THE SILENT EARTHQUAKE MENACE • A SCIENCE OF FAIR VOTING

SCIENTIFIC AMERICAN

The Time Bomb of Global Warming (and How to Defuse It)

MARCH 2004 $4.95
WWW.SCIAM.COM

Robots on Two Worlds

On MARS, twin rovers explore baffling landscapes

On EARTH, robotic vehicles race across the Mojave Desert

How Addiction Reshapes Brains

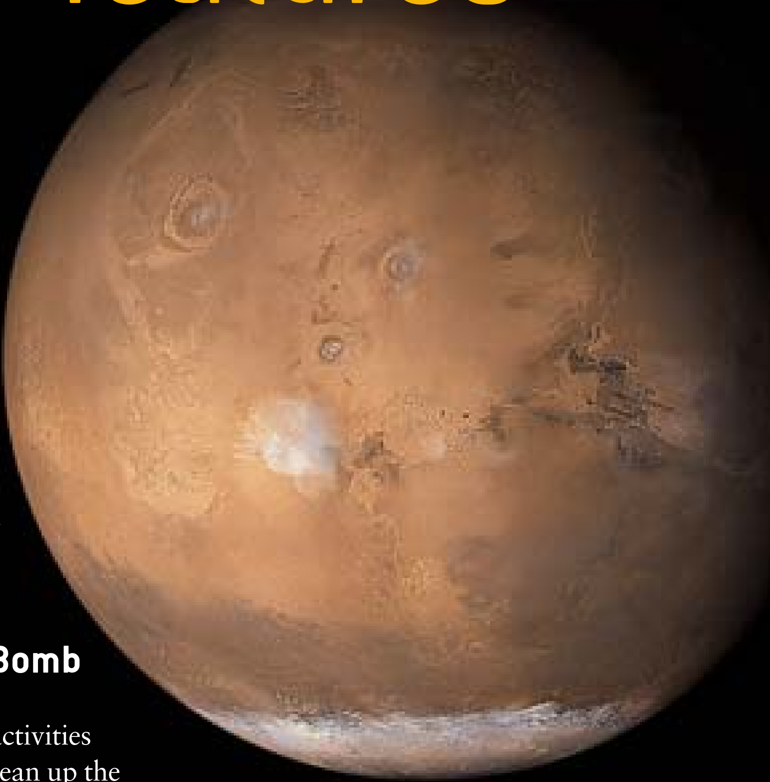
Flu Vaccines' Biotech Future
(see page 16)

march 2004

contents

SCIENTIFIC AMERICAN Volume 290 Number 3

features



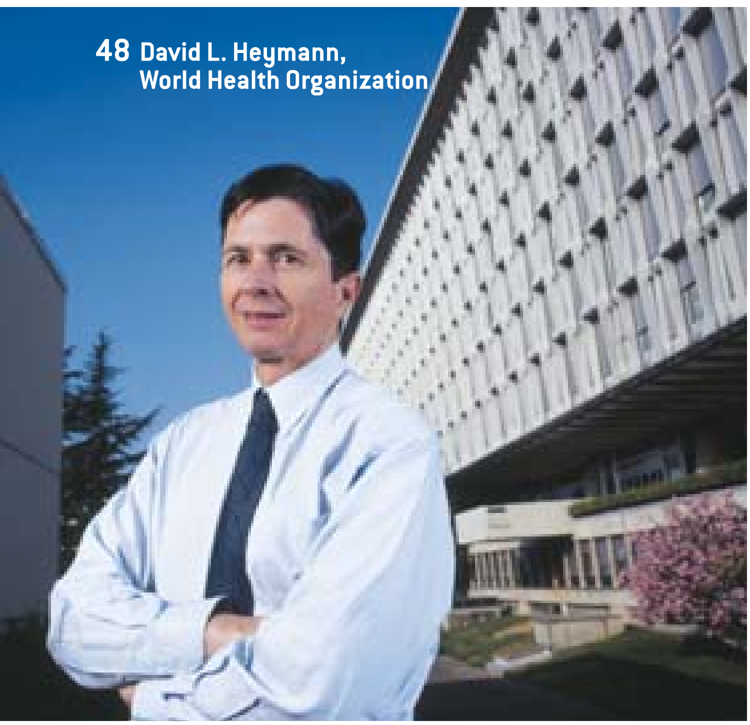
52 Mars yields grudgingly to robot probes

SCIENTIFIC AMERICAN Volume 290 Number 3

departments

42

111



48 David L. Heymann, World Health Organization

columns

Cover image by Daniel Maas, Maas Digital LLC, NASA/JPL/Cornell University; preceding page: NASA/JPL/Malin Space Science Systems.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

The Climate Leadership Vacuum

If you still doubt that global warming is real and that humans contribute to it, read the article beginning on page 68. Its author, James Hansen of the NASA Goddard Institute for Space Studies, is no doomsayer. Instead of relying on just computer climate models, which skeptics don't trust, Hansen builds a powerful case for global warming based on the geologic record and simple thermodynamics. He sees undeniable signs of danger, especially from rising ocean levels, but he also believes that we can slow or halt global warming affordably—if we start right away.

DIESEL SOOT is worth chasing.

ED JACKSON

Politically, that's the rub. As time slips by, our leverage over the problem melts away. Even small reductions in gas and aerosol emissions today forestall considerable warming and damage in the long run. In our view, the international community needs a leader, but the obvious nation for the job still has its head in the sand.

President George W. Bush's administration implies that it will get more serious about global warming after further years of study determine the scope of the problem (tick … tick … tick …). The Kyoto Protocol is the most internationally acceptable approach to a solution yet devised. Largely at the insistence of American negotiators, it adopts a market-based strategy. Nevertheless, the White House in 2001, like the U.S. Senate in 1997, rejected the treaty as economically ruinous and environmentally inadequate. The administration has yet to propose a workable alternative.

Two years ago the president committed the country to reducing its greenhouse gas "intensity"—the emissions per unit of economic output—by 18 percent in 10 years. But he has not enunciated a clear and credible strategy for doing even that. The White House boasts of the $4.3 billion budgeted for climate change–related programs in 2004 as well as its backing for hydrogen-based energy. But those initiatives don't set any goals by which they can be judged. All they do is throw money at new technologies in the hope that businesses might eventually adopt them. In other areas of environmental policy, the administration insists on cost-benefit analyses—but not for climate change policy.

A real action plan is feasible. Current technology can stop the increase of soot emissions from diesel combustion at a reasonable cost. Reductions in airborne soot would boost the reflection of sunlight from snow back into space. Minimizing soot also directly benefits human health and agricultural productivity.

Suitably controlling greenhouse gases is a greater challenge, but it can be done. Kyoto establishes a cap-and-trade program for carbon dioxide and other emissions. The administration has favored programs to trade credits for industrial pollutants such as mercury. Carbon dioxide is an even more appropriate subject for such an effort: creating environmental mercury "hot spots" raises local health risks, but concentrating carbon dioxide production is harmless.

The expense of reducing carbon dioxide could be kept low by letting the marketplace identify cost-effective ways to meet targets. Domestic emissions trading for sulfur dioxide under the first Bush administration was highly successful. Output levels were cut ahead of schedule and at half the expected cost.

The only significant U.S. activity in carbon dioxide trading now is at the state level. Ten northeastern states have established a regional initiative to explore such a market. Meanwhile the administration sits on the sidelines. That's not good enough: it needs to show specific, decisive, meaningful leadership today.

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
212-754-0550
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

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

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

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

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

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

Direct Response
A & A Media Options, LLC
203-267-4251
fax: 203-267-1552

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
Yogesh Rao
Convergence Media
+91-22-2414-4808
fax: +91-22-2414-5594

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

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

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

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

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

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

COURTESY OF NASA/JPL-CALTECH

Stardust Space Probe Flies by Comet

The first mission to collect a sample from a body beyond the moon and return it to Earth, the Stardust space probe has successfully made its close approach to Comet Wild-2. During the encounter, Stardust deployed a dust collector roughly the size and shape of a large-head tennis racket. The gathered dust, ranging in size from a few to a few hundred microns, is thought to be a piece of the swirling cloud from which the planets emerged.

Unmaking Memories: Interview with James McGaugh

In the recent sci-fi movie *Paycheck,* a crack reverse engineer helps companies to steal and improve on the technology of their rivals and then has his memory of the time he spent working for them erased. The plot, based on Philip K. Dick's short story of the same name, is set in the near future, but such selective memory erasure is still highly speculative at best. ScientificAmerican.com asked neurobiologist James McGaugh of the University of California at Irvine, who studies learning and memory, to talk about what kinds of memory erasure are currently possible.

Ask the Experts

Why do people snore?
Lynn A. D'Andrea, a sleep specialist at the University of Michigan Medical School, explains.

SIGN UP NOW AND GET INSTANT ONLINE ACCESS TO SCIENTIFIC AMERICAN DIGITAL
www.sciamdigital.com
MORE THAN JUST A DIGITAL MAGAZINE!

Three great reasons to subscribe:

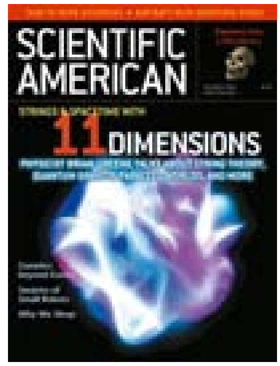
Get it first—all new monthly issues before they reach the newsstand.
10+ years of *Scientific American*—more than 140 issues, from 1993 to the present.
Enhanced search—easy, fast, convenient.

SUBSCRIBE TODAY AND SAVE!

Letters

EDITORS@SCIAM.COM

SCIENCE IS A PROJECT in a constant state of revision. Theories are tweaked, probabilities adjusted, limits pushed, elements added, maps redrawn. And every once in a while, a whole chapter gets a rewrite. In the November 2003 issue of *Scientific American,* "The Unseen Genome," by W. Wayt Gibbs, reviewed one such change currently under way in genetics as new research challenges the long-respected central dogma. In the field of space technology, "The Asteroid Tugboat," by Russell L. Schweickart, Edward T. Lu, Piet Hut and Clark R. Chapman, posited a new way to divert unpredictable Earth-bound asteroids. Reader reactions to these and other innovative ideas from the issue follow.



GENOME REVIEW

In "The Unseen Genome," W. Wayt Gibbs deplores the dogmatism that led biologists to write off large parts of the genome as junk and prevented them from recognizing several processes that may play an important role in heredity. I want to suggest a different perspective: This narrow focus by the research community led to detailed discoveries that have, in turn, challenged the guiding dogma and done so in a relatively short time on the scale of human history.

Closely constrained communal research may be a more effective long-term means of pursuing knowledge than research in which resources are continually diverted to following up any apparent lead. The idea that tightly organized research leads (despite itself) to the recognition of anomalies that generate new approaches was one of the themes of Thomas S. Kuhn's *The Structure of Scientific Revolutions*. This theme was largely forgotten by those who read Kuhn as attacking science, whether their aim was to defend science or join in the supposed attack.

Harold I. Brown
Department of Philosophy
Northern Illinois University

After reading "The Unseen Genome," we were surprised and disappointed that the author gave all credit for the discovery of riboswitches to Ronald R. Breaker's lab. We made this finding independently of Breaker; our paper in *Cell* describing two riboswitch families at once was published at the same time as the Breaker group's ("Sensing Small Molecules by Nascent RNA," by Mironov et al. in *Cell,* Vol. 111, No. 5, pages 747–756; November 27, 2002). Moreover, Gibbs refers to Breaker's August 2003 paper reporting that one family of riboswitches regulates the expression of no fewer than 26 genes. Our paper describing that same family of riboswitches ran several months earlier ("The Riboswitch-Mediated Control of Sulfur Metabolism in Bacteria," by Epshtein et al. in *PNAS USA,* Vol. 100, No. 9, pages 5052–5056; April 29, 2003).

Evgeny Nudler
Department of Biochemistry
New York University School of Medicine

SOLAR SOLUTIONS

"The Asteroid Tugboat," by Russell L. Schweickart, Edward T. Lu, Piet Hut and Clark R. Chapman, discussed using larger launch vehicles and possibly nuclear push mechanisms to deflect threatening asteroids into unthreatening orbits. These ideas unnerved my sense of simplicity. After reading Philip Yam's story about solar sails ["Light Sails to Orbit," News Scan], I wonder if painting the asteroid silver would turn the whole spinning nugget into a "solar sail" opposed to the sun and if this method would alter the orbit. Would the solar wind be enough to push such a painted asteroid away?

David T. Hanawalt
via e-mail

SCHWEICKART AND CHAPMAN REPLY: *A similar proposal was raised by J. N. Spitale in the April 5, 2002, issue of* Science *(Vol. 295, page*

77]. Spitale's proposal calls on the potentially more powerful Yarkovsky effect, in which emission of thermal photons changes an asteroid's momentum, rather than pressure from the solar wind [light pressure], but it is roughly the same idea. Recent and relevant information about the Yarkovsky effect is online at http://neo.jpl.nasa.gov/news/news141.html. There are practical problems with painting a whole asteroid, and no design has been looked at seriously yet. Attaching an actual, separate and necessarily large solar sail to an asteroid has also been proposed but likewise presents serious engineering challenges.

ASTRO LOTTO

When reflecting on the odds estimate presented in "Penny-Wise, Planet-Foolish" [SA Perspectives]—"every year Earth has a one-in-600,000 chance of getting whacked by an asteroid wider than one kilometer"—I found the lottery ticket in my hand to be quite disconcerting. To harvest the $160-million bounty on my ticket, I would have to beat the winning odds of 1:120,526,770, yet I'm willing to invest. While looking over the odds assigned to the remaining prizes, I find I have a similar chance of winning the $5,000 as perishing in the wake of an asteroid this year. Thanks for making me aware, I think.

Nicholas Kulke
Madison, Wis.

CALL FOR BETTER BAFFLERS

"Baffling the Bots," by Lee Bruno [Innovations], left one important question unanswered: How do Web visitors with visual impairments use a service that is guarded with such visual trickery? Web sites that use CAPTCHAs (for "completely automated public Turing test to tell computers and humans apart") and similar barriers to bots need to provide alternative access paths for users who are no less human for being visually impaired!

Carl Zetie
Waterford, Va.

SOLAR-SAIL SUPPORT

"Light Sails to Orbit," by Philip Yam [News Scan], correctly described the emerging interest in solar-sail technology in the aerospace community but incorrectly leaves the impression that NASA is unwilling to support solar-sail development efforts in the private sector. Further, the article's claim that the Cosmos 1 mission is the "lone player" in the private development of solar sails for spaceflight is also incorrect.

COSMOS 1 **is one of the many team efforts to harness the solar winds.**

COURTESY OF THE PLANETARY SOCIETY AND COSMOS STUDIOS; IMAGE PREPARED BY BABAKIN SPACE CENTER

Since 1999 Team Encounter has been developing a series of privately financed solar-sail missions. Our sailcraft technology, developed with our partner L'Garde, represents a significantly different approach from that of Cosmos 1 and has been well received and supported by NASA as well as the National Oceanic and Atmospheric Administration.

Charles M. Chafer
President, Team Encounter
Houston

YAM RESPONDS: Certainly many groups around the world are committed to solar sailing besides the Cosmos 1 team. The German space agency, for instance, is close to a test launch. And, as I noted in the story, NASA spends millions every year researching such advanced propulsion systems. I also wrote that NASA chose to be a bystander in the Cosmos 1 flight, not in solar-sail technology as a whole. Indeed, I described the kinds of goals NASA seeks in a test flight. Such goals are not part of the Cosmos 1 flight, which is meant to demonstrate feasibility and helps to explain

Letters

why NASA is not participating. Future test flights of more complex sail designs by Team Encounter and other groups would do much to push solar-sail technology forward.

Thomas Gold's assertion, noted in the marginalia of "Light Sails to Orbit," that the solar sail cannot work because "perfect mirrors do not create temperature differences, which are necessary to convert heat into kinetic energy," is false, because the force results from radiation pressure, not heat. Radiation pressure, given by the power flux divided by the speed of light, follows from 19th-century physics, specifically electrodynamics. The existence of this force was verified at least as early as 1901 using a torsional balance and has been used recently to manipulate small objects. The solar-sail concept is on firm theoretical and experimental ground.

Thomas G. Moran
NASA Goddard Space Flight Center

TWO TAKES ON TELLER

As a longtime reader of your magazine, I was appalled at the bad taste of Gary Stix's obituary of Edward Teller [News Scan]. Contrary to Isidor Rabi's ill-tempered political opinion, Teller's contributions were significant in keeping the Soviet threat in check and preserving the freedoms of the West.

Georgette P. Zoltani
Lutherville, Md.

I find it hard to believe that Stix defended Teller, stating that Isidor Rabi's comment that the world would have been a better place without Teller was "unquestionably harsh." I might also add that most of the important breakthroughs regarding the hydrogen bomb were the result of Stanislaw Ulam's work and brains, not Teller's.

Joseph Michael Cierniak
Glen Burnie, Md.

ERRATUM In "The Unseen Genome," by W. Wayt Gibbs, the statement that riboswitches have been extracted from species "in all three kingdoms of life" should have read "in all three domains of life."

Gamow ▪ Darwin ▪ Faraday

MARCH 1954

CRUNCH, BANG—"A theory which suggests that our Universe started from an extremely compressed concentration of matter and radiation naturally raises the question: How did it get into that state? Relativistic formulae tell us that various parts of the Universe are flying apart with an energy exceeding the forces of Newtonian attraction between them. Extrapolating these formulae to the period before the Universe reached the stage of maximum contraction, we find that the Universe must then have been collapsing, with just as great speed as it is now expanding! Thus, we conclude that our Universe has existed for an eternity of time; that until about five billion years ago it was collapsing uniformly from a state of infinite rarefaction; and that the Universe is currently on the rebound, dispersing irreversibly toward a state of infinite rarefaction. —George Gamow"

MARCH 1904

DARWIN'S ATOLL—"Darwin had earnestly desired a fuller examination of coral reefs, *in situ,* and in fact went so far as to express his conviction (in a letter to Agassiz in 1881) that nothing really satisfactory could be brought forward as contributory evidence on their origin until a boring was made in one of the Pacific or Indian atolls, and a core obtained down to a depth of at least 500 feet. That hoped-for consummation has, however, been over-achieved, since the boring of Funafuti was carried down to a limit of 1,114 feet, during the third expedition to this ring-shaped spot of land in the South Pacific. The evidence derived goes to show that the material appears to be entirely of organic character, traceable to the calcareous skeletons of marine invertebrate animals and calcareous algae."

ABRUZZI IN THE ARCTIC—"Great interest attaches to the polar expeditions of His Royal Highness Luigi Amedeo of Savoy, Duke of the Abruzzi. The 'Polar Star' was to sail as far to the north as possible along some coast line, and then a party was to travel on sledges toward the pole. The pole was not reached, but a latitude was reached which no man had previously attained, and it was proved that with determination and sturdy men and a number of well-selected dogs, the frozen Arctic Ocean can actually be crossed to the highest latitude. However, at the Emperor Franz Josef archipelago, the ice field trapped and threatened to sink the boat. Therefore, the crew were obliged to land with the utmost haste the stores for winter [*see illustration*], and to secure the necessary materials for building a dwelling. A retreat was carried out in the following spring."

POLAR STAR trapped in the ice, Arctic Ocean, 1904

MARCH 1854

A FARADAY LECTURE—"The opening lecture of the Royal Institution of London was delivered by Michael Faraday to a very crowded audience. The subject was the development of electrical principles produced by the working of the electric telegraph. To illustrate the subject, there was an extensive apparatus of voltaic batteries, consisting of 450 pairs of plates, and eight miles of wire covered with gutta-percha, four miles of which were immersed in tubs of water. The principal point which Professor Faraday was anxious to illustrate was the confirmation—which experiments on the large scale of the electric telegraph have afforded—of the identity of dynamic or voltaic electricity with static or frictional electricity."

DINO DINER—"Professor Richard Owen was recently entertained at dinner in the garden of the Crystal Palace at Sydenham, in the model of an Iguanadon. The animal in whose mould the dinner was given was one of the former inhabitants of Sussex, several of his bones having been found near Horsham. His dimensions have been kept strictly within the limits of anatomical knowledge. The length from the snout to the end of the tail was 35 feet. Twenty-one gentlemen dined comfortably within the interior of the creature, and Professor Owen sat in his head as substitute for brains."

VACCINES

Egg Beaters

FLU VACCINE MAKERS LOOK BEYOND THE CHICKEN EGG **BY KAREN HOPKIN**

If you want to make an omelet, you have to break some eggs. And if you want to supply the U.S. with flu vaccine, you have to break about 100 million.

That may change someday, as leading vaccine manufacturers explore the possibility of trading their chicken eggs for stainless-steel culture vats and growing their flu virus in cell lines derived from humans, monkeys or dogs. The technology could allow companies to produce their vaccines in a more timely and less laborious manner and to respond more quickly in an emergency.

OVER EASY? Researchers hope to replace the decades-old way of making flu vaccines, which involves injecting viruses into fertilized eggs pierced with a drill.

BETTMANN/CORBIS

Today's flu vaccines are prepared in fertilized chicken eggs, a method developed more than 50 years ago. The eggshell is cracked, and the influenza virus is injected into the fluid surrounding the embryo. The egg is resealed, the embryo becomes infected, and the resulting virus is then harvested, purified and used to produce the vaccine. Even with robotic assistance, "working with eggs is tedious," says Samuel L. Katz of the Duke University School of Medicine, a member of the vaccine advisory committee for the U.S. Food and Drug Administration. "Opening a culture flask is a heck of a lot simpler."

Better yet, using cells could shave weeks off the production process, notes Dinko Valerio, president and CEO of Crucell, a Dutch biotechnology company developing one of the human cell lines. Now when a new strain of flu is discovered, researchers often need to tinker with the virus to get it to reproduce in chicken eggs. Makers using cultured cells could save time by skipping that step, perhaps even starting directly from the circulating virus isolated from humans. As an added bonus, the virus harvested from cells rather than eggs might even look more like the virus encountered by humans, making it better fodder for a vaccine, adds Michel DeWilde, executive vice president of R&D at Aventis, the world's largest producer of flu vaccines and a partner with Crucell in developing flu shots made from human cells.

Whether vaccines churned out by barrels of cells will be any better than those produced in eggs "remains to be seen," says the FDA's Roland A. Levandowski. And for a person getting jabbed in the arm during a regular flu season, observes Richard Webby, a virologist at St. Jude Children's Research Hospital in Memphis, Tenn., "it's not going to matter where the vaccine came from."

Where the cell-based vaccine will become invaluable, Webby states, is in the case of a global pandemic. Should a new strain of flu crop up outside the normal season—one that is different enough from previous strains that people will have no immunity—cell-based systems will allow health officials to respond more rapidly. "Cell cultures are a lot easier to scale up faster," he explains. Technicians would simply remove cells from a freezer and grow them in large volumes—something that is not possible with chicken eggs. Although flocks of chickens kept in clean environments are available almost year-round, companies generally place their egg orders six months before they start vaccine production. And preventing a pandemic could require 10 times as much vaccine as a normal flu season. "If halfway into manufacturing, you need a billion more eggs, you're not going to get them," remarks Wayne Morges, a vice president at Baxter in Deerfield, Ill.

Preparing vaccines in cell cultures is not new. Aventis, for example, currently produces polio vaccines in the same monkey kidney cells that Baxter is gearing up to use to produce flu injections. And Baxter used the monkey cell line to replenish the U.S. supply of smallpox vaccine. So converting to cell-based systems, Katz says, would be "moving flu vaccine production into the 20th century at the beginning of the 21st."

Why has it has taken manufacturers so long to come around to considering cell-based systems? Perhaps because current egg-based systems work so well, Webby surmises. Up-front costs for preparing production plants to function with cells rather than eggs might also be an impediment.

Clinical trials of cell-based flu vaccines won't begin in the U.S. until this fall, and if approved, the new vaccines will at first probably just supplement those produced in chicken eggs. Having several different formulations of flu vaccine can't hurt. Except maybe for that muscle soreness that lingers for a day or two after you roll up your sleeve.

Karen Hopkin is based in Somerville, Mass.

ALTERNATIVE MEDICINE: Researchers, including Richard Webby, a virologist at St. Jude Children's Research Hospital in Memphis, Tenn., hope to speed influenza vaccine manufacturing by coming up with new options to the chicken egg as a virus growth medium.

GREG CAMPBELL *AP Photo*

VIRAL TIMETABLES

For the Northern Hemisphere, the flu season typically runs from November through March. Based on collected virus samples and infection activity, the World Health Organization decides which influenza strains to include in a vaccine in mid-February. By mid-March, high-growth strains of vaccine virus are provided to manufacturers, and the materials needed to test the identity and potency of the resulting vaccine are supplied in mid-May. Vaccines become available in clinics in October.

Number of U.S. flu cases per season: 29 million to 58 million

Number of Americans hospitalized per season: 114,000

Number of deaths: 36,000

Number of vaccine doses produced this season: 87.1 million

PHYSICS

Dream Machine

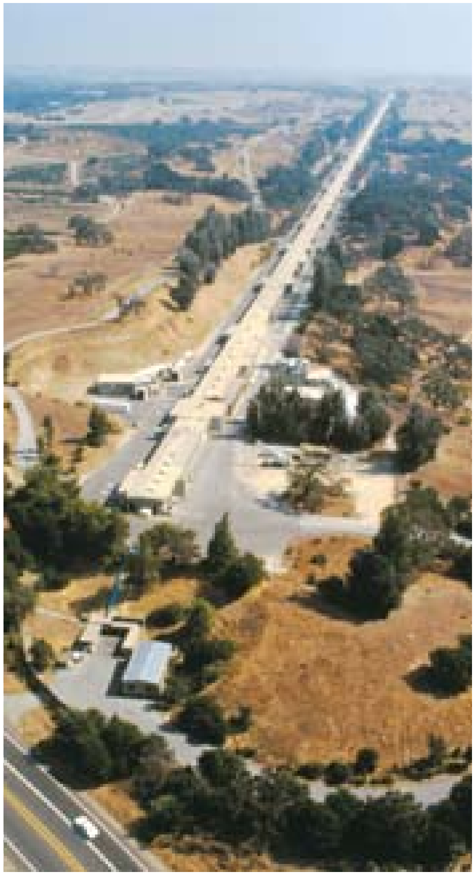
HOPES FOR A GIANT COLLIDER LIE IN A WORLDWIDE APPEAL **BY DAVID APPELL**

High-energy physicists have a new machine in mind: an unprecedented accelerator 30 kilometers long that would offer a precise tool to explore some of the most important unanswered questions in physics. But the specter of the defunct Superconducting Supercollider—and the money the project ended up wasting—looms large. Advocates of the machine, however, think they can overcome national doubts by going global.

Since they first began discussing a linear collider in earnest at a 2001 conference at Snowmass, Colo., the world's physicists have consistently and vigorously planned an international effort. Their hopes recently rose when U.S. Secretary of Energy Spencer Abraham named it the highest "midterm" priority in a

20-year outlook of new science facilities. The report estimates that were the project to be approved and funded, peak spending would occur sometime between 2010 and 2015.

The vision is of one machine built by the world and shared by the world. "Many people have been working very hard to make this more than an empty slogan," says theorist Chris Quigg of the Fermi National Accelerator Laboratory in Batavia, Ill., because no one government seems likely to spend the estimated $5 billion to $7 billion that such a facility would cost.



DOWN THE LINE: The 3.2-kilometer-long tunnels of the Stanford Linear Accelerator Collider would be dwarfed by the proposed International Linear Collider, which would be five times as long.

The plan is to accelerate electrons and positrons (the antimatter version of the electron) down dual 15-kilometer pipes and smash them together inside a large detector. The total energy would be up to one trillion electron volts (TeV). This energy may appear much less than the 2-TeV Tevatron at Fermilab and the 14-TeV Large Hadron Collider to be completed at CERN in 2007, but because the particles in those machines share their energy among their constituent quarks, their effective energy drops by about a factor of 10. By design, the international linear collider will have higher interaction rates, and because the spins of the particles in its beams are aligned—something that cannot be done at the Tevatron or Large Hadron Collider—it will be much more precise in dissecting and analyzing particle interactions.

The collider could reveal the specifics of Higgs bosons (particles that imbue all other particles with mass) and light supersymmetric particles (shadowy particles such as the neutralino, which may account for the dark matter that constitutes 23 percent of the universe). That knowledge could in turn open the door to exotica such as extra dimensions and low-energy superstring phenomena. "That's the exciting thing about the linear collider," says theorist Joseph Lykken of Fermilab. "It gives you a window into this whole other realm of physics that we're really interested in."

But opening that window requires cold, hard cash. The last time particle physicists asked for dollars for an accelerator, two billion of them ended up underneath the Texas prairie in now water-filled tunnels meant for the Superconducting Supercollider. "The story of its demise is so complicated, it's fair to say it died of fluctuations," Quigg remarks. "Our community hopes to have learned from the experience to organize future projects so they will be less vulnerable to fluctuations and political tussles."

In fact, several groups in the U.S., Europe and Japan are committed to the linear collider. "We are all behind it," states Albrecht Wagner, director of the DESY high-energy laboratory in Hamburg, Germany, acknowledging that in the end the project's site will be a political decision, not unlike that now being made about the fusion reactor called ITER.

So far the early politics involve technology recommendations. To accelerate particles, DESY backs a superconducting, lower-radiofrequency cavity; a higher-frequency, room-temperature structure is being championed by a collaboration between the Stanford Linear Accelerator and the KEK Accelerator Laboratory in Tsukuba, Japan. Given the history of grand accelerators, deciding on which approach to take will no doubt be the easy part.

David Appell is based in Lee, N.H.

LINDA A. CICERO *Stanford News Service*

PHYSICS WISH LIST

A linear collider came in at 13th on a list of 28 future science facilities, behind the international fusion reactor project ITER (first), and the UltraScale Scientific Computing Capability (second), which aims to increase scientific computing capacity 100-fold. Four projects tied for third: the Joint Dark Energy Mission; an intense x-ray laser called the Linac Coherent Light Source; a facility to mass-produce, characterize and tag tens of thousands of proteins; and the Rare Isotope Accelerator. Notably, the linear collider ranked ahead of several other competing physics projects, such as a superneutrino beam and upgrades to Brookhaven National Laboratory's Relativistic Heavy Ion Collider. The entire list is at www.er.doe.gov/Sub/Facilities_for_future/facilities_future.htm

DEFENSE

The Fog of War

CAN HIGH-TECH SENSORS FIGHT THE INSURGENCY IN IRAQ? BY MARK ALPERT

U.S. soldiers in Iraq face a bewildering array of threats. Since American and British troops occupied the country last spring, Iraqi insurgents have downed helicopters with heat-seeking missiles, detonated roadside bombs along the routes of army convoys and launched mortar rounds at U.S. bases. One of the biggest frustrations is the elusiveness of the enemy: the insurgents typically slip away before American forces can respond to an attack.

WHILE PATROLLING the streets in Iraqi cities, U.S. soldiers have proved vulnerable to sniper attacks.

Now the Pentagon's R&D arm, the Defense Advanced Research Projects Agency (DARPA), is trying to provide some high-tech assistance. The agency is pushing to deploy experimental systems that could quickly locate the positions of enemy snipers and mortar crews. One of the most startling examples is a ground-based carbon dioxide laser designed to pinpoint a sniper by measuring the movements of dust particles in the air caused by the shock wave of a speeding bullet. DARPA director Anthony J. Tether announced last fall that the anti-sniper laser, which would reportedly have a range in the tens of kilometers, would be sent to Iraq early this year.

Developed by Mission Research Corporation, a defense contractor based in Santa Barbara, Calif., the system relies on a Doppler lidar, a laser radar that can measure the velocity of moving objects in much the same way that a radar gun gauges the speed of cars on the highway. Because the wavelength of the laser light is roughly comparable to the diameter of a dust particle—about one to 10 microns—some of the light will scatter when it encounters airborne dust. The frequency of the scattered light will be higher if the dust particles are moving toward the laser and lower if the particles are moving away. By analyzing the returning signals, the Doppler lidar can determine wind velocities; in fact, these systems already find use in studies of the atmosphere and at airports to detect wind shear and other turbulence.

Some defense analysts, however, are skeptical that such a device could track a bullet. Because the shock wave would be so localized and short-lived, the system would need to crisscross the sky with laser beams to pick up signs of the atmospheric disturbance and determine the bullet's trajectory. Another challenge would be distinguishing between a sniper's gunshot and bullets fired by friendly forces or by civilians shooting into the air in celebration (a fairly common occurrence in Baghdad and other Iraqi cities). Says Philip E. Coyle, who was the Pentagon's director of testing and evaluation during the Clinton administration: "Before you can let the troops shoot back, you need a high-confidence system producing accurate results."

Although it is unusual for the military to field experimental prototypes in war zones, DARPA spokesperson Jan Walker notes that it is not unprecedented. For example, the airborne surveillance system known as JSTARS was deployed in Bosnia in 1996, and the unmanned Global Hawk reconnaissance aircraft was rushed into battle in Afghanistan in 2001. But the success rate for new military technologies is not inspiring: during the 1990s, the great majority of army systems that went into operational testing achieved less than half their required reliability, and most air force tests had to be halted because the systems were simply not ready.

Walker says the Pentagon is confident that the anti-sniper laser will prove useful to the soldiers in Iraq. But Coyle, who is now a senior adviser at the Center for Defense Information, a Washington, D.C., think tank, is less optimistic. "There's nothing wrong with trying it to see if it works," he says. "But often these things don't pan out."

HIDDEN ENEMIES

Iraq is not the first place where the U.S. military has attempted to use novel sensors to detect an elusive enemy. During the Vietnam War, the U.S. Air Force dropped 20,000 battery-powered devices into the jungle along the Ho Chi Minh Trail, the main supply route for the North Vietnamese army. The devices—seismic detectors implanted in the ground and camouflaged acoustic sensors hanging from the trees—picked up the movements of troops and supply trucks, and the transmitted signals were used to target bombing runs. The air force claimed that the operation, dubbed Igloo White, destroyed tens of thousands of trucks, but later studies indicated that the kill figures had been wildly inflated. North Vietnamese soldiers apparently disabled many of the devices and deceived others with tape-recorded truck noises.

ALI HAIDER *EPA/AP Photo*

CHEMISTRY

Drawing the Lines

IS A PRE-COLUMBUS MAP OF NORTH AMERICA TRULY A HOAX? BY CHARLES CHOI

In a Yale University library sits a map depicting the New World that predates the landing of Columbus by 60 years—if it isn't a fake. Although the lines on the so-called Vinland map are faded, those between scientists on the controversy are sharp. New salvos regarding its authenticity now come from both sides.

The parchment map, about 11 by 16 inches large, was uncovered in a Geneva bookshop in 1957 with no records of prior ownership. To the west of the inscriptions of Europe, Africa and the Far East are the words "a new land, extremely fertile and even having vines." The writing also says the crew of Leif Eriksson named the land "Vinland."

In 2002 Jacqueline S. Olin, retired from the Smithsonian Center for Materials Research and Education in Suitland, Md., and her colleagues reported results of carbon dating indicating that the map dates from 1434, give or take 11 years. That finding bolstered three decades of speculation linking it to the Council of Basel, convened in Switzerland by the Catholic Church from 1431 to 1449. There scholars from around Europe assembled to discuss important affairs, such as the rift in the papacy and the possible reunion of the Eastern and Western Churches. "The fact that it existed in the 15th century certainly presents the very real possibility of Columbus, or someone in contact with him, having some knowledge of the map," Olin says.

But since the map's discovery, critics have called it a clever fake. What lies in dispute is not the pre-Columbian age of the parchment but that of the map drawn on it. At the same time Olin and colleagues dated the map's parchment, chemists Katherine Brown and Robin Clark of University College London argued that the map's ink dated from after 1923. The ink contained jagged yellow crystals of anatase, a titanium-bearing mineral rarely found in nature that became commercially available in 20th-century printing ink. "The whole points to an elaborate forgery," Clark states.

Dueling papers appeared again in recent months. With medieval methods, Olin made iron gall inks, which were used before the printing press. She found that her inks contained anatase, results she discusses in the December 1, 2003, issue of *Analytical Chemistry*. She adds that the anatase crystals in the map and her inks were the same size, citing the electron microscope work of geologist Kenneth M. Towe, retired from the Smithsonian Institution. Those crystals found in modern inks should be about 10 times as large.

VINLAND MAP contains references to a new world to the west.

BEINECKE RARE BOOK AND MANUSCRIPT LIBRARY, YALE UNIVERSITY

Towe vociferously disagrees with Olin's interpretation of his work in a report appearing online in January in *Analytical Chemistry*. He concludes that the map's anatase crystals look modern in size. Moreover, he notes that whereas a map drawn with iron gall inks would reasonably be expected to contain iron, "there's hardly any there."

Olin responds by suggesting that iron might have disappeared as the inks deteriorated. Regarding the anatase crystal sizes, she concurs with Towe but says many other inks contain titanium and should be researched further to see what sizes are present. She adds that the presence of copper, zinc, aluminum and gold in the map's ink are also consistent with medieval manufacturing.

