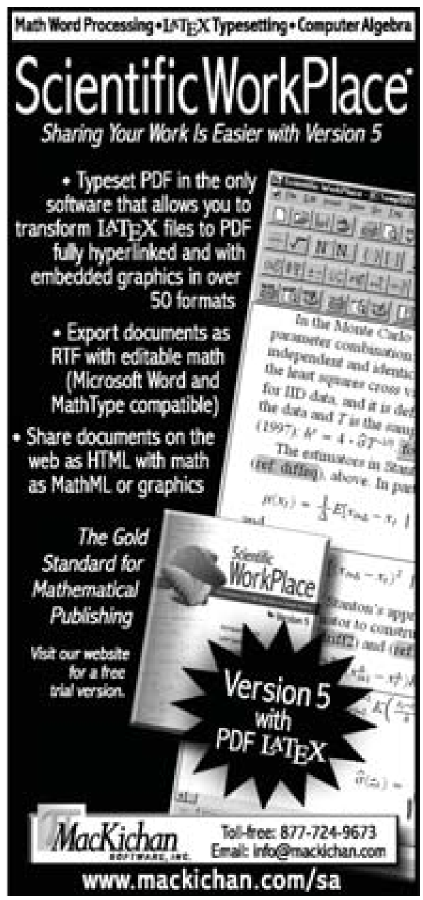


As the world of photography grows ever more digitized, Morell offers a glorious and surprising reminder of its classical roots. The well-known Cuban-born photographer essentially turns a room into the interior of a camera. He blacks out the windows, leaving a pinhole opening in one of them. Because of the nature of refracted light, the scene outside the window is projected upside down into the dim room. Morell then captures the room on film with a large-format view camera; exposures can take eight hours or more. The juxtapositions in the book's 60 duotones are eerily beautiful: New England clapboard houses hang serenely on the walls of a child's bedroom strewn with toy dinosaurs; Times Square throws a patchwork over the walls and bed of a Manhattan hotel room; the cityscape of Havana spreads across the crumbling interior wall of an apartment.

BOSTON'S Old Custom House in a hotel room, 1999

The books reviewed are available for purchase through www.sciam.com

COURTESY OF BULFINCH PRESS



Six-Legged Cinema

THE BIG SCREEN HAS BEEN BESET BY BUGS SINCE THE BEGINNING **BY STEVE MIRSKY**

They're baaaack.

I'm not referring to some horror movie monster, although that's what the line most likely conjures up in the imagination. I'm talking about May R. Berenbaum and Richard J. Leskosky, who have once again teamed up to write a scholarly piece sure to interest all fans of the science in science fiction.

Berenbaum is head of the entomology department at the University of Illinois, and Leskosky is assistant director of the Unit for Cinema Studies at the same institution. Over a decade ago the married couple wrote "Life History Strategies and Population Biology in Science Fiction Films," an article showing that the overwhelming majority of invading aliens in sci-fi movies are doomed by their own biology. [See "Nothing Personal, You're Just Not My Type," Science and the Citizen, Scientific American, February 1995.] Most sci-fi flicks feature more or less human-size creatures that show up and expect to be running things in no time. True, that approach actually worked for Arnold Schwarzenegger, but in real life, successful colonizers tend to be small organisms that produce huge numbers of offspring—for example, insects and studio executives.

More recently, Berenbaum and Leskosky wrote a piece for movie bugs on movie bugs: "Insects in Movies" appears in the new volume *Encyclopedia of Insects*. The entry reveals that moviemakers sometimes step all over bugs. "What constitutes an insect in cinema is not necessarily consistent with scientific standards," Berenbaum and Leskosky write. "In the taxonomy of cinema, any jointed-legged, segmented organism with an exoskeleton is likely to be classified as an insect, irrespective of how many legs or how few antennae it possesses." No less a polymath than Sherlock Holmes seemingly can't count. He incorrectly refers to a spider as an insect in 1944's *Sherlock Holmes and the Spider Woman,* which is about as elementary an entomological error as you can make. (And, yes, bugs should not technically be used as a synonym for insects, but we're taking etymological license.)

"Insect pheromones figure prominently in insect fear films," Berenbaum and Leskosky note, although in the 1978 movie *The Bees,* the characters call the chemical communication compounds "pherones." Lose the "er," too, and the insects could just call each other.

