TEST TUBE BABIES AND CLONES • THE NEXT STEP FOR PHYSICS

SCIENTIFIC AMERICAN

Mad Cow–Type Plague Strikes Wild Deer

JUNE 2003 $4.95
WWW.SCIAM.COM

COMPUTER, HEAL THYSELF

AN END TO DISASTROUS CRASHES?

Martian Mysteries

How Chain Letters Evolve

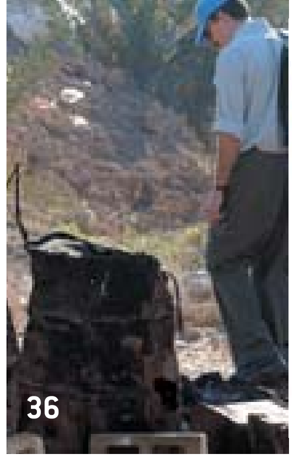
june 2003 contents

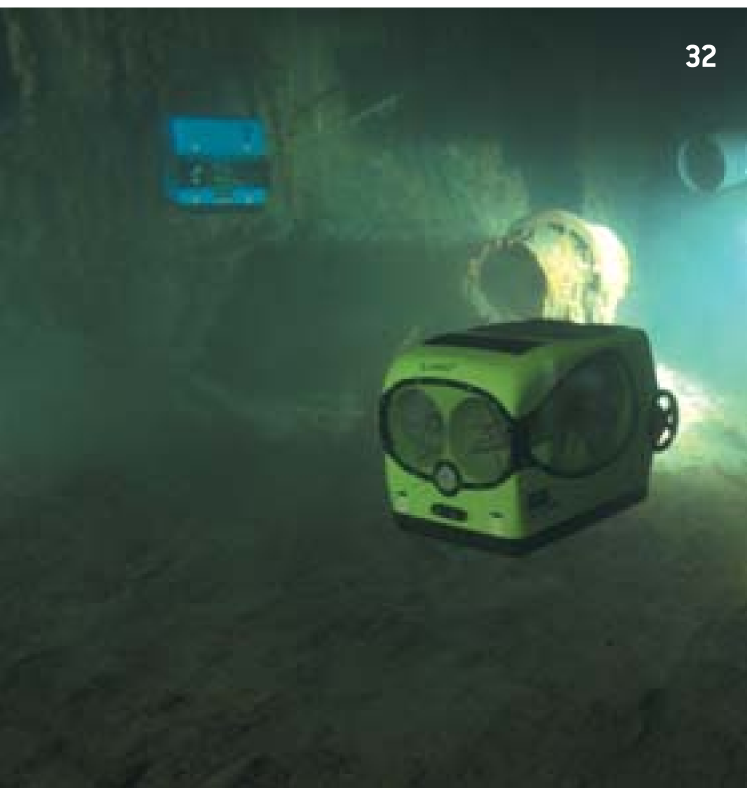
SCIENTIFIC AMERICAN Volume 288 Number 6

features

54 The key to crash recovery

departments





columns

Cover image by Kenn Brown.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2003 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 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.

A Pound of Flesh

In 2001 more than 6,000 people in the U.S. died while waiting for an organ transplant. The dire shortfall of organs compared with patient demand is growing as the population ages and more people experience organ failure. Although new immunosuppressive drugs have helped bridge the gap by allowing surgeons to transplant an organ that is a less than perfect match, there just aren't enough organs to go around.

The reasons vary. Some people have religious or cultural objections to organ donation; many families simply have a tough time making a decision at a time of personal tragedy. Living donors—those who volunteer a kidney or parts of their liver or lungs—are understandably reluctant: they must undergo potentially life-threatening surgery and put their own future health at risk.

The organ shortage has led various policymakers to propose radical steps. Several programs under consideration in the U.S. and elsewhere provide financial incentives to living donors or to the families of deceased donors. One approach, which has been instituted in Pennsylvania and is supported by the American Society of Transplant Surgeons, offers families who donate a loved one's organs $300 in food and lodging expenses. Editorials in medical journals advocating the program assert that the amount of money is intentionally small to "express appreciation" for the donation but not to serve as a payment. It is akin to the token coffee mug or umbrella one receives after donating to public radio or television.

Evidence that such programs will boost the organ supply is lacking, largely because of a paucity of studies. More important, some worry that these programs would mark the first step in encouraging an inhumane and subtly coercive market for spare body parts.

Although the outright purchase of organs is illegal in nearly every country in the world, a number have black markets for living-donor organs, and the results have been chilling. A study of 305 living kidney donors in Chennai, India, found that 96 percent sold a kidney to pay off debts, receiving about $1,070 apiece. But three fourths of the respondents soon faced debt and penury once again, and 79 percent would not recommend organ selling to others. Permitting trade in organs has already led to the exploitation of the poor.

This is an extreme example, but it illustrates the danger of attaching monetary value to whole organs. Society should redouble its support of less drastic steps to encourage families and to reduce the dangers to living donors. A host of bills now in Congress would create a "medal of honor" for donors, offer medical leave for living donors, or establish life and disability insurance for living donors in case they experienced negative side effects.

These initiatives could be paired with expanded public education campaigns that would explain the need for organ donation and demystify the process. Physicians and hospital personnel also require more training in encouraging organ donation. Many European countries either have implemented or are experimenting with "opt-out" laws, whereby the deceased is presumed to have consented to an organ donation unless he or she indicated otherwise. (Family members still have the final say.) These laws raise their own questions, but they bear watching.

Studies have shown that more than 95 percent of families would consent to organ donation if they knew it was the wish of their loved one. Appealing to people's better natures may not be the only way to raise the number of organs available for transplantation, but it is the best place to start.

THE EDITORS editors@sciam.com

VICTOR DE SCHWANBERG *Science Photo Library*

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

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

Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577

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

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

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

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

Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +49-8-442-7059

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

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

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

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

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

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

Safeguarding GPS

The space-based Global Positioning System (GPS) signal that guides "smart" bombs and cruise missiles to their targets is the underpinning of U.S. technological superiority on the battlefield. Yet because it is relatively easy to jam, the system is also the Achilles' heel of U.S. military might. Although the integrity of the signal was maintained in the war with Iraq, attempts to corrupt it underscored the need to protect GPS-dependent weapons and navigation systems. Against a more capable enemy, GPS might find itself among the first casualties of any new conflict.

NASA

Strung Out on the Universe: Interview with Raphael Bousso

The Holy Grail for many of today's theoretical physicists is a complete quantum-mechanical theory of gravity—useful for understanding the behavior of black holes, big bangs and entire universes. But bridging the gap between the smallest and largest constituents of reality will probably require a few totally new concepts (and shake our faith in some old ones). One researcher looking for these missing pieces is Raphael Bousso of Harvard University. The 31-year-old shared first prize in an international competition for young physicists last year for his work on the so-called holographic principle, which aims to reconcile quantum mechanics with black hole physics. His research has led him to think hard about string theory and cosmology, too.

Ask the Experts

What causes insomnia?

Henry Olders, who studies sleep patterns, depression and fatigue at McGill University, explains.

FREE Newsletters from ScientificAmerican.com

Delivering information to your in-box, when you want it and how you want it.

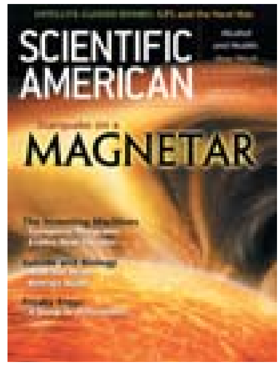
Subscribe today:

http://sciam.rsc03.net/servlet/campaignrespondent

Letters

EDITORS@SCIAM.COM

WHEN *SCIENTIFIC AMERICAN* runs an article that addresses evolution in any fashion, we can count on receiving spirited replies from all areas of the opinion spectrum. This is no less true when the subject is technological, rather than biological, in nature. "Evolving Inventions," by John R. Koza, Martin A. Keane and Matthew J. Streeter [February], which discussed a way to develop new devices with software, served as something of a Rorschach test for people's views. Some saw the authors' work as strongly supporting the Darwinian explanation, whereas others thought that it did *not* support the idea of natural evolution. These and additional reactions to the February issue appear below.



EVOLVING IDEAS

A conclusion one can draw from genetic programming is that this evolution is in its essence teleological ["Evolving Inventions," by John R. Koza, Martin A. Keane and Matthew J. Streeter]: the entire process is organized to realize a set of goals expressed in "high-level statements." If the authors are correct in claiming that their successful genetic programming emulates evolution in our world, then the natural processes must likewise be thought to operate successfully to achieve a conceptual goal. Perhaps modern biology is in need of a fundamental revision.

Ted Krasnicki
Richford, Vt.

Regarding "Evolving Inventions": Could there be any clearer evidence that intelligent designs can occur given raw materials and a selection process—without the need for an intelligent designer?

Wil Stark
Santa Rosa, Calif.

The article raises interesting questions about patent law. Any of the designs created by genetic programming would, by the standards applied by the U.S. patent office, be regarded as novel, and therefore patentable, had they been conceived in the ordinary way. But with this genetic programming machine, the obvious—and hence, by definition, nonpatentable—thing to do is to input your wish list for a widget and wait for the design to come out. Where is the inventive step? If such machines get common, patents could become a thing of the past. I am pleased that I have just retired as a patent attorney.

David L. McNeigh
Cheshire, England

That some machine may one day circumvent my livelihood as an inventor is disturbing. Why are we so determined to make ourselves obsolete? The only thing we have left is creativity. I beg you, please stop this research. I do not wish to have to make cheeseburgers to sustain myself.

Robert La Dellacruz
via e-mail

KOZA, KEANE AND STREETER REPLY: Genetic programming is patterned after natural evolution, but it is definitely not the same. Artificial evolution holds up an explicit goal in the hope of solving a particular problem by harnessing the problem-solving abilities of natural evolution. In nature, self-replicating entities evolve over time and acquire traits that enable them to survive and prosper in their environment (which also contains competing organisms and predators), but there is no prespecified final goal.

DRINK UP?

I enjoyed "Drink to Your Health?" by Arthur L. Klatsky, but I believe some important caveats are in order. First, observational studies, such as those quoted in support of the benefits to cardiovascular health of moderate alcohol drinking, are fraught with difficulties. Until recently, physicians advised postmenopausal women—based on observational studies—that hormone replacement therapy

SCIENTIFIC AMERICAN Help Desk

For your convenience, we have summarized the answers to often-asked customer service questions and how to get in touch with us.

Scientific American Frontiers
The magazine's PBS television series, Scientific American Frontiers, is hosted by actor and life long science buff, Alan Alda. There are now 10 one-hour episodes each season, with Frontiers becoming an integral part of a new PBS science programming initiative. Visit the Scientific American Frontiers Web site at www.pbs.org/saf

www.sciam.com
Our award-winning Internet resource, updated daily and thoroughly linked to related sites. Here you will find timely and interesting current news, recent illustrated scientific articles, current issue highlights, ask the editors, our on-line store, subscriptions, e-mail, other features not found in Scientific American, and much more.

Scientific American subscription rates
In the U.S., Scientific American is $34.97. In Canada, $49. Elsewhere, $55. Your 12-issue subscription includes the annual September single-topic issue.

Subscription inquiries
Call us: In the U.S. and Canada: (800) 333-1199. In other countries: (515) 247-7631. E-mail: subscriptions@sciam.com. Or write: Subscription Manager; Scientific American; PO Box 3187; Harlan, IA 51537. The date of the last issue of your subscription appears on the mailing label. For change of address, please notify us at least eight weeks in advance, and include both old and new addresses.

Back issues
In U.S. $10.95 each. Elsewhere $13.95 each. Many issues are available. Fax your order with your Visa, MasterCard or AMEX information to: (212) 355-0408.

Reprints
$4 each, minimum order 25 copies, prepaid. Limited availability. Write to: Reprint Department; Scientific American; 415 Madison Avenue; New York, NY 10017-1111; fax: (212) 355-0408 or tel. (212) 451-8877.

Photocopying rights
Granted by Scientific American, Inc., to libraries and others registered with the Copyright Clearance Center (CCC) to photocopy articles in this issue for the fee of $3.50 per copy of each article plus $0.50 per page. Such clearance does not extend to photocopying articles for promotion or other commercial purpose.

with estrogen would reduce their risks of cardiovascular disease. Now randomized, controlled trials demonstrate that such therapy actually increases the risk of coronary heart disease and stroke. As Klatsky notes, the question of alcohol and coronary health could be answered only by a randomized, controlled trial, a lengthy and probably impractical undertaking.

Second, observational studies are hampered by the low proportion of North American and European adults who do not drink—a proportion of these people have quit drinking because of previous alcohol-related problems, and their health outcomes cannot be extrapolated to the wider population. The reliability of observational studies can thus be questioned. In Scandinavia (with its higher proportion of alcohol abstainers), health outcome comparisons are less pronounced.

Third, as Klatsky points out, alcohol wreaks serious damage on individuals, communities and society. As a primary care physician, I regularly see patients whose lives have been ruined by excess alcohol. It behooves us to be extremely cautious about alcohol consumption for perceived cardiovascular benefits.

Steve Cottam
Great Eccleston
Health Center
Lancashire, England

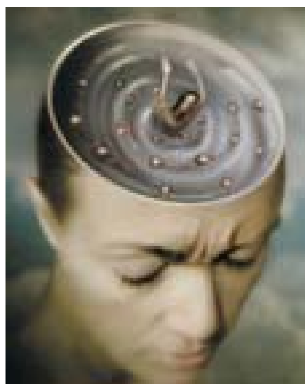
KLATSKY REPLIES: Cottam is right that observational data cannot completely rule out confounders for associations. Undoubtedly, a confounder of the observational association between hormone replacement therapy and cardiovascular disease was that women who chose such therapy because they believed it to be beneficial also had a generally healthy lifestyle. This situation was long suspected, and that fact influenced the decision to perform clinical trials. It is unlikely, though, that moderate drinkers were similarly motivated, because most reports of the inverse alcohol-coronary relationship predated any widespread knowledge of benefit, and drinking is not typically a prescribed treatment.

I cannot agree, however, with the implication that the alcohol-coronary data are inconsistent or unreliable. I'm not sure which Scandinavian studies are exceptions, but the Copenhagen Heart Study, for one, has shown strong evidence for protection conferred by moderate drinking. As Eric B. Rimm of the Harvard School of Public Health recently wrote: "Few other associations are so uniformly reported in the literature despite diverse population samples, varying exposure, and inconsistent control for confounding."

Finally, I emphatically agree that all considerations of benefit by moderate drinking need to be considered in light of the terrible toll of heavy uncontrolled intake.

CAUTION ABOUT ANTIDEPRESSANTS

I would like to mention a danger of antidepressants such as lithium that wasn't covered in "Why? The Neuroscience of Suicide," by Carol Ezzell. These drugs can cause a person with bipolar disorder to "overshoot," triggering a manic episode. It is suspected that a significant number of patients attempt suicide at the start of such an episode, as they come out of their depression. Among the newly approved mood stabilizers that don't have this problem are antiseizure medications originally used for epilepsy, including Depakote, Tegretol, Neurontin and Lamictal.



UNDERLYING neurobiological factors may increase suicide risk.

MATT MAHURIN

R. Tim Coslet
Sunnyvale, Calif.

MISSING THE TARGET?

Michael Puttré, in "Satellite-Guided Munitions," missed a major class of guided weapons: army missiles fired by the Multiple Launch Rocket System (MLRS). MLRS fires the Army Tactical Missile System, which has a range of up to 300 kilometers and was first used in Desert Storm.

The army missiles have a significant advantage over air-delivered ones, because pilots and expensive aircraft are not at risk, and reaction time to a call for support from the ground is considerably less. Last, the article should have mentioned dud rates, which are important for any weapon.

James B. Lincoln
Colonel, U.S. Army (retired)
Annandale, Va.

For the sake of national security, it is unwise to include detailed information about weapons, such as payload, range and accuracy. Even if this information is unclassified and readily available, I still question the need for it to be published.

Jeff Korpa
via e-mail

The term "smart" bomb is indeed an oxymoron. The $20,000 for a kit to outfit a $3,000 "dumb" bomb could send a student to college for a year or support an underprivileged American family. In the developing world, that money could build a clean-water well system for a village or provide vaccines for many people.

Nigel Mackenzie
Vancouver, B.C.

SCIENTISTS AND PSYCHICS

In "Psychic Drift" [Skeptic], Michael Shermer asks why most scientists do not believe in ESP and psi phenomena. An important factor must be the way their knowledge of the subject is in general limited to unscientific articles in the media, plus the very limited number of research papers and articles, mainly hostile in character, published in the major journals. Although the latter might appear to demonstrate that there is in essence no valid research in the area, in reality this situation is much more a reflection of negatively biased publication policies. Scientists are in a situation similar to that of citizens of countries where those in power have complete control over what they are allowed to read.

Brian D. Josephson
Department of Physics, Cavendish Laboratory
University of Cambridge

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Chemical Claw ▪ Horse Ancestor ▪ Chinese Cheese

JUNE 1953

TRUTH OR DAZE?—"Two lawyers and two psychiatrists on the Yale University faculty recently issued a joint warning against the use of 'truth serums' in criminal investigations. The psychiatrists cited clinical evidence to show that 'normal' subjects readily hide what they wish to hide when under the influence of one of these drugs (sodium amytal), and that 'neurotic' subjects frequently confess to deeds of which they are innocent. The statements elicited by drugs, they said, are more apt to be symbolically significant than objectively true."

CHEMICAL SCRUBBER—"Chelation is not a brand-new discovery, but there is now rising a flourishing industry which produces made-to-order chelate compounds for many purposes, from softening water to dissolving kidney stones. The various uses of the chelate compounds all depend on one fascinating property: the ability of the crablike claw to seize and sequester atoms of metal. Suppose that our water supply contains dissolved salts of iron. The iron forms a sediment on standing; it discolors bathtubs and linens; it spoils the taste of tea. On the domestic scale it is very difficult to remove. We may, instead, add a chemical called EDTA to the water. Now the iron will leave no stains. The iron is still there, yet it cannot be detected even by sensitive chemical tests. It is tightly imprisoned—'sequestered,' in the poetic language of chelation technology—by EDTA's chelate rings. The softening of water so far has been the largest use of chelation."

JUNE 1903

THE DAWN HORSE—"The Paleontological Department of the American Museum of Natural History, under the supervision of Prof. Henry F. Osborn, the curator, has recently prepared a remarkable exhibit depicting the ancestry and evolution of the horse. The blue-ribbon high-stepper of today is authentically traced back three million years or more. At this remote time he was about the size of a fox, only sixteen inches high, having four and five toes, with which he scampered over the marshes and shores of primeval earth. This noteworthy exhibit, the only one of its kind in America or elsewhere, has material from a special expedition for the search of fossil horses that was equipped and kept in the field for the past two seasons. A series of fine water-color paintings by Charles R. Knight [*see illustration*] complete the display."

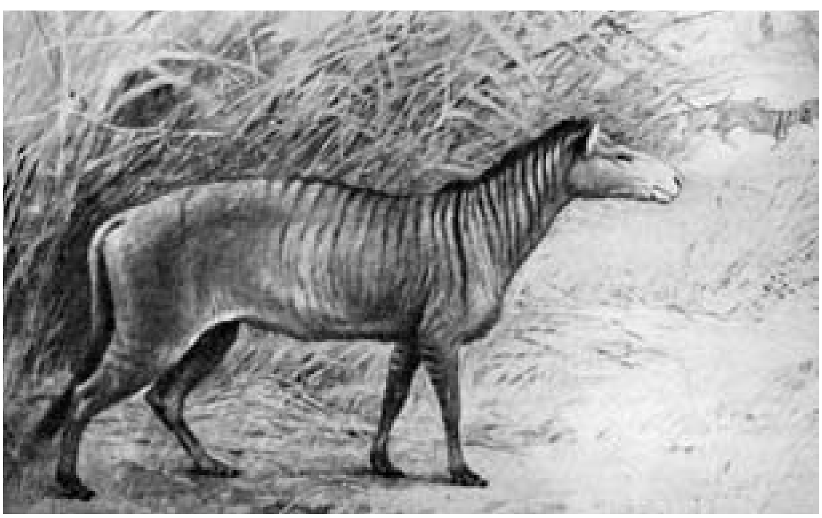
FROM MARVEL TO JUNK—"The Ferris wheel, one of the attractions of the Chicago Exposition of 1893, was recently sold at public auction for $1,800, engines, boilers, and all. Originally the contrivance cost $362,000. It is said there are about $300,000 worth of bonds outstanding against the owners of the wheel, as well as another $100,000 of debt."

