SPECIAL REPORT: MIRROR NEURONS AND THE MIND

ALSO: SUPERCONDUCTOR EYES • DEAD ZONES • PHONE VIRUSES

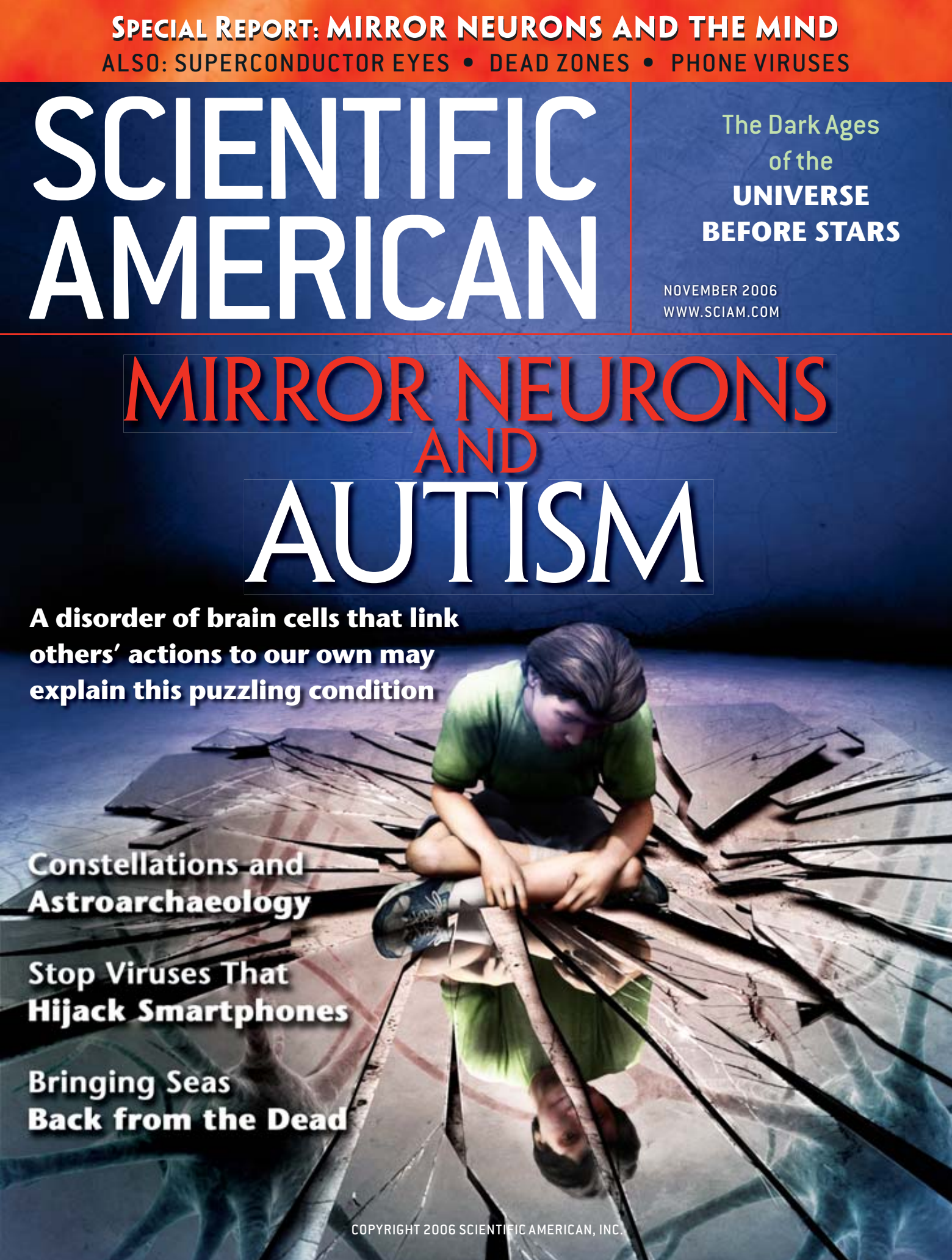
SCIENTIFIC AMERICAN

The Dark Ages of the **UNIVERSE BEFORE STARS**

NOVEMBER 2006
WWW.SCIAM.COM

MIRROR NEURONS AND AUTISM

A disorder of brain cells that link others' actions to our own may explain this puzzling condition



Constellations and **Astroarchaeology**

Stop Viruses That **Hijack Smartphones**

Bringing Seas **Back from the Dead**

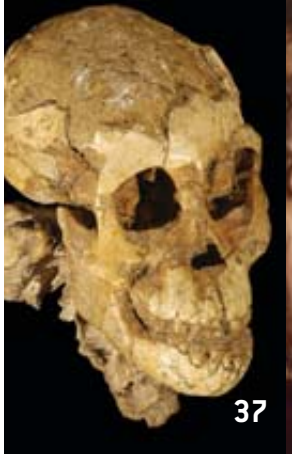
november 2006

contents features

SCIENTIFIC AMERICAN Volume 295 Number 5

54 Neurons of understanding

departments



columns

Cover image by Richard Marchand.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 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. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 248-7684. Send e-mail to sacust@sciam.com** Printed in U.S.A.

California, Here We Come

At the end of *An Inconvenient Truth,* the well-received documentary about global warming that appeared in theaters this past summer, the filmmakers provided a list of practical steps that anyone can take to tackle this looming environmental problem. Perhaps the most memorable suggestion was this one: "Vote for leaders who pledge to solve this crisis. Write to Congress. If they don't listen, run for Congress."

Unfortunately, the federal government is lagging behind other nations in the effort to control climate-warming gases—so now state lawmakers are taking the lead. In August the California legislature approved a bill calling for a 25 percent cut in emissions of carbon dioxide and other greenhouse gases by 2020. Although other states have pledged to make similar cuts, California is the first to mandate the emission limits. Moreover, the proposal won the support of both the Democratic-controlled legislature and Republican Governor Arnold Schwarzenegger.

POWER COMPANIES in California must cut their greenhouse gas emissions.

CHRIS CARLSON *AP Photo*

California's Chamber of Commerce opposed the bill, claiming that it would have no significant effect on global climate because polluting businesses would simply relocate outside the state. But this argument is based on a false assumption. Curbing greenhouse gases need not be costly. Improving the energy efficiency of a factory or office building can slash carbon use and save money at the same time. And thanks to recent technological advances, electricity generated from renewable energy sources—wind turbines, solar-thermal systems, geothermal facilities, and so on—is becoming economically competitive with power produced by burning coal or natural gas. Pacific Gas and Electric, one of the biggest utilities in California, broke with the rest of the industry by supporting the emission limits; the company currently generates 12 percent of its electricity from renewable sources (excluding large hydroelectric plants) and plans to increase that share to 20 percent by 2010.

Perhaps the greatest boon of the new California law is that it will encourage other states to take their own stands against global warming. The governors of seven northeastern states in the Regional Greenhouse Gas Initiative (Connecticut, Delaware, Maine, New Hampshire, New Jersey, New York and Vermont) have already agreed to lower their greenhouse gas emissions by 10 percent by 2019; the group recently issued a set of model rules that must now be approved by the states' legislatures or regulatory agencies. Meanwhile a dozen states have joined a federal lawsuit against the Environmental Protection Agency arguing that the Clean Air Act requires the agency to regulate greenhouse gases. The U.S. Supreme Court is expected to rule on that case next year.

But the federal courts may frustrate the states' efforts rather than furthering them. In 2002 California enacted a law calling for a 30 percent cut in the greenhouse gas emissions from cars and trucks sold in the state; two years later trade groups representing the auto industry filed a lawsuit claiming that the pending regulations are illegal because they effectively mandate an improvement in fuel economy and that only the federal government has the power to set such standards. We fervently hope that the courts reject this argument and recognize the fundamental right of states to protect their citizens from the catastrophic consequences of global warming.

THE EDITORS editors@sciam.com

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
212-451-8200
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
515-248-7684
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
415-503-3630
fax: 415-437-2892

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

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

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

Direct Response
Special Aditions Advertising, LLC
914-461-3269
fax: 914-461-3433

Australia
IMR Pty Ltd.
Tel: +612-8850-2220
Fax: +612-8850-0454

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

The Netherlands
Insight Publicitas BV
+31-35-539-5111
fax: +31-35-531-0572

Scandinavia and Finland
M&M International Media AB
+46-8-24-5401
fax: +46-8-24-5402

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:

NEWS

Colliding Clusters Shed Light on Dark Matter

A collision of two galaxy clusters—the so-called Bullet Cluster—has lifted the veil between ordinary and dark matter, proving that the latter must exist because of its outsized gravitational effect.

Scientists Identify Gene Difference between Humans and Chimps

A new computational technique has identified 49 regions in our genome that have changed particularly quickly between humans and chimps. It may have revealed at least one gene critical to the development of our larger brains.

Landscape Influences Human Social Interaction

Preliminary results from an ongoing study show that neighbors are more likely to be social when living among lush lawns, boding ill for efforts to encourage indigenous ground cover.

Hobbit Hubbub

A new paper has intensified the debate over the tiny human skeleton from Flores known as the Hobbit. At issue is whether it represents a new species or is instead a diseased member of our own kind.

PODCAST

Visit www.sciam.com/podcast to download our free daily and weekly audio programs, *60-Second Science* and *Science Talk*. Recent topics include the demotion of Pluto, a vaccine for weight loss and "Crocodile Hunter" Steve Irwin's tragic death.

NASA

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

IF A THEME emerged from the July issue, it was the technology that enhances our observational abilities. In "The Quest for the Superlens," John B. Pendry and David R. Smith wrote about building a lens that could overcome the resolution limits of the illuminating light wavelengths. "Hubble's Top 10," by Mario Livio, reviewed the greatest discoveries made with the orbiting Hubble Space Telescope. In "CSI: The Reality," Max M. Houck looked at how the portrayal of criminal investigation tools and methodologies on television forensics dramas might be making viewers unreasonable jurors in real-life courtrooms.

But even with better observational tools, can we trust the initial conclusions we draw from them? Not likely. In "The Political Brain" [Skeptic], Michael Shermer reported that a brain-imaging study showed our predilections to be a product of unconscious confirmation bias. Henry McDonald of Danville, Calif., found everyday evidence: "I believe that confirmation bias is an evolutionary trait that bonds the family together, thus increasing our chance of survival. The fact that my wife of 51 years still loves me and thinks I am terrific is positive proof of confirmation bias."

COLD CASE

In "A Power Grid for the Hydrogen Economy," by Paul M. Grant, Chauncey Starr and Thomas J. Overbye, the authors left unanswered an obvious question: Is a hybrid hydrogen-electricity SuperCable superior to a pipeline that carries just cryogenic hydrogen? In the latter case, nuclear, thermal or other energy conversion stations optimized for the production of hydrogen would feed the pipeline, and hydrogen-fueled generating plants would be located along its length or at its terminus. Clearly, the single-conduit hydrogen pipeline would be much less costly to build than the double-conduit hybrid SuperCable with its superconducting components. Are there superior energy-transmission efficiencies of the hybrid conduit configuration that are sufficient to offset the great advantages—lower capital costs, greater flexibility and lower risks—of the cryogenic pipeline alternative?

Leonard Miller
New York City

GRANT, STARR AND OVERBYE REPLY: *Miller's letter addresses indirectly an issue much discussed in energy-future circles: How much electricity and how much hydrogen will be needed? For our article, we tacitly assumed electricity generation—its transmission, distribution and consumption—will remain proportionate into the future relative to the present, perhaps including growing use for home heating if nuclear power becomes as inexpensive as we anticipate. On the other hand, hydrogen will principally target transportation, supplanting fossil fuels. If our crystal ball is clear, it will be critical that large amounts of electricity be delivered over short and long distances as efficiently as possible—and that means using superconductivity. Today's superconductors need to be cooled to 77 kelvins or below, and cryogenic hydrogen will do the job quite well, so why not generate and deliver both forms of power? Reducing the cost of superconducting wire is a work in progress by the Department of Energy and industry, and it is expected by 2010 that this wire will be price/performance competitive with copper held at the same low temperature but with essentially zero resistance. The energy savings alone could offset the refrigeration cost, especially if avoidance of a potential carbon tax is included.*

REFLECTIONS ON REFRACTION

With respect to "The Quest for the Superlens," by John B. Pendry and David R. Smith, please allow me to inquire if it is conceivable that the (interstellar) space around heavy objects, such as black holes, also has a negative refractive index.

Thomas Urbanek
Calgary, Alberta

The authors remark on the use of a thin layer of silver to focus light. Does the "lustrous" quality of polished metal have something to do with this phenomenon?

Robert Pfingston
Morganfield, Ky.

PENDRY AND SMITH REPLY: *Negative refraction occurs only at a nonzero frequency that is determined by how electrical charges in the metamaterial respond to electromagnetic waves. A black hole does not produce the kind of oscillating electrical response that is required. Alternatively, a quantity called the metric determines the refractive effects of curved space. A negative metric is not possible; it would lead to all kinds of paradoxes.*

Pfingston's guess is exactly right. Silver is lustrous because the material's dielectric function is negative, the same property that leads to the focusing effect of a thin layer.

HUBBLE'S HIT PARADE

Mario Livio's "Hubble's Top 10" makes the excellent point that the Hubble Space Telescope is an important adjunct to a constellation of ground-based telescopes, not, as the public often believes, a replacement for them.

A second point, often lost even on professional astronomers, is that Hubble's research in many cases builds on the work of small, not giant, telescopes. For example, the first three items on Livio's list are based on the discoveries of small telescopes: Shoemaker-Levy 9 was discovered with a 0.4-meter Schmidt telescope; the first extrasolar planet was found with a 1.9-meter telescope, and most of the early finds were spotted via a 0.6-meter telescope; and Supernova 1987A was discovered with a 0.3-meter telescope, with the only precursor data gathered by a 0.6-meter Schmidt.

The lower cost of operating small telescopes permits astronomers the latitude to undertake riskier or longer-term research, which results in unexpected discoveries, the exciting hallmark of astronomy. Unfortunately, the modest dollars needed for maintenance, instrumentation and use of such telescopes—a small fraction of the costs of large ones—have for the most part been diverted to big telescope projects.

William Bruce Weaver
Director, Monterey Institute for Research in Astronomy
Marina, Calif.

PEERLESS PEERING: **Hubble's infamous mirror flaw was corrected by shuttle astronauts during its first pit stop in 1993. Its performance has been stellar ever since.**

AS SEEN ON TV

In "CSI: The Reality," Max M. Houck implies that if the "CSI effect" exists in jury rooms, it is entirely harmful. Yet as we see ever more death penalty convictions overturned and as studies show the terrible fallibility of eyewitness testimony, it becomes clear that more scientifically valid forms of evidence are long overdue. The allegedly ignorant public may be onto something.

Paul Cantrell
Minneapolis

HOUCK REPLIES: *As long as juries are influenced by the media and inappropriately apply the results of scientifically valid evidence, the CSI effect is harmful. The best evidence in the world does no good if it is misinterpreted and misapplied. Juries are not ignorant, perhaps just misled. The response to improvements in forensic science—to forestall inaccurate convictions of any kind—is education, training, quality assurance, certification (of scientists) and laboratory accreditation. Only half the smaller forensic labs in the U.S. are accredited, and typically less than 1 percent of lab budgets is for training or quality assurance. Only 12 percent of U.S. forensic facilities have research funding. Both certification of forensic scientists and accreditation of their labs are voluntary. Clearly, our national funding priorities must be aligned with our demands for the criminal justice system.*

HAZARDS OF BIOTINKERING

I've been a reader since the 1970s, and after all these years I was quite excited to encounter what appeared to be a typo in "Reprogramming Biology" [Forum]. The second-to-last paragraph begins: "As an information technology, biology is subject to …" Of course, I thought it should have read, "As *in* information technology …" But then I identified the author, Ray Kurzweil. His tendency to analogize biology to information technology is arrogant and naive. Nature has repeatedly shown us that we are not as smart as we think we are. Global warming, pollution and species extinction are a result of our ignorant disruption of a delicate, eons-old balance. The same unexpected consequences await our intrusion into the ancient inner workings of genes and cell machinations.

Kurzweil points out that life expectancy was only 37 years in 1800. Yet its dramatic rise since then has been (merely) the result of macroknowledge of biological systems, (relatively) rudimentary surgical techniques, improved sanitation and nutrition, and the use of drugs based on serendipity and empirics. Given our penchant for spoiling the stew, we'd be wise to take Kurzweil's grandiose predictions with a grain of salt—and stick with taking care of ourselves "the old-fashioned way" for a great while longer.

John Herman
Dix Hills, N.Y.

ERRATUM In "Brief Points" [News Scan], the genus name *Giganotosaurus* was misspelled in the item comparing its size with that of the newly discovered carnivorous dinosaur *Mapusaurus roseae.*

NASA

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

Experimental Violence ▪ "Father of Aviation" ▪ The New Iron Age

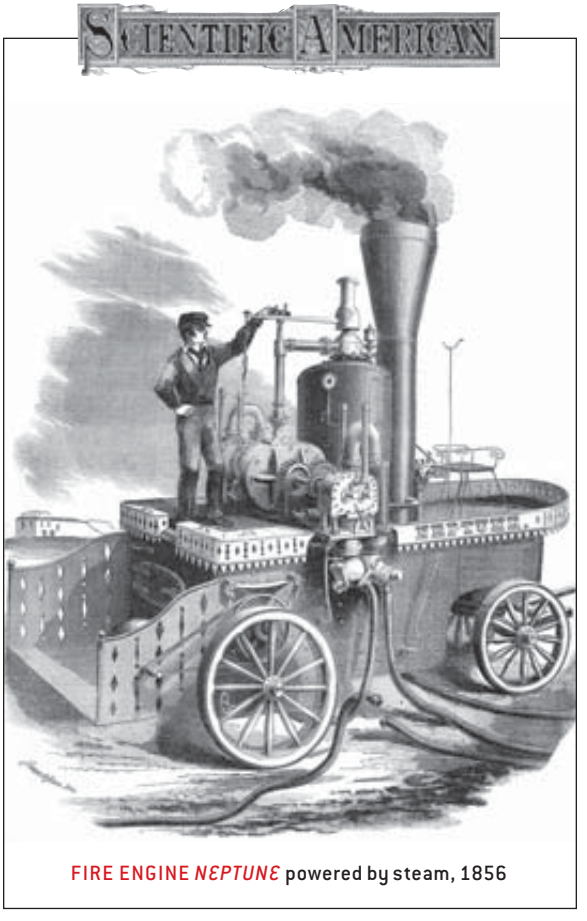
NOVEMBER 1956

RESEARCHING "THE GAME"—"Our working hypothesis was that when two groups have conflicting aims—i.e., when one can achieve its ends only at the expense of the other—their members will become hostile to each other even though the groups are composed of normal well-adjusted individuals. To produce friction between the groups of boys we arranged a tournament of games: baseball, touch football, a tug-of-war, a treasure hunt and so on. The tournament started in a spirit of good sportsmanship. But as it progressed good feeling soon evaporated. The members of each group began to call their rivals 'stinkers,' 'sneaks' and 'cheaters.' The rival groups made threatening posters, planned raids, and collected secret hoards of green apples for ammunition.—Muzafer Sherif"

NOVEMBER 1906

SANTOS-DUMONT FLIES—"On October 23 Alberto Santos Dumont, in the afternoon, drove his aeroplane through the air a distance of 150 feet at an elevation of about 20 feet from the ground. The experiment took place near Paris, and was witnessed by a crowd of people, including representatives of the Aero Club of France. According to the cabled account, the stability of the machine appeared to be good. At any rate, this is the first flight of a motor-driven, man-carrying aeroplane that has been witnessed by a considerable number of people. In comparing these results with those which the Wright brothers claim to have attained, there is one striking fact, viz., the young Brazilian found that a 50-horse-power motor was necessary to drive his flier up into the air; while the Wrights, with a machine of twice the weight and half the power, claim to have made nearly double the speed."

AGE OF STEEL—"Rarely, if ever, in the history of modern industries has there been witnessed such an extraordinary condition as confronts the steel industry in the United States. The mills, which are devoted to the production of structural steel, are overloaded with work. Even more acute conditions prevail at the plate mills, the demand for whose output is to be attributed very largely to the growing popularity of steel cars."



FIRE ENGINE *NEPTUNE* powered by steam, 1856

MEAT FOR THE POOR—"A peculiar German institution, which appears anything but attractive to American eyes, is the *freibank*. It is a shop devoted to the public and official sale of condemned meat which has, theoretically, been made wholesome by sterilization. For example, meat which contains tapeworm larvae is submitted to prolonged refrigeration, while tuberculous meat is heated to a high temperature in closed vessels. The prices of meats sold at the *freibank* are fixed by the local authorities, and the quantity sold to a single purchaser is limited. The *freibank* does not exist in France, but it is pointed out that the law of supply and demand creates *freibänke* of a very different sort, in which meat unfit for food is sold to customers who can not afford to buy any other."

NOVEMBER 1856

STEAM FIRE ENGINE—"The accompanying engraving represents the Rotary Steam Fire Engine *Neptune,* manufactured at the 'Island Works,' Seneca Falls, N.Y. The elliptical rotary pistons are applied in this fire extinguisher for both steam engine and force pump. At the Firemen's Tournament held at Seneca Falls, on the 10th and 11th September last, it threw two 1½ inch streams 170 feet horizontally."

CHEAP STEEL—"H. Bessemer has obtained an American patent for his improvement in the manufacture of iron. This claim is based upon a scientific discovery in the manufacture of the iron; namely, producing combustion without fuel, by forcing air, steam, or other gases through molten iron in a vessel, to supply oxygen to the carbon in the molten crude metal, and thus produce combustion to burn out the excess of carbon. The claim is not that he was the first that used air or steam in this manner, but that he discovered the cause of the effect produced by driving oxygenated gases through molten iron."

BIOTECHNOLOGY

Blastomere Blowup

A NOVEL WAY TO HARVEST STEM CELLS INTRIGUES AND INFLAMES **BY CHARLES Q. CHOI**

A method that can generate human embryonic stem cells without harming embryos? In August biotechnology firm Advanced Cell Technology (ACT) in Worcester, Mass., claimed it had developed just such a procedure. The company touted it as a way around the firestorm of controversy surrounding the conventional technique for growing these cells, which destroys human embryos. Most researchers find the method intriguing, because it might lead to new and maybe better stem cell lines. But several also argue that it raises fresh dilemmas.

Scientists investigate human embryonic stem cells because they can become any other kind of cell. As such, they hold great promise in regenerating body parts. Conventionally, researchers rely on embryos produced during in vitro fertilization (IVF) attempts, removing the inner cell mass from embryos that have grown to 50 to 100 cells. Currently at least 400,000 frozen embryos lie unused in U.S. fertility clinics, and thousands are typically discarded annually.

ACT's vice president of research and scientific development, Robert Lanza, and his colleagues experimented with 16 unused human embryos from an earlier, eight- to 10-cell stage. Physicians routinely pluck single cells from IVF embryos at that stage for preimplantation genetic diagnosis (PGD), which scans these cells, known as blastomeres, for diseases. Embryos can compensate for the loss and have so far developed into some 1,500 apparently healthy children since this procedure was introduced more than a decade ago.

Starting with a technique similar to PGD, the ACT researchers used micro eye-

OUTRAGE, AGAIN: New work has produced some fresh complaints about human embryonic stem cells. This colored scanning electron micrograph shows such a cell magnified about 4,200 times.

DAVID SCHARF *Photo Researchers, Inc.*

HYPE THAT HURTS

In announcing Advanced Cell Technology's new method of generating stem cells, *Nature*'s press office reported that single cells were removed from embryos and that embryos were left intact, errors repeated later in many news stories that led to *Nature* issuing corrections twice and to an angry backlash from journalists. The ire prompted a Senate hearing in September, in which senators who back stem cell research blasted ACT. Even if the method "actually works, it will take years for it to produce a substantial number of new lines," Senator Tom Harkin of Iowa said at the hearing. "We shouldn't make the mistake of holding out all our hope for one new, unproven method."

SINGLE CELL, called a blastomere, is suctioned from an embryo by the pipette on the left during preimplantation genetic diagnosis. It could lead to stem cells without harm to the embryo.

droppers to extract 35 blastomeres one or two at a time from the six highest-grade embryos they had. They then generated two stable human embryonic stem cell lineages, which they reported online August 23 via *Nature*. These cell lines apparently behave like their conventional counterparts, capable of replicating for months on end and developing into other cell types.

ACT ended up destroying its embryos—to make the most of the few embryos it had, Lanza's team eviscerated them by taking several cells from each (up to seven). But because standard PGD does not kill the embryo, the scientists declared that their method in principle can deliver stem cell lines without killing the embryos.

Researchers regard the cells generated by this method as scientifically intriguing, because they come from earlier-stage embryos. Development neurobiologist Arnold Kriegstein of the University of California, San Francisco, noted that the 50- to 100-cell-stage embryo cell lines cannot split to become twins, whereas eight- to 10-cell-stage embryos such as those ACT used can. It suggests that the younger cells "might have better potential for turning into certain kinds of tissue than currently used lines," he says. "Only further experiments will tell."

Even if investigators can get the method to work, however, the controversy will not abate. It remains uncertain whether single human blastomeres could develop into living beings. If they can, this technique would land the cell lines back in the original controversy. "In sheep and rabbits, for example, single cells are capable of developing into viable animals; however, the same does not appear to be true for mice," said James Battey, chair of the National Institutes of Health stem cell task force, at a Senate hearing in September.

In addition, the ACT study has already destroyed 16 embryos. Hence, whether the method can add to the number of federally funded human embryonic stem cell lines is unclear, the U.S. President's Council on Bioethics reported, because the law prevents funding of any additional lines produced through the destruction of embryos. Lanza says that ACT probably would file for federal funding to continue the work. The company has announced that—with the WiCell Research Institute in Madison, Wis., which hosts the National Stem Cell Bank—it would distribute ACT's new cell lines to U.S. scientists, provided that the federal government will fund research using them.

In the end, even if ACT overcomes the technical and ethical issues, it may face one last obstacle: resistance from physicians and expectant parents. In the *Nature* paper, the extracted cells were not grown in isolation but incubated near one another to help them thrive from chemical signals they each release. (A single blastomere isolated from other cells is less likely to lead to a cell line.)

Lanza says that ACT has some evidence that lone blastomeres placed next to the embryos from which they were extracted could grow without risk to the embryos. But whether such as-

GAMMA PRESSE

surance would be good enough for IVF doctors and their patients is not clear, points out IVF gynecologist Outi Hovatta of the Karolinska Institute in Sweden. They may not want to take any chances with their future children, even if ACT ultimately becomes ready, willing and able.

Charles Q. Choi, a frequent contributor, is based in New York City.

ENVIRONMENT

Keeping CO_2 Down

THE FIRST U.S. PROJECT FOR CARBON DIOXIDE BURIAL GIVES MIXED RESULTS BY REBECCA RENNER

Deep in the heart of Texas—just east of Houston, actually—a research project designed to answer some of the technical questions about storing carbon dioxide underground is producing its first results. Scientists hope that carbon capture and sequestration could help ameliorate global warming without harming the local environment. The preliminary data at this geologic site, called the Frio Formation, are thus far providing mixed lessons.

Interring carbon dioxide may become an important option for curbing global warming. "By 2050 nations could be burying five to 10 billion tons of carbon dioxide every year," says geologist Julio Friedmann of Lawrence Livermore National Laboratory. The capacity to store many decades' worth of carbon dioxide exists underground.

Saline aquifers—sandstone formations whose pores are full of salty water—are one of the best, most widespread candidates. But scientists need more information about them. "We know about injecting carbon dioxide—for 30 years oil companies have gotten extra oil out of depleted wells that way," says Sally Benson of Lawrence Berkeley National Laboratory. "What we don't know are the details of how to get lots in and keep it there for a long time."

Commercial efforts do exist—about 10 years ago Sleipner, an offshore gas platform in the North Sea run by Norwegian firm Statoil, began annually in-

BURIAL SERVICE: Carbon dioxide in a supercritical fluid state is pumped into the Frio Formation in Texas as a test of underground carbon storage.

SUSAN D. HOVORKA *Bureau of Economic Geology, University of Texas at Austin*

jecting one million tons of carbon dioxide into a thick saline aquifer. But projects such as Sleipner are not particularly well suited for research, says geologist Susan D. Hovorka of the University of Texas at Austin. To improve their models for what happens underground, researchers need to monitor injection experiments closely. Commercial sites are too busy to conduct such work, she notes.

That's where the Frio project comes in. In 2004 the $6-million project, led by Hovorka, began by compressing 3,000 tons of carbon dioxide into a supercritical fluid, warming it to 15 degrees Celsius and pumping it about 1.6 kilometers below the surface into a 23-meter-thick sandstone layer. "We want the carbon dioxide to move through the pores in the rock, get stuck in some of the pores, dissolve in the brine and even create new minerals" that trap the CO_2, Hovorka says. She is pleased that three-dimensional seismic imaging and other monitoring methods show that almost all the carbon dioxide injected at Frio has been caught in the pores or dissolved in the saltwater.

But Hovorka, U.S. Geological Survey geochemist Yousif Kharaka and their colleagues have also found that the dissolved carbon dioxide made the water more acidic. That water in turn dissolved some of the minerals in the sandstone, releasing calcite and metals, mainly iron. This could be good or bad. Dissolving some of the rock leaves more space to store carbon dioxide. But released metals might migrate to the surface and pose an environmental hazard. Some saline aquifers, for instance, may contain arsenic and uranium, which are best left undisturbed.

Kharaka states that saline aquifers are still excellent locations for carbon storage, noting that in well-sealed aquifers, such fluids will not escape. The acidic brines could attack cements used to install injection wells, so such wells need to use acid-resistant cements, and old abandoned wells should be avoided, he points out.

The Frio data also help to evaluate other sites, comments chemical engi-

SLEIPNER GAS PLATFORM in the North Sea was the first to pump carbon dioxide underground, although as a commercial site it does not generate the kind of data that scientists need to evaluate the strategy.

STATOIL

neer Howard Herzog of the Massachusetts Institute of Technology. "It's one thing to cherry-pick only good locations and another thing to have enough knowledge to implement carbon storage across a large variety of sites," Herzog says. Detailed data about how CO_2 moves through a formation can come only from a research project like Frio, he adds.

Hovorka and her colleagues may need to work quickly: commercial carbon storage is coming fast. This summer Japan announced plans to bury 200 million tons of carbon dioxide annually by 2020. Next year oil giant BP intends to start work on a $1-billion plant near Los Angeles to convert petroleum coke, a by-product of oil refining, into hydrogen and store about four million tons of carbon dioxide a year.

Rebecca Renner is based in Williamsport, Pa.

OPTICS

Muscling Up Color

POLYMER SPLITS LIGHT FOR TRUE COLOR IN DISPLAYS

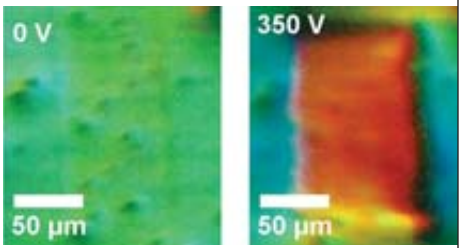
BY STEVEN ASHLEY

Artificial muscles—plastics that expand and relax when exposed to electric fields—could help produce truly lifelike colors in future television and computer screens. Tiny "tunable prisms" based on these materials could form the pixels of improved video displays within a decade.

Existing screens, such as those based on TV tubes, flat-screen LCDs or plasma displays, cannot faithfully reproduce the full range of colors that humans can see. Each pixel in those technologies consists of three light-emitting elements, one for each of the fundamental colors: red, green and blue. The displays combine the colors at various brightness levels to generate other colors but can achieve only a limited range.

Manuel Aschwanden and Andreas Stemmer of the Swiss Federal Institute of Technology in Zurich have devised a new way to paint the screen. They rely on arrays of reflective diffraction gratings—tiny optical elements, each composed of a set of fine, parallel and equally spaced grooves on a surface. These grooves spread white light into a rainbow just as prisms do. "You get the same effect when you tilt the bottom side of a CD in direct sunlight," Aschwanden says. "The spectrum reflects off the ruled surface."

The two researchers have built a proof-of-concept array containing 10 pixels, each with a diffraction grating. White light first enters a grating, which is about 75 microns on a side, Aschwanden explains. The grooves—a micron apart—are molded on one side of a thin polymer membrane. When dif-



VOLTAGE BEGETS COLOR when the light-diffracting grooves on this tiny artificial muscle expand in an electric field.

COURTESY OF MANUEL ASCHWANDEN

NO HARD BUT FAST RULE

Diffraction gratings, which were first developed in the early 1800s, are commonly employed in projector systems and fiber-optic telecommunications devices. But unlike existing hard, solid gratings, a new type made of a flexible polymer (an acrylic dielectric elastomer) expands in the planar directions when voltage is applied. In contrast to the hard materials, which must shift position rapidly via a mechanical actuator, the "artificial muscle" grating lengthens by 26 to 32 percent almost instantly in the presence of an electric field.

ferent voltages are applied, the grating expands or contracts, so the incoming light encounters grooves that are spaced farther or closer apart. This effect changes the angles at which the light reflects back, which effectively shifts the outgoing spectrum from side to side. The system isolates specific colors with a small hole placed in front of the grating. Altering the voltage makes different colors visible when they line up with the hole.

To produce composite colors in a full-size display, every pixel would incorporate two or more diffraction gratings. This would be necessary because some colors, such as brown, are not part of the spectrum that makes up white light.

Though too small to be useful, the system has the same pixel density as a high-quality LCD display, says Aschwanden, who readily acknowledges that his invention is still is a long way from application in a video product. The next prototype will feature an array of 400 gratings. The current display operates at 300 volts, far too high for home use, but new materials now under development could allow the activating voltage to drop.

"This is a very interesting approach to a color display," comments Olav Solgaard, an electrical engineer at Stanford University and one of the founders of Silicon Light Machines, a pioneer in micro-optoelectronic technology, "but it's a very steep hill to climb to get to a practical display." He cited several potential obstacles, which include achieving a "truly dark pixel" to create good contrast and efficiently maintaining image brightness given that the grating "throws away a good portion of the light." The technique might be most useful for passive displays that reflect back ambient white light, such as might be used in a cell phone.

In any case, the Zurich researchers are going beyond displays. They have already developed a prototype high-resolution microscope that redirects monochromatic light beams with artificial muscle membranes. "Tuning or steering light is basic to many optical systems," Aschwanden notes. "This approach offers a cheap, accurate way to do it."

Catchy Carbon

WHAT FLOYD LANDIS HAS IN COMMON WITH OCEAN SEDIMENT **BY SARAH SIMPSON**

When marine chemist John Hayes pioneered a way to scrutinize the carbon atoms in seafloor mud 15 years ago, he was trying to unravel mysteries about how dead microbes once lived. Never did he guess that sporting officials would one day use his invention to catch drug cheats.

By modifying Hayes's method to examine the carbon atoms in athletes' urine, medical researchers at the U.C.L.A. Olympic Laboratory have developed the first definitive screen for synthetic testosterone, a popular anabolic steroid banned by most sports organizations since the 1970s. The new test—known as the carbon isotope ratio (CIR) test—figures prominently in several recent, high-profile doping cases, including the disqualification of sprinter Justin Gatlin's world record in the 100 meters. It may also strip cyclist Floyd Landis of his 2006 Tour de France title.

Before the CIR test became standard protocol in the late 1990s, analysts relied on a more primitive urine screen that could recognize only higher-than-average ratios of testosterone to a related compound known as epitestosterone. But because these so-called T/E ratios are naturally elevated in some people, accusing an athlete of doping required follow-up exams, U.C.L.A. lab director Don H. Catlin explains. If subsequent urine samples contained lower T/E ratios, officials assumed a synthetic substance caused the initial spike. But if later screens matched the preliminary results, officials had to declare the athlete drug-free—even though a person could beat the test by ingesting artificial testosterone throughout the testing period, Catlin says.

What sport needed was a way to differentiate testosterone made in a factory from that produced in the body. Catlin and his

NEED TO KNOW: EAT TO CHEAT?

Cheating athletes who understand the basic premise behind the carbon isotope ratio drug test may be tempted to eat more tofu. The goal would be to mask their use of synthetic testosterone, often derived from soy plants, by changing the distinct carbon isotope fingerprint of their body's natural hormones. But Don H. Catlin of the U.C.L.A. Olympic Laboratory says don't bother. Results from his team's experiments suggest that even dramatic changes in diet hardly budge the average isotopic makeup of carbon atoms in a person's body.

Mission Control

CLOUD OF SUSPICION: Floyd Landis pulls away from the competition in stage 17 of the 2006 Tour de France. He ultimately won the three-week race but then failed drug tests, which indicated that he may have taken artificial testosterone to enhance his performance.

ALESSANDRO TROVATI *AP Photo*

colleagues knew they needed to find a distinct chemical fingerprint in the hormone that would allow them to trace the carbon atoms to one source or the other—just the way Hayes was able to determine whether his dead microbes ate the carbon dioxide in seawater or the methane bubbling up through gooey seafloor sediment.

As standard practice, pharmaceutical companies build testosterone molecules from a framework of carbon atoms derived from wild yams or soybeans. These warm-climate plants and their kin process carbon differently than do the more abundant, temperate-climate plants such as corn. This processing leaves the soy relatives notably depleted in carbon 13, an isotope that has one extra neutron and thus a slightly greater mass than the common isotope, carbon 12.

If the testosterone in an athlete's urine carries the soylike fingerprint, as Gatlin's and Landis's did, it sticks out like a sore thumb when compared with cholesterol or other hormones in the same urine sample. Those reference hormones, like all compounds produced by the human body, contain more carbon 13 (relative to carbon 12) than soy does. The variety of plant material in the food we eat—everything from apples to zucchini and even to the corn a feedlot cow consumes before becoming a hamburger—gives the naturally produced hormones their diagnostic enrichment.

The CIR test "strikes me as pretty close to bulletproof," says Hayes, now at the Woods Hole Oceanographic Institution. Changing the carbon isotope fingerprint of naturally secreted testosterone without also changing that of the reference hormones would be nearly impossible, he explains. Dehydration, alcohol consumption and many of the other excuses bandied about by accused athletes would simply have no effect. Only if the body were able to make testosterone from an artificial compound—such as the cortisone athletes sometimes inject to reduce muscle inflammation—might the natural hormone carry a synthetic-looking fingerprint, Hayes notes.

Catlin predicts that the efficacy of the CIR test, which also identifies a range of compounds related to testosterone, will make athletes think twice before taking steroids—unless dopers find a new way to beat the system. Until then, it seems they should heed an old adage: you are what you eat.

GEOSCIENCE

Martian Field Test

DRILLING IN GREENLAND TO PREPARE FOR MARS **BY CHRISTINA REED**

Summit Camp, Greenland—The guts of a thermal drill spilled out over a table and onto the wood-planked floor of a domed five-person tent. Five highly caffeinated planetary scientists from the Jet Propulsion Laboratory in Pasadena, Calif., continued updating the 122-centimeter-long, 7.6-centimeter-wide device with the software, firmware and hardware needed for its first field test. A square hole cut in the floor revealed the material that would serve as a Martian analogue.