Historian Kirsten A. Seaver, a fellow of the Royal Geographical Society in London, states that the map's writing contains historical anachronisms such as mention of Bishop Eirik of Greenland of the early 12th century reporting to superiors, although he would have had none, because Greenland had not yet become part of the Church hierarchy. "This map absolutely screams 'fake,'" Seaver remarks. In fact, she believes she has found the culprit—a German Jesuit priest, Father Josef Fischer, a specialist in mid-15th-century world maps. Her theory is that Fischer created the map in the 1930s to tease the Nazis, playing on their claims of early Norse dominion of the Americas and on their loathing of Roman Catholic Church authority. The map, she supposes, vanished during postwar looting. Seaver's book on her search will appear this June.

Charles Choi is based in New York City.

GEOMAGNETICS

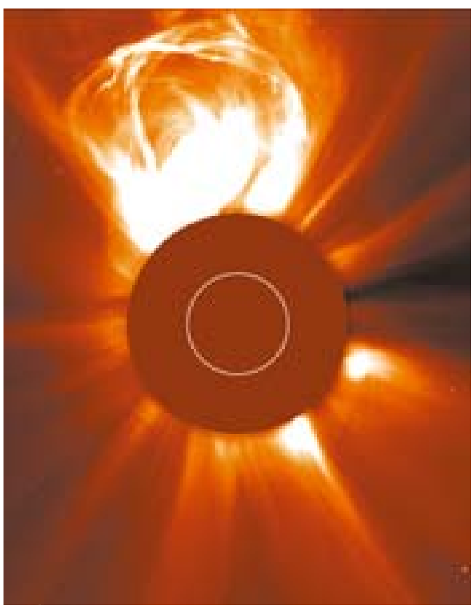
Storm Spotting

A STEP CLOSER TO FORECASTING DISRUPTIVE SOLAR ACTIVITY **BY KRISTA WEST**

On October 19, 2003, a large solar flare erupted from the surface of the sun, drawing scientists' attention to three massive sunspot groups that, over the next two weeks, produced a total of 124 flares. Three of them were the biggest flares ever recorded. Along with these bursts of electromagnetic radiation came enormous clouds of plasma mixed with magnetic fields. Known as coronal mass ejections (CMEs), these unpredictable clouds consist of billions of tons of energetic protons and electrons. When directed earthward, CMEs can create problems. At last count, the fall's flares and CMEs affected more than 20 satellites and spacecraft (not including classified military instruments), prompted the Federal Aviation Administration to issue a first-ever alert of excessive radiation exposure for air travelers, and temporarily knocked out power grids in Sweden.

Historically, CMEs have struck the earth with little or vague warning. If they could be forecast accurately, like tomorrow's weather, then agencies would have time to prepare expensive instruments in orbit and on the ground for the correct size and moment of impact. Such precise predictions could soon emerge: last December researchers announced the early success of a forecasting instrument, called the Solar Mass Ejection Imager (SMEI), that can track CMEs through space and time.

Launched in January 2003 on a three-year test run, SMEI (affectionately known as "schmee") orbits the planet over the poles, along the earth's terminator, once every 101 minutes. On each orbit, three cameras capture



SUN BURPS UP a bulb-shaped cloud called a coronal mass ejection, as seen in February 2000 by the sun-watching satellite SOHO. The mask blots out direct sunlight; the white circle denotes the sun.

SIMON PLUNKETT *SOHO/LASCO* (*ESA and NASA*)

BRIGHT LIGHTS, BIG PROBLEMS

Autumn 2003 saw two weeks of intense solar activity. The most serious disruptions of the earth's electronics systems stem from coronal mass ejections (CMEs).

October 19
Three massive sunspots rotate to face the earth.

October 22–23
First geomagnetic storm, triggered by a CME, strikes the earth.

October 28
The second-largest flare ever recorded erupts from the sun.

October 28–30
First-ever radiation alert goes out to air travelers above 25,000 feet.

October 29
Second CME-triggered geomagnetic storm hits the earth.

November 4
The biggest solar flare ever recorded erupts; fortunately, the sun has rotated enough so that no disruptive radiation strikes the earth.

images that, when pieced together, provide a view of the entire sky with the sun in the middle. The scattering plasma electrons of CMEs appear on SMEI images as bright clouds.

Other sun-watching instruments can image CMEs, but they work like still cameras, taking single pictures of the sun. NASA's Solar and Heliospheric Observatory (SOHO), for example, can "see" CMEs erupting from the sun quickly but is soon blind to the path of the clouds. SOHO came in handy last fall when it caught two large CMEs headed for the earth, but it could not follow the ejecta nor provide an accurate impact time.

Instead of a SOHO-style snapshot camera, SMEI works more like a 24-hour surveillance system, constantly scanning and tracking. SMEI begins looking about 18 to 20 degrees from the sun and continues imaging beyond the earth. SMEI can determine the speed, path and size of a CME, allowing for refined and reliable impact forecasts. Such information is particularly useful, scientists say, in predicting small CME events. Such ejections can take anywhere from one to five days to reach our planet. Since its launch, SMEI has detected about 70 CMEs.

During last fall's solar storms, SMEI had its first big chance to prove worthy of its estimated $10-million price tag. Managed primarily by the Air Force Research Laboratory at Hanscom Air Force Base in Massachusetts, about 20 air force and university scientists have been developing SMEI over the past 20 years. At the December 2003 American Geophysical Union meeting in San Francisco, Janet Johnston, SMEI's program manager, proudly announced that SMEI had successfully detected two of the autumn's largest CMEs about 21 and 10 hours, respectively, before they struck the earth.

Unfortunately, scientists didn't know of the detection and tracking potential until after the storms hit the earth. Right now it takes about 24 hours for SMEI data to reach Hanscom because they travel through multiple ground-tracking stations. According to David F. Webb, a physicist at Boston College who is part of the SMEI team, precise forecasting demands a reduction in data-transmitting time from 24 to six hours. Such a reduction will require more researchers at ground-tracking stations to move information along and to inspect SMEI's output.

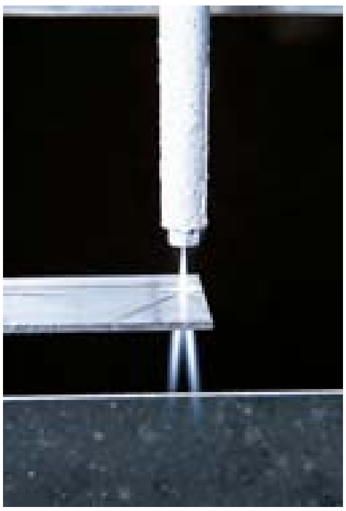
SMEI's data gathering may also need perfecting. Lead forecaster Christopher Balch of the Space Environment Center in Boulder, Colo., emphasizes that the CME signal must stand out better against other background light. Once improved, SMEI "could potentially fill a gap in our observations," Balch says, by allowing scientists to track CMEs precisely, thereby making "real-time" forecasts possible.

Krista West is based in Las Cruces, N.M.

TOOLS

Cryogenic Cutting

LIQUID-NITROGEN JET SLICES AND SCOURS ALMOST ANYTHING BY STEVEN ASHLEY



LIKE A KNIFE through warm butter, a high-pressure jet of liquid nitrogen hews through hard materials, then disappears into thin air.

Late-night television was once awash in a commercial hawking the "amazing Ginsu knife" that never needed sharpening. In the infamous ad, the blade carved through tin cans with ease and then deftly cut paper-thin slices of tomato. Engineers have recently produced an innovative industrial cutting device with Ginsu knife–like capabilities that uses a supersonic stream of high-pressure liquid nitrogen. The so-called Nitrojet slices through just about anything—steel girders, concrete slabs, stacks of fabric, meat carcasses—and never gets dull.

Nitrojet technology was originally developed in the 1990s by scientists at the Idaho National Engineering Laboratory (INEL) as a nonthermal method to cut open barrels of combustible waste. Ron Warnecke, president of TRUtech, an Idaho Falls–based firm that handles decontamination and decommissioning efforts for nuclear weapons facilities, stumbled on the still developmental system in the late 1990s when he was searching for an environmentally safe way to clean and cut up plutonium-processing equipment. TRUtech later licensed the technology and developed INEL's prototype into a salable product. Warnecke has since set up a new company, NitroCision, to market the device.

The supercooled nitrogen jet, which emerges from special nozzles fitted to a handheld or robotically positioned wand, seems to cleave materials so well because the dense liquefied gas enters a solid's cracks and crevices and then expands rapidly, breaking it up from the inside. The effectiveness of the process for various applications depends on the pressure (6,000 to 60,000 pounds per square inch), temperature (300 to –290 degrees Fahrenheit) and distance to the workpiece chosen by the user. Lower pressures enable the nozzle stream to strip tough-to-remove coatings off even delicate surfaces better than almost any other cleaning process.

Moreover, the cryogenic jet does not create secondary waste or cross-contamination; as the nontoxic, supercooled "blade" warms, it simply vanishes into the air. Hazardous refuse created by stripping or cutting can be vacuumed up at the point of impact.

NASA technicians are now employing a Nitrojet system at the Kennedy Space Center to precisely peel thermal-protection coatings off the inside surfaces of the space shuttle's solid-rocket boosters. Water-jet or similar abrasive-blasting methods would have required the entire internal surface to be processed, Warnecke reports. The U.S. Navy meanwhile has contracted to use Nitrojet units to remove anticorrosion coatings from ship decks and hulls, antennas and radomes. Others testing the technology include aerospace firms Boeing and Northrop Grumman, semiconductor manufacturers Semitool and Rogers, paint producer Sherwin-Williams, Merrimac Industries (makers of polyurethane parts) and meat packers Hormel and ConAgra.

Nitrojet systems, which come on skids measuring four feet by four feet by eight feet, start from $200,000 to $300,000 for a low-pressure unit and go to $450,000 for a full system. These figures represent a considerable premium over the $150,000-plus price tag for a conventional water-jet unit, but advocates of the technology say its unique capabilities are worth the extra cost. But don't expect it to appear on late-night infomercials, no matter how many easy payments are offered.

NITROCISION

BY THE NUMBERS

Rise of the Black Ghetto

HOW TO CREATE AN AMERICAN VERSION OF APARTHEID **BY RODGER DOYLE**

The North was once alive with the abolitionist spirit and open to the possibility of integration. Yet this passion yielded to several forces that marginalized African-Americans in the 20th century.

Before World War I, blacks were relatively few in the North, which together with people's need to be near their factories and offices, helped to reduce any tendency toward housing segregation. In New York City, for example, largely black neighborhoods were usually only a few blocks long and interspersed with the homes of working-class white families. The modern ghetto, with its sharply defined racial lines, generally did not begin to form until blacks in substantial numbers migrated north beginning in 1916. There they found themselves competing for jobs and housing with immigrants from Europe. The competition was often violent, as in the Chicago riot of 1919, when 38 people were killed. Violence and the threat of violence, together with agreements among white homeowners not to sell to blacks, increasingly left African-Americans in separate neighborhoods.

Because blacks had fewer choices, landlords could charge them more than whites. Crowding increased as tenants took in lodgers, and many landlords allowed their properties to become run down. The Federal Housing Administration and the Veterans Administration condoned redlining, the practice of denying mortgages to those in minority neighborhoods, until well into the 1960s.

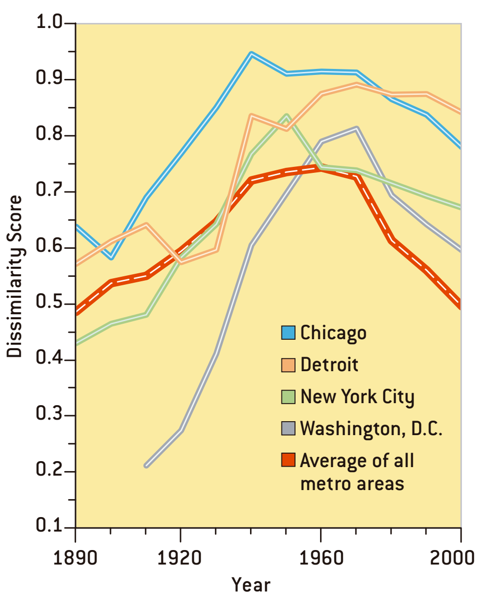
Despite the problems, several communities, notably Harlem, were vibrant, at least until the manufacturing economy began to decline in the 1970s. Other factors in the deterioration include the increasing availability of crack cocaine, the growth of unwed motherhood, higher crime rates as the baby boomers came of age, and the disruptive effects of urban renewal. Churches, social clubs, newspapers and unions in black communities withered, and banks closed their branches, to be replaced by currency exchanges that charged up to $8 for cashing a check.

To measure segregation, economists David M. Cutler and Edward L. Glaeser of Harvard University and Jacob L. Vigdor of Duke University calculated dissimilarity scores, which are defined as the proportion of blacks who would need to move across census-tract lines to achieve the same proportion of blacks in every tract of a metropolitan area. By convention, a dissimilarity index above 0.6 is high, whereas an index of less than 0.3 is low. A score of 0 represents perfect integration and 1.0 complete segregation.

As the chart shows, the average index for all metropolitan areas rose steadily to reach a peak of 0.74 in 1960 and then declined to 0.5 by 2000. But the largest metropolitan areas, particularly in the North, are still on average far above 0.6. Of 291 metropolitan statistical areas, 72 had dissimilarity scores above 0.6 in 2000 and 28 had scores below 0.3. Some of the fastest-growing cities, such as Las Vegas and Phoenix, had low and declining scores. Decreasing scores, however, reflect primarily the dispersion of more affluent blacks into previously white neighborhoods. The northern ghettos and their poverty remain, arguably, the number-one problem in the U.S.

Rodger Doyle can be reached at rdoyle2@adelphia.net

SEGREGATION IN U.S. METROPOLITAN AREAS



LIVING APART

A dissimilarity score is a measure of segregation: above 0.6 represents high segregation, and below 0.3, low. Data are for 2000.

MOST SEGREGATED	Score
Detroit, Mich.	0.84
Gary, Ind.	0.81
Milwaukee, Wis.	0.81
Chicago, Ill.	0.78
Cleveland, Ohio	0.77
Flint, Mich.	0.77
Buffalo, N.Y.	0.76
Cincinnati, Ohio	0.74
LEAST SEGREGATED	
Bellingham, Wash.	0.21
Santa Cruz, Calif.	0.22
Boulder, Colo.	0.23
Boise, Idaho	0.24
Jacksonville, N.C.	0.24
Redding, Calif.	0.25
San Angelo, Tex.	0.25
San Jose, Calif.	0.25

FURTHER READING

Harlem: The Making of a Ghetto. Gilbert Osofsky. Harper & Row, 1966.

Urban Injustice: How Ghettos Happen. David Hilfiker. Seven Stories Press, 2002.

How East New York Became a Ghetto. Walter Thabit. New York University Press, 2003.

SOURCE: "The Rise and Decline of the American Ghetto," by David M. Cutler, Edward L. Glaeser and Jacob L. Vigdor in Journal of Political Economy, Vol. 107, No. 3; June 1999. Data prior to 1950 are based on cities rather than metropolitan areas. Additional segregation data are at http://trinity.aas.duke.edu/~jvigdor/segregation

RODGER DOYLE

DATA POINTS: MADDENING WORLD

The discovery of mad cows in Canada and in the U.S. last year continues the global spread of bovine spongiform encephalopathy (BSE). Assuming that the North American cases represent the same strain of BSE as seen in the U.K., then the risk of getting the human form of BSE, called variant Creutzfeldt-Jakob disease, appears to be low.

BSE cases identified in the U.K., up to December 2003: 180,343

Number thought to have entered the food chain undetected: 1.6 million

Variant Creutzfeldt-Jakob disease cases in the U.K.: 143

Number worldwide: 153

Number of countries that had detected native BSE cases by

1986: 1
1990: 3
1995: 5
2000: 12
2003: 23

Pounds of U.S. beef produced, 2002: 27.1 billion

Pounds exported, 2002: 2.45 billion

Pounds of beef consumed annually, per capita: 67.7

Percent consumed as ground beef: 43.2

SOURCES: U.K. Department for Environment, Food and Rural Affairs; Proceedings of the Royal Society, November 7, 2002; U.K. Department of Health; Centers for Disease Control and Prevention; World Organization for Animal Health; CattleFax; National Cattlemen's Beef Association; U.S. Department of Agriculture.

MECHANICS

Getting into the Swing

SWING PHASE during running uses more energy than previously thought.

Experiments designed to study running mostly take an external view of the mechanics. Biologists at Northeastern University have peered directly at running muscles by measuring blood flow in the legs of the helmeted guinea fowl *Numida meleagris*. Researchers previously suggested that during running, virtually all energy fueling the muscles went to generating force when the foot is on the ground (the stance phase). Now they find that bringing the legs forward (the swing phase) consumed roughly a quarter of the energy used by the hind limbs. Because running birds are the second-best bipedal sprinters after humans, the investigators say their research should provide valuable clues to understanding human locomotion, with potential benefits to rehabilitative medicine. Their report appears in the January 2 *Science*. *—Charles Choi*

BIOLOGY

Making and Unmaking Memories

Prions lie at the root of many disorders, such as mad cow disease and fatal insomnia. But the prion ability to adopt a secondary shape—and force other proteins into that shape—does not always cause cellular malfunctions, as indicated by a protein called CPEB. Experiments show that CPEB, whose normal job involves creating other proteins at synapses during memory formation, has an alternative conformation. Its alter ego is still functional, and it can also reshape other proteins, as described in the December 26, 2003, issue of *Cell*. The prionlike nature of CPEB may help lock in long-term memories, considering that the prion state is typically durable.

Biological activity may also undergird the voluntary suppression of long-term memories, which has remained controversial since Freud. In an experiment, volunteers first memorized pairs of unrelated nouns, such as "ordeal/roach." Then, when looking at the first word of each pair, they were told not to recall its partner. As detailed in the January 9 *Science*, when suppression successfully impaired the recall of the second word, the prefrontal cortex was more active, following a pattern similar to one seen when that brain region stops physical actions. At the same time, the memory-forming hippocampus activated less, suggesting that the prefrontal cortex controlled its behavior. *—Charles Choi*

BURIED: The brain has a biological mechanism to suppress memories.

PHYSICS

Strangeness in Our Midst?

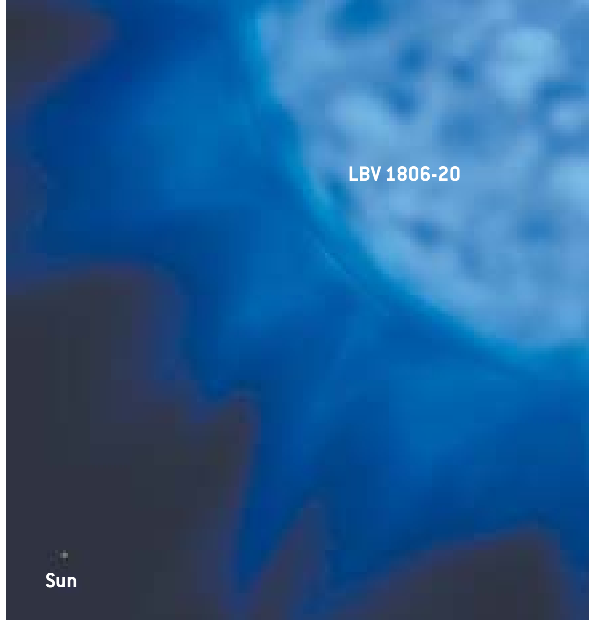
The hot early universe or colliding neutron stars may have coughed up so-called strange quark matter, an extremely dense mix of up, down and strange quarks. If they exist, wayfaring nuggets of strange matter might pierce the earth every few years and, like stones dropped in water, trigger seismic ripples in their wake. Because a strange nugget would far outpace sound underground, seismographs would record it as a simultaneous tremble from many points along a line. Careful sifting through one million seismic reports between 1990 and 1993 revealed one set of reports from November 1993 that has the right properties for a nugget strike, say Vigdor L. Teplitz and his colleagues at Southern Methodist University. Corroborating the result would require scrutinizing new readings in nearly real time. The findings appear in the December 2003 *Bulletin of the Seismological Society of America*. *—JR Minkel*

JON FEINGERSH *Corbis* (*top*); MARC TAUSS *Photonica* (*bottom*); ILLUSTRATION BY MATT COLLINS

ASTRONOMY

A Super Superstar

The Palomar telescope has spied what appears to be the brightest star yet known, a giant so oversized that it defies current theories. The star LBV 1806-20 shines up to 40 million times brighter than the sun. The previous record holder, the Pistol Star, was just roughly six million times as bright. Some 45,000 light-years from Earth, LBV 1806-20 weighs about 150 times as much as the sun, although present theory holds that stars of more than 120 solar masses could not coalesce, because their nuclear fires should burn off the excess. The colossus is surrounded by what the astronomers call "a zoo of freak stars," such as a rare magnetic neutron star. Rather than collapsing under their own gravity, LBV 1806-20 and its freaky neighbors may have formed when a supernova shock wave crushed a nearby molecular cloud into stars. The scientists presented their findings at the January meeting of the American Astronomical Society. *—Charles Choi*



BIGGEST AND BRIGHTEST: The star LBV 1806-20 could swallow at least eight million suns.

GEOCHRONOLOGY

Turning Back the Clock

The decay of radioactive carbon is the chief way to date ancient samples. The radiocarbon clock drifts, however, because isotopes do not accumulate consistently year to year. So researchers calibrate the clock by dating tree rings and other absolute age measures. A group led by Konrad Hughen of the Woods Hole Oceanographic Institution has extended the calibration from 26,000 to the maximum 50,000 years ago (radioactive carbon becomes scarce beyond that age). The researchers matched up radiocarbon-dated layers in marine sediment to annual layers in a Greenland ice core. They had previously shown that the two sets of layers are synchronous from 10,000 to 15,000 years ago, and a French group has obtained evidence for a similar preliminary trend. The January 9 issue of *Science* has more. *—JR Minkel*

ZOOPLANKTON called foraminifera, when fossilized, are used to calibrate radiocarbon dating.

HYDROGEN STORAGE

All Gassed Up

Storing elemental hydrogen for use as a clean fuel requires impractically low temperatures or high pressures. In search of a better storage medium, the daughter-father team of Wendy and David Mao of the University of Chicago and the Carnegie Institution compressed crystals of hydrogen and water or methane with a so-called diamond anvil and cooled them with liquid nitrogen. In one instance, the result was a hydrogen-water clathrate, or cagelike crystal, that retained its 5.3 percent hydrogen by weight when it returned to atmospheric pressure. The amount of hydrogen caged is reasonably high—today's metal hydride batteries hold about 2 to 3 percent—and could easily be released by warming the clathrate. Different additives and pressure and temperature pathways might make such storage crystals more practical. The research appeared online January 7 in the *Proceedings of the National Academy of Sciences USA*. *—JR Minkel*

BRIEF POINTS

- **Supersolid: A new state of matter seems to have emerged after helium 4 was sufficiently chilled under pressure. It turned into a solid whose atoms could, like a superfluid, flow without resistance.**

  Nature, January 15, 2004

- **NASA's Stardust spacecraft flew within 240 kilometers of Comet Wild-2 to collect microscopic grains coming off the object. The samples should reach the earth—specifically, Utah—on January 15, 2006.**

  NASA announcement, January 2, 2004

- **Prostate cancer cells start resisting drugs by making more receptors for androgens, which the cells ordinarily need to proliferate. Blocking those receptors could restore drug efficacy.**

  Nature Medicine online, December 21, 2003

- **Forget about tar levels: The risk of lung cancer for people who smoked even very low tar cigarettes was the same as for those who puffed the conventional variety.**

  BMJ, January 10, 2004

MEGHAN KENNEDY *University of Florida* (top); ALFRED PASIEKA *Science Photo Library* (bottom)

Working the System II

Corporate greed no longer remains the sole domain of the corporation By GARY STIX

Last month this column detailed how a recent lawsuit charged biotech giant Genentech with attempting to retain rights to a technology for more than a decade beyond the original patent's expiration date. These days, however, this type of behavior is by no means confined to the corporate sector. As university patenting has increased dramatically in the years since the Bayh-Dole Act of 1980, the law that encouraged such activity, academic institutions have taken a lesson or three from the corporations whose convoluted tactics keep a white-knuckled lock around valuable patents. Among the ivory tower set, Columbia University, that august Ivy League institution that is now marking its 250th anniversary, may be lighting the way for other centers of learning.

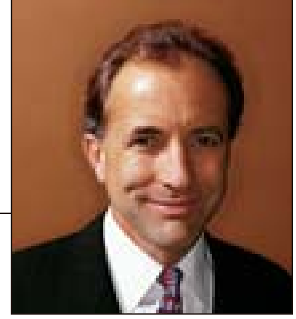
JENNIFER KANE

A parade of biotech heavyweights—among them Amgen, Biogen, Genzyme and, yes, Genentech—filed suits against Columbia last year for allegedly trying to prolong for an additional 17 years what is said to be one of the most lucrative university patent estates ever. Three biotech patents that expired in 2000 brought the academic institution almost $300 million in royalties and licensing fees during their lifetime. But Columbia received another patent in 2002 on what the various plaintiffs claim is essentially the same technology covered by those that had expired: a method for inserting human genes into hamster cells to identify cells that will produce large volumes of proteins from those genes. And Columbia, which maintains that the new patent covers a different invention, has already notified previous licensees of its intention to keep the cash flowing. But the plaintiffs in the various suits want the new patent invalidated.

The patent fight demonstrates that a university is as able as any corporation to do anything in its power to continue milking an intellectual-property cash cow. In devising a strategy to maintain a grip on its blockbuster, Columbia may even be able to teach corporate patent holders a few lessons. It enlisted Columbia alumnus Judd Gregg, now a senator from New Hampshire, to stick a provision in a few bills in 2000 that would extend its patent protection for 15 months. Moreover, even while the school begged legislators for an extension, it was secretly pursuing new patents, a fact never revealed to Congress, according to the complaint filed by Foley Hoag, the Boston-based law firm retained by Biogen, Genzyme and Baxter Healthcare. The patent in dispute "surfaced" in 2002 (another one is still pending) after the unsuccessful lobbying effort was completed.

This classic "submarine" patenting strategy will probably be remembered for years to come. The funding for the research for the original three patents came from the National Institutes of Health. At the time, Columbia had to obtain title to the invention from the NIH. But in doing so, the NIH stipulated that the university "shall include adequate safeguards against unreasonable royalties and repressive practices."

The Columbia imbroglio illustrates that at least for universities, the size of revenues expected from patents does matter. The era of university patenting has led to many fruitful collaborations in which schools license their discoveries to industry. Often university patents receive only modest royalties or fees. But Columbia's patents were different. The almost $100 million they garnered in 1999—a large chunk of the money came toward the end of the patents' term—reportedly constituted nearly 25 percent of the university's research budget. The Columbia patents go to prove that when the stakes are high enough, an institution of "higher" learning can get down and connive with the best of them. SA



None So Blind

Perceptual-blindness experiments challenge the validity of eyewitness testimony and the metaphor of memory as a video recording By MICHAEL SHERMER

Picture yourself watching a one-minute video of two teams of three players each. One team wears white shirts and the other black shirts, and the members move around one another in a small room tossing two basketballs. Your task is to count the number of passes made by the white team—not easy given the weaving movement of the players. Unexpectedly, after 35 seconds a gorilla enters the room, walks directly through the farrago of bodies, thumps his chest and, nine seconds later, exits. Would you see the gorilla?

Most of us believe we would. In fact, 50 percent of subjects in this remarkable experiment by Daniel J. Simons of the University of Illinois and Christopher F. Chabris of Harvard University did not see the gorilla, even when asked if they noticed anything unusual (see their paper "Gorillas in Our Midst" at http://viscog.beckman.uiuc.edu/djs_lab/). The effect is called inattentional blindness. When attending to one task—say, talking on a cell phone while driving—many of us become blind to dynamic events, such as a gorilla in the crosswalk.

SEE anything unusual?

I've incorporated the gorilla video into my lecture on science and skepticism given at universities around the country. I always ask for a show of hands of those who did not see the gorilla during the first viewing. About half of the more than 10,000 students I encountered last year confessed their perceptual blindness. Many were stunned, accusing me of showing two different clips. Simons had the same experience: "We actually rewound the videotape to make sure subjects knew we were showing them the same clip."

These experiments reveal our perceptual vainglory, as well as a fundamental misunderstanding of how the brain works. We think of our eyes as video cameras and our brains as blank tapes to be filled with sensory inputs. Memory, in this model, is simply rewinding the tape and playing it back in the theater of the mind, in which some cortical commander watches the show and reports to a higher homunculus what it saw.

This is not the case. The perceptual system and the brain that analyzes its data are far more complex. As a consequence, much of what passes before our eyes may be invisible to a brain that is focused on something else. "The mistaken belief that important events will automatically draw attention is exactly why these findings are surprising; it is also what gives them some practical implications," Simons told me. "By taking for granted that unexpected events will be seen, people often are not as vigilant as they could be in actively anticipating such events."

Driving is an example. "Many accident reports include claims like, 'I looked right there and never saw them,' " Simons notes. "Motorcyclists and bicyclists are often the victims in such cases. One explanation is that car drivers expect other cars but not bikes, so even if they look right at the bike, they sometimes might not see it." Simons recounts a study by NASA research scientist Richard F. Haines of pilots who were attempting to land a plane in a simulator with the critical flight information superimposed on the windshield. "Under these conditions, some pilots failed to notice that a plane on the ground was blocking their path."

Over the years in this column I have pounded paranormalists pretty hard, so they may rightly point to these studies and accuse me of inattentional blindness when it comes to ESP and other perceptual ephemera. Perhaps my attention to what is known in science blinds me to the unknown.

Maybe. But the power of science lies in open publication, which, with the rise of the Internet, is no longer constrained by the price of paper. I may be perceptually blind, but not all scientists will be, and out of this fact arises the possibility of new percepts and paradigms. There may be none so blind as those who will not see, but in science there are always those whose vision is not so constrained. But first they must convince the skeptics, and we are trained to look for gorillas in our midst. SA

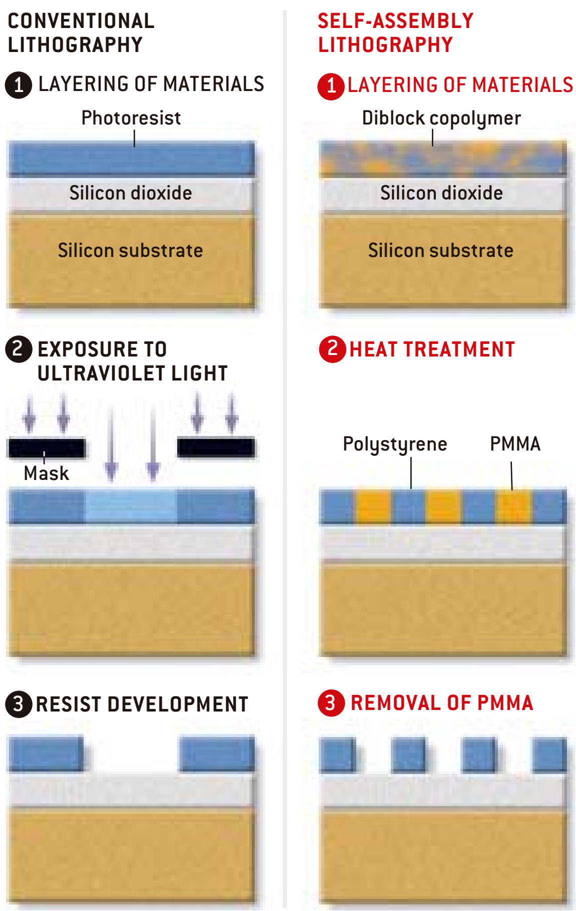
Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* The Science of Good and Evil.

BRAD HINES; DANIEL J. SIMONS (*group shot*)

Nano Patterning

IBM brings closer to reality chips that put themselves together By GARY STIX

Self-assembly has become a critical implement in the toolbox of nanotechnologists. Scientists and engineers who explore the nano realm posit that the same types of forces that construct a snowflake—the natural attractions and repulsions that prompt molecules to form intricate patterns—can build useful structures—say, medical implants or components in electronic chips. So far much of the work related to self-assembling nanostructures has been nothing more than demonstrations in university laboratories. To go beyond being a scientific curiosity, these nanotech materials and techniques will have to get from benchtop to a $2-billion semiconductor fabrication facility.



Four years ago two members of the technical staff at the IBM Thomas J. Watson Research Center in Yorktown Heights, N.Y., began to contemplate how they might transform the vision of self-assembly into a practical reality. The collaborators, Charles Black and Kathryn Guarini, knew that the grand academic ambitions of making an entire set of chip circuits from self-assembly had to be set aside. Instead the best way to begin, they thought, might be to replace a single manufacturing step. "The idea was that if we could ease the burden in any of the hundreds of steps to make a chip, we should take advantage of that," Black says.

They first had to select what type of molecules might self-construct without disrupting routine silicon manufacturing practices. Polymers were an obvious choice. They make up the "resist" used in photolithography—the material that, once exposed to ultraviolet or shorter-wavelength light, is washed away to form a circuit pattern. During the first two years of their quest, the duo spent time learning about polymers and the optimal temperatures and thicknesses at which they would self-assemble. They built on the work of Craig J. Hawker of the IBM Almaden Research Center in San Jose, Calif., and that of former IBMer Thomas P. Russell, a polymer scientist at the University of Massachusetts at Amherst. Both had done research on how polymers self-assemble on silicon. With this knowledge, Black and Guarini even started making things.

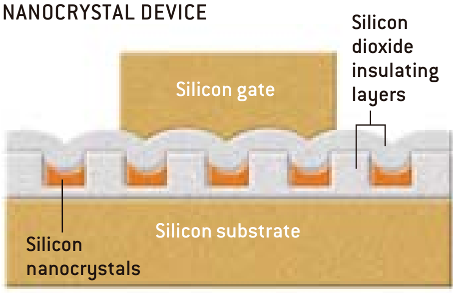
OLD AND NEW: Conventional lithography exposes a photoresist to ultraviolet light. An etchant then removes the exposed part of the photoresist. Self-assembly patterning occurs when a diblock copolymer is heated, thereby separating the two polymers in the material into defined areas before the PMMA is etched away. The template of cyclindrical holes is transferred into the silicon dioxide before the holes are filled with nanocrystalline silicon used to store data (*steps not shown*).

SAMUEL VELASCO

The two researchers appeared at conferences, giving presentations about honeycomb patterns that had self-assembled. But that accomplishment consisted of little more than PowerPoints, the type of through-the-microscope images found in abundance at any academic conference on nanotechnology. What would the nano patterns be good for? How could they be integrated into a fabrication line? Could they best circuit-patterning techniques that had already received hundreds of millions of dollars of investment?

Finally, last year, the pair demonstrated how a self-assembled honeycomb pattern might work in a real manufacturing facility. The material chosen for the demo was a diblock copolymer, one in which two polymers—in this case, polystyrene (Styrofoam) and polymethylmethacrylate (Plexiglas, or PMMA)—are tied together by chemical bonds. When spun onto the surface of a rotating silicon wafer, the two polymers separate, as if they were oil and water. Although the molecules stretch out, the chemical bonds keep them attached. Subsequent heat treatment exacerbates this elongation. In the end, PMMA ends up concentrated in small cylinders surrounded on all sides by the polystyrene. The diblock copolymer thus forms on its own into a nearly complete honeycomblike template.

NANOCRYSTAL DEVICE



FLASH MEMORY: A layer of self-assembled silicon nanocrystals is inserted into an otherwise standard device as part of a novel IBM manufacturing process.

To finish creating the 20-nanometer-wide pores, an organic etching solvent removes the PMMA. A subsequent etching step transfers the same honeycomb pattern into an underlying layer of more robust silicon dioxide. Then a coating of amorphous silicon gets deposited across the surface of the wafer. A gas etches away the silicon except for that deposited in the holes. All that is left are nanocrystalline cylinders surrounded by silicon dioxide. The final steps place an insulating layer and a block of silicon atop the structure, the block forming a "gate" that turns the electronic device off and on. Black and Guarini's honeycomb results in a nanostructure that is part of a working flash-memory device, the kind that retains digital bits even when a camera or a voice recorder is turned off. The nanocrystalline cylinders form capacitors where data are stored.

Manufacturing engineers are leery of introducing new technologies unless a researcher can make a very good case for their adoption. Self-assembly potentially fits the bill. Creating closely spaced holes for a flash memory would prove exceedingly difficult with ordinary lithographic and deposition methods. Forming nanocrystals using conventional techniques creates elements of different sizes that are all jumbled together. In contrast, the self-assembled nanocrystals are evenly spaced and of uniform size, improving their durability and their capacity to retain a charge while allowing the cylinders to shrink to smaller than 20 nanometers.

The IBM demonstration served as proof of principle in the strictest sense of the expression. The company has not made commercial flash memories for years, so the invention could not be applied immediately to improve its own manufacturing operations. But the nanocrystals enabled the pair of researchers to flaunt this type of nano patterning. "Politically in the company maybe it wasn't the smartest demonstration we could have done, but everybody was supportive and could see the power of the technology," Black says.

The understanding gained of how to integrate nanomanufacturing with conventional chipmaking may provide new approaches to fabricating other IBM electronic components. Making holes and filling them could create "decoupling" capacitors recessed into the chip substrate that smooth out fluctuations in the power supplied to a chip.