JUNE 1853

FISH CORNUCOPIA—"The 'Sacramento Union' says of the Sacramento river: 'the water of the river must be alive with salmon, or such quantities caught daily would sensibly reduce their numbers. But experienced fishermen inform us, while the run lasts, that no matter how many are employed in the business, or how many are taken daily, no diminution can be perceived. Estimates give the number of men employed at about 600; the number of fish taken daily on average, 2,000, which would give as each man's catch a fraction over three a day."

THE NEWS ON TOFU—"The Chinese prepare an actual cheese—legumin cheese—from peas, called 'tao-foo,' which they sell in the streets of Canton. In preparing this cheese, the paste from steeped ground peas is boiled, which causes the starch to dissolve with the casein. After straining, the liquid is coagulated by a solution of gypsum. This coagulum is worked up like sour milk, salted, and pressed into moulds to make cheese."

SCIENTIFIC AMERICAN



THE EXTINCT EOHIPPUS, as depicted by Charles R. Knight, 1903 news SCAN

EMERGING DISEASES

Caught Off Guard

SARS REVEALS GAPS IN GLOBAL DISEASE DEFENSE **BY CHRISTINE SOARES**

"**The hospitals have been closed** and people are dying." A brief but chilling dispatch from the city of Guangzhou provided one of the outside world's first hints of the chaos in southern China's Guangdong Province as the mysterious disease now known as SARS spread unchecked. "When I got [the message], the province was already in disarray, with wholesale demonstrations in the streets," says retired U.S. Navy infectious disease investigator Stephen Cunnion, of his friend's report that he posted to ProMED-mail, an international infectious disease listserv.

Chinese officials have issued an extraordinary apology, effectively admitting that months of secrecy and denial after the new illness appeared last November created a case study in how *not* to handle an infectious disease outbreak. But in the end, China might have done the world a favor of sorts by providing a test of global readiness for an even more devastating future epidemic, whether naturally occurring or unleashed in an act of terrorism. With SARS (severe acute respiratory syndrome) having hit 22 countries by mid-April, world preparedness looks decidedly mixed.

"This was not the big one," says David Heymann, executive director for communicable diseases for the World Health Organization. His global alerts helped most countries to gird for SARS. But Heymann, whose group keeps a lookout for killer influenza strains that might emerge from the same region, admits that he is worried. "We've always had confidence in Hong

WORLDWIDE SPECTER of SARS leads flight attendants from Qatar Airways to don masks on arrival at Jakarta, Indonesia.

SUPRI *Reuters*

FAST FACTS: VIRAL DEATH

- Severe acute respiratory syndrome, or SARS, kills about 5 percent of its victims; another 10 to 15 percent survive only because of modern intensive-care practices. (Influenza typically has a 1 to 2 percent mortality rate.)

- The disease is caused by a new coronavirus—one of a family of large RNA viruses with glycoprotein "crowns"—that invades immune cells. In SARS, the resulting inflammation of lung tissue can lead to severe pneumonia and even hemorrhage. Two other coronaviruses cause about one third of common colds.

- As of April 15, the number of probable SARS cases had reached 3,235 worldwide—2,650 of them in Hong Kong and mainland China, 162 in Singapore, 100 in Canada and 35 in the U.S. Deaths worldwide: 154.

Up-to-date figures are at www.who.int/csr/sarscountry/en/

Kong," with its more modern and better-financed facilities, Heymann states, "whereas we didn't have confidence that China was prepared. Now I think we'll have to reassess Hong Kong."

The spread of SARS may have been no surprise to the Institute of Medicine (IOM), which in March quietly released a report, "Microbial Threats to Health: Emergence, Detection, and Response," an update to its startling 1992 analysis of gaping holes in U.S. defenses against a natural microbial assault. The IOM's latest analysis takes a global view and adds the risk of bioterrorism. Although the new study finds a few areas of improvement, it concludes that "the outlook is bleak on a number of fronts."

The report notes that global surveillance has improved—certainly none of the electronic systems that alerted the world to SARS existed in 1992—but experts judge it inadequate. Stephen Morse of Columbia University, an author of the original IOM report, believes that health surveillance is not comprehensive enough. "There's still a lot of fragmentation of knowledge. Nature isn't standing still; neither are potential terrorists." For instance, even with its knowledge of the severe pneumonia outbreak in Guangdong as early as February 9, WHO could do nothing without Chinese cooperation. "What we never have is the teeth to go in if a country refuses information," Heymann laments. A full month passed until WHO learned about pneumonia patients who had infected an unusually high number of hospital staff in Hanoi, Hong Kong and Singapore. By then it was too late for Hong Kong, where authorities were caught off guard by the ability of SARS to spread not only through close contact but also through contaminated surfaces and possibly sewage. With dozens of new cases daily—many of them hospital workers—city health officials admitted defeat in containment efforts by early April.

Preventing such a meltdown in medical infrastructure is the rationale behind the current U.S. program to vaccinate health workers against smallpox. But there are no vaccines against some of the other "Category A" bioagents, considered the most dangerous potential weapons, such as the plague, tularemia and Ebola pathogens. And no vaccine or drug is effective against the suspected SARS agent, a coronavirus so different from others in that diverse viral family, it earned its own group designation.

SARS is unlikely to have been a product of bioengineering. Terrorists today are more apt to use known pathogens than to invent one, Morse says, but "a few years down the road, those who are technically sophisticated will be able to do things that are more imaginative."

LUIS ASCUI *Getty*

SARS INFECTION is suspected because of the presence of fluid congestion, which appears on x-rays as diffuse white areas in the lungs, most noticeable between the ribs.

The U.S. research enterprise must become equally imaginative, the IOM report warns. Because the threat of bioterrorism is part of a continuum with naturally occurring disease, the authors urge a national "comprehensive infectious disease research agenda." As if to make their point, a virus-gene microarray originally conceived for infectious disease research provided the Centers for Disease Control and Prevention's first clue that a coronavirus could be the SARS culprit. (Animal studies, too, have now implicated the coronavirus.)

Armed with a recently sextupled biodefense budget, the National Institute of Allergy and Infectious Disease (NIAID) is aggressively recruiting researchers to develop novel antimicrobials and vaccines. The proposed Project BioShield, targeted to Category A pathogens, would entice biotechnology firms with a $6-billion pool and a guaranteed customer in the form of the federal government. NIAID is also inviting biotech companies to work on a SARS vaccine, for which the potential worldwide market is probably incentive enough. Like the international laboratory network that identified the SARS agent with unprecedented speed and cooperation, the endeavor could demonstrate what modern science can accomplish when seriously applied to combating microbial threats.

Heymann hopes the world does take a lesson from the SARS experience: "It's excellent practice for what *might* be coming. When you think of other diseases that have spread, like AIDS, it's going to be very important that when the next one comes, we do it even better—much better."

Christine Soares is based in New York City.

SPREADING THE WORD

The Internet helped the world to learn quickly about SARS:

February 9:

The World Health Organization first gets word of pneumonia cases in China's Guangdong Province through an e-mail from a former staffer's son living there.

A mention was picked up by the Web-crawling program of the Global Public Health Intelligence Network, a joint WHO/Canadian surveillance project created in 1998.

February 10:

A message by Stephen Cunnion, a retired U.S. Navy medical investigator, showed up on ProMED-mail, an international listserv launched in 1994. A terse notice from Hong Kong health officials stating their awareness of a pneumonia outbreak in mainland China accompanied it.

GLOBAL WARMING

Acting Locally

IN CURBING GREENHOUSE GAS EMISSIONS, STATES GO IT ALONE BY DAVID APPELL

Frustrated by federal inaction on preventing climate change, states and municipalities have begun reducing greenhouse gas emissions on their own. In fact, their influence could be greater than that of many countries that have ratified the Kyoto Protocol, the international agreement that set reductions of carbon emissions but that the U.S. has refused to ratify. In the process, the local-area policies are serving as incubators for new procedures and technologies that will be important to a coordinated national effort.

REPLACING SPORT UTILITY VEHICLES with fuel-efficient autos is one strategy of states trying to reduce carbon emissions.

KRT

"There's been a remarkable turn of events in the past two to four years," observes Susan Tierney of Lexecon, an economics consulting firm in Cambridge, Mass., and past assistant secretary for policy in the U.S. Department of Energy. Traditional first actors on air-quality issues, such as California, New Jersey and the New England states, have initiated programs to reduce emissions. States are motivated not only by the danger of climate change but by the hope of cleaner air, cost savings from energy efficiency, and marketing opportunities for renewable energy.

Such a "bottom-up" approach has a large global potential: "If they were considered as independent nations, U.S. states would comprise about 25 of the top 60 countries that emit greenhouse gases," remarks Barry Rabe of the University of Michigan at Ann Arbor, whose "Greenhouse and Statehouse," a Pew Center report, presents case studies of initiatives in nine states. Texas alone exceeds France in emissions.

Raab reveals a surprising range of situations among those states working to cut emissions. States moving ahead have been successful, he says, in couching the climate change as a more immediate problem, such as New Hampshire's concern over the possible loss of maple trees and the concomitant loss of tourism dollars from autumn's leaf peepers. Many states have a champion pushing the issue, such as Robert Shinn, former administrator of the Department of Environmental Protection in New Jersey. California's historic Pavley Bill of 2002, requiring strict limits on vehicle emissions in 2009, could serve to force redesigns of entire automobile fleets. Sixteen states now require utilities to purchase "green power." Texas, for instance, sells renewable-energy credits and has seen a sixfold increase in wind power generation between 1999 and 2002.

The six New England states (Connecticut, Maine, Massachusetts, New Hampshire, Rhode Island and Vermont) have banded together with five Canadian provinces (New Brunswick, Newfoundland and Labrador, Nova Scotia, Prince Edward Island and Quebec) to enact a Climate Change Action Plan. Written in 2001, the scheme aims to curb greenhouse emissions to 1990 levels by 2010 and then by an additional 10 percent by 2020. (Under the Kyoto Protocol, the U.S. would have had to reduce average emissions in 2008 through 2012 to 7 percent below 1990 levels.)

The first step calls for states to assess the amount of their greenhouse gas emissions; only 38 states have completed these inventories, which account for 87 percent of U.S. emissions. Then, to reduce emissions, planners are focusing initially on "low-hanging fruit," including replacing sport utility vehicles in state government fleets, acquiring more energy-efficient office equipment, and using light-emitting diodes for traffic lights. Seven activities in the region reported emissions reductions or sequestrations totaling 1.2 million metric tons of CO_2-equivalent (MMTCE).

Cities, too, are acting on their own. Thirty-one specific plans have been filed by 141 U.S. members of the International Council for Local Environmental Initiatives, representing 16 percent of U.S. emissions. Ten MMTCE of emissions have been eliminated, according to the council's Susan Ode, in which western cities such as San Diego, Port-

FAST FACTS: HOT AIR

Greenhouse gas emissions are calculated in millions of metric tons of CO_2-equivalent (MMTCE), a measure that adds together the climate warming potential of the different atmospheric greenhouse gases in units relative to that of carbon dioxide.

- Estimated U.S. greenhouse gas emissions, 2001: 1,883.3 MMTCE
- Emissions in 1990: 1,683 MMTCE
- Completed state action plans: 20
- Annual greenhouse gas reductions, 2000: 3.2 MMTCE
- Potential reductions by 2010: 71 MMTCE
- Potential by 2020: 96 MMTCE
- Estimated cost savings by 2010: $8 billion

land, Ore., and Salt Lake City are prominent.

Although individual states cannot replace a federal initiative, their patchwork regulatory approach could compel businesses to seek more consistent, predictable nationwide standards. States, however, often encounter the same reluctance that has dominated the national climate change scene. "We think, whether it's federal, state or local, they're ill-advised policies that are not going to help state or national economies and only succeed in putting more Americans out of work," says Darren McKinney of the National Association of Manufacturers, an industrial trade organization opposed to the Kyoto Protocol.

Still, the collective effort of the states is already beginning to compensate for the lack of reductions by the Bush administration. "You may have some American states that are better prepared, from a policy standpoint, to reduce greenhouse gases than a number of nations that have ratified Kyoto," Raab comments. The earth's atmosphere will take whatever help it can get.

David Appell lives in Ogunquit, Me.

ROCKETS

Hybrids Take Off

ENGINEERS RECONSIDER CROSS-BRED PROPULSION BY STEVEN ASHLEY

With little fanfare last December, Lockheed Martin Space Systems launched a suborbital sounding rocket from a NASA pad in Virginia. Forty-four miles over the Atlantic, the five-story-tall, two-foot-diameter craft released an 800-pound payload. The package, containing aerodynamic reentry experiments, was nothing particularly special. The booster itself, however, was rather exceptional—it was the first launch of a rocket powered by a large-scale hybrid rocket motor.

Such rockets attempt to combine the best of solid and liquid propulsion, the traditional engine types. In a liquid-fuel rocket, the fuel and oxidizer, often liquid hydrogen and oxygen, are stored separately and then mixed to create combustion. Liquid-fuel rocket motors burn efficiently, provide high thrust and, critically, can be throttled and even stopped and restarted. Such control permits planners to tailor the rocket's trajectory. Complexity, though, is high, and so tends to be the price tag.

POWERFUL PLUME blasts out of a hybrid rocket motor during a ground test conducted in 1999 at NASA's Stennis Space Center in Mississippi. An aerospace industry consortium developed the 250,000-pound thrust engine prototype as part of a $20-million program.

NASA

Simpler and cheaper are solid-fuel engines; their fiery impetus comes from burning premixed fuel and oxidizer grains that are packed like coffee grounds into a cylindrical casing. Unfortunately, the solid propellants—usually aluminum fuel and ammonium perchlorate oxidizer—burn fairly inefficiently, are toxic to the environment, and are difficult to fabricate and handle safely. A solid rocket cannot be throttled, either—once lit, it runs until the fuel is expended.

Hybrid propulsion offers significant advantages, claims Randy Tassin, a vice president at Lockheed Martin's Michoud Operations in New Orleans. "Hybrids are nonexplosive, can be throttled, are low cost and environmentally benign," he says. In addition, the compact power plant can produce nearly as much thrust as liquid-fuel motors. In a typical hybrid rocket motor, a rubbery fuel—a synthetic polymer called hydroxyl-terminated polybutadiene—cast into the tubular hull combusts fiercely when ignited in the presence of oxygen, pumped in from a separate tank as a liquid or a gas.

"The main difficulty in hybrid rocket technology is controlling the way the propellant burns," Tassin explains. The performance of hybrid fuels is not well understood, because the specific interaction of the semisolid fuel and oxygen is complex. "To get a high burn rate," he continues, "you need to add more surface area to the propellant, which results in relatively complicated fuel geometries." Some hybrid-fuel structures look like wagon wheels in cross section, with multiple oxidizer-injection ports set between "spokes" of fuel. During burning, the fuel segments get thinner and thinner, sometimes breaking off, which makes the motor run rough or even become unstable.

Researchers at the U.S. Air Force, aerospace firms and universities have worked for years to address the unresolved technical issues. From 1999 to 2002, NASA and an industry group spent about $20 million developing and ground-testing a hybrid rocket engine that generated 250,000 pounds of thrust. For the 60,000-pound-thrust hybrid motor that powered the Lockheed Martin sounding rocket, engineers configured the fuel to burn in a staged, and hence more stable, fashion than previous designs.

As a result of this slow but steady progress, hybrid motor technology is now coming under greater consideration for various missions, including targets for "Star Wars" antimissile systems and upper stages of larger boosters. NASA is contemplating using hybrids to propel the crew-escape capsule on a next-generation Orbital Space Plane. The inert fuel could be safely stored until an emergency, when onboard liquid oxygen could be pumped in, making the escape module ready to blast away to safety.

In the meantime, Lockheed Martin managers are focusing on suborbital sounding rockets for hybrid motors. NASA launches 25 to 30 sounding rockets annually; the U.S. military uses them as well. Says Tassin: "It's our belief that hybrids will eventually supplant solid rockets and even some liquid types in many future applications."

CANDLESTICK ROCKETS

Scientists at Stanford University are investigating paraffin wax as a potential fuel for hybrid rockets. Its high burn rate could produce thrust equivalent to that generated by the best liquid-fuel rockets. When researchers ignite their small demonstration motors, the resulting heat causes nearby wax to melt. As the liquefied paraffin mixes with injected oxygen, the surface area for combustion increases, thus yielding a burn rate triple that of other hybrid propellants.

SECURITY

Boxed Out

SCIENCE LOSES AS THE U.S. TIGHTENS VISA RULES **BY MADHUSREE MUKERJEE**

After the September 11 attacks, the clampdown on those from overseas wishing to study in the U.S. was inevitable. The Patriot Act of 2001 quickly implemented an electronic system for tracking foreign students, and officials are now extensively reviewing visa applications of scientists, engineers and students in technical fields. These and other ongoing efforts are creating a "viscous" visa system, notes William F. Brinkman, president of the American Physical Society (APS). Although such a system makes it harder for would-be terrorists to slip through, Brinkman maintains that it could hobble the U.S. economy and actually compromise national security.

The most visible effect of the visa restrictions may be on the highly international endeavors of physics. At Fermi National Accelerator Laboratory in Batavia, Ill., scientists from Vietnam, China, India and Russia, who all had supplied equipment for one of the detectors, were unable to arrive and operate it. A dozen scientists missed a September 2002 conference at Brookhaven National Laboratory. Vladimir B. Braginsky, a Russian who heads a research group at the California Institute of Technology, could not return in time for a meeting; Rashid A. Sunyaev, another Russian and director of the Max Planck Institute for Astrophysics in Garching, Germany, had to abandon his fall visit to Caltech. Many institutions are advising their foreign scientists to avoid leaving the U.S. (Similar delays are plaguing the biomedical field.)

American colleagues are frustrated that the U.S. Department of State, which in the past was responsive to their concerns, now seems to be turning a deaf ear. Kip S. Thorne of Caltech, who vainly sought updates on the Russian scientists' applications, likens the "visa bureaucracy" to a black hole: "You can't get any information out." The proposed Patriot Act II would reportedly classify background material on visas as confidential, which could make it impossible to fix a flawed application.

Senior scholars face delays, but students

PATRIOTIC RESTRICTIONS

The academic endeavors of foreign students who enter the U.S. to study may be limited because of Patriot Act II. The law would allow agencies to engage in surveillance without court consent to ensure that universities comply with the electronic registration system, which also controls a student's field of study. That could deter foreign scholars from working on any projects outside their stated field.

meet with denials. In 2002 the number of student visas granted was 234,322, down 20 percent from 2001. Stuart Patt, a spokesperson for the State Department, contends that this drop reflects an overall downturn of visa applications since 9/11. At the same time, an APS survey of physics departments learned that 13 percent of foreign students admitted (including 20 percent of those from China) were denied visas, with almost half of those from China facing some kind of difficulty. These problems left science faculty scrambling to fill teaching and research positions and threaten the viability of some small programs.

VISA RESTRICTIONS may also hinder American science.

C. SHERBURNE *PhotoLink/Getty*

Part of the reason for the increase in visa denials is that consular officials are being held personally responsible—and possibly criminally liable—if they grant a visa to someone who goes on to commit a terrorist act. Irving A. Lerch of the APS points out that consular officials may not be able to distinguish a benign field of study from a related but dangerous one and would deem it safer to deny visas to most applicants in certain broad categories, such as condensed-matter physics or biotechnology. (In practice, most students are rejected because they cannot prove an intent to return home after they complete their programs.) Even the secretary of state is not empowered to overturn a consular officer's denial of a visa. Yet sadly, the State Department's inspector general discovered in December 2002 that inadequately trained junior officials made most visa decisions and that they were too inconsistent in their background checks to foil a determined terrorist.

What may not make it through is U.S. preeminence in the physical sciences. In science and engineering fields, between 35 to 50 percent of doctoral degrees go to foreigners, many of whom stay: in physics, a third of the faculty is foreign born. U.S. science gathered momentum during World War II, thanks to the influx of trained Europeans. Several estimates attribute fully half the growth in the American economy since then to innovation in science and technology, with "aliens" having played no small part. Lerch fears that a significant, permanent reduction in the numbers of visas for scientists and engineers could cause a long-term downturn in the economy.

Outsiders have also contributed to defense: Albert Einstein and Enrico Fermi, whose ideas lay behind the atom bomb, were originally citizens of then enemy countries. But new regulations prevent foreigners from being employed on a host of "unclassified but sensitive" projects in academia and industry. "There are categories of people who can't work in certain categories of knowledge," notes the CEO of a security-related software firm, who requested anonymity. "There is difficulty getting talented people, across the board." Concern about staff qualifications at the national weapons labs was already running high after the Wen Ho Lee affair at Los Alamos, which promulgated perceptions of racial profiling that made even some U.S. citizens reluctant to apply for positions at the labs. The new restrictions are exacerbating the problem and, according to the trade group Information Technology Association of America, could undermine long-term security.