The mission: to have the device, named Chronos, melt its way into Earth's ice sheet above the Arctic circle as a test of its worthiness for the coveted NASA mission to Mars in 2011. If chosen, Chronos would drill into our neighbor's northern ice cap. "Ice is probably the only accessible climate record on Mars; if there is a climate record, it is preserved in those ice caps" as variations in isotope abundance, explains lead engineer Greg Cardell.

Sitting on the accumulation of frozen precipitation at least 100,000 years old, Summit Camp reaches an elevation of 3.5 kilometers and provides an ideal location for testing ice drills. The team hoped to go 100 meters down, or about 200 years into the past. The goal for Mars is to drill 30 to 70 meters deep. Because of the Red Planet's thin, dry atmosphere, the ice on Mars has accumulated extremely slowly, so even 30 meters could date back thousands of years.

Chronos delivers power down a tether to heat a flat nose plate that pro-

MARS WORTHY: Ice drill named Chronos, operated inside a tent heated for human benefit (*left*), melted its way down about 50 meters. Blue LEDs illuminated the borehole (*below*).

CHRISTINA REED

STAYING WARM IN THE MARTIAN COLD

Researchers testing an ice-melting drill had to power down periodically to better mimic the frostier Martian climate. But toggling the heat also mimics the power conditions the drill would actually face. "On Mars we do not intend to melt continuously," explains task manager Miles Smith of the Jet Propulsion Laboratory. Instead solar cells will charge batteries at the surface that will in turn deliver electricity to the drill in short bursts of 200 to 400 watts. Aerogel insulation inside the tether will prevent the meltwater from refreezing on its journey back up. The spooling tether would be kept in a black box at the surface, where sunlight will keep it toasty during the summer months they plan to drill.

vides the only contact the drill has with the ice. A primary pump directs the meltwater from the borehole into the drill, and a secondary pump sends the meltwater up the tether to the surface for analysis. This keeps the borehole dry and prevents the hole from freezing over the drill. The side of the drill is also equipped with blue LEDs and a small camera to illuminate and document the ice along the way.

At 1:30 A.M. on July 19, Chronos began melting the ice at a rate of one centimeter a minute. But it quickly became clear that this Martian melt drill had some Earth hurdles to overcome. The field tent, usually heated for human benefit, was too warm even with the space heaters turned off. On Mars the drill will encounter temperatures that plunge to –110 degrees Celsius. At Summit Camp the air temperature was –15 degrees C and the ice –30 degrees C. To compensate, task manager Miles Smith and engineer Claus Mogensen programmed temperature controls to turn the drill heaters on and off remotely to relieve the overheating. "That fix is exactly the kind of software patch we could have done on Earth if this problem happened on Mars," Smith notes.

But Greenland has other ice properties unique to Earth: as deep as 70 meters, air bubbles create a porous ice layer called firn. Until water flowed through the tether, the team could not tell if the meltwater was draining into the firn or accumulating around the drill, which could freeze the borehole shut. Engineer Bob Kowalczyk added a temperature alarm to the outside of the drill that would signal them if the borehole rose above 0 degrees C. Still, the team remained tense until water flowed up for the first time, sending a collective cheer through the tent.

At that point, Mogensen started the laser-illuminated analyzer, which vaporized the incoming water and looked for variations in oxygen and hydrogen isotopes. "The mass spectrometer that used to take up an entire lab can now be done on a tabletop," he says.

Chronos went down almost 50 meters. "All and all a tremendous success," remarks Cardell, who stopped the mission when the drill hit a significant layer of mysterious dust that choked the filters on the last day. "We didn't prepare for coarse material, only fine particulates. It's a good lesson learned that we should prepare for the unexpected," Smith notes. "If Chronos hits a volcanic layer on Mars as thick as this, it would be front-page news. Of course, anything we get back from Chronos on Mars would be front-page news."

Christina Reed, based in Seattle, received permission from the National Science Foundation to travel with the NASA-sponsored JPL team.

PHARMACOLOGY

Not Imagining It

RESEARCH INTO HALLUCINOGENS CAUTIOUSLY RESUMES

BY DAVID BIELLO

Flashback: A middle-aged man enters a comfortably furnished room, sits on a couch and receives a pill. After swallowing the drug, his medical monitors place a mask over his eyes and headphones over his ears and encourage him to lie back. Soothing classical music plays, and during the next eight hours the self-identified religious man embarks on an inward journey occasioned by the drug: psilocybin, the active ingredient in magic mushrooms.

Thirty-six of these latter-day psy-

chonauts participated in research designed to explore the pharmacological and psychological effects of psilocybin as compared with an active control—methylphenidate hydrochloride, commonly known as Ritalin. The double-blind study revealed that psilocybin produced a self-described mystical experience in 22 of the volunteers, and 24 of them rated it as among the most meaningful of their lives, comparing it with the birth of a first child or the death of a parent. In follow-ups, "they continue to report positive changes in attitude and behavior," says study leader Roland Griffiths of Johns Hopkins University. "And the experience continues to be personally and spiritually meaningful to them."

The Griffiths work is part of a resurgence in the exploration of the psychological and physiological effects of psychedelics that seem to mimic the neurotransmitter serotonin. Charles Grob of the Harbor-U.C.L.A. Medical Center has been testing psilocybin's effect in easing the anxiety of terminally ill cancer patients. Francisco Moreno of the University of Arizona has studied its ability to alleviate the symptoms of obsessive-compulsive disorder. And studies with other hallucinogens—such as MDMA, or ecstasy—have begun exploring the potential benefits.

This resurgence actually started in 1990, when the U.S. Food and Drug Administration ended a two-decade-long moratorium by approving an investigation of the effects of DMT—dimethyltryptamine, a quick-acting, short-lived and powerful hallucinogen—in humans, according to Rick Doblin, founder of the Multidisciplinary Association for Psychedelic Studies, an advocacy group for such research. The FDA was willing to fund it because, as Grob puts it: "Sufficient time had elapsed from the time when there was so much turmoil."

Of course, reputable institutions were home to some of the disreputable research that created that turmoil; Timothy Leary rose to national prominence on the basis of his LSD experiments undertaken while a professor at Harvard

MAGIC MUSHROOMS produce psilocybin, a hallucinogen compound that has become the focus of some pharmacology research.

MARTIN BOND *Photo Researchers, Inc.*

University. And some fear a repeat: ecstasy became widely popular after some psychiatrists used it in an ostensibly therapeutic context in California and elsewhere.

Partially as a result, funding for such investigations remains a challenge. Wealthy individuals, such as computer industry millionaires Robert Jesse and the late Robert Wallace, have seeded this small flowering. Although the National Institute on Drug Abuse (NIDA) paid for the psilocybin study as part of a comparative pharmacology grant, the agency is uninterested in potential therapeutic uses. "We are open to studying all classes of illicit drugs of abuse," states David Shurtleff, director of the division of basic neuroscience behavioral research at the NIDA. "We are a long way from thinking about treatment with these compounds. There are just too many questions about how they interact with the individual."

Being illicit substances, psilocybin and its psychedelic counterparts might run into similar resistance to therapeutic indications as experienced in recent years with marijuana. But even observers from outside the small community of hallucinogen researchers feel that a qualitative difference exists between the two: "A lot of the marijuana work was exactly what was wrong with the psychedelic work in the past: there were not good, controlled studies," says Herbert Kleber, who directs the division of substance abuse at Columbia University. "One had the distinct impression that a lot of the money and effort was aimed at legitimizing the drug for recreational use."

Good, controlled studies are exact-

ly what hallucinogen researchers are striving to provide, focusing on effects in humans. "Animals will take all the drugs that humans abuse, but they do not self-inject psilocybin, LSD or any of the psychedelics," explains former NIDA director Charles Schuster of Wayne State University. "This is a uniquely human phenomenon." That makes such research fraught with peril: MDMA has shown some damaging effects in animal studies, and psilocybin may trigger schizophrenia and other mental illnesses in susceptible subjects. "Despite all of our preparation, 30 percent of our volunteers had a significant experience of fearfulness, and some of that was accompanied by paranoia," Griffiths notes. "It is easy to see how that could escalate to dangerous behavior."

But in the appropriate setting with adequate supervision, psychedelics such as psilocybin may prove effective in treating existential angst in the terminally ill or even drug addiction, by providing the type of spiritual experience that programs such as Alcoholics Anonymous rely on. Already psilocybin has shown some therapeutic potential in other health areas: researchers at McLean Hospital in Belmont, Mass., found that the compound seemed to halt cluster headache attacks. And the hallucinogenic tranquilizer ketamine proved fast and effective in alleviating depression in patients who had not responded to other treatments, according to a recent study at the National Institute of Mental Health.

Hallucinogens might also provide a window into the workings of the human brain, including spirituality. "To my knowledge, there is no class of psychoactive drug that has not found some medical utility except for hallucinogens," argues David Nichols of Purdue University, founder of the Heffter Institute, which supports some of these psychedelic studies. "Let's use them as tools to understand behavior and brain function."

"BULLET CLUSTER" formed by the collision of two galaxy clusters provides solid evidence of dark matter.

NASA/CXC/CfA/M.MARKEVITCH ET AL. (*x-ray*); NASA/STScI; MAGELLAN/UNIVERSITY OF ARIZONA/D.CLOWE ET AL. (*optical*); NASA/STScI; ESO WFI; MAGELLAN/UNIVERSITY OF ARIZONA/D.CLOWE ET AL. (*lensing map*)

ASTROPHYSICS

Dark Matter Discovery

The collision of two galaxy clusters provides the best evidence yet of dark matter, the invisible substance theorized to dominate the mass of the universe. Images taken by the Chandra X-ray Observatory show that the larger cluster dragged out the bulk of the observable mass from the smaller one, in the form of hot plasma. According to theory, the collision should have separated the plasma from dark matter. Scientists determined how parts of the collided clusters bent light behind them via gravitational lensing—the more refraction, the greater the gravity in that area. They found that the plasma did not produce the strongest lensing, indicating that some other unseen mass was doing most of the bending. The latest results, described in the September 10 *Astrophysical Journal Letters,* do not indicate what the dark matter is, but they do undermine approaches that radically rework conventional gravitational theory.

—David Biello

BIOTECH

Engineered Cells Beat Back Cancer

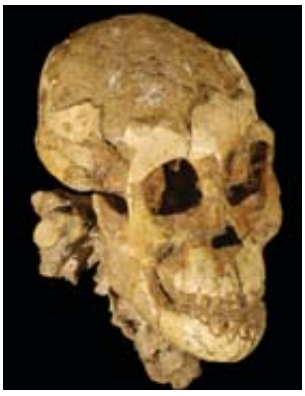
Immune cells from some patients can recognize spreading tumors and attack them. Steven Rosenberg of the National Cancer Institute and his colleagues cloned the genes governing the cancer-recognizing receptor in these immune cells—called tumor-infiltrating lymphocytes—from a patient who had successfully beaten back melanoma. They introduced this genetic information via retrovirus into regular T cells from 17 melanoma patients. After chemotherapy, the scientists infused the engineered lymphocytes back into their patients and discovered that such cells could persist, making up between 9 and 56 percent of the T cell population one month after treatment in 15 of the 17 patients. Even better, the cancer disappeared in two patients, who remain disease-free 18 months after treatment and continue to exhibit high levels of the engineered immune cells in their blood. The work was published online August 31 by *Science*.

—David Biello

Extended coverage of these News Scan briefs can be found at *www.sciam.com/ontheweb*

FOSSILS

Lucy's Baby



FAMILY ADDITION: Fossil skull of an *Australopithecus afarensis* toddler, found in Ethiopia.

Paleoanthropologists have found the fossilized remains of a baby *Australopithecus afarensis,* the ancient human ancestor most famously represented by the Lucy fossil. The 3.3-million-year-old partial skeleton turned up at a site called Dikika in Ethiopia's Afar region, a few kilometers from where Lucy herself was discovered. Largely encased in sandstone when it was found, the skeleton—believed to be that of a three-year-old female—is the earliest juvenile yet found. It preserves elements previously unknown for a species of such antiquity, including both shoulder blades and a small bone called the hyoid, which anchors the throat muscles. Analysis of the bones has revealed a number of apelike characteristics that lend tentative support to the notion that *afarensis,* though fully capable of upright walking, also spent time in the trees, where it may have sought food or evening refuge. The findings appear in the September 21 *Nature.* —*Kate Wong*

OCEANS

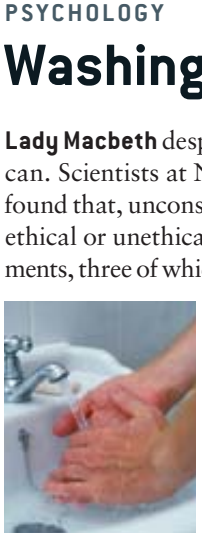
Anemic Phytoplankton

A long-standing puzzle in ocean photosynthesis was why phytoplankton failed to grow fast in parts of the Pacific Ocean; after all, the microscopic plants have access to plenty of carbon dioxide thanks to upwelling water. Lack of iron is the answer, marine scientists conclude. Photosynthesis is reduced among phytoplankton in water with poor iron concentrations as compared with those in iron-rich conditions, even though phytoplankton in both conditions make the same amount of chlorophyll. "When these little ocean plants are starved for iron, they produce more chlorophyll than they need," says lead researcher Michael Behrenfeld of Oregon State University. That way, if environmental changes cause iron levels to rise, they can then take advantage of the increase immediately. Prior studies relied on satellite images, which measure chlorophyll levels alone, so they would not have revealed this distinction. The iron deprivation means that estimates of global ocean carbon uptake are probably 2 to 4 percent too high, the group reports in the August 31 *Nature.* —*JR Minkel*

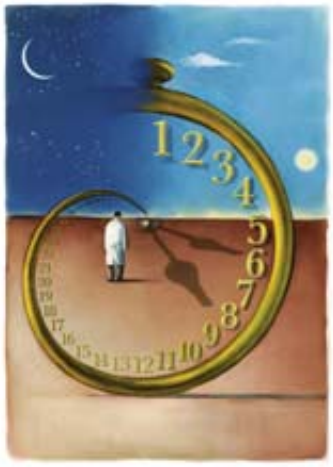
PSYCHOLOGY

Washing Hands to Remove Moral Taint

Lady Macbeth desperately attempted to wash away a spot of blood after the murder of Duncan. Scientists at Northwestern University have put the "Macbeth effect" to the test and found that, unconsciously at least, people can wash away their sins. Volunteers described an ethical or unethical action they had undertaken in their lives. Then they saw six word fragments, three of which—W_ _H, for example—could be completed in a cleansing way (WASH) or an unrelated way (WISH). Those who recollected an unethical deed were more likely to produce a cleansing word; moreover, they often chose an antiseptic wipe over a pencil when given the choice. A need to be clean might drive behavior, too. Among those who recalled dirty deeds but were not allowed to wash up, 74 percent later offered to help in another project, versus only 41 percent of those who had cleaned up, according to the September 8 *Science.* —*David Biello*



SCRUBBING UP is also psychologically cleansing.



DATA POINTS: TOO MANY HOURS

In the U.S., physicians fresh out of medical school regularly work far more hours than those in the U.K. and the European Union. Despite efforts by the Accreditation Council for Graduate Medical Education to cut hours to curtail mistakes, medical residents reported routinely exceeding the limits, based on a survey of 4,000 residents conducted between July 2002 and May 2004.

Limit of consecutive work hours for residents in:
U.K. and E.U.: 13
U.S.: 30

Limit of work hours per week in:
U.K. and E.U.: 58
U.S.: 80

Percent in the U.S. who worked more than 30 hours consecutively at least once: 67

Percent who worked more than 80 hours per week: 43

Percent of monthly work logs that showed some violation: 44

SOURCES: Journal of the American Medical Association, *September 6; British Medical Association; "Medical Residents Overworked despite New Work Rules" on www.sciam.com (September 5).*

BY THE NUMBERS

Myth: Red-Blue States

AMERICAN SPLIT IN "CULTURE WAR" UNFOUNDED BY RODGER DOYLE

Beginning in the 2000 presidential election, the broadcast networks began showing Republican-won states in red and Democratic-won states in blue. Soon pundits were talking about the red states and the blue states as if they were two different countries, one dominated by evangelicals and the other by secularists. The nation, some claimed, was engaged in a culture war.

The 2004 presidential election map followed much the same pattern, reinforcing the notion. But the split is not an intrinsic feature of U.S. politics, as suggested by the map that shows marginal states in white. The 12 white states accounted for 25 percent of the electoral vote. This map shows that the election geography could change substantially with a relatively small shift in votes.

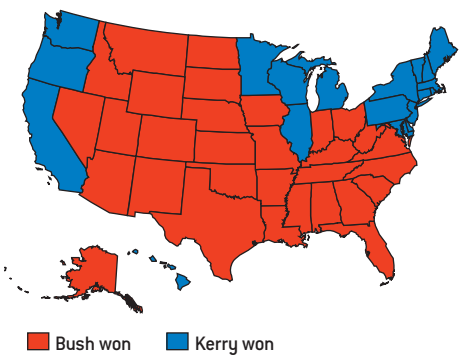
Survey data show that the two sides are closer together than is usually thought. On some economic issues, Americans agree. For example, almost two thirds in red and blue states think that big corporations have too much power. Furthermore, in both sections a substantial majority want to protect the environment. On "gut" issues, red and blue are not that far apart either, with survey differences running up to 12 percent on gun control, school vouchers, homosexual adoption and the death penalty. On the most contentious issue—abortion—about 20 percent believe it should be legal under all circumstances, and about 20 percent believe it should be illegal under all circumstances. For the rest, abortion may be important, but it apparently does not rise to the level of a fighting issue.

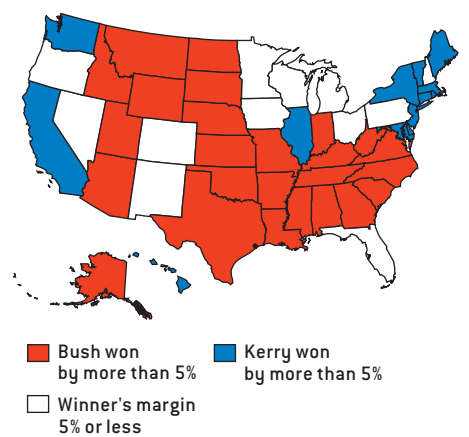
Harvard University economists Edward L. Glaeser and Bryce A. Ward have extensively analyzed the historical data on voting and cultural differences. They find little evidence that the red states are becoming more Republican or the blue states more Democratic, and they have determined that the two parties are no more spatially segregated now than in the past. They see no signs of a culture war, noting that most people are in the middle on most issues. They do note, however, that the U.S. is in a strongly partisan period because of opposition to President George W. Bush and that the alliance of the Republican party with evangelicals is fairly new. That alliance, however, is not necessarily a sign of increased divisiveness, considering that religious groups historically can swing from one party to another—evangelicals used to be more closely allied with Democrats.

As Glaeser and Ward note, the U.S. is remarkable for the geographic diversity of habits and beliefs. Those, for example, who believe in the Rapture do not live exclusively in Kansas: they live all over. About 45 percent of voters in red states are more liberal on economic issues than the typical blue state voter, whereas about 40 percent of blue staters are more conservative on cultural issues than the typical red stater.

Rodger Doyle can be reached at rodgerpdoyle@verizon.net

VOTE IN 2004 PRESIDENTIAL ELECTION





RODGER DOYLE; SOURCE: CLERK OF THE HOUSE OF REPRESENTATIVES

FAST FACTS: SWING VOTES

Percent of people who voted for George W. Bush or John Kerry in the 2004 presidential election:

	Bush	Kerry
Bush's Best States		
Utah	71.5	26.0
Wyoming	68.9	29.1
Idaho	68.4	30.3
Closest States		
Iowa	49.9	49.2
New Mexico	49.8	49.1
Wisconsin	49.3	49.7
Kerry's Best States		
Massachusetts	36.8	61.9
Rhode Island	38.7	59.4
Vermont	38.8	58.9

FURTHER READING

Culture War? The Myth of a Polarized America. Morris P. Fiorina, with Samuel J. Abrams and Jeremy C. Pope. Pearson Longman, 2004.

Myths and Realities of American Political Geography. Edward L. Glaeser and Bryce A. Ward in *Journal of Economic Perspectives*, Vol. 20, No. 2, pages 119–144; Spring 2006.

Purple America. Stephen Ansolabehere, Jonathan Rodden and James M. Snyder, Jr., in *Journal of Economic Perspectives*, Vol. 20, No. 2, pages 97–118; Spring 2006.

Wronger Than Wrong

Not all wrong theories are equal

By MICHAEL SHERMER

In belles lettres the witty literary slight has evolved into a genre because, as 20th-century trial lawyer Louis Nizer noted, "A graceful taunt is worth a thousand insults." To wit, from high culture, Mark Twain: "I didn't attend the funeral, but I sent a nice letter saying I approved of it." Winston Churchill: "He has all the virtues I dislike and none of the vices I admire." And from pop culture, Groucho Marx: "I've had a perfectly wonderful evening. But this wasn't it." Scientists are no slouches when it comes to pitching invectives at colleagues. Achieving almost canonical status as the ne plus ultra put-down is theoretical physicist Wolfgang Pauli's reported harsh critique of a paper: "This isn't right. It's not even wrong." I call this Pauli's proverb.

Scientists' wrongness attenuates with time.

Columbia University mathematician Peter Woit recently employed Pauli's proverb in his book title, a critique of string theory called *Not Even Wrong* (Basic Books, 2006). String theory, Woit argues, is not only based on nontestable hypotheses, it depends far too much on the aesthetic nature of its mathematics and the eminence of its proponents. In science, if an idea is not falsifiable, it is not that it is wrong, it is that we cannot determine if it is wrong, and thus it is not even wrong.

Not even wrong. What could be worse? Being wronger than wrong, or what I call Asimov's axiom, well stated in his book *The Relativity of Wrong* (Doubleday, 1988): "When people thought the earth was flat, they were wrong. When people thought the earth was spherical, they were wrong. But if you think that thinking the earth is spherical is just as wrong as thinking the earth is flat, then your view is wronger than both of them put together."

Asimov's axiom holds that science is cumulative and progressive, building on the mistakes of the past, and that even though scientists are often wrong, their wrongness attenuates with continued data collection and theory building. Satellite measurements, for instance, have shown precisely how the earth's shape differs from a perfect sphere.

The view that all wrong theories are equal implies that no theory is better than any other. This is the theory of the "strong" social construction of science, which holds that science is inextricably bound to the social, political, economic, religious and ideological predilections of a culture, particularly of those individuals in power. Scientists are knowledge capitalists who produce scientific papers that report the results of experiments conducted to test (and usually support) the hegemonic theories that reinforce the status quo.

In some extreme cases, this theory that culture shapes the way science is conducted is right. In the mid-19th century, physicians discovered that slaves suffered from *drapetomania,* or the uncontrollable urge to escape from slavery, and *dysaethesia aethiopica,* or the tendency to be disobedient. In the late 19th and early 20th centuries, scientific measurements of racial differences in cognitive abilities found that blacks were inferior to whites. In the mid-20th century, psychiatrists discovered evidence that allowed them to classify homosexuality as a disease. And until recently, women were considered inherently inferior in science classrooms and corporate boardrooms.

Such egregious examples, however, do not negate the extraordinary ability of science to elucidate the natural and social worlds. Reality exists, and science is the best tool yet employed to discover and describe that reality. The theory of evolution, even though it is the subject of vigorous debates about the tempo and mode of life's history, is vastly superior to the theory of creation, which is not even wrong (in Pauli's sense). As evolutionary biologist Richard Dawkins observed on this dispute: "When two opposite points of view are expressed with equal intensity, the truth does not necessarily lie exactly halfway between them. It is possible for one side to be simply wrong."

Simply wrong. When people thought that science was unbiased and unbound by culture, they were simply wrong. On the other hand, when people thought that science was completely socially constructed, they were simply wrong. But if you believe that thinking science is unbiased is just as wrong as thinking that science is socially constructed, then your view is not even wronger than wrong. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com). His new book is* Why Darwin Matters.

BRAD HINES

Welfare States, beyond Ideology

Are higher taxes and strong social "safety nets" antagonistic to a prosperous market economy? The evidence is now in By JEFFREY D. SACHS

One of the great challenges of sustainable development is to combine society's desires for economic prosperity and social security. For decades economists and politicians have debated how to reconcile the undoubted power of markets with the reassuring protections of social insurance. America's supply-siders claim that the best way to achieve well-being for America's poor is by spurring rapid economic growth and that the higher taxes needed to fund high levels of social insurance would cripple prosperity. Austrian-born free-market economist Friedrich August von Hayek suggested in the 1940s that high taxation would be a "road to serfdom," a threat to freedom itself.

Most of the debate in the U.S. is clouded by vested interests and by ideology. Yet there is by now a rich empirical record to judge these issues scientifically. The evidence may be found by comparing a group of relatively free-market economies that have low to moderate rates of taxation and social outlays with a group of social-welfare states that have high rates of taxation and social outlays.

Friedrich von Hayek was wrong.

Not coincidentally, the low-tax, high-income countries are mostly English-speaking ones that share a direct historical lineage with 19th-century Britain and its theories of economic laissez-faire. These countries include Australia, Canada, Ireland, New Zealand, the U.K. and the U.S. The high-tax, high-income states are the Nordic social democracies, notably Denmark, Finland, Norway and Sweden, which have been governed by left-of-center social democratic parties for much or all of the post–World War II era. They combine a healthy respect for market forces with a strong commitment to anti-poverty programs. Budgetary outlays for social purposes average around 27 percent of gross domestic product (GDP) in the Nordic countries and just 17 percent of GDP in the English-speaking countries.

On average, the Nordic countries outperform the Anglo-Saxon ones on most measures of economic performance. Poverty rates are much lower there, and national income per working-age population is on average higher. Unemployment rates are roughly the same in both groups, just slightly higher in the Nordic countries. The budget situation is stronger in the Nordic group, with larger surpluses as a share of GDP.

The Nordic countries maintain their dynamism despite high taxation in several ways. Most important, they spend lavishly on research and development and higher education. All of them, but especially Sweden and Finland, have taken to the sweeping revolution in information and communications technology and leveraged it to gain global competitiveness. Sweden now spends nearly 4 percent of GDP on R&D, the highest ratio in the world today. On average, the Nordic nations spend 3 percent of GDP on R&D, compared with around 2 percent in the English-speaking nations.

The Nordic states have also worked to keep social expenditures compatible with an open, competitive, market-based economic system. Tax rates on capital are relatively low. Labor market policies pay low-skilled and otherwise difficult-to-employ individuals to work in the service sector, in key quality-of-life areas such as child care, health, and support for the elderly and disabled.

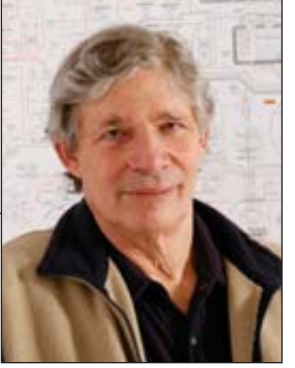
The results for the households at the bottom of the income distribution are astoundingly good, especially in contrast to the mean-spirited neglect that now passes for American social policy. The U.S. spends less than almost all rich countries on social services for the poor and disabled, and it gets what it pays for: the highest poverty rate among the rich countries and an exploding prison population. Actually, by shunning public spending on health, the U.S. gets much less than it pays for, because its dependence on private health care has led to a ramshackle system that yields mediocre results at very high costs.

Von Hayek was wrong. In strong and vibrant democracies, a generous social-welfare state is not a road to serfdom but rather to fairness, economic equality and international competitiveness. SA

An expanded version of this essay is available online at www.sciam.com/ontheweb

Jeffrey D. Sachs is director of the Earth Institute at Columbia University.

BRUCE GILBERT/EARTH INSTITUTE



The Evolution of Future Wealth

Technologies evolve much as species do, and that underappreciated fact is the key to growth By STUART A. KAUFFMAN

When the world changes unpredictably over the course of centuries, no one is shocked: Who blames the Roman centurions for not foreseeing the invention of rocket launchers? Yet monumental and surprising transformations occur on much shorter timescales, too. Even in the early 1980s you would have been hard-pressed to find people confidently predicting the rise of the Internet or the fall of the U.S.S.R. Unexpected change bedevils the business community endlessly, despite all best efforts to anticipate and adapt to it—witness the frequent failure of companies' five-year plans.

Economists have so far not been able to offer much help to firms trying to be more adaptive. Although economists have been slow to realize it, the problem is that their attempts to model economic systems focus on those in market equilibrium or moving toward it. They have drawn their inspiration predominantly from the work of physicists in this respect (often with good results, of course). For instance, the Black-Scholes model used since the 1970s to predict the volatility of stock prices was developed by trained physicists and is related to the thermodynamic equation that describes heat.

Economics should shift its attention from physics to biology.

As economics attempts to model increasingly complicated phenomena, however, it would do well to shift its attention from physics to biology, because the biosphere and the living things in it represent the most complex systems known in nature. In particular, a deeper understanding of how species adapt and evolve may bring profound—even revolutionary—insights into business adaptability and the engines of economic growth.

One of the key ideas in modern evolutionary theory is that of preadaptation. The term may sound oxymoronic but its significance is perfectly logical: every feature of an organism, in addition to its obvious functional characteristics, has others that could become useful in totally novel ways under the right circumstances. The forerunners of air-breathing lungs, for example, were swim bladders with which fish maintained their equilibrium; as some fish began to move onto the margins of land, those bladders acquired a new utility as reservoirs of oxygen. Biologists say that those bladders were preadapted to become lungs. Evolution can innovate in ways that cannot be prestated and is nonalgorithmic by drafting and recombining existing entities for new purposes—shifting them from their existing function to some adjacent novel function—rather than inventing features from scratch.

A species' suite of adaptive features defines its ecological niche through its relations to other species. In the same way, every economic good occupies a niche defined by its relations to complementary and substitute goods. As the number of economic goods increases, the number of ways in which to adaptively combine those goods takes off exponentially, forging possibilities for all-new niches. The autocatalytic creation of niches is thus a main driver of economic growth.

We do not yet know what makes some systems more adaptable than others, but research on complexity has yielded some clues. Some of my own work on physical systems called spin glasses suggests that the level of central control over subsidiary parts of a system is an important consideration. Too much control freezes the system into limited configurations; too little causes it to wander aimlessly. Only systems that hover on the border between order and chaos exhibit the needed general stability and capacity to explore the universe of possible solutions to challenges.

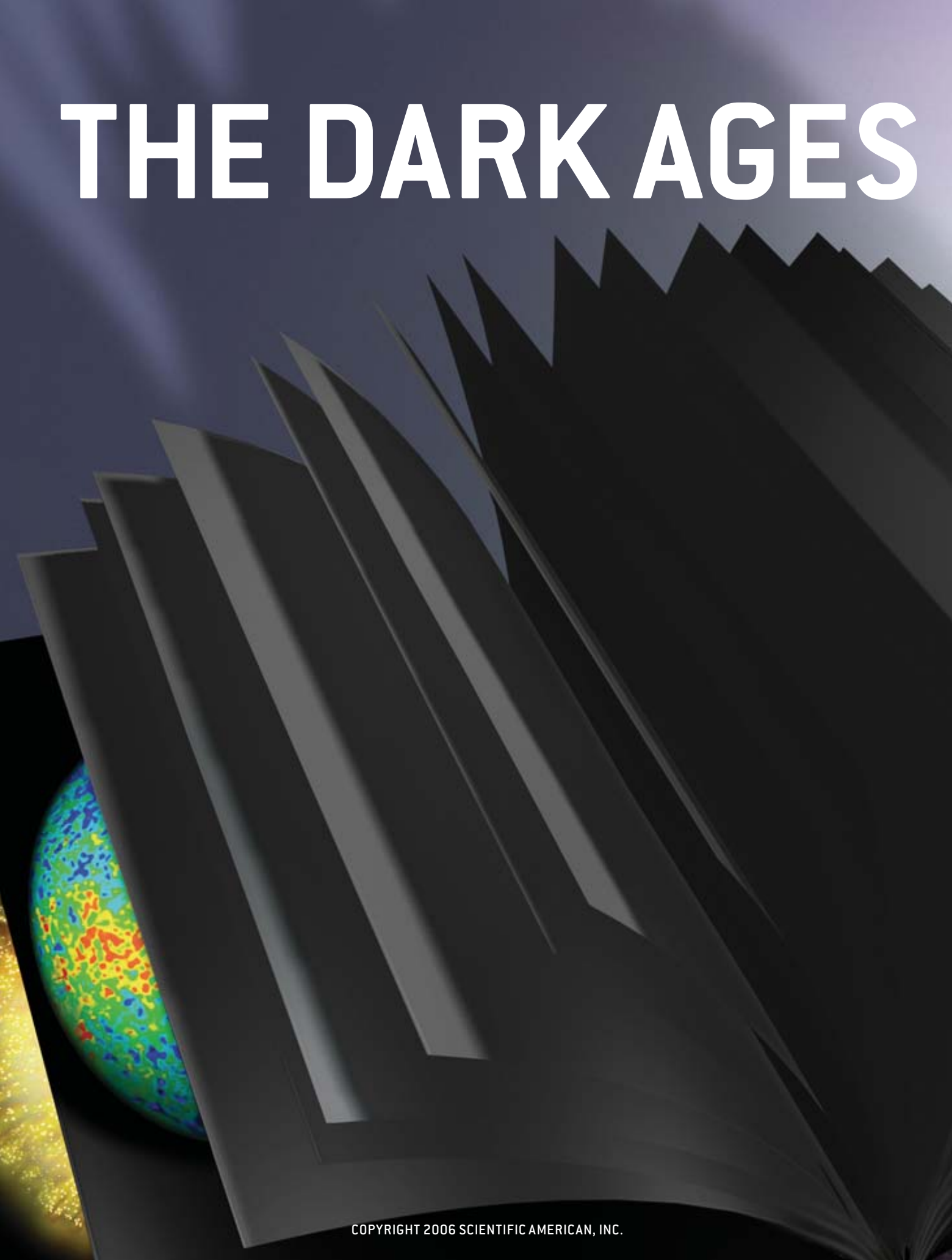
The path to maximum prosperity will depend on finding ways to build economic systems in which new niches will generate spontaneously and abundantly. Such an approach to economics is indeed radical. It is based on the emergent behavior of systems rather than on the reductive study of them. It defies conventional mathematical treatments because it is not prestatable and is nonalgorithmic. Not surprisingly, most economists have so far resisted these ideas. Yet there can be little doubt that learning to apply these lessons from biology to technology will usher in a remarkable era of innovation and growth. SA

Stuart A. Kauffman is professor of biocomplexity and informatics at the University of Calgary and external professor at the Santa Fe Institute.

JASON DZIVER/ALBERTA INFORMATICS CIRCLE OF RESEARCH EXCELLENCE (iCORE), © iCORE

Page Intentionally Blank

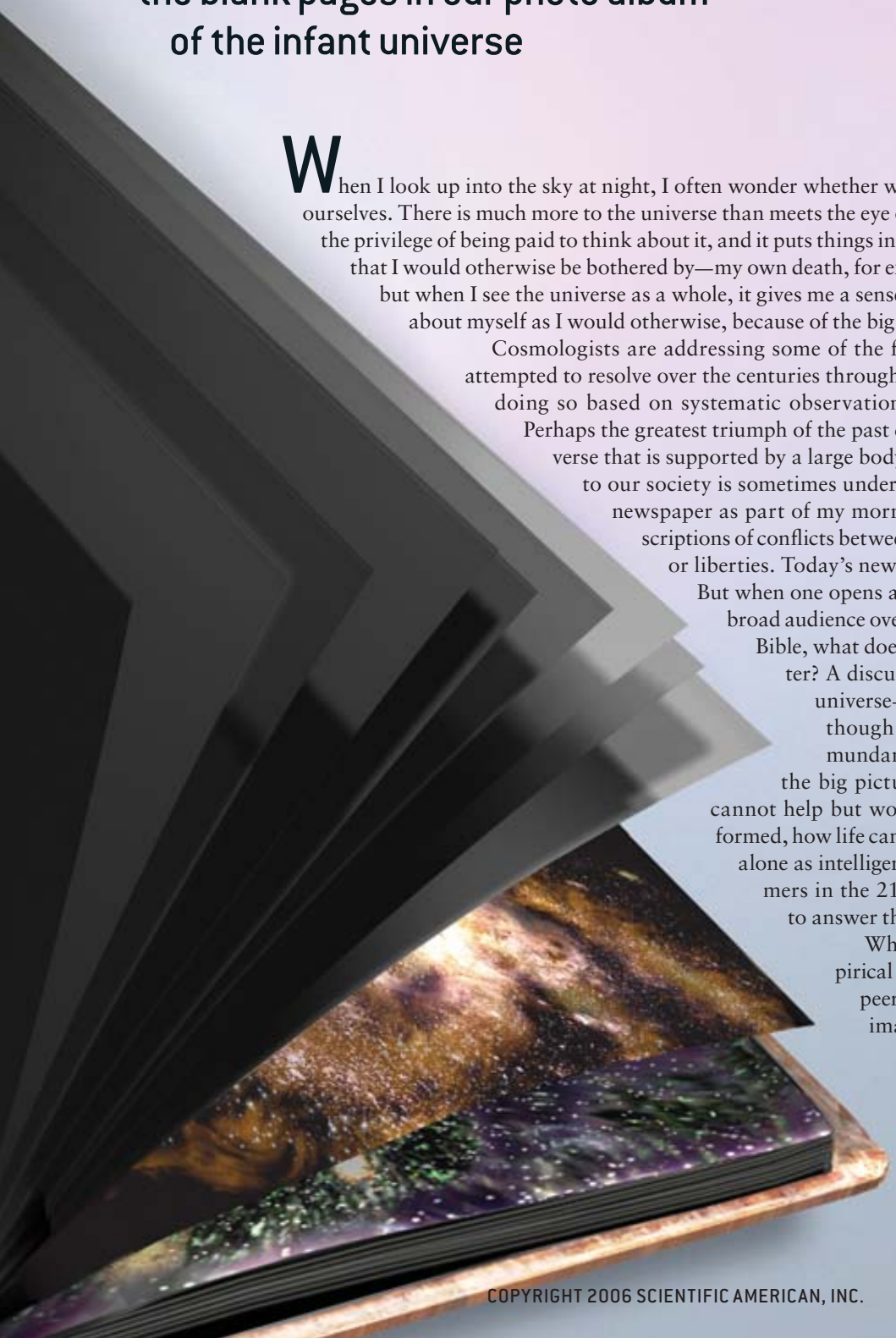
THE DARK AGES



of the Universe

Astronomers are trying to fill in the blank pages in our photo album of the infant universe

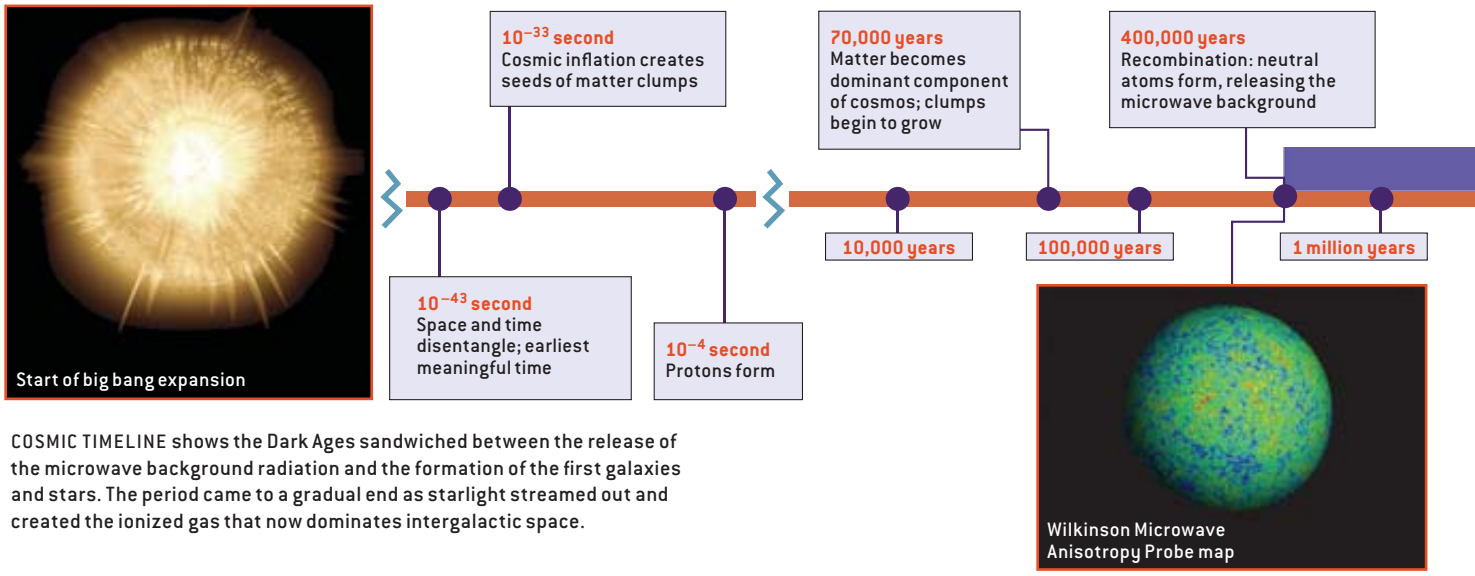
By Abraham Loeb



When I look up into the sky at night, I often wonder whether we humans are too preoccupied with ourselves. There is much more to the universe than meets the eye on earth. As an astrophysicist I have the privilege of being paid to think about it, and it puts things in perspective for me. There are things that I would otherwise be bothered by—my own death, for example. Everyone will die sometime, but when I see the universe as a whole, it gives me a sense of longevity. I do not care so much about myself as I would otherwise, because of the big picture.