Using a variant of nano patterning, a self-assembling polymer could also create tiny, finger-shaped silicon protrusions sticking up from the underlying substrate. These fingers would constitute the "channel" in a transistor through which electrons flow—but one in which electrons flow vertically instead of across a chip, as in today's devices. The gate to turn the transistor off and on could encircle the silicon finger. The geometry might prevent electrons from "tunneling," or leaking, through the channel when the transistor is in the off state, a constant threat when feature sizes become very small.

Ultimately, self-assembly might play a much bigger role in fashioning electronic circuits. But the incrementalist approach of Black and Guarini may represent the most promising way to get nanotechnology adopted as a real manufacturing tool. "The greatest excitement is that these materials aren't just in the polymer-science laboratory anymore," Black says. A small step for small manufacturing. SA

SAMUEL VELASCO

A Strategy of Containment

Pathogens take windows of opportunity, and so must humans, says David L. Heymann, who helped to create a global early-warning and response network By CHRISTINE SOARES

Late last spring World Health Organization officials talked about putting severe acute respiratory syndrome, or SARS, "back in the box" before it could become endemic in China and the other countries to which it had spread. The virus infected more than 8,000 people worldwide and killed nearly 800 last year. But so far this season, it had caused just a handful of possible cases by mid-January, with only two confirmed, one the result of a laboratory accident. If SARS has indeed been tamed, without a vaccine or any effective drug treatment, it will be a triumph for the good old-fashioned public health tactics of surveillance and infection control.

YVES LERESCHE *Lookat Photos*

DAVID L. HEYMANN: PATHOGEN PATROL

- **On being called a "roustabout epidemiologist": "That's the beauty of understanding a little bit about epidemiology and many different diseases—you can jump from one to another. You can figure out which principles you can apply and which you can't apply and take a fresh look at a new issue."**
- **SARS lesson learned: The world's health ministers voted unanimously last May to allow the World Health Organization to act on information from all sources, not just official reports; all countries must now report any disease outbreak of "international concern."**
- **At least 34 new pathogens have been identified in the past three decades.**

"Identify cases, isolate, contact tracing, and when contacts get sick, [do it] all over again" is the not so secret formula for containing disease outbreaks, according to David L. Heymann, the veteran pathogen fighter who led WHO's response to SARS last year as executive director of the agency's communicable diseases division. Whether it's SARS, smallpox or polio, the fundamentals of stopping infectious disease are the same, he says: find it and break its chain of transmission. He is not declaring victory against SARS just yet, though. Only another full year of surveillance will tell whether the virus has become endemic, he says, "so we need to have the mechanisms in place to detect this one and to detect any new one that emerges, too."

The 58-year-old American has learned the value of vigilance over 30 years of battling infectious diseases, both new and old, around the world. Fresh out of the London School of Hygiene and Tropical Medicine in 1974, he was recruited, along with hundreds of other idealistic young doctors, by Donald A. Henderson, who was running WHO's global smallpox eradication program. Heymann spent two years in India administering smallpox vaccinations. In 1976, thoroughly hooked on international public health, he returned to the U.S. to join the Centers for Disease Control and Prevention's epidemic intelligence service.

That year "swine flu" provoked fears of a killer influenza pandemic, prompting the CDC to bolster influenza surveillance. When the agency heard about an unusual respiratory infection spreading at an American Legion convention in Philadelphia, Heymann was sent on his first outbreak investigation. Instead of flu, the illness turned out to be a new one, later dubbed Legionnaire's disease. Just a few weeks later Heymann "got lucky" again, he says. It was Christmastime and he was single, so he was sent to Zaire (now Congo-Kinshasa) to investigate a highly lethal hemorrhagic fever ravaging patients and health care workers. That virus would be named Ebola.

Heymann spent the next 13 years in West Africa working for the CDC and crossing paths again with Henderson, who always found the tireless epidemiologist to be "a really positive person, optimistic, *very* intelligent" and a critical thinker who "could examine what's being done and how to improve it."

In 1995, when WHO asked Heymann to create an emerging and infectious disease program, it was clear that the agency "really didn't have a useful tool in outbreak alert and response," Heymann remarks. Having chosen to be loaned to WHO, rather than climb the CDC senior management ladder in Atlanta, Heymann was by then living just outside Geneva, married and a father of three but still jetting off to help contain disease outbreaks. Often WHO's aid arrived late because the agency relied on member nations to voluntarily report domestic outbreaks, with the exception of yellow fever, cholera and plague, for which reporting was mandatory. The problem was, the very developing countries where diseases were most likely to flare up had little systematic surveillance. By the time the central government realized that an outbreak was happening, it could have reached crisis proportions.

INDIA, 1974: David L. Heymann receives a smallpox vaccination to demonstrate its safety.

Once WHO did learn of an outbreak, the agency could only deal directly with national governments to offer advice and, if invited, assistance, albeit with limited resources. But earlier in 1995 Heymann had been in Kikwit, Zaire, during a large Ebola outbreak, and he was struck by the number of "other actors out there waiting to help." The Red Cross, Doctors Without Borders and additional nongovernmental organizations could act as eyes and ears for WHO, he realized, as well as extra hands during emergencies.

So Heymann and his team set out to create what he calls "a network of networks." It would include laboratories and experts around the world pledged to work with WHO when called on and a semiformal array of informants. Also determined to tap into the digital information stream, Heymann's group collaborated with Canada in 1998 to create a Web-crawling program that searches for hints of disease outbreaks. "He's been innovative in a number of ways," says emerging-disease specialist Stephen S. Morse of Columbia University, simply by "connecting up sources of information—in the intelligence community they call it 'all-source information'—and 'stovepiping' existing information, making sure it gets to the right people."

The WHO formally unveiled its Global Alert and Response Network in 2000, but SARS was the first multicountry outbreak the coalition faced [see "Caught Off Guard," by Christine Soares; News Scan, SCIENTIFIC AMERICAN, June 2003]. "We had a vision of a world on alert and able to respond to emerging and other infectious diseases," Heymann says. "This was its international rollout, and it worked." Scientists from 17 countries worked on SARS, he notes, "and when you have real-time information, you can make evidence-based decisions and WHO can play that role."

Heymann, too, has a new role, having been charged with WHO's current attempt to completely eradicate an old disease from the world—this time, polio. He took over the job last July from Jong Wook Lee when the latter became WHO's director general, and in January, Heymann made a bold public promise to stop the transmission of wild poliovirus in all countries by the end of this year. The move was calculated to draw world attention and to put on the spot the leaders of the nations where polio is still endemic. "We have to do it," Heymann says of the self-imposed deadline. "If we don't, we might have to admit that it might not be feasible to do. The only thing that may be lacking now is political will."

In its 16th year, the eradication program has already cost $4.6 billion. Just six countries have wild poliovirus transmission within their borders, but political squabbles have bogged down immunization efforts in some areas. Polio is also much harder to ferret out, notes Henderson, who served as WHO's adviser for polio eradication in the Americas. Unlike smallpox, which produces dramatic symptoms in all victims, polio causes a distinctive "acute flaccid paralysis" in only one of every 200 cases. "You just didn't know where it was until you found that first case," Henderson explains. "I wish him well," he sighs. "If anyone can do it, it's David."

Henderson, Morse and other observers are less confident that international support for WHO's efforts to bolster global disease surveillance will continue now that the program's charismatic leader is gone. "If it doesn't go on without me, it was pretty poorly conceived, and I think there's no question that it will," Heymann declares. Besides, he enjoys starting things more than maintaining them and relishes the chance to reinvigorate the polio program.

The challenge is rejuvenating him in turn, Heymann says, by getting him into the field more often. "Somebody told me once that you have idealism candles that burn, and those candles slowly go out, but you can rekindle them. The fire burns brightly again when you get out and see, really, the need in this world." SA

COURTESY OF DAVID L. HEYMANN

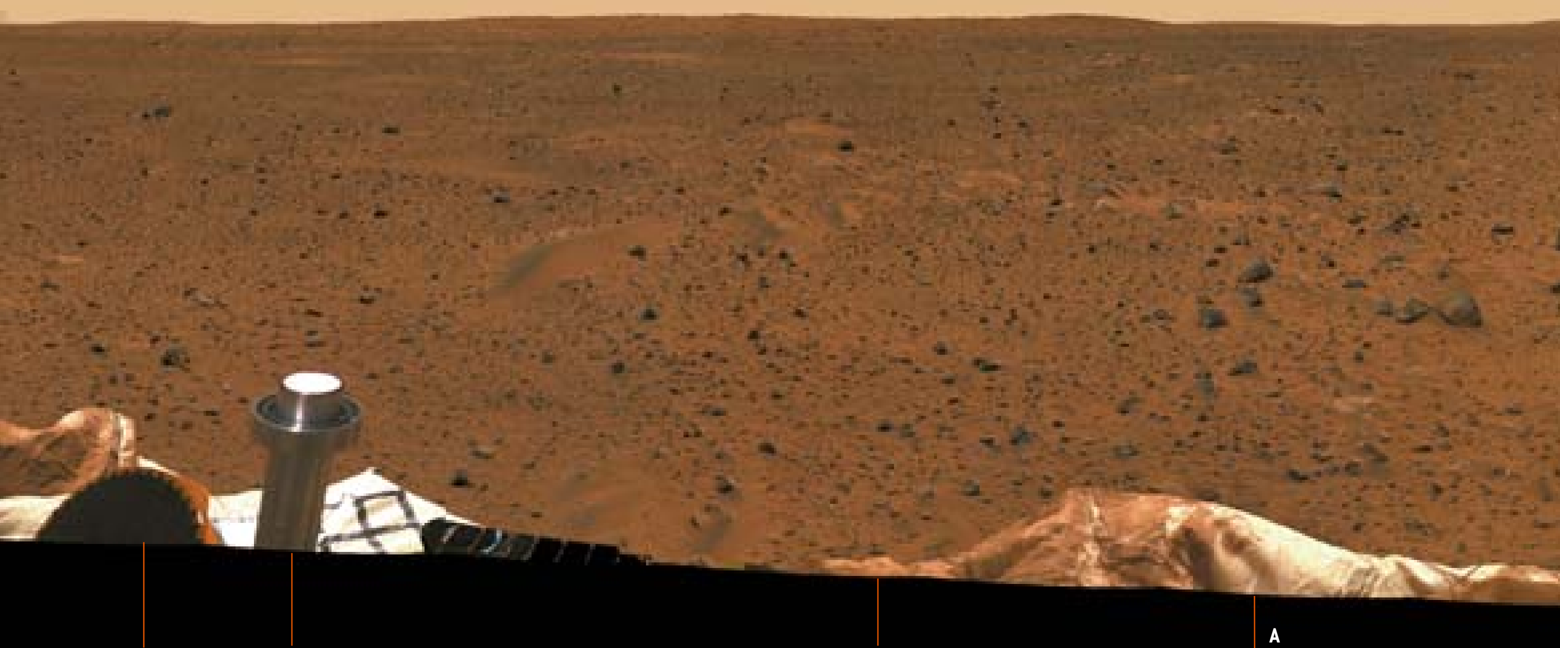
THE SPIRIT OF

At 8:15 P.M. Pacific time on January 3, the Spirit rover, tucked inside its protective capsule, separated from its interplanetary mother ship and prepared to enter the atmosphere of Mars. For weeks, mission engineers and scientists had been listing in grim detail everything that could go wrong. Explosive bolts might not blow on time; strong winds might slam the capsule against the ground; the lander might settle with its nose down, wedged helplessly between rocks; radio links might fail. As the final days ticked by, a dust storm on the planet erupted, reducing the density of the upper atmosphere. To compensate, controllers reprogrammed the parachute to deploy earlier. Eight hours before the capsule's entry, deputy mission manager Mark Adler said, "We're sending a complicated system into an unknown environment at very high speed. I feel calm. I feel ready. I can only conclude it's because I don't have a full grasp of the situation."

This candid doom-mongering was reassuring. If the team had said there was nothing to worry about, it would have been time to start worrying. Between 1960 and 2002 the U.S., Russia and Japan sent 33 missions to the Red Planet. Nine made it. By the standards of planetary exploration, the failure rate is not unusually high: of the first 33 missions to the moon, only 14 succeeded. But the blunders that damned the Mars Climate Orbiter in 1999—neglecting to convert imperial to metric units, then failing to diagnose the error when the spacecraft kept drifting off course—are hard to live down. And just a week before Spirit reached Mars, the British Beagle 2 lander bounded into the Martian atmosphere never to be heard from again.

Controllers at NASA's Jet Propulsion Laboratory (JPL) have a tradition of opening a bag of peanuts for good luck, and the moment had come to do so. At 8:29 P.M., Spirit started its meteoric descent. (To be precise, that is when the confirmation signal reached Earth. By then, Spirit had already landed on Mars; the only question was whether it had landed in one piece or in many.) Within two minutes, the lander had survived the peak atmospheric heating and maximum g-force. After another two minutes, it deployed its chute and emerged from its capsule. Two minutes later its cushion of air bags inflated and controllers announced, "We have signs of bouncing on the surface of Mars."

The control room became a blur of cheering and hugging. It didn't take long, though, for people to wonder whether they had cheered and hugged too soon. The radio signal had flatlined. Rob Manning, the leader of the group that devised the landing sequence, recalls: "The signal disappeared. That



NASA's rover fights the curse of the Angry Red Planet

BY GEORGE MUSSER

EXPLORATION

caused us some pause. I was trying to act calm. It was nerve-wracking." Up until then, he says, the entry had felt just like one of the team's many test runs. "It was only when the signal started going away that I said, 'Uh-oh, this is not a rehearsal.'"

Engineers had warned that Spirit might go silent for 10 minutes or so until it rolled to a stop. A tumbling lander does not make a good transmission platform. But the 10th minute came and went without contact, then the 11th and the 12th. People swiveled in their chairs, crossed their arms, chewed gum. A thin jittery line, representing radio static, ran across the bottom of controllers' computer screens. Manning says he was watching the bottom of his screen so intently that it took him a moment to notice when the line jumped to the top. At 8:52 P.M., or 2:51 P.M. local time at the landing site, Spirit proclaimed its safe arrival on the Red Planet.

Squyres's Odyssey

LIKE SAILORS ROUNDING Cape Horn, scientists and engineers willingly put themselves in the capricious hands of fate for a reason: to put life on our planet into context, either as a singular phenomenon or as an exemplar of a universal process. Steve Squyres, principal investigator of the rover's scientific instruments, has been trying to get to Mars for 17 years. The Cornell University professor has something of a wunderkind reputation. He did his Ph.D. from start to finish in three years and, during the 1980s, became an expert on half the solid bodies of the solar system, from the icy satellites of Jupiter to the volcanic plains of Venus to the water-cut highlands of Mars. But he came to feel that his career was missing something.

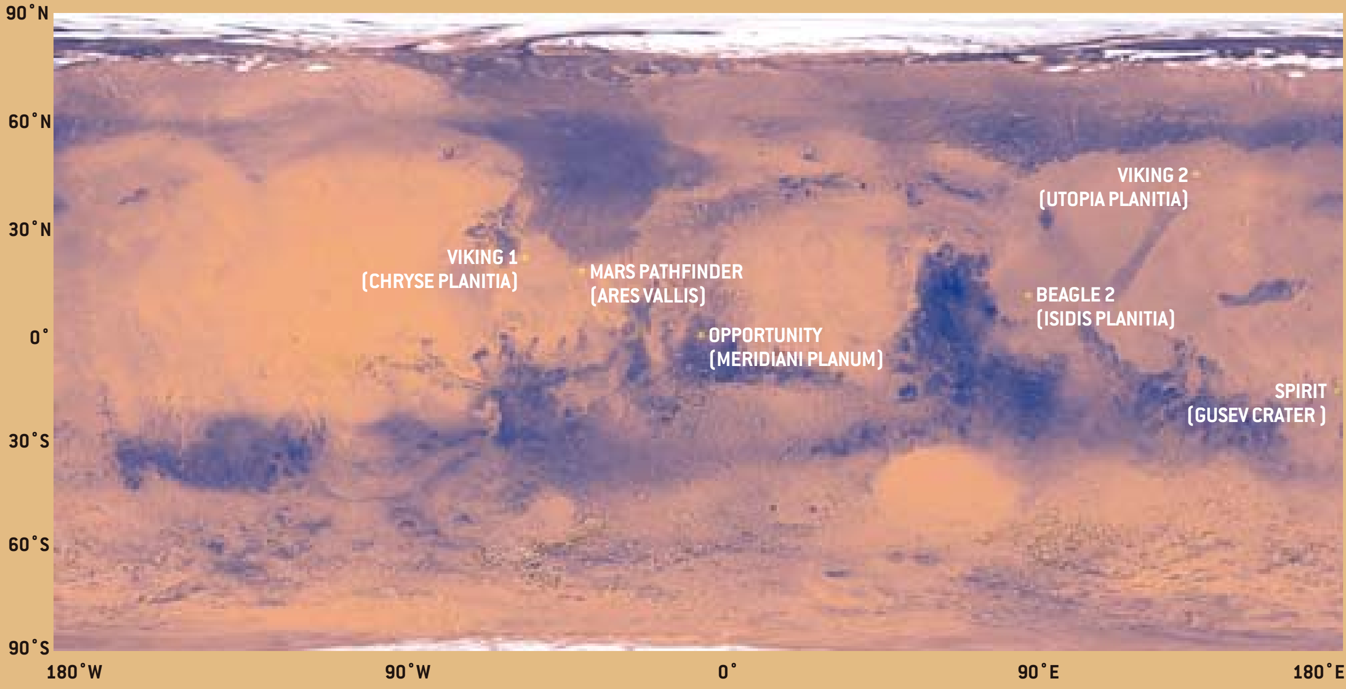
"The real advances in our business come from people who build instruments and put them on spacecraft and send them to the planets," he says. "I worked on Voyager; I worked on Magellan. I didn't think of those missions, I didn't design those instruments, I didn't calibrate them. I just parachuted in at the end, scooped up some data and went off and wrote a bunch of papers. It was a very enjoyable, satisfying way to do a career, in a lot of respects, but I did feel that I was profiting by the efforts of others. For just once—and it *is* going to be just once; this is an experience neither to be missed nor repeated—for just once I wanted to do one where at the end I could say, You

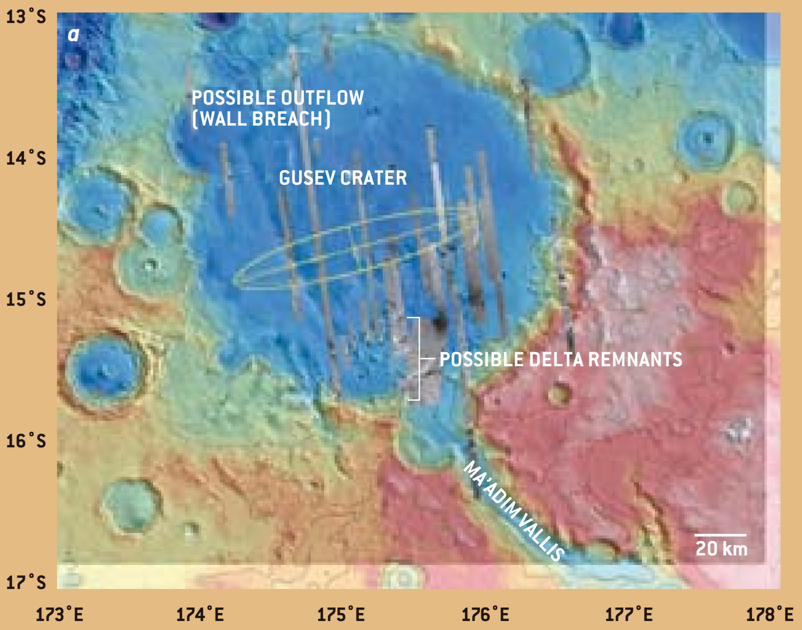
EASTERN PANORAMA from the Spirit landing site runs from due north at the left to due south at the right. The first major goal of the rover is to reach a crater about 250 meters to the northeast. Later it could drive toward the East Hills, which lie three to four kilometers away and are about 100 meters high.

NASA/JPL/CORNELL UNIVERSITY

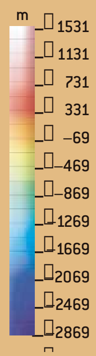
HUMANITY'S NEW BEACHHEAD ON MARS

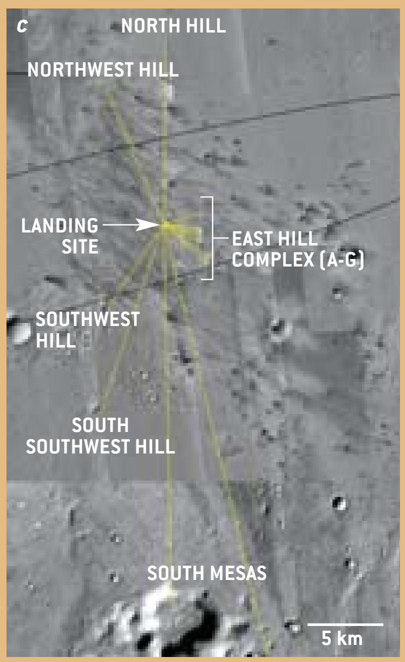
SPIRIT'S LANDING SITE, Gusev Crater, is only the fourth place on Mars that humans have seen in any detail. The crater lies on the boundary between the southern highlands and northern plains. It is one of half a dozen possible lake beds that scientists have identified on the Red Planet. The landing sites of the ill-fated Beagle 2 and of Opportunity, Spirit's twin, may also have been ancient lakes. The earlier Mars Pathfinder rover roamed the mouth of a large outflow channel. The Viking landers set down on featureless plains.

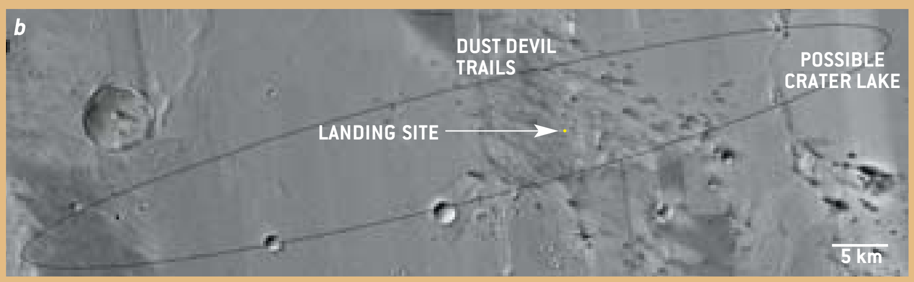




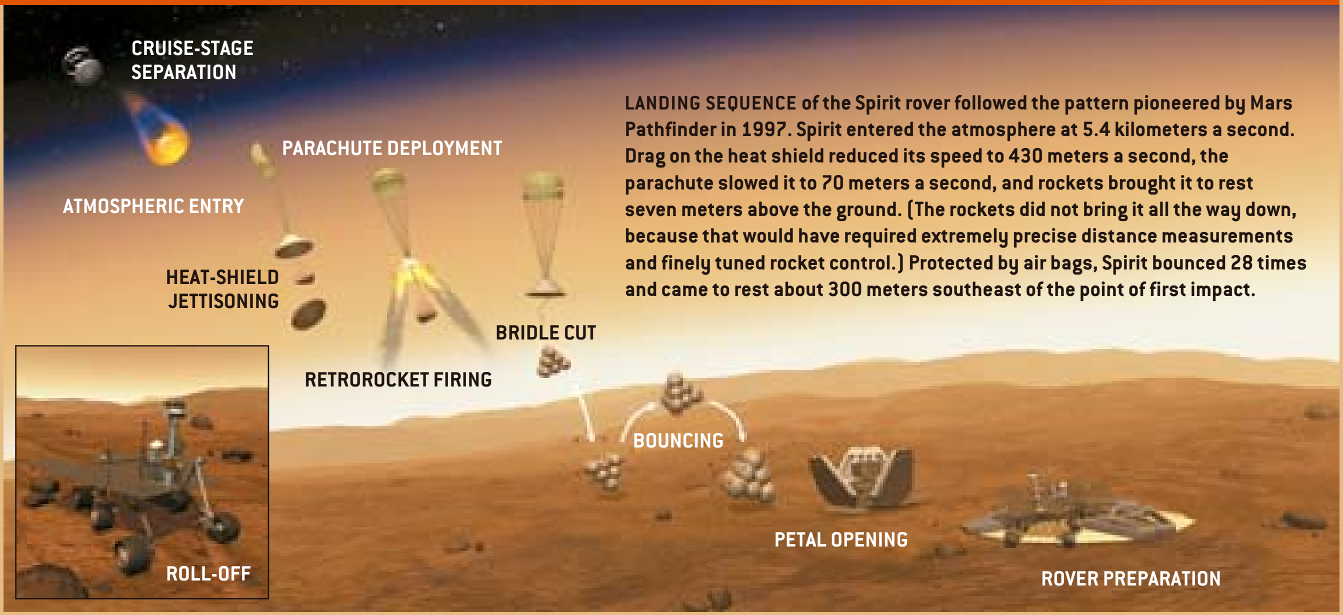
GUSEV CRATER is just north of Ma'adim Vallis, a canyon 900 kilometers long. The regional view (*a*) shows topography (*colors*) and strips of high-resolution images. The high density of craters implies an ancient terrain, perhaps four billion years old. Mosaics of high- and low-resolution images (*b, c*) zoom in on the landing site. The ellipses represent the targeted landing area (which changed slightly over time); the yellow lines are sight lines from the rover's initial position.







NASA/JPL/MALIN SPACE SCIENCE SYSTEMS; NASA PLANETARY DATA SYSTEM (*top view*)



LANDING SEQUENCE of the Spirit rover followed the pattern pioneered by Mars Pathfinder in 1997. Spirit entered the atmosphere at 5.4 kilometers a second. Drag on the heat shield reduced its speed to 430 meters a second, the parachute slowed it to 70 meters a second, and rockets brought it to rest seven meters above the ground. (The rockets did not bring it all the way down, because that would have required extremely precise distance measurements and finely tuned rocket control.) Protected by air bags, Spirit bounced 28 times and came to rest about 300 meters southeast of the point of first impact.

SMOOTH ROCK SURFACES may have been polished by windblown sand grains. This is one of the first color images taken by Spirit.

EAST HILLS E (*left*) and F (*right*) were imaged several hours apart, showing how dust affects the visibility. The atmosphere above Gusev is dustier than predicted; consequently, the rover is warmer but has less solar power. Hill E is 3.1 kilometers away, and F is 4.2 kilometers away.

DRAG MARKS, left by the air bags as they were retracted, indicate a cohesive soil—perhaps electrostatically charged dust or a weakly cemented "duricrust" like that seen by Viking.

THERMAL SCAN shows the area from the East Hill Complex to Sleepy Hollow. Dust is warmer (*red*) because it has a low thermal inertia, which means it heats up quickly in the sun. Rocks, with their higher thermal inertia, stay cooler (*blue*). Other data from the infrared spectrometer reveal magnesium carbonate and hydrated minerals, but no one yet knows what it means for the history of water at Gusev.

ALFRED T. KAMAJIAN; SOURCE: NASA/CORNELL UNIVERSITY (*top*); NASA/JPL/CORNELL UNIVERSITY (*center left* and *right*; *bottom left*); NASA/JPL/ARIZONA STATE UNIVERSITY/CORNELL UNIVERSITY (*bottom right*) know, okay, that was something that I helped make happen."

In 1987 Squyres put together a team, built a camera and proposed it to NASA for what became the Mars Pathfinder mission. It had the wrong dimensions and was disqualified. He also joined one of the instrument teams for the Mars Observer spacecraft. Shortly after it lifted off in September 1992, its booster rocket fired to break out of Earth orbit, and the fragility of spaceflight intruded. The radio signal went dead. Sitting in the auditorium at launch control, Squyres put his head in his hands and said, "I think we may have lost it. I think we may have lost it." Forty minutes later the spacecraft reappeared. It vanished for good when it got to Mars the following year.

In 1993 Squyres and his team proposed another instrument package and were again turned down. As they were developing yet another set of plans, for a full-blown mobile geology lab called Athena, news broke that a meteorite discovered in Antarctica might contain hints of past life on Mars. The hoopla reenergized Mars exploration. The Pathfinder mission in 1997 showed what a rover could do, and in November of that year NASA gave the go-ahead to Athena. Squyres found himself the leader of 170 scientists and 600 engineers.

Two years later NASA lost the Mars Climate Orbiter and the Mars Polar Lander. Although Squyres's team was not directly involved, the fiascoes convulsed the entire Mars program. In response to an investigation panel, which put the blame largely on a caustic mix of underfunding and overconfidence, the agency increased the budget for the rovers; they eventually cost $820 million. Redesigned and refocused, Spirit and its twin, Opportunity, finally blasted off last summer. "To get through something like what we went through, you have to be optimistic by nature," Squyres says. "To be prepared for every eventuality, you also have to be pessimistic by nature."

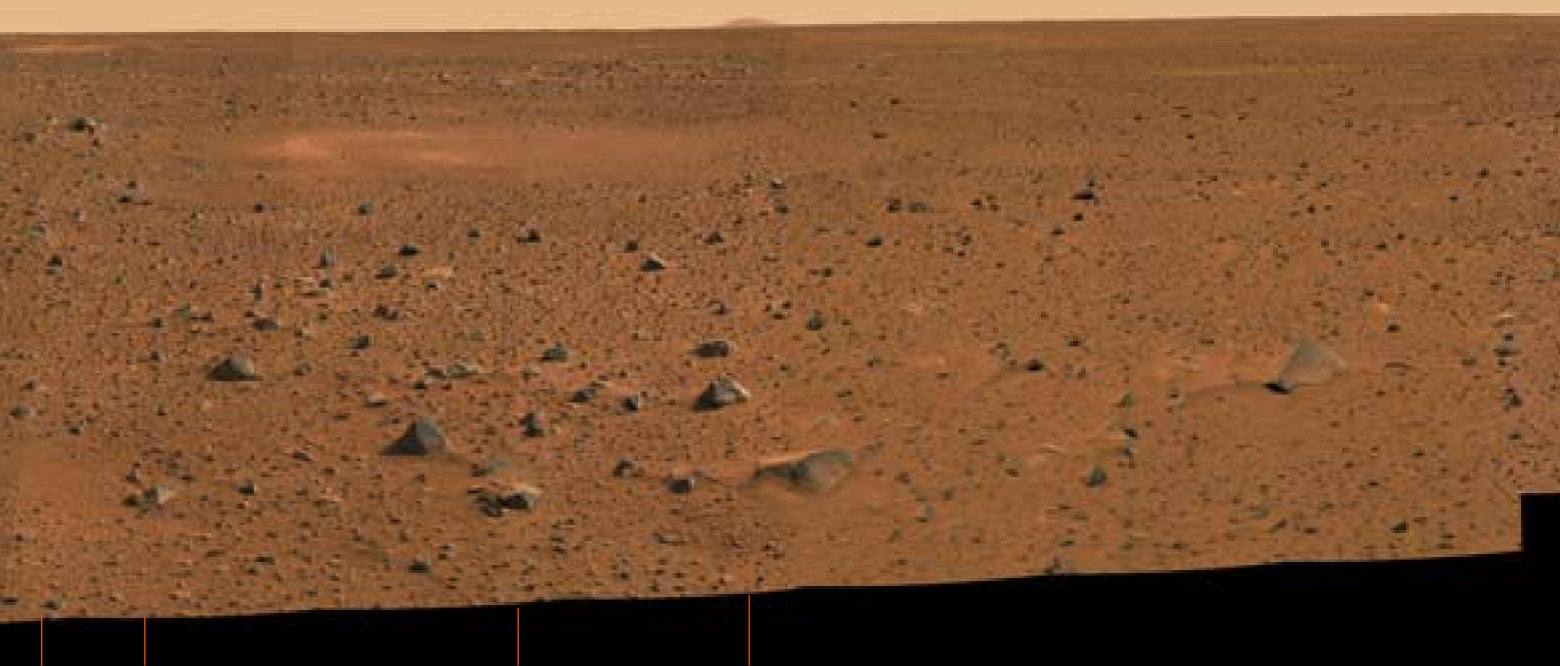
Freeze-Dried Planet

AS THE TWO Mars Exploration Rovers (MERs) were coming together, Martian science went through an upheaval. The Mariner and Viking missions of the 1960s and 1970s revealed a cold, dry and lifeless world, but one etched with remnants of past vigor: delicate valley networks from the distant past and vast flood channels from the intermediate past. Researchers expected that when new space probes assayed the planet, they would find water-related minerals: carbonates, clays, salts.

Over the past six and a half years, the Mars Global Surveyor and Mars Odyssey orbiters—bearing duplicates of the instruments that the ill-fated Mars Observer carried—have looked for and detected essentially none of those minerals. They have found layers of olivine, a mineral that liquid water should have degraded. And yet the orbiters have also seen fresh gullies, old lake beds and shorelines, and an iron oxide mineral, gray hematite (as opposed to red hematite, otherwise known as rust), that typically forms in liquid water. The planet holds extensive reservoirs of ice and bears the marks of recent geologic and glacial activity. Scientists are more baffled than ever.

"There's a fairly raging debate about how the environment of early Mars differed from now," says Matt Golombek, the JPL planetary geologist who led the Pathfinder science team and is a member of the Mars Exploration Rover team. "MER is really the first attempt to go to the surface and try to verify what the environment was really like."

The notoriously risk-averse Viking planners sent their two landers to the most boring places on Mars. (To be fair, you'd



WESTERN PANORAMA runs from due south at the left to due north at the right. The prominent light-colored area is Sleepy Hollow, a shallow depression about nine meters in diameter and located about 12 meters away. Dark marks on the dusty surface of the hollow may be places where the rover bounced before settling down.

probably do the same if you had a $3.5-billion, easily toppled spacecraft and knew almost nothing about the terrain.) Pathfinder, though bolder, was really just a test flight. Beyond a desire to study as many different rocks as possible, Golombek's team didn't much care where it went. Spirit and Opportunity are the first landers to visit places that scientists actively wanted to go.

From orbit, Spirit's new home, Gusev Crater, looks like a lake bed. It has fine layering, deltalike deposits and sinuous terracing, and it sits at the northern end of Ma'adim Vallis, one of the largest valleys on the planet. Opportunity has gone for the gray hematite, which is concentrated in Meridiani Planum. Phil Christensen, a planetary geologist at Arizona State University, recently studied the topography of the hematite outcrops and concluded that the mineral forms a thin, flat layer—as though Meridiani, like Gusev, was once a lake bed.

Only on the surface can these hypotheses be tested. For instance, because wind cannot transport sand grains larger than half a centimeter, the discovery of bigger grains would imply another agent of erosion, probably water. When hematite crystallizes in lake water (as opposed to, say, a hot spring), the chemical reaction often involves the mineral goethite, which spectrometers on the rovers can look for. Piece by piece, datum by datum, the rovers should help resolve how Mars can be both so Earth-like and so alien.

Mars under the Earthlings

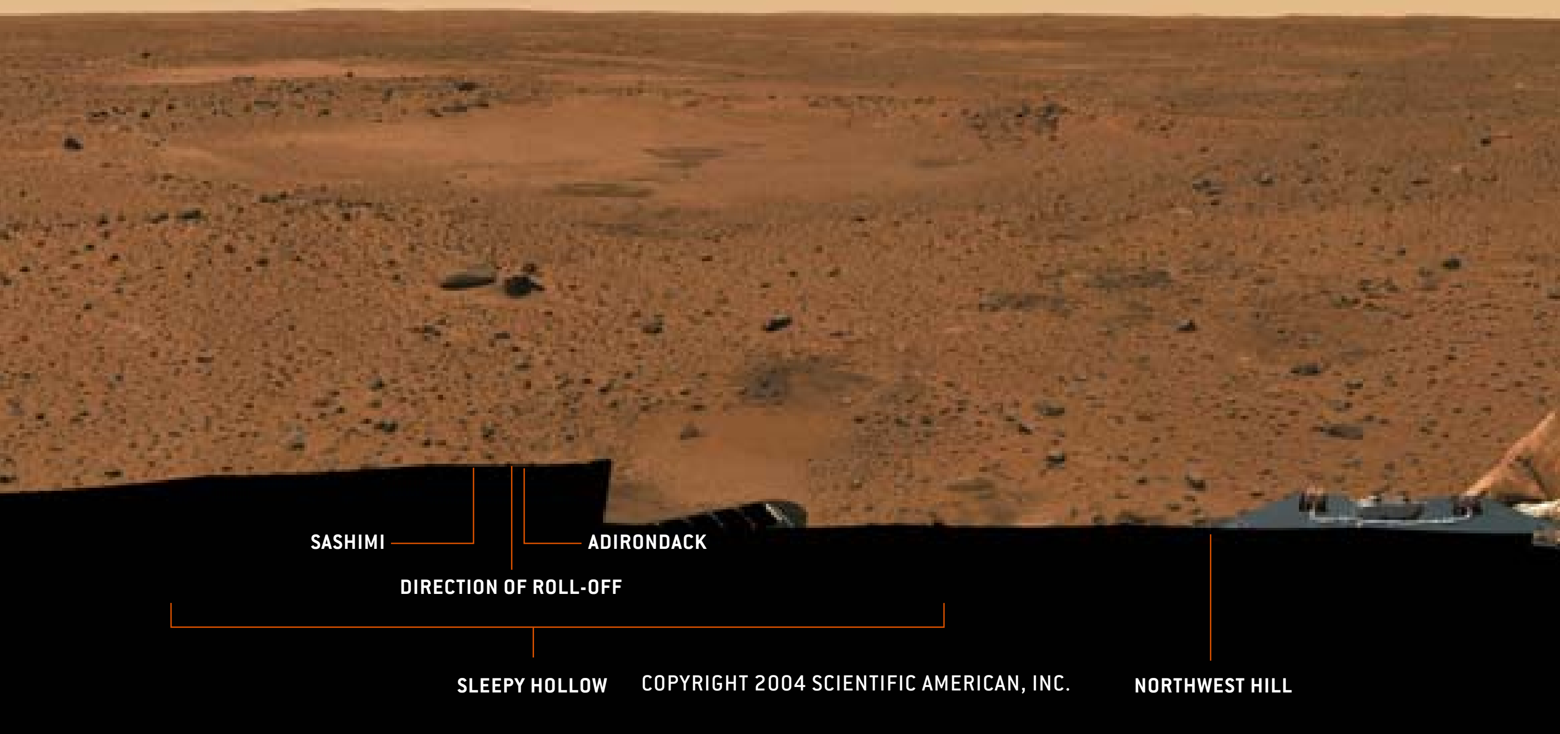
ABOUT THREE HOURS after Spirit landed, at 11:30 P.M. Pacific time on January 3, the data started to pour in, relayed by the Odyssey orbiter. For observers used to earlier missions, when images slowly built up line by line like a curtain rising on another world, it was startling. The first pictures flashed up on the screen, and Gusev Crater leapt into the control room.