The most immediate concern, however, is the insensitive implementation of existing regulations. In January the U.S. Immigration and Naturalization Service arrested and detained Pakistani journalist Ejaz Haider for failing to report for fingerprinting. (All men from certain countries must register for background checks.) Haider, who had issued warnings about Islamic holy warriors long before 9/11, was a visiting scholar at the Brookings Institution, a Washington, D.C., think tank, and had apparently been assured by consular officials that he need not register. The affair made headlines in Pakistan. "Everyone here is surprised that the INS is not able to distinguish between friend and foe," comments A. H. Nayyar, a physicist at Quaid-e-Azam University in Pakistan. "This is very scary for friends."

Madhusree Mukerjee, who holds a Ph.D. in physics, lives in Montclair, N.J.

BARRIERS TO STUDY

According to Stuart Patt, a spokesperson for the U.S. Department of State, visas are being delayed by "interagency review": the department seeks information on the individual that other federal agencies might have. All visa applicants from so-called state sponsors of terrorism—Cuba, Libya, Iran, Iraq, North Korea, Sudan and Syria—and some applicants from nations that pose a risk of nuclear proliferation—China, India, Israel, Pakistan and Russia—face such scrutiny. Also, consular officials have to be alert to 16 kinds of potentially dangerous technologies (specific aspects of nuclear science, biotechnology, propulsion systems, lasers, robotics, materials science, advanced computation and others) that a scholar from any nation might seek to acquire.

PHYSICS

Law and Disorder

A QUANTUM STEAM ENGINE GETS AROUND THE SECOND LAW **BY JR MINKEL**

The fathers of thermodynamics got a lot of mileage from thought experiments about gas-filled engines conjoined to reservoirs of hot and cold. Today a few physicists are playing with quantum mechanics in hopes of finding new methods to control and create energy flow in quantum versions of the steam engine. Their research suggests that it is possible to "beat" the inviolable second law of thermodynamics with some quantum sleight-of-hand.

The second law limits the efficiency of any physical process. In essence, it states that, to perform work, energy must flow between two reservoirs set at different temperatures. The flow introduces disorder into the system. The temperature difference between the two baths determines the engine's maximum, or Carnot, efficiency, named after 19th-century French physicist Sadi Carnot.

Marlan O. Scully's quantum optics group at Texas A&M University has calculated a way to extract work from a single heat bath, thereby surmounting the Carnot limit and giving the appearance of breaking the second law. This setup would rely on photons rebounding in a small cavity between two mirrors, one of which would act as a piston. The bath is a circulating gas of atoms that emits heat in the form of photons as it passes through the mirrored cavity. The atoms are prepared in a special fashion. Each atom has three electron states: an excited state and two nearly identical relaxed states that are quantum-mechanically mixed. This so-called coherence interferes with the absorption of photons but permits the emission of photons to proceed unfettered, causing an excess of heat beyond what the temperature of the atoms alone would dictate.

Coherence is a unique property of quantum mechanics that allows laser photons to march in lockstep and atoms to be in two states at once. Thermodynamically speaking, coherence is an extra dose of order, which pushes the engine out of the uniform state of equilibrium, where the second law applies. An incoherent photon gas, in contrast, pumps a piston with the usual Carnot efficiency, as shown in a study by physicist M. Howard Lee of the University of Georgia.

Scully states that his team's analysis demonstrates effects that classical heat engines cannot produce—a tiny bit of coherence can cause a significant boost in work output. Coherence is as fragile as a house of cards, however, and building it up in this case costs three or four times as much energy, in the form of microwaves, as the engine puts out. These handicaps may mean that the effect probably will not readily find applications. Nevertheless, Seth Lloyd of the Massachusetts Institute of Technology comments that Scully's investigation may help point the way toward using coherence to bring lasers or thermoelectric refrigerators closer to their ideal efficiency limits. "If quantum coherence could do that for you, that would be great," Lloyd notes.

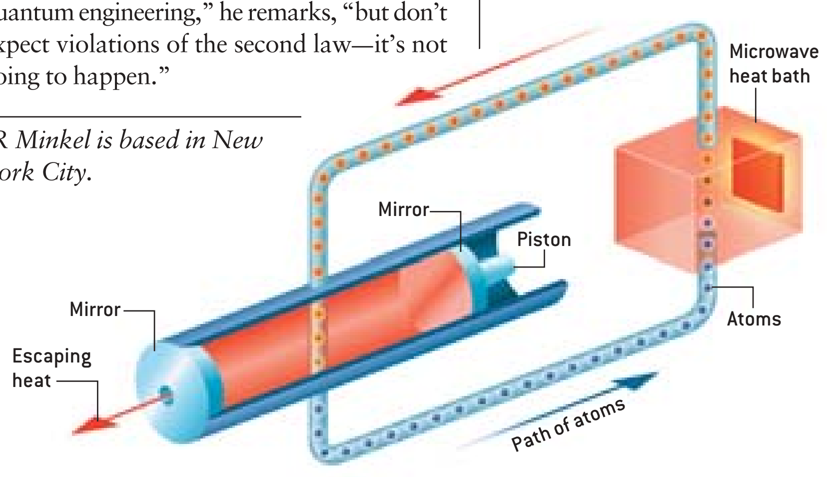
Scully is also attempting to construct and test a laser-engine hybrid. In 2002 he proposed that a quantum "afterburner" could squeeze extra work from some ideal engines that operate below the Carnot limit if the exhaust atoms were stimulated to produce laser light.

Most physicists see no reason why quantum mechanics should damage Carnot's result. Quantum changes preserve disorder, so the second law is built in from the beginning, Lloyd observes. Looking for ways to improve heat engines is "a praiseworthy branch of quantum engineering," he remarks, "but don't expect violations of the second law—it's not going to happen."

JR Minkel is based in New York City.

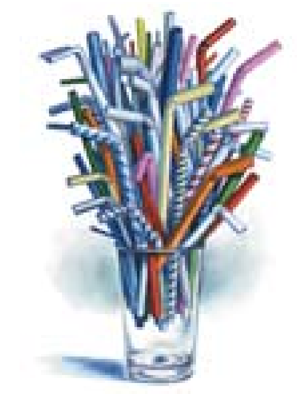
NEED TO KNOW: RANDOM WORK

Some researchers have focused on quantum heat engines that harness the random twitchings of particles—that is, their Brownian motion. When two reservoirs of electrons at different energies make contact, electrons flow toward the colder bath. But random back-and-forth electron motions spoil the efficiency of most devices designed to turn that flow into work. In 2002 Heiner Linke of the University of Oregon and his colleagues reasoned that a semiconductor filter for electrons of specific wavelengths could prevent wasteful energy flow and allow a Brownian heat engine to operate at or near the Carnot limit. Martin A. Green of the Center for Third Generation Photovoltaics in Sydney, Australia, says that his group is currently exploring whether a similar effect could enhance the efficiency of solar cells.



QUANTUM POWER: A microwave bath puts atoms in a special state, one that enables some atoms to emit photons (as heat) into the mirrored cavity but not to absorb them. The photons push the piston to do work; some heat escapes to enable the piston to recompress.

SAMUEL VELASCO



DATA POINTS: NOT ALL WET

In March the United Nations reported on the state of the world's freshwater. Population growth could mean that by the middle of this century, seven billion people in 60 countries could be affected by a lack of clean water. Yet little is being done to confront the impending crisis.

Percent of the world's accessible freshwater used by humans: 54
Percent estimated to be used by 2025: 70

Percent used for agriculture: 69

For industry (average): 22

For industry, high-income countries: 59
For industry, low-income countries: 8

Annual number of deaths from water-related diseases: 5 million
Annual number sickened by poor water: 2.3 billion

Available water per person, in liters per day:

Countries with the least:

Bahamas: 181

United Arab Emirates: 159

Gaza Strip: 142

Kuwait: 27

Country with the most:

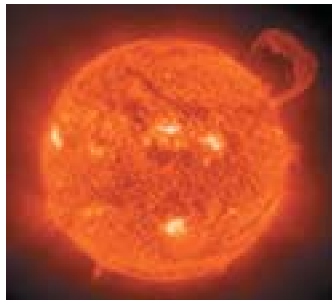
Greenland: 29.5 million

U.S. (contiguous): 20,300

SOURCE: World Water Assessment Program; see www.wateryear2003.org

CLIMATE

Rising Sun



GLOBAL WARMER? The sun as seen in extreme ultraviolet.

Humans may be shouldering too much of the blame for global warming, according to a new look at data from six sun-gazing satellites. They suggest that Planet Earth has been drenched in a bath of solar radiation that has been intensifying over the past 24 years—an increase of about 0.05 percent each decade. If that trend began early last century, it could account for a significant component of the climatic warm-up that is typically attributed to human-made greenhouse gases, says Richard C. Willson of Columbia University's Center for Climate Systems Research in Coronado, Calif. Willson concedes that the climate's sensitivity to such subtle solar changes is still poorly understood, but the evidence merits keeping a close eye on both the sun and humans to better gauge their relative influences on global climate. "In 100 years I think we'll find the sun is in control," he says. His team's report appears in the March 4 *Geophysical Research Letters.* *—Sarah Simpson*

PHYSICS

Why Neutrons Outweigh Protons

Protons and neutrons would be like twins if not for charge symmetry breaking (CSB), a subtle effect that causes neutrons to be 0.1 percent heavier than protons. Had the imbalance gone the other way, hydrogen would not have survived to form stars. Physicists believe that CSB hinges on the repulsion strength and mass difference between the up and down quarks inside nuclear particles, but they haven't been able to pin down the exact values. Now all the puzzle pieces are in place. Researchers working at the Indiana University Cyclotron Facility in Bloomington have made the first observation of a long-sought, rare reaction in which two heavy hydrogen nuclei produce a helium nucleus and a neutral pion, which partially mediates the force holding nuclei together. The reaction rate depends on the mass difference and repulsion interaction. An experiment at the Tri-University Meson Facility in Vancouver has measured another key sign of the cracked symmetry: a slight preference for pions and heavy hydrogen nuclei to fly off in one direction when formed from proton-neutron collisions. Theorists have already begun the arduous calculations required to extract the quark properties from the results. Both groups announced their findings at the April meeting of the American Physical Society. *—JR Minkel*

ECOLOGY

Sexy and Delicious

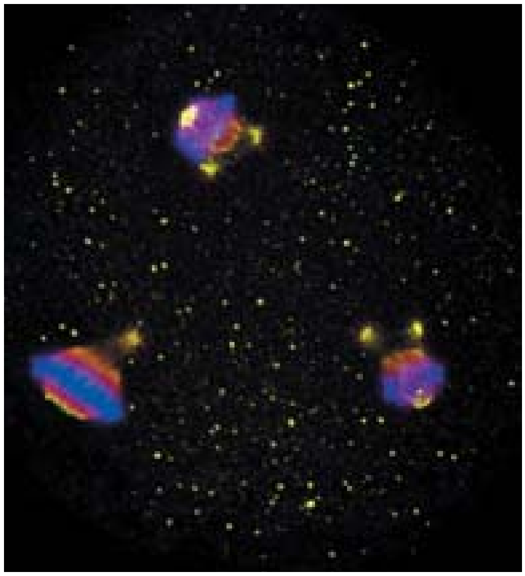
PAINTED BUNTING and other colorful birds go locally extinct 25 percent more often than their drab cousins.

Having extra color may help male birds woo mates, but it also attracts predators. Scientists analyzed 21 years of data from the North American Breeding Bird Survey, which thousands of volunteers across the continent gathered by counting all birds seen or heard during breeding-season mornings. On average, wildlife ecologists find that "dichromatic" bird species die out nearly 25 percent more often than their monochromatic relatives. Two-toned birds in general don't go completely extinct, because as soon as one species vanishes from a neighborhood, other colorful ones take its place. The report, appearing in the April 17 *Proceedings of the National Academy of Sciences USA,* supports the theory that species under strong sexual-selection pressure face greater risks of becoming locally extinct and suggests that human activities that block migrations could jeopardize the survival of dichromatic species. *—Charles Choi*

ILLUSTRATION BY MATT COLLINS; SOHO/NASA (*top*); JEREMY WOODHOUSE *Photodisc/Getty*

BIOTECH

Corrupted Clones



BAD SPLITTING: Mitotic spindles (*red*) separate chromosomes (*blue*) in cloned macaque eggs undergoing cell division. The chromosomes do not divide properly in primates.

Despite claims by a UFO cult and maverick physician Severino Antinori, most scientists think today's cloning methods cannot make a viable baby. Now new research suggests that cloning of primates could very well be impossible. Scientists have attempted to clone rhesus macaques, but none of the resulting embryos survived implantation in a surrogate mother. The researchers found that although cell division appeared superficially normal, chromosomes were divvied up unevenly: some cells ended up with too many and others too few. Usually spindles of protein tubes help to pull opposite ends of a dividing cell apart and ensure that chromosomes split up equally. In normal rhesus egg cells, vital spindle proteins are concentrated near the eggs' chromosomes, which are inadvertently removed during the first steps of each of the four different nuclear-transfer techniques the investigators tried. The location of the spindle proteins could make cloning embryonic stem cells difficult "and reproductive cloning unachievable," the researchers say in the April 11 issue of *Science*. *—Charles Choi*

PSYCHOLOGY

The Unusual Suspects

The more confidence an eyewitness has when identifying a suspect, the stronger that evidence typically becomes in court. But revelations about the malleability of people's memories are upsetting this conventional wisdom. In a recent experiment conducted at Iowa State University, all 253 participants who watched a staged crime video chose a suspect from a six-man lineup—even though the true culprit was not among them. Unaware that they were mistaken, witnesses who were then told, "Good, you identified the suspect," tended to further overstate their confidence and recollection of details, including the criminal's facial features. The false certainty prevailed whether they heard the affirmation immediately following the lineup or a full 48 hours later. The Iowa researchers conclude that law enforcers must curb on-the-spot comments about suspects and secure statements about a witness's confidence right away to avoid tainting future testimony. The report appears in the March *Journal of Experimental Psychology: Applied.* *—Sarah Simpson*

CRIMINAL LINEUPS have been around for decades—this one is from Chicago in 1927—but police comments to witnesses afterward may produce false certainty in the testimony.

COURTESY OF CALVIN SIMERLY (*top*); UNDERWOOD & UNDERWOOD *Corbis* (*bottom*)

BRIEF POINTS

- **The oldest verifiably original DNA has been found in Siberian sediment. The plant genetic material may be 400,000 years old and harbor clues about the paleoenvironment.**

  Science, April 18, 2003

- **A survey has found catastrophic decreases in gorillas and chimpanzees in western equatorial Africa, the last stronghold of these apes. Hunting and, more recently, an Ebola outbreak appear to be the primary causes.**

  Nature, April 10, 2003

- **In a mouse study, the drug Accutane prevented the accumulation of lipofuscin, a toxin that causes the macular degeneration of Stargardt's disease. The drug, which can produce night blindness as a side effect in acne treatment, apparently mimicked the effects of light deprivation.**

  Proceedings of the National Academy of Sciences USA (online), March 17, 2003

- **Fertilizing the ocean with iron to encourage the growth of plankton, which absorb carbon dioxide from the air, may not be very effective in sequestering the greenhouse gas, because the carbon may not sink deep enough to remain locked away.**

  Science, April 4, 2003

BY THE NUMBERS

Trade Globalization

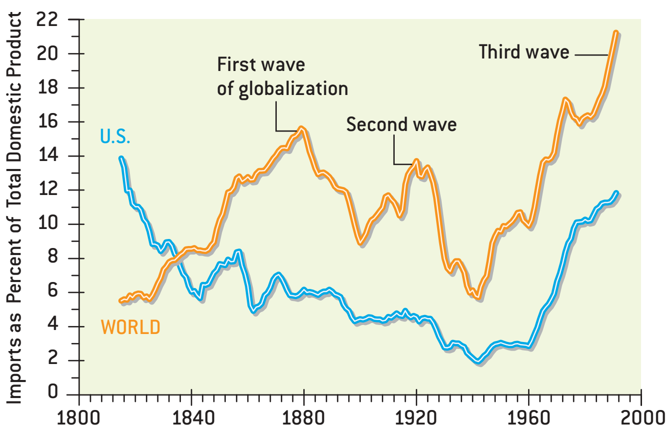
IT IS NEARLY TWO CENTURIES OLD AND LIKELY TO CONTINUE **BY RODGER DOYLE**

Many think that globalization is a recent development, but its origins go back to the early 19th century. This fact is apparent from a new study by sociologist Christopher Chase-Dunn of the University of California at Riverside and his colleagues. Their data, which are based on the relation between imports and gross domestic product, show that the initial wave of globalization began about 1830 and peaked about 1880. During this time, international commerce, with the abandonment of mercantilism, first became a force in the lives of ordinary people. Before the 19th century, international trade was a paltry affair mostly confined to luxuries, such as spices and tobacco. This early wave is associated with the growth of railroads, more efficient ocean transport, and the political victory of manufacturing and trading interests over those of the landowners, signaled by the 1846 repeal of the British corn laws. (Those laws imposed duties on imported corn and thereby kept prices high.) The second wave coincided with the rise of electricity and steel around 1900 and peaked in the 1920s. The current wave began after World War II as a result of the creation of international institutions such as the General Agreement on Tariffs and Trade, the predecessor of the World Trade Organization.

Decreasing costs of transport and communication underlie the long-range increase in world trade, but no satisfactory reason explains the wave pattern. Chase-Dunn cites "hegemonic stability," in which a great power provides stable conditions. The first and third coincide with, respectively, the eras of British and U.S. hegemony, but, as he notes, the theory does not account for the second wave, which occurred when Britain was in relative decline and the U.S. had not yet asserted its power.

During most of the 19th and early 20th centuries, America did not follow Britain's free-trade policy but instead imposed high tariffs to protect manufacturing. The U.S. became more open to imports only after World War II. But it still lagged behind other major countries in trade participation—not surprising considering its vast domestic market, which could supply a larger variety of demands than smaller economies could. Nevertheless, the greater involvement of the U.S. has been the primary factor in world trade expansion since World War II.

In the long run it is very likely that international trade will continue to expand as the costs of transport and communication continue to decline. Perhaps the most formidable obstacle to trade growth in the near future is failure to reform government practices that foster doubt and mistrust. Transparency International, an organization funded by several European governments, polls well-informed individuals in more than 100 countries regarding the extent of misuse of public power for private benefit. Its 2003 report shows that trust in the institutions of industrial nations averages 7.3 out of a perfect score of 10; developing countries average only 2.3. How governments deal with this issue of integrity could largely determine the next phase of worldwide trade.



SOURCE: Christopher Chase-Dunn et al., 2000. Data are shown as five-year moving averages.

Next: Globalization's winners and losers.

Rodger Doyle can be reached at rdoyle2@adelphia.net

BIG-TIME PLAYERS

The 12 Largest Trading Countries	Percent of World Goods and Services in 2000: Imports	Exports
U.S.	18.7	14.1
Germany	8.1	8.3
Japan	6.3	7.2
U.K.	5.5	5.2
France	4.6	5.0
Canada	3.6	4.1
Italy	3.7	3.9
China*	2.6	3.3
Netherlands	3.2	3.4
Belgium	2.7	3.0
Korea	2.5	2.7
Spain	2.3	2.2

** Excludes services, which account for about 20 percent of international trade worldwide.*

SOURCES: International Trade Statistics Yearbook, *2000, United Nations;* OECD Statistics on International Trade in Services, 1999–2000.

FURTHER READING

Trade Globalization since 1795: Waves of Integration in the World-System. Christopher Chase-Dunn, Yukio Kawano and Benjamin D. Brewer in *American Sociological Review*, Vol. 65, No. 1, pages 77–95; February 2000.

Globalization, Trade, and Development: Some Lessons from History. Alan M. Taylor. National Bureau of Economic Research working paper 9326; November 2002.