Cosmologists are addressing some of the fundamental questions that people attempted to resolve over the centuries through philosophical thinking, but we are doing so based on systematic observation and a quantitative methodology. Perhaps the greatest triumph of the past century has been a model of the universe that is supported by a large body of data. The value of such a model to our society is sometimes underappreciated. When I open the daily newspaper as part of my morning routine, I often see lengthy descriptions of conflicts between people about borders, possessions or liberties. Today's news is often forgotten a few days later. But when one opens ancient texts that have appealed to a broad audience over a longer period of time, such as the Bible, what does one often find in the opening chapter? A discussion of how the constituents of the universe—light, stars, life—were created. Although humans are often caught up with mundane problems, they are curious about the big picture. As citizens of the universe we cannot help but wonder how the first sources of light formed, how life came into existence and whether we are alone as intelligent beings in this vast space. Astronomers in the 21st century are uniquely positioned to answer these big questions.

What makes modern cosmology an empirical science is that we are literally able to peer into the past. When you look at your image reflected off a mirror one meter



COSMIC TIMELINE shows the Dark Ages sandwiched between the release of the microwave background radiation and the formation of the first galaxies and stars. The period came to a gradual end as starlight streamed out and created the ionized gas that now dominates intergalactic space.

away, you see the way you looked six nanoseconds ago—the light's travel time to the mirror and back. Similarly, cosmologists do not need to guess how the universe evolved; we can watch its history through telescopes. Because the universe appears to be statistically identical in every direction, what we see billions of light-years away is probably a fair representation of what our own patch of space looked like billions of years ago.

The ultimate goal of observational cosmology is to capture the entire history of the universe, providing a seamless picture of our descent from a shapeless gas of subatomic particles. We have a snapshot of the universe as it was 400,000 years after the big bang—the cosmic microwave background radiation—as well as pictures of individual galaxies a billion years later. By the middle of the next decade, NASA plans to launch a new space telescope, named the James Webb Space Telescope (JWST), that should be able to pick up the first galaxies, which theorists predict formed at a cosmic age of hundreds of millions of years.

But that still leaves a tremendous gap. In between the release of the microwave background and the first rays of starlight was a period when the universe was dark and the microwave background no longer traced the distribution of matter. It might sound like a languid, gloomy time, a boring interlude between the immediate aftermath of the big bang and the bustling cosmos of the present day. Yet a great deal happened in these Dark Ages: the primordial soup evolved into the rich zoo of celestial bodies we now see. Within the inky blackness, gravitational forces were assembling objects in the cosmos.

The situation that astronomers face is similar to having a photo album of a person that contains the first ultrasound image of him or her as an unborn baby and some additional photos as a teenager and an adult. If you tried to guess from these pictures what happened in the interim, you could be seriously wrong. A child is not simply a scaled-up fetus or scaled-down adult. The same is true with galaxies. They did not follow a straightforward path of development from the incipient matter clumping evident in the microwave background. Observations hint that the universe underwent a wrenching transition during the Dark Ages.

Astronomers are currently searching for the missing pages of the cosmic photo album, which will show how the universe evolved during its infancy and made the building blocks of galaxies like our own Milky Way. A decade ago, when I started to work on this effort, only a handful of researchers were interested in it. Now it motivates a major fraction of future observational projects and promises to be one of the most exciting frontiers in cosmology over the next decade.

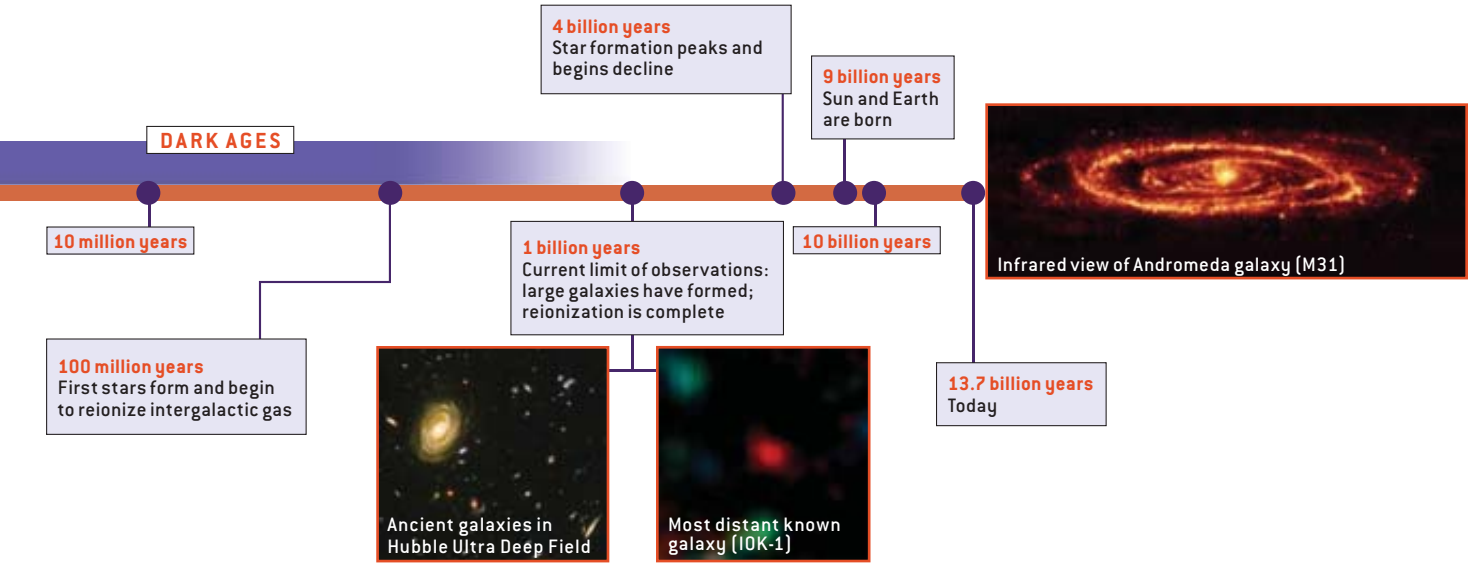
Overview/*Epoch of Reionization*

- Much of the attention in cosmology over the past several years has focused on the cosmic microwave background radiation, which provides a snapshot of the universe at an age of 400,000 years. But between this moment and the appearance of the first galaxies was a period of almost total darkness, broken by not so much as a glimmer of starlight. Hidden in the shadows of this era are the secrets of how galaxies took shape.
- Clearly, it is hard to probe a period that is by its very nature practically invisible. The key is to look for the feeble radio waves emitted by electrically neutral hydrogen gas as it interacts with the background radiation. Observers are now starting to do so.
- The result should be an even more interesting map than that of the microwave background. It will be fully three-dimensional and will show, step by step, how form emerged from formlessness.

From Ions to Ions

ACCORDING TO the big bang theory, the early universe was filled with hot plasma—a cauldron of protons, electrons and photons, with a smattering of other particles. The freely moving electrons interacted with photons through a process known as Thomson scattering, which coupled matter and

JEAN-FRANCOIS PODEVIN (*preceding pages and start of big bang, this page*); MAX TEGMARK *Massachusetts Institute of Technology* (*WMAP*)



radiation tightly together. As the universe expanded in size, it cooled, and when temperatures fell to 3,000 kelvins, the protons and electrons combined to make electrically neutral hydrogen atoms. The Thomson scattering process ended, and the photons ceased interacting with matter so intensively, becoming the microwave background. Cosmic expansion continued to cool the gas, so one might expect that the cosmic gas would still be cold and neutral today.

Surprisingly, it is not. Although the world around us is composed of atoms, the bulk of the universe's ordinary matter today is in the form of plasma, located deep in intergalactic space. The observed spectra of the most distant (and hence oldest) known quasars, galaxies and gamma-ray bursts indicate that this diffuse cosmic hydrogen was fully ionized by a cosmic age of one billion years [see "The Emptiest Places," by Evan Scannapieco, Patrick Petitjean and Tom Broadhurst; Scientific American, October 2002]. A tantalizing hint of what happened came three years ago, when the Wilkinson Microwave Anisotropy Probe (WMAP) confirmed that the microwave background radiation is slightly polarized. Neutral hydrogen does not polarize this radiation; only ionized hydrogen does. The amount of polarization suggests that the gas was ionized as early as a few hundred million years after the big bang. Thus, the atoms must have been broken back down into their constituent protons and electrons as the Dark Ages came to an end.

Most researchers associate this process of reionization with the first generation of stars. Ionizing an atom of hydrogen takes an energy of 13.6 electron volts, an amount delivered by a photon of ultraviolet light. It is not a great deal of energy—equivalent to about 10^9 joules per kilogram of hydrogen, much less than the 10^{15} joules released by the nuclear fusion of the same amount of hydrogen. If just one millionth of the gas in the universe underwent fusion inside stars, it would have produced enough energy to ionize all the rest. Other researchers conjecture that material plummeting into black holes gave off the ionizing radiation. Falling into a black hole releases as much as 10^{16} joules per kilogram, so only a 10-millionth of the cosmic hydrogen would need to fall into black holes to ionize the rest.

Stars and black holes arise within galaxies, so before reionization could take place, galaxies must have formed. Although most people think of galaxies as collections of stars, cosmologists regard them simply as large clumps of matter in which stars are relative latecomers. In fact, galaxies consist mostly of dark matter, an as yet unidentified type of material that is inherently invisible. Galaxies are thought to have formed when a region of the universe that started slightly denser than average pulled itself together by its own gravity. Although the region initially expanded like the rest of the universe, its extra gravity slowed its expansion down, turned it around and made the region collapse on itself to create a bound object: a galaxy.

According to current models, dwarf galaxies started to take shape when the universe was 100 million years old. They merged and built up bigger galaxies as time went on. A modern galaxy such as the Milky Way involved the coalescence of a million such building blocks. Within the embryonic galaxies, gas cooled and fragmented to create stars [see "The First Stars in the Universe," by Richard B. Larson and Volker Bromm; Scientific American, December 2001]. The stars' ultraviolet radiation leaked into intergalactic space, broke electrons out from their atoms and created an expanding bubble of ionized gas. Ever more bubbles appeared as new galaxies took root, and the intergalactic gas looked like Swiss cheese. The

The universe underwent a WRENCHING TRANSITION during the Dark Ages.

NASA, EUROPEAN SPACE AGENCY, STEVEN BECKWITH *Space Telescope Science Institute* AND HUBBLE ULTRA DEEP FIELD TEAM (*ancient galaxies*); M. IYE AND K. OTA, © SUBARU TELESCOPE, NATIONAL ASTRONOMICAL OBSERVATORY OF JAPAN (*IOK-1*); NASA/JPL (*M31*)

bubbles started to overlap and eventually filled all of space.

Although the above sequence of events sounds plausible, its only residence so far has been in the minds of theorists. Practical cosmologists would like to see direct evidence for the reionization epoch before adding the missing chapter to their textbooks. Moreover, only observations can settle whether stars or black holes dominated reionization and what the properties of the dark matter were. But how are such observations possible if, initially at least, the Dark Ages were dark?

Seeing in the Dark

FORTUNATELY, EVEN COLD HYDROGEN can emit a form of light. Subatomic particles have an intrinsic orientation known as spin, which can point in one of two directions, conventionally called "up" and "down." The electron and proton in a hydrogen atom can point either in the same direction (aligned) or in opposite directions (antialigned). In the antialigned state the atom has a lower energy. If, for example, both the electron and proton point up, and the electron then flips so that it points down, the atomic state will decrease in energy and give off a photon with a wavelength of 21 centimeters. Conversely, if the atom absorbs a photon of this wavelength, the electron will flip up.

A 21-centimeter photon is much less energetic than the photons typically emitted by hydrogen as electrons jump between orbits. For this reason, the spin-flipping process was able to operate even when no stars yet shone. Energy from the microwave background radiation and from collisions among the atoms would have sufficed to flip electrons and induce the hydrogen to glow feebly. The relative numbers of atoms with aligned and antialigned spins defined the so-called spin temperature of the gas. A high spin temperature, for example, would indicate that a high fraction of atoms were aligned.

Theory therefore indicates that the Dark Ages were defined by three distinct temperatures: the spin temperature (a measure of the relative abundance of atoms with different spin states); the ordinary, kinetic temperature (a measure of the motions of the atoms); and the radiation temperature (a measure of the energy of the background photons). These three temperatures could deviate from one another, depending on which physical processes operated.

In a strange ménage à trois, the spin temperature matched first the kinetic temperature, then the radiation temperature and finally the kinetic temperature once again [*see box on opposite page*]. As space expanded, both the gas and the radiation cooled. Left to its own devices, the gas would have cooled faster, but initially a small residual number of free electrons left over from the formation of hydrogen atoms counteracted this tendency. These electrons acted as middlemen to convey energy from the microwave background to the atoms, keeping all three temperatures equal. Ten million years after the big bang, however, the electrons faltered in their role because the microwave background had become too dilute. The equilibrium between gas and radiation broke down, and the gas started to cool rapidly. Atomic collisions kept the kinetic and spin temperatures equal. In this phase the hydrogen was a net absorber of 21-centimeter photons and soaked up energy from the microwave background (though never enough to restore equilibrium).

One hundred million years after the big bang, a second transition occurred. Cosmic expansion had diluted the density of the gas to the point where collisions were too infrequent to equalize the spin and kinetic temperatures. The spins then picked up energy from the microwave background. When the spin temperature returned to equilibrium with the radiation temperature, hydrogen was neither a net absorber nor a net emitter of 21-centimeter photons. During this period the gas could not be seen against the microwave background.

When the first stars and black holes lit up, a third transition took place. The x-rays they gave off raised the kinetic temperature. Their ultraviolet light was absorbed and reradiated by the hydrogen, and the ensuing hopscotching of electrons among atomic orbits brought the spin and kinetic temperatures into equilibrium. The spin temperature increased beyond the microwave background temperature, so the hydrogen outshone the background. Flipping electrons takes much less energy than ionizing atoms, so galaxies caused the hydrogen to glow well before they reionized it. Eventually, as the hydrogen became ionized, it gave off light by different means and the intergalactic 21-centimeter emission faded away.

The new map may carry MORE INFORMATION than even the cosmic microwave background radiation.

Primeval Tomography

BECAUSE OF THIS MÉNAGE À TROIS, the 21-centimeter sky will be either brighter or darker than the microwave background, depending on time and location. Another phenomenon that observers need to take into account is that cosmic expansion has stretched the photons to longer wavelengths.

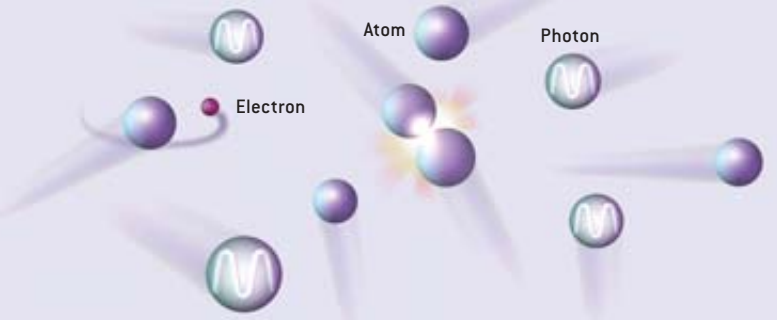
THE AUTHOR

ABRAHAM LOEB is a world leader in the theoretical study of the first stars and black holes and the epoch of reionization. What drives him, he says, is an interest in ancient philosophical questions; these inspired him to enter physics as a young person. He is now an astronomy professor at Harvard University and a visiting professor at the Weizmann Institute of Science in Rehovot, Israel. Loeb has also been a pioneer of the detection of extrasolar planets by gravitational microlensing and the production of gamma rays in intergalactic space. He served on the first science working group for the James Webb Space Telescope and received a Guggenheim Fellowship in 2002.

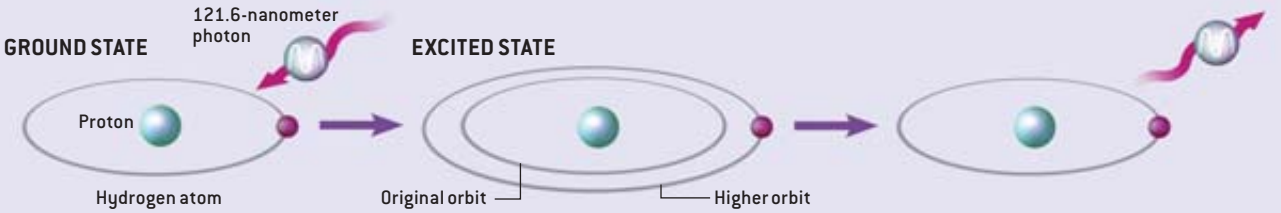
HOW TO SEE IN THE DARK

Despite the lack of stars, the Dark Ages were not completely dark. A rare process caused hydrogen gas to glow dimly.

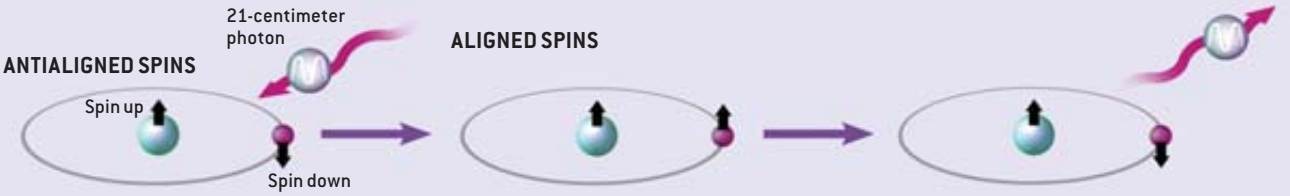
For hydrogen to glow, there had to be a source of energy. The only available ones were the atoms' own kinetic energy (released by collisions between atoms) and the photons of the cosmic background radiation. A smattering of unattached electrons was available to help transfer energy between the atoms and the photons.



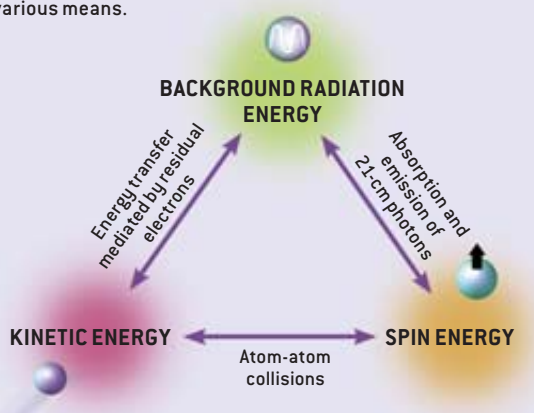
Neither source, however, was strong enough to cause the hydrogen to glow by the usual means, in which an electron gets bumped into a higher orbit (a so-called excited state) and falls back down, releasing a photon.



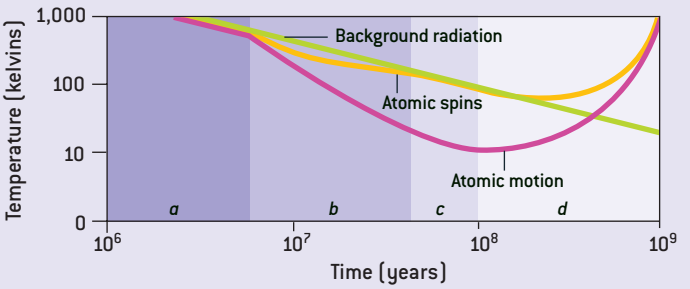
The collisions and photons did, however, pack just enough punch to flip an electron so that its spin pointed the same way as the proton's. When the electron flipped back, it released a photon with a wavelength of 21 centimeters.



The kinetic energy, photon energy and spin energy were three reservoirs that interchanged energy by various means.



The amount of energy in each reservoir can be represented in terms of temperature: the higher the temperature, the greater the energy. At the start of the Dark Ages, all three temperatures were the same (*a*). Then the kinetic and spin temperature began to fall faster than the photon energy (*b*). After a while, the spin temperature returned to equilibrium with the photon temperature (*c*). Finally, stars and quasars warmed the gas, pumping up the kinetic and spin temperatures (*d*). The relative temperatures determine how (and whether) the hydrogen can be observed.



GEORGE RETSECK (*illustrations*); DANIELA NAOMI MOLNAR (*graph*)

Since the start of the Dark Ages, the universe has expanded in size by a factor of 1,000, so a 21-centimeter photon emitted at that time arrives on earth with a wavelength of 210 meters. A photon emitted toward the end of the Dark Ages is shifted to a wavelength of one to two meters.

This range of wavelengths falls into the radio part of the electromagnetic spectrum. The emission can be picked up by arrays of low-frequency antennas similar to those used for television and radio communications. Several groups are currently constructing such arrays. The Mileura Widefield Array (MWA) in Western Australia will consist of 8,000 antennas scattered across a region 1.5 kilometers long and sensitive to a wavelength of one to 3.7 meters. It has an angular resolution of a few arcminutes, which corresponds to a physical scale of about three million light-years during the Dark Ages. Other efforts include the Low-Frequency Array (LOFAR), the Pri-

meval Structure Telescope (PaST) and, in the more distant future, the Square Kilometer Array (SKA).

By scanning across wavelengths, these arrays will probe the 21-centimeter emission at different times in cosmic history. Astronomers will be able to build a three-dimensional map of the neutral hydrogen distribution. They will be able to watch density fluctuations of one part in 100,000 (as in the microwave background) become orders of magnitude greater. At the locations of greatest density, galaxies should take shape and create bubbles of ionized hydrogen. The bubbles will proliferate and merge, eventually clearing intergalactic space of neutral hydrogen [*see box at right*]. The sharpness of the bubbles' boundaries will answer the question of whether reionization was caused by massive stars or by black holes. Massive stars pour out most of their energy in ultraviolet light, which is readily blocked by intergalactic hydrogen, whereas black holes generate mostly x-rays, which penetrate deeply into the gas. So black holes produce fuzzier boundaries.

For several reasons, the 21-centimeter map may carry more bits of information than any other survey in cosmology—more than even the cosmic microwave background. First, whereas an image of the microwave background is two-dimensional, because it originated at a single moment in time (when the universe cooled below 3,000 kelvins), the 21-centimeter map, as mentioned above, will be fully three-dimensional. Second, the microwave background is somewhat blurry because its release did not occur at the same time everywhere. The universe went through a period when it was neither fully opaque nor fully transparent, like a fog that dissipated gradually. During that time, the radiation diffused across short distance scales, smearing the fine print in the microwave background sky. In contrast, when 21-centimeter radiation emerged from hydrogen atoms, nothing blocked its propagation through space, so it traces the gas distribution without such blurring. Third, the microwave background carries information about the matter density fluctuations that seeded galaxies, whereas the 21-centimeter map will depict both the seeds of galaxies and the effect that the galaxies, once formed, had on their surroundings.

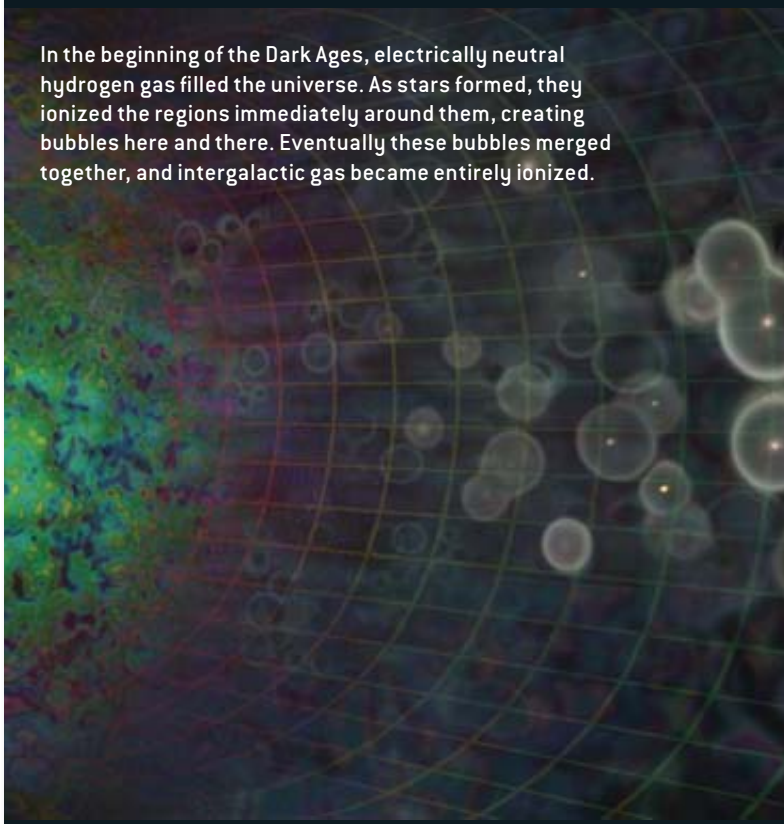
To detect the 21-centimeter signal, observers will have to overcome numerous challenges. Low-frequency radio broadcasts on earth have to be filtered out. Even more difficult will be dealing with foreground radio emission from our galaxy, which is 10,000 times more intense than the signal from the epoch of reionization. Fortunately, the galactic noise is roughly the same at slightly different wavelengths, whereas the signal fluctuates with wavelength, reflecting the spatial structure of the ionized bubbles. This difference makes it possible to extract the signal. Astronomers should be able to compare the 21-centimeter maps with images from instruments such as JWST. The galaxies seen in infrared light should correlate with ionized bubbles in the neutral hydrogen.

In addition to the above observational challenges, a number of tasks remain for theorists. Most important, they need to run bigger computer simulations to track events in a volume of space large enough (a billion light-years across) to be a representative statistical sample of our universe and with high enough resolution to capture dwarf galaxies. The simulation also needs to trace the propagation of the ionizing radiation from the galaxies through the surrounding gas, a process modeled only very crudely so far. Observers may well see reionization before theorists are able to forecast what they should see.

This combined observational and theoretical effort should shed light on various mysteries that now plague the theory of galaxy formation. One set of questions concerns the massive black holes in the centers of galaxies. Over the past decade astronomers have realized that almost every galaxy in the present-day universe, including our own Milky Way, hosts a massive black hole. These holes are believed to be fed with gas in episodic events, triggered by mergers of galaxies. During these

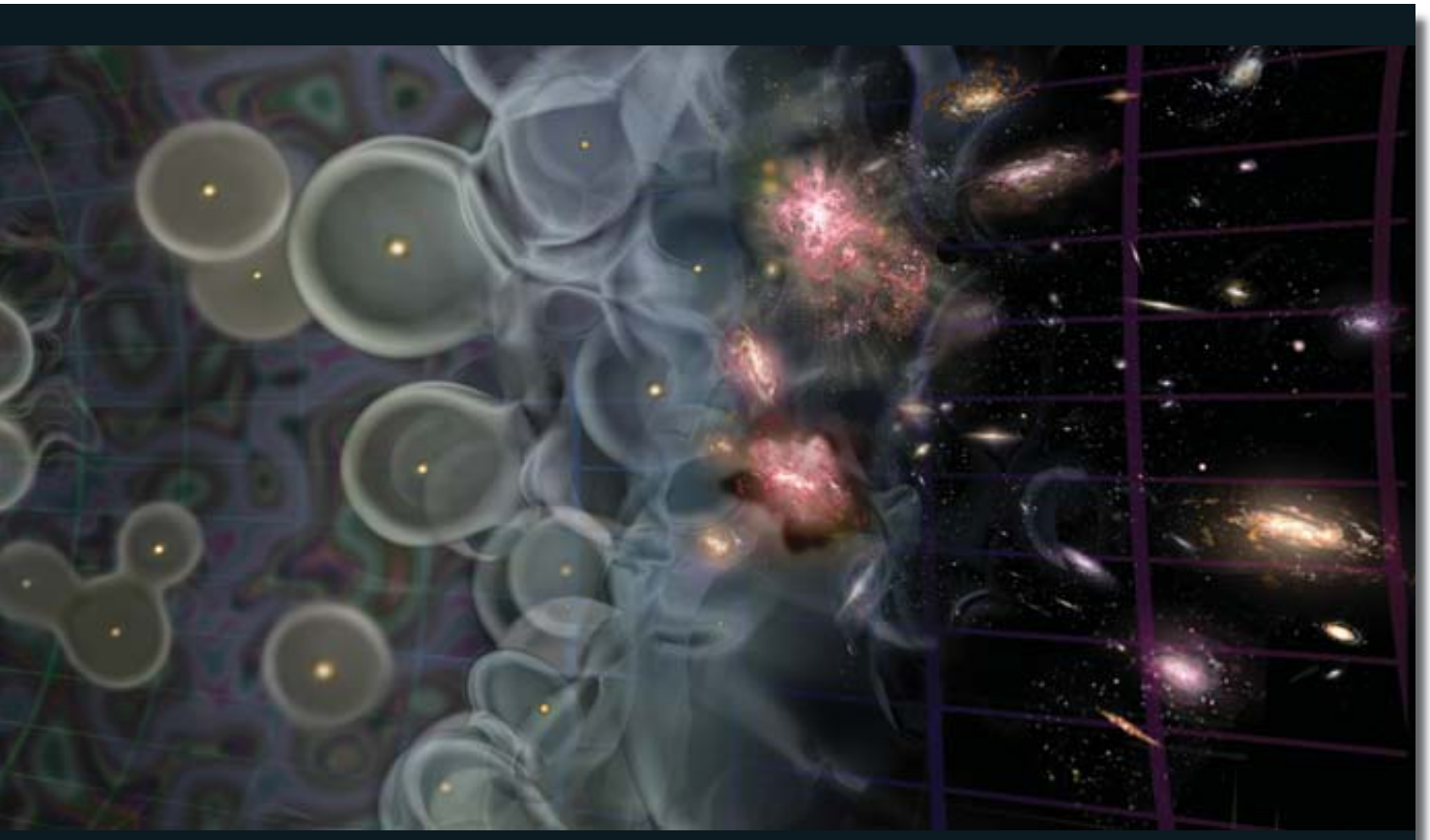
LIGHTING UP THE COSMOS

In the beginning of the Dark Ages, electrically neutral hydrogen gas filled the universe. As stars formed, they ionized the regions immediately around them, creating bubbles here and there. Eventually these bubbles merged together, and intergalactic gas became entirely ionized.



Simulated images of 21-centimeter radiation show how hydrogen gas turns into a galaxy cluster. The amount of radiation (*white is highest; orange and red are intermediate; black is least*) reflects both the density of the gas and its degree of ionization: dense, electrically neutral gas appears white; dense, ionized gas appears black. The images have been rescaled to remove the effect of cosmic expansion and thus highlight the cluster-forming processes. Because of expansion, the 21-centimeter radiation is actually observed at a longer wavelength; the earlier the image, the longer the wavelength.

JEAN-FRANCOIS PODEVIN (*illustration*); STEVEN FURLANETTO, AARON SOKASIAN AND LARS HERNQUIST *Harvard University* (*simulations*)



210 million years 2.4 million light-years 4.1 meters	290 million years 3.0 million light-years 3.3 meters	370 million years 3.6 million light-years 2.8 meters	460 million years 4.1 million light-years 2.4 meters	540 million years 4.6 million light-years 2.1 meters	620 million years 5.0 million light-years 2.0 meters	710 million years 5.5 million light-years 1.8 meters
All the gas is neutral. The white areas are the densest and will give rise to the first stars and quasars.	Faint red patches show that the stars and quasars have begun to ionize the gas around them.	These bubbles of ionized gas grow.	New stars and quasars form and create their own bubbles.	The bubbles are beginning to interconnect.	The bubbles have merged and nearly taken over all of space.	The only remaining neutral hydrogen is concentrated in galaxies.

growth spurts, the accreting gas shines much more brightly than the entire rest of the galaxy, producing a quasar. The Sloan Digital Sky Survey has revealed that quasars with black holes of more than a billion solar masses already existed at a cosmic age of one billion years. How did such massive black holes come to exist so early? Why did they stop growing?

Another set concerns the size distribution of galaxies. Theorists believe that the ultraviolet radiation produced by dwarf galaxies during the epoch of reionization heated the cosmic gas and suppressed the formation of new low-mass galaxies. How did this suppression unfold over time? Which of the dwarf galaxies we find today were already in existence at the beginning? These are only a few of the many questions whose answers lie in the Dark Ages. SA

MORE TO EXPLORE

Measuring the Small-Scale Power Spectrum of Cosmic Density Fluctuations through 21 cm Tomography Prior to the Epoch of Structure Formation. Abraham Loeb and Matias Zaldarriaga in *Physical Review Letters,* Vol. 92, No. 21, Paper No. 211301; May 25, 2004. Preprint available at **arxiv.org/abs/astro-ph/0312134**

The State of the Universe. Peter Coles in *Nature,* Vol. 433, pages 248–256; January 25, 2005.

First Light. Abraham Loeb. Lecture notes for the SAAS-Fee Winter School, April 2006. **arxiv.org/abs/astro-ph/0603360**

Chasing Hubble's Shadows: The Search for Galaxies at the Edge of Time. Jeff Kanipe. Hill and Wang, 2006.

Cosmology at Low Frequencies: The 21 cm Transition and the High-Redshift Universe. Steven Furlanetto, S. Peng Oh and Frank Briggs in *Physics Reports* (forthcoming). **arxiv.org/abs/astro-ph/0608032**

MIRRORS IN THE MIND

A special class of brain cells reflects the outside world, revealing a new avenue for human understanding, connecting and learning

By Giacomo Rizzolatti, Leonardo Fogassi and Vittorio Gallese

John watches Mary, who is grasping a flower. John knows what Mary is doing—she is picking up the flower—and he also knows why she is doing it. Mary is smiling at John, and he guesses that she will give him the flower as a present. The simple scene lasts just moments, and John's grasp of what is happening is nearly instantaneous. But how exactly does he understand Mary's action, as well as her intention, so effortlessly?

A decade ago most neuroscientists and psychologists would have attributed an individual's understanding of someone else's actions and, especially, intentions to a rapid reasoning process not unlike that used to solve a logical problem: some sophisticated cognitive apparatus in John's brain elaborated on the information his senses took in and compared it with similar previously stored experiences, allowing John to arrive at a conclusion about what Mary was up to and why.

Although such complex deductive operations probably do occur in some situations, particularly when someone's behavior is difficult to decipher, the ease and speed with which we typically understand simple actions suggest a much more straightforward explanation. In the early 1990s our research group at the University of Parma in Italy, which at the time included Luciano

ACTION PERFORMED by one person can activate motor pathways in another's brain responsible for performing the same action. The second understands viscerally what the first is doing because this mirror mechanism lets her experience it in own her mind.

CARY WOLINSKY (*photograph*); JEN CHRISTIANSEN (*photoillustration*)

Fadiga, found that answer somewhat accidentally in a surprising class of neurons in the monkey brain that fire when an individual performs simple goal-directed motor actions, such as grasping a piece of fruit. The surprising part was that these same neurons also fire when the individual sees someone else perform the same act. Because this newly discovered subset of cells seemed to directly reflect acts performed by another in the observer's brain, we named them mirror neurons.

Much as circuits of neurons are believed to store specific memories within the brain, sets of mirror neurons appear to encode templates for specific actions. This property may allow an individual not only to perform basic motor procedures without thinking about them but also to comprehend those acts when they are observed, without any need for explicit reasoning about them. John grasps Mary's action because even as it is happening before his eyes, it is also happening, in effect, inside his head. It is interesting to note that philosophers in the phenomenological tradition long ago posited that one had to experience something within oneself to truly comprehend it. But for neuroscientists, this finding of a physical basis for that idea in the mirror neuron system represents a dramatic change in the way we understand the way we understand.

Instant Recognition

OUR RESEARCH GROUP was not seeking to support or refute one philosophical position or another when we first noticed mirror neurons. We were studying the brain's motor cortex, particularly an area called F5 associated with hand and mouth movements, to learn how commands to perform certain actions are encoded by the firing patterns of neurons. For this purpose, we were recording the activity of individual neurons in the brains of macaques. Our laboratory contained a rich repertoire of stimuli for the monkeys, and as they performed various actions, such as grasping for a toy or a piece of food, we could see that distinct sets of neurons discharged during the execution of specific motor acts.

The pattern of activity was a true representation in the brain of the act itself, regardless of who was performing it.

Then we began to notice something strange: when one of us grasped a piece of food, the monkeys' neurons would fire in the same way as when the monkeys themselves grasped the food. At first we wondered whether this phenomenon could be the result of some trivial factor, such as the monkey performing an unnoticed movement while observing our actions. Once we managed to rule out this possibility and others, including food expectation by the monkeys, we realized that the pattern of neuron activity associated with the observed action was a true representation in the brain of the act itself, regardless of who was performing it.