The main cameras sit on a mast 1.5 meters tall, so the view closely matches what you'd see if you stood on the planet. But it still takes some getting used to. Jim Bell, a Cornell scientist who has worked on the color panoramic camera, Pancam, since 1994, says: "One thing that I learned through all the testing we did is when you experience a place through the eyes of a rover, and then go yourself, it's pretty different. The sense of depth is very different, because you're looking at this flat projection of the world, and there's nothing in it for human reference. There's no trees, no fire hydrants—you're missing all the cues we have all around us that tell us how far away things are."

Even so, the first images have an eerily familiar quality, showing rocks, hollows, hills and mesas. "It's beautiful in the same way the desert is beautiful," aerospace engineer Julie Townsend says. "It's a beautiful vacantness, the beauty of an undisturbed landscape."

But space exploration is like plucking the petals of a daisy: it works, it works not, it works, it works not. You never know how it will end. Early morning Pacific time on January 21, controllers were preparing Spirit to analyze its first rock, named Adirondack. They instructed the rover to test part of the infrared spectrometer, and Spirit sent the robotic equivalent of "roger." But then it went silent. For two days, controllers tried nearly a dozen times to reach it. When they finally reestablished contact, the situation was serious. Though in no imminent danger, Spirit had rebooted itself more than 60 times trying to shake off a fault it could not diagnose. Pete Theisinger, the project manager, says, "The chances it will be perfect again are not good." But he adds, "The chances that it will not work at all are also low." And that, in the business of planetary science, is a victory. SA

George Musser, a staff writer, was a graduate student of Steve Squyres' in the early 1990s. For updates on the Spirit and Opportunity missions, see www.sciam.com



NASA/JPL/CORNELL UNIVERSITY

A New Race of ROBOTS

SCOTT GOLDSMITH

DECEMBER 2, 2003: The Red Team prepares its robotic vehicle, Sandstorm, for its maiden voyage. As Nick Miller, one of several dozen Carnegie Mellon University undergraduates on the team, drives the robot around a test loop between abandoned steel mills in Pittsburgh, onboard computers (*in metal box*) record the test path. Five days later the robot drives the loop with no one at the wheel.

Around the U.S., engineers are finishing one-year crash projects to create robots able to dash 200 miles through the Mojave Desert in a day, unaided by humans. *Scientific American* tailed the odds-on favorite team for 10 months and found that major innovations in robotics are not enough to win such a contest. Obsession is also required

BY W. WAYT GIBBS

PITTSBURGH, DECEMBER 10, 2003: A cold rain blows sideways through the night into the face of Chris Urmson as he frets over Sandstorm, the robotic vehicle idling next to him on an overgrown lot between two empty steel mills. Urmson checks a tarp protecting the metal cage full of computers and custom electronics that serves as the sensate head of the chimeric robot, which has the body of an old Marine Corps Humvee. His ungloved hands shivering and his body aching from three sleep-deprived days and nights of work in the field, Urmson stares glumly at the machine and weighs his options. None of them are good.

He and his teammates had vowed months ago that by midnight tonight Sandstorm would complete a 150-mile journey on its own. It seemed a reasonable goal at the time: after all, 150 miles on relatively smooth, level ground would be but a baby step toward the 200-mile, high-speed desert crossing that the robot must be ready for on March 13, 2004, if it is to win the U.S. Department of Defense's Grand Challenge race, as well as the $1-million prize and the prestige that accompanies an extraordinary leap in mobile robotics.

But after 20 hours of nonstop debugging, Sandstorm's navigational system is still failing in mystifying ways. Two days ago the machine was driving itself for miles at a time. Last night it crashed through a fence, and today it halts after just a few laps around the test path. The dozen or so team members here are wet, cold and frazzled, hunched over laptops in a makeshift lean-to or hunkered down in a van. The 28-year-old Urmson has hardly seen his wife and two-month-old baby for weeks. Continuing under these wretched conditions seems pointless.

On the other hand, an hour ago he and the rest of the group huddled around William "Red" Whittaker, the leader of the Red Team—and Urmson's Ph.D. adviser at Carnegie Mellon University (CMU)—and acceded to his decision that they would continue fixing and testing through the night and into the day and through the night again, if need be, until Sandstorm completed the 150-mile traverse they had promised. For the umpteenth time, Red repeated the team's motto: "We say what we'll do, and we do what we say." Their reputations, their morale—and for the students, their final-exam grades—are on the line.

But at the moment, Whittaker is not around, so Urmson, as the team's technical director, is in charge. He looks at the rivulets streaming over the tarp, considers how many weeks of work could be undone by one leak shorting the circuits inside, and aborts the test, sending everyone home to their beds.

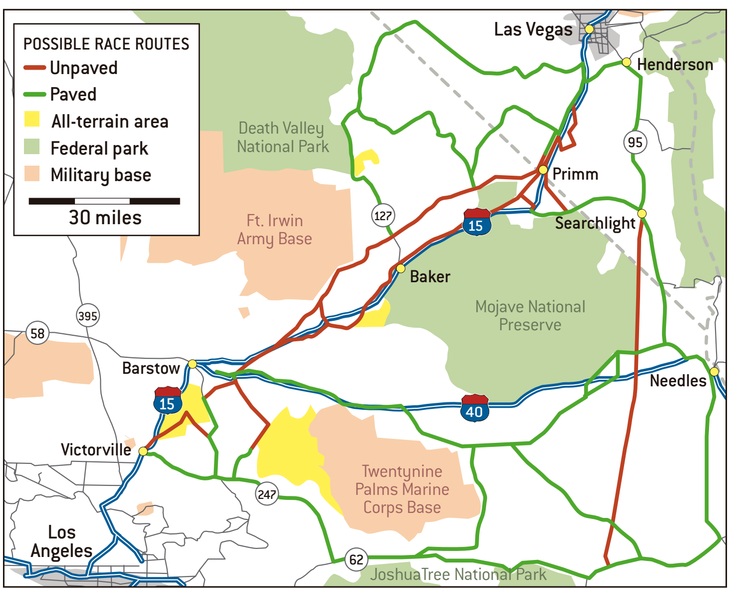
The Grand Challenge Race

DARPA ANNOUNCED in February 2003 that it was organizing a desert race for self-navigating robotic vehicles to be held on March 13, 2004. The race was named the Grand Challenge because its requirements—cross 200 miles of unfamiliar, rough terrain in 10 hours or less, without any human assistance—fell well beyond the capabilities of any robot yet designed.

THE PRIZE: $1 million to the team whose vehicle completes the course in the shortest time less than 10 hours.

THE RULES: The robotic racers must be fully autonomous; during the race they cannot receive signals of any kind (except a stop command) from humans. The vehicles must stay on the ground and within the boundaries of the course. No robot may intentionally interfere with another. The race will begin with a staggered start; a qualifying event will determine who goes first. If no vehicle wins in 2004, the race will be repeated each year until there is a winner or the funding runs out (after 2007).

THE COURSE: Two hours before the race begins, DARPA officials will give each team a CD-ROM containing a series of GPS coordinates, called waypoints, spaced 150 to 1,000 feet apart. The width of the route between waypoints will also vary: in some sections of the course, racers will have to remain within a 10-foot-wide corridor, whereas in other sections they will be able to roam more freely. Depending on how officials mix and match from various potential routes through the Mojave Desert (*map*), the course may be as short as 150 miles or as long as 210 miles.



RACE OFFICIALS have warned participants to expect sandy trails, narrow underpasses, power line towers and hairpin turns. The Red Team is creating a test course in Pittsburgh that includes all of these hazards.

LUCY READING; SOURCES: MDEP, BLM, SRA/DARPA (*map*); COURTESY OF THE RED TEAM, CARNEGIE MELLON UNIVERSITY (*top left and bottom left*); ALEX GUTIERREZ (*top right*); JASON DIVENERE (*bottom right*)

The next day brings hell to pay. Like an angry coach at halftime, Whittaker castigates the team for giving up and for missing other self-imposed goals. "A great deal of what we agreed to do got lost as the team focused monotonically on the 150-mile objective," he rebukes. "The vehicle body didn't get painted; the Web site didn't get updated; the sensor electronics weren't completed. And do we win the race if we don't have better shock isolation than we have now?" Heads shake. "No, we'll lose the race. Is the condition of this shop consistent with who we are?" he asks, waving at the tools and parts scattered over every flat surface. Eyes avert. He clenches his jaw.

"Yesterday we lost that sense deep inside of what we're all about," Whittaker continues. "What we have just been through was a dress rehearsal of race day. This is exactly what the 13th of March will be like. We're in basic training; this is all about cranking it up a notch. Come March, *we* will be the machine, an impeccable machine."

Whittaker concludes his pep talk and asks for a show of hands of all those willing to devote every minute of the next four days to another grueling attempt to complete a 15-hour, 150-mile autonomous traverse. Fourteen hands shoot up. Sometime between the first team meeting eight months ago and today, each person in the room had passed his own point of no return.

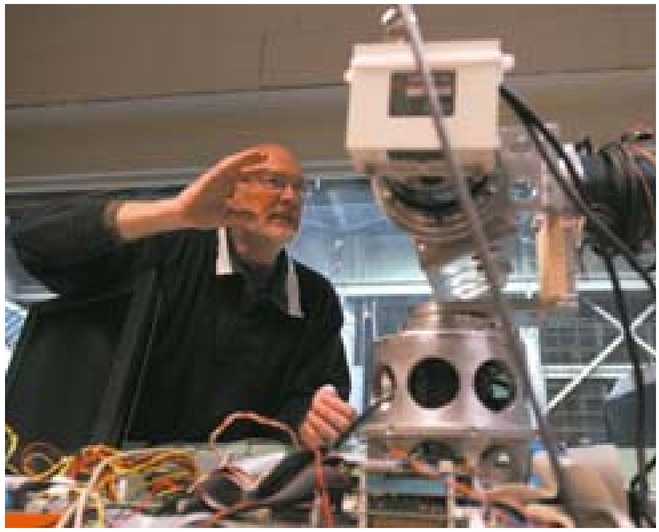
A Grand Challenge Indeed

APRIL 30, 2003: In a conference room at CMU's Robotics Institute, a tall man rises to his feet. He wears the blue blazer and tan chinos of an academic but has the bravado of a heavyweight who used to box for the marines. "Welcome to the first meeting of the Red Team," he booms. "I'm Red Whittaker, director of the Fields Robotics Center, and I am committed to leading this team to victory in Las Vegas next year."

Whittaker attended the conference last February at which officials from the Defense Advanced Research Projects Agency (DARPA) announced their first-ever prize contest, a robot race from Barstow, Calif., to Las Vegas [*see box on opposite page*]. DARPA set up the competition to spur progress toward a vehicle that could enter a battlefield with minimal human supervision. "It could be delivering supplies or taking out wounded. It could also be a tank," says Anthony J. Tether, the agency's director.

A different vision moved Whittaker to be among the first of more than 100 teams that would sign up to enter the race. To him, the principal attractions are the public attention it will bring to robotics and the difficulty of the task, which he often compares to Lindbergh's first transatlantic flight. "The race defies prevailing technology, and many hold that the challenge prize is unwinnable in our time," he wrote in an e-mail on March 13 to potential volunteers and sponsors.

Building an autonomous robot would not be the hard part. With colleagues at the Robotics Institute, Whittaker has created self-driving vehicles that haul boulders, harvest crops, map underground mines, and hunt for meteorites in Antarctica. What makes the Grand Challenge aptly named is its speed—the speed at which the robot must move over rough, unfamiliar terrain and the haste with which it must be built.



NOVEMBER 29: Team leader Red Whittaker helps to tackle major problems with a gimbal meant to give the robot a steady gaze despite bounces and bumps.

DECEMBER 2: Sandstorm takes its first independent steps, driving four miles in 30 minutes. It reaches a leisurely top speed of 15 miles an hour.

DECEMBER 8: After navigating well for four hours and 46 miles, Sandstorm veers off course and into a fence. The next night it rams through the fence.

SCOTT GOLDSMITH (*top* and *middle*); ALEX GUTIERREZ (*bottom*)

"In order to win, Sandstorm will have to average better than 10 meters per second [22 miles per hour]," CMU engineer Scott Thayer points out. That is roughly 10 times the speed of the prototype robots that DARPA has acquired through a four-year, $22-million program to develop unmanned ground vehicles.

"Just getting it to move that fast will be a profoundly challenging problem," Thayer says. "Maintaining those speeds safely for almost 10 hours straight is just mind-boggling." He ventures that "it will take a fundamental innovation to win. And the professional roboticists like me may be the last to come up with a breakthrough like that. After doing this for decades, we tend to think more incrementally. So who knows—one person with a dune buggy may win it."

Blueprint for the Red Team

JUNE 24: "The last time we met, we considered a tricycle with giant wheels seven feet in diameter," Whittaker reports at the team's third meeting. "We also looked at a four-wheel-drive, four-wheel-steered vehicle with a chassis that can change shape. We gave these hard technical looks, but each is too bold a technical step for a yearlong program."

Three months into that year, the team has not yet decided whether to base its robot on a tortoise, such as a military Humvee, or on a hare, such as a professional pickup truck or a low-slung Chenowth combat buggy. Whittaker presents a mathematical analysis of how each vehicle would perform on a course composed mainly of dirt roads and rough trails. "A tough consistent vehicle could go 250 miles in 9.3 hours; a sprinter would take 10.6 hours," he concludes. The choice seems clear, yet it will be September before they will raise the door on the Planetary Robotics Building, where the team has set up shop, and push in a 1986 Hummer M998.

But the group—which now numbers more than 50, thanks to the dozens of CMU graduate and undergraduate students working on the project for credit—has prepared a 58-page technical paper describing how Sandstorm will track its position, plan its route, and detect and avoid hazards in its way. Alex Gutierrez, one of the graduate students at the core of the team,

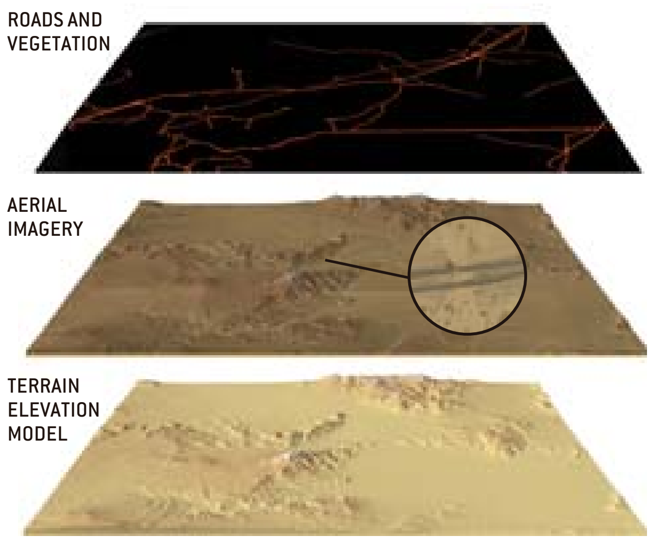
Planning to Win

THE RED TEAM concluded early on that the most feasible way to win the race is to give the Sandstorm robot an extremely detailed and accurate navigational plan to guide it over the race route. The exact course will be held secret until two hours before the starting gun, however. So the team has spent thousands of hours assembling maps, models and aerial imagery of the entire potential race area, which spans 400 times the area shown in this illustration. The engineers overlay, align and hand-correct several distinct views of the terrain.

From the U.S. Geological Survey, the team obtained relatively rough three-dimensional profiles of the land and aerial photography that can distinguish objects as small as one meter. To these they add custom-made road and vegetation maps, then fuse these layers of information into an enormous geographic database several terabytes in size.

A computer program can use this database to calculate the "cost" for Sandstorm to traverse every square meter in the region. Some areas, such as cliffs or course boundaries, have an infinite cost because they would disable or disqualify the racer. Dry lake beds, in contrast, might have a cost of zero.

On race day, the actual course data (*simulated below as circles and blue lines*) will be sent through a high-speed link to the Red Team's control center. There a fleet of computers will use the cost map to compute the optimal route. A dozen or more trained volunteers will then divide the route into sections and will tweak the computed plan as needed so that it does not mistakenly send Sandstorm into harm's way. The final navigation instructions (*yellow dots*) will be beamed to the robot shortly before the race begins.



W. WAYT GIBBS; SOURCES: USGS, THE RED TEAM hands out copies to executives from SAIC, Boeing, Caterpillar, Seagate and other corporate partners as they enter the room.

"First we will work for eight months to create the best possible maps of the course terrain," Whittaker explains. "When DARPA hands out the race route, two hours before the race starts, we will use those maps to calculate the optimal route and do a simulated flight though it" [*see box on opposite page*]. The resulting list of thousands of GPS coordinates will be copied to computers on the robot, giving it "little seeds of corn to aim for every meter or so," Whittaker says. "Sandstorm will just go along like Pac-Man, gobbling up these little virtual dots."

The budget now sums to an astonishing bottom line: $3,539,491. Nearly $2.5 million of that is for personnel expenses that will probably never get paid. The $725,000 for the vehicle itself is not optional, however, and so far only Caterpillar and a local foundation have written checks. But many others are donating valuable equipment and expertise.

Applanix, for example, delivered a $60,000 position-tracking system that not only will allow Sandstorm to know where it is as it bounces along the desert but also will help it to solve one of the toughest problems in mobile robotics: watching where it is going with a steady gaze. "It will know what the world outside looks like through lasers, what it looks like in radar, and what it looks like through a stereo, or two-eyed, camera—provided by our good friends at SAIC," Whittaker declares. Each of these sensors will be mounted on motorized platforms connected to the Applanix system in a tight feedback loop. These gimbals, as engineers call them, will compensate for the motion of the vehicle much like the neck and eye muscles of a human driver [*see box on next two pages*].

Many of the competing teams have similar plans. One composed of undergraduates at the California Institute of Technology is forgoing radar and relying heavily on four video cameras mounted to the front of their modified Chevrolet Tahoe. The Red Team's Navtech radar is worth its $47,000 price because "it works through dust, which can blind the other sensors," Whittaker says. For that very reason, Ohio State University's Team Terramax is mounting two radars—plus six cameras and four laser scanners—on the robot it is building from a huge six-wheeled Oshkosh truck.

More sensors are not necessarily better. Each one streams data like a fire hose; too many can choke a robot's computers. As the vehicle jolts and shakes, overlapping scans may confuse more than they inform. And merging sensor data of different types is notoriously tricky. Laser scanners produce "point clouds," radars emit rectangular blips, a stereo camera generates a so-called disparity map. "If you aren't careful," says Jay Gowdy, a CMU scientist on the Red Team, "you can end up combining the weaknesses of each sensor instead of combining their strengths."

Reality Checks In

NOVEMBER 6: Whittaker, Urmson and Philip Koon, one of two engineers that Boeing Phantom Works has embedded with the team, sit down for the weekly teleconference with the team's

DECEMBER 9: As a cold rain begins to fall outside the workshop, team members scramble to prepare the vehicle for another long night of testing.

DECEMBER 10: Sleep-deprived and frustrated, Chris Urmson and Kevin Peterson struggle to debug the robot's hardware and software.

DECEMBER 18: Engineers from Boeing Phantom Works join part of the team in the Mojave Desert to test an innovative radar system for Sandstorm.

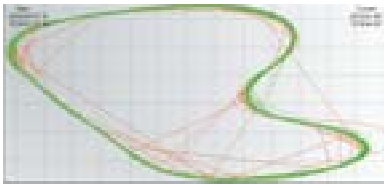
SCOTT GOLDSMITH (*top*); TAKEHITO ETANI (*middle*); JASON DIVENERE (*bottom*)

How Sandstorm Works

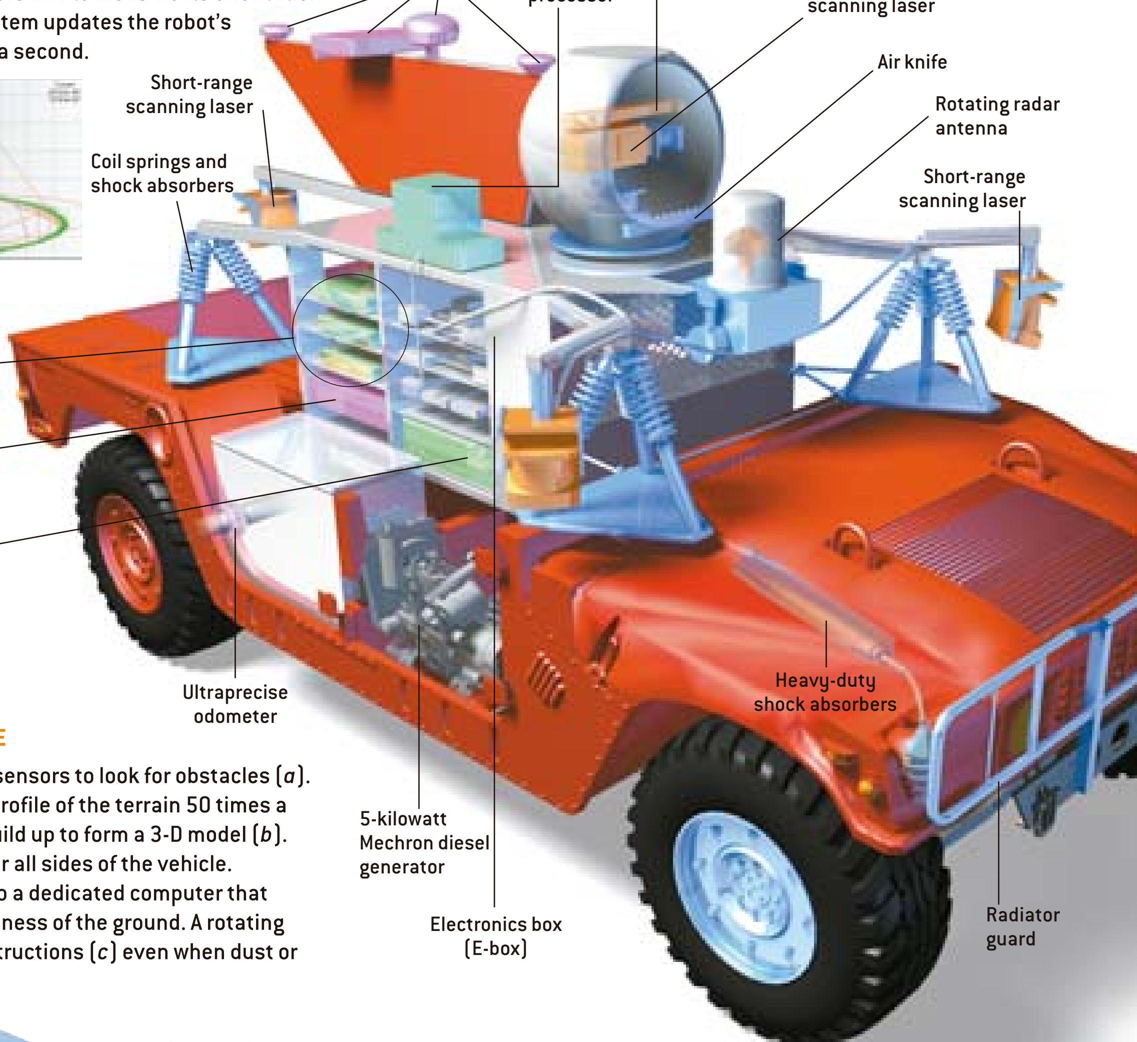
JUST BEFORE THE RACE BEGINS, the Red Team will calculate the best route and send a detailed itinerary (in the form of geographic coordinates for every meter of the course) to the Sandstorm robot. The vehicle will try to follow this virtual trail of breadcrumbs from the starting line to the finish as closely as it can, while detecting and avoiding any unexpected obstacles, such as a disabled racer in the road ahead. To succeed, the robot must solve four challenging problems.

1. TRACKING ITS POSITION

An Applanix navigation computer contains two GPS receivers, three fiber-optic gyroscopes, three silicon accelerometers and an ultraprecise odometer, which it uses to pin down the robot's position to within 50 centimeters and to measure its orientation in space to 0.4 degree. The system updates the robot's sense of where it is 200 times a second.

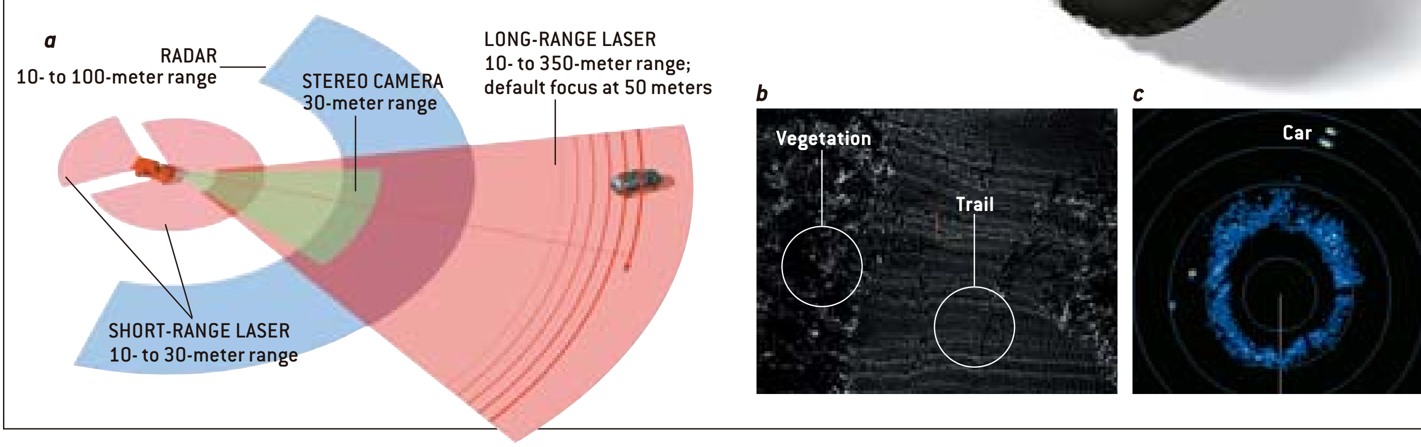


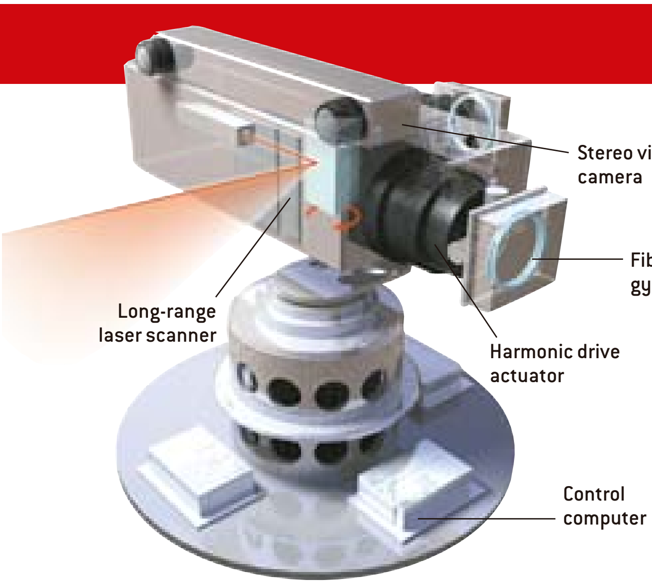
GPS trace of 119-mile test (red lines are sensor glitches)



2. PERCEIVING AN OBSTACLE

Sandstorm uses four kinds of sensors to look for obstacles (*a*). A long-range laser traces the profile of the terrain 50 times a second. Successive profiles build up to form a 3-D model (*b*). Shorter-range lasers also cover all sides of the vehicle. A stereo camera sends video to a dedicated computer that estimates the slope and roughness of the ground. A rotating radar antenna will pick up obstructions (*c*) even when dust or glare blinds the other sensors.



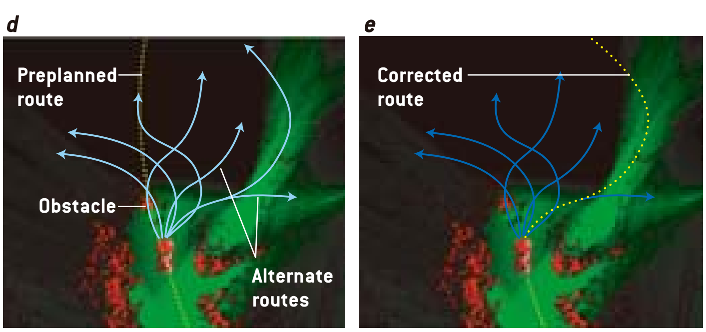


4. ENDURING THE DUST AND BUMPS

Back roads through the Mojave are rough, so the team has equipped the Humvee with racing shocks and springs, a radiator guard and run-flat wheels. To protect the computers, the electronics box is suspended on tripods of spring-reinforced shock absorbers and strapped in place by superstrong bungee cords. A dozen "ruggedized" hard disks inside will operate in redundant pairs. As Sandstorm bounces over a washboard dirt road at 30 miles an hour, it must hold its forward sensors steady. Red Team engineers built a computer-controlled stabilizer, or gimbal (*above*), that both aims and steadies the camera and long-range laser. The gimbal uses three fiber-optic gyroscopes and three precise actuators to measure and compensate for the vehicle's pitch, roll and yaw. The radar is similarly bolted to a one-axis gimbal.

3. REVISING ITS ROUTE

Even the best maps are not up to the minute. So three onboard Xeon computers will use data from each sensor to update the "cost" assigned to each square meter in the area. A paved road carries a cost of zero; a cliff or competing racer warrants an infinite cost. Several times a second, a fourth Itanium2 computer checks whether the "breadcrumb trail" (*d, yellow dots*) passes through high-cost territory (*red areas*). If so, the planner program prices alternative routes (*blue arcs*) and shifts the breadcrumbs to the shortest safe path (*e*).



BRYAN CHRISTIE DESIGN (*illustrations*); SOURCE: THE RED TEAM

partners. "We were maybe 50–50 on our goals this week—this is the first time we have really missed the mark," Whittaker announces. The radar was hung up in customs en route from the U.K. After more than 100 hours of work, the mapping group has completed less than 4 percent of the area they aim to cover. And money is getting tight. "At the moment, we're short about $950,000 and burning through eight grand a day," Whittaker reports. He hopes to sell advertising space on the robot's hood and fin for half a million dollars but has found no buyers.

Two weeks later the team meets to confront other problems. A superprecise optical odometer built to slide on the robot's axle does not fit together properly. "And this is troubling," Whittaker says as he points to a large spike on a graph of how the computer cage—they call it the E-box—bounced around as the vehicle ran over a railroad tie at five miles an hour. "That reads seven g's, which is very bad," he continues. Hard disks will crash and chips may pop from their sockets unless the E-box is isolated from all shocks greater than about three g's. They must figure out a better way to suspend the E-box within the chassis.

"Engineering is always a series of failures to get to success," points out Bryon Smith, one of the few seasoned roboticists on the team. "It takes iteration after iteration to get it right." But iterations take time. The 100 days that Whittaker scheduled for development are almost up, and the team has yet to install and wire all the onboard computers, construct the gimbals, finish the software or mount the sensors.

"This vehicle hasn't rolled so much as a foot under its own control," Whittaker says. "You have promised to get 150 miles on that beast in two weeks. Just so we're clear on the ambition here: DARPA's Spinner vehicle program, based right here at CMU, has a team of pros and a budget of $5 million and is now in its second year. So far the furthest it has driven is 15 miles. Okay, anyone who thinks it is not appropriate for us to go for 150 miles by December 10, raise your hand." No one does. "There it is," he smiles. "We're now heading into that violent and wretched time of birthing this machine and launching it on its maiden voyage."

DECEMBER 1: "There were a bunch of us here all day on Thanksgiving and through the weekend—me, Alex, Philip, Yu [Kato] and several others. But it was worth it," Urmson says. So ends any semblance of normal life as these young engineers are drawn into their leader's constructive obsession. "Around 3 or 4 A.M. Sunday morning, as all the pieces started coming together and getting connected, it felt damn good," Whittaker adds, casting critical looks at those who spent the holiday with their families.

The robot now has several of its sensory organs attached and a rudimentary nervous system working. Smith and Kato have assembled the three-axis gimbal that will aim and steady the stereo camera and long-range laser only to discover "very strange behavior with the fiber-optic gyroscopes" that measure the device's motion, Smith reports. Whittaker listens intently to the details. "The gimbal is an essential device to win the race," he reminds the team. "Its main purpose is to suppress jitter. Right now when we turn it on, it induces jitter." For the next week, Kato will hardly leave the shop as he valiantly attempts to correct through software a fundamental flaw in the gyroscope hardware.

At 7:51 the next evening, after handling well in a piloted test run, Sandstorm is allowed to take its own wheel. It is driving blind, simply following a recorded list of GPS waypoints that trace an oval loop. The computer is doing the steering, but Urmson is on board as a "human kill switch," to hit the emergency stop button if something goes wrong. Four miles and half an hour later programmer Kevin Peterson clicks a button on his laptop, a command travels wirelessly to the robot, and Sandstorm brakes to a halt. "Very well done," Whittaker congratulates, and sends the vehicle back to the shop for another night of modifications.

"From now on we need everybody here 24 hours so that as soon as the vehicle returns from the field, people jump on it and start working," Whittaker says in the morning. "It is exciting to see a robot first spring into action. But the point is to make this kind of driving boring. A 150-mile traverse in the next five days, while taking sensor data: that's the final exam, and it's pass/fail."

DECEMBER 8: The Red Team has set up camp by the empty blast furnaces to watch the robot make its 15-hour nonstop, unguided journey. They record a figure-eight test path, but the machine gets confused at the crossing point; sometimes it goes left and sometimes right. So they go back to the oval loop.

But before the test can begin, a short circuit sends current surging through a wireless "E-stop" receiver that DARPA has provided so that race officials can disable any robot that goes berserk. With that receiver fried, the team has no fail-safe way to force Sandstorm to stop—only a piece of software. Peterson and Martin Stolle, two of the team's software gurus, urge Whittaker not to rely on the software.

Urmson arrives with a servomotor borrowed from a radio-controlled airplane and proceeds to jury-rig a wireless kill switch. But that transmitter also shorts out. "So now we have just Martin's software stop," Whittaker sighs. "Martin, how many hours do you have on your controller?" he asks.

"We've tested it for about half an hour," Martin replies. Moreover, he warns, if the onboard computer fails, "we will lose all control, and the vehicle will just plow ahead until it hits an obstacle larger than a Humvee."

Urmson huddles the team together. "We can go ahead, but we all need to understand and agree that—"

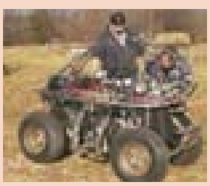
"Everyone understands it, and I'm accountable," Red interrupts. "It's not a question of pros or cons; we're going to do it." The sun has set, and the slush on the track is refreezing. Whittaker insists that two team members stay in the open to keep watch as the robot drives 792 laps around its short test loop.

With a puff of gray smoke, Sandstorm zooms forward. As it rounds the first two turns and enters a straightaway, sparks appear in the undercarriage. It skids to a stop on command, and team members sprint out with a fire extinguisher. The cause is

The Competitors

MORE THAN 100 TEAMS registered for the Grand Challenge; 86 sent technical applications to DARPA, which approved 45. DARPA officials later culled the field to 25 vehicles, which fall into roughly four categories. No more than 20 will be allowed to race.

Modified All-Terrain Vehicles

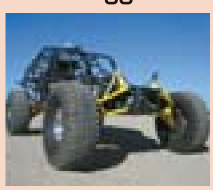


Pros: Inexpensive; off-road suspensions are standard; can stop, turn and accelerate quickly; small size provides a margin of error on narrow trails.

Cons: Sensors are low and thus limited in their range of view; high risk of critical damage in a collision; very limited ability to generate electrical power; small fuel tanks; overturn easily.