RODGER DOYLE

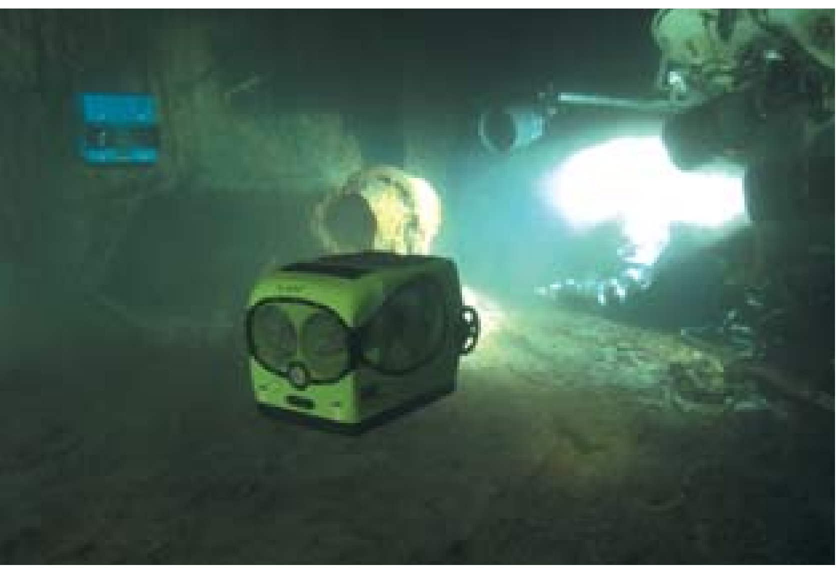
The Abyss Transit System

James Cameron commissions the making of robots for a return to the *Titanic*   By GARY STIX

At the beginning of the movie that made Leonardo DiCaprio a megastar, a camera-toting unmanned robot ventured into a cavernous hole in the wreck that sits on the bottom of the Atlantic, 12,640 feet from the surface. The 500-pound vehicle, christened Snoop Dog, could move only about 30 feet along a lower deck, hampered by its bulky two-inch-diameter tether hitched to a submarine that waited above. The amount of thrust needed to move its chunky frame stirred up a thick cloud. "The vehicle very quickly silted out the entire place and made imaging impossible," director James Cameron recalls.

But the eerie vista revealed by Snoop Dog on that 1995 expedition made Cameron hunger for more. He vowed to return one day with technology that could negotiate anyplace within the *Titanic*'s interior.

In the past six months two documentaries—one for IMAX movie theaters called *Ghosts of the Abyss,* the other, *Expedition: Bismarck,* for the Discovery Channel—demonstrated the fruits of a three-year effort that Cameron financed with $1.8 million of his own money to make this vision materialize. The payoff was two 70-pound robots, named after Blues Brothers Jake and Elwood, that had the full run of two of the world's most famous wrecks, the *Titanic* and the *Bismarck,* which they visited on separate expeditions.



LITTLE EYES: Remotely controlled robots use onboard cameras to explore the deepest innards of sunken vessels.

The person who took Jake and Elwood from dream to robot is Mike Cameron, James's brother, an aerospace engineer who once designed missiles and who also possesses a diverse background as a helicopter pilot, stunt photographer and stuntman. (Remember the corpse in the movie *The Abyss,* from whose mouth a crab emerges?) Giving the remotely operated vehicles freedom of movement required that they be much smaller than Snoop Dog and that the tether's width be tapered dramatically so as not to catch on vertical ship beams.

Mike Cameron took inspiration from the wire-guided torpedoes used by the military that can travel for many miles. His team created vehicles operable to more than 20,000 feet (enough to reach as much as 85 percent of the ocean floor). The dimensions of the front of the robot are 16 inches high by 17 inches across, small enough to fit in a B deck window of the *Titanic*. The bots have an internal battery so that they do not need to be powered through a tether. Instead the tether—fifty-thousandths of an inch in diameter—contains optical fibers that relay control signals from a manned submersible vehicle hovering outside and that also send video images in the other direction. The tether pays out from the robot, a design that prevents it from snagging on objects in the wreck.

James Cameron thought the project would be a straightforward engineering task, not much harder than designing a new camera system. "This turned out to be a whole different order of magnitude," he says. "There was no commercial off-the-shelf hardware that would work in the vehicles. Everything had to be built from scratch." If the team had known this early on, he added, "we wouldn't have bothered." Water pressure on the

PHOTOGRAPHS COURTESY OF WALT DISNEY PICTURES cable that carried the optical fibers could create microscopic bends in the data pipe, completely cutting off the control signals from the submersibles. Dark Matter in Valencia, Calif. (Mike Cameron's company), had to devise a fluid-filled sheath around the fiber to displace the minuscule air pockets in the cable that could lead to the microbending.

To save weight, the frame—similar to a monocoque body of a race car—was made up of small glass hollow spheres contained in an epoxy matrix. The thruster contained a large-diameter, slowly rotating blade with nozzles that diffused the propulsive flow, minimizing the churning that would otherwise disturb the caked silt. A high-resolution video camera, along with an infrared camera for navigation, was placed in the front of the craft along with three light-emitting-diode arrays for fill lighting and two quartz halogen lamps for spotlighting.

The winter of 2001 marked a critical juncture. It was six months before dives to the *Titanic* could be safely attempted, and James had to determine whether to proceed or wait another year. "Mike was really, really negative on the idea, but I decided to go for it," the director says. He felt he couldn't afford to wait longer and thought that a fixed deadline would focus the engineering staff at Dark Matter. For his part, Mike was contending with an unending series of design challenges. "It was such an overwhelming set of problems that I had very little confidence that certain parts would be solvable in the time we had," Mike says.

A few weeks before the dives commenced in the summer of 2001, the robots' lithium sulfur dioxode–based batteries caught fire while being tested in a pressure tank, destroying what was to have been a third robot. Mike wanted to delay the dives, but James found a supplier of another type of lithium battery and pressed ahead.

At the dive site, Jake and Elwood took starring roles with their 2,000-foot tethers, exploring for the first time in about 90 years remote parts of the ships, including the engine room, the firemen's mess hall and the cabins of first-class passengers—even focusing in on a bowler hat, a brass headboard and an intact, upright glass decanter. The images lack the resolution and novel quality of the high-definition, three-dimensional IMAX images, the other major technological innovation of *Ghosts of the Abyss*. Jake and Elwood's discoveries, however, draw the viewers' interest because of what they convey of the *Titanic*'s mystique. "You actually feel like you're out there in the wreck," Mike says. He remembers his brother piloting the robots with the helicopter stick that had been installed in the Russian submersible from which the robots were launched. "Jim ended up being a cowboy pilot," Mike says. "He was far more aggressive with the system than I was."

One scene in *Ghosts of the Abyss* reveals the tension that sometimes erupted between the brothers. James contemplates moving one of the robots through a cabin window that is still partially occluded by a shard of glass that could damage the vehicle or cut the data tether. When James declares that he is going to take Jake in, moviegoers can hear Mike pleading with his brother not to do it, ultimately relenting once the bot has negotiated the opening.

The decision to install a new type of battery at the last minute came to haunt the expedition; Elwood's lithium-polymer battery ignited while in the bowels of the ship. James manipulated the remaining robot into the *Titanic* to perform a rescue operation by hooking a cord to the grill of the dead bot and towing it out. At the surface—on the deck of the Russian scientific vessel the *Keldysh,* from which the two submarines carrying Jake and Elwood to the *Titanic* were launched—Mike rebuilt Elwood with a backup battery. During the next dive, the robot caught fire again while it was still mounted on the submarine, endangering the crew. Finally, Mike worked for an 18-hour stretch to adapt a lead-acid gel battery used for devices onboard the mother ship into a power source for Elwood, enabling the expedition to continue.

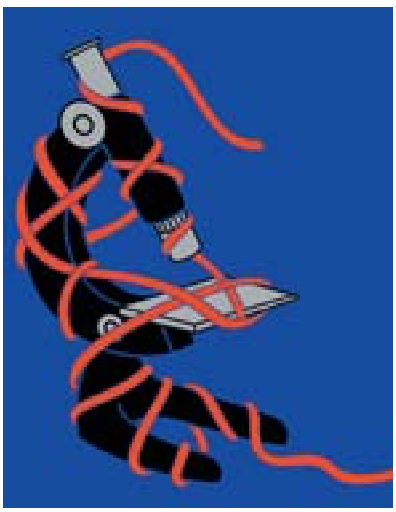
Robots Jake and Elwood, with their 2,000-foot tethers, took on starring roles, exploring the remotest reaches of the *Titanic.*

The bots, now fitted with a new, nonflammable battery that Mike designed, may find service beyond motion pictures. The U.S. Navy has funded Dark Matter to help it assess the technology for underwater recovery operations of ships or aircraft. The bots also have potential for scientific exploration of deep-sea trenches. After traveling to the *Titanic* and the *Bismarck,* the team went on to probe mid-Atlantic hydrothermal vents, discovering mollusks in a place where scientists had never encountered them before. As adventure aficionados, the brothers speculate that a descendant of Jake and Elwood might even be toted on a mission to Europa, one of Jupiter's moons, to investigate the waters that are suspected to exist below its icy shell. The Cameron siblings, who tinkered with home-built rafts and rockets as children in Ontario near Niagara Falls, hope to be around long enough to witness their robotic twins go from the bottom of the ocean to the depths of space. SA

Sign Here

Will a scientist need a legal opinion before starting the next experiment? By GARY STIX

Academic biologists routinely work with genes without so much as a second thought. They focus their attention on determining the function of the gene and the protein it produces, not on whether the relevant DNA is patented or not. According to some universities and scholarly associations, a recent federal appeals court decision means that the entire scholarly research community, not just biologists, will be spending a lot more time with lawyers to determine whether their investigations violate someone's patent rights.



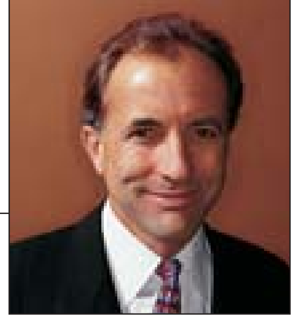
JENNIFER KANE

Universities have often labored under the assumption that using research tools and materials is a permissible practice: noncommercial uses fall under a research exemption that precludes liability for patent infringement. The Court of Appeals for the Federal Circuit (CAFC), which hears appeals of patent cases, issued a ruling last year that defines this safe haven for researchers so narrowly that it becomes virtually useless. It reiterated that the exemption applies only "for amusement to satisfy idle curiosity, or for strictly philosophical inquiry." But noncommercial, academic research, the court decided, serves to further the "legitimate business objectives" of the university, so patented equipment and materials do not warrant an exemption.

Some universities fear that researchers will now have to devote time and grant money to conducting patent searches and arranging licensing agreements before proceeding with their experiments. The rationale for a research exemption is based on legal opinions issued by judges that date back to 1813. But many members of the legal community argue that despite these precedents, the exception for academics was always very narrow. "It's a widespread urban legend that this research exemption would protect university research," notes Lynn H. Pasahow, an intellectual-property attorney with the law firm Fenwick & West in Mountain View, Calif. The case for the exemption has also been weakened because universities are now involved more than ever in obtaining licensing revenues for discoveries made by their researchers.

What spurred the CAFC's ruling was a suit brought by John Madey, the inventor of the free-electron laser. He brought a claim against Duke University after that institution removed him as head of a laboratory in 1997 and he had moved to another university. In the complaint, he charged that Duke had, among other things, violated his patent rights by continuing to use the laser equipment. On appeal, the CAFC sent the case back to a lower court, saying that it had erred by using an overly broad interpretation of the research exemption to decide that Duke had not infringed Madey's patents. Duke has asked the U.S. Supreme Court to review the CAFC's decision, and a number of other universities and associations have supported Duke by filing friend of court briefs.

Not everyone is worried. Lita Nelsen, director of the technology licensing office at the Massachusetts Institute of Technology, expects the fallout from the case to be minimal—most individuals and companies are not interested in bringing high-profile suits against academic researchers. "Most of the time the researcher doesn't know a patent exists and is not doing any harm," Nelsen says. "If somebody comes to us and says, 'I hold the patent, and I want you to stop doing something,' the first thing I'd say is 'Do you care?' And if the person did, we'd respect the patent." Even if the Supreme Court does not take up the case, *Madey v. Duke* may trigger useful debate—and may lead to a push for legislation that clarifies when universities and nonprofits may conduct research without first making a call to their attorney. SA



Codified Claptrap

The Bible Code is numerological nonsense masquerading as science By MICHAEL SHERMER

In the epilogue of *In Memoriam A.H.H.*, Alfred, Lord Tennyson captured the essence of the quest for a single unifying principle and purpose in nature: "One God, one law, one element,/And one far-off divine event,/To which the whole creation moves."

The noble dream of finding teleological succor in the march of time has become big business, as demonstrated by works from Hal Lindsey's 1970s blockbuster *The Late Great Planet Earth* to today's Left Behind series, by Tim LaHaye and Jerry B. Jenkins. (Both are said to have sold in the tens of millions.) And if you can sprinkle your homiletics with scientistic jargon, so much the better. The latest and most egregious example of the (mis)use of science in the (dis)service of religion is Michael Drosnin's *Bible Code II*, enjoying a lucrative ride on the *New York Times* best-seller list, as did the 1997 original.

Just like the prophecies of soothsayers past and present, all such predictions are actually postdictions.

According to proponents of the Bible Code—itself a subset of the genre of biblical numerology and Kabbalistic mysticism popular since the Middle Ages—the Hebrew Pentateuch can be decoded through an equidistant-letter-sequencing software program. The idea is to take every *n*th letter, where *n* equals whatever number you wish: 7, 19, 3,027. Print out that string of letters in a block of type, then search left to right, right to left, top to bottom, bottom to top, and diagonally in any direction for any interesting patterns. Seek and ye shall find.

Predictably, in 1997 Drosnin "discovered" such current events as Yitzhak Rabin's assassination, Benjamin Netanyahu's election, Comet Shoemaker-Levy 9's collision with Jupiter, Timothy McVeigh and the Oklahoma City bombing, and, of course, the end of the world in 2000. Because the world did not end and current events dated his first book, Drosnin continued the search and learned—lo and behold—that the Bible predicted the Bill and Monica tryst, the Bush-Gore election debacle and, of course, the World Trade Center cataclysm.

Just like the prophecies of soothsayers past and present, all such predictions are actually postdictions (note that not one psychic or astrologer forewarned us about 9/11). To be tested scientifically, Bible codes would need to predict events *before* they happen. They won't, because they can't—as Danish physicist Niels Bohr averred, predictions are difficult, especially about the future. Instead, in 1997 Drosnin proposed this test of his thesis: "When my critics find a message about the assassination of a prime minister encrypted in *Moby Dick*, I'll believe them."

Australian mathematician Brendan McKay did just that, locating no fewer than nine political assassinations secreted in the great novel, along with additional discoveries in *War and Peace* and other tomes (see cs.anu.edu.au/~bdm/dilugim/moby.html). American physicist David E. Thomas predicted the Chicago Bulls's NBA championship in 1998 from his code search of Leo Tolstoy's novel. He also recently unearthed "the Bible code is a silly, dumb, fake, false, evil, nasty, dismal fraud and snake-oil hoax" from *Bible Code II* (see www.nmsr.org/biblecod.htm).

If there is an encrypted message in all this numerological poppycock it is this: there is a deep connection between how the mind works and how we perceive the world works. We are pattern-seeking animals, the descendants of hominids who were especially dexterous at making causal links between events in nature. The associations were real often enough that the ability became engrained in our neural architecture. Unfortunately, the belief engine sputters occasionally, identifying false patterns as real. The habit of faltering may not be enough to prevent you from passing on your genes for detecting false positives to the next generation, but it does create superstitious and magical thinking. This process is coupled to the law of large numbers that accompanies our complex world, where, as it is said, million-to-one odds happen eight times a day in New York City.

Given our propensity to look for patterns in a superfluity of data, is it any wonder that so many are taken in by such codified claptrap? The problem is pervasive and a permanent part of our cognitive machinery. The solution is science, our preeminent pattern-discriminating method and our best hope for detecting a genuine signal within the noise of nature's cacophony. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* Why People Believe Weird Things.

BRAD HINES

One Last Look

Although United Nations weapons inspector Rocco Casagrande and his colleagues found no bioweapons in Iraq, they could sense that the government had not come clean   By GARY STIX

The Saddam Center for Biotechnology on the campus of Baghdad University boasted a state-of-the-art facility, replete with surreptitiously imported equipment for amplifying tiny amounts of DNA and running tests with gels to determine protein sizes. "It looked like you were walking into a laboratory in one of the better-equipped U.S. institutions," remembers Rocco Casagrande, who began his trips as a United Nations inspector to various Iraqi facilities in mid-December 2002.

The lab was ideal for performing DNA amplification using the polymerase chain reaction (PCR) to make countless copies of genes. Oddly, though, the only thing these expensive machines were being used for was genetic fingerprinting of goats involved in what the Iraqis said were in vitro fertilization experiments. Iraq does not suffer from problems with goat fertility. An infertile goat is eaten for dinner, not sent to an IVF clinic. Casagrande and the others took samples from the lab and combed through records on a computer hard disk, to no avail. No evidence of cloning genes for making bioweapons was found. They speculated that the facility could be used for human cloning, but in the end they never figured out its real purpose.

This experience was not the only time during his three-month stay that Casagrande encountered projects that did not quite make sense. But neither did the biological weapons inspection team come across the anthrax, botulinum or any pathogen that had been part of the notorious program that the Iraqi government claimed was now defunct.

Casagrande was one of about 10 U.S. representatives on the roughly 100-member team of nuclear, chemical, biological and missile inspectors, a contrast to the investigations in the 1990s, when many more of the officials were American. Every day the team received lists of sites to visit from U.N. headquarters in New York City. Some destinations were obvious, such as the biotechnology center; others were gleaned from intelligence reports. Once they arrived, a few inspectors conducted interviews while the rest looked for suspicious activity. Casagrande and his colleagues became familiar faces to Iraqis in the months immediately preceding the war. At night Iraqi television broadcast extensive coverage, identifying the inspectors by name and country of origin. Casagrande couldn't go into a restaurant or shop without being recognized.

This 29-year-old—only a few years beyond a doctorate in biology from the Massachusetts Institute of Technology—had the job of refitting the biological

JAMES SALZANO

ROCCO CASAGRANDE: WITNESS TO HISTORY

- **Member of the 20-person U.N. bioweapons inspection team that visited about 150 Iraqi sites from December 2002 through early March of this year.**
- **Sites visited included: breweries, dairies, hospitals, airfields, ammunition dumps, pharmaceutical manufacturers and a tomato cannery.**
- **"I'd never be comfortable leaving the country and saying Iraq doesn't have biological agents. It wasn't behaving like a country that doesn't have biological weapons."**

analysis laboratory used to test samples taken during its daily tours. U.N. inspectors had last operated the lab in 1998, before they were expelled from Iraq; in the interim, it had become a nesting place for pigeons.

As a child growing up in a Philadelphia suburb, Casagrande was fascinated with both science and history. While doing his doctoral work, he became involved with the Harvard Sussex Program's Chemical and Biological Warfare Colloquium, led by Harvard University microbiologist Matthew Meselson, and realized that cultivating an expertise in biowarfare would serve as a means to combine his two interests. Before he left for Iraq, Casagrande was developing weapons biodetectors for Surface Logix, in Cambridge, Mass.

The U.S. State Department took notice of Casagrande after he wrote articles for *Nonproliferation Review* and *Bioscience* chronicling the potential threat of biowarfare against crops and livestock. The government later recommended him for the U.N. post. (He is now employed by Abt Associates in Cambridge, setting up a homeland security consulting practice.)

In Iraq, Casagrande was always aware that a positive result on any of the countless assays performed in the refurbished lab could reverberate around the world. "We tried not to think about what the implications were for what we might find," he says. "But we couldn't help but realize that this could be a turning point in history." The 20 or so members of the bioweapons team, one fifth of the total complement of inspectors, which also included chemical, missile and nuclear contingents, visited sites ranging from breweries to munitions facilities. The inspectors had been taught during training sessions in Geneva how mundane equipment for producing such routine items as beer or tomato sauce could also be employed for culturing anthrax or another bioweapons agent.

JUST OIL? U.N. weapons inspector Rocco Casagrande examines dilapidated oil barrels on a farm in Juwesma, Iraq, 26 miles southwest of Baghdad, in January.

HUSSEIN MALLA *AP Photo*

Interviews often took on a surreal quality that reflected the deep-seated fear that gripped the populace. A simple question—"How long have you been head of this facility?"—could elicit a five-minute answer that never addressed the original query. Reactions could turn hostile. Casagrande recalls speaking with the head of an agricultural research center whose director accused him of being personally responsible for sanctions against the country. Technicians sometimes flatly disavowed the presence of certain microbes. In one instance, Casagrande had to reprimand a worker in a university lab who had refused to acknowledge possession of a strain of anthrax that was found by the inspectors (the strain could be used only for making a vaccine, not a bioweapon).