Often in biological research, the most direct way to establish the function of a gene, protein or group of cells is simply to eliminate it and then look for deficits in the organism's health or behavior afterward. We could not use this technique to determine the role of mirror neurons, however, because we found them spread across important regions on both sides of the brain, including the premotor and parietal cortices. Destroying the entire mirror neuron system would have produced such broad general cognitive deficits in the monkeys that teasing out specific effects of the missing cells would have been impossible.

So we adopted a different strategy. To test whether mirror neurons play a role in understanding an action rather than just visually registering it, we assessed the neurons' responses when the monkeys could comprehend the meaning of an action without actually seeing it. If mirror neurons truly mediate understanding, we reasoned, their activity should reflect the meaning of the action rather than its visual features. We therefore carried out two series of experiments.

First we tested whether the F5 mirror neurons could "recognize" actions merely from their sounds. We recorded the mirror neurons while a monkey was observing a hand motor act, such as ripping a sheet of paper or breaking a peanut shell, that is accompanied by a distinctive sound. Then we presented the monkey with the sound alone. We found that many F5 mirror neurons that had responded to the visual observation of acts accompanied by sounds also responded to the sounds alone, and we dubbed these cell subsets audiovisual mirror neurons.

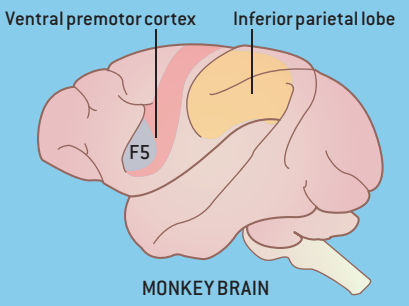
Next we theorized that if mirror neurons are truly involved in understanding an action, they should also discharge when the monkey does not actually see the action but has sufficient clues to create a mental representation of it. Thus, we first showed a monkey an experimenter reaching for and grasping a piece of food. Next, a screen was positioned in front of the monkey so that it could not

Overview/*Meeting of Minds*

- Subsets of neurons in human and monkey brains respond when an individual performs certain actions and also when the subject observes others performing the same movements.
- These "mirror neurons" provide a direct internal experience, and therefore understanding, of another person's act, intention or emotion.
- Mirror neurons may also underlie the ability to imitate another's action, and thereby learn, making the mirror mechanism a bridge between individual brains for communication and connection on multiple levels.

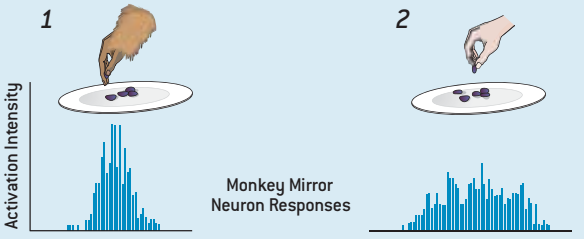
REALITY REFLECTED

In experiments with monkeys, the authors discovered subsets of neurons in brain-motor areas (*right*) whose activation appeared to represent actions themselves. Firing by these "mirror neurons" could therefore produce in one individual an internal recognition of another's act. Because the neurons' response also reflected comprehension of the movement's goal, the authors concluded that action understanding is a primary purpose of the mirror mechanism. Involvement of the mirror neurons in comprehending the actor's final intention was also seen in their responses, which distinguished between identical grasping actions performed with different intentions.



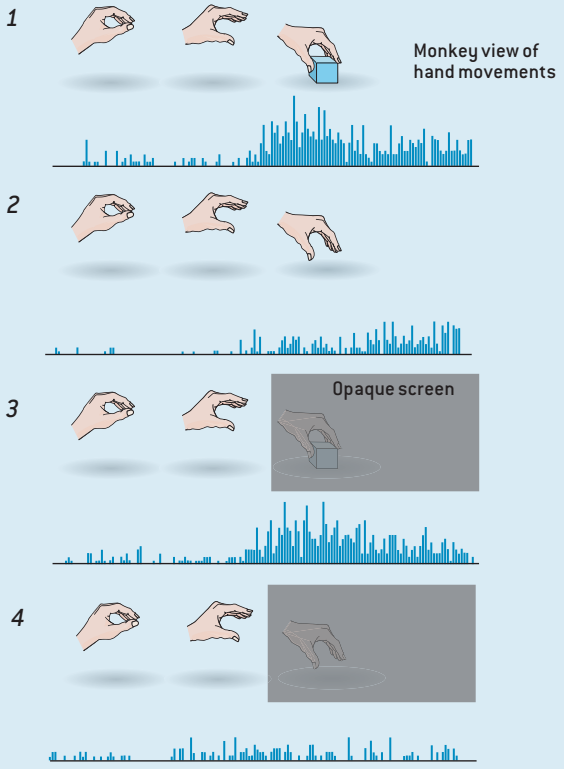
UNDERSTANDING ACTION

In early tests, a neuron in the premotor area F5, associated with hand and mouth acts, became highly active when the monkey grasped a raisin on a plate (*1*). The same neuron also responded intensely when an experimenter grasped the raisin as the monkey watched (*2*).



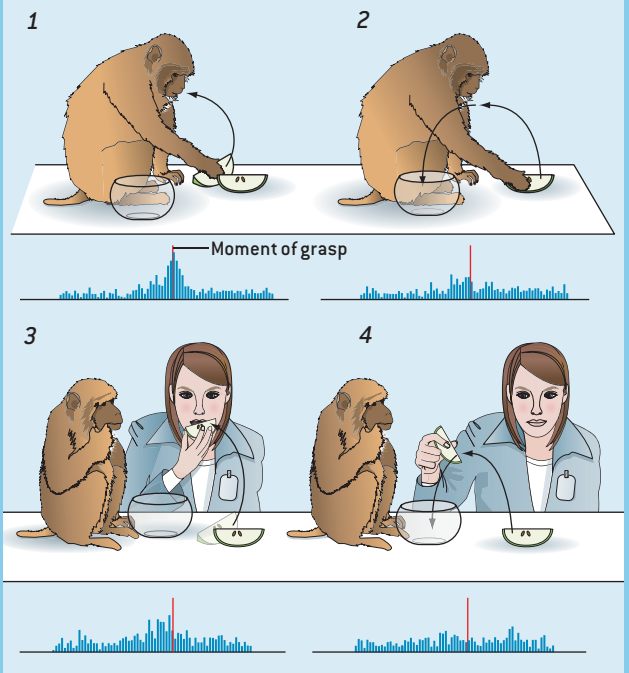
DISCRIMINATING GOAL

An F5 mirror neuron fired intensely when the monkey observed an experimenter's hand moving to grasp an object (*1*) but not when the hand motioned with no object as its goal (*2*). The same neuron did respond to goal-directed action when the monkey knew an object was behind an opaque screen, although the animal could not see the act's completion (*3*). The neuron responded weakly when the monkey knew no object was behind the screen (*4*).



DISCERNING INTENTION

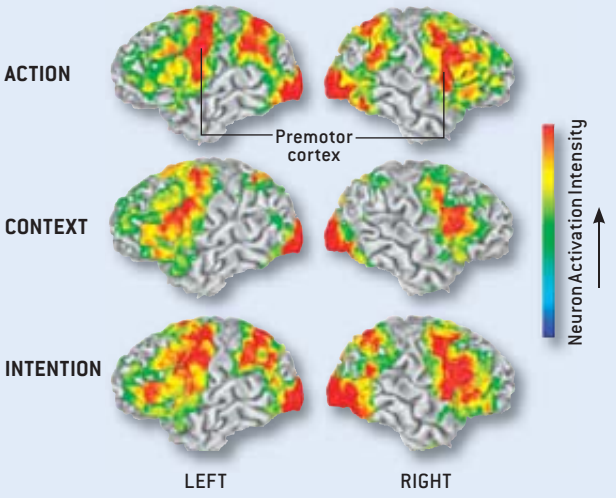
In the inferior parietal lobe, readings from one neuron show intense firing when the monkey grasped a fruit to bring it to its mouth (*1*). The neuron's response was weaker when the monkey grasped the food to place it in a container (*2*). The same mirror neuron also responded intensely when the monkey watched an experimenter perform the grasp-to-eat gesture (*3*) and weakly to the grasp-to-place action (*4*). In all cases, the responses were associated with the grasping act, indicating that the neuron's initial activation encoded an understanding of final intention.

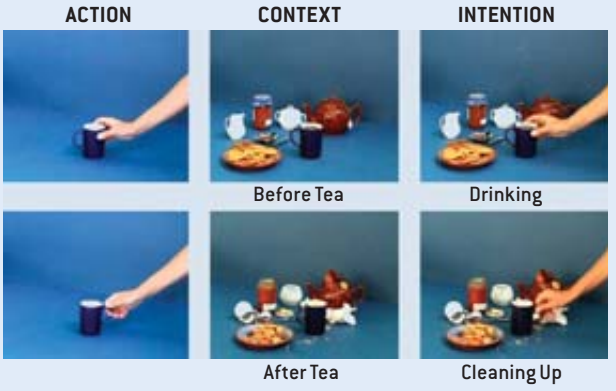


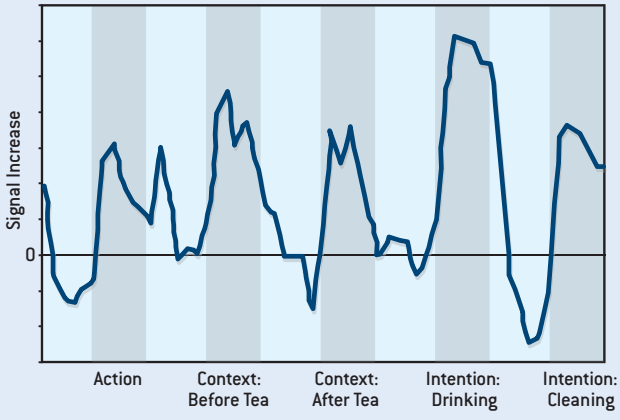
LUCY READING-IKKANDA

GRASPING INTENTION

Understanding the intentions of others is fundamental to human social behavior, and human mirror neurons appeared to confer that ability in an experiment designed to test their intention recognition. Volunteers were shown film clips (*below left*) depicting two similar cup-grasping actions without context, two contexts without action, and combinations of acts and context that signaled the action's intention: settings for afternoon tea that suggested the cup was being grasped for the purpose of drinking from it or that tea was over and the cup was being cleaned up. Activation of mirror neuron populations in premotor cortex areas in both hemispheres of subjects' brains (*right*) increased most strongly in response to scenes of action with a clear intention. Mirror neurons also distinguished between possible intentions, responding more intensely to the basic biological function of drinking than to the culturally acquired act of cleaning (*below right*).







see the experimenter's hand grasping the food but could only guess the action's conclusion. Nevertheless, more than half the F5 mirror neurons also discharged when the monkey could just imagine what was happening behind the screen.

These experiments confirmed, therefore, that the activity of mirror neurons underpins understanding of motor acts: when comprehension of an action is possible on a nonvisual basis, such as sound or mental representation, mirror neurons do still discharge to signal the act's meaning.

Following these discoveries in the monkey brain, we naturally wondered whether a mirror neuron system also exists in humans. We first obtained strong evidence that it does through a series of experiments that employed various techniques for detecting changes in motor cortex activity. As volunteers observed an experimenter grasping objects or performing meaningless arm gestures, for example, increased neural activation in their hand and arm muscles that would be involved in the same movements suggested a mirror neuron response in the motor areas of their brains. Further investigations using different external measures of cortical activity, such as electroencephalography, also supported the existence of a mirror neuron system in humans. But none of the technologies we had used up to this point allowed us to identify the exact brain areas activated when the volunteers observed motor acts, so we set out to explore this question with direct brain-imaging techniques.

In those experiments, carried out at San Raffaele Hospital in Milan, we used positron-emission tomography (PET) to observe neuronal activity in the brains of human volunteers as they watched grasping actions performed with different hand grips and then, as a control, looked at stationary objects. In these situations, seeing actions performed by others activated three main areas of the brain's cortex. One of these, the superior temporal sulcus (STS), is known to contain neurons that respond to observations of moving body parts. The other two—the inferior parietal lobule (IPL) and the inferior frontal gyrus (IFG)—correspond, respectively, to the monkey IPL and the monkey ventral premotor cortex, including F5, the areas where we had previously recorded mirror neurons.

These encouraging results suggested a mirror mechanism at work in the human brain as well but still did not fully reveal its scope. If mirror neurons permit an observed act to be directly understood by experiencing it, for example,

SOURCE: "GRASPING THE INTENTIONS OF OTHERS WITH ONE'S OWN MIRROR NEURON SYSTEM," BY M. IACOBONI ET AL., IN *PLoS BIOLOGY*, VOL. 3, NO. 3; 2005; LUCY READING-IKKANDA (*graph illustration*)

we wondered to what extent the ultimate goal of the action is also a component of that "understanding."

On Purpose

RETURNING TO our example of John and Mary, we said John knows both that Mary is picking up the flower and that she plans to hand it to him. Her smile gave him a contextual clue to her intention, and in this situation, John's knowledge of Mary's goal is fundamental to his understanding of her action, because giving him the flower is the completion of the movements that make up her act.

When we perform such a gesture ourselves, in reality we are performing a series of linked motor acts whose sequence is determined by our intent: one series of movements picks the flower and brings it to one's own nose to smell, but a partly different set of movements grasps the flower and hands it to someone else. Therefore, our research group set out to explore whether mirror neurons provide an understanding of intention by distinguishing between similar actions with different goals.

For this purpose, we returned to our monkeys to record their parietal neurons under varying conditions. In one set of experiments, a monkey's task was to grasp a piece of food and bring it to its mouth. Next we had the monkey grasp the same item and place it into a container. Interestingly, we found that most of the neurons we recorded discharged differently during the grasping part of the monkey's action, depending on its final goal. This evidence illustrated that the motor system is organized in neuronal chains, each of which encodes the specific intention of the act. We then asked whether this mechanism explains how we understand the intentions of others.

We tested the same grasping neurons for their mirror properties by having a monkey observe an experimenter performing the tasks the monkey itself had done earlier [*see box on page 57*]. In each instance, most of the mirror neurons were activated differently, depending on whether the experimenter brought the food to his mouth or put it in the container. The patterns of firing in the monkey's brain exactly matched those we observed when the monkey itself performed the acts—mirror neurons that discharged most strongly during grasping-to-eat rather than grasping-to-place did the same when the monkey watched the experimenter perform the corresponding action.

A strict link thus appears to exist between the motor organization of intentional actions and the capacity to understand the intentions of others. When the monkeys observed an action in a particular context, seeing just the first grasping component of the complete movement activated mirror neurons forming a motor chain that also encoded a specific intention. Which chain was activated during their observation of the beginning of an action depended on a variety of factors, such as the nature of the object acted on, the context and the memory of what the observed agent did before.

To see whether a similar mechanism for reading intentions exists in humans, we teamed with Marco Iacoboni and his colleagues at the University of California, Los Angeles, for a functional magnetic resonance imaging (fMRI) experiment on volunteers. Participants in these tests were presented with three kinds of stimuli, all contained within video clips. The first set of images showed a hand grasping a cup against an empty background using two different grips. The second consisted of two scenes containing objects such as plates and cutlery, arranged in one instance as though they were ready for someone to have afternoon tea and in the other as though they were left over from a previously eaten snack and were ready to be cleaned up. The third stimulus set showed a hand grasping a cup in either of those two contexts.

We wanted to establish whether human mirror neurons would distinguish between grasping a cup to drink, as suggested by the ready-for-tea context, and grabbing the cup to take it away, as suggested by the cleanup setting. Our results demonstrated not only that they do but also that the mirror neuron system responded strongly to the intention component of an act. Test subjects observing the hand motor acts in the "drinking" or "cleaning" contexts showed differing activation of their mirror neuron systems, and mirror neuron activity was stronger in both those situations than when subjects observed the hand grasping a cup without any context or when looking only at the place settings [*see box on opposite page*].

When people use the expression "I feel your pain," they may not realize how literally it could be true.

Given that humans and monkeys are social species, it is not difficult to see the potential survival advantage of a mechanism, based on mirror neurons, that locks basic motor acts onto a larger motor semantic network, permitting the direct and immediate comprehension of others' behavior without complex cognitive machinery. In social life, however, understanding others' emotions is equal-

THE AUTHORS

GIACOMO RIZZOLATTI, LEONARDO FOGASSI and *VITTORIO GALLESE* work together at the University of Parma in Italy, where Rizzolatti is director of the neurosciences department and Fogassi and Gallese are associate professors. In the early 1990s their studies of motor systems in the brains of monkeys and humans first revealed the existence of neurons with mirror properties. They have since continued to investigate those mirror neurons in both species as well as the role of the motor system in general cognition. They frequently collaborate with the many other research groups in Europe and the U.S. now also studying the breadth and functions of the mirror neuron system in humans and animals.

ly important. Indeed, emotion is often a key contextual element that signals the intent of an action. That is why we and other research groups have also been exploring whether the mirror system allows us to understand what others feel in addition to what they do.

Connect and Learn

AS WITH ACTIONS, humans undoubtedly understand emotions in more than one way. Observing another person experiencing emotion can trigger a cognitive elaboration of that sensory information, which ultimately results in a logical conclusion about what the other is feeling. It may also, however, result in direct mapping of that sensory information onto the motor structures that would produce the experience of that emotion in the observer. These two means of recognizing emotions are profoundly different: with the first, the observer deduces the emotion but does not feel it; via the second, recognition is firsthand because the mirror mechanism elicits the same emotional state in the observer. Thus, when people use the expression "I feel your pain" to indicate both comprehension and empathy, they may not realize just how literally true their statement could be.

A paradigmatic example is the emotion of disgust, a basic reaction whose expression has important survival value for fellow members of a species. In its most primitive form, disgust indicates that something the individual tastes or smells is bad and, most likely, dangerous. Once again using fMRI studies, we collaborated with French neuroscientists to show that experiencing disgust as a result of inhaling foul odorants and witnessing disgust on the face of someone else activate the same neural structure—the anterior insula—at some of the very same locations within that structure [*see box below*]. These results indicate that populations of mirror neurons in the insula become active both when the test participants experience the emotion and when they see it expressed by others. In other words, the observer and the observed share a neural mechanism that enables a form of direct experiential understanding.

Tania Singer and her colleagues at University College London found similar matches between experienced and observed emotions in the context of pain. In that experiment, the participants felt pain produced by electrodes placed on their hands and then watched electrodes placed on a test partner's hand followed by a cue for painful stimulation. Both situations activated the same regions of the anterior insula and the anterior cingulate cortex in the subjects.

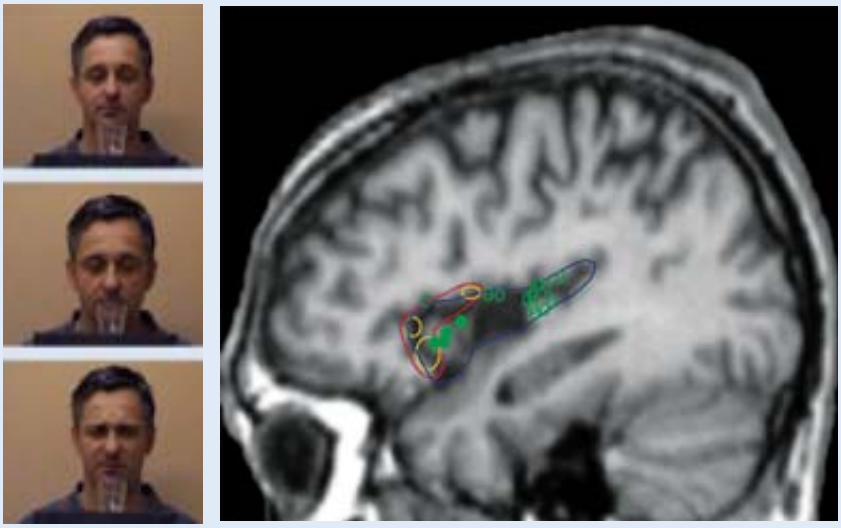
Taken together, such data strongly suggest that humans may comprehend emotions, or at least powerful negative emotions, through a direct mapping mechanism involving parts of the brain that generate visceral motor responses. Such a mirror mechanism for understanding emotions cannot, of course, fully explain all social cognition, but it does provide for the first time a functional neural basis for some of the interpersonal relations on which more complex social behaviors are built. It may be a substrate that allows us to empathize with others, for example. Dysfunction in this mirroring system may also be implicated in empathy deficits, such as those seen in children with autism [see "Broken Mirrors: A Theory of Autism," by Vilayanur S. Ramachandran and Lindsay M. Oberman, on page 62].

Many laboratories, including our own, are continuing to explore these questions, both for their inherent interest and their potential therapeutic applications. If the mirror neuron template of a motor action is partly inscribed in the brain by experience, for instance, then it should theoretically be possible to alleviate motor impairments, such as those suffered following a stroke, by potentiating undamaged action templates. Recent evidence indicates, in fact, that the mirror mechanism also plays a role in the way we initially learn new skills.

Although the word "ape" is often used to denote mimicry, imitation is not an especially well developed ability among nonhuman primates. It is rare in monkeys and limited in the great apes, including chimpanzees and gorillas. For

EMOTIONAL MIRRORS

Feeling disgust activated similar parts of the brain when human volunteers experienced the emotion while smelling a disgusting odor or when the same subjects watched a film clip (*left*) of someone else disgusted. In this brain cross section, neuron populations activated by the experience of disgust are outlined in red, and those activated by seeing disgust are circled in yellow. (Blue outlines the region of investigation, and green indicates areas examined in a previous study.) These overlapping neuron groups may represent a physical neural mechanism for human empathy that permits understanding the emotions of others.



IMITATION requires reproduction of actions performed by another person. If mirror neurons underlie the uniquely human facility for imitation, the mirror system may serve as a bridge that allows us to teach and learn new skills.

humans, in contrast, imitation is a very important means by which we learn and transmit skills, language and culture. Did this advance over our primate relatives evolve on the neural substrate of the mirror neuron system? Iacoboni and his group provided the first evidence that this might be the case when they used fMRI to observe human subjects who were watching and imitating finger movements. Both activities triggered the IFG, part of the mirror neuron system, in particular when the movement had a specific goal.

In all these experiments, however, the movements to be imitated were simple and highly practiced. What role might mirror neurons play when we have to learn completely new and complex motor acts by imitation? To answer this question, Giovanni Buccino at our university and collaborators in Germany recently used fMRI to study participants imitating guitar chords after seeing them played by an expert guitarist. While test subjects observed the expert, their parietofrontal mirror neuron systems became active. And the same area was even more strongly activated during the subjects' imitation of the chord movements. Interestingly, in the interval following observation, while the participants were programming their own imitation of the guitar chords, an additional brain region became active. Known as prefrontal area 46, this part of the brain is traditionally associated with motor planning and working memory and may therefore play a central role in properly assembling the elementary motor acts that constitute the action the subject is about to imitate.

Many aspects of imitation have long perplexed neuroscientists, including the basic question of how an individual's brain takes in visual information and translates it to be reproduced in motor terms. If the mirror neuron system serves as a bridge in this process, then in addition to providing an understanding of other people's actions, intentions and emotions, it may have evolved to become an important component in the human capacity for observation-based learning of sophisticated cognitive skills.

Scientists do not yet know if the mirror neuron system is unique to primates or if other animals possess it as well. Our own research group is currently testing rats to see if that species also demonstrates mirror neuron responses. Such internal mirroring may be an ability that developed late in evolution, which would explain why it is more extensive in humans than in monkeys. Because even newborn human and monkey babies can imitate simple gestures such as sticking out the tongue, however, the ability to create mirror templates for observed actions could be innate. And because lack of emotional mirroring ability appears to be a hallmark of autism, we are also working with young autistic children to learn whether they have detectable motor deficits that could signal a general dysfunction of the mirror neuron system.

Only a decade has passed since we published our first discoveries about mirror neurons, and many questions remain to be answered, including the mirror system's possible role in language—one of humanity's most sophisticated cognitive skills. The human mirror neuron system does include Broca's area, a fundamental language-related cortical center. And if, as some linguists believe, human communication first began with facial and hand gestures, then mirror neurons would have played an important role in language evolution. In fact, the mirror mechanism solves two fundamental communication problems: parity and direct comprehension. Parity requires that meaning within the message is the same for the sender as for the recipient. Direct comprehension means that no previous agreement between individuals—on arbitrary symbols, for instance—is needed for them to understand each other. The accord is inherent in the neural organization of both people. Internal mirrors may thus be what allow John and Mary to connect wordlessly and permit human beings in general to communicate on multiple levels. SA

MORE TO EXPLORE

Action Recognition in the Premotor Cortex. Vittorio Gallese, Luciano Fadiga, Leonardo Fogassi and Giacomo Rizzolatti in *Brain,* Vol. 119, No. 2, pages 593–609; April 1996.

A Unifying View of the Basis of Social Cognition. V. Gallese, C. Keysers and G. Rizzolatti in *Trends in Cognitive Sciences,* Vol. 8, pages 396–403; 2004.

Grasping the Intentions of Others with One's Own Mirror Neuron System. Marco Iacoboni et al. in *PLoS Biology,* Vol. 3, Issue 3, pages 529–535; March 2005.

Parietal Lobe: From Action Organization to Intention Understanding. Leonardo Fogassi et al. in *Science,* Vol. 302, pages 662–667; April 29, 2005.

GETTY IMAGES

BROKEN MIRRORS

A THEORY OF AUTISM

Studies of the mirror neuron system may reveal clues to the causes of autism and help researchers develop new ways to diagnose and treat the disorder

By Vilayanur S. Ramachandran and Lindsay M. Oberman

At first glance you might not notice anything odd on meeting a young boy with autism. But if you try to talk to him, it will quickly become obvious that something is seriously wrong. He may not make eye contact with you; instead he may avoid your gaze and fidget, rock his body to and fro, or bang his head against the wall. More disconcerting, he may not be able to conduct anything remotely resembling a normal conversation. Even though he can experience emotions such as fear, rage and pleasure, he may lack genuine empathy for other people and be oblivious to subtle social cues that most children would pick up effortlessly.

In the 1940s two physicians—American psychiatrist Leo Kanner and Austrian pediatrician Hans Asperger—independently discovered this developmental disorder, which afflicts about 0.5 percent of American children. Neither researcher had any knowledge of the other's work, and yet by an uncanny coincidence each gave the syndrome the same name: autism, which derives from the Greek word *autos*, meaning "self." The name is apt, because the most conspicuous feature of the disorder is a withdrawal from social interaction. More recently, doctors have adopted the term "autism spectrum disorder" to make it clear that the illness has many related variants that range widely in severity but share some characteristic symptoms.

Ever since autism was identified, researchers have struggled to determine what causes it. Scientists know that susceptibility to autism is inherited, although environmental risk factors also seem to play a role [see "The Early Origins of Autism," by Patricia M. Rodier; SCIENTIFIC

CARY WOLINSKY (*photograph*); JEN CHRISTIANSEN (*photoillustration*)

CHILDREN WITH AUTISM may struggle with social interaction because their mirror neuron systems are not functioning properly.

AMERICAN, February 2000]. Starting in the late 1990s, investigators in our laboratory at the University of California, San Diego, set out to explore whether there was a connection between autism and a newly discovered class of nerve cells in the brain called mirror neurons. Because these neurons appeared to be involved in abilities such as empathy and the perception of another individual's intentions, it seemed logical to hypothesize that a dysfunction of the mirror neuron system could result in some of the symptoms of autism. Over the past decade, several studies have provided evidence for this theory. Further investigations of mirror neurons may explain how autism arises, and in the process physicians may develop better ways to diagnose and successfully treat the disorder.

Mirror neurons appear to be performing precisely the same functions that are disrupted in autism.

Explaining the Symptoms

ALTHOUGH THE CHIEF diagnostic signs of autism are social isolation, lack of eye contact, poor language capacity and absence of empathy, other less well known symptoms are commonly evident. Many people with autism have problems understanding metaphors, sometimes interpreting them literally. They also have difficulty miming other people's actions. Often they display an eccentric preoccupation with trifles yet ignore important aspects of their environment, especially their social surroundings. Equally puzzling is the fact that they frequently show an extreme aversion to certain sounds that, for no obvious reason, set off alarm bells in their minds.

The theories that have been proposed to explain autism can be divided into two groups: anatomical and psychological. (Researchers have rejected a third group of theories—such as the "refrigerator mother" hypothesis—that blame the disorder on poor upbringing.) Eric Courchesne of U.C.S.D. and other anatomists have shown elegantly that children with autism have characteristic abnormalities in the cerebellum, the brain structure responsible for coordinating complex voluntary muscle movements. Although these observations must be taken into account in any final explanation of autism, it would be premature to conclude that damage to the cerebellum is the sole cause of the disorder. Cerebellar damage inflicted by a stroke in a child usually produces tremors, swaying gait and abnormal eye movements—symptoms rarely seen in autism. Conversely, one does not see any of the symptoms typical of autism in patients with cerebellar disease. It is possible that the cerebellar changes observed in children with autism may be unrelated side effects of abnormal genes whose *other* effects are the true causes of the disorder.

Perhaps the most ingenious of the psychological theories is that of Uta Frith of University College London and Simon Baron-Cohen of the University of Cambridge, who posit that the main abnormality in autism is a deficit in the ability to construct a "theory of other minds." Frith and Baron-Cohen argue that specialized neural circuitry in the brain allows us to create sophisticated hypotheses about the inner workings of other people's minds. These hypotheses, in turn, enable us to make useful predictions about others' behavior. Frith and Baron-Cohen are obviously on the right track, but their theory does not provide a complete explanation for the constellation of seemingly unrelated symptoms of autism. Indeed, saying that people with autism cannot interact socially because they lack a "theory of other minds" does not go very far beyond restating the symptoms. What researchers need to identify are the brain mechanisms whose known functions match those that are disrupted in autism.

One clue comes from the work of Giacomo Rizzolatti and his colleagues at the University of Parma in Italy, who in the 1990s studied neural activity in the brains of macaque monkeys while the animals were performing goal-directed actions [see "Mirrors in the Mind," by Giacomo Rizzolatti, Leonardo Fogassi and Vittorio Gallese, on page 54]. Researchers have known for decades that certain neurons in the premotor cortex—part of the brain's frontal lobe—are involved in controlling voluntary movements. For instance, one neuron will fire when the monkey reaches for a peanut, another will fire when the animal pulls a lever, and so on. These brain cells are often referred to as motor command neurons. (Bear in mind that the neuron whose activity is recorded does not control the arm by itself; it is part of a circuit that can be monitored by observing the signals in the constituent neurons.)

What surprised Rizzolatti and his coworkers was that a subset of the motor command neurons also fired when the monkey watched another monkey or a researcher perform the same action. For example, a neuron involved in controlling the reach-for-the-peanut action fired when the monkey saw one of his fellows

Overview/*Mirror Neurons and Autism*

- Because mirror neurons appear to be involved in social interaction, dysfunctions of this neural system could explain some of the primary symptoms of autism, including isolation and absence of empathy.
- Studies of people with autism show a lack of mirror neuron activity in several regions of the brain. Researchers speculate that treatments designed to restore this activity could alleviate some of autism's symptoms.
- A complementary hypothesis, the salience landscape theory, could account for secondary symptoms of autism such as hypersensitivity.

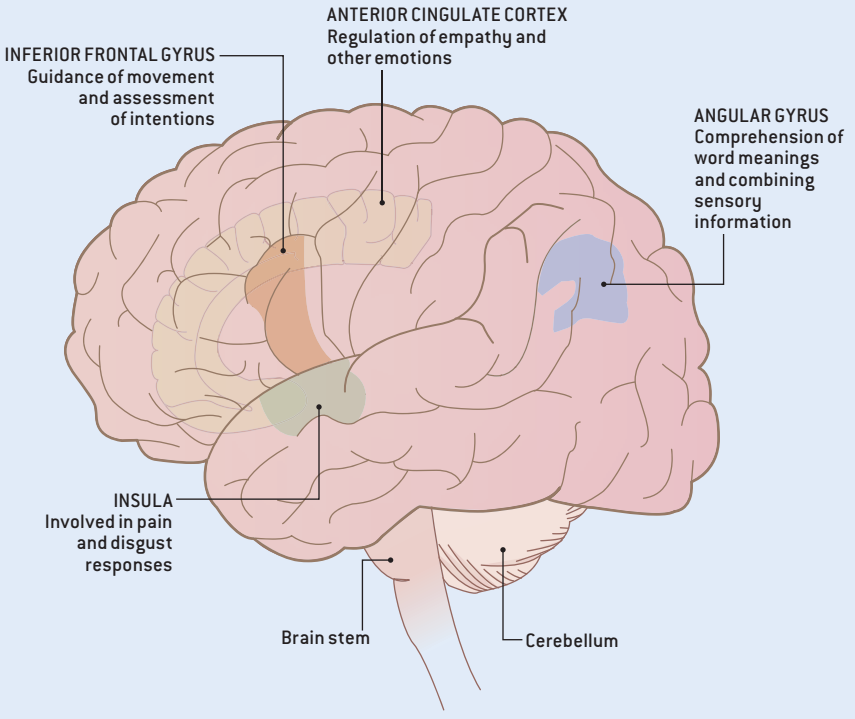
making that movement. Brain-imaging techniques subsequently showed that these so-called mirror neurons also exist in the corresponding regions of the human cortex. These observations implied that mirror neurons—or, more accurately, the networks they are part of—not only send motor commands but also enable both monkeys and humans to determine the intentions of other individuals by mentally simulating their actions. In monkeys, the role of the neurons may be limited to predicting simple goal-directed actions, but in humans the mirror neuron system may have evolved the ability to interpret more complex intentions.

Later research showed that mirror neurons are located in other parts of the human brain, such as the cingulate and insular cortices, and that they may play a role in empathetic emotional responses. While studying the anterior cingulate cortex of awake human subjects, investigators found that certain neurons that typically fire in response to pain also fired when the person saw someone else in pain. Mirror neurons may also be involved in imitation, an ability that appears to exist in rudimentary form in the great apes but is most pronounced in humans. The propensity to imitate must be at least partly innate: Andrew Meltzoff of the University of Washington has shown that if you stick your tongue out at a newborn baby, the infant will do the same. Because the baby cannot see its own tongue, it cannot use visual feedback and error correction to learn the skill. Instead there must be a hardwired mechanism in the child's brain for mapping the mother's visual appearance—whether it be a tongue sticking out or a smile—onto the motor command neurons.

Language development in childhood also requires a remapping of sorts between brain areas. To imitate the mother's or father's words, the child's brain must transform auditory signals in the hearing centers of the brain's temporal lobes into verbal output from the motor cortex. Whether mirror neurons are directly involved in this skill is not known, but clearly some analogous process must be going on. Last, mirror neurons may enable humans to see themselves as others see them, which may be an essential ability for self-awareness and introspection.

THE ANATOMY OF AUTISM

People with autism show reduced mirror neuron activity in the inferior frontal gyrus, a part of the brain's premotor cortex, perhaps explaining their inability to assess the intentions of others. Dysfunctions of mirror neurons in the insula and anterior cingulate cortex may cause related symptoms, such as the absence of empathy, and deficits in the angular gyrus may result in language difficulties. People with autism also have structural changes in the cerebellum and brain stem.



LUCY READING-IKKANDA

Suppressing Mu Waves

WHAT HAS ALL THIS to do with autism? In the late 1990s our group at U.C.S.D. noted that mirror neurons appear to be performing precisely the same functions that seem to be disrupted in autism. If the mirror neuron system is indeed involved in the interpretation of complex intentions, then a breakdown of this neural circuitry could explain the most striking deficit in people with autism, their lack of social skills. The other cardinal signs of the disorder—absence of empathy, language deficits, poor imitation, and so on—are also the kinds of things you would expect to see if mirror neurons were dysfunctional. Andrew Whitten's group at the University of St. Andrews in Scotland made this proposal at about the same time we did, but the first experimental evidence for the hypothesis came from our lab, working in collaboration with Eric L. Altschuler and Jaime A. Pineda of U.C.S.D.

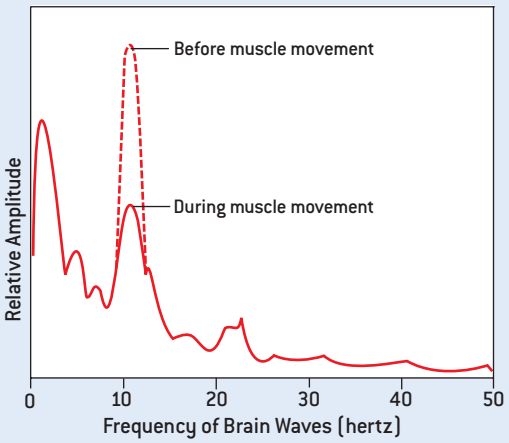
To demonstrate mirror neuron dysfunction in children with autism, we needed to find a way to monitor the activity of their nerve cells without putting electrodes in their brains (as Rizzolatti and his colleagues did with their monkeys). We realized that we could do so using an electroencephalogram (EEG) measurement of the children's brain waves. For more than half a century, scientists have known that an EEG component called the mu wave is blocked anytime a person makes a voluntary muscle movement, such as opening and closing one's hands. Interestingly, this component is also blocked when a person watches someone else perform the same

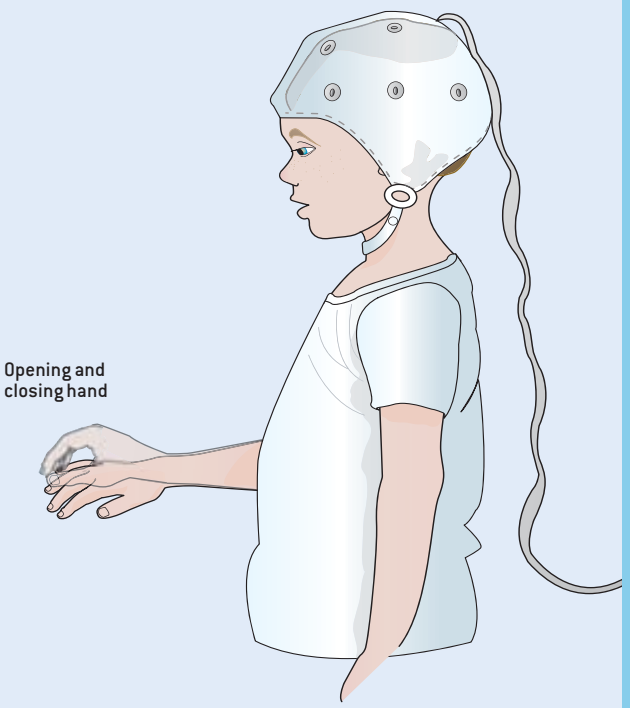
FOCUSING ON MU WAVES

To study the mirror neuron system in people with autism, researchers relied on the observation that the firing of neurons in the premotor cortex suppresses the mu wave, a component of the electroencephalogram (EEG) measurement of the brain's activity. (Mu waves range from eight to 13 hertz.) Investigators monitored the mu waves of children with autism and control subjects as they made voluntary muscle movements and then watched the same actions on video.