Teams: ENSCO, Phantasm (*pictured*), Virginia Tech

Dune Buggies

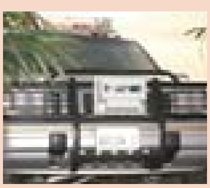


Pros: Very low center of gravity prevents overturning; frame and suspension are customized for desert racing; lightweight, agile and fast.

Cons: Sensors are low and vulnerable to collisions and dust; small wheels; low mass and electrical budgets limit onboard computing.

Teams: AI Motorvator, CyberRider (*pictured*), LoGHIQ, Sciautonics (which is fielding two robots)

Modified Sport-Utility Vehicles



Pros: Easily acquired; good ground clearance; large enclosed interior for electronics; powerful engines; high mounting points for sensors.

Cons: Expensive; high rollover risk; complex electrical system; suspension is designed for paved roads rather than trails.

Teams: Arctic Tortoise, Axion Racing (*pictured*), Caltech, Digital Auto Drive, Insight Racing, Navigators, Overbot, Palos Verdes Road Warriors

Modified Military Vehicles

Pros: Very high ground clearance, stability and crash tolerance; powerful engines and large chassis can easily carry a large payload of electronics and computers; high vantage point for sensors.

Cons: Expensive and hard to obtain; parts are difficult to find; stiff suspension creates problems for sensors; wide turning radius; relatively slow acceleration and braking.

Teams: The Red Team, Terramax (*pictured*)

TEAM PHANTASM (*ATV*); TEAM CYBERRIDER (*dune buggy*); AXION RACING (*SUV*); TEAM TERRAMAX (*military vehicle*)

JANUARY 20: Sandstorm grows faster, smarter and more robust almost every day. Yet Whittaker still gives it only 40 percent odds of finishing the race.

W. WAYT GIBBS

innocuous: someone had forgotten to refill a gas cylinder that keeps the parking brake released, so it was driving with its brake on. They push the vehicle back onto the course, only to find that the batteries have failed.

And so it went for the next several days, with one thing after another going wrong. While Smith and Kato managed to conquer the bugs in the gimbal and get one of its three arms working for 15 hours, gremlins bedeviled the rest of the Red Team, sending Sandstorm careening into, and later through, a chain-link fence. In the wee hours of December 13, the robot was just clearing 119 miles when it headed for the hills and had to be stopped. They persisted for two more days, through a snowstorm and bitter cold, persisted and failed.

Sprinting to the Starting Line

DECEMBER 21: "We didn't do the 150," Whittaker acknowledges, as the diehards meet to take stock. "But it was a hell of a four days. It was our battle cry, and it was magnificent."

On Christmas Eve a new shock isolation design for the E-box is tested. It works, as do all three arms on the gimbal. Christmas Day brings—what else?—test, fail, rework, repeat.

Within two weeks, as industry partners fly in for the last full team meeting on January 6, the robot is ready for its public unveiling before politicians and television cameras. Behind closed doors, Whittaker acknowledges that "in the last six months we've fallen behind a month. Following GPS waypoints, the vehicle is now rock-solid, to the point where you can turn your back on it." Sandstorm has graduated from a paved lot to an open field, where it now safely drives by itself at more than 30 miles an hour.

But although the machine can see the world, it cannot yet reason enough to avoid obstacles. Even with 10 of the most powerful processors that Intel makes installed in Sandstorm, the computers formulate their plans about a third too slowly.

In February the robot and its creators will head to the desert. "We need to put 10,000 miles of testing on it," Whittaker says. "This fancy stuff could shake apart because it's all prototype. Just inside the E-box there are 5,000-odd components, a failure in any one of which could screw us up. Any team could beat us."

And if the Red Team wins? The best thing about building a new race of robots, Whittaker said one frigid night in December as we watched Sandstorm do its laps beneath a nearly perfect full moon, is not the act of creation. "What's most fun is exploring the space of possibilities you have opened with your invention. I'm thinking about proposing a mission to NASA to launch a lunar rover that could circumnavigate the pole of the moon, searching for ice." Other team members have suggested building a robot to run the Iditarod in Alaska or to serve as an ambulance in Antarctica.

More likely, however, the $1-million prize will go unclaimed this year and the contest will repeat in 2005. "If no one wins this race and we recommit for next year, who's in?" Whittaker asks at the end of the meeting. Up go a roomful of hands. SA

Senior writer W. Wayt Gibbs has been in Pittsburgh covering the progress of the Red Team since March 2003.

MORE TO EXPLORE

The Red Team Web site: **redteamracing.org**

For links to other teams: **www.darpa.mil/grandchallenge/teams.htm**

For more information on the Grand Challenge race: **www.darpa.mil/grandchallenge/**

ICEBERG BREAKS OFF the San Rafael Glacier in Chile. Global disintegration of ice masses has the potential to raise sea level by several meters or more. The grim consequences of a rising sea level set a low threshold for how much the planet can warm without disrupting human society.

Global warming is real, and

Nevertheless, practical actions, which

Defusing the Global Warming TIME BOMB

BY JAMES HANSEN

the consequences are potentially disastrous.
would also yield a cleaner, healthier atmosphere, could slow, and eventually stop, the process

A paradox in the notion of human-made global warming

became strikingly apparent to me one summer afternoon in 1976 on Jones Beach, Long Island. Arriving at midday, my wife, son and I found a spot near the water to avoid the scorching hot sand. As the sun sank in the late afternoon, a brisk wind from the ocean whipped up whitecaps. My son and I had goose bumps as we ran along the foamy shoreline and watched the churning waves.

That same summer Andy Lacis and I, along with other colleagues at the NASA Goddard Institute for Space Studies, had estimated the effects of greenhouse gases on climate. It was well known by then that human-made greenhouse gases, especially carbon dioxide and chlorofluorocarbons (CFCs), were accumulating in the atmosphere. These gases are a climate "forcing," a perturbation imposed on the energy budget of the planet. Like a blanket, they absorb infrared (heat) radiation that would otherwise escape from the earth's surface and atmosphere to space.

Our group had calculated that these human-made gases were heating the earth's surface at a rate of almost two watts per square meter. A miniature Christmas tree bulb dissipates about one watt, mostly in the form of heat. So it was as if humans had placed two of these tiny bulbs over every square meter of the earth's surface, burning night and day.

The paradox that this result presented was the contrast between the awesome forces of nature and the tiny lightbulbs. Surely their feeble heating could not command the wind and waves or smooth our goose bumps. Even their imperceptible heating of the ocean surface must be quickly dissipated to great depths, so it must take many years, perhaps centuries, for the ultimate surface warming to be achieved.

This seeming paradox has now been largely resolved through study of the history of the earth's climate, which reveals that small forces, maintained long enough, can cause large climate change. And, consistent with the historical evidence, the earth has begun to warm in recent decades at a rate predicted by climate models that take account of the atmospheric accumulation of human-made greenhouse gases. The warming is having noticeable impacts as glaciers are retreating worldwide, Arctic sea ice has thinned, and spring comes about one week earlier than when I grew up in the 1950s.

Yet many issues remain unresolved. How much will climate change in coming decades? What will be the practical consequences? What, if anything, should we do about it? The debate over these questions is highly charged because of the inherent economic stakes.

Objective analysis of global warming requires quantitative knowledge of three issues: the sensitivity of the climate system to forcings, the forcings that humans are introducing, and the time required for climate to respond. All these issues can be studied with global climate models, which are numerical simulations on computers. But our most accurate knowledge about climate sensitivity, at least so far, is based on empirical data from the earth's history.

Overview/*Global Warming*

- At present, our most accurate knowledge about climate sensitivity is based on data from the earth's history, and this evidence reveals that small forces, maintained long enough, can cause large climate change.
- Human-made forces, especially greenhouse gases, soot and other small particles, now exceed natural forces, and the world has begun to warm at a rate predicted by climate models.
- The stability of the great ice sheets on Greenland and Antarctica and the need to preserve global coastlines set a low limit on the global warming that will constitute "dangerous anthropogenic interference" with climate.
- Halting global warming requires urgent, unprecedented international cooperation, but the needed actions are feasible and have additional benefits for human health, agriculture and the environment.

The Lessons of History

OVER THE PAST few million years the earth's climate has swung repeatedly between ice ages and warm interglacial periods. A 400,000-year record of temperature is preserved in the Antarctic ice sheet, which, except for coastal fringes, escaped melting even in the warmest interglacial periods. This record [*see box on opposite page*] suggests that the present interglacial period (the Holocene), now about 12,000 years old, is already long of tooth.

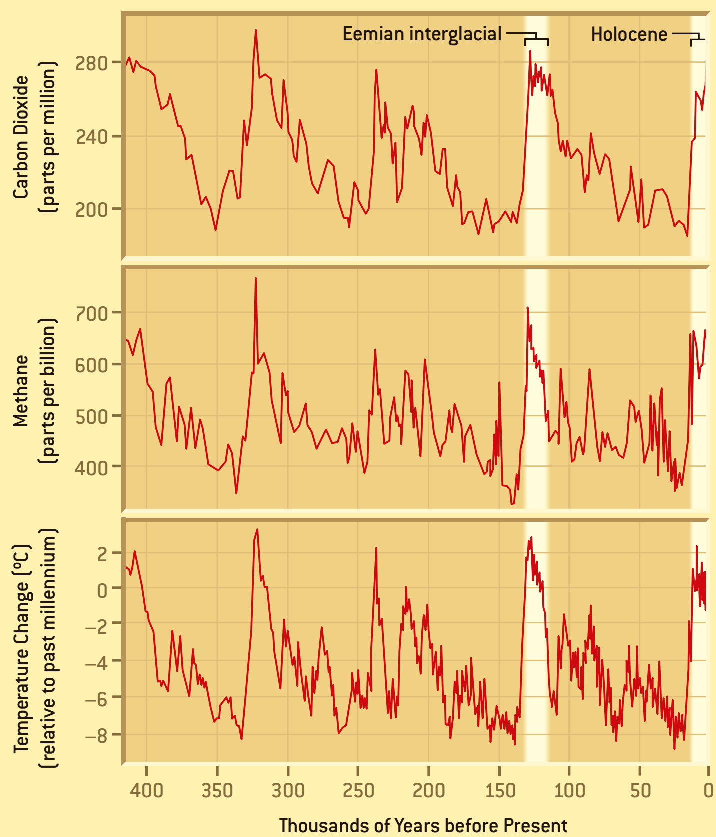
The natural millennial climate swings are associated with slow variations of the earth's orbit induced by the gravity of other planets, mainly Jupiter and Saturn (because they are so heavy) and Venus (because it comes so close). These perturbations hardly affect the annual mean solar energy striking the earth, but they alter the geographical and seasonal distribution of incoming solar energy, or insolation, as much as 20 percent. The insolation changes, over long periods, affect the building and melting of ice sheets.

Insolation and climate changes also affect uptake and release of carbon dioxide and methane by plants, soil and the ocean. Climatologists are still developing a quantitative understanding of the mechanisms by which the ocean and land release carbon dioxide and methane as the earth warms, but the paleoclimate data are already a gold mine of information. The most critical insight that the ice age climate swings provide is an empirical measure of climate sensitivity.

The composition of the ice age atmosphere is known precisely from air bubbles trapped as the Antarctic and Greenland ice sheets and numerous mountain glaciers built up from annual snowfall. Furthermore, the geographical distributions of the ice sheets, vegetation cover and coastlines during the ice age are well mapped. From these data we know that the change of

RIC ERGENBRIGHT *Corbis* (*preceding pages*)

400,000 YEARS OF CLIMATE CHANGE



ANTARCTIC ICE has preserved a 400,000-year record of temperature and of levels of carbon dioxide and methane in the atmosphere. Scientists study gases trapped in air bubbles in the ice—generally using ice cores (*photograph*) extracted from the ice sheet and transported to a laboratory. The historical record provides us with two critical measures: Comparison of the current interglacial period (the Holocene) with the most recent ice age (20,000 years ago) gives an accurate measure of climate sensitivity to forcings. The temperature in the previous interglacial period (the Eemian), when sea level was several meters higher than today, defines an estimate of the warming that today's civilization would consider to be dangerous anthropogenic interference with climate.

JEN CHRISTIANSEN; SOURCE: J. R. PETIT ET AL. IN *NATURE*, VOL. 399, PAGES 429–436; JUNE 3, 1999 (*graphs*); CSIRO/SIMON FRASER *Photo Researchers, Inc.* (*photograph*)

climate forcing between the ice age and today was about 6.5 watts per square meter. This forcing maintains a global temperature change of 5 degrees Celsius (9 degrees Fahrenheit), implying a climate sensitivity of 0.75 ± 0.25 degrees C per watt per square meter. Climate models yield a similar climate sensitivity. The empirical result is more precise and reliable, however, because it includes all the processes operating in the real world, even those we have not yet been smart enough to include in the models.

The paleodata provide another important insight. Changes of the earth's orbit instigate climate change, but they operate by altering atmosphere and surface properties and thus the planetary energy balance. These atmosphere and surface properties are now influenced more by humans than by our planet's orbital variations.

Climate-Forcing Agents Today

THE LARGEST change of climate forcings in recent centuries is caused by human-made greenhouse gases. Greenhouse gases in the atmosphere absorb heat radiation rather than letting it escape into space. In effect, they make the proverbial blanket thicker, returning more heat toward the ground rather than letting it escape to space. The earth then is radiating less energy to space than it absorbs from the sun. This temporary planetary energy imbalance results in the earth's gradual warming.

Because of the large capacity of the oceans to absorb heat, it takes the earth about a century to approach a new balance—that is, for it to once again receive the same amount of energy from the sun that it radiates to space. And of course the balance is reset at a higher temperature. In the meantime, before it achieves this equilibrium, more forcings may be added.

The single most important human-made greenhouse gas is carbon dioxide, which comes mainly from burning fossil fuels (coal, oil and gas). Yet the combined effect of the other human-made gases is comparable. These other gases, especially tropospheric ozone and its precursors, including methane, are ingredients in smog that damage human health and agricultural productivity.

Aerosols (fine particles in the air) are the other main human-made climate forcing. Their effect is more complex. Some "white" aerosols, such as sulfates arising from sulfur in fossil fuels, are highly reflective and thus reduce solar heating of the earth; however, black carbon (soot), a product of incomplete combustion of fossil fuels, biofuels and outdoor biomass burning, absorbs sunlight and thus heats the atmosphere. This aerosol direct climate forcing is uncertain by at least 50 percent, in part because aerosol amounts are not well measured and in part because of their complexity.

small forces, maintained long enough, can cause large climate change

Aerosols also cause an indirect climate forcing by altering the properties of clouds. The resulting brighter, longer-lived clouds reduce the amount of sunlight absorbed by the earth, so the indirect effect of aerosols is a negative forcing that causes cooling.

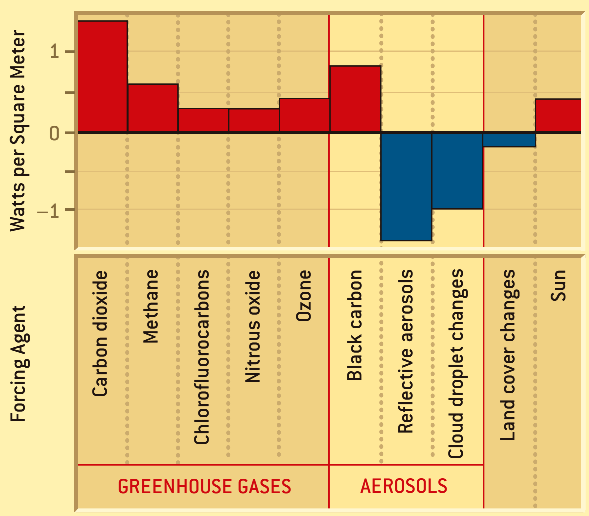
Other human-made climate forcings include replacement of forests by cropland. Forests are dark even with snow on the ground, so their removal reduces solar heating.

Natural forcings, such as volcanic eruptions and fluctuations of the sun's brightness, probably have little trend on a timescale of 1,000 years. But evidence of a small solar brightening over the past 150 years implies a climate forcing of a few tenths of a watt per square meter.

The net value of the forcings added since 1850 is 1.6 ± 1.0 watts per square meter. Despite the large uncertainties, there is evidence that this estimated net forcing is approximately correct. One piece of evidence is the close agreement of observed global temperature during the past several decades with climate models driven by these forcings. More fundamentally, the observed heat gain by the world ocean in the past 50 years is consistent with the estimated net climate forcing.

CLIMATE FORCINGS

A CLIMATE FORCING is a mechanism that alters the global energy balance. A forcing can be natural—fluctuations in the earth's orbit, for example—or human-made, such as aerosols and greenhouse gases. Human-made climate forcings now dominate natural forcings. Carbon dioxide is the largest forcing, but air pollutants (black carbon, ozone, methane) together are comparable. (Aerosol effects are not known accurately.)



Global Warming

GLOBAL AVERAGE surface temperature has increased about 0.75 degree C during the period of extensive instrumental measurements, which began in the late 1800s. Most of the warming, about 0.5 degree C, occurred after 1950. The causes of observed warming can be investigated best for the past 50 years, because most climate forcings were observed then, especially since satellite measurements of the sun, stratospheric aerosols and ozone began in the 1970s. Furthermore, 70 percent of the anthropogenic increase of greenhouse gases occurred after 1950.

The most important quantity is the planetary energy imbalance [*see box on page 75*]. This imbalance is a consequence of the long time that it takes the ocean to warm. We conclude that the earth is now out of balance by something between 0.5 and one watt per square meter—that much more solar radiation is being absorbed by the earth than is being emitted as heat to space. Even if atmospheric composition does not change further, the earth's surface will therefore eventually warm another 0.4 to 0.7 degree C.

Most of the energy imbalance has been heat going into the ocean. Sydney Levitus of the National Oceanic and Atmospheric Administration has analyzed ocean temperature changes of the past 50 years, finding that the world ocean heat content increased about 10 watt-years per square meter in the past 50 years. He also finds that the rate of ocean heat storage in recent years is consistent with our estimate that the earth is now out of energy balance by 0.5 to one watt per square meter. Note that the amount of heat required to melt enough ice to raise sea level one meter is about 12 watt-years (averaged over the planet), energy that could be accumulated in 12 years if the planet is out of balance by one watt per square meter.

The agreement with observations, for both the modeled temperature change and ocean heat storage, leaves no doubt that observed global climate change is being driven by natural and anthropogenic forcings. The current rate of ocean heat storage is a critical planetary metric: it not only determines the amount of additional global warming already in the pipeline, but it also equals the reduction in climate forcings needed to stabilize the earth's present climate.

The Time Bomb

THE GOAL OF the United Nations Framework Convention on Climate Change, produced in Rio de Janeiro in 1989, is to stabilize atmospheric composition to "prevent dangerous anthropogenic interference with the climate system" and to

JEN CHRISTIANSEN; SOURCE: JAMES HANSEN

HUMAN-MADE climate forcings, mainly greenhouse gases, heat the earth's surface at a rate of about two watts per square meter—the equivalent of two tiny one-watt bulbs burning over every square meter of the planet. The full effect of the warming is slowed by the ocean, because it can absorb so much heat. The ocean's surface begins to warm, but before it can heat up much, the surface water is mixed down and replaced by colder water from below. Scientists now think it takes about a century for the ocean to approach its new temperature.

RANDY HARRIS

achieve that goal in ways that do not disrupt the global economy. Defining the level of warming that constitutes "dangerous anthropogenic interference" is thus a crucial but difficult part of the problem.

The U.N. established an Intergovernmental Panel on Climate Change (IPCC) with responsibility for analysis of global warming. The IPCC has defined climate-forcing scenarios, used these for simulations of 21st-century climate, and estimated the impact of temperature and precipitation changes on agriculture, natural ecosystems, wildlife and other matters. The IPCC estimates sea-level change as large as several tens of centimeters in 100 years, if global warming reaches several degrees Celsius. The group's calculated sea-level change is due mainly to thermal expansion of ocean water, with little change in ice-sheet volume.

These moderate climate effects, even with rapidly increasing greenhouse gases, leave the impression that we are not close to dangerous anthropogenic interference. I will argue, however, that we are much closer than is generally realized, and thus the emphasis should be on mitigating the changes rather than just adapting to them.

The dominant issue in global warming, in my opinion, is sea-level change and the question of how fast ice sheets can disintegrate. A large portion of the world's people live within a few meters of sea level, with trillions of dollars of infrastructure. The need to preserve global coastlines sets a low ceiling on the level of global warming that would constitute dangerous anthropogenic interference.

The history of the earth and the present human-made planetary energy imbalance together paint a disturbing picture about prospects for sea-level change. Data from the Antarctic temperature record show that the warming of the past 50 years has taken global temperature back to approximately the peak of the current interglacial (the Holocene). There is some additional warming in the pipeline that will take us about halfway to the highest global temperature level of the previous interglacial (the Eemian), which was warmer than the Holocene, with sea level estimated to have been five to six meters higher. One additional watt per square meter of forcing, over and above that today, will take global temperature approximately to the maximum level of the Eemian.

The main issue is: How fast will ice sheets respond to global warming? The IPCC calculates only a slight change in the ice sheets in 100 years; however, the IPCC calculations include only the gradual effects of changes in snowfall, evaporation and melting. In the real world, ice-sheet disintegration is driven by highly nonlinear processes and feedbacks. The peak rate of deglaciation following the last ice age was a sustained rate of melting of more than 14,000 cubic kilometers a year—about one meter of sea-level rise every 20 years, which was maintained for several centuries. This period of most rapid melt coincided, as well

THE AUTHOR

JAMES HANSEN is director of the NASA Goddard Institute for Space Studies and a researcher at the Columbia University Earth Institute. He received his Ph.D. in physics and astronomy from the University of Iowa, where he studied under James Van Allen. Hansen is best known for his testimony to congressional committees in the 1980s that helped to raise awareness of the global warming issue.

R. J. BRAITHWAITE, REPRINTED WITH PERMISSION FROM *SCIENCE*, VOL. 297, NO. 5579; JULY 12, 2002. © 2002 AAAS

as can be measured, with the time of most rapid warming.

Given the present unusual global warming rate on an already warm planet, we can anticipate that areas with summer melt and rain will expand over larger areas of Greenland and fringes of Antarctica. Rising sea level itself tends to lift marine ice shelves that buttress land ice, unhinging them from anchor points. As ice shelves break up, this accelerates movement of land ice to the ocean. Although building of glaciers is slow, once an ice sheet begins to collapse, its demise can be spectacularly rapid.

The human-induced planetary energy imbalance provides an ample supply of energy for melting ice. Furthermore, this energy source is supplemented by increased absorption of sunlight by ice sheets darkened by black-carbon aerosols, and the positive feedback process as meltwater darkens the ice surface.

ON A SLIPPERY SLOPE to disaster, a stream of snowmelt cascades down a moulin on the Greenland ice sheet during a recent summer. The moulin, a near-vertical shaft worn in the ice by surface water, carries water to the base of the ice sheet. There the water is a lubricating fluid that speeds motion and disintegration of the ice sheet. Ice sheet growth is a slow, dry process, inherently limited by the snowfall rate, but disintegration is a wet process, driven by positive feedbacks, and once well under way it can be explosively rapid.

These considerations do not mean that we should expect large sea-level change in the next few years. Preconditioning of ice sheets for accelerated breakup may require a long time, perhaps many centuries. (The satellite ICESat, recently launched by NASA, may be able to detect early signs of accelerating ice-sheet breakup.) Yet I suspect that significant sea-level rise could begin much sooner if the planetary energy imbalance continues

we are much closer to dangerous anthropogenic interference than is generally realized

to increase. It seems clear that global warming beyond some limit will make a large sea-level change inevitable for future generations. And once large-scale ice-sheet breakup is under way, it will be impractical to stop. Dikes may protect limited regions, such as Manhattan and the Netherlands, but most of the global coastlines will be inundated.

I argue that the level of dangerous anthropogenic influence is likely to be set by the global temperature and planetary radiation imbalance at which substantial deglaciation becomes practically impossible to avoid. Based on the paleoclimate evidence, I suggest that the highest prudent level of additional global warming is not more than about one degree C. This means that additional climate forcing should not exceed about one watt per square meter.

Climate-Forcing Scenarios

THE IPCC defines many climate-forcing scenarios for the 21st century based on multifarious "story lines" for population growth, economic development and energy sources. It estimates that added climate forcing in the next 50 years is one to three watts per square meter for carbon dioxide and two to four watts per square meter with other gases and aerosols included. Even the IPCC's minimum added forcing would cause dangerous anthropogenic interference with the climate system based on our criterion.

The IPCC scenarios may be unduly pessimistic, however. First, they ignore changes in emissions, some already under way, because of concerns about global warming. Second, they assume that true air pollution will continue to get worse, with ozone, methane and black carbon all greater in 2050 than in 2000. Third, they give short shrift to technology advances that can reduce emissions in the next 50 years.

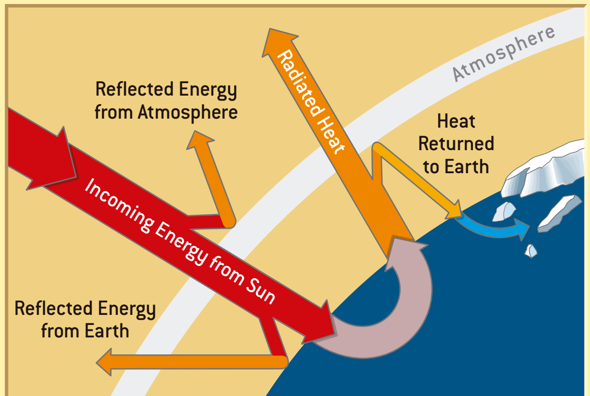
An alternative way to define scenarios is to examine current trends of climate-forcing agents, to ask why they are changing as observed, and to try to understand whether reasonable actions could encourage further changes in the growth rates.

The growth rate of the greenhouse-gas climate forcing peaked in the early 1980s at almost 0.5 watt per square meter per decade but declined by the 1990s to about 0.3 watt per square meter per decade. The primary reason for the decline was reduced emissions of chlorofluorocarbons, whose production was phased out because of their destructive effect on stratospheric ozone.

The two most important greenhouse gases, with chlorofluorocarbons on the decline, are carbon dioxide and methane. The growth rate of carbon dioxide surged after World War II, flattened out from the mid-1970s to the mid-1990s, and rose moderately in recent years to the current growth rate of about two parts per million per year. The methane growth rate has declined dramatically in the past 20 years, by at least two thirds.

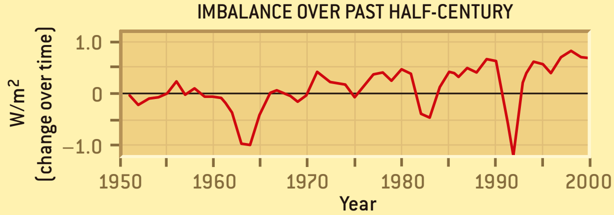
EARTH'S ENERGY IMBALANCE

THE EARTH'S ENERGY is balanced when the outgoing heat from the earth equals the incoming energy from the sun. At present the energy budget is not balanced (*diagram* and *table*). Human-made aerosols have increased reflection of sunlight by the earth, but this reflection is more than offset by the trapping of heat radiation by greenhouse gases. The excess energy—about one watt per square meter—warms the ocean and melts ice. The simulated planetary energy imbalance (*graph*) is confirmed by measurements of heat stored in the oceans. The planetary energy imbalance is a critical metric, in that it measures the net climate forcing and foretells future global warming already in the pipeline.



TOTAL INCOMING SOLAR ENERGY	340 W/m^2
TOTAL OUTGOING ENERGY	339 W/m^2
REFLECTED ENERGY (from atmosphere and surface)	101 W/m^2
100 W/m^2 because of natural processes	
1 W/m^2 because of human-made aerosols	
RADIATED HEAT (from land and ocean sinks)	238 W/m^2
240 W/m^2 because of natural processes	
–2 W/m^2 because of human-made greenhouse gases, which return heat to the surface	
NET RESULT	1 W/m^2
1 W/m^2 of excess energy, which warms the oceans and melts glaciers and ice sheets	

IMBALANCE OVER PAST HALF-CENTURY



JEN CHRISTIANSEN; SOURCE: JAMES HANSEN

the emphasis should be on mitigating the changes rather than just adapting to them

These growth rates are related to the rate of global fossil-fuel use. Fossil-fuel emissions increased by more than 4 percent a year from the end of World War II until 1975 but subsequently by only about 1 percent a year. The change in fossil-fuel growth rate occurred after the oil embargo and price increases of the 1970s, with subsequent emphasis on energy efficiency. Methane growth has also been affected by other factors, including changes in rice farming and increased efforts to capture methane at landfills and in mining operations.

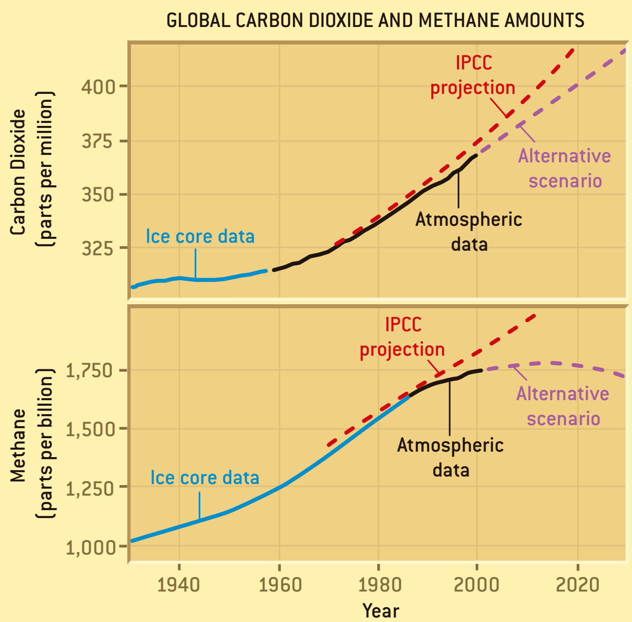
If recent growth rates of these greenhouse gases continued, the added climate forcing in the next 50 years would be about 1.5 watts per square meter. To this must be added the change caused by other forcings, such as atmospheric ozone and aerosols. These forcings are not well monitored globally, but it is known that they are increasing in some countries while decreasing in others. Their net effect should be small, but it could add as much as 0.5 watt per square meter. Thus, if there is no slowing of emission rates, the human-made climate forcing could increase by two watts per square meter in the next 50 years.

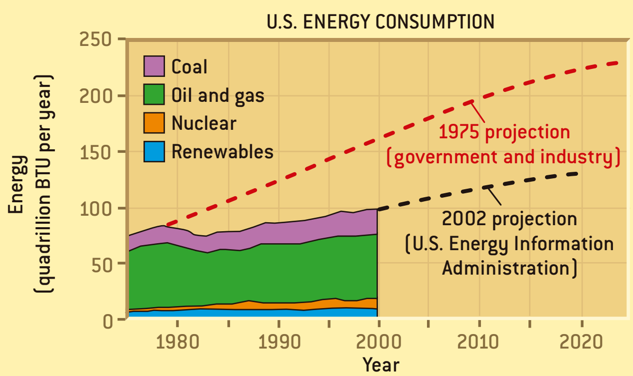
This "current trends" growth rate of climate forcings is at the low end of the IPCC range of two to four watts per square meter. The IPCC four watts per square meter scenario requires 4 percent a year exponential growth of carbon dioxide emissions maintained for 50 years and large growth of air pollution; it is implausible.

Nevertheless, the "current trends" scenario is larger than the one watt per square meter level that I suggested as our current best estimate for the level of dangerous anthropogenic in-

REDUCING EMISSIONS

OBSERVED AMOUNTS of carbon dioxide and methane (*top two graphs*) fall below IPCC estimates, which have proved consistently pessimistic. Although the author's alternative scenario agrees better with observations, continuation on that path requires a gradual slowdown in carbon dioxide and methane emissions. Improvements in energy efficiency (*bottom graph*) have allowed energy use in the U.S. to fall below projections in recent decades, but more rapid efficiency gains are needed to achieve the carbon dioxide emissions of the alternative scenario, unless nuclear power and renewable energies grow substantially.





JEN CHRISTIANSEN; SOURCE: JAMES HANSEN (*graphs*); MARK BOULTON *Photo Researchers, Inc.* (*photograph*) fluence. This raises the question of whether there is a feasible scenario with still lower climate forcing.

A Brighter Future

I HAVE DEVELOPED a specific alternative scenario that keeps added climate forcing in the next 50 years at about one watt per square meter. It has two components: first, halt or reverse growth of air pollutants, specifically soot, atmospheric ozone and methane; second, keep average fossil-fuel carbon dioxide emissions in the next 50 years about the same as today. The carbon dioxide and non–carbon dioxide portions of the scenario are equally important. I argue that they are feasible and at the same time protect human health and increase agricultural productivity.

In addressing air pollution, we should emphasize the constituents that contribute most to global warming. Methane offers a great opportunity. If human sources of methane are reduced, it may even be possible to get the atmospheric methane amount to decline, thus providing a cooling that would partially offset the carbon dioxide increase. Reductions of black-carbon aerosols would help counter the warming effect of reductions in sulfate aerosols. Atmospheric ozone precursors, besides methane, especially nitrogen oxides and volatile organic compounds, must be reduced to decrease low-level atmospheric ozone, the prime component of smog.

Actions needed to reduce methane, such as methane capture at landfills and at waste management facilities and during the mining of fossil fuels, have economic benefits that partially offset the costs. In some cases, methane's value as a fuel entirely pays for the cost of capture. Reducing black carbon would also have economic benefits, both in the decreased loss of life and work-years (minuscule soot particles carry toxic organic compounds and metals deep into lungs) and in increased agricultural productivity in certain parts of the world. Prime sources of black carbon are diesel fuels and biofuels (wood and cow dung, for example). These sources need to be dealt with for health reasons. Diesel could be burned more cleanly with improved technologies; however, there may be even better solutions, such as hydrogen fuel, which would eliminate ozone precursors as well as soot.

Improved energy efficiency and increased use of renewable energies might level carbon dioxide emissions in the near term. Long-term reduction of carbon dioxide emissions is a greater challenge, as energy use will continue to rise. Progress is needed across the board: continued efficiency improvements, more renewable energy, and new technologies that produce little or no carbon dioxide or that capture and sequester it. Next-generation nuclear power, if acceptable to the public, could be an important contributor. There may be new technologies before 2050 that we have not imagined.

Observed global carbon dioxide and methane trends [*see box on opposite page*] for the past several years show that the real world is falling below all IPCC scenarios. It remains to be proved whether the smaller observed growth rates are a fluke, soon to return to IPCC rates, or are a meaningful difference. In contrast, the projections of my alternative scenario and the observed growth rates are in agreement. This is not surprising, because that scenario was defined with observations in mind. And in the three years since the alternative scenario was defined, observations have continued on that path. I am not suggesting, however, that the alternative scenario can be achieved without concerted efforts to reduce anthropogenic climate forcings.

How can I be optimistic if climate is closer to the level of dangerous anthropogenic interference than has been realized? If we compare the situation today with that 10 to 15 years ago, we note that the main elements required to halt climate change have come into being with remarkable rapidity. I realize that it will not be easy to stabilize greenhouse-gas concentrations, but I am optimistic because I expect that empirical evidence for climate change and its impacts will continue to accumulate and that this will influence the public, public-interest groups, industry and governments at various levels. The question is: Will we act soon enough? SA

For an expanded version of this article, including more data and additional sources, see www.sciam.com/ontheweb

BUT WHAT ABOUT . . .

"Last winter was so cold! I don't notice any global warming!"
Global warming is ubiquitous, but its magnitude so far is only about one degree Fahrenheit. Day-to-day weather fluctuations are roughly 10 degrees F. Even averaged over a season this natural year-to-year variability is about two degrees F, so global warming does not make every season warmer than a few decades ago. But global warming already makes the probability of a warmer than "normal" season about 60 percent, rather than the 30 percent that prevailed from 1950 to 1980.

"The warming of the past century is just a natural rebound from the little ice age."
Any rebound from the European little ice age, which peaked in 1650–1750, would have been largely complete by the 20th century. Indeed, the natural long-term climate trend today would be toward a colder climate were it not for human activities.

"Isn't human-made global warming saving us from the next ice age?"
Yes, but the gases that we have added to the atmosphere are already far more than needed for that purpose.

"The surface warming is mainly urban 'heat island' effects near weather stations."
Not so. As predicted, the greatest warming is found in remote regions such as central Asia and Alaska. The largest areas of surface warming are over the ocean, far from urban locations [see maps at www.giss.nasa.gov/data/update/gistemp]. Temperature profiles in the solid earth, at hundreds of boreholes around the world, imply a warming of the continental surfaces between 0.5 and one degree C in the past century.

JAMES HANSEN

The Addicted BRAIN

Drug abuse produces long-term changes in the reward circuitry of the brain. Knowledge of the cellular and molecular details of these adaptations could lead to new treatments for the compulsive behaviors that underlie addiction

By Eric J. Nestler and Robert C. Malenka

White lines on a mirror. A needle and spoon. For many users, the sight of a drug or its associated paraphernalia can elicit shudders of anticipatory pleasure. Then, with the fix, comes the real rush: the warmth, the clarity, the vision, the relief, the sensation of being at the center of the universe. For a brief period, everything feels right. But something happens after repeated exposure to drugs of abuse—whether heroin or cocaine, whiskey or speed.