Berenbaum and Leskosky point out that although pheromones exist throughout nature, including in humans, they are rarely encountered in film in organisms other than insects. But 1977's *Empire of the Ants* acknowledges the human susceptibility to pheromonic influence: "Giant ants," Berenbaum and Leskosky explain, "use pheromones to enslave the local human population and to compel the humans to operate a sugar factory for them." In Florida, this same phenomenon is called agribusiness.

And let's talk about humongous ants. "A recurring conceit in insect films is the violation of the constraint imposed by the ratio of surface area to volume," write the arthropod aficionados. As body size increases, the ratio of surface area to volume decreases. Insects get their oxygen by taking in air through openings on their body surface, and if they got big enough the demands of all that volume on the relatively small surface would suffocate them. And when they molted, losing their chitinous corset, they would literally sag to death. Gregor Samsa did not awake one morning to find that he had been turned into a gigantic insect.

Obstinate insects have even played a fundamental role in movie history, inspiring one of the earliest efforts in stop-action animation, the 1910 short film *Battle of the Stag Beetles.* Entomologist Wladyslaw Starewicz first tried filming stag beetles in action, but like many temperamental actors, they objected to the hot lights and refused to perform. "Accordingly," Berenbaum and Leskosky write, "[Starewicz] dismembered the beetles and wired their appendages back onto their carcasses, painstakingly repositioning them for sequential shots." Using that same technique, just think of the performances you could get out of the Hilton sisters. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

Why do some expectant fathers experience pregnancy symptoms?

—D. Barrera, McAllen, Tex.

Katherine E. Wynne-Edwards, a professor of biology at Queen's University in Kingston, Ontario, who studies hormonal changes in expectant fathers, offers an answer:

Many factors—from social to hormonal—could play a role when an expectant father experiences pregnancy side effects such as nausea, weight gain, mood swings and bloating. The condition is called couvade, from the French verb *couver,* which means "to hatch" or "to brood."

The prevalence of couvade is difficult to measure because of its varied definitions and the skewed rates of symptom reporting. In modern Western populations, estimates of couvade's frequency range from less than 20 percent to more than 80 percent of expectant fathers.

Only recently has this phenomenon received attention from biologists, spawning a variety of hypotheses. For one, change in one partner's lifestyle can affect the other's: the cravings and increased appetite of a pregnant woman may pave the way for the father's weight gain, heartburn and indigestion. The mother's feelings can range from frustrated incapacitation to boundless anticipatory joy, fostering in the father jealousy of the ability to carry a child, guilt over having caused this sometimes unwelcome transformation in his partner and selfish attention seeking. Changes in sexual activity, shifts in social priorities, time off from work, or the arrival of a relative for a potentially stressful extended visit may also contribute.

Recent studies have shown that some of the same hormones that fluctuate in pregnant women may be affected in future fathers, making male hormones another factor. Men with higher levels of prolactin, which causes lactation in women, report more couvade symptoms, and their paternal prolactin peaks just before delivery. Levels of cortisol (a steroid hormone secreted in response to stress) and the sex steroids estradiol and testosterone also change in the father, though not as significantly as those in the mother.

Unfortunately, we don't yet know whether such hormonal changes and behaviors are cause-and-effect patterns or just coincidental. It is certainly tempting to look to hormones as the definitive root of couvade, but social and emotional factors could be equally influential. Either way, questions in this area have quietly expanded the horizons for research on male hormone levels—testosterone alone is clearly no longer the sum of the man.

Why does a shaken soda fizz more than an unshaken one?

—P. Biesemeyer, Malone, N.Y.

Chemist Chuck Wight of the University of Utah explains:

The short answer is that shaking = bubbles = fizz = sticky mess on your floor. But there's more to it than that. Creating bubbles in a soft drink takes a surprising amount of energy, and the turbulence caused by shaking a can or pouring it quickly into a glass speeds up their formation.

The tingly texture of a soft drink comes from dissolved carbon dioxide molecules that escape as bubbles from an open can, causing the soda to go flat. Forming bubbles requires the gas to overcome the surface tension of the liquid and part its surface molecules, a feat that requires a relatively large amount of energy. Once a bubble is formed, a smaller amount of work is needed to expand the bubble by vaporizing additional carbon dioxide molecules. (Less energy is needed because as a bubble grows, its volume increases faster than its surface area.)

Shaking the can introduces lots of small bubbles into the liquid, allowing the dissolved gas to vaporize more easily by joining existing bubbles. Having circumvented the difficult step of bubble formation, the gas can escape more quickly from shaken soda, resulting in more fizz. SA

For a complete text of these and other answers from scientists in diverse fields, visit www.sciam.com/askexpert

SARA CHEN