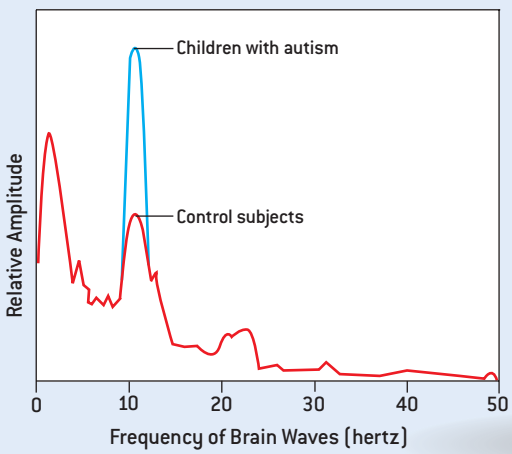
TAKING ACTION

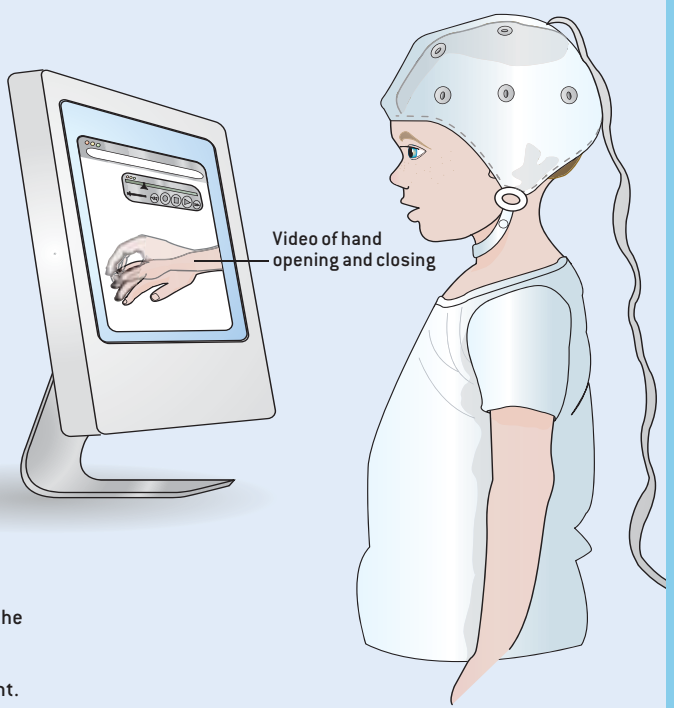
Motor command neurons fire whenever a person makes a voluntary muscle movement. Researchers asked all the subjects to open and close their right hands. In the children with autism and the control subjects, this action suppressed the amplitude of their mu waves, as expected.





SIMULATING ACTION





Mirror neurons in the premotor cortex also fire when a person observes someone else performing an action. The investigators took EEG measurements of brain activity while the subjects observed a video of a hand opening and closing. The mu waves of the control subjects plummeted (*red*), but those of the children with autism showed no suppression (*blue*). This finding suggests that the mirror neuron systems of the children with autism are deficient.

LUCY READING-IKKANDA

action. One of us (Ramachandran) and Altschuler suggested that mu-wave suppression might provide a simple, noninvasive probe for monitoring mirror neuron activity.

We decided to focus our first experiments on a high-functioning child with autism—that is, a child without severe cognitive impairments. (Very young, low-functioning children did not participate in this study because we wanted to confirm that any differences we found were not a result of problems in attention, understanding instructions or the general effects of mental retardation.) The EEG showed that the child had an observable mu wave that was suppressed when he made a simple, voluntary movement, just as in normal children. But when the child watched someone else perform the action, the suppression did not occur. We concluded that the child's motor command system was intact but that his mirror neuron system was deficient. This observation, which we presented at the annual meeting of the Society for Neuroscience in 2000, provided a striking vindication of our hypothesis.

One has to be careful, however, of generalizing from a single case, so our lab group later conducted a more systematic series of experiments in 10 high-functioning individuals with autism spectrum disorder and 10 age- and gender-matched control subjects. We saw the expected suppression of mu waves when the control subjects moved their hands and watched videos of a moving hand, but the EEGs of the subjects with autism showed mu suppression only when they moved their own hands.

Other researchers have confirmed our results using different techniques for monitoring neural activity. A group led by Riitta Hari of the Helsinki University of Technology found mirror neuron deficits in children with autism by employing magnetoencephalography, which measures the magnetic fields produced by electric currents in the brain. More recently, Mirella Dapretto of the University of California, Los Angeles, and her colleagues used functional magnetic resonance imaging to show a reduction in mirror neuron activity in the prefrontal cortices of individuals with autism. And Hugo Théoret of the University of Montreal and his co-workers used transcranial magnetic stimulation, a technique that induces electric currents in the motor cortex to generate muscle movements, to study mirror neuron activity in subjects with autism. In the control subjects, induced hand movements became more pronounced when the subjects watched videos of the same movements; this effect was much weaker in the subjects with autism.

Taken together, these findings provide compelling evidence that people with autism have dysfunctional mirror neuron systems. Scientists do not yet know which genetic and environmental risk factors can prevent the development of mirror neurons or alter their function, but many research groups are now actively pursuing the hypothesis because it predicts symptoms that are unique to autism. In addition to explaining the primary signs of autism, deficiencies in the mirror neuron system can also account for some of the less well known symptoms. For instance, researchers have long known that children with autism often have problems interpreting proverbs and metaphors. When we told one of our subjects to "get a grip on yourself," he took the message literally and started grabbing his own body. Though seen in only a subset of children with autism, this difficulty with metaphors cries out for an explanation.

These findings provide compelling evidence that people with autism have dysfunctional mirror neuron systems.

Understanding metaphors requires the ability to extract a common denominator from superficially dissimilar entities. Consider the bouba/kiki effect, which was discovered by German-American psychologist Wolfgang Köhler more than 60 years ago. In this test, a researcher displays two crudely drawn shapes, one jagged and one curvy, to an audience and asks, "Which of these shapes is bouba and which is kiki?" No matter what languages the respondents speak, 98 percent will pick the curvy shape as bouba and the jagged one as kiki. This result suggests that the human brain is somehow able to extract abstract properties from the shapes and sounds—for example, the property of jaggedness embodied in both the pointy drawing and the harsh sound of kiki. We conjectured that this type of cross-domain mapping is analogous to metaphors and must surely involve neural circuits similar to those in the mirror neuron system. Consistent with this speculation, we discovered that children with autism perform poorly at the bouba/kiki test, pairing the shapes and sounds incorrectly.

But which part of the human brain is involved in this skill? The angular gyrus, which sits at the crossroads of the brain's vision, hearing and touch centers, seemed to be a likely candidate—not only because of its strategic location but because nerve cells with mirror neuron–like properties have been identified there. When we studied nonautistic subjects with damage to this area of the brain, we found that many of them fail the bouba/kiki test and have a disproportionate dif-

THE AUTHORS

VILAYANUR S. RAMACHANDRAN and *LINDSAY M. OBERMAN* have investigated the links between autism and the mirror neuron system at the Center for Brain and Cognition at the University of California, San Diego. Ramachandran, director of the center, earned his Ph.D. in neuroscience from the University of Cambridge. A renowned expert on brain abnormalities, he has also studied the phenomena of phantom limbs and synesthesia, for which he won the 2005 Henry Dale Prize and a lifetime fellowship from the Royal Institution of Great Britain. Oberman is a graduate student in Ramachandran's laboratory at U.C.S.D., joining the group in 2002.

ficulty understanding metaphors, just like people with autism. These results suggest that cross-domain mapping may have originally developed to aid primates in complex motor tasks such as grasping tree branches (which requires the rapid assimilation of visual, auditory and touch information) but eventually evolved into an ability to create metaphors. Mirror neurons allowed humans to reach for the stars, instead of mere peanuts.

Can the Mirrors Be Repaired?

THE DISCOVERY of mirror neuron deficiencies in people with autism opens up new approaches to diagnosing and treating the disorder. For example, physicians could use the lack of mu-wave suppression (or perhaps the failure to mimic a mother sticking out her tongue) as a diagnostic tool to identify children with autism in early infancy, so that the currently available behavioral therapies can be started as quickly as possible. Timely intervention is critical; the behavioral therapies are much less effective if begun after autism's main symptoms appear (typically between ages two and four).

An even more intriguing possibility would be to use biofeedback to treat autism or at least alleviate its symptoms. Doctors could monitor the mu waves of a child with autism and display them on a screen in front of the patient. If the child's mirror neuron functions are dormant rather than completely lost, it may be possible for him or her to revive this ability by learning—through trial and error and visual feedback—how to suppress the mu waves on the screen. Our colleague Pineda is pursuing this approach, and his preliminary results look promising. Such therapies, though, should supplement rather than replace the traditional behavioral-training techniques.

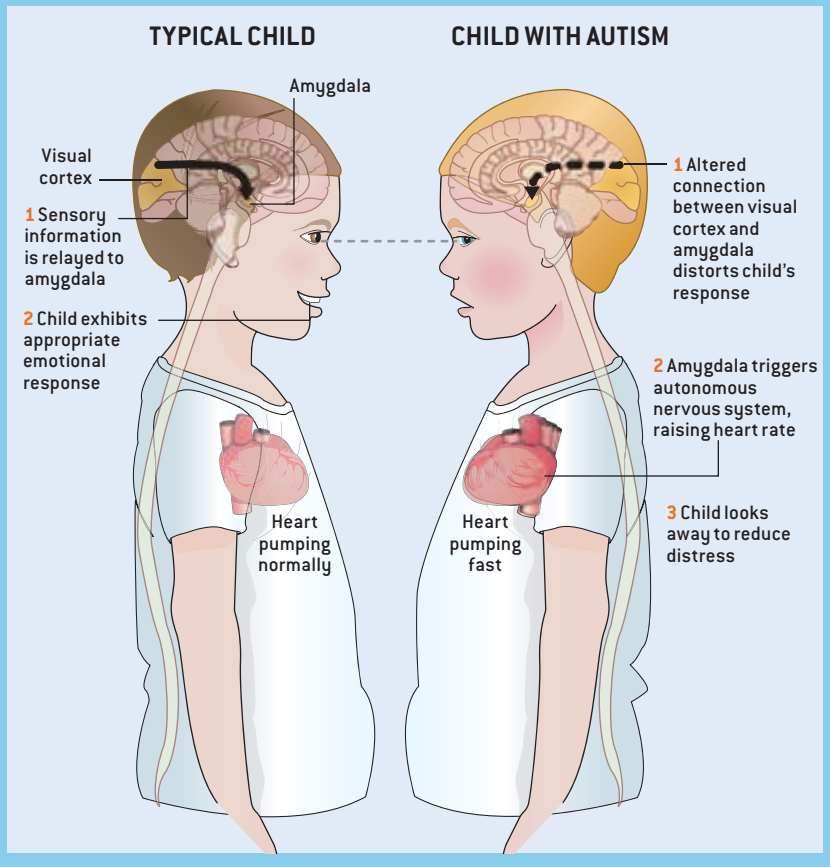
Another novel therapeutic approach might rely on correcting chemical imbalances that disable the mirror neurons in individuals with autism. Our group (including students Mikhi Horvath and Mary Vertinsky) has suggested that specialized neuromodulators may enhance the activity of mirror neurons involved in emotional responses. According to this hypothesis, the partial depletion of such chemicals could explain the lack of emotional empathy seen in autism, and therefore researchers should look for compounds that stimulate the release of the neuromodulators or mimic their effects on mirror neurons. One candidate for investigation is MDMA, better known as ecstasy, which has been shown to foster emotional closeness and communication. It is possible that researchers may be able to modify the compound to develop a safe, effective treatment that could alleviate at least some of autism's symptoms.

Such treatments, however, may offer only partial relief, because other symptoms of autism cannot be explained by the mirror neuron hypothesis—for example, repetitive motions such as rocking to and fro, avoidance of eye contact, hypersensitivity, and aversion to certain sounds. In an attempt to determine how these secondary symptoms might arise, our lab group (in collaboration with William Hirstein of Elmhurst College and Portia Iversen of Cure Autism Now, a nonprofit foundation based in Los Angeles) has developed what we call the salience landscape theory.

When a person looks at the world, he

THE SALIENCE LANDSCAPE THEORY

To account for some of the secondary symptoms of autism—hypersensitivity, avoidance of eye contact, aversion to certain sounds, and so on—researchers have developed the salience landscape theory. In a typical child, sensory information is relayed to the amygdala, the gateway to the emotion-regulating limbic system. Using input from stored knowledge, the amygdala determines how the child should respond emotionally to each stimulus, creating a salience landscape of the child's environment. In children with autism, though, the connections between the sensory areas and the amygdala may be altered, resulting in extreme emotional responses to trivial events and objects.



LUCY READING-IKKANDA

or she is confronted with an overwhelming amount of sensory information—sights, sounds, smells, and so on. After being processed in the brain's sensory areas, the information is relayed to the amygdala, which acts as a portal to the emotion-regulating limbic system. Using input from the individual's stored knowledge, the amygdala determines how the person should respond emotionally—for example, with fear (at the sight of a burglar), lust (on seeing a lover) or indifference (when facing something trivial). Messages cascade from the amygdala to the rest of the limbic system and eventually reach the autonomic nervous system, which prepares the body for action. If the person is confronting a burglar, for example, his heart rate will rise and his body will sweat to dissipate the heat from muscular exertion. The autonomic arousal, in turn, feeds back into the brain, amplifying the emotional response. Over time, the amygdala creates a salience landscape, a map that details the emotional significance of everything in the individual's environment.

Our group decided to explore the possibility that children with autism have a distorted salience landscape, perhaps because of altered connections between the cortical areas that process sensory input and the amygdala or between the limbic structures and the frontal lobes that regulate the resulting behavior. As a result of these abnormal connections, any trivial event or object could set off an extreme emotional response—an autonomic storm—in the child's mind. This hypothesis would explain why children with autism tend to avoid eye contact and any other novel sensation that might trigger an upheaval. The distorted perceptions of emotional significance might also explain why many children with autism become intensely preoccupied with trifles such as train schedules while expressing no interest at all in things that most children find fascinating.

We found some support for our hypothesis when we monitored autonomic responses in a group of 37 children with autism by measuring the increase in their skin conductance caused by sweating. In contrast with the control subjects, the children with autism had a higher overall level of autonomic arousal. Although they became agitated when exposed to trivial objects and events, they often ignored stimuli that triggered expected responses in the control group.

But how could a child's salience landscape become so distorted? Investigators have found that nearly one third of children with autism have had temporal lobe epilepsy in infancy, and the proportion may be much higher given that many epileptic seizures go undetected. Caused by repeated random volleys of nerve impulses traversing the limbic system, these seizures could eventually scramble the connections between the visual cortex and the amygdala, indiscriminately enhancing some links and diminishing others. In adults, temporal lobe epilepsy results in florid emotional disturbances but does not radically affect cognition; in infants, however, the seizures may lead to a more profound disability. And, like autism, the risk of temporal lobe epilepsy in infancy appears to be influenced by both genetic and environmental factors. Some genes, for example, could make a child more susceptible to viral infections, which could in turn predispose the child to seizures.

Our findings on autonomic responses may help explain the old clinical observation that high fever sometimes temporarily alleviates the symptoms of autism. The autonomic nervous system is involved in controlling body temperature; because fever and the emotional upheavals of autism appear to be regulated by the same neural pathways, perhaps the former can mitigate the latter.

The salience landscape theory could also provide an explanation for the repetitive motions and head banging seen in children with autism: this behavior, called self-stimulation, may somehow damp the child's autonomic storms. Our studies found that self-stimulation not only had a calming effect but also led to a measurable reduction in skin conductance. This result suggests a possible symptomatic therapy for autism. Hirstein is now developing a portable device that could monitor an autistic child's skin conductance; when the device detects autonomic arousal, it could turn on another device, called a squeeze vest, that provides a comforting pressure by gently tightening around the child's body.

If the child's mirror neuron functions are dormant rather than lost, it may be possible to revive this ability.

Our two candidate theories for explaining the symptoms of autism—mirror neuron dysfunction and distorted salience landscape—are not necessarily contradictory. It is possible that the same event that distorts a child's salience landscape—the scrambled connections between the limbic system and the rest of the brain—also damages the mirror neurons. Alternatively, the altered limbic connections could be a side effect of the same genes that trigger the dysfunctions in the mirror neuron system. Further experiments are needed to rigorously test these conjectures. The ultimate cause of autism remains to be discovered. In the meantime, our speculations may provide a useful framework for future research. SA

MORE TO EXPLORE

Autonomic Responses of Autistic Children to People and Objects. William Hirstein, Portia Iversen and Vilayanur S. Ramachandran in *Proceedings of the Royal Society of London B,* Vol. 268, pages 1883–1888; 2001.

EEG Evidence for Mirror Neuron Dysfunction in Autism Spectrum Disorders. Lindsay M. Oberman, Edward M. Hubbard, Joseph P. McCleery, Eric L. Altschuler, Jaime A. Pineda and Vilayanur S. Ramachandran in *Cognitive Brain Research,* Vol. 24, pages 190–198; 2005.

A Brief Tour of Human Consciousness. New edition. Vilayanur S. Ramachandran. Pi Press, 2005.

MALWARE GOES MOBILE

BY MIKKO HYPPONEN

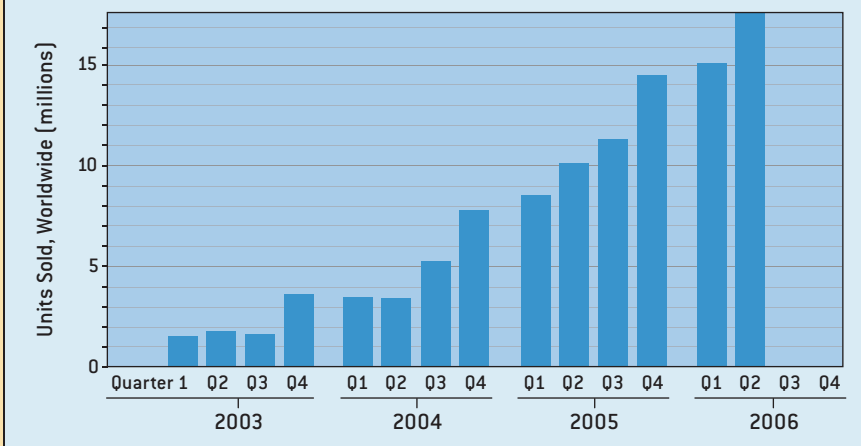
Computer viruses are now airborne, infecting mobile phones in every part of the globe. Security companies, cellular operators and phone makers are moving to quash these threats before they spiral out of control

The day the computer security community had anticipated for years finally arrived in June 2004. I and other researchers who study malicious forms of software knew that it was only a matter of time until such malware appeared on mobile phones as well. As cell phones have evolved into smartphones—able to download programs from the Internet and share software with one another through short-range Bluetooth connections, worldwide multimedia messaging service (MMS) communications and memory cards—the devices' novel capabilities have created new vulnerabilities. Scoundrels were bound to find the weaknesses and exploit them for mischief or, worse, for criminal gain.

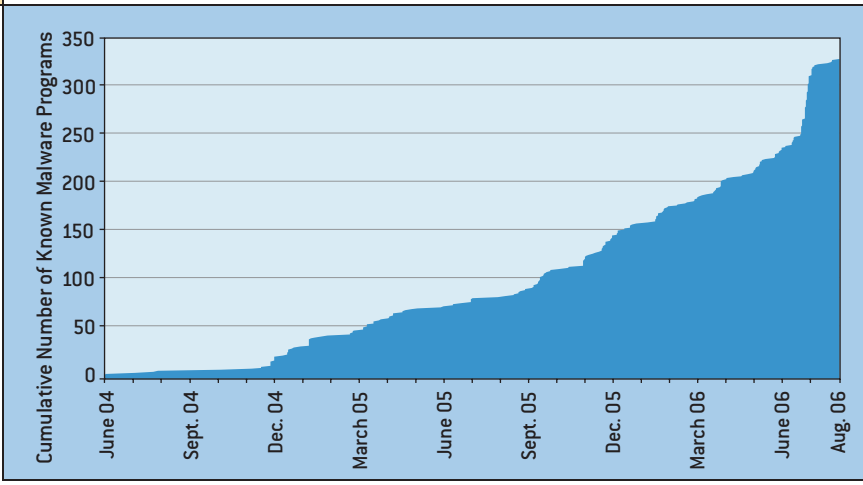
Sure enough, three summers ago security experts found the first rogue program written specifically for smartphones. Dubbed Cabir, it was a classic proof-of-concept virus, clearly created to

INFECTION of one smartphone by malicious software—malware—could bring down others in a domino effect.

SMARTPHONES ON THE RISE



GROWTH IN MOBILE MALWARE



MORE PHONES, MORE TARGETS

The number of smart mobile devices in the world has expanded dramatically in recent years, and so has the amount of malware set loose to attack them. That mix is a recipe for disaster: as the size of a target audience increases, so, too, does the likelihood that miscreant programmers will attack it. And audience size is expected to soar in the years ahead. Industry analysts predict that more than 200 million smartphones will be sold in 2009.

capture bragging rights. It caused no damage to an infected device, other than running down the phone's battery as the virus tried to copy itself to another smartphone by opening a Bluetooth connection. The anonymous author, most likely somewhere in Spain, chose to post Cabir on a Web site rather than releasing it into the wild. But within two months other scofflaws had turned it loose in Southeast Asia. It soon spread worldwide.

Even though we had been on the lookout for viruses such as Cabir, security experts were not fully prepared to deal with it. As soon as the alert was sounded, my co-workers and I at F-Secure, a computer security firm, started inspecting the new virus, which was a type known as a worm [*see box on opposite page for definitions of terms*]. But we had no safe place to study it; unlike a computer virus that can be observed and dissected on a machine that is disconnected from any network, wireless malware can spread—in some cases, even make transoceanic leaps—the moment the infected phone is powered up.

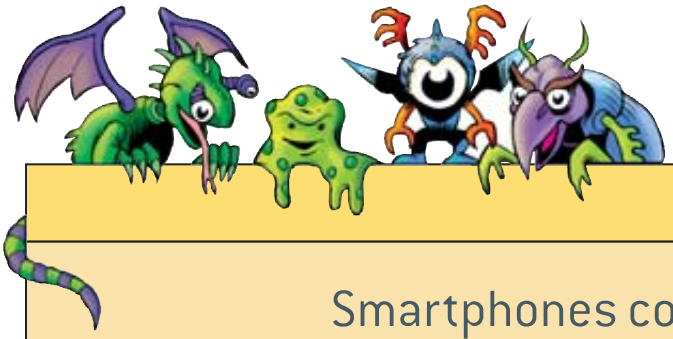
So we took four cell phones hit by Cabir to the basement bomb shelter in our office building and posted a guard at the door before turning them on, lest an unsuspecting employee walk in and catch the bug. Later that year F-Secure built two aluminum-and-copper-encased laboratories, impenetrable to radio waves, to study this contagious new form of malware.

Although the initial version of Cabir was relatively innocuous, some unscrupulous malware writers rushed to modify it into forms that are more virulent and damaging, while others began crafting novel kinds of attacks. Mobile viruses on the loose now can completely disable a phone, delete the data on it or force the device to send costly messages to premium-priced numbers. Within two years the number of viruses targeting smartphones soared from one to

Overview/*Imperiled Phones*

- The first malicious software aimed at smartphones hit in 2004. Smartphones are mobile phones that permit users to install software applications from sources other than the cellular network operator.
- Today more than 300 kinds of malware—among them worms, Trojan horses, other viruses and spyware—have been unleashed against the devices.
- As sales of such sophisticated phones soar worldwide, the stage is being set for the massive spread of malware. Steps are being taken to prevent that scenario, but the opportunity to block the onslaught is unlikely to last long.

MIRACLE STUDIOS (*preceding pages* and *opposite page*); LUCY READING-IKKANDA (*this page*); SOURCES: CANALYS (*top graph*) AND F-SECURE SECURITY RESEARCH (*bottom graph*)



more than 200, a rate of growth that roughly paralleled that of computer viruses in the first two years after the first PC virus, called Brain, was released in 1986.

Despite Herculean efforts to rein it in, PC malware continues at a gallop: more than 200,000 forms have been identified so far, and today an unprotected PC is often infected within minutes of connecting to the Internet. The economic costs of the 20-year onslaught have been steep, and they are spiraling higher as old-school malware written for glory has given way to a new era of "crimeware" designed for spamming, data theft or extortion.

Mobile malware, though little more than a nuisance today, could quickly escalate into an even more formidable problem than PC malware in the years ahead unless the security community, cellular network operators, smartphone designers and phone users all work together to hold it in check. The history of PC malware is humbling, but it offers lessons that will help us to anticipate some of the ways in which mobile virus writers will strike next and to take steps to thwart them.

Smartphones could in the very near future make up most of the world's computers.

A Rising Tide

IN 1988 many computer experts dismissed viruses as inconsequential novelties. That assessment proved regrettably naive. For mobile malware, the time is now 1988, and we have a brief window in which to act to avoid repeating the mistakes of the past.

One such mistake was to underestimate how quickly malware would grow in prevalence, diversity and sophistication. Prevalence is a function of both the population of potential hosts for virtual pathogens and of their rate of infection. The target population for malicious mobile software is enormous and growing by leaps. There are now more than two *billion* mobile phones in the world.

It is true that the great majority of these are older cell phones running closed, proprietary operating systems that are largely immune from viral infection. But customers are quickly abandoning these devices for newer generations of smartphones that run more open operating systems, Web browsers, e-mail and other messaging clients and that contain Flash memory card readers and short-range Bluetooth radios. Each of these features offers a conduit through which malware can propagate.

Bluetooth, for example, allows certain mobile worms to spread among vulnerable phones by mere proximity, almost like the influenza virus. A Bluetooth-equipped smartphone can identify and exchange files with other Bluetooth devices from a distance of 10 meters or more. As victims travel, their phones can leave a trail of infected bystanders in their wake. And any event that attracts a large crowd presents a perfect breeding ground for Bluetooth viruses.

A particularly nasty form of Cabir, for example, spread so rapidly through the audience at the 2005 world track and field championships in Helsinki that stadium operators flashed warnings on the big screen. Most smartphones can put Bluetooth into a "nondiscoverable" mode that protects them from invasion by worms. But few users avail themselves of this feature. While giving a talk at a computer security conference this spring, I conducted a quick scan of the room and found that almost half the professionals in the audience had left the Bluetooth radios in their phones wide open. The proportion is even higher among the general population, so these devices offer a disturbingly effective vector for invisible parasites.

And this host population is growing rapidly. Smartphones got started as expensive business models, but their popularity with consumers has recently taken off. With each generation the devices accrete more PC-like functionality. At the same time that smartphones have begun sporting features such as video cameras, GPS navigation and MP3 players, their prices have dropped—subsidized in part by network operators, who hope the new capabilities will encourage customers to spend more on cellular services. Manufacturers sold more than 40 million smartphones last year, and industry analysts expect to see 350 million units in service by 2009.

In the medium term, these devices may be adopted most quickly in emerging economies, where computer owner-

A Malware Primer

PHISHING SCAM
Fraudulent Web page, e-mail or text message that entices the unwary to reveal passwords, financial details or other private data.

SPYWARE
Software that reveals private information about the user or computer system to eavesdroppers.

TROJAN HORSE
A program that purports to be useful but actually harbors hidden malicious code.

VIRUS
Originally, computer code that inserts itself into another program and replicates when the host software runs. Now often used as a generic term that also includes Trojan horses and worms.

WORM
Self-replicating code that automatically spreads across a network.

ANATOMY OF AN ATTACK

Even an astute person can fall victim to a well-designed mobile worm, such as CommWarrior. Some 15 variants of this worm have been seen since the malware was first spotted in March 2005. CommWarrior exploits the Bluetooth user interface to persuade victims to install the malware on their phones. Once active, it can spread rapidly via Bluetooth connections, multimedia (MMS) messages and memory cards.

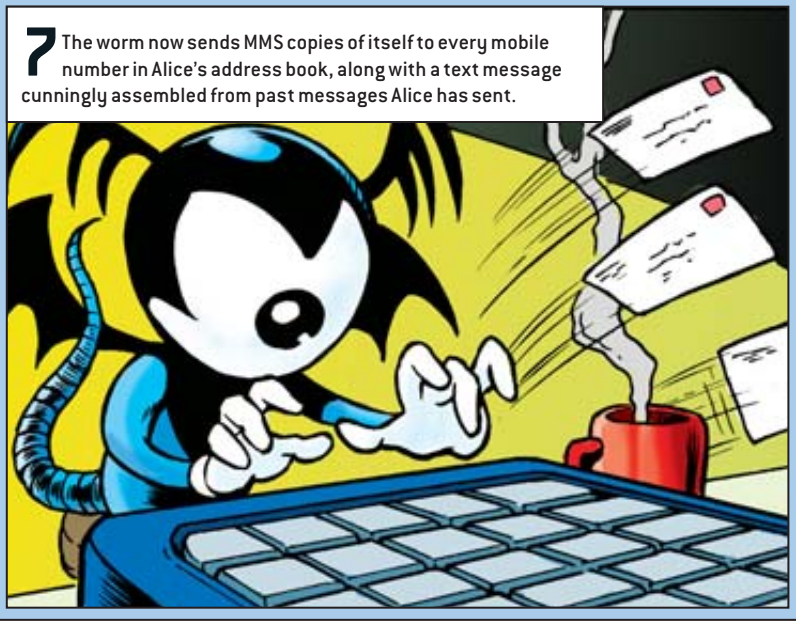
1 As Bob boards a bus, his smartphone beeps. Another phone in the vehicle is carrying CommWarrior.Q, which is attempting to copy itself onto Bob's phone via Bluetooth.

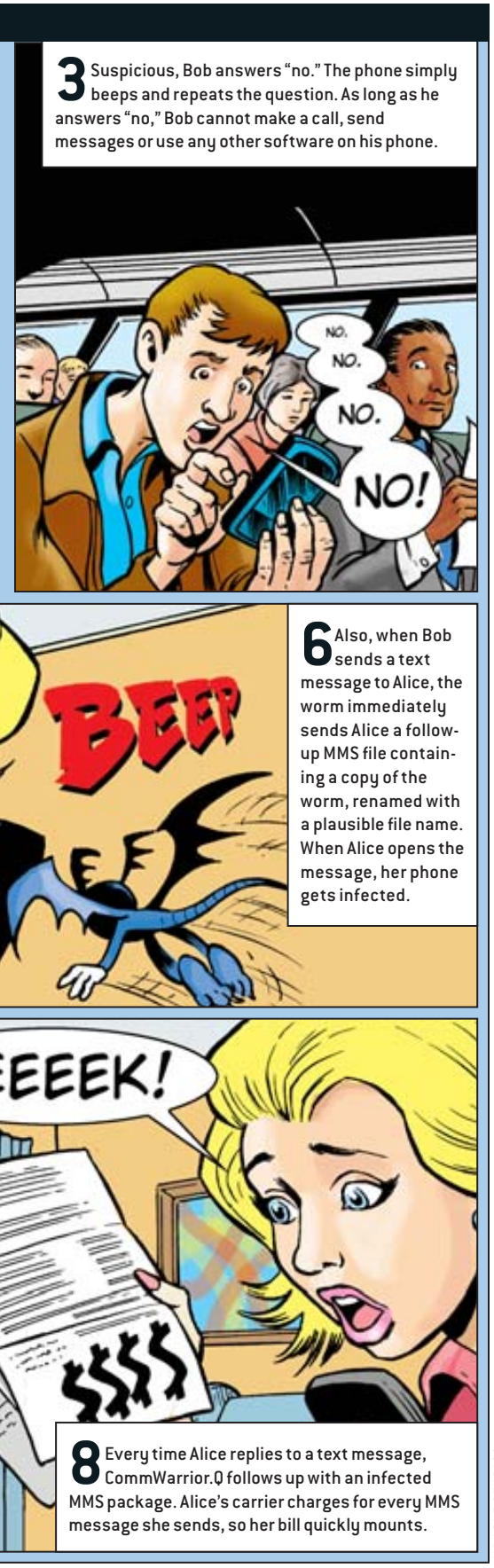
2 Bob's phone alerts him that it is about to receive a file and asks his permission to accept the transmission.

4 Bob needs to make an urgent call so he finally answers "yes" to the transmission query and to the installation and security queries after it. His phone now becomes infected. If Bob should place his phone's memory card into another phone to transfer an application, the second device would become infected.

5 CommWarrior.Q begins scanning for other Bluetooth devices nearby and attempts to copy itself onto any it finds, sometimes onto several at once.

7 The worm now sends MMS copies of itself to every mobile number in Alice's address book, along with a text message cunningly assembled from past messages Alice has sent.





MIRACLE STUDIOS

ship is still relatively low. Research by Canalys, a high-tech consultancy near Reading, England, found that smartphone sales in the first quarter of this year grew twice as fast in eastern Europe, Africa and the Middle East as they did in western Europe. Industry analysts predict that some developing nations will choose to forgo construction of a wired Internet infrastructure and will instead upgrade their digital wireless networks and promote smartphones as affordable computers. The wireless route can be much less expensive to construct and maintain (and, from a censor's perspective, much easier to monitor and control).

If these forecasts prove accurate, smartphones could in the very near future make up most of the world's computers. And huge populations of users who have little or no experience with computers could soon be surfing the Web and sharing files with their phones. They would present mobile malware creators with an irresistibly large and unwary target.

One lesson from PC viruses is that the bigger the target, the bigger the attraction for nefarious programmers. The vast majority of desktop malware works only on the ubiquitous Microsoft Windows operating system. For the same reason, nearly all the mobile worms and Trojan horses released so far infect the Symbian operating system, which runs some 70 percent of smartphones worldwide—including phones made by Nokia, Samsung, Sony Ericsson and Motorola. In contrast, only a few varieties of malware infect Microsoft's PocketPC or Windows Mobile, Palm's Treo, or Research in Motion's BlackBerry devices. The Symbian bias partly explains why mobile malware is currently most prevalent in Europe and Southeast Asia, where Symbian is commonplace, but is rarer in North America, Japan and South Korea. Cellular operators in North America have spread their markets more equally across the various platforms. The Japanese and Korean markets were dominated for a long time by Linux-based phones, and carriers there heavily restrict the types of applications that users can install on their phones.

Carriers would be wise to begin educating cellular customers now about how to identify and avoid mobile viruses, rather than waiting until these infections become epidemic. Phone makers should install antivirus software by default, just as PC manufacturers now do. And regulators and phone companies can also help avoid the monoculture problem that plagues PCs by encouraging a diverse ecosystem for smartphones in which no single variety of software dominates the market.

From Kicks to Crime

DIVERSITY CUTS both ways, of course. Over time malware, too, inevitably mutates into new species that attack and subvert useful software in an ever widening variety of ways. On the PC, the early viruses were eventually joined by Trojans, worms, spyware and most recently phishing attacks. Since 2003 much of the new malware appearing on PCs has been written for profit rather than for mere mischief. Organized gangs of cyber-criminals now operate all over the world. Thieves use crimeware to make money by stealing financial data, business secrets or computer resources. Spammers assemble "botnets" of hacked machines to forward bulk e-mail and phishing scams. And blackmailers extort money with threats of digital destruction or of virtual blockades that shut down a company's Web or e-mail servers. In some countries, cyber-criminals are virtually untouchable because authorities lack the technical

THE AUTHOR

MIKKO HYPPONEN is chief research officer for F-Secure, a computer security company in Helsinki that consults for mobile phone makers and network operators. His team of virus fighters has been first to identify and combat dozens of viruses in the 15 years he has worked at F-Secure, including the infamous LoveLetter worm in 2000. A co-author of two books on computer security, Hypponen has assisted with investigations by Microsoft, the U.S. Federal Bureau of Investigation, the U.S. Secret Service and Scotland Yard in the U.K.

Computers do not have a built-in billing system; mobile phones do. The bad guys will exploit this feature before long.

MIRACLE STUDIOS

expertise, resources or will to enforce laws against computer crimes.

As for-profit virus writing increases, the likelihood of severe mobile malware attacks escalates as well. After all, every phone call placed and every text or multimedia message sent is also a financial transaction. That opens up a flood of potential earning opportunities for profiteer hackers and virus authors. Computers do not have a built-in billing system; mobile phones do. The bad guys will exploit this feature before long.

Indeed, at least one already has. A Trojan called RedBrowser sends a continuous stream of text messages from any phone it infects to a number in Russia until the user disables the phone. Each message is charged at a premium rate of about five dollars, resulting in huge bills for the unfortunate victims. Some cellular carriers hold their customers liable for such unauthorized transactions, and when they do, the criminals, who own the premium number, collect the premium fees. Luckily, RedBrowser has so far only been spotted inside Russia.

Meanwhile service providers in North American markets are beginning to introduce "mobile wallets." Customers will be able to use their phones to transfer funds from their accounts to others by sending specially formatted text messages. PayPal, a digital payments firm, offers a similar service that allows users to buy items using their phones. Such services could be of intense interest to malware authors.

With both the sophistication of mobile malware and the technological and financial capabilities of mobile phones on the rise, we will have to move rapidly in the next couple of years. Actions now could thwart mobile malware while it is in its infancy and while smartphone services are still fairly flexible in their design. But that window of opportunity will not stay open for long.

More Dangers Ahead

THE REASON FOR HASTE is clear when one considers all the ways that hackers could—but have yet to—wreak havoc with smartphones. On personal computers, many of the worst culprits spread via e-mail or force infected machines to spew spam onto the Internet. None of the miscreant programs released so far for smartphones capitalize on the devices' ability to send e-mail. It is only a matter of time until malware appears that can propagate as e-mail attachments or can turn phones into spam-sending robots.

Spyware is another mushrooming problem in the PC arena, and the potential for surreptitious software on phones to destroy privacy is obvious. Only a handful of such programs have been seen as yet. One, called FlexiSpy, periodically and invisibly sends a log of phone calls and multimedia messages, both sent and received, to a third party. The eavesdropper needs to gain physical access to the phone to download and install the spying program.

It may not be long, however, before hackers incorporate this kind of eavesdropping behavior into viruses that replicate on their own. With new phones featuring voice recorder capability, manufacturers should take extra care to ensure that these features cannot easily be exploited by malware to record conversations and then beam the recordings to a snoop.

Then there is the surprising fact that not one of the more than 300 forms of mobile malware released as yet exploits programming bugs or security design flaws to insert itself into a vulnerable machine. This has long been a standard modus operandi for many PC viruses and Trojans.

So far mobile malware writers have instead relied exclusively on "social engineering"—in other words, tricking users into actively allowing installation of the malicious program on their phones. Some camouflage themselves as useful utilities or desirable games. But some, especially ones like Cabir and CommWarrior that spread via Bluetooth, do not. Many people accept the files even when the device warns of the security risk and gives them a chance to refuse the foreign software.