Iraqi "minders" constantly tracked the inspectors and followed them everywhere. A certain warmth developed between the two groups. On a trip south of Baghdad the inspectors had to wait endlessly as the Iraqis fished in two bags of unmarked keys to enter 150 triple-padlocked cinderblock buildings that turned out to house just conventional ammunition. One minder told his charges that if an invasion ever came, it would be important to give the Iraqis three months' notice so that they would have enough time to open the bunkers. "It was actually a friendly relationship, but we understood they weren't our friends," Casagrande comments. Chumminess had its risks. An escort, whom the inspectors knew as Mr. Wa'ad, remarked to the team that he envied his relatives who had emigrated to the U.S. "He soon disappeared," Casagrande says.

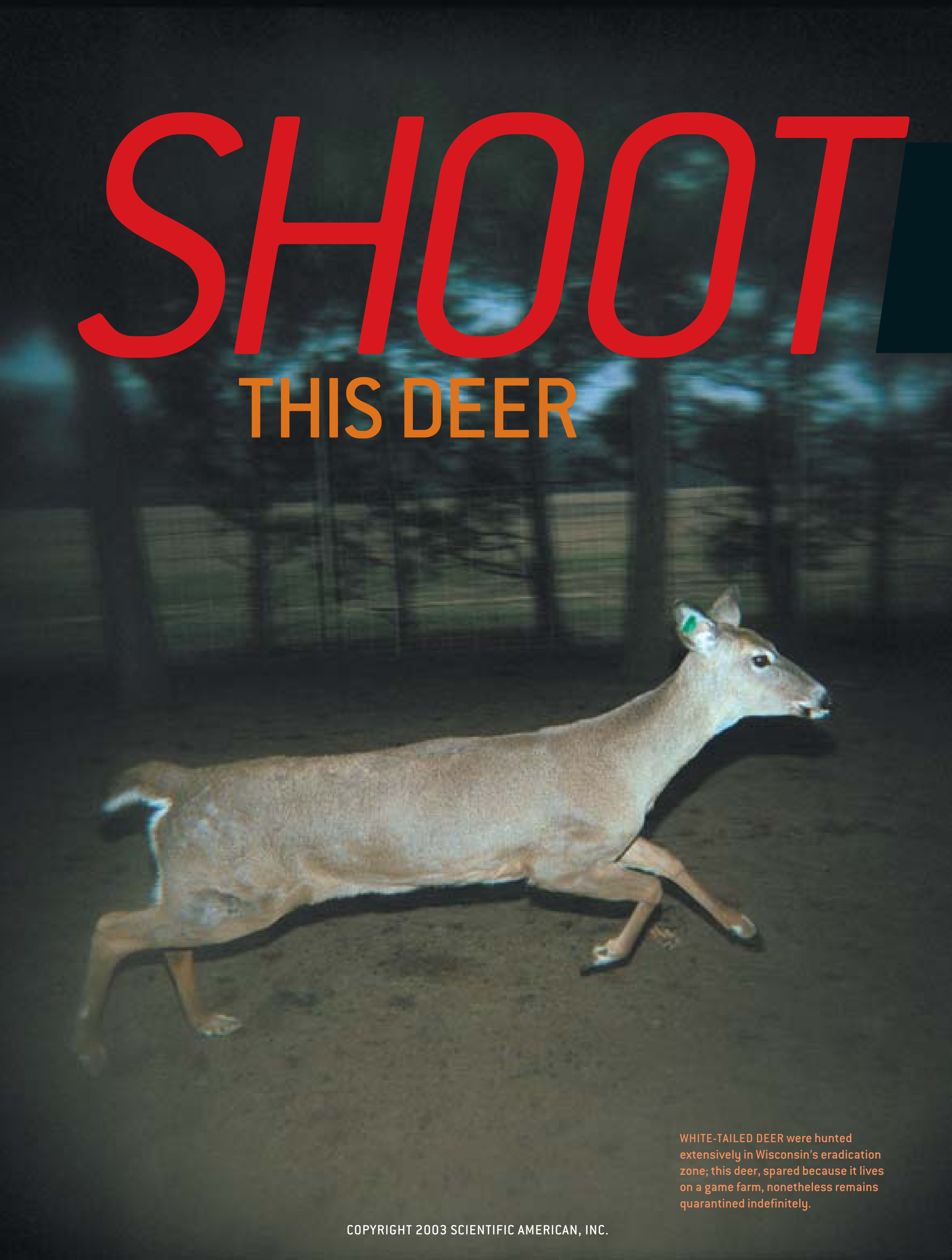
Some of the Iraqi scientists were eager to exchange information with outsiders and might launch into a discussion with Casagrande about bacterial indicators of soil health. He even met the infamous Rihab Taha, the former director of the country's biological weapons program, who related to the inspectors that she now spends her time caring for her children as a Baghdad homemaker. (Women headed about a third of the civilian laboratories visited.)

During their stay, the team members never uncovered what they were seeking. Still, Casagrande came away with a distinct uneasiness. It seems unimaginable to him that a government so obsessed with documentation—the moving of a centrifuge from one room to another required extensive paperwork—would be unable to account for how it disposed of pathogens from its previous biowarfare program and to reveal what it did with large quantities of growth media used to culture pathogenic agents.

There were places the inspectors did not look. Bioweapons could have been secreted in off-limits religious sites. Iraq was, in fact, in the midst of a mosque-building boom, including the recently completed Mother of All Battles mosque, with minarets shaped to resemble Scud missiles. Also questionable was the discovery of a possible smallpox vaccination program. "It makes you wonder why someone in Iraq thought they needed to be vaccinated against smallpox," says Casagrande, who as a U.N. inspector chose not to offer an opinion about the U.S.-Iraq war.

Despite the frustrations, Casagrande feels that the work was not wasted. The inspectors had their stay cut short. But information that they gathered might help in conducting follow-up investigations to unravel the extent of the regime's conjectured clandestine programs to cultivate anthrax, botulinum and other mass killers. After all, those supposed weapons stocks would be the after-the-fact basis for waging a war. SA

SHOOT THIS DEER



WHITE-TAILED DEER were hunted extensively in Wisconsin's eradication zone; this deer, spared because it lives on a game farm, nonetheless remains quarantined indefinitely.

Chronic wasting disease, a cousin of mad cow disease, is spreading among wild deer in parts of the U.S. Left unchecked, the fatal sickness could threaten North American deer populations—and maybe livestock and humans

By Philip Yam

A place called the eradication zone, lying about 40 miles west of Madison, Wis., covers some 411 square miles. There thousands of white-tailed deer live—or rather, used to live.

Last year the Wisconsin Department of Natural Resources instituted special hunting periods to try to wipe out upward of 18,000 deer. During the fall, dead deer were taken to registration areas, where state employees in protective suits and gloves dragged carcasses from pickup trucks and lifted them onto plastic-covered picnic tables. With hacksaws, they severed the heads, double-bagged them and sent them for testing; the bodies themselves were incinerated.

The Dairy State's massacre is an attempt to keep a fatal ailment known as chronic wasting disease (CWD) from infecting its other 1.6 million deer. The testing enables wildlife officials to ascertain the scope of the epidemic—running at nearly 1.6 percent—and determine whether the culling can slow the spread. Currently no practical live test exists to check whether an apparently healthy, wild animal is actually incubating the sickness; only a brain sample will do.

The disease occurs because a pathogen peppers neural tissue full of microscopic holes and gums up the brain with toxic clumps of protein called amyloid plaques. Long confined to a patch of land near the Rocky Mountains, the disease has shown up in 12 states and two Canadian provinces. The sickness passes readily from one deer to another—no deer seem to have a natural resistance. "From everything we've seen," comments Michael W. Miller, a CWD expert with the Colorado Division of Wildlife, "it would persist. It would not go away on its own."

The urgency also reflects concern about the nature of CWD, which belongs to the same family as a better-known scourge: bovine spongiform encephalopathy (BSE), or mad cow disease. Spread by animal-based feed inadvertently containing tissue from sick cows and sheep, BSE emerged in the U.K. in the 1980s and continues to plague that country at a low level. (Nearly two dozen other countries have now also reported cases.) In 1996 scientists realized that BSE can pass to humans who eat infected meat, leading to a fatal condition: variant Creutzfeldt-Jakob disease, or vCJD (distinct from the more common sporadic CJD, which arises spontaneously in one in a million people). Researchers are now trying to figure out whether CWD could infect humans and livestock and thereby create an American version of the U.K.'s mad cow disaster.

Pathological Protein

THE DISEASE AGENT COMMON to all these maladies is the prion ("PREE-on"), a term coined in 1982 by Stanley B. Prusiner of the University of California at San Francisco. The prion is a protein that exists in all animals, although the exact amino acid sequence depends on the species. It takes one of two shapes. Folded correctly, it is the normal prion protein (PrP), which is especially abundant in brain cells and may help process copper. Folded incorrectly, the prion protein becomes a pathogenic entity that kills. The malformed protein has the ability to refold copies of normal PrP in its own image, thereby making more of itself.

Prusiner's conception of prions initially met with great skepticism. That a pathogen could replicate and pass on its traits without assistance from nucleic acids (DNA or RNA) violated the orthodoxy of molecular biology. But enough evidence has accumulated to prove that some proteins can in fact copy themselves and that variants of PrP are essential players in spongiform encephalopathies.

That prions lack any DNA or RNA is also the prime reason

DAVID NEVALA

why they are so tough. Germicidal light, formaldehyde baths and boiling water all promptly disrupt bacterial and viral nucleic acids, yet such treatments have little effect on malformed prions. Researchers have exposed prion-contaminated tissue to a dry heat of 600 degrees Celsius and left it buried for three years, only to find that the material, though greatly weakened, was still infectious. Indeed, physicians have unwittingly passed prion diseases on to patients via surgical instruments and transplanted organs that had undergone standard sterilization procedures. (Prion disinfection requires extended heating or corrosive chemicals such as sodium hydroxide.)

Foothold in the Foothills

THE RESILIENCE OF misfolded prions appears to be a key reason why chronic wasting disease has persisted and spread from its presumed starting point near Fort Collins, Colo. There, in 1967, at the state's Foothills Wildlife Research Facility, CWD made its first recorded appearance, in captive mule deer that were being maintained for nutritional studies (mule deer are the most common type in the West). As the name of the disease suggests, affected deer lose weight over the course of weeks or months. They often become thirsty, which drives them to drink large amounts of water and, consequently, to urinate a great deal; they also start slobbering and drooling. They may stop socializing with fellow deer, become listless or hang their heads. Death typically ensues three to four months after symptoms start, although some victims expire within days and others in about a year. The incubation period, during which the animals show no symptoms, ranges from about 20 to 30 months.

The Fort Collins facility became a CWD death trap. Between 1970 and 1981, 90 percent of the deer that stayed more than two years died from the disease or had to be euthanized. In 1980 the scourge emerged outside Colorado, at the Sybille Research Unit in southeastern Wyoming, 120 miles northwest of Fort Collins. The two facilities had exchanged deer for breeding purposes, thus indicating that the disease was infectious—even to a different species: soon the elk at the facilities contracted the disease. (Deer and elk both belong to the cervid family.)

For years, researchers thought CWD resulted from nutritional deficiencies, poisoning, or stress from confinement. But in 1977 Elizabeth S. Williams, studying for her doctorate at Colorado State University, discovered that this view was mistaken. When Williams looked at brain slices from infected animals, she saw that the tissue was full of microscopic holes. "I happened to be taking a course in neuropathology and had studied a lot of brain lesions," she recalls. The holes were unmistakably like scrapie, the sheep sickness that was the first documented spongiform encephalopathy.

In fact, CWD appears to have originated from scrapie. Richard E. Race of the National Institutes of Health Rocky Mountain Laboratories in Hamilton, Mont., conducted test tube studies that revealed no distinction between the malformed PrP of scrapie sheep and CWD cervids. Consistent with this discovery, Amir Hamir of the U.S. Department of Agriculture's National Animal Disease Center in Ames, Iowa, found no difference in the appearance of brain samples from elk with CWD and elk experimentally infected with scrapie. (BSE also probably arose from scrapie, after cows ate feed derived from infected sheep.)

Geography provided natural barriers, but humans provided a way past them: along the roads, IN A TRUCK.

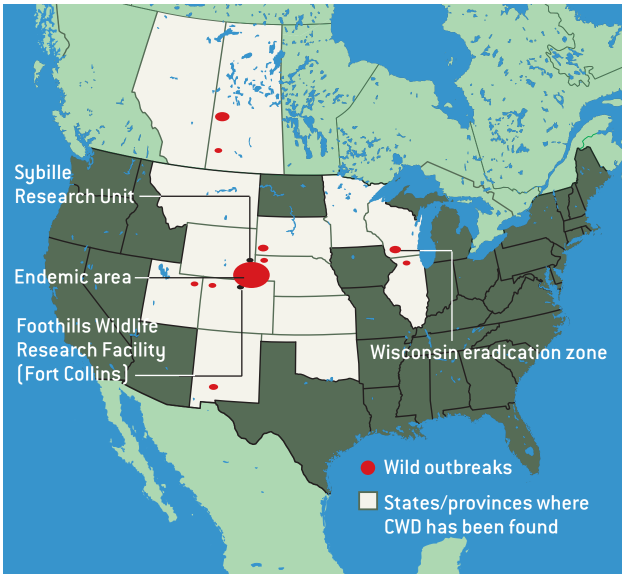
But unlike BSE in cows (or vCJD in humans), the cervids were not getting ill from their food. CWD behaves more like scrapie, in that the sickness spreads among individuals, although no one really knows how it does. The prions could lurk in the urine. During rutting season, deer bucks lap up the urine of perhaps dozens of does to find out which are in heat. Elk females lick males that have sprayed themselves with urine. Saliva could be a vector, too; in both deer and elk, individuals meet and greet by licking each other's mouths and noses, thus exchanging drool. Ranched elk may swap saliva when they feed in close quarters. It is also possible that animals take in the pathogen while grazing in areas where sick animals have shed prions on the ground in their feces, urine and saliva.

By 1985 veterinarians discovered CWD in free-ranging deer and elk, generally within about 30 miles of the two wildlife facilities. Whether the disease originated in the wild and spread to the captives, or vice versa, is not known. The two populations had plenty of time to mingle. Especially during mating season, wild cervids nosed up to captives through the chain-link fences. Incubating deer could also have escaped or been released.

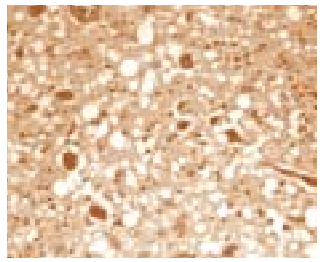
Both facilities tried hard to eradicate CWD. The Sybille center killed all the deer and elk in the affected area and waited a year to introduce new animals; four years later deer and elk started coming down with CWD. The Fort Collins facility acted more aggressively. Officials first

Overview/*Chronic Wasting Disease*

- Chronic wasting disease (CWD) is a fatal condition spreading among wild deer in some parts of North America. It kills in part by making holes in the brain.
- Malformed proteins called prions trigger the disease. The extreme durability of prions and CWD's long incubation times make controlling the spread of the sickness difficult.
- Studies are under way to see whether CWD can infect humans and livestock.



CHRONIC WASTING DISEASE roughens the coat of an infected elk, which also shows early signs of emaciation (*top left*). The illness, which has appeared in 14 states and provinces (*map*), fills brain tissue with holes (*white areas in micrograph*). In a temporary lab near Black Earth, Wis. (*above*), a researcher extracts tissue samples from the head of a white-tailed deer killed in the state's eradication zone.



COURTESY OF WYOMING GAME AND FISH DEPARTMENT (*elk*); DAVID NEVALA (*sampling station*); ELIZABETH WILLIAMS *University of Wyoming* (*micrograph*); CORNELIA BLIK (*map*)

killed off all the resident deer and elk; then they turned several inches of soil and repeatedly sprayed structures and pastures with swimming-pool chlorine, which readily wipes out bacteria and viruses. After waiting a year, they brought in 12 elk calves, but a few years afterward two of those elk contracted CWD.

The disease's persistence has permanently contaminated an area of about 15,000 square miles in northeastern Colorado, southeastern Wyoming and (beginning in 2001) southwestern Nebraska. The incidence of CWD among the cervids in this so-called endemic area averages about 4 to 5 percent but has reached 18 percent in some places. To help keep the disease confined here, the research facilities stopped trading captive animals with each other. In fact, no captive cervids now leave the endemic area alive: "They're only allowed out to come to my necropsy room," wryly remarks Williams, now at the University of Wyoming. More important were the mountains and other natural barriers, which scientists expected would keep CWD from spreading rapidly out of the endemic area. There was, however, an easy way past those natural barriers: along the roads, in a truck.

Out and About

SOME 11,000 GAME FARMS and ranches holding hundreds of thousands of deer and elk dot the U.S. and Canada. Besides harvesting the meat, ranchers can sell the antlers—those from elk are marketed as a supplement in vitamin stores ("velvet antler") and as an aphrodisiac in Asia ("velvet Viagra"). To start such farms, ranchers must buy breeding cervids. Somewhere along the line, businesses must have picked up incubating animals from the endemic area. And the interstate trade of cervids continued the spread, west across the Continental Divide and east across the Mississippi River. (These days most states regulate such trade.)

The first farmed cervid to display signs of CWD was an elk that fell ill in 1996 on a ranch in Saskatchewan. By 2001 some 20 ranches reported cases across six states (Colorado, Kansas, Montana, Nebraska, Oklahoma and South Dakota) and one other Canadian province (Alberta). Quick, aggressive measures—namely, killing off the herds—appear to have eliminated the problems on the ranches.

Nevertheless, the transport of incubating cervids may have carried CWD to wild populations in those states and beyond—such as to white-tailed deer in Wisconsin's eradication zone. But precisely when and how mule deer gave it to white-tailed deer, the most common type in the eastern U.S., is unknown and may never be clear. "By the time these problems are discovered," Miller says, "they have probably been sitting there for decades, which makes it difficult to go back and retrace how things came about." Based on epidemiological models and on Wisconsin's roughly 1.6 percent incidence in the eradication zone, Miller thinks CWD had probably been lurking there since the early 1990s.

Wisconsin's approach makes sense to scientists studying prion diseases. "The

DAVID NEVALA

DEER CARCASSES from March 2003 hunts in Wisconsin's eradication zone were loaded into a refrigerated storage unit. The heads were sent for CWD testing; the bodies were incinerated.

idea is to find a fairly small focus and get rid of all the animals in the area," in the hopes of preventing CWD from attaining a permanent hold in the region, Williams says. A rapid spread is possible in Wisconsin because the deer population in the state's southwestern corner is dense: Thomas Givnish, an expert on the ecology of diseases at the University of Wisconsin–Madison, notes that it runs about 50 to 100 deer per square mile, or 10 times that of the endemic area around Fort Collins. "The alternative is to do nothing," Williams observes, and then "you know it's going to be established." By the end of March, Wisconsin hunters had bagged 9,287 deer—which will cut the fall population by 25 percent but will not eliminate CWD, notes state wildlife biologist Tom Howard. A few more seasons of liberal hunting may be needed.

Considering the persistence of prions, Wisconsin may have to live with CWD, as Colorado does. "The disease has been here a long time," Miller comments of CWD around northeastern Colorado. "We can't get rid of it here. We try to get infection rates down so that it can't spread." Miller says that Colorado had hoped to purge CWD through culling. But "we discovered we were 10 to 20 years too late. It was already out there; we didn't realize it." That statement may apply to other states that have found CWD among wild deer, including Illinois and New Mexico.

Venison and Beyond

NO ONE KNOWS whether CWD can pass to humans. A test tube study mixed CWD prions with normal prion proteins from cervids, humans, sheep and cows. The CWD prions had a hard time converting normal human PrP—less than 7 percent of the protein was changed. The downside is that CWD prions converted human PrP about as efficiently as BSE prions do. And because BSE has infected humans, CWD might pose a similar risk. But because beef is far more popular than venison, CWD doesn't present quite the same public health threat.

To see if CWD has already infected people, the Centers for Disease Control and Prevention investigated the deaths of the three young venison eaters who succumbed to sporadic Creutzfeldt-Jakob disease. All were younger than 30 years, which is exceedingly rare in CJD. In fact, through May 31, 2000, just one other U.S. case of sporadic CJD occurred in this age group since surveillance began in 1979.

The first was a 28-year-old cashier, who died in 1997; she had eaten deer and elk as a child, from her father's hunts in Maine. The second was a 30-year-old salesman from Utah who had been hunting regularly since 1985 and who died in 1999. The third was a 27-year-old truck driver from Oklahoma who died in 2000; he had harvested deer at least once a year. Tests of the 1,037 deer and elk taken during the 1999 hunting season from the regions where the victims' meat originated all turned up CWD negative (none of the meat came from the endemic area). The victims' brains showed no unique damage or distinct biochemical signs, as is the case with other prion diseases in humans.