The amount that once produced euphoria doesn't work as well, and users come to need a shot or a snort just to feel normal; without it, they become depressed and, often, physically ill. Then they begin to use the drug compulsively. At this point, they are addicted, losing control over their use and suffering powerful cravings even after the thrill is gone and their habit begins to harm their health, finances and personal relationships.

Neurobiologists have long known that the euphoria induced by drugs of abuse arises because all these chemicals ultimately boost the activity of the brain's reward system:

ILLUSTRATION BY JANA BRENNING (*photograph used for illustration purposes only*)

ADDICTION ARISES in part because habit-forming drugs cause the brain's circuit for assessing reward to deem the drugs more desirable than anything else in life.

a complex circuit of nerve cells, or neurons, that evolved to make us feel flush after eating or sex—things we need to do to survive and pass along our genes. At least initially, goosing this system makes us feel good and encourages us to repeat whatever activity brought us such pleasure.

But new research indicates that chronic drug use induces changes in the structure and function of the system's neurons that last for weeks, months or years after the last fix. These adaptations, perversely, dampen the pleasurable effects of a chronically abused substance yet also increase the cravings that trap the addict in a destructive spiral of escalating use and increased fallout at work and at home. Improved understanding of these neural alterations should help provide better interventions for addiction, so that people who have fallen prey to habit-forming drugs can reclaim their brains and their lives.

Drugs to Die For

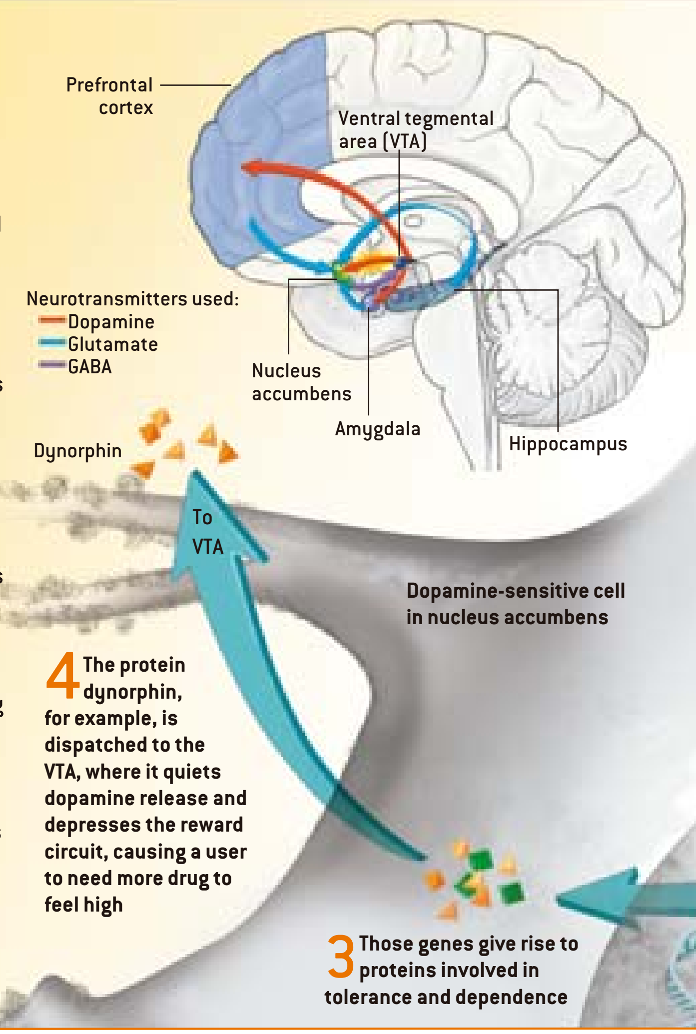
THE REALIZATION that various drugs of abuse ultimately lead to addiction through a common pathway emerged largely from studies of laboratory animals that began about 40 years ago. Given the opportunity, rats, mice and nonhuman primates will self-administer the same substances that humans abuse. In these experiments, the animals are connected to an intravenous line. They are then taught to press one lever to receive an infusion of drug through the IV, another lever to get a relatively uninteresting saline solution, and a third lever to request a food pellet. Within a few days, the animals are hooked: they readily self-administer cocaine, heroin, amphetamine and many other common habit-forming drugs.

What is more, they eventually display assorted behaviors of addiction. Individual animals will take drugs at the expense of normal activities such as eating and sleeping—some even to the point that they die of exhaustion or malnutrition. For the most addictive substances, such as cocaine, animals will spend most of their waking hours working to obtain more, even if it means pressing a lever hundreds of times for a single hit. And just as human addicts experience intense cravings when they encounter drug paraphernalia or places where they have scored, the animals, too, come to prefer an environment that they associate with the drug—an area in the cage in which lever pressing always provides chemical compensation.

When the substance is taken away, the animals soon cease to labor for chemical satisfaction. But the pleasure is not forgotten. A rat that has remained clean—even for months—will immediately return to its bar-pressing behavior when given just a taste of cocaine or placed in a cage it associates with a drug high. And certain

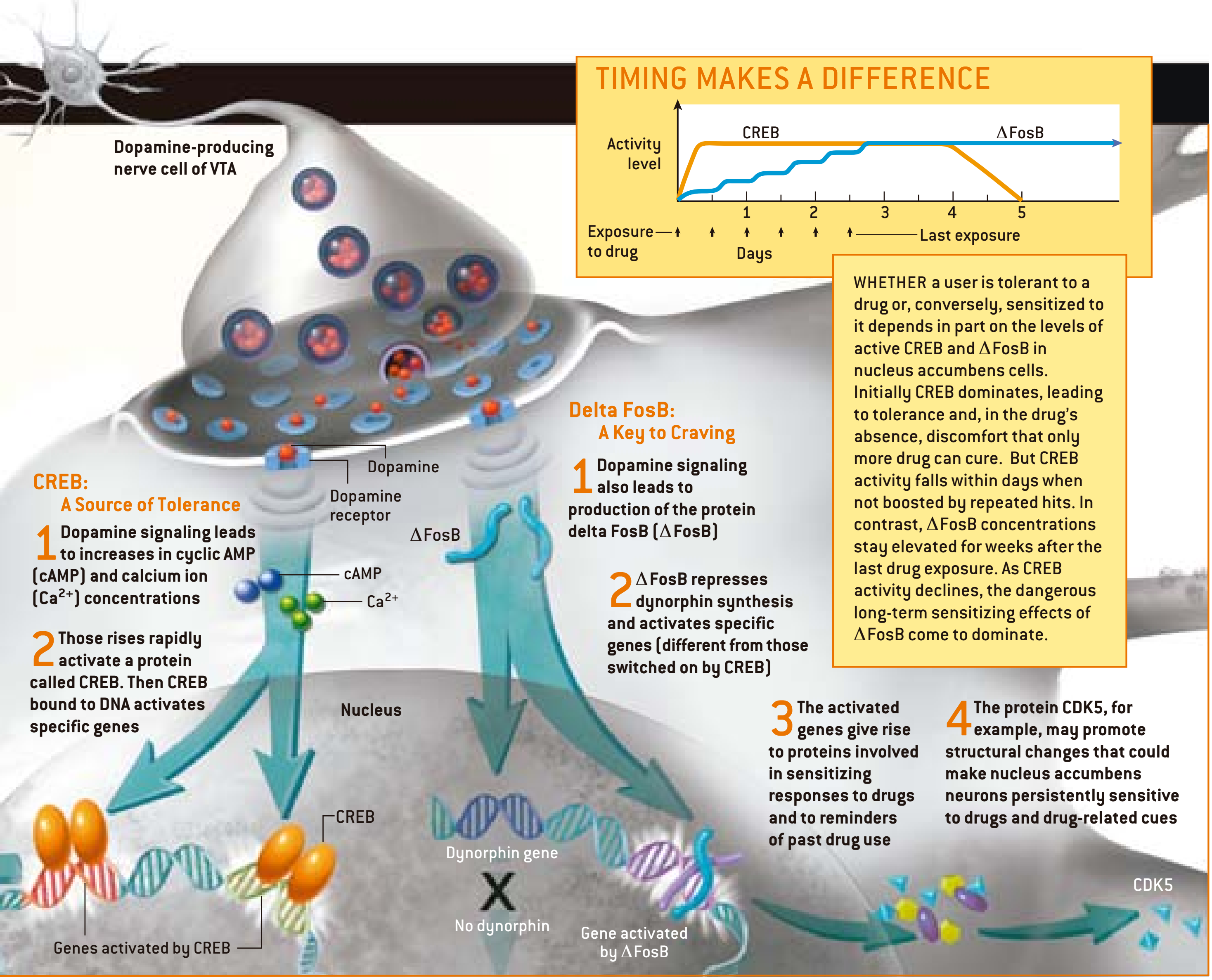
THE BRAIN UNDER THE INFLUENCE

CHRONIC USE of addictive substances can change the behavior of a key part of the brain's reward circuit: the pathway extending from the dopamine-producing nerve cells (neurons) of the ventral tegmental area (VTA) to dopamine-sensitive cells in the nucleus accumbens. Those changes, induced in part by the molecular actions depicted at the right and in the graph, contribute significantly to the tolerance, dependence and craving that fuel repeated drug use and lead to relapses even after long periods of abstention. The colored arrows on the brain indicate some of the pathways linking the nucleus accumbens and VTA with other regions that can help to make drug users highly sensitive to reminders of past highs, vulnerable to relapses when stressed, and unable to control their urges to seek drugs.



Overview/*The Evolution of Addiction*

- Drugs of abuse—cocaine, alcohol, opiates, amphetamine—all commandeer the brain's natural reward circuitry. Stimulation of this pathway reinforces behaviors, ensuring that whatever you just did, you'll want to do again.
- Repeated exposure to these drugs induces long-lasting adaptations in the brain's chemistry and architecture, altering how individual neurons in the brain's reward pathways process information and interact with one another.
- Understanding how chronic exposure to drugs of abuse reshapes an addict's brain could lead to novel, more broadly effective ways to correct the cellular and molecular aberrations that lie at the heart of all addiction.



TERESE WINSLOW

psychological stresses, such as a periodic, unexpected foot shock, will send rats scurrying back to drugs. These same types of stimuli—exposure to low doses of drug, drug-associated cues or stress—trigger craving and relapse in human addicts.

Using this self-administration setup and related techniques, researchers mapped the regions of the brain that mediate addictive behaviors and discovered the central role of the brain's reward circuit. Drugs commandeer this circuit, stimulating its activity with a force and persistence greater than any natural reward.

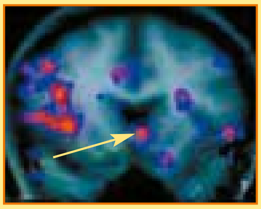
A key component of the reward circuitry is the mesolimbic dopamine system: a set of nerve cells that originate in the ventral tegmental area (VTA), near the base of the brain, and send projections to target regions in the front of the brain—most notably to a structure deep beneath the frontal cortex called the nucleus accumbens. Those VTA neurons communicate by dispatching the chemical messenger (neurotransmitter) dopamine from the terminals, or tips, of their long projections to receptors on nucleus accumbens neurons. The dopamine pathway from the VTA to the nucleus accumbens is critical for addiction: animals with lesions in these brain regions no longer show interest in substances of abuse.

Rheostat of Reward

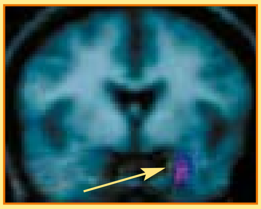
REWARD PATHWAYS are evolutionarily ancient. Even the simple, soil-dwelling worm *Caenorhabditis elegans* possesses a rudimentary version. In these worms, inactivation of four to eight key dopamine-containing neurons causes an animal to plow straight past a heap of bacteria, its favorite meal.

In mammals, the reward circuit is more complex, and it is integrated with several other brain regions that serve to color an experience with emotion and direct the individual's response to rewarding stimuli, including food, sex and social interaction. The amygdala, for instance, helps to assess whether an experience is pleasurable or aversive—and whether it should be repeated or avoided—and helps to forge connections between an experience and other cues; the hippocampus participates in recording the memories of an experience, including where and when and with whom it occurred; and the frontal regions of the cerebral cortex coordinate and process all this information and determine the ultimate behavior of the

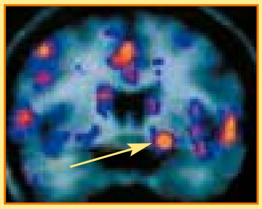
INSIGHTS FROM IMAGING



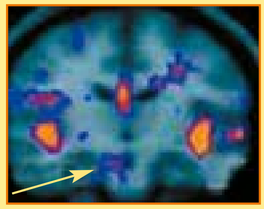
Nucleus accumbens



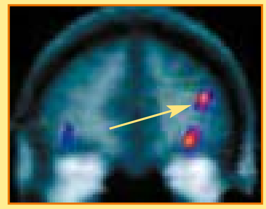
Amygdala



Sublenticular extended amygdala

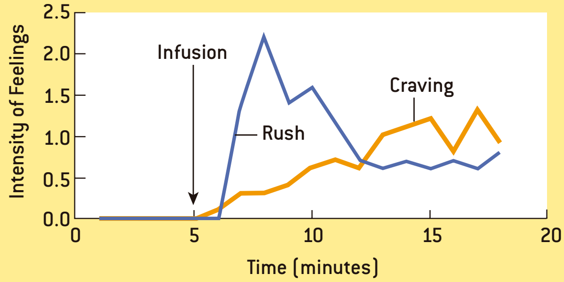


Ventral tegmental area



Prefrontal cortex

SPOTS OF COLOR in brain scans of cocaine addicts (*above*) confirm animal studies indicating that drug intake can induce profound immediate activity changes in many brain regions, including those shown; brightest spots show the most significant change. While being scanned, the subjects rated their feelings of rush and craving on a scale of zero to three—revealing that the VTA and the sublenticular extended amygdala are important to the cocaine-induced rush and that the amygdala and the nucleus accumbens influence both the rush and the craving for more drug, which becomes stronger as the euphoria wears off (*graph*).



individual. The VTA-accumbens pathway, meanwhile, acts as a rheostat of reward: it "tells" the other brain centers how rewarding an activity is. The more rewarding an activity is deemed, the more likely the organism is to remember it well and repeat it.

Although most knowledge of the brain's reward circuitry has been derived from animals, brain-imaging studies conducted over the past 10 years have revealed that equivalent pathways control natural and drug rewards in humans. Using functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) scans (techniques that measure changes in blood flow associated with neuronal activity), researchers have watched the nucleus accumbens in cocaine addicts light up when they are offered a snort. When the same addicts are shown a video of someone using cocaine or a photograph of white lines on a mirror, the accumbens responds similarly, along with the amygdala and some areas of the cortex. And the same regions react in compulsive gamblers who are shown images of slot machines, suggesting that the VTA-accumbens pathway has a similarly critical role even in nondrug addictions.

Dopamine, Please

HOW IS IT POSSIBLE that diverse addictive substances—which have no common structural features and exert a variety of effects on the body—all elicit similar responses in the brain's reward circuitry? How can cocaine, a stimulant that causes the heart to race, and heroin, a pain-relieving sedative, be so opposite in some ways and yet alike in targeting the reward system? The answer is that all drugs of abuse, in addition to any other effects, cause the nucleus accumbens to receive a flood of dopamine and sometimes also dopamine-mimicking signals.

When a nerve cell in the VTA is excited, it sends an electrical message racing along its axon—the signal-carrying "highway" that extends into the nucleus accumbens. The signal causes dopamine to be released from the axon tip into the tiny space—the synaptic cleft—that separates the axon terminal from a neuron in the nucleus accumbens. From there, the dopamine latches onto its receptor on the accumbens neuron and transmits its signal into the cell. To later shut down the signal, the VTA neuron removes the dopamine from the synaptic cleft and repackages it to be used again as needed.

Cocaine and other stimulants temporarily disable the transporter protein that returns the neurotransmitter to the VTA neuron terminals, thereby leaving excess dopamine to act on the nucleus accumbens. Heroin and other opiates, on the other hand, bind to neurons in the VTA that normally shut down the dopamine-producing VTA neurons. The opiates release this cellular clamp, thus freeing the dopamine-secreting cells to pour extra dopamine into the nucleus accumbens. Opiates can also generate a strong "reward" message by acting directly on the nucleus accumbens.

But drugs do more than provide the dopamine jolt that induces euphoria and mediates the initial reward and reinforcement. Over time and with repeated exposure, they initiate the gradual adaptations in the reward circuitry that give rise to addiction.

An Addiction Is Born

THE EARLY STAGES of addiction are characterized by tolerance and dependence. After a drug binge, an addict needs more of the substance to get the same effect on mood or concentration and so on. This tolerance then provokes an escalation of drug use that engenders dependence—a need that manifests itself as painful emotional and, at times, physical reactions if access to a drug is cut off. Both tolerance and dependence occur because frequent drug use can, ironically, suppress parts of the brain's reward circuit.

At the heart of this cruel suppression lies a molecule known as CREB (cAMP response element-binding protein). CREB is a transcription factor, a protein that regulates the expression, or activity, of genes and thus the overall behavior of

MICROGRAPHS of nucleus accumbens neurons in animals exposed to nonaddictive drugs display dendritic branches with normal numbers of signal-receiving projections called spines (*left* and *center*). But those who become addicted to cocaine sprout additional spines on the branches, which consequently look bushier (*right*). Presumably, such remodeling makes neurons more sensitive to signals from the VTA and elsewhere and thus contributes to drug sensitivity. Recent findings suggest that delta FosB plays a part in spine growth.

nerve cells. When drugs of abuse are administered, dopamine concentrations in the nucleus accumbens rise, inducing dopamine-responsive cells to increase production of a small signaling molecule, cyclic AMP (cAMP), which in turn activates CREB. After CREB is switched on, it binds to a specific set of genes, triggering production of the proteins those genes encode.

Chronic drug use causes sustained activation of CREB, which enhances expression of its target genes, some of which code for proteins that then dampen the reward circuitry. For example, CREB controls the production of dynorphin, a natural molecule with opiumlike effects. Dynorphin is synthesized by a subset of neurons in the nucleus accumbens that loop back and inhibit neurons in the VTA. Induction of dynorphin by CREB thereby stifles the brain's reward circuitry, inducing tolerance by making the same-old dose of drug less rewarding. The increase in dynorphin also contributes to dependence, as its inhibition of the reward pathway leaves the individual, in the drug's absence, depressed and unable to take pleasure in previously enjoyable activities.

But CREB is only a piece of the story. This transcription factor is switched off within days after drug use stops. So CREB cannot account for the longer-lasting grip that abused substances have on the brain—for the brain alterations that cause addicts to return to a substance even after years or decades of abstinence. Such relapse is driven to a large extent by sensitization, a phenomenon whereby the effects of a drug are augmented.

Although it might sound counterintuitive, the same drug can evoke both tolerance and sensitization. Shortly after a hit, CREB activity is high and tolerance rules: for several days, the user would need increasing amounts of drug to goose the reward circuit. But if the addict abstains, CREB activity declines. At that point, tolerance wanes and sensitization sets in, kicking off the intense craving that underlies the compulsive drug-seeking behavior of addiction. A mere taste or a memory can draw the addict back. This relentless yearning persists even after long periods of abstention. To understand the roots of sensitization, we have to look for molecular changes that last longer than a few days. One candidate culprit is another transcription factor: delta FosB.

Road to Relapse

DELTA FOSB APPEARS to function very differently in addiction than CREB does. Studies of mice and rats indicate that in response to chronic drug abuse, delta FosB concentrations rise gradually and progressively in the nucleus accumbens and other brain regions. Moreover, because the protein is extraordinarily stable, it remains active in these nerve cells for weeks to months after drug administration, a persistence that would enable it to maintain changes in gene expression long after drug taking ceased.

Studies of mutant mice that produce excessive amounts of delta FosB in the nucleus accumbens show that prolonged induction of this molecule causes animals to become hypersensitive to drugs. These mice were highly prone to relapse after the drugs were withdrawn and later made available—a finding implying that delta FosB concentrations could well contribute to long-term increases in sensitivity in the reward pathways of humans. Interestingly, delta FosB is also produced in the nucleus accumbens in mice in response to repetitious nondrug rewards, such as excessive wheel running and sugar consumption. Hence, it might have a more general role in the development of compulsive behavior toward a wide range of rewarding stimuli.

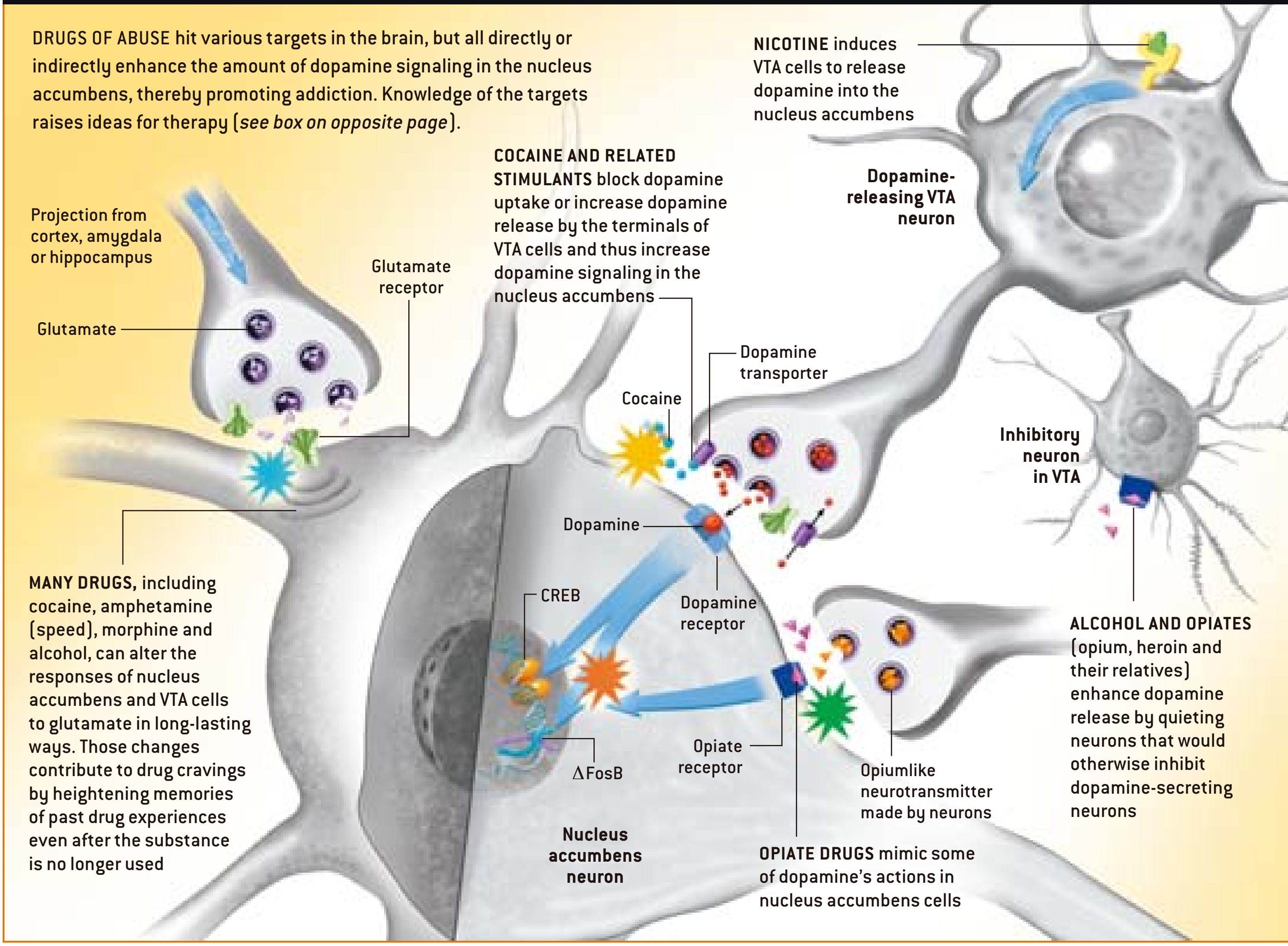
Recent evidence hints at a mechanism for how sensitization could persist even after delta FosB concentrations return to normal. Chronic exposure to cocaine and other drugs of abuse is known to induce the signal-receiving branches of nucleus accumbens neurons to sprout additional buds, termed dendritic spines, that bolster the cells' connections to other neurons. In rodents, this sprouting can continue for some months after drug taking ceases. This discovery suggests that delta FosB may be responsible for the added

THE AUTHORS

ERIC J. NESTLER and *ROBERT C. MALENKA* study the molecular basis of drug addiction. Nestler, professor in and chair of the department of psychiatry at the University of Texas Southwestern Medical Center at Dallas, was elected to the Institute of Medicine in 1998. Malenka, professor of psychiatry and behavioral sciences at the Stanford University School of Medicine, joined the faculty there after serving as director of the Center for the Neurobiology of Addiction at the University of California, San Francisco. With Steven E. Hyman, now at Harvard University, Nestler and Malenka wrote the textbook *Molecular Basis of Neuropharmacology* (McGraw-Hill, 2001).

DIFFERENT DRUGS, SAME ULTIMATE EFFECT

DRUGS OF ABUSE hit various targets in the brain, but all directly or indirectly enhance the amount of dopamine signaling in the nucleus accumbens, thereby promoting addiction. Knowledge of the targets raises ideas for therapy (*see box on opposite page*).



spines. Highly speculative extrapolation from these results raises the possibility that the extra connections generated by delta FosB activity amplify signaling between the linked cells for years and that such heightened signaling might cause the brain to overreact to drug-related cues. The dendritic changes may, in the end, be the key adaptation that accounts for the intransigence of addiction.

Learning Addiction

THUS FAR WE HAVE focused on drug-induced changes that relate to dopamine in the brain's reward system. Recall, however, that other brain regions—namely, the amygdala, hippocampus and frontal cortex—are involved in addiction and communicate back and forth with the VTA and the nucleus accumbens. All those regions talk to the reward pathway by releasing the neurotransmitter glutamate. When drugs of abuse increase dopamine release from the VTA into the nucleus accumbens, they also alter the responsiveness of the VTA and nucleus accumbens to glutamate for days. Animal experiments indicate that changes in sensitivity to glutamate in the reward pathway enhance both the release of dopamine from the VTA and responsiveness to dopamine in the nucleus accumbens, thereby promoting CREB and delta FosB activity and the unhappy effects of these molecules. Furthermore, it seems that this altered glutamate sensitivity strengthens the neuronal pathways that link memories of drug-taking experiences with high reward, thereby feeding the desire to seek the drug.

The mechanism by which drugs alter sensitivity to glutamate in neurons of the reward pathway is not yet known with certainty, but a working hypothesis can be formulated based on how glutamate affects neurons in the hippocampus. There certain types of short-term stimuli can enhance a cell's response to glutamate over many hours. The phenomenon, dubbed long-term potentiation, helps memories to form and appears to be mediated by the shuttling of certain glutamate-binding receptor proteins from intracellular stores, where they are not functional, to the nerve cell membrane, where they can respond to glutamate released into a synapse. Drugs of abuse influence the shuttling of glutamate receptors in the

TREATMENT POSSIBILITIES

Hypothetical anticocaine agent might reduce dopamine signaling in the nucleus accumbens by interfering with cocaine's ability to block dopamine uptake by VTA neuron terminals.

Hypothetical broad-spectrum agent would mute dopamine's effects by preventing CREB or ΔFosB from accumulating or from activating the target genes of these molecules.

Hypothetical broad-spectrum agent might interfere with the unhelpful changes in glutamate signaling that occur in nucleus accumbens cells with chronic drug use.

Opiate antagonists (such as naltrexone), already on the market, block opiate receptors. They are used against alcoholism and cigarette smoking because alcohol and nicotine trigger release of the brain's own opiumlike molecules.

reward pathway. Some findings suggest that they can also influence the synthesis of certain glutamate receptors.

Taken together, all the drug-induced changes in the reward circuit that we have discussed ultimately promote tolerance, dependence, craving, relapse and the complicated behaviors that accompany addiction. Many details remain mysterious, but we can say some things with assurance. During prolonged drug use, and shortly after use ceases, changes in the concentrations of cyclic AMP and the activity of CREB in neurons in the reward pathway predominate. These alterations cause tolerance and dependence, reducing sensitivity to the drug and rendering the addict depressed and lacking motivation. With more prolonged abstention, changes in delta FosB activity and glutamate signaling predominate. These actions seem to be the ones that draw an addict back for more—by increasing sensitivity to the drug's effects if it is used again after a lapse and by eliciting powerful responses to memories of past highs and to cues that bring those memories to mind.

The revisions in CREB, delta FosB and glutamate signaling are central to addiction, but they certainly are not the whole story. As research progresses, neuroscientists will surely uncover other important molecular and cellular adaptations in the reward circuit and in related brain areas that will illuminate the true nature of addiction.

A Common Cure?

BEYOND IMPROVING understanding of the biological basis of drug addiction, the discovery of these molecular alterations provides novel targets for the biochemical treatment of this disorder. And the need for fresh therapies is enormous. In addition to addiction's obvious physical and psychological damage, the condition is a leading cause of medical illness. Alcoholics are prone to cirrhosis of the liver, smokers are susceptible to lung cancer, and heroin addicts spread HIV when they share needles. Addiction's toll on health and productivity in the U.S. has been estimated at more than $300 billion a year, making it one of the most serious problems facing society. If the definition of addiction is broadened to encompass other forms of compulsive pathological behavior, such as overeating and gambling, the costs are far higher. Therapies that could correct aberrant, addictive reactions to rewarding stimuli—whether cocaine or cheesecake or the thrill of winning at blackjack—would provide an enormous benefit to society.

Today's treatments fail to cure most addicts. Some medications prevent the drug from getting to its target. These measures leave users with an "addicted brain" and intense drug craving. Other medical interventions mimic a drug's effects and thereby dampen craving long enough for an addict to kick the habit. These chemical substitutes, however, may merely replace one habit with another. And although nonmedical, rehabilitative treatments—such as the popular 12-step programs—help many people grapple with their addictions, participants still relapse at a high rate.

Armed with insight into the biology of addiction, researchers may one day be able to design medicines that counter or compensate for the long-term effects of drugs of abuse on reward regions in the brain. Compounds that interact specifically with the receptors that bind to glutamate or dopamine in the nucleus accumbens, or chemicals that prevent CREB or delta FosB from acting on their target genes in that area, could potentially loosen a drug's grip on an addict.

Furthermore, we need to learn to recognize those individuals who are most prone to addiction. Although psychological, social and environmental factors certainly are important, studies in susceptible families suggest that in humans about 50 percent of the risk for drug addiction is genetic. The particular genes involved have not yet been identified, but if susceptible individuals could be recognized early on, interventions could be targeted to this vulnerable population.

Because emotional and social factors operate in addiction, we cannot expect medications to fully treat the syndrome of addiction. But we can hope that future therapies will dampen the intense biological forces—the dependence, the cravings—that drive addiction and will thereby make psychosocial interventions more effective in helping to rebuild an addict's body and mind. SA

MORE TO EXPLORE

Incentive-Sensitization and Addiction. Terry E. Robinson and Kent C. Berridge in *Addiction,* Vol. 96, No. 1, pages 103–114; January 2001.

Molecular Basis of Long-Term Plasticity underlying Addiction. Eric J. Nestler in *Nature Reviews Neuroscience,* Vol. 2, No. 2, pages 119–128; February 2001.

Addiction: From Biology to Drug Policy. Second edition. A. Goldstein. Oxford University Press, 2001.

National Institute on Drug Abuse Information on Common Drugs of Abuse: **www.nida.nih.gov/DrugPages/**

The Threat of Silent Earthquakes

A LACK OF RUMBLING DOES NOT NECESSARILY MAKE AN EARTHQUAKE HARMLESS. SOME OF THE QUIET TYPES COULD PRESAGE DEVASTATING TSUNAMIS OR LARGER, GROUND-SHAKING SHOCKS

By Peter Cervelli

In early November 2000 the Big Island of Hawaii experienced its largest earthquake in more than a decade. Some 2,000 cubic kilometers of the southern slope of Kilauea volcano lurched toward the ocean, releasing the energy of a magnitude 5.7 shock. Part of that motion took place under an area where thousands of people stop every day to catch a glimpse of one of the island's most spectacular lava flows. Yet when the earthquake struck, no one noticed—not even seismologists.

How could such a notable event be overlooked? As it turns out, quaking is not an intrinsic part of all earthquakes. The event on Kilauea was one of the first unambiguous records of a so-called silent earthquake, a type of massive earth movement unknown to science until just a few years ago. Indeed, I would never have discovered this quake if my colleagues at the U.S. Geological Survey's Hawaiian Volcano Observatory had not already been using a network of sensitive instruments to monitor the volcano's activity. When I finally noticed that Kilauea's south flank had shifted 10 centimeters along an underground fault, I also saw that this movement had taken nearly 36 hours—a turtle's pace for an earthquake. In a typical tremor, opposite sides of the fault rocket past each other in a matter of seconds—quickly enough to create the seismic waves that cause the ground to rumble and shake.

But just because an earthquake happens slowly and quietly does not make it insignificant. My co-investigators and I realized immediately that Kilauea's silent earthquake could be a harbinger of disaster. If that same large body of rock and debris were to gain momentum and take the form of a gigantic landslide—separating itself from the rest of the volcano and sliding rapidly into the sea—the consequences would be devastating. The collapsing material would push seawater into towering tsunami waves that could threaten coastal cities along the entire Pacific Rim. Such catastrophic flank failure, as geologists call it, is a potential threat around many island volcanoes worldwide.

Unexpected Stir

FORTUNATELY, the discovery of silent earthquakes is revealing more good news than bad. The chances of catastrophic flank failure are slim, and the instruments that record silent earthquakes might make early warnings possible. New evidence for conditions that might trigger silent slip suggests bold strategies for preventing flank collapse. Occurrences of silent earthquakes are also being reported in areas where flank failure is not an issue. There silent earthquakes are inspiring ways to improve forecasts of their ground-shaking counterparts.

The discovery of silent earthquakes and their link to catastrophic flank collapse was a by-product of

GIANT LANDSLIDE (*upper left*) spawned by a silent earthquake could generate a fearsome tsunami hundreds of meters high (*below*).

efforts to study other potential natural hazards. Destructive earthquakes and volcanoes are a concern in Japan and the U.S. Pacific Northwest, where tectonic plates constantly plunge deep into the earth along what are called subduction zones. Beginning in the early 1990s, geologists began deploying large networks of continuously recording Global Positioning System (GPS) receivers in these regions and along the slopes of active volcanoes, such as Kilauea. By receiving signals from a constellation of more than 30 navigational satellites, these instruments can measure their own positions on the planet's surface at any given time to within a few millimeters.

The scientists who deployed these GPS receivers expected to see both the slow, relentless motion of the planet's shell of tectonic plates and the relatively quick movements that earthquakes and volcanoes trigger. It came as some surprise when these instruments detected small ground movements that were not associated with any known earthquake or eruption. When researchers plotted the ground movements on a map, the pattern that resulted very much resembled one characteristic of fault movement. In other words, all the GPS stations on one side of a given fault moved several centimeters in the same general direction. This pattern would have been no surprise if it had taken a year or longer to form. In that case, scientists would have known that a slow and steady process called fault creep was responsible. But at rates of up to centimeters a day, the mystery events were hundreds of times as fast as that. Beyond their relative speediness, these silent earthquakes shared another attribute with their noisy counterparts that distinguished them from fault creep: they are not steady processes but instead are discrete events that begin and end suddenly.

That sudden beginning, when it takes place on the slopes of a volcanic island, creates concern about a possible catastrophic flank event. Most typical earthquakes happen along faults that have built-in brakes: motion stops once the stress is relieved between the two chunks of earth that are trying to move past each other. But activity may not stop if gravity becomes the primary driver. In the worst-case scenario, the section of the volcano lying above the fault becomes so unstable that once slip starts, gravity pulls the entire mountainside downhill until it disintegrates into a pile of debris on the ocean floor.

The slopes of volcanoes such as Kilauea become steep and vulnerable to this kind of collapse when the lava from repeated eruptions builds them up more rapidly than they can erode away. Discovering the silent earthquake on Kilauea suggests that the volcano's south flank is on the move—perhaps on its way to eventual obliteration.

For now, friction along the fault is acting like an emergency brake. But gravity has won out in many other instances in the past. Scientists have long seen evidence of ancient collapses in sonar images of giant debris fields in the shallow waters surrounding volcanic islands around the world, including Majorca in the Mediterranean Sea and the Canary Islands in the Atlantic Ocean. In the Hawaiian Islands, geologists have found more than 25 individual collapses that have occurred over the past five million years—the blink of an eye in geologic time.

In a typical slide, the volume of material that enters the ocean is hundreds of times as great as the section of Mount St. Helens that blew apart during the 1980 eruption—more than enough to have triggered immense tsunamis. On the Hawaiian island of Lanai, for instance, geologists discovered evidence of wave action, including abundant marine shell fragments, at elevations of 325 meters. Gary M. McMurtry of the University of Hawaii at Manoa and his colleagues conclude that the most likely way the shells could have reached such a lofty location was within the waves of a tsunami that attained the astonishing height of 300 meters along some Hawaiian coastlines. Most of the tallest waves recorded in modern times were no more than one tenth that size.