I and other researchers have asked

Some Protective Software for Smartphones

COMPANY	PROGRAM NAME	SUPPORTED OPERATING SYSTEMS
F-Secure	Mobile Anti-Virus Mobile Security	PocketPC, Symbian, Windows Mobile Nokia Communicators
McAfee	VirusScan Mobile	PocketPC, Symbian, Windows Mobile
Symantec	AntiVirus for Handhelds Mobile Security	Palm, PocketPC, Windows Mobile Symbian
Trend Micro	Mobile Security	PocketPC, Symbian, Windows Mobile

A Bestiary of Mobile Malware

NAME	TYPE AND METHOD OF INFECTION	EFFECTS
Cabir (discovered June 2004)	Worm. Connects to other Bluetooth devices and copies itself	Constant Bluetooth scanning drains phone's battery
CommWarrior (discovered March 2005)	Worm. Replicates via Bluetooth; sends itself as an MMS file to numbers in phone's address book and in automatic replies to incoming SMS (text) and MMS messages; copies itself to the removable memory card and inserts itself into other program installation files on phone	Some users incur a charge for every MMS file the worm sends; variants of the worm disable phone entirely
Doomboot (discovered July 2005)	Trojan horse. Pretends to be a version of the Doom 2 video game, enticing users to download and install it	Prevents phone from booting and installs Cabir and CommWarrior on phone
RedBrowser (discovered February 2006)	Trojan horse. Deceptive description on a Web site offering many downloadable programs entices users to install this Java program, which runs on hundreds of phone models	Surreptitiously sends a stream of text messages, at a premium rate of $5 each, to a phone number in Russia
FlexiSpy (discovered March 2006)	Spyware. Internet download, typically installed by someone other than phone owner	Sends a log of phone calls and copies of text and MMS messages to a commercial Internet server for viewing by a third party

people victimized by such viruses: Why did you click "yes"? A common answer is that they did not at first—they chose "no." But then the question immediately reappeared on the screen. A worm, you see, does not take no for an answer, and it gives the user no time to hit the menu option to disable Bluetooth [*see box on pages 74 and 75*]. Unfortunately, even the newest versions of most smartphones permit the kind of Bluetooth harassment that effectively denies a person use of a phone until the individual accepts the file transfer (or until the user walks out of range of whatever infected device is sending the request—although few people realize they have this option).

Staying a Step Ahead

THE ONLY HOPE of stopping mobile malware before it seriously degrades the utility and value of smartphones is quick and concerted action on the part of all concerned. Antivirus software now available from many companies can immunize and disinfect smartphones. Yet few customers have installed such protection. That needs to change.

Phones should also incorporate firewall software that warns the user when a program on the phone seizes the initiative to open an Internet connection. This is an especially important form of protection for smartphones that can connect to Wi-Fi (also called 802.11) networks and thus directly to the public Internet. Many cellular companies aggressively filter traffic on the GPRS or UMTS data networks that their mobile devices use; open Wi-Fi networks have no such protection. And while some carriers already filter their MMS streams to remove messages bearing malicious attachments, all should do so.

Some of the biggest phone manufacturers have joined the Trusted Computing Group, which has been hammering out industry standards for microcircuitry inside phones that will make it harder for malware to get at sensitive data in the device's memory or to hijack its payment mechanisms. And Symbian recently released a new version of its operating system that does an improved job of protecting key files and that requires software authors to obtain digital certificates from the company. The new Symbian system refuses to install programs not accompanied by a certificate. Unless disabled by a user, the system effectively excludes all mobile malware discovered to date.

Governments could also play a more constructive role than they have so far. Even though most countries have passed laws against hacking both ordinary computers and the computers inside cell phones, enforcement is lax or nonexistent in most of the world. Many of the nations hit hardest so far by mobile malware outbreaks, such as Malaysia, Indonesia and the Philippines, do not always collect reliable and timely statistics that could be helpful for tracking software crimes.

For our part, my team and others in the security research community have been proactively studying Symbian and PocketPC, looking for vulnerabilities in the code and in the system designs that might afford entrée to malware. We hope to find these holes so that they can be patched before the bad guys exploit them in the inevitable next round of this constant battle. SA

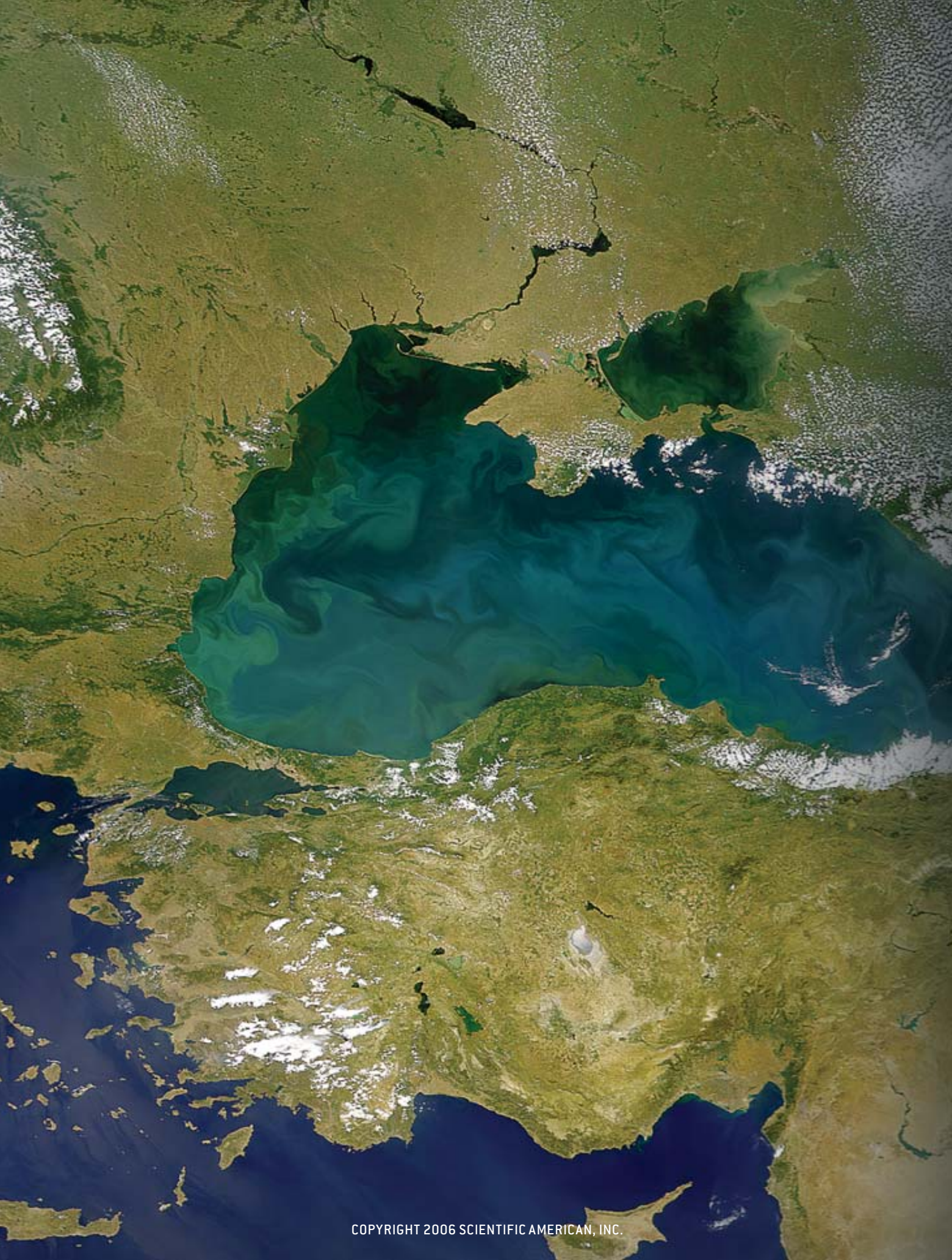
MORE TO EXPLORE

Mobile Phones as Computing Devices: The Viruses Are Coming! David Dagon, Tom Martin and Thad Starner in *IEEE Pervasive Computing,* Vol. 3, No. 4, pages 11–15; October–December 2004.

Mobile Phones: The Next Frontier for Hackers? Neal Leavitt in *Computer,* Vol. 38, No. 4, pages 20–23; April 2005.

Mikko Hypponen and his teammates blog at **www.f-secure.com/weblog/**

Trusted Computing Group: **www.trustedcomputinggroup.org/groups/mobile**



Reviving Dead Zones

BY LAURENCE MEE

How can we restore coastal seas ravaged by runaway plant and algae growth caused by human activities?

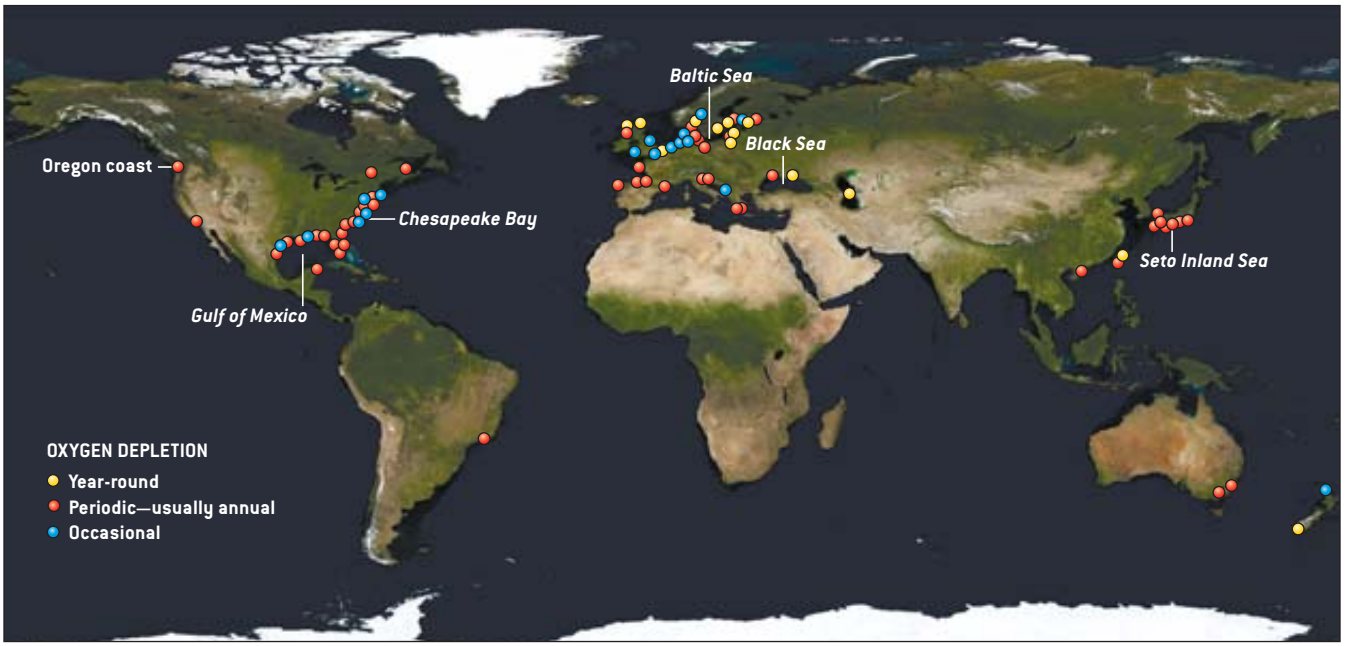
Imagine a beach crowded with vacationers enjoying the hot summer sun. As children paddle about in the shallows, foraging for shells and other treasures, dead and dying animals begin to wash ashore. First, a few struggling fish, then smelly masses of decaying crabs, clams, mussels and fish. Alerted by the kids' shocked cries, anxious parents rush to the water to pull their children away. Meanwhile, out on the horizon, frustrated commercial fishermen head for port on boats with empty nets and holds.

This scene does not come from a B horror movie. Incidents of this type actually occurred periodically at many Black Sea beach resorts in Romania and Ukraine in the 1970s and 1980s. During that period an estimated 60 million tons of bottom-living (or benthic) life perished from hypoxia—too little oxygen in the water for them to survive—in a swath of sea so oxygen-deprived that it could no longer support nonbacterial life.

At its extreme in 1990, the "dead zone," located in the northwestern part of the sea off the mouth of the Danube River, extended over an area the size of Switzerland (40,000 square kilometers). On the other side of the world, in the Gulf of Mexico off the Mississippi River delta, another huge dead zone appeared in the mid-1970s, which at its largest reached 21,000 square kilometers. In the past two decades, additional reports of dying or depleted areas in coastal seas and estuaries across the globe have emerged [*see map on next page*].

Determining the causes of this destruction, how it might be prevented and what to do to bring affected areas back to life has been a prime focus of my research since the early 1990s, when I published my first paper on the ecological crisis in the Black Sea. My work and that of others has now revealed important details of the events that degrade coastal ecosystems in many parts of the earth and has turned up new information that could help establish pathways to recovery.

BLACK SEA DEAD ZONE became evident when dying sea life began washing ashore near the mouth of the Danube River in the 1970s. Above, dead fish are left strewn across a Black Sea beach by the high tide. A satellite photograph of the western Black Sea from 2000 (*left*) shows the huge blooms of floating microscopic plants that resulted from the river's nutrient-rich discharge.



DEAD ZONES—areas starved of oxygen (by bacterial decay of overabundant plant growth) and thus of most animal life—occur in coastal seas, often near developed countries. The number of regions affected has doubled since 1990. Polluted runoff from the land often triggers dead zone conditions, although some result naturally. The dead zone in the northwestern Black Sea is now much smaller than it was at its largest a few decades ago.

Dead Zone Formation

OCEAN RESEARCHERS today link the creation of most dead zones to a phenomenon called eutrophication, the overenrichment of the sea by nutrients (principally compounds containing nitrogen and phosphorus) that promote plant growth. A certain amount of these "fertilizers" is essential to the health of phytoplankton—floating algae and other microscopic photosynthesizers that form the base for most marine food chains—as well as for the well-being of the sea grasses and algae that live on the floors of shallow, well-lit seas. But too much of these nutrients in illuminated waters greatly accelerates plant growth, leading to disruptive algal blooms and other unwanted effects.

Plants enter the food chain when tiny seaborne animals (zooplankton), herbivorous fish and filter-feeding bottom dwellers such as mussels and oysters graze them or when they die, decay, fall to the seabed, undergo bacterial decomposition and are incorporated into the underlying sediments. This organic bottom matter feeds the animals living there, including worms, shrimp and some fish.

Normally the numbers of phytoplankton are limited by the availability of light and nutrients and by grazing. But large increases in nitrogen and phosphorus concentrations enable these minute photosynthetic organisms to multiply in great profusion. The water eventually turns green or even brown as phytoplankton populations burgeon, and the shade they cast deprives plants living below them of essential sunlight. Sea grasses in shallow bays also become covered with small attached algae (epiphytes) and can ultimately be smothered and die. Algae can in addition envelop coral reefs, especially where heavy fishing pressure has thinned the ranks of resident grazers.

A major upsurge in the numbers of phytoplankton and epiphytes immediately causes difficulties for nearby sea life, but an even worse situation arises when oxygen levels in bottom waters decline. Lower oxygen concentrations appear when bacteria consume oxygen to break down the masses of organic matter that result from animal wastes and the dead bodies of organisms that multiply during eutrophication. Much of this material accumulates on the seafloor, where oxygen is relatively scarce to begin with.

Oxygen finds its way into the water from either photosynthesis or physical diffusion from the air at the sea surface. Should an area whose bottom is covered with dead plants also feature a strong density gradient that prevents mixing with the overlying water column, the oxygen at the bottom can soon become

Overview/*Coastal Seas in Trouble*

- River-borne plant nutrients from the land are killing off life in parts of shallow seas around the globe, resulting in so-called dead zones.
- Fertilizing chemicals cause microscopic plants floating near the surface to overgrow, depriving any plants living on the bottom of light and leading to large increases in the amount of decaying organic material falling to the seafloor. The bacteria living off the dead organisms use up seafloor oxygen, thereby causing the loss of most animal life there.
- Significant reductions in agricultural and sewage runoff, as well as controls on overfishing, can restore these key marine ecosystems.

THE VISIBLE EARTH/NASA [*satellite image*]; RADU MIHNEA [*dead fish*] [*preceding pages*]; THE VISIBLE EARTH/NASA [*satellite image*]; SOURCE: BASED ON *UNEP GEO YEARBOOK 2003* [*this page*]

exhausted, leading to the die-off of entire animal communities. (Such gradients can stem from temperature or salinity differences in the water at various depths.) This basic sequence—eutrophication leading to phytoplankton blooms, excess bacterial activity at the bottom, oxygen depletion, and the death of existing plants and animals—has occurred in almost every dead zone examined by researchers.

The details do vary, however, according to the local biological and physical conditions as well as the rate of supply of plant nutrients from the land. Poorly flushed estuaries, for example, are particularly vulnerable to the effects of eutrophication, because low water flows lead to slow replenishment of oxygen in bottom waters. This reduction in oxygen has been a persistent problem along the eastern seaboard of the U.S., where large estuaries, such as the Chesapeake Bay, have been affected.

The excess of nitrogen and phosphorus arriving in coastal seas results in large measure from the changing habits of people living in the areas draining into the sea. Rising fossil-fuel use (which releases nitrogen into the atmosphere), effluent from the mass breeding of food animals and intensive farming, and the construction of sewage systems that empty into bodies of water all lead to greater nutrient emissions into watersheds. The Millennium Ecosystem Assessment released by the United Nations in 2005 reported that the supply of nitrogen-containing compounds to the sea grew by 80 percent from 1860 to 1990. It predicted that the overall outflow to the oceans from human activities will increase by an additional 65 percent by midcentury. Dead zones are thus likely to become even more widespread unless society takes prompt action to reduce plant nutrient runoff.

Watery Graveyard

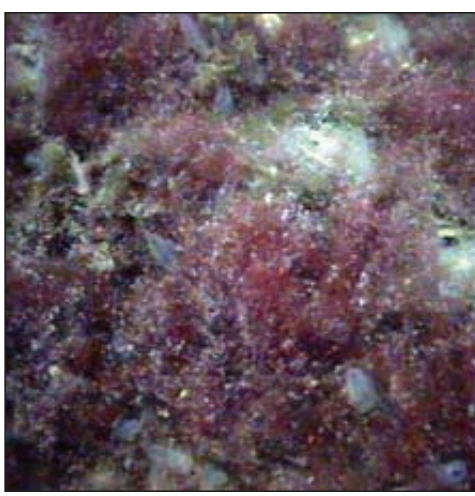
ALTHOUGH EMERGENCE of a dead zone is the final stage of the eutrophication process, marine systems, especially the animal populations, undergo changes long before then. A healthy coastal marine food chain often starts with silica-shelled phytoplankton called diatoms, which are consumed by copepods, minuscule zooplanktonic crustaceans. These animals, in turn, serve as food for fish. Increased nutrient concentrations affect the mix of phytoplankton species such that diatoms often become outnumbered by smaller or less digestible types. When eutrophication produces massive phytoplankton blooms, copepods often are unable to graze on the new, abundant phytoplankton species as well as the large quantities of organic detritus that result from the disruption of the natural ecosystem. This change favors the growth of highly tolerant gelatinous organisms such as *Noctiluca* (responsible for nighttime phosphorescence that occurs when the water surface is disturbed). Biologists sometimes call these jellyfishlike fauna "dead-end species" because higher-level predators have difficulty living off them. Their presence reduces the efficiency of the food chain, leading fish stocks to wane.

Such an imbalance in the food chain can be worsened by intense commercial fishing, particularly where these efforts target high-value "top predator" species such as cod, hake, dorado or horse mackerel. Loss of apex fish species leads to rises in the numbers of small prey fish, which results in fewer zooplankton (the food of the small fish) and, consequently, even more phytoplankton. Scientists call this sequential process "trophic cascading." An inefficient food chain engenders more organic matter on the seafloor, which enhances the risk that a dead zone will follow.

Ecosystems altered by eutrophication also become more vulnerable to invasion by foreign species, which can arrive, for example, in the ballast discharge from transoceanic ships. In the 1980s the comb jellyfish *Mnemiopsis leidyi,* which probably originated off the eastern coast of the U.S., took up residence in the Black Sea. By 1990 this voracious dead-end predator dominated the ecosystem completely, at its maximum attaining an astounding density of up to five kilograms per square meter.

Sometimes shellfish reefs can help stave off degradation of an ecosystem. In many estuaries on the eastern seaboard of the U.S., oysters act as ecosystem engineers by accumulating into huge reefs rising several meters from the seabed; these reefs support a diverse assemblage of organisms, including flounder, snapper, silver perch and blue crabs.

Hunter Lenihan of the University of California, Santa Barbara, and Charles H. Peterson of the University of North Carolina at Chapel Hill have shown, for example, that the tops of oyster reefs in North Carolina's Neuse River became refuges for displaced bottom-water species at the onset of dead zone formation because they projected above the deoxygenated water layer. Mechanical oyster



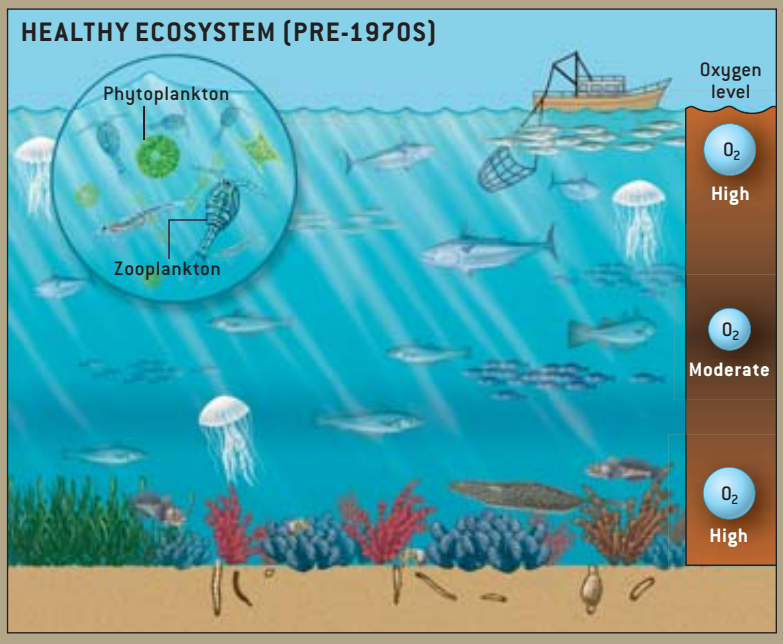
DEAD—AND RECOVERED—BOTTOM LIFE are apparent in these views of two locations on the Black Sea bed earlier this year. The photograph on the left shows a largely depleted area that is covered with the shells of mussels killed by a lack of dissolved oxygen. On the right is a recovering site with a thick and diverse cover of algae and large numbers of ascidians (sea squirts).

TIM STEVENS

KEY STAGES IN THE FORMATION OF A DEAD ZONE

The events underlying the creation of the Black Sea dead zone were fairly typical of how similar oxygen-depleted (hypoxic) areas form, although the details vary from case to case. At the root was eutrophication, an excessive inflow of plant nutrients that resulted in an overgrowth of algae and other small floating photosynthetic plants, which led indirectly to hypoxia and to the death of plants and animals in the lower depths.

The decline of the ecosystem followed three stages described initially by Tatsuki Nagai of Japan's Fisheries Research Agency, who studied one of the first known hypoxic regions. These conditions appeared in Japan's Seto Inland Sea in the early 1960s. He called the natural state the "sea of red bream" (a predator species targeted by local fishers). Then came the "sea of anchovies," when predator species declined, leaving mainly small prey fish behind. Finally, there came a "sea of jellyfish," in which most other species died or fled, leaving highly tolerant invading species dominant. Nagai also was among the first to identify overfishing as a contributor to degradation of the sea's food chain (by removing the top predator fish).



The near-surface coastal waters of the northwestern Black Sea initially contained a diverse mix of phytoplankton (floating algae and other microscopic plants) as well as varied fish and other organisms. Shallow waters close to shore featured shoals of immature anchovies, mackerel and bonito. At the middle depths lived large schools of top predators, such as whiting, and numerous prey fish, plus some jellyfish. At the bottom, communities of mussels as well as gobies, turbot, sturgeon and hermit crabs thrived among extensive meadows of sea grasses and brown and red algae.

harvesting frequently shortens the height of these reefs, however, which helps to destroy the natural resilience of these ecosystems.

Black Sea Catastrophe

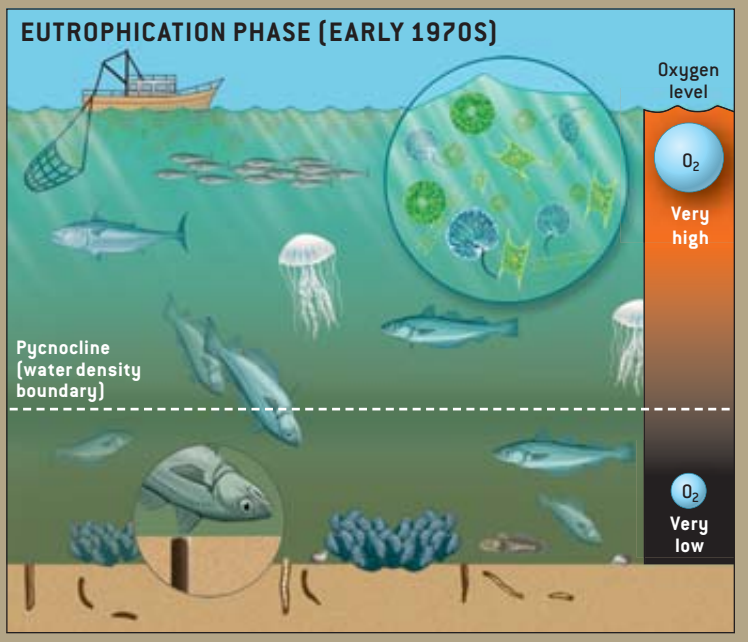
THE BLACK SEA offers a dramatic example of how undersea ecosystems can be destroyed by a surplus of nutrients and also provides insight into how they can be resuscitated. The waters of the northwestern area of the sea fell prey to eutrophication when flows of nitrogen and phosphorus compounds from the land doubled between the 1960s and 1980s. The principal pipeline for these chemicals was the Danube, which drains much of the watersheds in 11 countries across central Europe, from Germany to Romania. The main culprits were agricultural runoff, urban and industrial wastewater, and, in the case of nitrogen compounds, atmospheric transport. At least half the increased nitrogen poured into the Black Sea resulted from modernized farming practices, including intensive use of fertilizers and the establishment of huge animal production facilities. These agricultural activities also contributed to the rise in the phosphorus effluent, but industrial and urban waste discharges laden with polyphosphate detergents played an even more significant role.

Before the 1960s the shallow northwestern region of the Black Sea was a diverse and highly productive system that included extensive near-shore mats of bottom-living brown algae and, further offshore, the largest red algae community on earth—a field of *Phyllophora* the size of the Netherlands. These natural algae meadows coexisted with enormous beds of mussels and other bivalves, and the whole system supported a large number of invertebrate and fish species. The algae helped to oxygenate the bottom waters, and the mussels filtered the seawater, thereby maintaining good light conditions for photosynthesis. This ecosystem was highly resilient, being able to accommodate large variations in climate conditions and natural disturbances. As nutrient effluents escalated, however, dense phytoplankton blooms appeared in the surface waters. Such luxuriant growth lowered water transparency, which in turn deprived the bottom algae of light

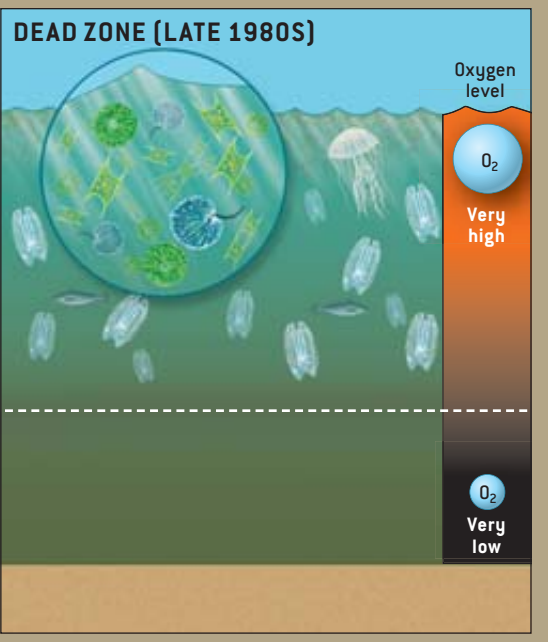
THE AUTHOR

LAURENCE MEE is director of the Marine Institute at the University of Plymouth in England. He also heads the university's interdisciplinary Marine and Coastal Policy Research Group. An oceanographer with a Ph.D. from the University of Liverpool, Mee has held research positions at the Institute of Marine and Limnological Sciences in Mexico and the IAEA Marine Environment Laboratory in Monaco and has coordinated the United Nations Global Environmental Facility–Black Sea Environmental Program. He became a Pew Fellow in Marine Conservation in 1998. Mee's current work focuses on ways to protect the marine environment and the associated drainage basins and coastal areas.

SHAWN GOULD



With growing influxes of the nutrients nitrogen and phosphorus from terrestrial runoff, the ecology of the Black Sea coastal region began to change. Massive phytoplankton blooms turned the water green or even brown, depriving plants living below of sunlight and depositing a steady stream of decaying organic matter in the bottom. Bacteria on the seabed then consumed great amounts of oxygen as they feasted on the organic matter and the dead plants, thus producing hypoxia on the seafloor, which killed many organisms.



Eventually, increased shading and rampant hypoxia on the seabed finally left the bottom waters devoid of life. Heavy fishing had already reduced the numbers of predator fish species, and ultimately these and most other large animals disappeared from the area. Opportunistic invading species, notably the comb jellyfish *Mnemiopsis leidyi,* multiplied in great numbers in the upper depths.

and eventually led to their loss, which altered the entire natural ecosystem.

During the summer months, when the water column became stratified, oxygen levels, particularly near the seabed, began to fall. Many of the affected bivalve communities could survive hypoxia for as long as 20 days by closing their shells and living on internal reserves of glycogen—an animal's chief energy-storage carbohydrate. But when these supplies were exhausted, the mollusks died en masse, which led bacteria and other organisms to consume the remaining local oxygen as they degraded the dead animals and to release even more plant nutrients. By the time nearly all the oxygen was gone, all the fauna that normally lived in the area had either migrated away in search of food and oxygen or perished. The region's natural ecosystem was thus seriously degraded.

The area began to recover only when the communist regimes in eastern Europe fell at the end of 1989, ending central economic planning. Suddenly farmers there had little capital to buy fertilizer, so agricultural activities slowed. Likewise, many giant animal farms closed, thus profoundly reducing nutrient runoff. One former pork production facility in Romania with more than a million pigs had generated the equivalent emissions of a city with five million inhabitants.

Within six years the profound drop-off in nutrient influx led to shrinkage of the dead zone [*see box on next page*]. Recovery of the sea bottom, however, was gradual. For example, studies by my Ukrainian colleagues have shown that mussel beds in devastated parts of the northwestern shelf became firmly reestablished only in 2002, many years after other communities had recovered substantially. This past August a research expedition we conducted to examine the state of the sea revealed major reestablishment of benthic algae communities, though generally not the same species that had dominated before the onset of dead zone conditions.

Long Road to Recovery

CLEARLY, RESTORING dead zones requires, at a minimum, reducing nutrient delivery from nearby lands. But marine ecosystems that have collapsed because of eutrophication and hypoxia may not simply bounce back when humans alter their activities to lower the amounts of plant nutrients reaching rivers. This resistance to recovery occurs for three reasons.

River catchments typically possess a huge capacity for storing nutrients—either dissolved in groundwater or adsorbed on soil particles. Years or even decades may pass before nitrogen and phosphorus fertilizers and other chemicals stop leaching out and passing to the sea. Nitrogen compounds in particular tend to accumulate in groundwater.

Dead zones can also linger if there is a dearth of nearby healthy populations

of marine plants and animals that can provide the "seed stock" from which the missing communities can be restored. Indeed, the flora and fauna that once lived in the affected area may even have gone extinct. It is possible for native marine animals to drift large distances as larvae from healthy ecosystems and to eventually reestablish themselves in a suitable vacant biological niche. Sometimes, however, these would-be returnee species find themselves supplanted by opportunistic invading organisms that have taken up all suitable habitats.

Finally, eutrophication often causes alterations in ecosystem composition that are not easily reversed [*see illustration on opposite page*]. As nutrient concentrations begin to climb early on, some species decline, but ecosystems as a whole may remain strong for a long while if the natural populations can withstand a relatively high amount of phytoplankton growth and the like. At some point, however, a threshold is reached at which the loss of key species yields an abrupt collapse to a new degraded state. This new equilibrium arises from the presence of some remaining species that are tolerant of eutrophication's effects and from the arrival of opportunistic creatures from elsewhere. Unfortunately, the new state is often quite stable. Consequently, simply cutting back the nutrient supply to its pre-eutrophication levels may not restore the original ecosystem; lowering nutrient concentrations to levels much below their starting points may be required.

To further complicate matters, the threshold for change from a natural state to a degraded one typically comes earlier if an ecosystem's resilience is diminished by overfishing. Therefore, it may also be necessary to reduce fishing activities markedly before the healthy state can return. If the species of the original system have been lost or invaders have appeared, the former pristine conditions may never be regained.

Eliminating Dead Zones

KNOWING WHAT TO DO to fix dead zones will not be enough; a key to reviving them is for governments to believe it is an important goal and to take the lead. Indeed, scientists have documented few cases of dead zone recovery, because reducing nutrient runoff from the land requires major changes in agricultural practices and wastewater treatment efforts. Most such programs have so far achieved only partial cuts in terrestrial nutrient outflows.

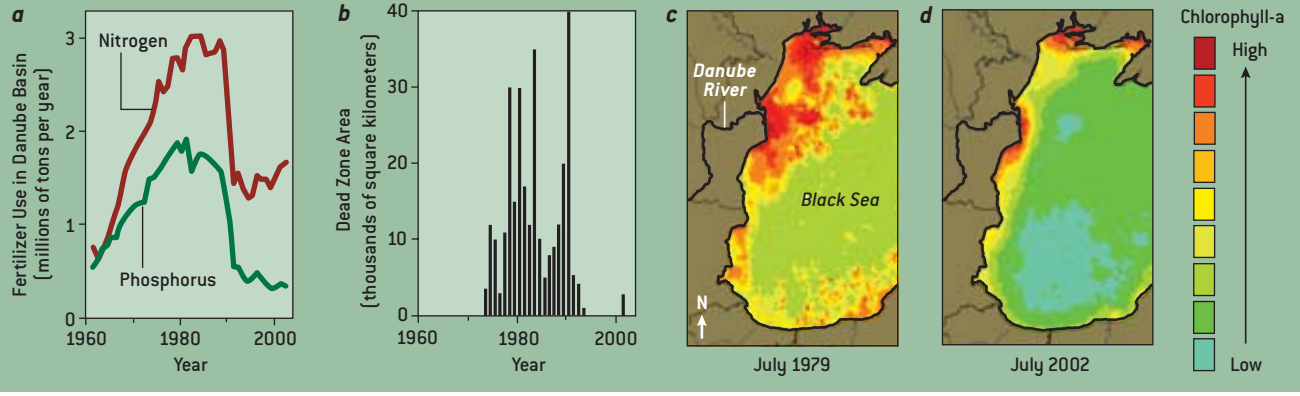
To shrink nutrient loads, comprehensive plans (at the scale of an entire river catchment system) must be put in place that keep nitrogen and phosphorus on the land and out of the water. Such efforts are currently under way in the Chesapeake Bay and in the Black Sea. In the latter case, surrounding governments, aided by the United Nations Global Environment Facility, have agreed to pursue a landmark initiative to maintain nutrient runoff levels at those of the mid-1990s, a scheme that seems to be aiding recovery through pilot-scale projects to improve farming practices and waste treatment.

Two serious problems must be over-

The Black Sea Comes Back

The recovery of the Black Sea dead zone underscores the need to reduce agricultural, sewage and other nutrient runoff from the land if affected areas are to be restored to health. The dead zone adjacent to the northwest coast of the Black Sea began to revive only after the communist system collapsed in 1989, which prevented the continuation of intensive farming—including large-scale raising of livestock and heavy application of fertilizers containing nitrogen and phosphorus (*a*)—that had been in place since the 1960s. Nutrient residues made their way into the Danube River and other watersheds and eventually down into the Black Sea, which caused the dead zone to appear in 1973 and to return in the summer for the next 21 years (*b*). Red color in a satellite image from 1979 (*c*), for instance, clearly reveals a large expanse of overfertilized water. (In that image and in *d*, eutrophication was assessed by determining the concentrations of chlorophyll-a, an indicator of plant growth, in surface waters.)

Within five years after the intensive farming ended, the degraded region had returned to life (*b* and *d*), relapsing only during the exceptionally hot summer of 2001. By 2002 mussel communities in the area had reestablished themselves. The sea may again be at risk, however, as central European economies recover and agriculture there begins to intensify again.



CZCS/SeaWiFS DATA PROCESSED BY ABIGAIL McQUATTERS-GOLLOP (*c* and *d*)

come, however, before full and sustainable recovery of the Black Sea ecosystem can occur. European authorities have to take steps to ensure that renewed economic development does not lead to the resurgence of terrestrial nutrient release to the sea. They should, for instance, invest in major waste-reduction projects that employ the latest technology. This point is particularly important for the Danube basin, where six of the countries have joined or are in the process of joining the European Union. Some farmers from western Europe, where intensive agricultural practices have often caused eutrophic rivers and coastal waters, are eager to buy up central European farmland.

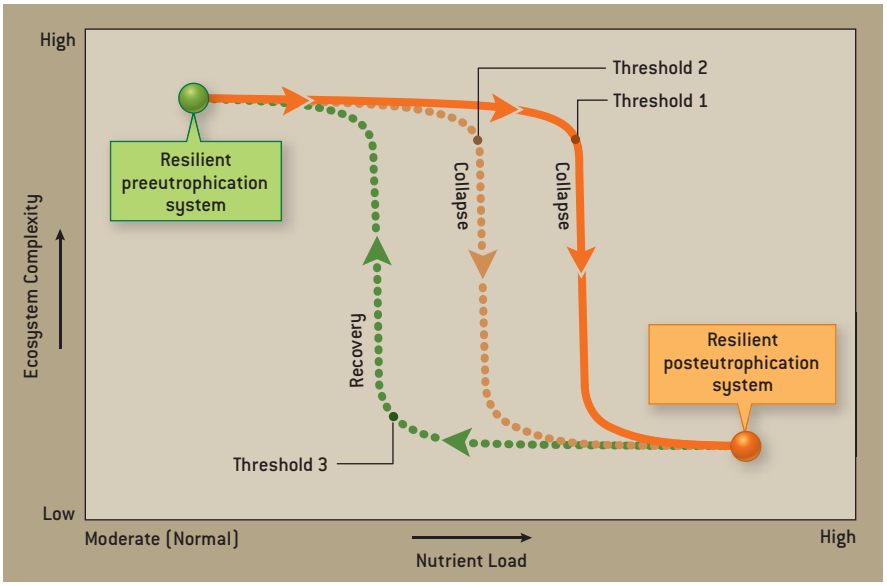
Next governments must reduce the intensity of commercial fishing activities enough to allow depleted stocks of piscine predators to recover. In addition, the trawls and dredges on fishing boats destroy key benthic communities and must be regulated more effectively.

In fact, maritime nations worldwide must work to lessen fishing pressure on eutrophic areas, difficult as that is to achieve with more than half the planet's fisheries being overexploited today. Although an international agreement has been reached to establish by 2012 a global network of marine protected areas—which would help curb overfishing and save the vital seed stock needed for dead zone recovery—the agreement's goals are unlikely to be met, because enforcement mechanisms are lacking.