Six other patients (all at least middle-aged) raised suspicions about the CWD risk to humans. Three were outdoorsmen from the Midwest who had participated in wild deer and elk feasts and died in the 1990s. The other cases were reported in April and include two from Washington State who hunted together. Researchers, however, could not find any connection with CWD. And states with CWD have not discovered a higher incidence of Creutzfeldt-Jakob disease.

These observations may seem reassuring, but it is too early to conclude that CWD does not pose a human health hazard. The incubation period of prion diseases may span upward of 40 years, and CWD has been spreading noticeably in only the past 10. The rarity of prion diseases and the low national consumption of deer and elk (compared with beef) make it hard to draw any firm conclusions. Because of the uncertainties with CWD and the fact that animal prion diseases have jumped to humans, the CDC warns against eating food derived from any animal with evidence of a spongiform encephalopathy.

Scientists are still trying to determine if CWD poses a threat to livestock. In an ongoing experiment begun in 1997, Hamir and his colleagues injected brain suspensions from CWD mule deer into the brains of 13 Angus beef calves. Two became ill about two years after inoculation, three others nearly five years after. Hamir began repeating the experiment in November 2002, this time with the brains of CWD white-tailed deer.

Under more natural conditions, bovines have not contracted CWD. Williams has kept cows with infected cervids, and more than five years on, the cows are still healthy. Bovines kept with decomposing CWD carcasses or isolated in pens that once housed CWD animals have also remained free of prion disease. (These reports are good news for pasture-grazing cows, which might find themselves in the company of wild deer.) To see whether CWD might pose a danger when eaten, Williams has begun feeding CWD brain matter to calves. The long incubation of these illnesses, however—BSE incubates for up to eight years—means these experiments must continue for several years.

If U.S. livestock so far seem to be safe from CWD, the same cannot be said of other animals. If an infected deer dies in the forest and nobody is there to see it, plenty of coyotes, bobcats and other carnivores will, and they will gladly scavenge what remains of the wasted carcass. Moreover, during the clinical phase, CWD animals undoubtedly make easier prey. The canine family is evidently immune to prion diseases, but felines can contract them. Transmission studies with mountain lions have begun, and local lions that die for unknown reasons end up on the pathology table, Williams says.

It is too early to conclude that chronic wasting disease does not pose a HUMAN HEALTH HAZARD.

Although many states have uncovered CWD, other states "are looking darn hard," Miller says, but have not found it—among them, Arizona, Kansas, Michigan, Montana, Nevada and New Jersey. Apparently, only pockets of outbreaks exist. Wildlife managers therefore have a fighting chance to keep CWD from gaining a permanent grip throughout the country, so long as control efforts begin promptly. Unfortunately, not all states with CWD are as aggressive as Wisconsin when it comes to surveillance and eradication.

To stop or at least slow the spread of the fatal sickness, extensive culling appears to be the best strategy. One could hope that CWD occurs naturally in deer and that the epidemics will run their course and leave behind CWD-resistant cervids. Some lines of sheep, for instance, are immune to scrapie. But so far all white-tailed and mule deer appear to be uniformly susceptible. "I don't think genetics is going to save us on this," remarks the NIH's Race. Sadly, the only way to save the deer, it seems, is to shoot them. SA

Philip Yam is Scientific American*'s news editor. This article is adapted from his book,* The Pathological Protein: Mad Cow, Chronic Wasting, and Other Deadly Prion Diseases, *published in June.*

MORE TO EXPLORE

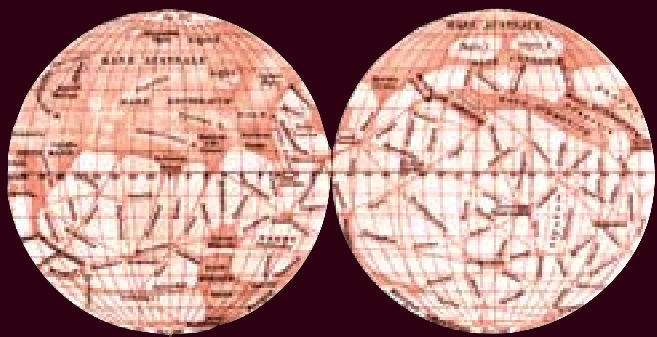
Risk Analysis of Prion Diseases in Animals. Edited by Corinne I. Lasmézas and David B. Adams. *Scientific and Technical Review,* Vol. 22, No. 1; April 2003.

The Pathological Protein: Mad Cow, Chronic Wasting, and Other Deadly Prion Diseases. Philip Yam. Copernicus Books, 2003. (**www.thepathologicalprotein.com**)

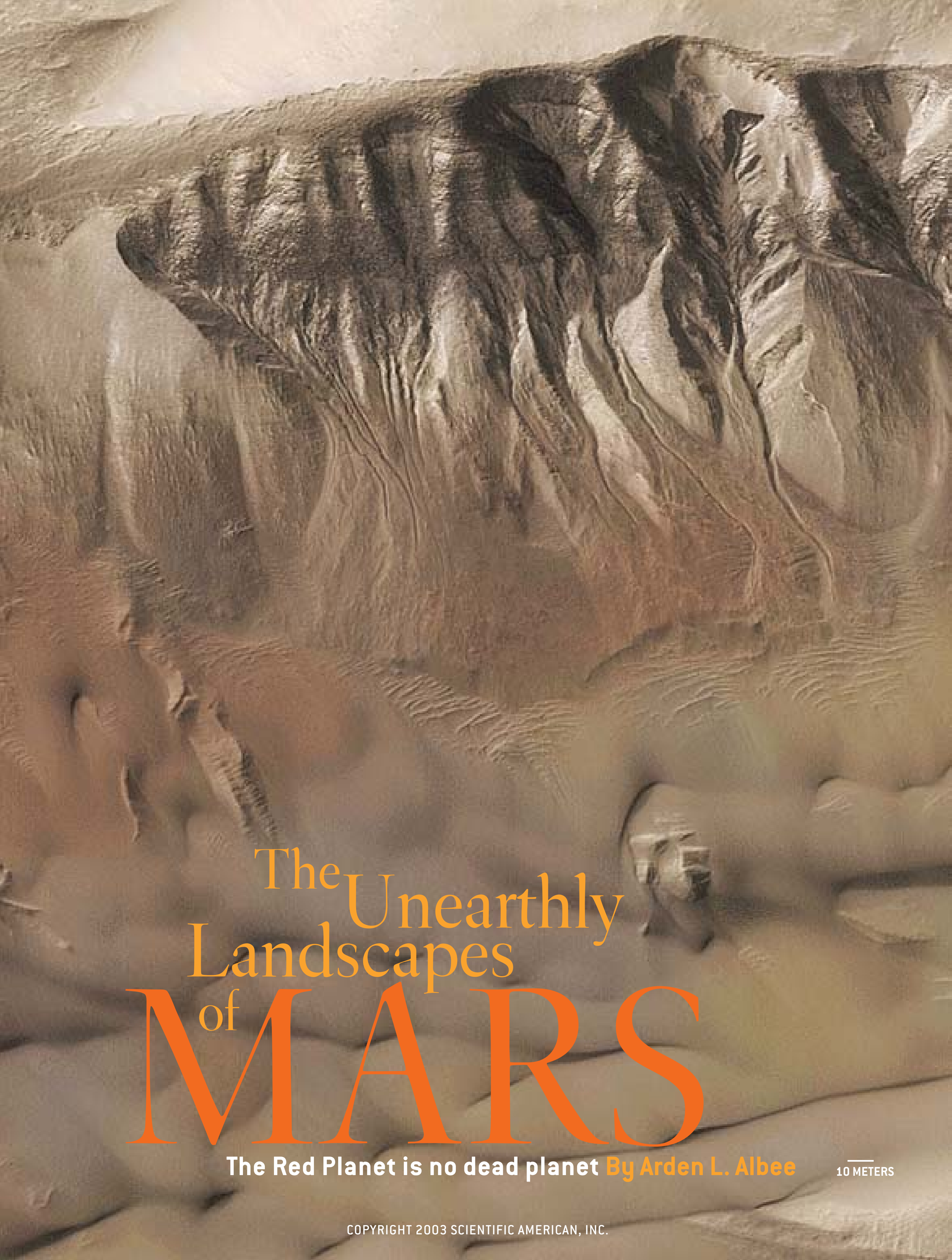
Chronic Wasting Disease Alliance Web site is at **www.cwd-info.org/**

U.S. Department of Agriculture Web site on CWD is at **www.aphis.usda.gov/lpa/issues/cwd/cwd.html**

Captain John Carter, the hero of the adventure novels of Edgar Rice Burroughs, was a gentleman of Virginia and an officer of the Confederacy. Impoverished after the Civil War, he went looking for gold in Arizona and, while being chased by Apache warriors, fell and struck his head. He returned to consciousness on an arid planet with twin moons, populated by six-legged creatures and beautiful princesses who knew the place as "Barsoom." The landscape bore an uncanny resemblance to southern Arizona. It was not entirely dissimilar to Earth, only older and decayed. "Theirs is a hard and pitiless struggle for existence upon a dying planet," Burroughs wrote in the first novel.



IT WOULD TAKE YOU ABOUT FIVE MINUTES to hike across the area shown in this image, on the north side of Newton Crater in the southern hemisphere of Mars. You would leave your footprints on lightly frosted soil (*bright areas*), clamber over windblown features such as sand dunes and jump across possibly water-carved features such as gullies. These landforms probably continue to form even today. Like other Mars Global Surveyor images, this one is a composite of high-resolution grayscale and low-resolution color; the colors are only approximate. It is a far cry from the vague (and often fanciful) view of Mars a century ago (*above*).



The Unearthly Landscapes of MARS

The Red Planet is no dead planet By Arden L. Albee

In science as well as science fiction, Mars is usually depicted as a version of Earth in its extreme—smaller, colder, drier, but sculpted by basically the same processes. Even well into the 20th century, many thought the planet had flowing water and proliferating plants. The resemblance to Earth fell apart when spacecraft in the late 1960s revealed a barren, cratered world, more like the moon. But it quickly returned with the subsequent discoveries of giant mountains, deep canyons and complex weather patterns. The Viking and Mars Pathfinder images from the surface look eerily Earth-like. Like Burroughs, researchers compare the equatorial regions of Mars to the American Southwest. For the polar regions, the model is the Dry Valleys of Antarctica, a frozen desert in a landscape of endless ice.

But if there is one thing researchers have learned from recent Mars exploration, it is to be careful about drawing such comparisons. In the past five years, spacecraft have collected more information about the Red Planet than all previous missions combined. Mars has proved to be a very different and more complicated planet than scientists thought beforehand. Even the single biggest question—Was Mars once warm and wet, possibly hospitable to the evolution of life?—is more nuanced than people have tended to assume. To make sense of Mars, investigators cannot be blinded by their experience of Earth. The Red Planet is a unique place.

Mars as the Abode of Dust

MARS EXPLORATION has certainly had its up and downs. In the past decade NASA has lost three spacecraft at Mars: Mars Observer, Mars Climate Orbiter (intended as a partial replacement for Mars Observer) and Mars Polar Lander. Lately, though, the program has had a run of successes. Mars Global Surveyor has been taking pictures and collecting infrared spectra and other data continuously since 1997. It is now the matriarch of a veritable family of Mars spacecraft. Another, Mars Odyssey, has been orbiting the planet for more than a year, mapping the water content of the subsurface and making infrared images of the surface. As this article goes to press, NASA plans to launch the Mars Exploration Rovers, successors to the famous Sojourner rover of Mars Pathfinder [*see box on page 50*]. Around the same time, the European Space Agency expects to send off the Mars Express orbiter, with its Beagle 2 lander. The Nozomi orbiter, sent by the Institute of Space and Astronautical Science in Japan, should arrive at Mars in December.

Never before have scientists had such a comprehensive record of the processes that operate on the surface and in the atmosphere [*see box on page 52*]. They have also studied the craters, canyons and volcanoes that are dramatic relics of the distant past. But there is a huge gap in our knowledge. Between ancient Mars and modern Mars are billions of missing years. No one is sure of the conditions and the processes that sculpted Mars during most of its history. Even less is known about the subsurface geology, which will have to be the subject of a future article.

Present-day Mars differs from Earth in a number of broad respects. First, it is enveloped in dust. Much of Earth's surface consists of soil derived by chemical weathering of the underlying bedrock and, in some regions, glacial debris. But much of Mars's surface consists of dust—very fine grained material that has settled out of the atmosphere. It drapes over all but the steepest features, smothering the ancient landscape. It is thick even on the highest volcanoes. The dustiest areas correspond to the bright areas of Mars long known to telescope observers.

LAYERED TERRAIN looks surreal, almost like a topographic map, but is quite real. It covers the floor of western Candor Chasma, a ravine that is part of the Valles Marineris canyon system. Scientists have identified 100 distinct layers, each about 10 meters thick. They could be sedimentary rock originally laid down by water, presumably before the canyon cut through the terrain. Alternatively, the layers could be dust deposited by cyclic atmospheric processes. This image was taken by Mars Global Surveyor.

Dust produces otherworldly landscapes, such as distinctively pitted terrain. As dust settles through the atmosphere, it traps volatile material, forming a mantle of icy dust. Later on, the volatile ices turn to gas, leaving pits. Intriguingly, the thickness of the icy, dusty mantle on Mars varies with latitude; near the poles, Mars Odyssey has shown, as much as 50 percent of the upper meter of soil may be ice. On slopes, the icy mantle shows signs of having flowed like a viscous fluid, much in the manner of a terrestrial glacier. This mantle is becoming the focus of intense scientific scrutiny.

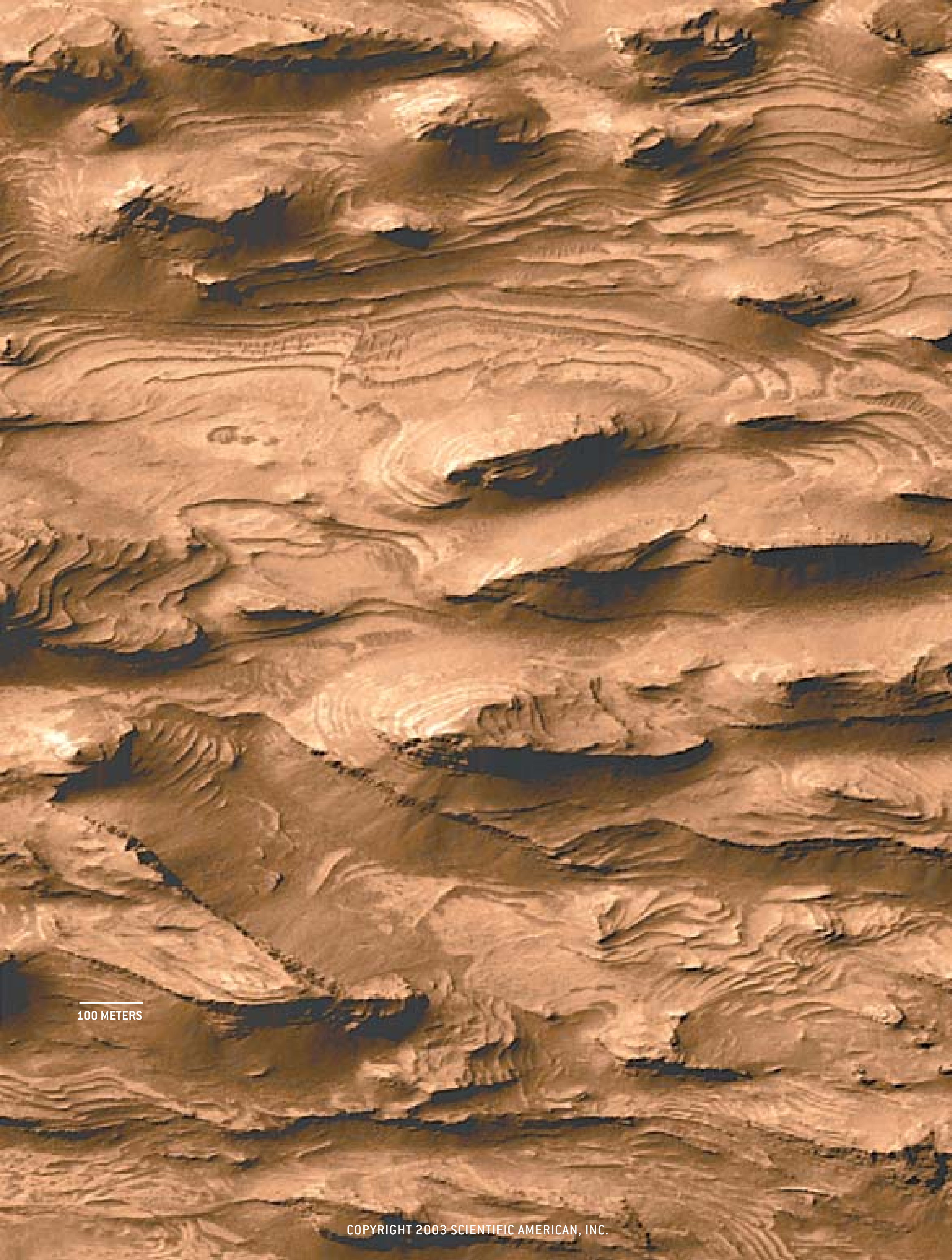
Second, Mars is extremely windy. It is dominated by aeolian activity in much the way that Earth is dominated by the action of liquid water. Spacecraft have seen globe-encircling dust storms, huge dust devils and dust avalanches—all wrought by the wind. Dust streaks behind obstacles change with the seasons, presumably because of varying wind conditions.

Where not dust-covered, the surface commonly shows aeolian erosion or deposition. Evidence for erosion shows up in craters, from which material appears to have been removed by wind, and in yardangs, bedrock features that clearly have been carved by windblown sand. Evidence for deposition includes sand sheets and moving sand dunes. The latter are composed of sand-size grains, which move around by saltation—multiple bounces of

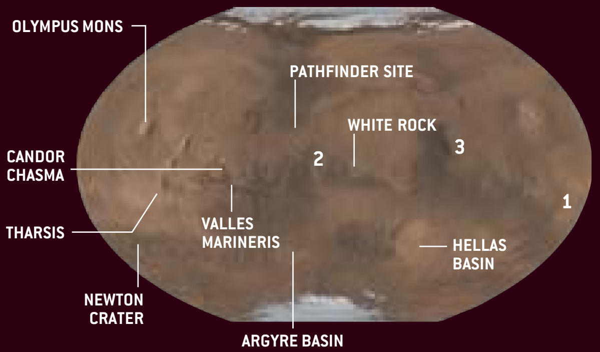
Overview/*The Martian Surface*

- Two ongoing missions to Mars, Mars Global Surveyor and Mars Odyssey, are raising difficult questions about the Red Planet. Flowing water, ice and wind have all helped carve the landscape over the past several billion years. The processes are both similar and dissimilar to those acting on Earth's surface. Scientists' experience of Earth have sometimes led them astray.
- The question of whether Mars was once hospitable is more confusing than ever. Spacecraft have gathered evidence both for and against the possibility. Three upcoming landers, two American and one European, could prove crucial to resolving the matter.