Tsunami-generating VOLCANIC COLLAPSES may occur once every 10,000 years.

Overview/*Slippery Slope*

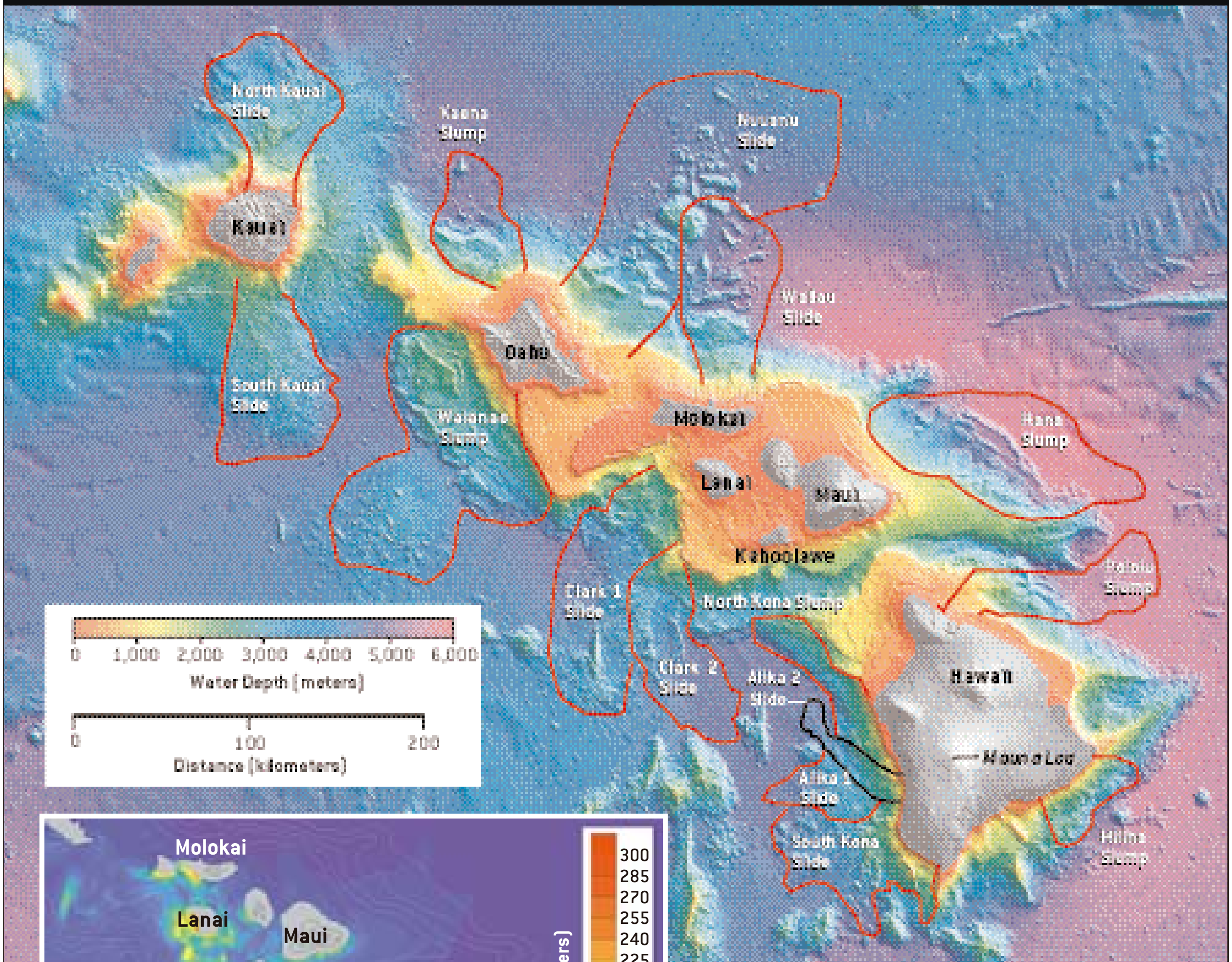
- Not all earthquakes shake the ground, it turns out. The so-called silent types are forcing scientists to rethink their understanding of the way quake-prone faults behave.
- In rare instances, silent earthquakes that occur along the flanks of seaside volcanoes may cascade into monstrous landslides that crash into the sea and trigger towering tsunamis.
- Silent earthquakes that take place within fault zones created by one tectonic plate diving under another may increase the chance of ground-shaking shocks.
- In other locations, however, silent slip may decrease the likelihood of destructive quakes, because they release stress along faults that might otherwise seem ready to snap.

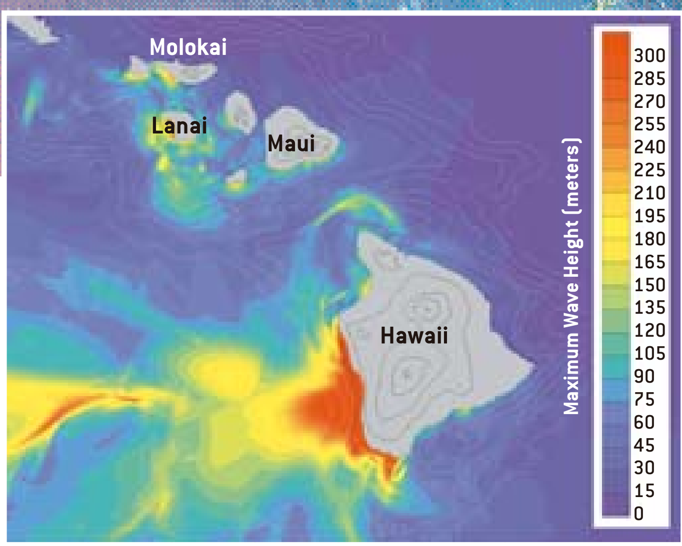
Preparing for the Worst

AS FRIGHTENING AS such an event may sound, this hazard must be understood in the proper context. Catastrophic failure of volcanic slopes is very rare on a human timescale—though far more common than the potential for a large asteroid or comet to have a damaging col-

DAVID FIERSTEIN (*preceding pages*)

GIANT LANDSLIDES AND TERRIFYING TSUNAMIS





UNDERWATER FIELDS of rock and debris (*examples outlined in red*) reveal that massive sections of Hawaiian volcanoes have crumbled into the ocean many times in the past. Some geologists suspect that one collapse (*black outline*) off the western flank of Mauna Loa volcano kicked up a gigantic tsunami that deposited shattered shell and rock as high as 800 meters along the nearby coast. A computer simulation (*left*) reveals that the same landslide may have produced waves up to 300 meters high. For a brief time, Maui could have been divided in two and the ocean floor west of Molokai could have been exposed.

USGS, WITH JAPAN MARINE SCIENCE AND TECHNOLOGY CENTER, UNIVERSITY OF HAWAII AND MONTEREY BAY AQUARIUM RESEARCH INSTITUTE (*top*); REPRINTED FROM GARY M. MCMURTRY IN *MARINE GEOLOGY*, VOL. 202, © 2002, WITH PERMISSION FROM ELSEVIER (*bottom*)

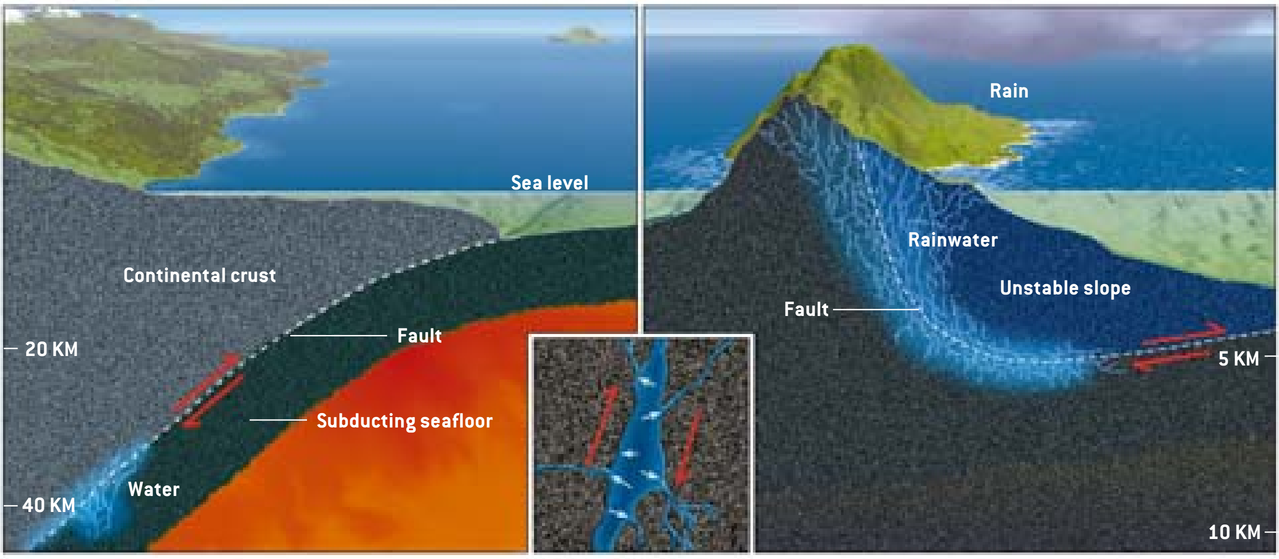
lision with the earth. Collapses large enough to generate a tsunami occur somewhere in the Hawaiian Islands only about once every 100,000 years. Some scientists estimate that such events occur worldwide once every 10,000 years. Because the hazard is extremely destructive when it does happen, many scientists agree that it is worth preparing for.

To detect deformation within unstable volcanic islands, networks of continuous GPS receivers are beginning to be deployed on Réunion Island in the Indian Ocean, on Fogo in the Cape Verde Islands, and throughout the Galápagos archipelago, among others. Kilauea's network of more than 20 GPS stations, for example, has already revealed that the volcano experiences creep, silent earthquakes as well as large, destructive typical earthquakes. Some scientists propose,

THE MECHANICS OF SILENT EARTHQUAKES

PERCOLATING WATER may trigger silent earthquakes if it finds a way into a vulnerable fault. Highly pressurized by the burden of overlying rock, water can push apart the two sides of the fault (*inset*), making it easier for them to slip past each other (*red arrows*). This kind of silent slip can occur within subduction zones and volcanic islands.

SUBDUCTION ZONE **VOLCANIC ISLAND**



WATER squeezed out of hydrous minerals in a slab of ancient seafloor may enter faults created as the slab dives underneath another tectonic plate.

WATER-FILLED fault

RAINWATER may seep down from the earth's surface into shallow faults, such as those that separate an unstable slope from the rest of a volcano.

DAVID FIERSTEIN

however, that Kilauea may currently be protected from catastrophic collapse by several underwater piles of mud and rock—probably debris from old flank collapses—that are buttressing its south flank. New discoveries about the way Kilauea is slipping can be easily generalized to other island volcanoes that may not have similar buttressing structures.

Whatever the specific circumstances for an island, the transition from silent slip to abrupt collapse would involve a sudden acceleration of the mobile slope. In the worst case, this acceleration would proceed immediately to breakneck velocities, leaving no chance for early detection and warning. In the best case, the acceleration would occur in fits and starts, in a cascade of silent earthquakes slowly escalating into regular earthquakes, and then on to catastrophe. A continuous GPS network could easily detect this fitful acceleration, well before ground-shaking earthquakes began to occur and, with luck, in plenty of time for a useful tsunami warning.

If the collapse were big enough, however, a few hours' or even days' warning might come as little comfort because it would be so difficult at that point to evacuate everyone. This problem raises the question of whether authorities might ever implement preventive measures. The problem of stabilizing the teetering flanks of oceanic volcanoes is solvable—in principle. In practice, however, the effort required would be immense. Consider simple brute force. If enough rock were removed from the upper reaches of an unstable volcanic flank, then the gravitational potential energy that is driving the system toward collapse would disappear for at least several hundred thousand years. A second possible method—lowering an unstable flank slowly through a series of small earthquakes—would be much cheaper but fraught with geologic unknowns and potential dangers. To do so, scientists could conceivably harness as a tool to prevent collapse the very thing that may be currently driving silent earthquakes on Kilauea.

Nine days before the most recent silent earthquake on Kilauea, a torrential rainstorm dropped nearly a meter of water on the volcano in less than 36 hours. Geologists have long known that water leaking into faults can trigger earthquakes, and nine days is about the same amount of time that they estimate it takes water to work its way down through cracks and pores in Kilauea's fractured basaltic rock to a depth of five kilometers—where the silent earthquake occurred. My colleagues and I suspect that

THE AUTHOR

PETER CERVELLI is a research geophysicist at the U.S. Geological Survey's Hawaiian Volcano Observatory, which sits along the rim of Kilauea Caldera on the Big Island. As leader of the observatory's crustal deformation project, Cervelli is responsible for interpreting data from a network of nearly 50 instruments that measure the tilt, strain and subtle movements within the island's two most active volcanoes, Mauna Loa and Kilauea. Cervelli discovered the silent earthquake that struck Kilauea's south flank in November 2000 while he was working on his Ph.D., which he received from Stanford University in 2001.

the burden of the overlying rock pressurized the rainwater, forcing the sides of the fault apart and making it much easier for them to slip past each other.

This discovery lends credence to the controversial idea of forcefully injecting water or steam into faults at the base of an unstable flank to trigger the stress-relieving earthquakes needed to let it down slowly. This kind of human-induced slip happens at very small scales all the time at geothermal plants and other locations where water is pumped into the earth. But when it comes to volcanoes, the extreme difficulty lies in putting the right amount of fluid in the right place so as not to inadvertently generate the very collapse that is meant to be avoided. Some geophysicists considered this strategy as a way to relieve stress along California's infamous San Andreas fault, but they ultimately abandoned the idea for fear that it would create more problems than it would solve.

Some SILENT EARTHQUAKES HAPPEN at such regular intervals that they can be predicted successfully.

Wedges of Water

APART FROM CALLING attention to the phenomenon of catastrophic collapse of the flank of a volcano, the discovery of silent earthquakes is forcing scientists to reconsider various aspects of fault motion—including seismic hazard assessments. In the U.S. Pacific Northwest, investigators have observed many silent earthquakes along the enormous Cascadia fault zone between the North American plate and the subducting Juan de Fuca plate. One curious feature of these silent earthquakes is that they happen at regular intervals—so regular, in fact, that scientists are now predicting their occurrence successfully.

This predictability most likely stems from the fact that water flowing from below subduction zones may exert significant control over when and where these faults slip silently. As the subducting plate sinks deeper into the earth, it encounters higher and higher temperatures and pressures, which release the significant amount of water trapped in water-rich minerals that exist within the slab. The silent earthquakes may then take place when a batch of fluid from the slab is working its way up—as the fluid passes, it will unclamp the fault zone a little bit, perhaps allowing some slow slip.

What is more, Garry Rogers and Herb Dragert of the Geological Survey of Canada reported last June that these silent tremors might even serve as precursors to some of the region's large, ground-shaking shocks. Because the slow slips occur deep and at discrete intervals, they regulate the rate at which stress accumulates on the shallower part of the fault zone, which moves in fits and starts. In this shallow, locked segment of the fault, it usually takes years or even centuries to amass the stress required to set off a major shock. Rogers and Dragert suggest, however, that silent slip may dramatically hasten this stress buildup, thereby increasing the risk of a regular earthquake in the weeks and months after a silent one.

Silent earthquakes are forcing scientists to rethink seismic forecasts in other parts of the world as well. Regions of Japan near several so-called seismic gaps—areas where fewer than expected regular earthquakes occur in an otherwise seismically active region—are thought to be overdue for a destructive shock. But if silent slip has been relieving stress along these faults without scientists realizing it, then the degree of danger may actually be less than they think. Likewise, if silent slip is discovered along faults that were considered inactive up to now, these structures will need careful evaluation to determine whether they are also capable of destructive earthquakes.

If future study reveals silent earthquakes to be a common feature of most large faults, then scientists will be forced to revisit long-held doctrines about all earthquakes. The observation of many different speeds of fault slip poses a real challenge to theorists trying to explain the faulting process with fundamental physical laws, for example. It is now believed that the number and sizes of observed earthquakes can be explained with a fairly simple friction law. But can this law also account for silent earthquakes? So far no definitive answer has been found, but research continues.

Silent earthquakes are only just beginning to enter the public lexicon. These subtle events portend an exponential increase in our understanding of the how and why of fault slip. The importance of deciphering fault slip is difficult to overstate because when faults slip quickly, they can cause immense damage, sometimes at a great distance from the source. The existence of silent earthquakes gives scientists a completely new angle on the slip process by permitting the detailed study of fault zones through every stage of their movement. SA

MORE TO EXPLORE

Sudden Aseismic Fault Slip on the South Flank of Kilauea Volcano, Hawaii. Peter Cervelli, Paul Segall, Kaj Johnson, Michael Lisowski and Asta Miklius in *Nature,* Vol. 415, pages 1014–1017; February 28, 2002.

Episodic Tremor and Slip on the Cascadia Subduction Zone: The Chatter of Silent Slip. Garry Rogers and Herb Dragert in *Science,* Vol. 300, pages 1942–1943; June 20, 2003.

Giant Landslides, Mega-Tsunamis, and Paleo-Sea Level in the Hawaiian Islands. G. M. McMurtry, P. Watts, G. J. Fryer, J. R. Smith and F. Imamura in *Marine Geology.* Available online at **www.sciencedirect.com/science/journal/00253227**

Visit the U.S. Geological Survey Hawaiian Volcano Observatory at **http://hvo.wr.usgs.gov**

The Fairest

All VOTING SYSTEMS have drawbacks. But by taking into rank candidates, one system gives the

By Partha Dasgupta and Eric Maskin

Most American and French citizens—indeed, those of democracies the world over—spend little time contemplating their voting systems. That preoccupation is usually left to political and electoral analysts. But in the past few years, a large segment of both these countries' populations have found themselves utterly perplexed. People in France wondered how a politician well outside the political mainstream made it to the final two-candidate runoff in the presidential election of 2002. In the U.S., many voters asked why the most popular candidate lost the election of 2000.

We will leave discussions of hanging chads, butterfly ballots, the electoral college and the U.S. Supreme Court to political commentators. But based on research by ourselves and colleagues, we can address a more fundamental issue: What kinds of systems, be they for electing national leaders or student council

Vote of all

account how voters
truest REFLECTION OF THE ELECTORATE'S VIEWS

presidents, go furthest toward truly representing the wishes of the voters? We argue that one particular system would be best in this sense—and it would be simple and practical to implement in the U.S., France and myriad other countries.

The Importance of Being Ranked

IN MOST NATIONAL presidential electoral systems, a voter chooses only his or her favorite candidate rather than ranking them all. If just two candidates compete, this limitation makes no difference. But with three or more candidates, it can matter a great deal.

The French presidential election of 2002 provides a case in point. In the first round, voters could vote for one of nine candidates, the most prominent being the incumbent Jacques Chirac of the Gaullist party, the Socialist leader Lionel Jospin and the National Front candidate Jean-Marie Le Pen. The rules dictated that if no candidate obtained an outright majority, the two candidates with the largest numbers of votes would face each other in a runoff. Chirac finished first (with 19.9 percent of the vote). The real surprise, however, lay in second place: the far-right-winger Le Pen took it (with 16.9 percent), while Jospin—who, with Chirac, had been heavily favored to reach the runoff—finished third (with 16.2 percent). In the second round, Chirac handily defeated Le Pen.

Despite Jospin's third-place finish, most available evidence suggests that in a one-to-one contest against Le Pen, he would have easily won. It is even plausible that he could have defeated Chirac had he made it to the second round. Yet by having voters submit only their top choice, the French electoral system cannot take account of such important information. Furthermore, it permits extremist candidates such as Le Pen—candidates who have no real chance of winning—to have an appreciable effect on the outcome.

The 2000 U.S. presidential election exposed similar shortcomings. To make this point most clearly, we will pretend that the election procedure was simpler than it actually was. We will consider just the four main candidates, and we will assume that there is no difference between the popular vote and the electoral college vote. (There have been many complaints about the electoral college, but even if it were replaced by popular vote, serious problems would remain.) We will also assume that there are only four kinds of voters: those who prefer Ralph Nader to Al Gore, Gore to George W. Bush, and Bush to Pat Buchanan (the "Nader" voters); those with the ranking Gore, Bush, Nader, Buchanan (the "Gore" voters); those with the ranking Bush, Buchanan, Gore, Nader (the "Bush" voters); and those with the ranking Buchanan, Bush, Gore, Nader (the "Buchanan" voters).

To be concrete, suppose that 2 percent of the electorate are Nader voters, 49 percent Gore voters, 48 percent Bush voters, and 1 percent Buchanan voters. If voters each choose one candidate, Gore will receive 49 percent and Bush 48 percent of the total (the actual percentages were 48.4 percent and 47.9 percent, respectively). Given that no candidate receives a majority (that is, more than 50 percent), how is the winner to be determined? Gore receives a plurality (the most votes short of 50 percent), so perhaps he should win.

On the other hand, the American Constitution stipulates that, absent a majority of the electoral votes, the House of Representatives should determine the winner. With a Republican majority in 2000, the House would presumably have gone for Bush. Clearly, having U.S. voters name solely their favorite candidate does not result in an outcome that is obviously right.

As in the French election, such ambiguity can be resolved by having voters submit complete rankings. Even though Gore is the favorite of only 49 percent of the electorate, the rankings show that a clear majority of 51 percent—the Gore and Nader voters combined—prefer Gore to either Bush or Buchanan. So Gore is the winner according to an electoral system called true majority rule (or simple majority rule), in which voters submit rankings of all the candidates and the winner is the one who beats each opponent in head-to-head competition based on these rankings.

True majority rule and rank-order voting result in DRAMATICALLY DIFFERENT outcomes.

Rankings can also be used in other electoral systems. Consider, for instance, "rank-order voting"—a procedure often used to elect committee officers that has been proposed to solve the problems inherent in the American and French presidential electoral systems. If four candidates are running, each voter assigns four points to his or her favorite, three to the next favorite, two to the next, and one to the least favorite. The winner is the candidate with the biggest total. The method appears to have been invented by Jean-Charles Borda, an 18th-century French engineer, and is sometimes known as the Borda count.

Imagine that 100 million people vote in the U.S. election. Based on our earlier assumptions, we know that 49 million of them will rank Gore first. So Gore will receive 196 million points—that is, 49 million times four points—from the Gore voters. The Nader voters place him second, so he picks up six million points from them. Finally, the Bush and Buchanan voters place him third, for an additional 98 million points. His grand total is 300 million points. If we make the corresponding computations for the others, we find that Nader gets 155 million points and Buchanan 199 million. Strikingly, Bush gets 346 million, even though a majority of the electorate prefer Gore [*see scenario A in box on opposite page*]. Only 2 percent of the electorate ranks Bush lower than second place, which is good enough to elect him under rank-order voting.

Thus, true majority rule and rank-order voting result in dramatically different outcomes. Considering this sharp contrast, it may seem hard to say which method is better at capturing the essence of voters' views. But we propose to do just that. We can evaluate these two systems—and any other—according to some

Overview/*Getting Voting Right*

- There is no such thing as a perfect voting system: every kind has one flaw or another.
- Nevertheless, one method could solve some of the problems that arose during recent elections in France and the U.S. Called true majority rule, this system incorporates information about the ranking of candidates, permitting a more accurate representation of voters' views.
- Our theoretical work shows that true majority rule more often avoids the flaws that arise for other voting methods. And, significantly, it could be easily implemented in countries the world over.

TOBY TALBOT *AP Photo;* PHOTOILLUSTRATION BY JOHNNY JOHNSON (*preceding pages*)

RANK-ORDER VOTING: SAMPLE SCENARIOS

IN THIS ELECTORAL SYSTEM, candidates are ranked and the corresponding points are tallied. Interestingly, even if a candidate were the true majority winner, as Gore is in scenario A, he or she would not necessarily win the rank-order vote. But a slight change in ranking, as happens with the Bush voters in scenario B, can make an enormous difference. In this case, it would lead to Gore winning.

Scenario A

CANDIDATE RANKING	POINTS ASSIGNED	VOTE TOTALS (in millions)
GORE VOTERS 49% (of 100 million votes)		
Gore	4	4 X 49 = 196
Bush	3	3 X 49 = 147
Nader	2	2 X 49 = 98
Buchanan	1	1 X 49 = 49
NADER VOTERS 2%		
Nader	4	4 X 2 = 8
Gore	3	3 X 2 = 6
Bush	2	2 X 2 = 4
Buchanan	1	1 X 2 = 2
BUSH VOTERS 48%		
Bush	4	4 X 48 = 192
Buchanan	3	3 X 48 = 144
Gore	2	2 X 48 = 96
Nader	1	1 X 48 = 48
BUCHANAN VOTERS 1%		
Buchanan	4	4 X 1 = 4
Bush	3	3 X 1 = 3
Gore	2	2 X 1 = 2
Nader	1	1 X 1 = 1
		Gore Total: 300
		Bush Total: 346

Scenario B

CANDIDATE RANKING	POINTS ASSIGNED	VOTE TOTALS (in millions)
BUSH VOTERS 48%		
Bush	4	4 X 48 = 192
Gore	3	3 X 48 = 144
Buchanan	2	2 X 48 = 96
Nader	1	1 X 48 = 48
		Gore Total: 348
		Bush Total: 346

fundamental principles that any electoral method should satisfy. Kenneth J. Arrow of Stanford University originated this axiomatic approach to voting theory in a 1951 monograph, a work that has profoundly shaped the voting literature.

Most voting analysts would agree that any good electoral method ought to satisfy several axioms. One is the consensus principle, often called the Pareto principle after Italian sociologist Vilfredo Pareto. It states that if everyone agrees that candidate A is better than B, then B will not be elected. This axiom does not help discriminate between true majority rule and rank-order voting, however, because both methods satisfy it—that is, both will end up with B losing. Moreover, the principle does not apply very often: in our U.S. election example, there is no unanimous preference for any one candidate over another.

Another important axiom holds that all voters should count equally—the "one-person, one-vote," or equal-treatment, principle. Voting theorists call it the principle of anonymity: who you are should not determine your influence on the election. True majority rule and rank-order voting also both satisfy anonymity.

A third criterion, however, does differentiate between the two. Neutrality, as this axiom is called, has two components. The first is symmetry, which means that the electoral rules should not favor one candidate over the other. The second requires that the voters' choice between candidates A and B should not depend on their views about some third candidate C. What would happen in our U.S. example if the Bush voters' ranking shifted to become Bush, Gore, Buchanan, Nader (instead of Bush, Buchanan, Gore, Nader)? From the standpoint of true majority rule, nothing important would change: the majority still prefer Gore to Bush. But look at what happens under rank-order voting: Gore now receives 348 million points, while Bush's total remains 346 million [*see scenario B in box at left*]. Gore now wins instead of Bush.

Obviously, rank-order voting can violate neutrality. Voters' preferences between Gore and Buchanan, a candidate who stands no chance of getting elected, determine the choice between Bush and Gore—and the outcome of the election. In contrast, true majority rule always satisfies neutrality. This last assertion may puzzle those readers who recall that in the actual election, discussion abounded about whether votes for Nader would affect the race between Bush and Gore. Indeed, in retrospect it appears that Nader—perhaps with help from the infamous butterfly ballot in Florida and even from Buchanan—may have siphoned off enough Gore votes to tip the election to Bush. But this effect was possible only because the U.S. election system is not actually true majority rule but its own unique system.

Majority Rule and the French Election

LET'S LOOK AT WHAT would happen to the French election of 2002 under true majority rule—which, for simplicity's sake, we will henceforth refer to as majority rule. Imagine Chirac, Jospin and Le Pen are the only candidates, and the electorate divides into three groups. Everyone in the first group, 30 percent of voters, has the ranking Jospin, Chirac, Le Pen. In the second group, 36 percent of the electorate, the ranking is Chirac, Jospin, Le Pen. In the remaining 34 percent, voters rank Le Pen over Jospin over Chirac. Chirac and Le Pen—with 36

THE AUTHORS

PARTHA DASGUPTA and *ERIC MASKIN* frequently collaborate in their research, including recent work on auction theory. Dasgupta is Frank Ramsey Professor of Economics at the University of Cambridge and past president of the Royal Economic Society. Maskin is Albert O. Hirschman Professor of Social Science at the Institute for Advanced Study in Princeton, N.J., and past president of the Econometric Society.

SLIM FILMS

FRENCH ELECTION OF 2002

CANDIDATE RANKING	PERCENTAGE OF VOTERS CHOOSING THIS RANKING
Jospin Chirac Le Pen	30
Chirac Jospin Le Pen	36
Le Pen Jospin Chirac	34


Jospin


Chirac


Le Pen

VOTERS' PREFERENCES BY PERCENT	
Prefer Jospin to Chirac	64
Prefer Jospin to Le Pen	66
Prefer Chirac to Jospin	36
Prefer Chirac to Le Pen	66
Prefer Le Pen to Chirac	34
Prefer Le Pen to Jospin	34

and 34 percent of the vote, respectively—would move forward into a runoff, where Chirac would easily prevail because 66 percent of voters prefer him to Le Pen.

The same outcome would result under yet another system, called instant-runoff voting (IRV), which is practiced in Ireland and Australia and which, like rank-order voting, has been advocated as an alternative to the French and U.S. systems. In IRV, simply put, rankings are used by election officials to successively eliminate the lowest-ranking candidates (and to incorporate their percentages into the voters' next-ranked choices) until only two candidates remain.

But the French and IRV systems conflict with majority rule. If you examine the configuration of voters' rankings, you see that Jospin actually commands an enormous majority: 64 percent of the electorate prefer him to Chirac, and 66 percent prefer him to Le Pen. Majority rule dictates that Jospin should win by a landslide [*see box above*].

When more than two choices present themselves, voters should SUBMIT A RANKING of candidates.

Recall that under majority rule a voter can make a political statement without harming the chances of any electable candidate. Someone who preferred Jospin to Chirac and knew that Le Pen had no chance of winning but wished to rank him first as a gesture of protest could do so without fear of knocking Jospin out of the race. (Except, of course, in the highly unlikely event that a majority of other voters made the same gesture.) The analogous point can be made about a voter who preferred Gore to Bush but wished to lend symbolic support to Nader.

Yet despite these virtues, majority rule has a flaw. It can violate another well-accepted voting principle: transitivity. Transitivity requires that if candidate A is chosen over B, and B is chosen over C, then A should be chosen over C. Now, ignoring Buchanan, pretend that 35 percent of the electorate prefer Gore to Bush to Nader, 33 percent rank Bush above Nader above Gore, and 32 percent go for Nader above Gore above Bush. Sixty-seven percent of voters rank Gore above Bush, 68 percent rank Bush above Nader, and 65 percent rank Nader above Gore. In other words, no matter which candidate is chosen, at least 65 percent of voters prefer somebody else! In this case, majority rule produces no winner.

This possibility, called the Condorcet paradox, was identified in the late 18th century by Marie-Jean-Antoine-Nicholas de Caritat, the Marquis de Condorcet, a colleague and arch-critic of Borda. The three rankings—Gore over Bush over Nader, Bush over Nader over Gore, and Nader over Gore over Bush—are collectively called a Condorcet cycle.

Our comparison of majority rule and rank-order voting appears to have resulted in a dead heat: majority rule satisfies every principle on our list except transitivity, and rank-order voting satisfies all but neutrality. This conundrum leads us to consider whether some other electoral system exists that satisfies all the principles. Arrow's celebrated impossibility theorem says no. It holds that any electoral method must sometimes violate at least one principle [see "Rational Collective Choice," by Douglas H. Blair and Robert A. Pollak; SCIENTIFIC AMERICAN, August 1983].

Beyond Impossibility

BUT ARROW'S THEOREM is unduly negative. It requires that an electoral method must satisfy a given axiom, no matter what voters' rankings turn out to be. Yet some rankings are quite unlikely. In particular, the Condorcet paradox—the bugaboo of majority rule—may not always be a serious problem in practice. After all, voters' rankings do not come out of thin air. They often derive from ideology.

To see what implications ideology holds for majority rule, think about each candidate's position on a spectrum ranging from the political left to the right. If we move from left to right, we presumably encounter the 2000 presidential candidates in the order Nader, Gore, Bush, Buchanan. And if ideology drives voters' views, then any voter who ranks Nader above Gore is likely to rank Gore above Bush and Bush above Buchanan. Similarly, any voter who ranks Bush above Gore can be anticipated to rank Gore above Nader. We would not expect to find a voter with the ranking Bush, Nader, Gore, Buchanan.

In a pioneering paper published in the 1940s, the late Duncan Black of the University College of North Wales showed that if voters' rankings are ideologically driven in the above manner—or at least if there are not too many nonideological voters—majority rule will satisfy transitivity. This discovery made possible a great deal of work in political science because, by posit-

JACK GUEZ *AFP/Corbis* (*Jospin*); DANIEL JANIN *AFP/Corbis* (*Chirac*); FRANÇOIS GUILLOT *Reuters NewMedia/Corbis* (*Le Pen*) ing ideological rankings of candidates on the part of voters, researchers could circumvent the Condorcet paradox and make clear predictions about the outcome of majority rule.

Of course, voters may not always conform to such a tidy left-right spectrum. But other situations also ensure transitivity. For another example, look again at the French election. Although Chirac and Jospin led the two major parties, it seems fair to say that they did not inspire much passion. It was the extremist candidate, Le Pen, who aroused people's repugnance or enthusiasm: evidence suggests that a huge majority of voters ranked him third or first among the three top candidates; few ranked him second. One can argue about whether such polarization is good or bad for France. But it is unquestionably good for majority rule. If voters agree that one candidate of three is not ranked second, transitivity is guaranteed. This property, called value restriction, was introduced in 1966 by Amartya Sen of Harvard University.

In our research on voting, we say that a voting system works well for a particular class of rankings if it satisfies the four axioms when all voters' rankings belong to that class. For instance, majority rule works well when all rankings are ideologically driven. It also works well when all rankings are "value restricted." Indeed, we have found that whenever any voting system works well, so does majority rule. Furthermore, majority rule works well in some cases in which other systems do not. We call this the majority dominance theorem.

To illustrate, we will imagine a three-way race between Gore, Bush and Nader. Suppose that every voter in fact ranks the candidates as either Gore, Bush, Nader or Bush, Gore, Nader. With voters' rankings belonging to this two-element class, rank-order voting satisfies its nemesis: the principle of neutrality (because voters' views on Nader do not affect whether Bush or Gore wins a rank-order election). Yet majority rule also works well here, because it satisfies its nemesis, transitivity.

But rank-order voting no longer works well if the situation becomes slightly more complicated. If we add a third ranking—Gore, Nader, Bush—majority rule is still transitive. These three rankings together do not constitute a Condorcet cycle. Rank-order voting, however, no longer satisfies neutrality. Suppose 51 percent rank Bush above Gore above Nader. If the remaining 49 percent rank Gore above Nader above Bush, Gore will win. If the remainder instead have the ranking Gore, Bush, Nader, however, then Bush wins—even though this group of 49 percent has the same ranking of Gore and Bush in either case.

BALLOT FROM SOUTH AFRICA'S first free elections in 1994. Sixty-two percent of the electorate chose Nelson Mandela and the African National Congress Party.

Majority rule still fails to work well sometimes, as the Condorcet paradox shows, though less often than other voting rules do. And in such cases, it has to be modified to identify a winner. There are many ways this can be done. Perhaps the simplest modification is as follows: If no one obtains a majority against all opponents, then among those candidates who defeat the most opponents in head-to-head comparisons, select as winner the one with the highest rank-order score.

Improving Future Elections

THE WAY most countries pick their presidents is faulty. Both the 2000 U.S. and 2002 French presidential elections were appreciably affected—perhaps decisively—by candidates who had no realistic chance of winning. These candidates were able to wield influence because, in each case, only a voter's top-ranked candidate was counted. We believe that when more than two choices present themselves, voters should submit a ranking of candidates and that majority rule—as we have discussed it—should determine the winner. Such a method would not be perfect; no method is. But as the majority dominance theorem shows, it would come closer to an accurate representation of the voters' wishes than any other system does. SA

MORE TO EXPLORE

Social Choice and Individual Values. Kenneth J. Arrow. Wiley, 1951. (Yale University Press, 1990.)

The Theory of Committees and Elections. Duncan Black. Cambridge University Press, 1958. (Kluwer Academic Publishers, 1998.)

Collective Choice and Social Welfare. Amartya Kumar Sen. Holden-Day, 1970. (North-Holland, 1984.)

On the Robustness of Majority Rule and Unanimity Rule. Partha Dasgupta and Eric Maskin. Available at **www.sss.ias.edu/papers/papers/econpapers.html**

QUARTZ WATCHES

Rock Clock

Step on a hunk of quartz in the dirt, and you'll think nothing of it. But carve one of the stone's crystals into a tiny tuning fork, and you'll have the key component for the watch ticking on your wrist.

Almost 90 percent of today's watches are electronic. Batteries provide the power to turn the hands or fire the liquid-crystal display, but quartz oscillators—essentially, vibrating tuning forks—provide the chronometers' steady beat. "Even a cheap quartz watch is accurate to one or two seconds a month," says Lou M. Galie, vice president of research and development at Timex in Middlebury, Conn. "Far more precise than expensive mechanical watches."