Even if a eutrophic ecosystem makes a partial comeback, authorities must be aware that a partial recovery may leave it in a highly unstable situation. Mussels, for example, have an extraordinary capacity for filtering water, and the establishment of mussel beds on artificial reefs has been employed to improve water quality. But bacterial decomposition of mussel excreta and dead individuals consumes large amounts of oxygen, which can result in boom-and-bust cycles in locations where water mixing is poor and oxygen replenishment is limited. In these cases, thriving mussel communities suddenly collapse, leaving a dead zone until the organic material has fully decayed and recovery begins again. Scientists have observed this phenomenon in Baltic Sea estuaries. The challenge for marine resource managers is to maintain conditions that sustain resilient and diverse systems—even where full recovery is no longer possible.

On a more subtle level, the entire concept of rating the health or quality of an ecosystem depends on the values of local inhabitants. For some, the desirable outcome of remedial action might be a sea of small prey fish; for others, only restoration of a sea filled with apex predators will be acceptable.

Coastal dead zones remind us that humanity cannot simply expect natural ecosystems to absorb our wastes without severe and often unexpected consequences. We now know how to revive dead zones, but ultimately the steps required to do so depend on our recognition of the ramifications to the environment of waste disposal and the degree to which we value the world's marine ecosystems. SA



REDUCING NUTRIENT LEVELS to those present before a dead zone has formed may not be enough to ensure recovery, as shown in this graph, which relates the health of an ecosystem (in terms of complexity, or species diversity) and the amount of nutrients with which it must contend. A system with high complexity and moderate terrestrial nutrient inflows tends to be highly resilient until the nutrient load finally surpasses some level (*threshold 1*), causing the system to collapse to a much less complex state. This breaking point arrives earlier (*threshold 2*) if overfishing has depleted the numbers of top predator fish, which reduces species diversity. Unfortunately, the new degraded state is also typically resistant to change and may recover its lost complexity only when nutrient influxes drop significantly below the starting levels (*threshold 3*). Even then, an ecosystem may never return to its previous state if key species have gone extinct.

MORE TO EXPLORE

Marine Benthic Hypoxia: A Review of Its Ecological Effects and the Behavioral Responses of Benthic Macrofauna. R. J. Diaz and R. Rosenberg in *Oceanography and Marine Biology: An Annual Review,* Vol. 33, pages 245–303; 1995.

National Estuarine Eutrophication Assessment: Effects of Nutrient Enrichment in the Nation's Estuaries. S. B. Bricker, C. G. Clement, D. E. Pirhalla, S. P. Orlando and D.R.G. Farrow. NOAA, National Ocean Service, Special Projects Office and the National Centers for Coastal Ocean Science, 1999.

Nutrient-Enhanced Productivity in the Northern Gulf of Mexico: Past, Present and Future. N. N. Rabelais, R. E. Turner, Q. Dortch, D. Justic, V. J. Bierman and W. J. Wiseman in *Hydrobiologia,* Vol. 475, No. 6, pages 39–63; 2002.

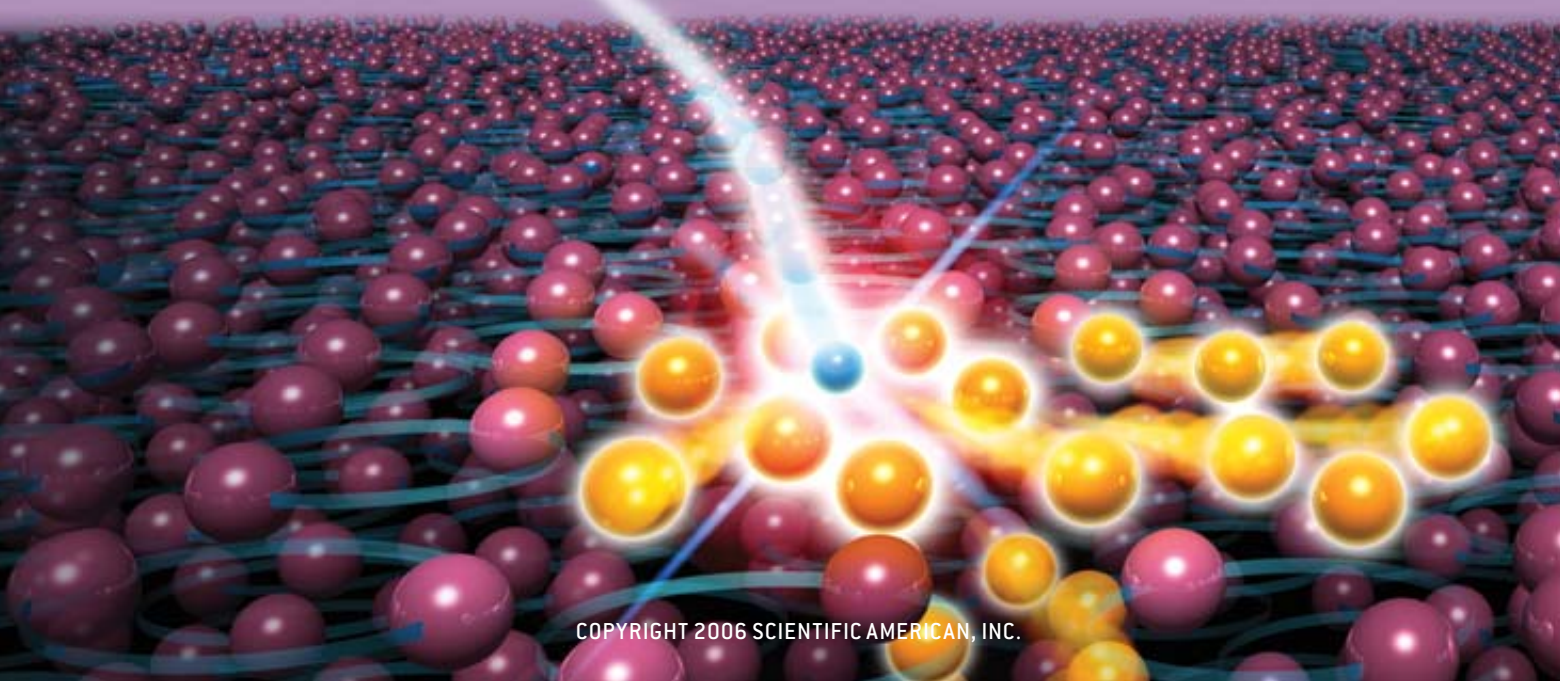
Ecosystems and Human Well-Being: Current State and Trends. Millennium Ecosystem Assessment. Island Press, 2005. Available online from **www.millenniumassessment.org/en/products.global.overview.aspx**

Restoring the Black Sea in Times of Uncertainty. L. D. Mee, J. Friedrich and M. T. Gomoiu in *Oceanography,* Vol. 18, pages 32–43; 2005.

JEN CHRISTIANSEN

SEEING *with* Superconductors

Tiny devices made of superconducting material that act as superb sensors of photons and other particles are revolutionizing a wide range of research and technology fields

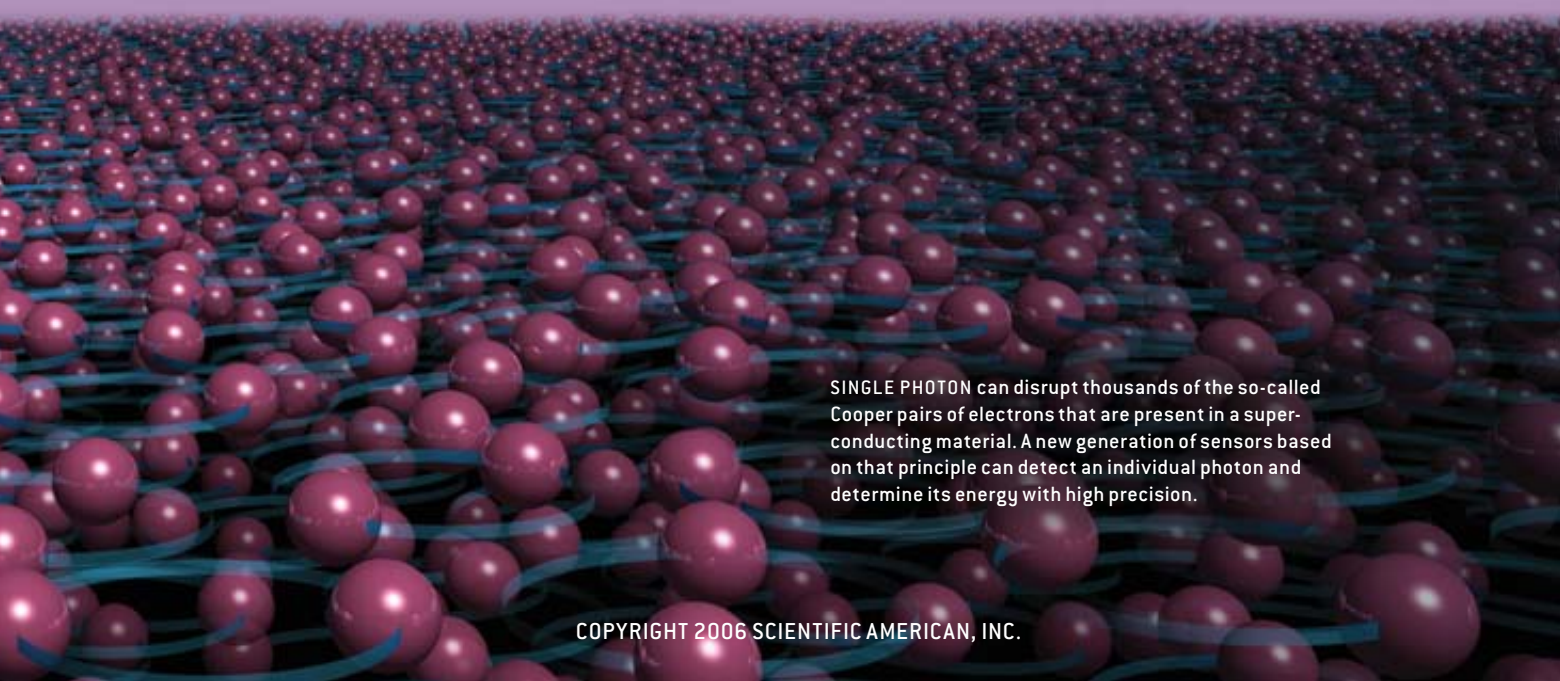


By Kent D. Irwin

Your eyes are exquisite light detectors, determining the intensity, color and spatial distribution of the rays incident on them. The human retina has more "pixels" than a consumer digital camera, containing about six million color-sensing cone cells and more than 100 million of the rod cells responsible for vision in the dark. And eyes are highly sensitive: a dark-adapted rod cell can fire off a signal to the brain on absorbing a single particle of light, or photon, the smallest quantum unit of an electromagnetic wave. As few as six of these single-photon signals are required for your brain to perceive a flash. But eyes and commercial cameras are far from ideal for many tasks, because they can detect only those photons whose frequencies lie in the narrow visible range. Furthermore, their color capabilities do not involve a measurement of each photon's precise frequency.

Scientific and industrial photon detectors, in contrast, peer into the electromagnetic realms beyond that of visible light—into the low-frequency (long-wavelength, low-energy) world of infrared and microwaves and into the high-frequency regime of x-rays and gamma rays. Yet they too are limited in their abilities. In particular, for visible and longer wavelengths scientists have lacked a detector able to "see" an individual photon and discern its frequency, and thus its energy, with any accuracy. Determining the frequency of photons opens the door to a wealth of information about the matter that emitted the photons.

A revolution in photon detection is now under way, with the advent of detectors based on superconductivity that are capable of such fine measurements and other prodigious feats. These new tools are dramatically improving the sensitivity of measurements across the electromagnetic spectrum, from radio waves through visible light to gamma rays. Improved devices for measuring the polarization of microwaves will soon probe the first moments of the universe by measuring the pattern that gravity waves from the big bang imprinted on the cosmic microwave background. Detectors capable of counting individual visible photons are improving the security of quantum communication. At synchrotrons, superconducting x-ray detectors are being used to study the chemical composition of materials. And researchers are developing gamma-ray detectors that can do a more discriminating job of identifying nuclear materials to stop them from being stolen or smuggled across international borders.



SINGLE PHOTON can disrupt thousands of the so-called Cooper pairs of electrons that are present in a superconducting material. A new generation of sensors based on that principle can detect an individual photon and determine its energy with high precision.

Beyond detecting photons, superconducting devices are also sensing biological polymers and searching for the weakly interacting particles that make up the mysterious dark matter constituting five sixths of the matter in the universe. Superconducting detectors are just beginning to fulfill their scientific and commercial potential.

Useful Fragility

ODDLY ENOUGH, a property of superconductors that has restricted their use in such applications as power transmission is exactly the feature that renders them valuable as photon detectors. Superconductivity, the flow of electric current without resistance, arises when electrons in a suitable material bind together in what are called Cooper pairs, which then flow en masse like a superfluid. A fragile quantum-mechanical effect, superconductivity occurs in a material only when it is cooled below an extremely low temperature, termed its critical transition temperature. Cooling a material reduces the vibrations of its atoms. If the temperature rises above that threshold, the thermal energy of the vibrations knocks apart the partners in the Cooper pairs, eliminating the superconductivity.

Because of this sensitivity to heat, many superconducting devices have to be refrigerated to within a few degrees above absolute zero (0 kelvins, equal to –273.15 degrees Celsius or –459.69 degrees Fahrenheit). Some types require temperatures as low as a few hundredths of a kelvin. These extreme temperatures can be achieved with commercially available refrigerators that use liquid helium or a process called adiabatic demagnetization, but the need for such cooling prohibits many applications. Scientists have struggled for years to develop materials with more robust superconductivity that survives at higher temperatures.

The very fragility of superconductivity is, however, the feature that makes it ideally suited for use in sensitive detectors. Superconducting photon detectors rely on the ability of the energy in a single photon to disrupt thousands of Cooper pairs. The change in the superconducting state can then be measured in a number of ways to reveal the energy deposited by the photon. Because a photon's energy is proportional to its frequency, this measure also indicates the photon's frequency, which is the key to learning about the object that the photon came from.

Room-temperature semiconductor detectors, such as the charge-coupled device (CCD) in a digital camera, also work by disturbing a quantum state in a material. In the case of a CCD, a photon of visible light knocks an electron out of an energy band in a semiconductor crystal. But electrons in these bands are so strongly bound that usually only a single electron is freed by each photon. That release is too coarse to reveal the frequency of the photon. As a result, a CCD cannot measure the color of a photon directly—digital cameras produce color images by using a system of red, green and blue filters that let through photons in ranges of frequencies.

In contrast, one visible photon can break thousands of Cooper pairs in a superconductor. The creation of thousands of excitations allows an accurate measurement of energy, much as an exit poll at an election is more accurate when thousands of people are interviewed.

The fragility of superconductivity is the feature that makes it IDEALLY SUITED for sensitive detectors.

Detectors that work by sensing the disruption of superconductivity come in two main classes. The thermal type is cooled right to its transition temperature, at which point it is only partly in the superconducting state and thermal excitations are on the verge of destroying superconductivity entirely. Any energy deposited into the superconductor increases its temperature and causes its electrical resistance to rise significantly. The pair-breaking detector type, in contrast, is cooled to well below the transition temperature and is fully in the superconducting state. This detector measures the number of Cooper pairs broken when energy is deposited into the superconductor.

One other variety of superconducting device must be mentioned for completeness: a superconducting "mixer" works as a kind of amplifier of lower-frequency electromagnetic signals such as microwaves. Because such devices do not employ the fragility of superconductivity, further discussion of them lies beyond the scope of this article.

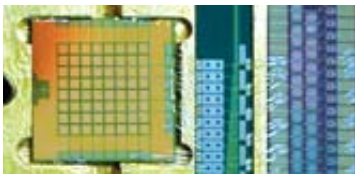
The thermal type of detector relies on the fact that the electrical resistance of a superconductor rises sharply from zero to its normal value in the very narrow

Overview/*Superconducting Detectors*

- Sensors capable of detecting an individual particle of light (a photon) and determining its energy or sensitively measuring a many-photon signal have myriad scientific and technological applications, including homeland security (detection of materials that could be used in a nuclear weapon), analysis of defects in microchips, astronomy, chemical analysis and particle physics.
- A revolutionary new generation of such sensors are based on the properties of superconductors and can measure a photon's energy with much greater precision. The new sensors also allow images to be gathered much faster.
- The new detectors come in two types. The type called thermal sensors depends on how a photon's energy increases the temperature of the detector material. The other kind, called pair-breaking detectors, senses how a photon disrupts some of the pairs of electrons that cause the superconductivity.

SLIM FILMS (*preceding pages*)

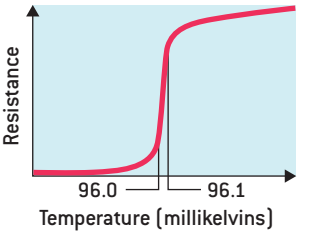
HOW A SUPERCONDUCTOR DETECTS LIGHT

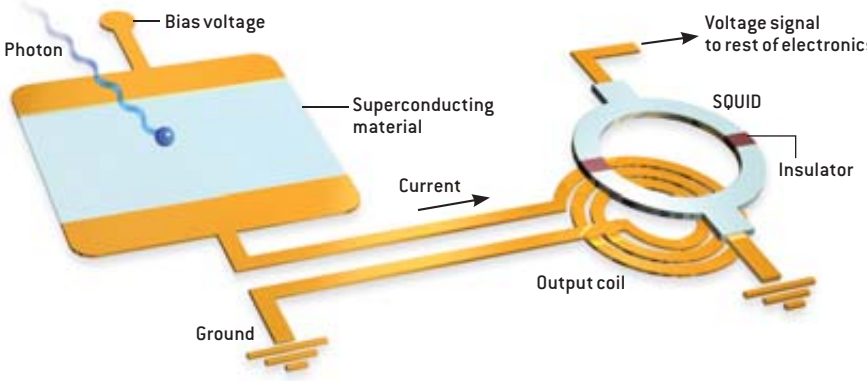


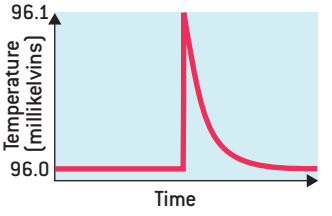
Superconducting photon detectors come in two main types, thermal and pair-breaking. With both types, the energy of an individual photon (and hence its frequency) is revealed by the strength of the device's output signal. At the far left is a 64-pixel array of thermal sensors that is used simultaneously as an x-ray camera and high-resolution x-ray spectrometer. The circuitry at the right side of the image is a SQUID multiplexer (*described at bottom*).

THERMAL DETECTORS

The most common kind of superconducting thermal detector is the transition-edge sensor, which is cooled to the middle of the very narrow temperature range over which its active material changes from superconducting to normal (*upper graph*). A bias voltage stabilizes it at that transition temperature and drives a steady current flow. On absorbing a photon, the sensor briefly heats up by a tiny amount (*lower graph*), which results in a significant increase in the resistance of the sensor. A superconducting device known as a SQUID detects the corresponding momentary drop in the current and converts this signal into a voltage pulse that can be further amplified by conventional electronics before data collection.

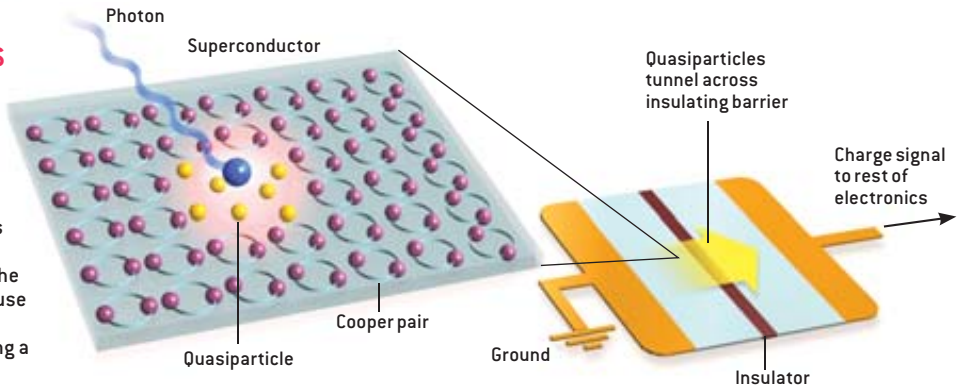






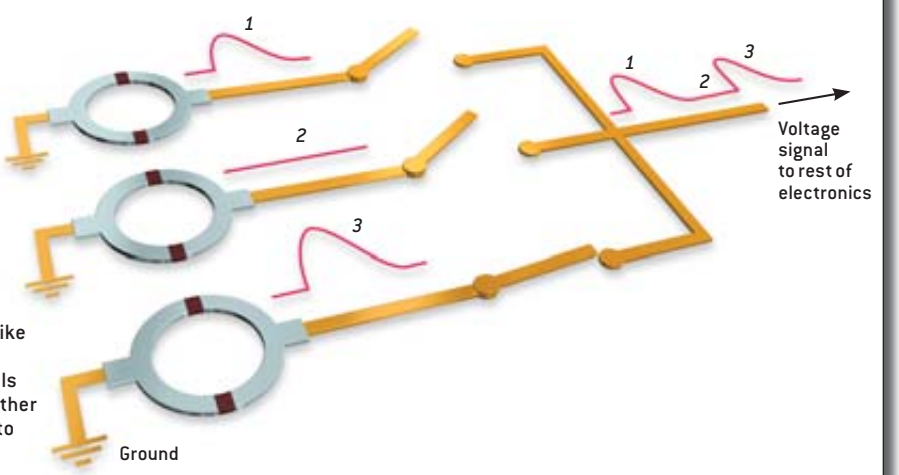
PAIR-BREAKING DETECTORS

The other type of sensor is cooled completely into its superconducting phase, in which its electrons form Cooper pairs. A single photon has enough energy to disrupt thousands of Cooper pairs, breaking them into excited electrons known as quasiparticles (*left*). Although the material remains superconducting, the quasiparticles can be detected because they, unlike the Cooper pairs, tunnel across an insulating barrier, producing a pulse of charge that is passed to conventional electronics (*right*).



CONVEYING AN IMAGE BY MULTIPLEXING

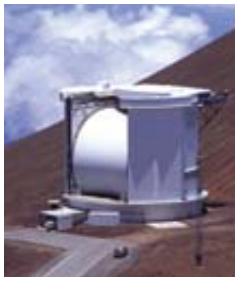
Imaging is carried out by large arrays of detectors, but all the output signals coming from the detectors must be combined into a smaller number of data lines, a process known as multiplexing. For example, in what is called time-division multiplexing of transition-edge sensors, the outputs of a number of SQUIDs are connected to a single data line. The SQUIDs act like switches that are closed sequentially one at a time, converting the detectors' pattern of signals (*1–3*) into a sequence of pulses. Electronics further downstream identify which pulses correspond to each detector by their precise timing.

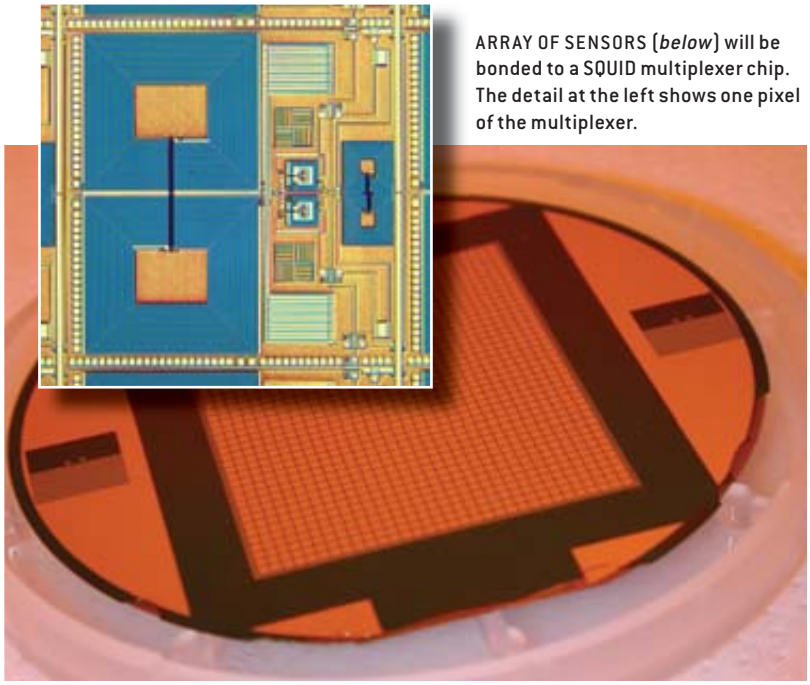


SLIM FILMS (*illustrations*); DANIELA NAOMI MOLNAR (*graphs*); KENT D. IRWIN (*photograph*)

SUBMILLIMETER ASTRONOMY

Astronomical observations at wavelengths somewhat shorter than a millimeter allow studies of regions inside molecular clouds that are hidden at other wavelengths; such clouds play important roles in planet-, star- and galaxy-forming regions. This problematic wavelength range will become much more accessible with the development of the SCUBA-2 camera, which will be installed in the James Clerk Maxwell Telescope (*right*) on Mauna Kea in Hawaii in 2007. The camera has eight arrays of 1,280 transition-edge sensors.





ARRAY OF SENSORS (*below*) will be bonded to a SQUID multiplexer chip. The detail at the left shows one pixel of the multiplexer.

temperature range in which the material switches from its superconducting to its normal state [*see box on preceding page*]. The abrupt change in resistance allows the superconductor to act as an ultrasensitive thermometer. A detector that uses the superconducting phase transition this way is called a transition-edge sensor (TES). When a TES absorbs a photon, the photon's energy gets converted into heat, which increases the temperature and thus the resistance of the material in proportion to the deposited energy. Depending on the photon-absorbing material used, a TES can serve as a spectrometer for measuring the energy of x-rays and gamma rays, a photon counter at near-infrared to visible wavelengths, or a total power detector for radiation at infrared and millimeter-wave bands.

The first TES detectors were developed in the 1940s but were impractical for many years. The problem was that the superconducting transition is often less than a thousandth of a degree wide, and it was very hard to keep the temperature of the device within that range. While I was a graduate student with Blas Cabrera at Stanford University, our research group was developing arrays of TES detectors for experiments aimed at studying neutrinos from nuclear reactors and detecting dark matter. We were able to get a few TES detectors to work, but small variations in the transition temperature of different sensors made it impossible to operate an array of sensors at the same temperature.

In 1993 I realized that a simple trick would solve this problem—apply a constant voltage across the detectors, a technique called voltage biasing. The applied voltage results in an electric current through the TES, which heats the detector. When the transition temperature is reached, the resistance goes up, decreasing the current and stopping the heating. The self-heating thus acts as a negative feedback that tends to keep the temperature of the film within its transition. In an array of voltage-biased sensors, each sensor will self-heat into its own transition, even if the transition temperatures vary slightly. The negative feedback also speeds the response of the detectors. The introduction of voltage biasing has led to an explosive growth in the development of TES detector arrays worldwide.

Counting Quasiparticles

UNLIKE A THERMAL SENSOR, a superconducting pair-breaking detector cannot rely on a change of electrical resistance to signal the absorption of a photon. An incoming photon breaks Cooper pairs and creates quasiparticles, which for most purposes can be thought of as free electrons in an otherwise superconducting metal. The number of quasiparticles created is proportional to the photon's energy. But because the detector is cooled to well below its transition temperature, a sea of intact Cooper pairs is still present, so the electrical resistance remains zero. A superconducting pair-breaking detector therefore must be able to distinguish between Cooper pairs and quasiparticles.

One device capable of that feat is the superconducting tunnel junction, which consists of two superconducting films separated by a thin layer of insulating material. If the insulator is thin enough (about two nanometers), electrons can cross from one side of the barrier to the other by a process known as quantum-mechanical tunneling. Applying a small magnetic field will stop Cooper pairs from tunneling across the junction, so only quasiparticles can cross. Now a voltage can be applied to the device and no current will flow until one of the superconducting films absorbs a photon, generating quasiparticles. The resulting current pulse is proportional to the num-

NIK SZYMANEK (*telescope*); GENE HILTON *NIST* (*chips*)

ber of quasiparticles created and thus to the energy and frequency of the photon.

A team led by Jonas Zmuidzinas of the California Institute of Technology and Peter Day of the Jet Propulsion Laboratory is developing another device to measure the number of quasiparticles in a superconductor. Called a microwave kinetic inductance detector, it takes advantage of the fact that a superconducting structure can have an electromagnetic resonance at a microwave frequency, much as a tuning fork has a mechanical resonance at an audio frequency. (Whereas the tuning fork vibrates, the superconductor admits an oscillating electric current.) When photons create quasiparticles in the superconductor, the resonance becomes less sharp and wave propagation slows down, reducing the resonant frequency. The shifts in both the resonant frequency and the sharpness of the resonance are proportional to the number of quasiparticles. Initial results with these devices are extremely promising.

Single superconducting detectors are useful for some applications, such as analyzing materials, but practical imaging requires a large array of detectors, or pixels, similar to a CCD array in a digital camera. A problem arises in connecting the ultracold detector array to its associated room-temperature electronics: if one simply runs a wire from every pixel, too much heat flows down the wires, undermining the superconductivity of the detectors. A better scheme is to use circuitry within the ultracold part of the device so that the signals from many pixels are combined, or multiplexed, into a few wires that lead from there to the warm electronics.

The most advanced multiplexed superconducting arrays available are based on TES detectors. As discussed above, when a TES absorbs a photon the current through the detector changes. The current, however, is very small, so it is measured by connecting each pixel to a superconducting quantum interference device, or SQUID, whose sensitivity is limited only by quantum mechanics [see "SQUIDs," by John Clarke; Scientific American, August 1994]. The SQUID changes the tiny current pulse into a voltage signal that is large enough for conventional electronics to measure. The outputs from many SQUIDs can be multiplexed by adding their voltages and feeding the result into a single wire, but something further must be done so that the signals from each individual detector can be distinguished. In time-division multiplexing the SQUIDs are turned on one at a time, whereas in frequency-division multiplexing the detectors operate at different frequencies, which allows their signals to be separated later.

Pixels in a microwave kinetic inductance detector (the type that oscillates like a tuning fork) can also be multiplexed by tuning them to different resonant frequencies, connecting them all in parallel, and reading them out with one cold transistor and one output line to a room-temperature amplifier. Multiplexing large arrays of tunnel junction detectors has so far proved difficult, although new microwave readout techniques may make multiplexed arrays feasible.

Myriad Applications

THE SUPERCONDUCTING detectors available today are 10 to 100 times more sensitive than conventional detectors operated at room temperature. These devices are improving measurements in a broad range of fields.

Nuclear nonproliferation and homeland defense. One of the most pressing international priorities is to control the dissemination of nuclear materials that could be used in attacks by terrorists or rogue states. Nuclear materials contain unstable isotopes, which emit x-rays and gamma rays. The characteristic energies of these photons provide a fingerprint revealing which radioactive isotopes are present. Unfortunately, some isotopes that occur in benign applications emit gamma rays with energies that are very similar to those emitted by materials used in weapons, which leads to ambiguous identifications and false alarms.

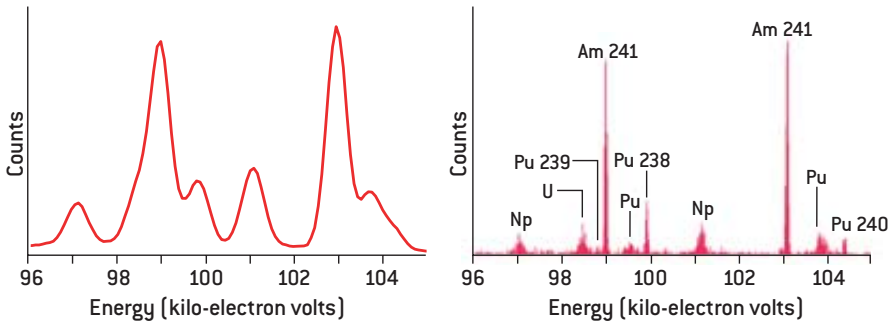
This problem has been vexing the U.S., which is installing thousands of radiation portal monitors to detect the gamma rays emitted by nuclear materials carried by vehicles crossing the Canadian and Mexican borders. One of our worst fears is that terrorists might smuggle highly enriched (weapons-grade) uranium into the country to build a crude Hiroshima-style atomic bomb [see "Thwarting Nuclear Terrorism," by Alexander Glaser and Frank N. von Hippel; Scientific American, February]. The primary signature of highly enriched uranium is the 185.7-kilo-electron-volt (keV) gamma ray from uranium 235. This gamma ray, though, has almost the same energy as the 186.1-keV gamma ray emitted from the radium 226 in clay in cat litter and other materials, making it very difficult to distinguish the two. This so-called kitty litter problem is the largest source of false alarms at the U.S. borders.

A team led by Michael Rabin of Los

Practical imaging requires a large array of detectors, similar to a CCD ARRAY in a digital camera.

THE AUTHOR

KENT D. IRWIN leads the quantum sensors project at the National Institute of Standards and Technology (NIST) in Boulder, Colo., and is an adjoint professor of astrophysical and planetary science at the University of Colorado at Boulder. Irwin received a B.S. at the California Institute of Technology and a Ph.D. from Stanford University. His research interests include precision measurements of electromagnetic signals for cosmological physics and the fundamental thermodynamic and quantum-mechanical limitations of sensors and detectors. His team has received the Department of Commerce gold medal and the NIST applied research award for its work with superconducting detectors.



AT PORTS and elsewhere today, detectors screen for nuclear materials being smuggled into the country (*photograph*). Conventional semiconductor detectors are too nonspecific to distinguish between some isotopes in a test sample (*left graph*). Measurements made with a superconducting detector can sharply resolve all the lines (*right graph*), including the presence of plutonium 239, the isotope preferred for nuclear weapons.

Alamos National Laboratory and Joel Ullom of the National Institute of Standards and Technology (NIST) in Boulder, Colo., and another team at Lawrence Livermore National Laboratory have developed gamma-ray detectors based on TES technology that have more than 10 times better energy resolution than conventional detectors. These detectors can resolve more lines in the complicated gamma-ray spectra of nuclear materials, such as uranium and plutonium isotopic mixtures [*see graphs in illustration above*]. The devices are being developed specifically to help in the verification of international nonproliferation treaties by determining the plutonium content of spent nuclear fuel. But they are also able to distinguish between the radium 226 in cat litter and the uranium 235 in highly enriched uranium. If a conventional handheld detector or portal monitor were to detect a gamma-ray signal, one of the superconducting devices could be used as a follow-up tool to distinguish unambiguously between these two isotopes, thus eliminating many false alarms.

Analysis of microchips. An application that is important for the semiconductor industry is electron-probe microanalysis. When a scanning electron microscope images a sample, the electron beam causes the sample to emit x-rays. One can determine the chemical composition of the sample within the nanometer-scale region of the beam by measuring the energies of the various x-rays emitted. As the beam scans across the entire sample, the resulting image shows where different chemical compositions are present, mapping out the structures that define how a microchip functions.

At present the semiconductor industry uses semiconductor x-ray detectors to study structures and defects on microchips. But as microchips use smaller features, new generations of microanalysis instruments with greater sensitivity are needed. My group at NIST has met this challenge by developing microanalysis systems based on TES detectors that have an energy resolution 50 times better than industrially available semiconductor detectors, enabling them to resolve many important x-ray spectral peaks. Such microanalysis systems are now becoming commercially available.

Submillimeter astronomy. Astronomy is a field rich with opportunities for superconducting detectors. Astronomers have often driven the development of new detector technologies because of their need to measure very faint signals from distant objects.

My group at NIST is working with the U.K. Astronomy Technology Center in Edinburgh, Raytheon Vision Systems in Goleta, Calif., and universities in the U.K. and Canada to develop a superconducting camera called SCUBA-2 for use at the James Clerk Maxwell Telescope atop Mauna Kea on the big island of Hawaii. SCUBA-2 will supersede the Submillimeter Common-User Bolometer Array (SCUBA), a semiconductor-based detector array developed by the U.K. Astronomy Technology Center. SCUBA images planet-, star- and galaxy-forming regions by detecting rays having submillimeter wavelengths, which are shorter than microwaves but longer than visible light.

Until recently, this wavelength range was largely inaccessible to astronomers because photons in the submillimeter band have too little energy to create electronic excitations in semiconductors but too high a frequency to be amplified effectively by receivers analogous to radio receivers. Yet the submillimeter range is well worth pursuing, as it contains fully 98 percent of the photons emitted since the big bang. Among other features, it allows observations of regions inside molecular clouds that are hidden at other wavelengths. SCUBA works by detecting heating of its 128 chilled semiconductor pixels, a process that is slower and much less sensitive than SCUBA-2's superconducting technology.

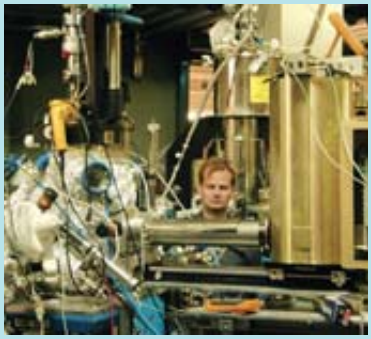
When complete, SCUBA-2 will consist of more than 10,000 TES detector

DANIELA NAOMI MOLNAR; SOURCE: JOEL ULLOM *NIST* AND MICHAEL RABIN *Los Alamos National Laboratory* (*graphs*); MEL EVANS *AP Photo* (*photograph*)

Other Uses

In addition to the applications discussed in the main text, superconducting detectors are already used for the following:

- X-ray spectroscopy at synchrotrons (*right*), including chemical analysis of metals in proteins and other samples.
- Efficient detection of large biological polymers and DNA fragments in mass spectrometers, which has applications in genomics, proteomics, drug discovery and analysis of natural compounds.
- Photon counting at telecommunications wavelengths (infrared) for quantum cryptography.
- The search for weakly interacting massive particles postulated to make up dark matter in the universe. —*K.D.I.*



EXPERIMENT at synchrotron at Lawrence Berkeley National Laboratory.

For more on photon counting and dark matter applications, see *www.sciam.com/ontheweb*

pixels read out with superconducting multiplexers and will allow astronomical objects to be imaged up to 1,000 times faster. The prototype 1,280-pixel SCUBA-2 subarrays have been fabricated [*see box on page 90*], and the full camera should be operational on the telescope by 2007. Many other submillimeter- and millimeter-wave superconducting detector systems are being developed for various observatories and satellites.

Cosmology. In recent years some of the most important breakthroughs in our understanding of the universe have come from measurements of the cosmic microwave background (CMB) radiation. The photons in the CMB are a snapshot of the universe roughly 400,000 years after the big bang, because for the past 13 billion years most CMB photons have passed undisturbed through the universe. Sound waves in the plasma of the early universe created patterns in the CMB radiation that astronomers see today [see "The Cosmic Symphony," by Wayne Hu and Martin White; SCIENTIFIC AMERICAN, February 2004]. Measurements of these patterns in the CMB, in concert with other cosmological observations, have shown that only about 5 percent of the universe today is made up of the ordinary matter and energy that we are familiar with; about 22 percent is dark matter and 73 percent is a mysterious field referred to as dark energy.