COURTESY OF www.nirgal.net/schemas/schiaparelli_map.html (*page 44*); NASA/JPL/MALIN SPACE SCIENCE SYSTEMS MOC 2–320 (*page 45*); NASA/JPL/MALIN SPACE SCIENCE SYSTEMS MOC 2–259 (*opposite page*)

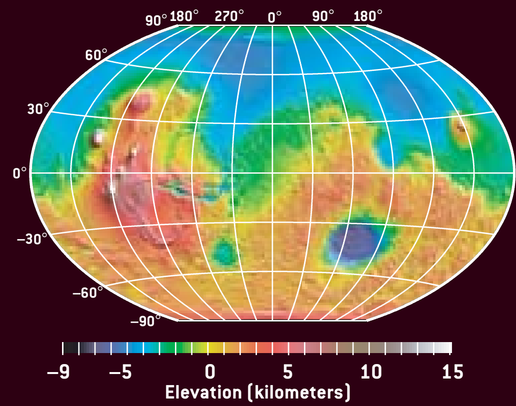

100 METERS

GLOBAL VIEWS of MARS

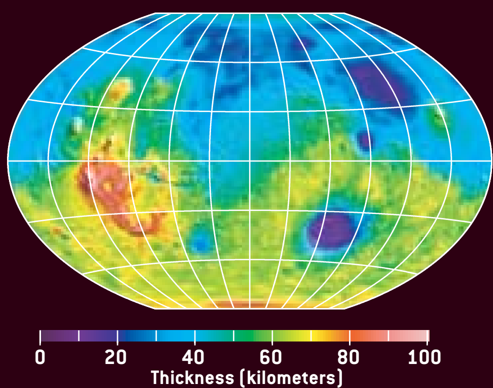


Landing Sites: Gusev Crater (1), *Meridiani Planum* (2), *Isidis Planitia* (3)

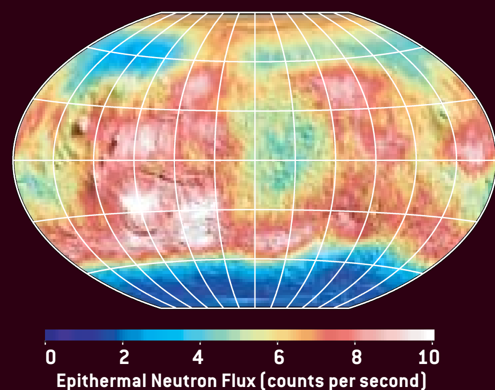
TRUE COLOR Mars is four worlds in one: the heavily cratered southern hemisphere (with riverlike valley networks), the smoother northern hemisphere (with hints of an ancient shoreline), the equatorial region (with giant volcanoes and canyons), and the polar caps (with bizarre, protean terrain). This map combines wide-angle camera images with altimetry, which brings out details. The color is realistic.



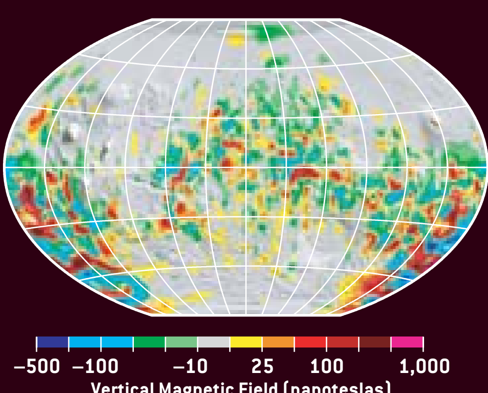
TOPOGRAPHY The elevation spans 30 kilometers from the lowest basins (*dark blue*) to the highest volcanoes (*white*). For comparison, the range of elevation on Earth is only 20 kilometers. The large blue circle in the southern hemisphere represents the Hellas impact basin, one of the biggest craters in the solar system. Girdling Hellas is a vast ring of highlands about two kilometers in elevation.



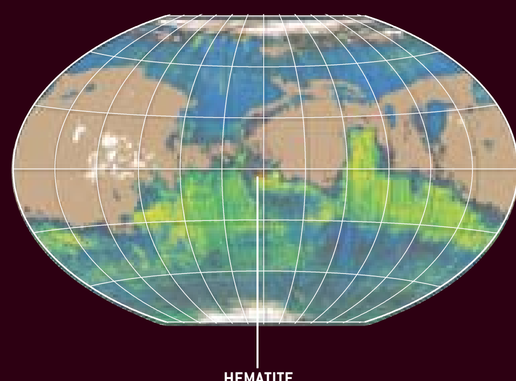
CRUSTAL THICKNESS Combining the topographic map with measurements of Mars's gravity, researchers have deduced the thickness of the Martian crust: roughly 40 kilometers under the northern plains and 70 kilometers under the far southern highlands. The crust is especially thick (*red*) under the giant Tharsis volcanoes and thin (*purple*) under the Hellas impact basin.



WATER Neutrons reveal the presence of water in the top meter of soil. The energy of these particles, which are produced when cosmic radiation bombards the soil, is sapped by the hydrogen within water molecules. A dearth of medium-energy ("epithermal") neutrons means water-rich soil (*blue*). The implied amount of water, most of it in the far south, would fill two Lake Michigans. More may lie deeper underground.



MAGNETISM Mars lacks a global magnetic field, yet areas of its crust are magnetized up to 10 times as strongly as Earth's crust. In these areas, iron-rich rocks have become bar magnets, suggesting that Mars had a global field at the time the rocks solidified from a molten state. The east-west banding resembles patterns produced by plate tectonics on Earth, but its origin is unknown.



GEOLOGY Infrared spectral measurements reveal rock types. Basalt (*green*), a primitive volcanic rock, dominates the southern hemisphere. Andesite (*blue*), a more complex volcanic rock, seems to be common in the north. Near the equator is an outcrop of hematite (*red*), a mineral typically produced in the presence of water. In large regions, dust (*tan*) or clouds (*white*) hide the underlying rock types.

MARS ORBITER LASER ALTIMETER SCIENCE TEAM, MALIN SPACE SCIENCE SYSTEMS (*true color mosaic*); ODED AHARONSON *California Institute of Technology* (*all other mosaics*);
SOURCE: MARS ODYSSEY MISSION (*water*); PHILIP CHRISTENSEN *Arizona State University* (*geology*); MARS GLOBAL SURVEYOR MISSION (*all other mosaics*)

windblown grains along the ground. It takes a stronger wind to loft dust directly than to initiate saltation, so this phenomenon accounts for most of the dust kicked into the atmosphere.

Aeolian activity seems to have persisted since the time of heavy cratering, back when the solar system was still young. Many images show craters with varying degrees of erosion: some are shallow and partially filled with deposits and sand dunes, whereas others are pristine—deeper and bowl-shaped. Michael Malin and Kenneth Edgett of Malin Space Science Systems in San Diego, the research firm that operates the Mars Global Surveyor camera, have inferred a sequence of processes: Sand was blown through the region, and some of it got trapped in craters; other craters formed later. Where and how such a volume of sand was produced and how it was blown around remain a mystery, however.

The Angry Skies of Mars

A THIRD WAY in which Mars differs from Earth is in its amazing variety of weather and climate cycles, many of which are similar to those on Earth, many like nothing on Earth. The Martian day is almost the same as an Earth day, but the Martian year is 687 Earth days. The tilt of Mars's rotation axis, which produces seasons, is very close to that of Earth's. Mars lacks the precipitation and oceans that are so crucial to weather on Earth. But the atmospheric pressure (less than 1 percent of that on Earth) varies seasonally by about 25 percent, driven by the condensation and sublimation of carbon dioxide frost at the poles. The thin atmosphere has a very low heat capacity, so the surface temperature swings by more than 100 degrees Celsius from day to night. The thermal properties of the thin atmosphere are dramatically affected by dust and ice particles suspended in the air. The upshot is that, despite being so thin, the atmosphere has complex circulation patterns and dynamics. A daily weather report might talk of strong winds, high-level ice clouds, low-level fogs, seasonal frost, dust devils and massive dust storms.

As on Earth, storm systems often spiral southward from the northern polar regions. But the largest dust storms typically start during the southern spring as the planet rapidly heats up. Periodically they coalesce and come to encircle the entire planet. Mars Global Surveyor closely followed the evolution of a four-month global dust storm that started in June 2001. Contrary to scientists' expectations, it was not, in fact, one single global storm, but the confluence of several regional storms. Malin has compared the climatic effect of the dust raised by this storm with the aftermath of Mount Pinatubo's eruption on Earth in 1991—namely, a brief but widespread cooling.

The polar ice caps play a key role in the atmospheric cycles. Their size and shape, as shown by topographical measurements, indicate that the caps are predominantly water ice, as opposed to so-called dry ice, made of carbon dioxide: dry ice is not as rigid as water ice, and it could not support the observed domelike shape. A major new discovery has been that the layer of dry ice that covers much of the south polar cap is being eroded away at a high rate. Clearly, the erosion cannot go on forever. Nor can the current dust sinks and sources remain in their current states indefinitely. To replenish the ice and dust, other cycles must be occurring, perhaps tied to orbital variations. Malin and Edgett have suggested that wind conditions may be less intense now than in the fairly recent past, another hint that the Martian climate changes with time.

A fourth major difference between Earth and Mars is the behavior of liquid water. Liquid water is unstable at the surface under present pressure and temperature conditions. It does not rain. Still, water ice can—and does—persist at some depth within the Martian soil during all or much of the year. On Mars, as on Earth, several types of patterned ground mark the presence of ice-rich soil. Mars Odyssey has detected ground ice over most of the planet outside the equatorial regions, and models predict that the ice extends to considerable depths.

Liquid water can sometimes leak onto the surface. In 2000 Malin and Edgett described fresh gullies that look like water-carved features on Earth [see "Gully Gee Whiz," by George Musser; SCIENTIFIC AMERICAN, September 2000]. In the ensuing excitement, researchers advanced many theories to explain them: leaking aquifers (which would be inexplicably perched high on crater rims); pressurized geysers of water; high-pressure outbursts of carbon dioxide gas; volcanic heat sources at depth. Finally, earlier this year Philip Christensen of Arizona State University discovered gullies that clearly emerge from underneath a bank of snow and ice. He concluded that they are related to Martian climate cycles. In colder periods, slopes become blanketed with a mixture of snow and dust. Sunlight penetrates this insulating blanket, heating it enough for water to melt under the snow and to run down the slope, creating gullies. In warmer periods, the blanket melts or evaporates entirely, exposing the gullies.

Layer upon Layer

DESPITE THE ABUNDANCE of water, however, Mars is arid. It has the mineralogy of a nearly waterless surface. On Earth, the action of warm liquid water produces weathered, quartz-rich soils, hydrated clays, and salts such as calcium carbonate and sulfate. Beach sand and sand dunes are largely quartz. On Mars, spacecraft have yet to find any deposits of these minerals. The darker Martian dunes are basaltic, consisting mainly of minerals such as pyroxene and plagioclase, which on Earth would readily weather away. It follows that the present cold and dry atmospheric conditions have persisted since a time far back into the planet's history.

Has Mars always differed so much

THE AUTHOR

ARDEN L. ALBEE is the project scientist—that is, overall leader of the science team—for the Mars Global Surveyor mission. He had the same role for the ill-fated Mars Observer mission, for which Surveyor is a partial replacement. Albee is emeritus professor of geology and planetary science at the California Institute of Technology and served as chief scientist of NASA's Jet Propulsion Laboratory from 1978 until 1984. His research interests run from field geology to compositional analysis of rocks, meteorites, comets and lunar samples. He still makes time for his eight children and 11 grandchildren.

from Earth? Below the mantles of dust and sand are numerous signs that the Red Planet has transformed over time. To begin with, the planet has a striking dichotomy in landscape between its northern and southern hemispheres. The southern hemisphere is high in altitude and heavily cratered (indicating an ancient surface). The northern one is a vast, low-lying plain with fewer craters (indicating a younger age). In between is the immense Tharsis Plateau, intermediate in age and capped by volcanoes that dwarf any on Earth. Using the new high-resolution data on these volcanoes, James W. Head III of Brown University has found flow patterns that look strikingly like mountain glaciers—and that may suggest the presence of ice under a blanket of rock and dust.

The northern lowlands are exceedingly level, leading to speculation that they were lake beds during a significant chunk of Martian history. They appear to be covered with multiple layers of volcanic flows and sedimentary debris that originated in the south. Detailed new topographic maps have unveiled "stealth craters"—faint circular expressions, evidently part of an ancient cratered surface that has been buried by a thin layer of younger deposits.

Along the edge of the southern highlands are features that could only have been carved by liquid water. These features are tremendously larger than their counterparts on Earth. The famous canyon Valles Marineris would run from Los Angeles to New York with a width extending from New York to Boston and a depth similar to the elevation of Mount McKinley. No terrestrial canyon comes close. At its head is a jumbled terrain, intimating that water flowed not in a steady trickle but in concentrated, catastrophic outflows, scouring the surface along its path. Other Martian outflow channels have similar features. Because these features are carved into the Tharsis Plateau, they must have an intermediate age.

Streamlined islands and other features in these channels look much like the scablands in the northwestern U.S., which were gouged by the Spokane Flood toward the end of the last ice age, about 10,000 years ago. During the massive deluge, a lake roughly the size of one of the Great Lakes burst its ice dam and rushed out within just a few days. On Mars, such calamities were 10 to 100 times as devastating. They may have been triggered by volcanic heat sources or by the general heat flow from the interior of the planet. Heat would have melted ice underneath the thick permafrost layer, building up tremendous pressures until the water finally burst out.

The most contentious water-related features of all are the valley networks. Located throughout the southern highlands, they have a branching, dendritic pattern reminiscent of rivers on Earth—suggesting that they were formed by surface runoff from rainfall or snowfall. They are the strongest hint that Mars was once as warm as Earth. But these networks look rather different from rain-fed rivers on Earth. They more closely resemble river networks in desert areas, which are fed by water that slowly seeps from subterranean sources. Such streams typically originate in steep-walled amphitheaters rather than in ever smaller tributaries. Heated debates have been taking place at scientific meetings over the crucial question: Did it rain on early Mars?

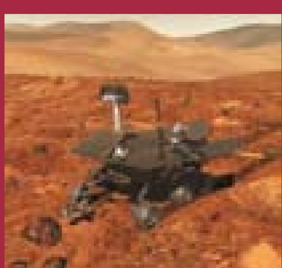
The timing of the water networks could be the key to making sense of them. Recent detailed studies of the northern edge of the highlands show that immense amounts of material eroded during—rather than after—the intense meteor bombardment that took place early in Martian history. These analyses imply that the distribution of water kept changing as impacts reworked the landscape. Craters filled with water and debris, and channels began to link them together into

Going for a Drive

THIS MONTH NASA PLANS to launch twin rovers to the Red Planet, and the European Space Agency will send off a lander, too. The three of them, scheduled to arrive next January, will be robotic geologists—studying the geologic history of landing sites, investigating what role water played there and determining how suitable past conditions would have been for life.

The rovers deserve particular mention, because they will give scientists unprecedented mobility. Each rover can travel 100 meters a day. For comparison, the Sojourner rover on the 1997 Mars Pathfinder lander traversed 100 meters over its entire mission [see "The Mars Pathfinder Mission," by Matthew Golombek; Scientific American, July 1998]. A mast, about 1.5 meters high, supports a binocular camera and a thermal-emission spectrometer, one of many instruments that can analyze the composition of materials. A robot arm holds the other instruments: a Mossbauer spectrometer, an alpha-particle detector, an x-ray spectrometer and a microscope. The arm also carries a scraper to clean off rock surfaces for study. A dish antenna beams signals directly to Earth, and a black rod antenna relays data through the Mars Global Surveyor or Mars Odyssey orbiter.

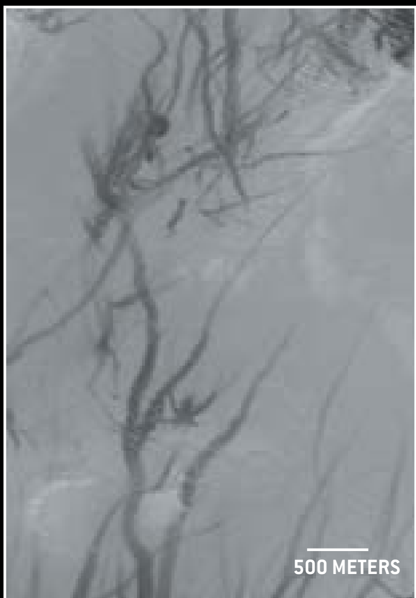
Deciding on their landing sites has been one of the most exciting uses of data from the orbiters. Unlike the selection of sites for the Viking landers in 1976—which was a matter of a few people's gut instinct—the choice for the rovers involved a long deliberation among dozens of scientists and engineers. Weighing tantalizing geology (such as suspected water-related features) against potential dangers (such as steep slopes and high winds), they winnowed an initial list of more than 150 possible sites down to seven, then four and finally, on April 11, two: Gusev Crater, whose layered deposits might be lake-bed sediments, and Meridiani Planum, which is rich in coarse-grained hematite, a mineral typically formed in association with liquid water. (The European Beagle 2 lander will touch down on Isidis Planitia, a possible sedimentary basin.) *—A.L.A.*

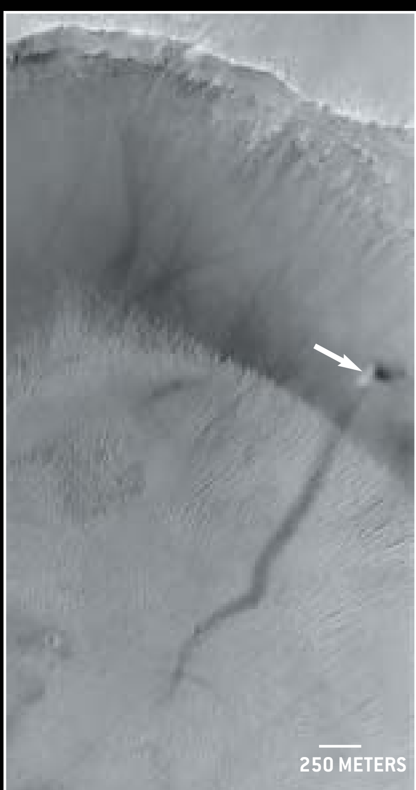


NASA's Mars Exploration Rover

NASA ARTIST CONCEPT/MARS EXPLORATION PROGRAM

IN THE PLAINS northwest of Olympus Mons, dust devils are sweeping over the land and leaving streaks in their wake (*right*). A similar scene has unfolded in the Argyre Basin (*below*) and east of Valles Marineris (*bottom*), where a devil was caught in the act. These tornadolike vortices—thought to be created as warm air rises off the surface—clear away light-colored dust and expose comparatively dark soil. They are one of the many wind-related processes that are continuously reshaping the Martian surface. The picture at the right is from Mars Odyssey; the two below are from Mars Global Surveyor.





NASA/ARIZONA STATE UNIVERSITY THEMIS SCIENCE TEAM, FEBRUARY 6, 2003 (*right*); NASA/JPL/MALIN SPACE SCIENCE SYSTEMS MOC 2-220 (*left top*) AND MOC 2-318 (*left bottom*) a network, but impacts continually disrupted this process. For instance, the Argyre Basin, 1,000 kilometers in diameter, may once have been filled to its brim with water. It is part of a valley system that brought water from near the South Pole, through the basin, into channels that crossed the equator. The roles of water and ice in these systems, both aboveground and underground, remain unclear. In any case, these networks are very different from hydrologic systems on Earth.

A final clue to Martian history comes from one of the biggest surprises delivered by Mars Global Surveyor, the extent to which the uppermost crust consists of layered deposits. Almost everywhere that the subsurface is exposed—on walls of canyons, craters, mesas and valleys—it is layered. The layers differ from one another in thickness, color and strength. They show that the Martian surface has undergone complex sequences of deposition, crater formation and erosion. The oldest layers are the most extensive. The higher layers have been partially stripped away, apparently blown by the wind.

Where did the layers come from? The lack of boulderlike blocks argues against their being volcanic flows, although they could be volcanic ash. Ultimately, however, most of the layers probably originated in impact debris. On the moon, scientists observe overlapping rings of impact debris, which mark craters of differing ages. Similarly, Mars is so heavily cratered that the upper crust has been stirred up like soil tilled by a gardener. Water and wind then scattered this material.

Blue Mars?

IN A SENSE, scientists' ideas about early Mars are more uncertain than they have ever been. This doubt comes to the fore when researchers address the question of liquid water. The presence or absence of liquid water is fundamental to geologic processes, climate change and the origin of life. The early valley networks and the later flood channels attest to an abundance of water. The evidence for early rainfall suggests that the atmosphere was once much denser. But spacecraft have found no evidence for deposits of carbonate minerals, which would be the vestiges expected from an early dense carbon dioxide atmosphere [see "The Climate of Mars," by Robert M. Haberle; SCIENTIFIC AMERICAN, May 1986].

At this point, scientists have three main hypotheses. Perhaps the early atmosphere was indeed thick. The planet might have had lakes, even oceans, free of ice. Robert A. Craddock of the National Air and Space Museum and Alan D. Howard of the University of Virginia recently suggested that the carbon dioxide was lost to space or locked up in carbonate minerals that have so far escaped detection. Intriguingly, Mars Odyssey spectra have revealed small amounts of carbonate in the dust.

Alternatively, perhaps Mars had a fairly thin atmosphere. It was a wintry world. Any standing bodies of water were covered

Monitoring Mars, 24/7/687

BOTH MARS GLOBAL SURVEYOR and Mars Odyssey circle the planet in paths that take them over both poles. Their orbits remain fixed as Mars spins below them, allowing the instruments to observe day and night swaths over the entire planet. This continuous coverage can track changes in the surface, atmosphere, gravity and magnetic field.