Since the Renaissance, interconnected gears and wheels—driven by pendulums, weights or springs wound tight by human hands—have turned the arms of clocks. By the early 1800s Swiss craftsmen were fabricating intricate wristwatches and founding companies that would dominate the trade for more than a century. Quartz clocks appeared for sale around 1940, and bulky watches tested the market in the 1960s, but most watchmakers saw the technology as a curiosity. A few Swiss firms improved the designs. Yet Japanese companies miniaturized the oscillator, battery, motor and circuitry and stormed the market in the 1970s. Traditional watchmakers took 20 years to recover and join the electronic movement.

Mechanical watches, finer than ever, are now limited to the high-price luxury market. Quartz watches owe their low cost to integrated circuits and their superior accuracy to the oscillator's high frequency of 32,768 vibrations a second. The spinning balance wheel that paces a mechanical watch typically rocks back and forth at about five beats a second.

Quartz oscillators were employed in the 1930s by military scientists to provide accurate timing for navigation equipment. Today Swatch and several Japanese firms supply most of the world's tuning fork oscillators, says Anton Bally, Swatch Group manufacturing president. They are mass-produced from artificial quartz using a photolithographic process devised by East German defector Juergen Staudte in 1968 at North American Aviation, now Rockwell.

—*Mark Fischetti*

DIGITAL quartz watch is paced by a crystal oscillator that vibrates when a battery applies voltage. The pulses, in turn, create a voltage that is fed back to the fork so it resonates at 32,768 beats a second. The beats time a microprocessor; it tells electrodes which shapes (numbers) to create on the liquid-crystal display. Electronics such as capacitors compensate for circuit feedback error.

KENT SNODGRASS *Precision Graphics*; SOURCES: SWATCH GROUP AND TIMEX

DID YOU KNOW . . .

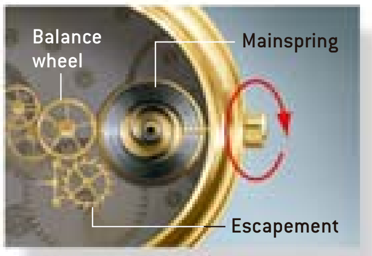
➤ NO BATTERY? Batteries power more than 95 percent of quartz watches. Early batteries lasted 18 months at best; now the finest can reach 10 years. Ironically, the current produced by the photovoltaic cells in most solar-powered watches is used to refresh a rechargeable battery hidden inside, although a few models store energy in a supercapacitor.

➤ GET WITH THE PROGRAM: The microprocessor in a $10 digital watch has more power than the processor in the Apple II that popularized personal computers around 1980. In that year, Timex employed no software engineers, says vice president Lou M. Galie, but today programmers make up more than half the engineering staff.

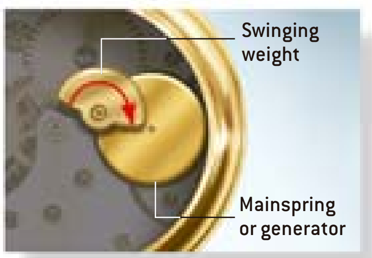
➤ COMPLICATED: Watchmakers call extra hardware functions such as chimes or date indicators “complications.” In 1783 Marie Antoinette’s lover commissioned watch pioneer Abraham Breguet to devise the most complicated watch ever, but it wasn’t completed until after the queen was beheaded. In 1927 auto magnate James Packard paid $2,500 for a watch that could show star positions from his Ohio home.

➤ QUARTZ, QUARTZ EVERYWHERE: Tiny quartz tuning forks provide precise reference frequencies for millions of computer chips, cellular phones, radio transmitters, satellite transceivers and music synthesizers. The shorter the tines, the higher the frequency. Music teachers might find them difficult to strike against a piano, however.

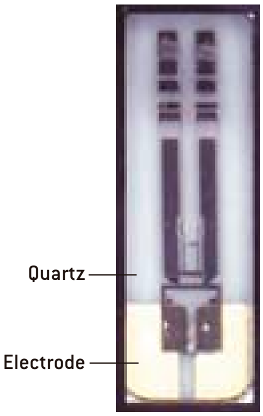
ANALOG quartz watch uses the same oscillator and feedback circuitry as a digital design, but integrated circuitry pares down the vibrations to two reliable beats a second. The beats instruct a stepper motor to turn on and off, which tells the second hand to start and stop each tick.



WINDUP mechanical watch is powered by turning the crown, which coils up the mainspring. Small quantities of the stored energy are released through an escapement. A rotating balance wheel provides the timing.



AUTOMATIC (also called kinetic or self-winding) watch is powered by an uneven weight that swings as the wearer’s arm moves, turning a generator in a quartz design or coiling the mainspring in a mechanical movement.



OSCILLATOR is manufactured in the shape of a tuning fork from man-made quartz using photolithography. Early oscillators were machine-cut from natural crystals. Because quartz is a piezoelectric material, it vibrates when voltage is applied to its gold-plated electrodes.

Topic suggested by reader Avraham Aharoni. Send ideas to workingknowledge@sciam.com

Passport in Time

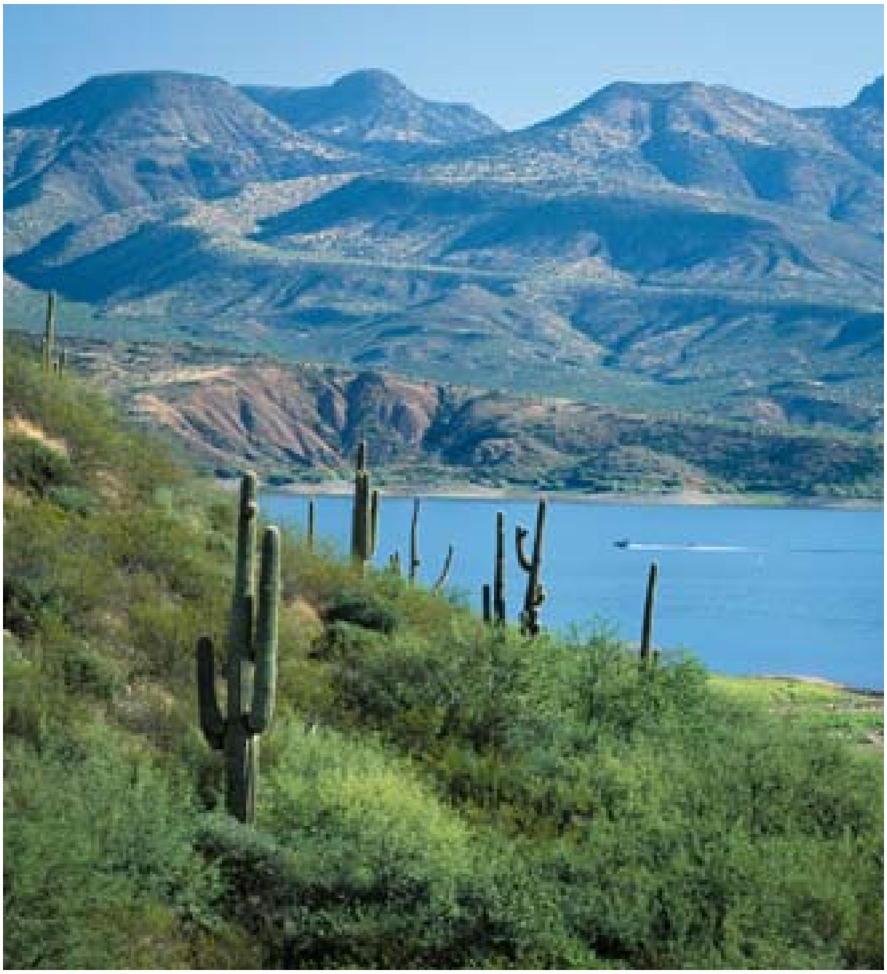
VOLUNTEERS JOIN ARCHAEOLOGICAL AND HISTORICAL FOREST SERVICE PROJECTS AROUND THE COUNTRY, LEARNING FIELD TECHNIQUES BY MARGUERITE HOLLOWAY

We stand 15 feet apart from one another until we form a long line that stretches through dry grass and around mesquite shrubs, and then we start walking through the scrubby Arizona desert under a midday sun, our eyes scanning the ground. Chris Reed, a Phoenix native who has done this before, calls out when he finds a circle of stones embedded in the earth, and the line breaks as all 14 of us cluster around for the official word.

"Very odd," says J. Scott Wood, chief archaeologist for the Tonto National Forest. "Record it as a feature." After scraping in the dirt for a while and lighting a cigar, Wood leans against his cane and begins to muse: it's a little big to be a storage pit, about right for a granary, but there is no compressed dirt floor. The Hohokam, Salado and other peoples who lived here between 850 and the late 1200s or so didn't use stone slabs as floors, he says, but still, the ground should be compacted. In this way, on an October day that reaches upward of 90 degrees Fahrenheit, an unusual weeklong field season begins.

Unusual, because those accompanying Wood and three other U.S. Forest Service archaeologists are not professionals—although some have had archaeological training—but rather volunteers, many of whom spend their vacations or retirement working alongside researchers in the field, surveying sites, making discoveries. They are all participants in Passport in Time (PIT), a U.S. Department of Agriculture program that began regionally in 1989 and went national in 1991. An average of 2,500 people a year join projects that include documenting petroglyphs on Kosciusko Island in Alaska, restoring old forest-fire lookouts in Washington State and excavating a sauropod in Colorado. (Visit www.passportintime.com for more information about PIT and how to apply.)

Since the program's inception, more than 12,800 people between the ages of 12 and 80 have worked on ventures in 38 states. Unlike similar projects in which individuals pay to do fieldwork—such as



ROOSEVELT LAKE in Tonto National Forest in Arizona stretches for about 25 miles. The U.S. Forest Service runs one or two Passport in Time projects in the forest every year.

DANNY LEHMAN *Corbis*

Earthwatch Institute, based in Maynard, Mass., or Crow Canyon Archaeological Center in Colorado—Passport in Time requires only that participants pay their way to the site and for their upkeep once there. Depending on the location of the study area, volunteers stay in motels, Forest Service quarters, campgrounds or their own RVs.

"Some people go from PIT to PIT," says Stephen Germick, an archaeologist at Tonto National Forest and its Passport in Time coordinator. "They leave in May and return home in September." Germick has come to rely on the expertise of these so-called PIT-heads, and he selects a mix of novices and returnees, many of whom have come to possess professional-level knowledge. One man in his 70s, for instance, has been working on another Tonto forest project—surveying the 1875 Silver King Mine—since 1996. "He is essential. He is our mapper," Germick says.

Mapping is an essential part of this October project, too, and those volunteers who have never done it before are initiated quickly. The site we walk during our first pedestrian survey is called the Armer Complex, after a ranching family that settled here in the 1870s. It was also the home of several Native American tribes for hundreds of years. As the largest site in the Tonto Basin and one that clearly experienced waves of settlement and trade, as well as a mysterious and sudden demise, it is important to understanding the prehistory of the area, Germick and Wood explain.

DROUGHT exposed ancient walls in September 2002. These ruins were covered again when rains replenished Roosevelt Lake.

U.S. FOREST SERVICE *AP Photo*

The site is often impossible to study. If Roosevelt Lake—a 25-mile-long reservoir formed in 1911 when the Roosevelt Dam was completed—is full, the ruins are submerged. But because of the severe drought the Southwest has been experiencing, parts of the Armer Complex have been exposed, and Germick has been using PIT volunteers to map remnant storage rooms, granaries, burial sites, housing assemblages, trash mounds and agricultural plots—some of them barely discernible circles or ovals or corners of embedded stones.

Participants also look for the many other records of various cultures. On our first day in the field, volunteers find all manner of pottery shards, and Wood analyzes the composition and pigment of several of them to determine their age and origin. They come from northeastern, east-central and southern Arizona, from the Anasazi, the Mogollon and the Hohokam. William Ramsey, a retired fire captain from California and frequent PIT-goer, discovers a lovely fragment: red with a series of black squares and lines—a music-bar design. Wood examines it, pointing out the difference between salt-glazed paint on one part and vegetable paint on another. He identifies it as Pinedale Black-on-red from between 1275 and 1350.

The expert eyes of others locate a metate (grinding slab) and hammer stones. Frances Mayse—on her 11th PIT and a steward for an archaeological site near her Tucson home—finds a beautiful mano (grinding stone). She then picks up what looks like a simple rock to me. “See the platform and the bulb of percussion? This would have been a scraper,” she explains. “I would have walked by this before. But now I know. You can never go out for a hike again without walking into the bushes.” It is true. We don’t look up at the mountains—the Sierra Ancha, Mazatzal and Apache—that frame this site. Our eyes only cast about at our feet for the past.

On the second day, after we have once again puttered and spluttered across the lake in a finicky boat provided by the nearby ranger station, Germick and Jennifer Berke, a Tonto forest ethnohistorian, set me to mapping a large compound. Germick puts pink flags in some of the corners and along walls and points out the dragon teeth—upright stones that were used as base supports for mud and poles. We stand in one corner and stretch measuring tape to the next. Once we have recorded the distance, we use the compass to figure out the direction of the far corner and then sketch the wall on graph paper, noting the right length in the right orientation. For a novice, it is not intuitive. But I have no choice but to get it and get it right: Germick and Berke are off measuring and discussing, setting flags, bantering and arguing about what this extensive, virtually erased compound could have been.

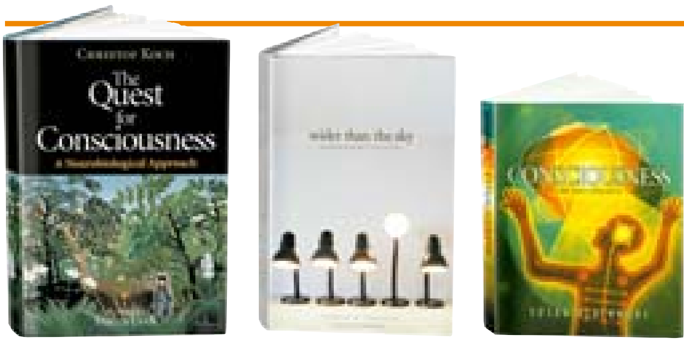
Soon I, too, can begin to see something of what they see. I see pit houses, then the later adobe compounds. I see plots of agave and traders coming in from over the mountains bringing different pottery and textiles. And I can see how PIT-heads are born. SA

NORTH SHORE of the lake is where the Armer Complex is found. Last October, PIT volunteers mapped part of the site.

MARGUERITE HOLLOWAY

The Major Unsolved Problem in Biology

THREE BOOKS TRY TO EXPLAIN CONSCIOUSNESS **BY MICHAEL SHERMER**



THE QUEST FOR CONSCIOUSNESS: A NEUROBIOLOGICAL APPROACH
by Christof Koch
Roberts & Company Publishers, Englewood, Colo., 2004 ($45)

WIDER THAN THE SKY: THE PHENOMENAL GIFT OF CONSCIOUSNESS
by Gerald M. Edelman
Yale University Press, New Haven, Conn., 2004 ($24)

CONSCIOUSNESS: AN INTRODUCTION
by Susan Blackmore
Oxford University Press, New York, 2004 (paperbound, $39.95)

When he was still a student, Richard Feynman hinted at a career to come as a scientific wonderer when he wrote: "I wonder why. I wonder why. / I wonder why I wonder / I wonder *why* I wonder why / I wonder why I wonder!"

Such wondering, and meta-wondering, takes us to the heart of what geneticist-cum-neuroscientist Francis Crick (who would know) calls "the major unsolved problem in biology"—explaining how billions of neurons swapping chemicals give rise to such subjective experiences as consciousness, self-awareness, and awareness that others are conscious and self-aware.

The body of literature attempting to solve this problem is extensive, and getting one's mind around the field is a herculean task successfully executed by psychologist Susan Blackmore in her delightful introduction, *Consciousness.* Presented as a textbook, it is so highly engaging that I recommend it for general readers, too. In many ways, the book is structured like a brain, with loads of independent modules (boxes and sidebars featuring profiles, concepts and activities) tied together by a flowing narrative and integrated into a conceptual whole.

The easy problem, Blackmore says, is explaining each of the functional parts of the brain, such as "the discrimination of stimuli, focusing of attention, accessing and reporting mental states, deliberate control of behavior, or differences between waking and sleep." In contrast, the hard problem in consciousness studies "is *experience:* what it is like to *be* an organism, or to *be in* a given mental state."

Adding up all of the solved easy problems does not equal a solution to the hard problem. Something else is going on in private subjective experiences—called qualia—and there is no consensus on what it is. Dualists hold that qualia are separate from physical objects in the world and that mind is more than brain. Materialists contend that qualia are ultimately explicable through the activities of neurons and that mind and brain are one. Blackmore, uniquely qualified to assess all comers (she sports multihued hair, is a devotee of meditation, and studies altered states of consciousness), allows the myriad theorists to make their case (including her own meme-centered theory) so that you can be the judge.

Making a strong case for the materialist position is Gerald M. Edelman's latest contribution, *Wider Than the Sky,* offered as a "concise and understandable" explanation of consciousness "to the general reader." Concise it is, but as for understandable, Edelman understates: "It will certainly require a concentrated effort on the part of the reader."

As director of the Neurosciences Institute in La Jolla, Calif., a Nobel laureate and author of several books on consciousness (*Neural Darwinism, The Remembered Present* and *Bright Air, Brilliant Fire*), Edelman has impeccable credentials. But science writing for a general audience involves more than expunging scholarly references and providing a glossary of technical terms as a substitute for clear exposition. To wit, on memory Edelman writes that "it is more fruitfully looked on as a property of degenerate nonlinear interactions in a multidimensional network of neuronal groups." Such prose is common throughout the book, which is a shame because Edelman is a luminously entertaining conversationalist, and his theory that the brain develops in a Darwinian fashion of neuronal variation and selection, and

REVIEWS

that consciousness is an emergent property of increasingly complex and integrated neuronal groups, has considerable support from neuroscience research.

An ideal combination of exquisite prose and rigorous science can be found in California Institute of Technology neuroscientist Christof Koch's *The Quest for Consciousness*. A rock climber adorned with a tattoo of the Apple Computer logo on his arm, Koch takes an unabashed neurobiological approach, the natural extension of what his longtime collaborator Francis Crick started in 1994 when he wrote in *The Astonishing Hypothesis* "that 'you,' your joys and your sorrows, your memories and your ambitions, your sense of personal identity and free will, are in fact no more than the behavior of a vast assembly of nerve cells and their associated molecules." To me, the most astonishing aspect of this theory is that it is astonishing to anyone. Where else could the mind be but in the brain?

Nevertheless, finding the neuronal correlates of consciousness (NCC) has proved elusive, so instead of concocting a grand unified theory, Koch and Crick un-

THE EDITORS **RECOMMEND**

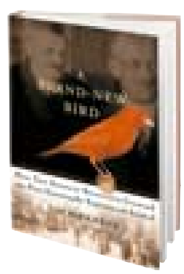
LONELY PLANETS: THE NATURAL PHILOSOPHY OF ALIEN LIFE

by David Grinspoon. Ecco, New York, 2003 ($39.95)

As he tells engagingly the story of humankind's long fascination with the possibility of extraterrestrial life, Grinspoon ponders the impact of a first contact in the form of a radio message from an intelligent civilization. "It might be frightening, liberating, uplifting, disturbing, or all of the above, but I say, 'Bring it on.' " And what if the first form of extraterrestrial life to be discovered turns out to be microbes? It "would enlarge our kingdom." Grinspoon, principal scientist in the department of space studies at the Southwest Research Institute and adjunct professor of astrophysical and planetary sciences at the University of Colorado at Boulder, concludes with his own belief: "I think our galaxy is full of species who have crawled up from the slime of their home worlds, evolved self-awareness and started to tinker, passed beyond the threat of technological self-extermination, and transcended their animal origins to move out into the cosmos."

A BRAND-NEW BIRD: HOW TWO AMATEUR SCIENTISTS CREATED THE FIRST GENETICALLY ENGINEERED ANIMAL

by Tim Birkhead. Basic Books, New York, 2003 ($26)



The brand-new bird is the red canary. It was the object of a quest that two Germans—Hans Duncker, a high school teacher interested in genetics, and Karl Reich, a bird keeper—carried on in Bremen for many years, beginning in 1921. Duncker's idea was to pluck the genes from a red siskin (a relative of the canary) and insert them into the yellow canary. His method was cross-breeding. The effort fell short of the goal, producing canaries of a reddish coppery hue. But it led to success years later by others who recognized the subtle connection between genes and the environment, in this case a diet containing carotenoids. Birkhead, professor of evolutionary biology at the University of Sheffield in England, makes a grand story by weaving in lore about genetics, bird keeping and the people involved in the quest.

All the books reviewed are available for purchase through www.sciam.com dertook a very specific research program focusing on the visual system, to understand precisely how photons of light striking your retina become fully integrated visual experiences. Koch and his colleagues, for example, discovered a single neuron that fires only when the subject sees an image of President Bill Clinton. If this neuron died, would Clinton be impeached from the brain? No, because the visual representation of Clinton is distributed throughout several areas of the brain, in a hierarchical fashion, eventually branching down to this single neuron. The visual coding of any face involves several groups of neurons—one to identify the face, another to read its expression, a third to track its motion, and so on.

This hierarchy of data processing allows the brain to economize neural activity through the use of combinatorics: "Assume that two face neurons responded either not at all or by firing vigorously. Between them, they could represent four faces (one face is encoded by both cells not firing, the second one by firing activity in one and silence in the other, and so on). Ten neurons could encode 2^{10}, or about a thousand faces.... It has been calculated that less than one hundred neurons are sufficient to distinguish one out of thousands of faces in a robust manner. Considering that there are around 100,000 cells below a square millimeter of cortex, the potential representational capacity of any one cortical region is enormous."

Given that the brain has about 100 billion neurons, consciousness is most likely an emergent property of these hierarchical and combinatoric neuronal connections. How, precisely, the NCC produce qualia remains to be explained, but Koch's scientific approach, in my opinion, is the only one that will solve the hard problem. SA

Michael Shermer writes the Skeptic column for Scientific American *and is publisher of* Skeptic *(www.skeptic.com) and author of* The Science of Good and Evil *(Henry Holt and Company, 2004).*

Grid Speed

BY DENNIS E. SHASHA

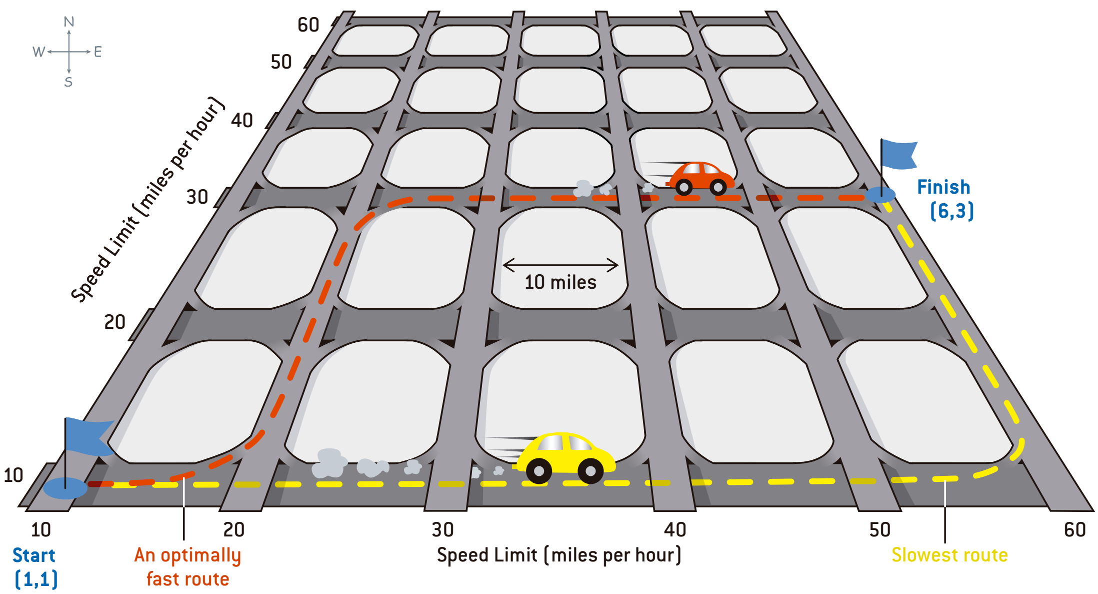
Consider a square grid in which six north-south streets, separated by gaps of 10 miles each, are elevated above six east-west streets laid out in a similar fashion [*see illustration below*]. Entrance and exit ramps connect the streets at every intersection. Because there are no traffic lights, switching from a north-south street to an east-west street (and vice versa) takes essentially zero time. The grid has very little traffic, but the local police patrol very carefully for speeders.

The speed limits follow an unusual pattern. The limit is 10 miles per hour for the southernmost east-west street, 20 miles per hour for the east-west street immediately to the north, and so on. (Therefore, the limit for the northernmost east-west street is 60 miles per hour.) Similarly, the speed limits for the north-south streets range from 10 miles per hour for the westernmost to 60 miles per hour for the easternmost. Let's label the intersections using their column and row numbers: the southwestern corner of the grid is (1,1), the southeastern corner is (6,1), the northwestern corner is (1,6), and so on. As a warm-up problem, can you determine the fastest legal way to get from (1,1) to (6,3)?

As it turns out, there are several optimally quick routes. One of them goes from (1,1) to (2,1), which takes one hour, then to (2,3), which takes another hour, and finally to (6,3), which takes an additional hour and 20 minutes. The slowest direct route (that is, covering the same distance as the quickest routes) goes from (1,1) to (6,1) in five hours and then to (6,3) in another 20 minutes.

Your challenge is to visit every intersection in as short a time as possible, starting from (1,1). How do you do it? And is there a better place to start in order to visit every intersection in a shorter time? My guess is no, but I would very much like to see a clever proof. (Conjectures can so easily turn out to be wrong.) SA

Dennis E. Shasha is professor of computer science at the Courant Institute of New York University.



CROSSTOWN TRAFFIC: The red route shows one answer to the warm-up problem. The yellow route shows the slowest path that covers the same distance. But what is the fastest way to visit every intersection?

Answer to Last Month's Puzzle

Encode your message as a number that will become the *x* coordinate of a point P in three-dimensional space. Choose two other coordinates randomly. Now select five planes that all intersect at point P and assign a different plane to each courier. Define each plane using the coordinates of three other (non-P) points in the plane.

Two nonparallel planes meet at a line, and any plane not containing that line but intersecting it will hit the line at a single point. Knowing the planes of any two couriers would give the enemy no useful information about point P, but any three couriers together could find the critical point readily and thus determine the *x* coordinate to decipher the message.

Web Solution

For a peek at the answer to this month's problem, visit www.sciam.com

CORNELIA BLIK

Regulation Rag

RED TAPE CAN BE A PAIN, UNTIL YOU ACTUALLY NEED TAPE **BY STEVE MIRSKY**

When I was a small boy, I was strongly in favor of reduced regulatory intrusion into private activity. I remember riding in the 1953 Dodge with my mother and asking her, "Wouldn't it be great if there were no stop signs?" She patiently explained to me that, no, it would be very bad because of all the carnage and chaos.

I recalled that incident recently when I received an e-mail entitled "If you are over 35, you should be dead." As I am in fact over 35, I decided to read it to clear up the mystery of why I was not dead. On opening the message, I thought (an act that provided axiomatic evidence of my being) that I recognized the contents as an antiregulatory diatribe I had already received a few times. And if I've gotten this note a few times, perhaps some *Scientific American* readers have received it as well. Therefore, I decided to respond here, because I don't have all your e-mail addresses.

First, some highlights of the e-mail: "According to today's regulators and bureaucrats, those of us who were kids in the 40's, 50's, 60's, or even maybe the early 70's probably shouldn't have survived. Our baby cribs were covered with bright colored lead-based paint. When we rode our bikes, we had no helmets. As children, we would ride in cars with no seatbelts or air bags. We fell out of trees, got cut and broke bones and teeth, and there were no lawsuits from these accidents. Please pass this on to others who have had the luck to grow up as kids, before lawyers and government regulated our lives, for our own good! People under 30 are WIMPS!"

I'm going to go out on a limb and assume that the target audience for this message is people who are alive. In data analysis, this is what's known as selection bias. Indeed, many kids didn't wear seat belts way back when. Some of them are now, in technical medical terminology, dead. The dead ones don't write such rants. Kids brain-damaged by lead or preventable blunt trauma may write, but they are probably not responsible for the above e-mail. Probably.

Still, life unfettered by bureaucratic interference remains tempting. And so I find myself musing on the good old days. I mean the really old days—30,000 years ago. Bureaucrats and lawyers didn't even exist yet. We were on our own and took responsibility for our actions. When we were attacked by a lion, we bled until we passed out and died. When we lost our teeth, we stopped eating and died. If we were painting a cave wall and scratched a finger that then got infected, we didn't complain to the Occupational Safety & Health Administration. And we didn't sue Og, who made the paint, or Oog, who chose the wall. We just waited for the infection to spread, and we keeled over. If we wanted to go somewhere, driven by mind-numbing desperation to find some scrap of decaying antelope before we starved to death, we went on our own two bare feet. If we lived to 35, we got stared at for being so elderly. And we talked about how kids were wimps because they wore animal skins on their feet.

Back in the present, all this talk about starvation has made me hungry. Ordinarily I might have a treat for which I first developed a taste when I was too young to appreciate stop signs: a burger. But on this early January day I have just learned that the U.S. Department of Agriculture has finally decided to ban the sale of "downer" cattle, animals too sick to walk, in the wake of the country's first case of mad cow disease. (Nearly 200,000 downers were slaughtered for sale last year.) The ban comes "only a few weeks after the department and allies in the powerful meat lobby blocked an identical measure in Congress," the *Washington Post* points out. The same article also quotes Representative Gary L. Ackerman of New York, who notes that the USDA "has seen the light, but that's only because they've been struck by lightning." Well, until I'm sure that the regulation banning downers is having the desired effect, perhaps this chicken will have turkey instead. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

Why are blood transfusions not rejected, as can happen with organs?

—K. Dahlke, Troy, Mich.

Robertson D. Davenport, associate professor of pathology and medical director of the Blood Bank and Transfusion Service at the University of Michigan at Ann Arbor, explains:

Blood transfusions from strangers are less likely to be rejected than organ transplants, for three key reasons. First, transfusions are given intravascularly, or into the circulatory system. Immune responses to antigens (foreign substances) received intravascularly tend to be less pronounced than responses from other routes. Second, transplanted organs contain immune cells that can stimulate the recipient's defenses, whereas most such cells in the blood are filtered out before transfusion. Third, the body replaces transfused red blood cells within three months, reducing the chances that they will be recognized as "alien," whereas transplants may remain in place for many years.

Rejection of blood is relatively rare. The risk of an acute hemolytic reaction is about one in 80,000 units transfused. To understand this occurrence, it helps to review some basic immunology. The two main types of immune responses are humoral and cellular. Humoral responses produce antibodies specific to a foreign antigen. These antibodies may attach to the antigen, forming immune complexes; the liver and spleen destroy the complexes. The complexes can also activate the complement pathway. Complement activation can punch holes in the membranes of bacteria or cells coated with antibodies. Immune responses elicited by blood transfusions are usually humoral. Organ transplants, in contrast, usually evoke cellular immune responses, which lead to the creation of cell-killing agents called cytotoxic lymphocytes.

The key antigens in blood transfusions are in the ABO system. Blood types can be A (having A but not B antigens on red cells), B (having B but not A), AB (both A and B), or O (neither). Virtually everyone over the age of six months has antibodies to the antigens they don't produce. A patient who receives blood with the wrong antigens can have a serious reaction, including breakdown of red cells and a strong inflammatory response that could lead to kidney failure or even death.

SARA CHEN

More common is a delayed hemolytic reaction, occurring in one in 5,000 units transfused. The antigens involved are usually not in the ABO system but in one of the other 25 known blood group systems, such as Rh (Rhesus). The antibodies produced tend not to activate complement, so the transfused cells are not usually broken down. Instead the spleen removes the cells, and a milder inflammatory response may occur days to weeks after the transfusion; sometimes this reaction can lead to renal failure.

How can deleted computer files be retrieved at a later date?

Clay Shields, professor of computer science at Georgetown University, offers this answer:

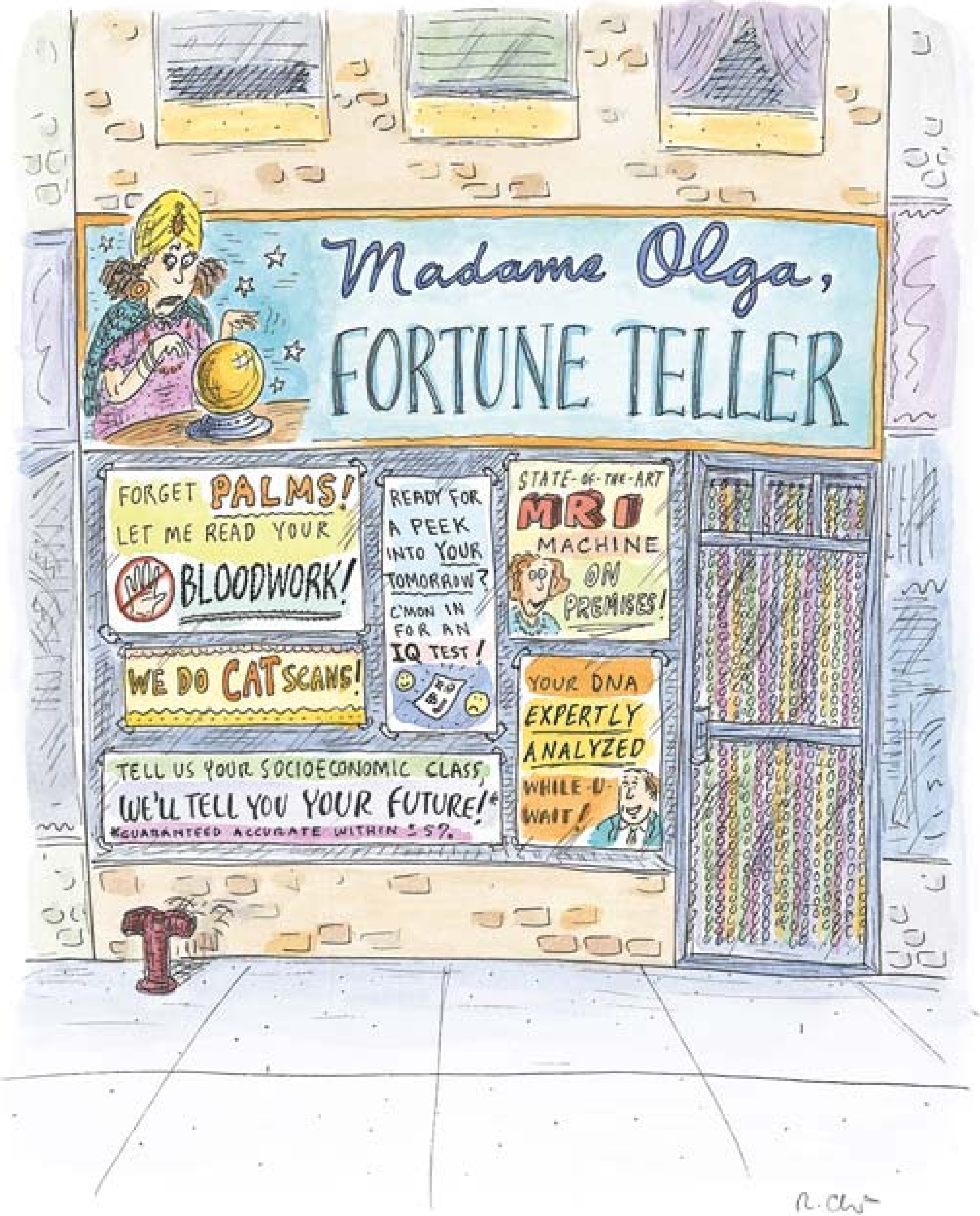
"Deleted" files can be restored because they aren't really gone—at least not right away. This is because it is faster and more efficient for computers to overwrite data only when necessary, when no other space is available to write new data.

A computer stores information in chunks called sectors. A file may be written across several sectors and might be scattered around the disk. The operating system keeps an index of which sectors belong to which files and a directory that maps the file names to the index entries.

When a user deletes a file, its directory entry is either removed or labeled as deleted. A deleted file can thus be salvaged if the index information and sectors have not yet been reused.

Such recovery is easy in operating systems that simply mark directory entries as deleted. A program scans the directory for deleted entries and presents a menu of files to recover. In other types of systems, recovery is more complicated. The directory entries may be lost, making it harder to find the file. The recovery program must look through all the index information and piece together files from various sectors. Because sectors may have been reused, only parts of the file may be accessible. SA

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


Madame Olga,
FORTUNE TELLER
FORGET PALMS! LET ME READ YOUR BLOODWORK!
WE DO CAT SCANS!
READY FOR A PEEK INTO YOUR TOMORROW? C'MON IN FOR AN IQ TEST!
STATE-OF-THE-ART MRI MACHINE ON PREMISES!
YOUR DNA EXPERTLY ANALYZED WHILE-U-WAIT!
TELL US YOUR SOCIOECONOMIC CLASS, WE'LL TELL YOU YOUR FUTURE!*
*GUARANTEED ACCURATE WITHIN ±5%
R. Chast