In addition to these sound-wave patterns, more subtle patterns should be imprinted on the polarization of the CMB radiation by gravity waves that were generated during a period of exponential cosmic expansion known as inflation—the so-called cosmic gravity-wave background. (In polarized light, the electric field of the electromagnetic wave is aligned in some way instead of oscillating randomly in all directions.) This polarization of the CMB radiation arose when the radiation scattered off the primordial plasma, much as visible light is polarized when it is reflected off a surface.

Soon astronomers will use superconducting detectors with polarization sensitivity to search for the cosmic gravity-wave background. At first, specialized telescopes both on the ground and in high-altitude balloons will employ the devices. Later, NASA plans for a satellite called the Inflation Probe to make the definitive CMB polarization measurements. Successful measurement of these gravity-wave patterns would offer insight into the physics that operated in the first trillionth of a trillionth of a trillionth of a second after the big bang, when interactions occurred at the energy scale at which all forces except gravity are unified into one. Ever since Albert Einstein, physicists have dreamed of directly investigating this regime of "grand unification," but the largest particle accelerators on earth are a trillion times too low in energy. Superconducting detectors will help scientists use the laboratory of the cosmos to access energy scales forever out of reach of terrestrial experiments.

Giant Arrays

DESPITE THE DRAMATIC advances in superconducting detector technology over the past decade, the full potential of superconducting detectors is still unrealized. The impact of CCDs was not truly felt until pictures were taken with extremely large arrays of pixels. Looking ahead, large-scale arrays of superconducting detectors—with up to 10,000 pixels at millimeter wavelengths and millions of pixels in the x-ray band—will be developed using new fabrication techniques and incorporating signal multiplexing at microwave frequencies, which allows many more pixels to be read out in a single wire. Engineers are developing smaller and cheaper cryogenic systems to cool them. These big superconducting arrays will affect an even broader range of disciplines. The arrays will still have fewer "pixels" than the human retina, but they will take human vision into exciting new realms of discovery. SA

MORE TO EXPLORE

Low-Temperature Particle Detectors. Norman E. Booth, Blas Cabrera and Ettore Fiorini in *Annual Reviews of Nuclear and Particle Science,* Vol. 46, pages 471–532; 1996.

Quantum Calorimetry. Caroline K. Stahle, Dan McCammon and Kent D. Irwin in *Physics Today,* Vol. 52, No. 8, pages 32–37; August 1999.

Seeing Single Photons. Graham P. Collins in *Scientific American,* Vol. 290, No. 1, pages 24–25; January 2004.

Transition-Edge Sensors. K. D. Irwin and G. C. Hilton in *Cryogenic Particle Detection.* Edited by Christian Enss. Springer-Verlag, 2005.

STEPHAN FRIEDRICH *Lawrence Livermore National Laboratory and Lawrence Berkeley National Laboratory*

The Origin of the Greek

By Bradley E. Schaefer

ASTRONOMY AND ARCHAEOLOGY HAVE TOGETHER UNCOVERED THE HISTORY OF HOW THE STAR PICTURES CAME TO BE—AND HOW PEOPLE HAVE USED THEM OVER TIME

MARY EVANS PICTURE LIBRARY

My grandfather first taught me about the Great Bear constellation. After that, I had fun wielding an old pair of binoculars and picking out other constellations in the wide sky over Colorado—or even inventing my own. At the time, of course, I gave no thought to the age or origin of the constellations, but the curious pictures in the sky present a fascinating scientific puzzle.

In 1922, when the International Astronomical Union officially defined 88 constellations, it drew the bulk of them from Ptolemy's *The Almagest*, which was written around A.D. 150 and described the traditions widespread among the Greeks. These traditions had been popularized in the "best-selling" poem *The Phaenomena*, by Aratus (275 B.C.). The great astronomer Hipparchus's sole surviving book, *The Commentary* (147 B.C.), tells us that Aratus's poem is for the most part a copy of a work with the same name by Eudoxus (366 B.C.), which no longer survives. These books held the earliest descriptions of the Greek skies, and in them the constellations are already fully formed. But where did the Greek constellations come from?

Paleolithic Hunters

TO TRACE THE HISTORY of the constellations, we need to go back to a time long before the Greeks. Constellations are ubiquitous throughout the world's cultures. In the millennia before light pollution dimmed the night sky, people would watch the ceaseless motion

Constellations

GREAT BEAR is shown as it was depicted by Alexander Jamieson in his 1822 celestial atlas. The three stars in the tail are often referred to as three hunters.

of the stars overhead. Humans are naturals at pattern recognition, and so it is not surprising that the collection of neighboring stars into groups was a universal preoccupation from the earliest times. The best place to start is where my grandfather did, with the Great Bear (or Ursa Major).

These seven moderately bright stars are known by other names—the Big Dipper (or Drinking Gourd), the Wagon (or Wain) and the Plow. Aratus calls the grouping both Bear and Wagon. The wagon name must, of course, have come after the invention of the wheel (roughly the fourth millennium B.C.), but the bear name is undoubtedly much older. Early societies throughout Eurasia recognized the Great Bear stars and myth. The most common version was that the four stars in the bowl of the dipper were the bear, which was perpetually chased by the three stars in the handle, which represented three hunters. The Greeks, Basques, Hebrews and many tribes in Siberia had this basic star/myth combination. Surprisingly, the same bear stars and stories surfaced throughout North America. With some variations, many tribes of the New World—including the Cherokee, Algonquin, Zuni, Tlingit and Iroquois—share the interpretation of the bear followed by three hunters.

How can we explain the close matching of widespread traditions between the Old World and the New? The Bear is unlikely to be an independent invention, because the stars do not look like a bear. Also, we can rule out the possibility of contamination from the first missionaries and explorers, because the Indian lore was often collected very early, and the stories do not mirror exactly the Greek version carried by European settlers. The most logical explanation to connect the traditions holds that the first settlers of the New World carried the basic myth across the Bering Strait.

Roughly 14,000 years ago Paleolithic hunters and gatherers first migrated across a land bridge that formed during the last ice age, when sea level was low, and connected Siberia and the Americas. Their culture lived on in their descendants, who populated the New World. It is easy to envision a chain of grandfathers stretching from Paleolithic Siberia to the mountains and plains of the New World and eventually to modern Colorado, telling about the Bear in the sky.

The actual origin of the Bear constellation could have been very long indeed before this migration. European cave paintings, artifacts and ensembles of cave bear skulls date to more than 30,000 years ago and suggest some kind of bear worship. The constellation may have originated as a folk depiction of a shamanistic icon. However it came to be, the Great Bear is quite likely one of the oldest inventions of humanity.

Assyrian Priests

THE EARLIEST DIRECT EVIDENCE for the constellations comes from inscribed stones and clay writing tablets dug up in Mesopotamia (within modern Iraq). A text called "Prayer to the Gods of the Night" from old Babylon, dated to around 1700 B.C., mentions four constellations, including the Wagon, plus three individual stars and the Pleiades. Stone inscriptions from before 1300 B.C. show icons that in later centuries referred to constellations, but these early icons do not appear in a celestial context, implying that the figures had not yet been projected onto the sky. Although the scarcity of evidence makes it hard to reach a confident conclusion, the Mesopotamians apparently had formed only a few constellations before 1300 B.C.

After that time, boundary markers and cylinder seals begin to depict constellation icons grouped together and in conjunction with known symbols for the sun, moon and planets. These close associations assure us that the icons refer to constellations. Starting around 1100 B.C., cuneiform tablets list more than 30 constellation names from three bands stretching around the sky. One series of three tablets called MUL.APIN contains long lists of observations about the positions and movements of almost all the Mesopotamian star groups. MUL.APIN was copied repeatedly, with little variation, apparently as a textbook or almanac; surviving copies date from 687 B.C. to the third century B.C.

EARLIEST DIRECT EVIDENCE FOR THE CONSTELLATIONS

COMES FROM INSCRIBED STONES AND CLAY WRITING TABLETS DUG UP IN MESOPOTAMIA.

Fortunately for those of us who study ancient astronomy, we can date constellation lore by a unique method based on the precession of the skies. Precession is the shifting of stars against the coordinate grid defined by the North Pole and the equinoxes [*see box on opposite page*]. We can read the date like a clock, with stars forming a very slow hour hand moving against the face of the sky coordinates. MUL.APIN gives the relative positions of stars on the sky, and we can translate these into approximate dates. For example, the tablets say that the spring equinox is in the eastern side of what we now call the Ram, and this reads as late in the second millennium B.C. Many of the observations also depend on the latitude of the observer, and thus a full analysis can produce both a date and the observer's latitude.

Hermann Hunger of the University of Vienna and the late David Pingree of Brown University analyzed several of the lists in MUL.APIN, in part by comparing them with a later accounting of star appearance dates from Ptolemy. They derive a date of 1000 B.C. and a latitude of 36 degrees, suggesting that the observer—or observers—might have worked in Assyria (the northern part of Mesopotamia). Because MUL.APIN mentions almost all the Mesopotamian constellations, which must have been formed before this date, it appears that the bulk of the Mesopotamian constellations were created within a relatively short interval from around 1300 to 1000 B.C.

Independently, I have identified 114 observations in MUL.APIN that imply the date and latitude for the observations. No one item will yield a date and latitude with enough precision

to be of much use, but the statistical combination of all 114 observations can lead to fairly accurate values. I find that the reports in MUL.APIN date to 1100 B.C. (with an uncertainty of 80 years) from a latitude of 33 degrees north (with an uncertainty of 1.5 degrees). My results agree with those of Hunger and Pingree in pointing to Assyrian observers and suggesting a formative period of less than two centuries.

A separate analysis of the constellation data can determine the time and place of the origin of the southernmost constellations visible from the Northern Hemisphere. The idea behind this calculation is that the stars too far south to be visible from a mid-northern site would form a roughly circular void centered on the South Pole. The position of the void's center, the implied site of the pole, would tell the date, and the radius of the void would indicate the latitude of the constellations' inventors. Using this basic concept, I calculated that the six southern constellations that define the edge of the void were invented early in the first millennium B.C. from a latitude of roughly 33 degrees.

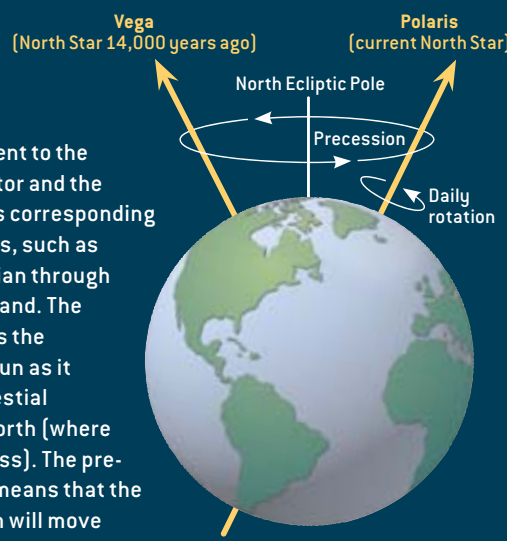
The archaeological evidence plus the epoch and latitude from MUL.APIN and from the six southernmost constellations all give a consistent story. Most of the Mesopotamian constellations and observational data were made from near a latitude of 33 to 36 degrees between 1300 and 1000 B.C., by people we would call Assyrians.

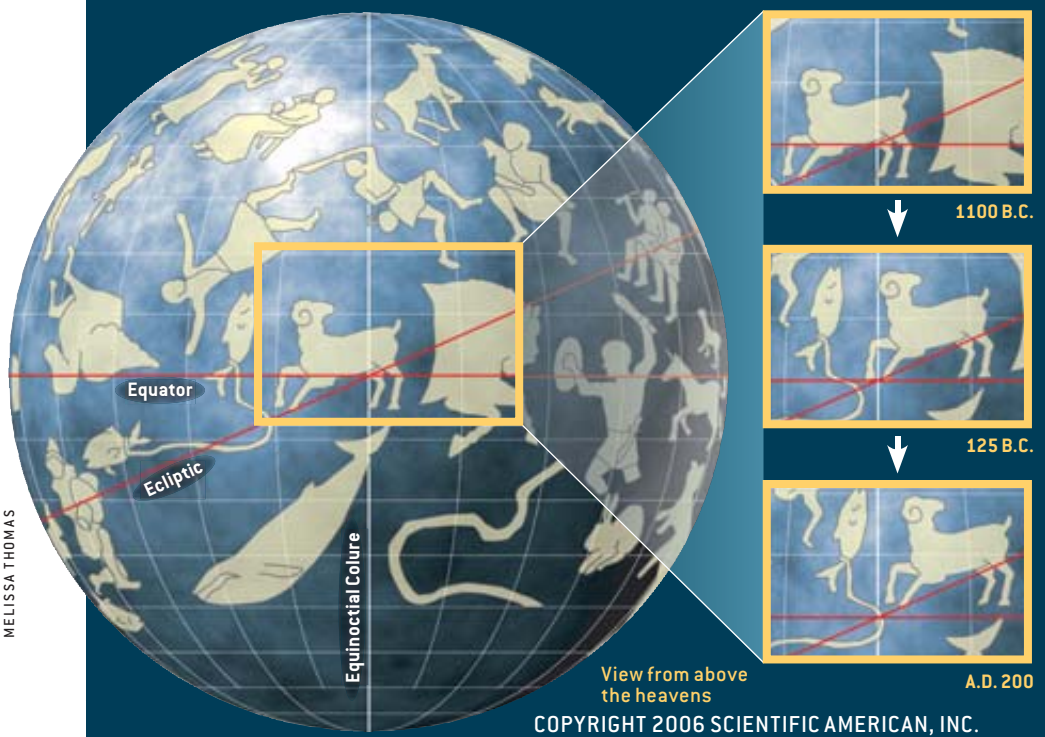
The constellations in MUL.APIN are a curious mixture. Some are gods, others represent animals, and the remainder depict everyday farm implements. The text gives many omens

PRECESSION: The Shifting Stars

The dating of constellations requires knowledge of a process called precession, the slow wobbling of the earth's pole, caused by tidal forces from the moon, that makes the positions of the stars shift over the centuries. A toy top will totter in much the same way as it slows down. One consequence of this wobble is that the earth's North Pole shifts in a big circle in the sky, and thus the so-called polestar, or North Star, the star nearest to the pole, changes from time to time (*diagram at upper right*). Another consequence is that the positions of the stars slide along the ecliptic (the path of the sun in the sky) at a rate of almost one degree (twice the moon's angular diameter) every 72 years.

Star positions are measured with respect to conceptual lines (actually great circles) on the sky. These lines are analogous to those on the earth's globe, with the celestial equator equivalent to the terrestrial equator and the celestial colures corresponding to longitude lines, such as the prime meridian through Greenwich, England. The spring equinox is the position of the sun as it crosses the celestial equator going north (where the red lines cross). The precessional shift means that the equinox position will move against the background stars, changing zodiacal constellations roughly once every two millennia. In classical Greek times the equinox was in Aries; it then moved into Pisces and more recently has moved into Aquarius (hence the "Age of Aquarius"). Hipparchus discovered precession sometime around 128 B.C., as he noted the shifting position of the equinoxes when compared with what were then old reports on its position—a stunning intellectual feat that establishes him as one of the greatest of ancient astronomers. *—B.E.S.*





MELISSA THOMAS

CONSTELLATION POSITIONS shift slowly over time with respect to the colures and equator, thus providing a way to determine dates. An analysis of the position of the Ram on the sky globe of the Farnese Atlas [*see box on next page*] shows that the original of the Atlas statue was made around 125 B.C.—when the edge of the Ram's horn was just past the colure.

THE FARNESE ATLAS

The oldest surviving depiction of the set of Greek constellations appears on a second-century A.D. Roman statue called the Farnese Atlas. Art historians conclude that the statue is a copy of a Greek original. Now in Naples, the marble carving depicts the god Atlas holding a sky globe on his shoulder.

My detailed analysis of the positions of the constellations on the globe reveals that the figures are placed with an accuracy of better than two degrees for a date of 125 B.C., with an uncertainty of 55 years. The accuracy indicates that the original data were systematic and precise—such as those in a star catalogue. Hipparchus's catalogue was the only one then in existence, and a careful comparison of the constellations on the globe with the verbal descriptions in Hipparchus's *Commentary* and in other works provides a match only with Hipparchus. Of course, another astronomer could have made an independent catalogue around that time, but no hint of one survives, and Hipparchus is almost certainly the source. —*B.E.S.*

RELIEF FIGURES on the sky globe trace the constellations in fine detail and include the celestial equator, tropics and colures.

based on the star groups and also uses them to form a calendar—vital for any farming economy. Omens, gods and calendars were the province of the priesthood, a persuasive clue that the constellations were developed by priests.

Greek Scholars

THE MESOPOTAMIAN GROUPINGS turn up in many of the classical Greek constellations. The stars of the Greek Capricorn and Gemini, for example, were known to the Assyrians by similar names—the Goat-Fish and the Great Twins. A total of 20 constellations are straight copies. Another 10 have the same stars but different names. The Assyrian Hired Man and the Swallow, for instance, were renamed Aries and Pisces.

The constellation called the Triangle provides a good example of how the Greeks adopted the Mesopotamian stars. In MUL.APIN, the stars of the Triangle were designated the Plow (a name that had also been used for the Bear). In Mesopotamia and Egypt, geometry was well known to scholars but only as a mundane engineering tool. Thales (585 B.C.) brought geometry from Egypt to Greece, where it was transformed from a practical set of rules of thumb to an organized logical system of great beauty and generality, culminating in Euclid's definitive book *The Elements* (300 B.C.). Only with this transformation would anyone seek to commemorate the triangle, as the basis of geometry, in the sky. The Triangle is thus certainly a Greek renaming of a Mesopotamian star group between the time of Thales and Eudoxus or sometime in the sixth to fourth centuries B.C.

Whether the Greeks had adopted many constellations at the time of the Triangle's introduction is unknown, but the limited information on the subject suggests not. The two earliest written sources from the Greeks—the epics of Homer (assumed to be eighth century B.C.) and the farmer's almanac of Hesiod (also dated to the eighth century)—both mention two prominent constellations (Orion and the Great Bear), two star clusters (the Pleiades and the Hyades), and two stars (Sirius and Arcturus). But nothing more. And all other Greek sources from before about 500 B.C. are silent on the stars. So the Greeks had the most prominent star pictures before 500 B.C., but perhaps not many more.

The first complete discussion of the Greek skies comes from Eudoxus's fourth-century B.C. book, now known only through extensive copying by Aratus and Hipparchus. Eudoxus contains many reports such as "that head [of the Dragon] wheels near where the limits of setting and rising blend," which is to say that the stars in the head of Draco are at a distance from the pole such that they skim the northern horizon. This observation is true only for a given latitude that varies with date. This constraint on latitude and epoch is rather loose and is not helpful by itself. Eudoxus also writes, "At the rising of the Scorpion in the East Orion flees at the

THE AUTHOR

BRADLEY E. SCHAEFER has worked for more than 20 years on the technical aspects of what is and is not visible in the sky and the application of this information to the history of astronomy. He earned undergraduate and graduate degrees at the Massachusetts Institute of Technology and is currently a professor in the department of physics and astronomy at Louisiana State University. He is on the editorial boards of the *Journal for the History of Astronomy* and of *Archaeoastronomy*.

Western verge"; in other words, the Scorpion and Orion appear simultaneously on opposite horizons. This second statement also leads to a fuzzy constraint on latitude and epoch.

The two constraints can be combined to derive a unique date and latitude for the observations, but with an accuracy still too poor to be useful. To get around this problem, I identified 172 statements in Eudoxus that depend on the latitude and epoch. The combined constraints from all observations result in a final uncertainty that is 0.9 degree (100 kilometers, or 62 miles) in latitude and 80 years in time. I find that all of Eudoxus's lore is consistent with coming from one time and place—1130 B.C. at 36 degrees latitude, which would have been in Assyria.

The derived time and place match well with those determined for the MUL.APIN observations. Indeed, I have noticed that MUL.APIN and Eudoxus share a substantial amount of information, leading me to conclude that the data used by both came from an original database made by an Assyrian observer around 1100 B.C. Because both surviving branches contain most of the ancient constellations (albeit some with different names), I conclude that the set of constellations was largely complete by sometime around 1100 B.C.

BOUNDARY STONE from Mesopotamia dates from sometime after 1300 B.C. It depicts constellation icons.

MUL.APIN, a cuneiform tablet with observations from 1100 B.C., lists the Mesopotamian constellations.

Therefore, it is reasonable to conclude that sometime after this and before the existence of Eudoxus's book (366 B.C.), the Greeks received the Mesopotamian star groups. The lack of any evidence for the Greek constellations (other than the Bear and Orion mentioned in Homer) before 500 B.C. suggests that most of the transfer happened after that time. We know from textual evidence that the Babylonian zodiac system came to Greece around 400 B.C. (The zodiac was the track that the sun took as it traveled around the earth; as the sun moved along the zodiac, it passed in front of 12 constellations that had animal and human forms.) Intellectual knowledge could have spread from Mesopotamia to Greece in many ways, but we do not know enough to choose between them.

Despite the heavy reliance on Mesopotamian star groups, the Greek system of constellations still has 18 star pictures with no roots in the East nor anywhere else that we can discover. Moreover, the nature of these constellations is characteristically Greek. There is the quintessential Greek hero Hercules, for example, joined in the sky by other constellations that represent creatures the great warrior had defeated—Leo and Draco among them. Ophiuchus, carrying the Serpent, is there as well, and the Dolphin, as would be appropriate for a seafaring people such as the Greeks. Six of these constellations depict a tableau from Greek mythology concerning Perseus's rescue of Andromeda. Most likely, these "new" star pictures were invented by the Greeks themselves.

Modern Scientists

OVER TIME, the way the Greeks used the constellations changed. The images started out telling stories about legendary heroes and animals. Then they became tools for calendars and navigators. Later the zodiac became a coordinate system for measuring planetary positions as part of astrology learned from the Babylonians.

The writings of Hipparchus mark the shift to the scientific study of the stars. Early in his career, he made detailed quantitative comparisons of the constellations in Eudoxus's book with what was actually visible in the sky—and he found many differences. In 135 B.C. he discovered a nova, an exploding star, and that discovery inspired him to construct a complete catalogue of the stars (now lost) so that later novae could be identified. With both a quantitative catalogue and the earlier Greek data in hand, Hipparchus was able to figure out the shifting of the stars caused by precession, a watershed discovery. Constellation study had become essentially scientific in the modern sense.

The origins of the Greek constellations from a patchwork of sources superimposed on the sky lie so far in the past that evidence is sketchy, and we must acknowledge that the story will have significant gaps and that some points can only be reasonable speculation. But the basic outline is clear, and it shows us how knowledge is transferred across time and space as changing parts of culture. Constellations provide a unique way of gaining insight into a culture (different from, say, pottery styles) because they consider an intellectual aspect of prehistoric society, something that is hard to discern from more typical archaeological sources. Furthermore, the sequence in the uses to which societies put constellations tracks the way one aspect of astronomy was transformed into a modern science. The sequence moves from religious to folk to practical to scientific usages, a consistent trend toward decreasing spirituality and increasing quantification. And yet today as in the past, the tales of our grandparents continue to link generations and cultures. SA

MORE TO EXPLORE

Asiatic Parallels in North American Star Lore: Ursa Major. W. B. Gibbon in *Journal of American Folklore,* Vol. 77, pages 236–250; 1964.

The Mithraic Mysteries. D. Ulansey in *Scientific American,* Vol. 261, No. 6, pages 130–135; December 1989.

Astral Sciences in Mesopotamia. H. Hunger and D. Pingree. Brill, 1999.

PAPER RECYCLING

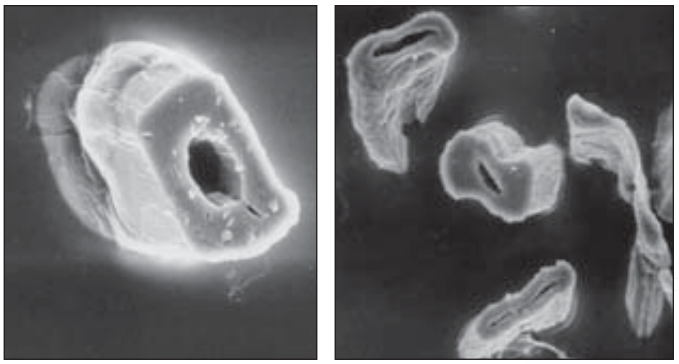
Gunk-Free Fiber

Thanks to citizens everywhere, the world now recycles just over 50 percent of the paper it uses, according to various estimates. Reprocessing plants are sprouting widely. But trees will never be fully spared because of a quirk of wood fibers themselves.

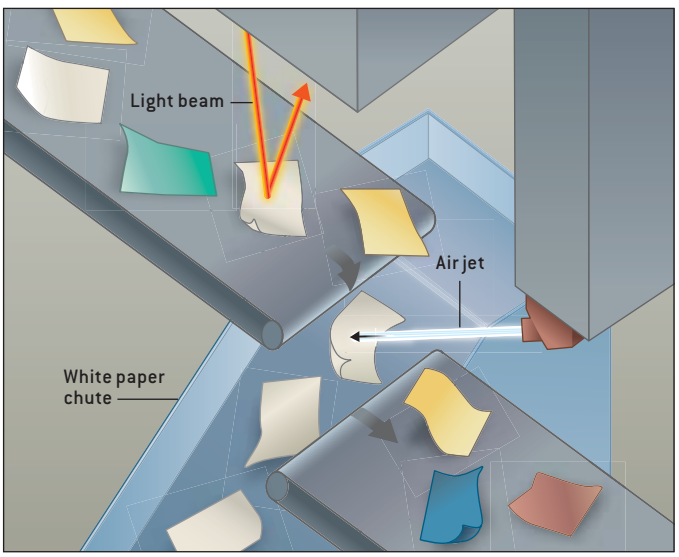
Virgin pulp is rich in water, which provides for ample hydrogen bonding that holds fibers together when made into paper. But each time a fiber is cleaned, de-inked and dried in a reprocessing plant, "only 80 percent of the bonds recover," explains Yulin Deng, associate professor of chemical and biomolecular engineering at the Georgia Institute of Technology. After four or five recyclings, he says, "a fiber can no longer make strong enough bonds" and becomes waste [*see photographs at right*].

Engineers can do little that is economically viable to overcome this physical limitation, so they focus on reducing the cost of reprocessing fresher fibers. One main challenge is finding a better way to neutralize "stickies"—the mess of adhesives from stamps, labels, seals, tape, magazine spines and myriad other sources. The gunk, highly deformable, slips through screens designed to catch it and clogs machinery, necessitating a slow and energy-intensive "fine screen" step to try to trap it [*see main illustration*]. The industry has been working for a decade to find a chemical process that will break down stickies, Deng says, "but we haven't fully solved it." Part of the challenge is to minimize contamination of the volumes of water used in reprocessing, which itself has to be cleansed and reused.

More effective sorting would also reduce cost, further raising demand for recycled paper and sparing landfills. Although some municipalities require residents to sort paper into classes such as cardboard and newsprint, most of the waste that arrives at collection centers is mixed. About 95 percent of it is unscrambled by hand, a costly effort. Manufacturers have only recently begun to sell automated sorting machines, and sensors to improve them are still being developed. Richard A. Venditti, associate professor of wood and paper science at North Carolina State University, says better equipment would not only reduce costs but would also feed a more homogeneous stream into the reprocessing line, allowing that operation "to use fewer chemicals, less water and less energy." —*Mark Fischetti*



TREE FIBER, or cellulose (*left*), contains plentiful water molecules that facilitate hydrogen bonding among neighboring fibers when pressed into paper. Each recycling collapses and dries the fiber until it is so contorted and rigid that it lacks sufficient surface area or bonding sites to bind (*right*). It ends up as refuse. (All images are cross sections; full fibers are elongated.)



AUTOMATED SORTERS, relatively new, use inclines and spinning disks to thin a waste stream into a single layer on a conveyor (*not shown*). Visible and near-infrared light beams identify each piece by color or gloss, and air jets propel it into the appropriate collection chute.

Topic suggested by readers David Beaver and Burckhard Mohr. Send ideas to workingknowledge@sciam.com

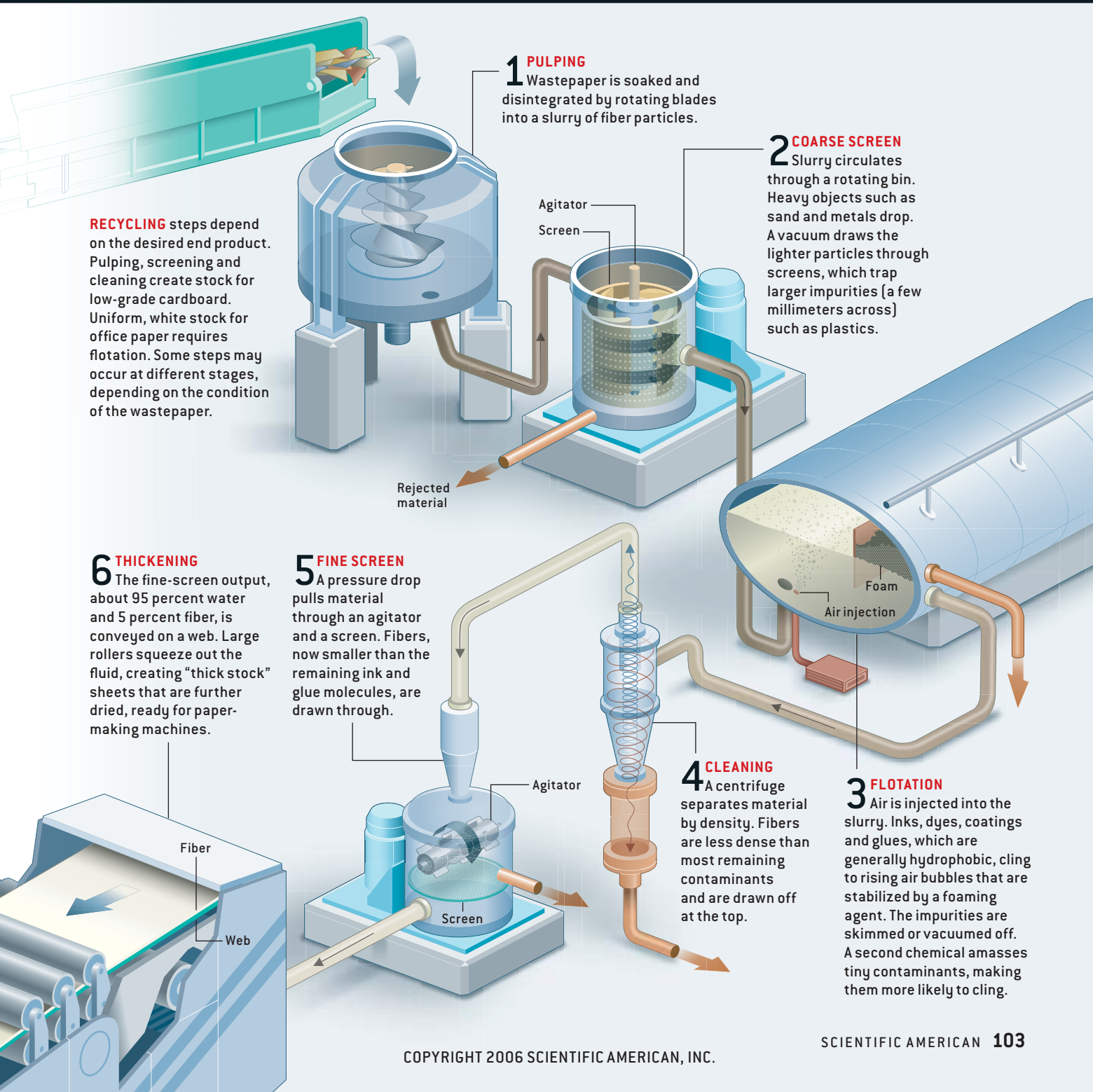
SAMUEL VELASCO *SW Infographics* (*opposite page*); JEN CHRISTIANSEN (*this page*); ELSA CHAIDEZ *Institute of Paper Science and Technology* (*micrographs*)

DID YOU KNOW...

COST CUTTER: Higher grades of recycled pulp require more reprocessing steps, progressively raising their cost to that of virgin pulp produced from trees. In general, cereal and shoe-box cardboard pressed from recycled material is significantly cheaper than a similar product made from new fiber, says Richard A. Venditti of North Carolina State University. Corrugated boxes and newsprint cost somewhat less. There are no real savings for office paper.

RESIDENTS RULE: Some 86 percent of U.S. residents have access to curbside or drop-off recycling, according to the American Forest and Paper Association. Participation is fairly high but could still improve. Schools are less likely to recycle; businesses even less so. Greater compliance would help collectors make money, in part because fast-growing countries such as China want to buy more wastepaper as an inexpensive raw material.

SEE IT, SORT IT: Sensors in the first automated sorting machines identify primarily the color of paper passing on a conveyor, allowing them to separate, say, brown bags from copier paper. Other sensors being perfected at companies such as MSS, Inc., in Nashville, Tenn., and in university projects funded by the Department of Energy can distinguish between the black and white of office documents and that of newsprint as well as differences in gloss among consumer packaging.



Uncommon Scents

A WHIFF AT THE PLATE BRINGS UP A BOUQUET OF POSSIBILITIES **BY STEVE MIRSKY**

New York Yankee great Derek Jeter usually comes up smelling like a rose. But according to mid-August news reports, Jeter will soon also smell like "chilled grapefruit, clean oakmoss and spice." Those odors are the elements of the shortstop's new men's perfume—I mean cologne—to be sold under the name Driven. (It's the scent that says, "I'm not stopping at second base.")

Athletes thus join movie stars and other celebrities in having their own signature fragrances, for sale to the malodorous masses. But, although scientists are vital to the fragrance industry, there are no fragrances honoring them. So here are some suggestions for a new line of scientist-inspired scents.

Isaac Newton's *Gravitas*
Ingredients: Fresh-cut grass, royal mint and, of course, apple spice.
Slogan: For the man who likes his heavenly bodies as far away as possible.

J. Robert Oppenheimer's *Cataclysm*
Ingredients: Desert sand, enriched geranium.
Slogan: Now I am become Death, the destroyer of worlds, but with a delicate floral hint.

Barbara McClintock's *Transposition*
Ingredients: Corn kernels, corn husk, Cornell.
Slogan: When your genes make that unexpected move.

Alan Turing's *Enigma*
Ingredients: Sptf qfubmt.
Slogan: Tell them it's the real you.

Nikola Tesla's *Genius*
Ingredients: Secret.
Slogan: Change the world.
(*Note: Removed from market; formula lost.*)

Alessandro Volta's *Charge*
Ingredients: Cat hair, amber, balloons.
Slogan: Make sparks fly.

Antoine Lavoisier's *Chemistry*
Ingredients: Hydrogen, oxygen.
Slogan: When the elements come together, you may lose your head.

Ludwig Wittgenstein's *Logic*
Ingredient: Silence.
Slogan: For when you don't have the words to say the things you mean.

Anton van Leeuwenhoek's *Hidden*
Ingredient: Pond water.
Slogan: Reveal … the little things.

Socrates' *Philosophy*
Ingredients: Olive oil, feta cheese, traces of hemlock.
Slogan: You fill me with … questions.

Albert Einstein's *Continuum*
(undoubtedly to be renamed Relativity after disappointing early sales)
Ingredient: Thyme.
Slogan: The faster you go, the shorter you get.

Stephen Hawking's *Universe*
Ingredients: Everything.
Slogan: You don't have to understand it.

Michael Faraday's *Dynamo*
Ingredients: Copper, mercury, rubber.
Slogan: Your magnetism will create … electricity.

Kurt Gödel's *Theorem*
Ingredients: Complete list unavailable.
Slogan: For the man who has almost everything.

Galileo Galilei's *Insight*
Ingredient: Patronage.
Slogan: Discover new worlds.

Benjamin Franklin's *Revolution*
Ingredient: Sage.
Slogan: Come in from the rain.

Stephen Jay Gould's *Evolution*
Ingredients: Flamingo smiles, hen's teeth, panda thumbs.
Slogan: Punctuate your equilibrium.

Leonardo da Vinci's *Invention*
Ingredients (written backward on label): Oil-based pigment.
Slogan: Leave a lady smiling. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

What is dark matter, and how is it affecting the universe?

Robert Caldwell, a cosmologist at Dartmouth College, offers this explanation:

Dark matter is a proposed solution to an as yet unresolved phenomenon—the mismatch between measurements of the gravitational mass and the luminous mass (the mass contributed by light-emitting matter) of galaxies and clusters, gravitationally bound groups of galaxies. This disparity suggests the presence of matter in the universe that does not efficiently produce light—hence, it is invisible, or "dark."

We can determine the gravitational mass of an object, such as a star, by measuring the velocity and radius of the orbits of its satellites. To find the luminous mass of a galaxy, we use the known relations between stellar mass, color and luminosity to translate the observed colors and intensity of light from the galaxy into the total mass of its constituent stars. This mass-to-light comparison indicates that the gravitational mass of galaxies and clusters far exceeds the luminous mass.

Thus, more matter exists than we can see. Other indicators, including recent NASA measurements of the cosmic microwave background radiation (which provides a glimpse of the universe at an early age), give us further information: dark matter outweighs normal matter by a factor of 6 to 1.

What could dark matter be? Many physicists and astronomers suspect it is a type of particle that they have not yet been able to detect. The prototypical dark matter candidate is something like a neutrino—a particle that is similar to an electron but has a much smaller mass and no electric charge. All known types of neutrinos, however, are too light and too rare to fit the theoretical description of dark matter.

How does dark matter affect the universe? It must be the basic building block of the largest structures in the universe: galaxies and clusters. And dark matter does not just explain the behavior of distant bodies in the cosmos; it must be abundant within our galaxy as well. Estimates of the Milky Way's makeup predict that our solar system is immersed in a fine sea of dark matter with a density as high as roughly 10^5 particles per cubic meter. As Earth travels around the sun, moreover, we experience dark "seasons" as we move with or against the flow of this dark sea.

Does the moon also have a tidal effect on Earth's atmosphere?

—T. Digenti, Somerville, Mass.

Rashid Akmaev, a research scientist at the University of Colorado at Boulder, replies:

The short answer is yes. At various times this question has occupied such famous scientists as English physicist Isaac Newton and French mathematician Pierre-Simon Laplace, whose theory describing the behaviors of oceans predicted the existence of atmospheric tides two centuries ago.

First, let us consider how ocean tides occur. At the point on the ocean's surface closest to the moon, the moon's gravitational force is strongest, pulling the ocean toward it. On the opposite side of Earth the moon's attractive force is weakest, which allows the ocean to bulge outward again, in this case away from the moon.

Now think of the atmosphere as an ocean whose seafloor is Earth's surface. Laplace's theory predicts two atmospheric pressure maxima—peaks in the amount of atmospheric material overhead—per lunar day corresponding to the two ocean bulges. As the ocean swells, so does the atmosphere above it.

Surprisingly, observations show that the sun causes much stronger semidaily atmospheric tides, although the solar gravitational influence is less than half that of the moon. Laplace suggested that the strong solar tide was primarily generated by solar heating and not by solar gravity—a hypothesis that scientists finally confirmed in the 1960s. SA

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

MATT COLLINS