Global Surveyor has five main instruments. Its laser altimeter has measured the overall shape and the topography of Mars with an altitude precision of about five meters, which means that Mars is now better mapped than most of Earth. The camera takes red and blue medium-resolution images of the entire surface, as well as high-resolution images—1.4 meters per pixel, as good as the pictures taken by the spy satellites of the 1960s—of limited areas. A Michelson interferometer measures the emitted thermal infrared spectrum with high spectral resolution but low spatial resolution, suitable for mapping the mineral composition and thermal properties of the surface. A magnetometer determines the magnetic field. Finally, the spacecraft itself counts as an instrument, because its motion is sensitive to variations in Martian gravity. The gravitational field reveals the thickness of the crust and changes in the size of the polar ice caps.

Odyssey complements Global Surveyor. Its camera lacks a high-resolution mode but takes images in five selected color bands. Its infrared imager has low spectral resolution but high spatial resolution. Another instrument measures gamma-ray and neutron fluxes, which are sensitive to hydrogen just below the planet's surface; Odyssey is therefore the first spacecraft capable of peeking under the surface of Mars, to a depth of about one meter.

These spacecraft also monitor the atmosphere. Cameras scan the entire planet daily, much like Earth-orbiting weather satellites. Twelve times a day, the thermal-emission spectrometer takes readings of temperature, pressure, cloud cover and dust abundance. Additionally, as radio transmissions pass through the Mars atmosphere, they are diffracted; signal processing can infer the variation of temperature and pressure with altitude.

—A.L.A.

MARS GLOBAL SURVEYOR (artist's conception)

NASA ARTIST CONCEPT/GLOBAL SURVEYOR PROGRAM in ice. Snow might have fallen, recharging the groundwater and leading to temporary trickles of water across the surface. Steven M. Clifford of the Lunar and Planetary Science Institute in Houston, among others, has conjectured that melting under a glacier or a thick layer of permafrost could also have recharged subterranean water sources. Although Mars was bitterly cold, periodic bursts of relatively warmer temperatures could have reinvigorated the planet. Orbital shifts, similar to those that trigger ice ages on Earth, drove these climate cycles. Head, John F. Mustard of Brown and others have pointed to the latitude dependence of the ice and dust cover as evidence for climate change.

Finally, perhaps the climate cycles were insufficient to make Mars warm enough to sustain liquid waters. The planet had clement conditions for only brief periods after major impacts. Each such impact deposited water-rich material and pumped enough heat and water into the atmosphere to permit rain. Soon, though, the planet returned to its usual frozen state. Victor Baker of the University of Arizona has argued that the intensive volcanism in the Tharsis region periodically made early Mars quite a temperate place.

It is also very possible that none of these options is correct. We simply do not yet know enough about early Mars to have any real understanding of its climate. We must wait for future exploration. Unlike Earth, Mars has preserved much of its ancient landscape, which may yield clues to the conditions under which it formed. Indeed, understanding how Mars became so different from Earth will help geologists grapple with Earth's own history. The new lander missions will soon provide some of these clues. SA

NASA/JPL/MALIN SPACE SCIENCE SYSTEMS MOC 2–264; CALVIN J. HAMILTON (*inset*)

"WHITE ROCK," seen by the Viking spacecraft in the 1970s (*inset*), is a prime example of how Mars's Earth-like appearance can deceive. The feature looks tantalizingly like a pile of salt deposited by liquid water. But spectral measurements now show it to be generic dust that somehow got compacted or cemented together. The reddish dust buried existing features, such as the crater at the top right, and was in turn buried by black sand. The image, taken by Mars Global Surveyor, attests to a bafflingly complex sequence of geologic events.

MORE TO EXPLORE

Mars 2000. Arden L. Albee in *Annual Review of Earth and Planetary Sciences,* Vol. 28, pages 281–304; 2000.

Mars: The Lure of the Red Planet. William Sheehan and Stephen James O'Meara. Prometheus Books, 2001.

The Mars Global Surveyor Mars Orbiter Camera: Interplanetary Cruise through Primary Mission. Michael C. Malin and Kenneth S. Edgett in *Journal of Geophysical Research,* Vol. 106, No. E10, pages 23429–23570; October 25, 2001.

Nature Insight: Mars. Special section in *Nature,* Vol. 412, pages 207–253; July 12, 2001.

For information on NASA's Mars missions, visit **www.jpl.nasa.gov**

For information on the European Space Agency's Mars Express, visit **sci.esa.int/marsexpress**

For information on the Japanese Nozomi mission, visit **www.isas.ac.jp/e/enterp/missions/index.html**

Self-Repairing Co

BY Armando Fox AND David Patterson

Digital computing performance has improved 10,000-fold in the past two decades: what took a year of number crunching in 1983 takes less than an hour nowadays, and a desktop computer from that era can't match the processing power of one of today's handheld organizers.

We pay a price for these enhancements, though. As digital systems have grown in complexity, their operation has become brittle and unreliable. Computer-related failures have become all too common. Personal computers crash or freeze up regularly; Internet sites go offline often. New software upgrades, designed to augment performance, may leave things worse than they were before. Inconvenience aside, the situation is also an expensive one: annual outlays for maintenance, repairs and operations far exceed total hardware and software costs, for both individuals and corporations.

By embracing the inevitability of system failures, RECOVERY-

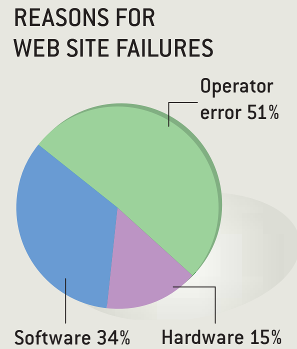
mputers

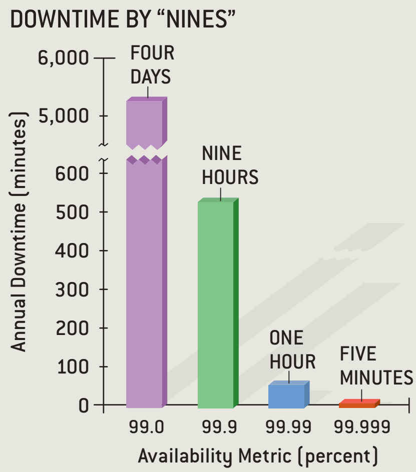
– ORIENTED computing returns service faster

WHOSE FAULT WAS THAT?

TRADITIONAL ENGINEERING approaches to raising the reliability of computer systems largely ignore the possibility of operator error. But in many cases, human mistakes account for more downtime (time during which the system is not functioning) than hardware problems or software bugs. The pie chart (*right*) depicts a breakdown of typical failure causes for three Internet sites.

For many industries, computer system downtime can be costly or even life-threatening. Engineers call the proportion of time a computing system functions correctly its availability, which is measured by "nines" (*graph*). A system that runs without crashing 99.999 percent of the time, for example, has an availability of "five nines," which corresponds to about two hours of downtime over 25 years of operation. Rather than seeking to reduce the number of failures, proponents of recovery-oriented computing advocate methods to shorten the time needed to bring systems back online. Boosting availability from two nines to five nines, for instance, shrinks total recovery time from 90 hours to five minutes a year.





Our group of research collaborators at Stanford University and the University of California at Berkeley has taken a new tack, by accepting that computer failure and human operator error are facts of life. Rather than trying to eliminate computer crashes—probably an impossible task—our team concentrates on designing systems that recover rapidly when mishaps do occur. We call our approach recovery-oriented computing (ROC).

We decided to focus our efforts on improving Internet site software. This kind of highly dynamic computing system must evolve and expand quickly in response to consumer demands and market pressures—while also serving users who expect instant access at any time. Consider the example of the Google search engine, which in just a few years has gone from locating hundreds of millions of Web pages of English text to three billion pages in more than 20 languages in a dozen formats, plus images. Meanwhile the number of daily Google searches has grown from 150,000 to 150 million—the site is now 1,000 times busier than it was at the outset.

Because of the constant need to upgrade the hardware and software of Internet sites, many of the engineering techniques used previously to help maintain system dependability are too expensive to be deployed. Hence, we expect Internet software to be a good proving ground for our ideas and perhaps a model for other computing systems, including desktop and laptop machines. If ROC principles can help the big animals in the computational jungle, they might do the same for the smaller species.

Following a proven engineering strategy first adopted during the era of cast-iron-truss railroad bridges in the 19th century, our initial step was to see what we could learn from previous failures. Specifically, we asked: Why do Internet systems go down, and what can be done about it? We were a bit surprised to find out that operator error was a leading cause of system problems. Traditional efforts to boost the dependability of software and hardware have for the most part overlooked the possibility of human mistakes, yet in many cases operators' miscues accounted for more downtime than any other cause [*see box above*].

Operators may face such difficulties because computer designers and programmers have frequently sacrificed ease of use in the quest for better performance. Database software, for example, can require a full-time staff of trained administrators to manage it. Ironically, because hardware and software have grown cheaper over time, operator salaries are now often the biggest expense in running complex Internet sites.

With these issues in mind, our team is exploring four principles to guide the construction of "ROC-solid" computing systems. The first is speedy recovery: problems are going to happen, so engineers should design systems that recover quickly. Second, suppliers should give operators better tools with which to pinpoint

Overview/*High-Dependability Computing*

- Despite the undoubted power of today's computers, users continue to be tormented by their systems' stubborn unreliability. Recovery-oriented computing (ROC) design practices could do much to solve this predicament.
- ROC principles—which comprise efforts to engineer rapid-recovery capabilities, software tools to locate faults quickly, "undo" functions to reverse human operators' mistakes, and the means to inject errors to test systems' ability to return to service—may eventually take much of the frustration out of computing.
- Benchmarking programs that evaluate the speed with which computer systems return to full service would encourage industry efforts at improving dependability.

SLIM FILMS (*preceding pages*); JOHNNY JOHNSON (*above*), SOURCE FOR PIE CHART: DAVID OPPENHEIMER

Methods to ACHIEVE HIGH DEPENDABILITY: ensure fewer breakdowns or bring systems back online faster.

the sources of faults in multicomponent systems. Third, programmers ought to build systems that support an "undo" function (similar to those in word-processing programs), so operators can correct their mistakes. Last, computer scientists should develop the ability to inject test errors; these would permit the evaluation of system behavior and assist in operator training. We plan to release all the ROC-inspired software applications we write to the computing community at no cost.

To foster the adoption of our approach, we also advocate the development and distribution of benchmark programs that would test the speed of a computing system's recovery. This software would measure the computer industry's progress in raising dependability and encourage companies to work toward achieving this end.

Quick Comeback

MANY A USER REBOOTS his or her personal computer routinely—either preemptively, because the machine is behaving strangely, or reactively, because it has crashed or seized up. Rebooting works for large computers, too, because it wipes the slate clean and fixes a whole class of so-called transient failures—that is, problems that appear intermittently.

Unfortunately, most systems take a long time to reboot and, worse, may lose data in the process. Instead we believe that engineers should design systems so that they reboot gracefully. If one were to look inside a computer, one would see that it is running numerous different software components that work together. During online shopping, for instance, some software modules let customers search through the available merchandise; others permit items to be added to a "shopping cart." Still other software enables the completion of the purchase. Yet another layer of programming choreographs all these functions to produce the overall experience of using the Web site by ensuring that each portion of code does its job when required.

Frequently, only one of these modules may be encountering trouble, but when a user reboots a computer, all the software it is running stops immediately. If each of its separate subcomponents could be restarted independently, however, one might never need to reboot the entire collection. Then, if a glitch has affected only a few parts of the system, restarting just those isolated elements might solve the problem. Should that prove ineffective, reinitializing a larger set of subcomponents might work. The trick is to be able to restart one module without accidentally confusing its peers into thinking something is really wrong, which is akin to swapping out the bottom plate in a stack without allowing the other plates to fall—a difficult but doable feat.

George Candea and James Cutler, Stanford graduate students on our team, have focused on developing this independent-rebooting technique, which we call micro-rebooting. Cutler's experience includes building systems that use inexpensive ground receivers assembled from off-the-shelf PCs, low-cost ham radios and experimental software to capture incoming satellite data. In use, ground-station failures were common, and if a human operator was not available to reactivate the equipment manually, the satellite signal could be lost—and with it, all the data for that orbit.

Last year Candea and Cutler tested micro-rebooting on the ground-station software. They and others modified each receiving-station software module so that it would not "panic" if other subcomponents were reinitialized. The students first consulted the human operators to learn about the most frequent causes of failures and then experimented to determine which sets of subcomponents would have to be reinitialized to cure those specific maladies. They succeeded in automating the recovery process for a range of recurring problems, cutting the average restoration time from 10 minutes to two—fast enough for a ground station that has faltered to reacquire the satellite signal and continue collecting data for the latest orbit.

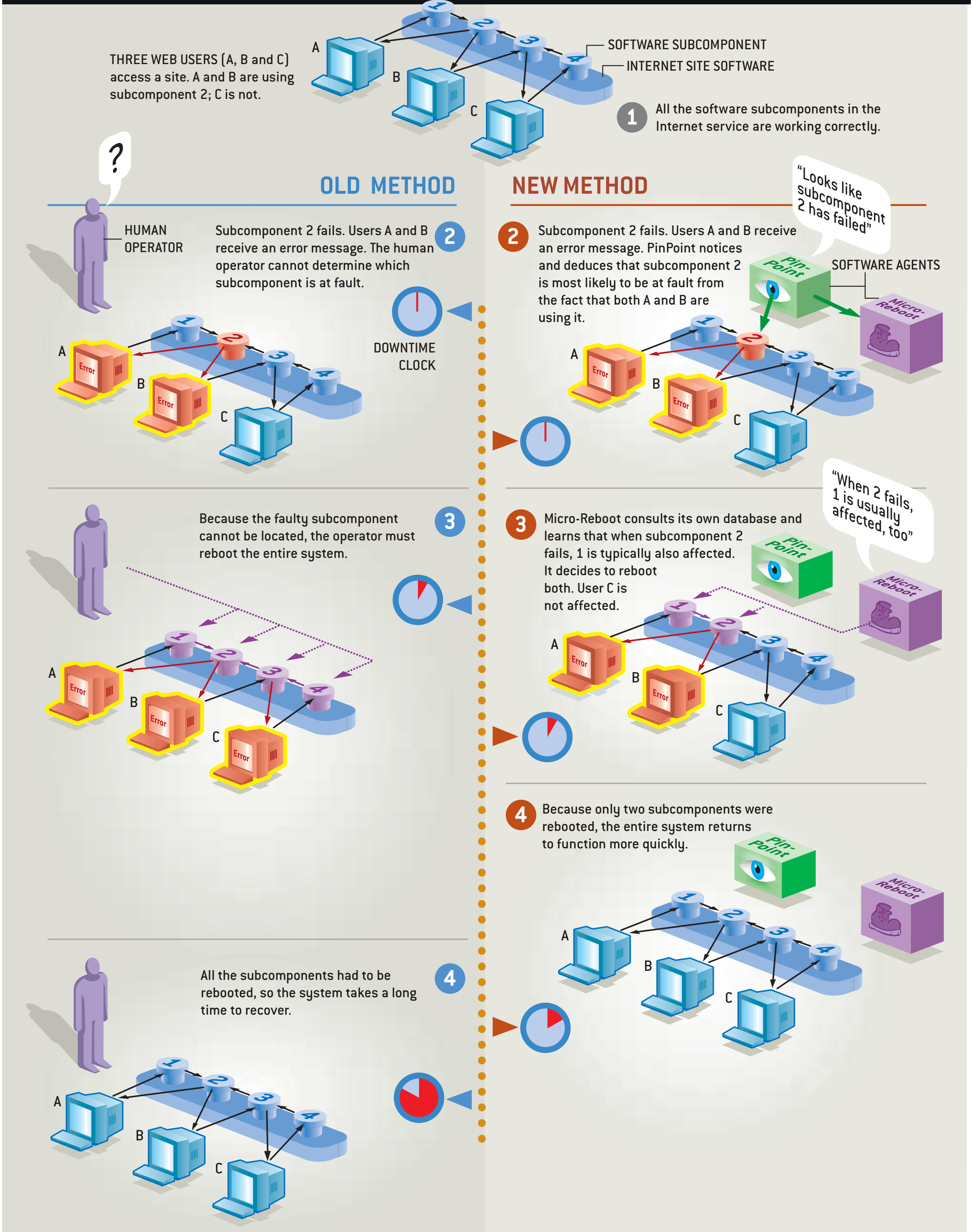
In dependability lingo, the percentage of time a computer system is functioning correctly is termed its availability, which is typically measured in "nines." A system that is functioning correctly 99.999 percent of the time, for example, has an availability of "five nines," corresponding to about two hours of downtime over 25 years of operation. In contrast, well-managed mainstream computing systems are available only 99 to 99.9 percent of the time ("two to three nines"). Going from two nines to five nines would save almost 90 hours of downtime a year, which is easy to appreciate when loss of service costs big money [*see box on opposite page*]. Two methods exist to achieve high dependability: ensuring fewer breakdowns or, failing that, bringing systems back online faster.

For the satellite ground-station operators, a fivefold-faster return to service was much more valuable than a fivefold increase in the time between failures (better reliability), even though either measure would yield the same level of improved availability. We believe that a variety of computing systems exhibit such a threshold.

Although much effort had to be expended to modify the ground-station software manually, Candea and one of us (Fox) are now investigating whether the technique can be applied in an automated fashion to Web sites programmed using the Java 2 Enterprise Edition, a popular development framework for Internet software.

The most common way to fix Web site faults today is to reboot the entire system, which takes anywhere from 10 seconds (if the application alone is rebooted) to a minute (if the whole thing is restart-

HOW SELECTIVE REBOOTING CUTS DOWNTIME



SAMUEL VELASCO

Computers may end up **10,000 TIMES FASTER** yet no more dependable than today's machines.

ed). According to our initial results, microrebooting just the necessary subcomponents takes less than a second. Instead of seeing an error message, a user would experience a three-second delay followed by resumption of normal service.

Pinpointing Problems

IT'S ONE THING to fix known or likely errors, but a tougher, related challenge is to find the unanticipated ones. System operators could use assistance in tracking down problems more quickly, the second of our ROC principles.

When building a traditional high-dependability computer system, programmers start with a complete description of all hardware and software elements. They then follow with a carefully constructed failure-analysis tree, which traces the myriad ways the system can break down—so that those may be prevented or, if they do occur, corrected. In contrast to single-source systems, Internet services are heterogeneous, using components from multiple vendors. Further, these modules often change rapidly as the service evolves. Failures frequently arise from unexpected interactions between components, rather than resulting from a bug in a single piece of software. When this kind of dynamic fault occurs, a Web user who happens to be accessing the service at the time may receive an error message.

To help analyze these complex malfunctions, graduate students Emre Kiciman and Eugene Fratkin of Stanford and Mike Chen of Berkeley created PinPoint, a ROC-based computer program that attempts to determine which components are at fault. Every time someone surfs to a PinPoint-enabled Web site, the program traces which software components participated in delivering service to that user. When a particular access request fails—for example, the user gets an error message from the site—PinPoint notes this fact. Over time, the monitoring application analyzes the mix of components that were activated in both failed and successful requests, using standard data-mining techniques. By doing this, PinPoint can help find out which components are suspected of causing most of the failures. The additional information gathered by the failure-analysis code slows the system down by at most 10 percent. Unlike the traditional solution—which requires elaborate preplanning every time the software suite changes—PinPoint works with any combination of software components.

Wiping Away Errors

PERHAPS THE GREATEST challenge in boosting system reliability is ensuring a margin of safety against random errors inputted by the operator; this rationale underlies our third ROC principle, which concerns the undo command. The first word processors did not have this capability, which made them frustrating, if not terrifying, to use. A single erroneous global substitution could destroy an entire file. The undo function—which affords users the ability to cancel any command—removed the anxiety from word processing.

Operators of today's large computing systems have no such option. When the foundations of information technology were laid, no one considered it important to be able to expunge mistakes. That's because an undo function requires more work to construct, consumes a significant amount of storage space and probably slows systems down somewhat.

To demonstrate a better approach, our group is working on an undo capability for e-mail systems that is aimed at the place where messages are stored. Berkeley graduate student Aaron Brown and one of us (Patterson) have recently completed the prototype of an e-mail system featuring an operator undo utility. We are testing it now [*see box on page 61*].

Suppose a conventional e-mail storage server gets infected by a virus. The system operator must disinfect the server, a laborious job. Our system, however, would record all the server's activities automatically, including discarded messages. If the system gets infected, the operator could employ the undo command to "turn back the clock" to before the arrival of the virus. Software that attacks that virus could then be downloaded. Finally, the operator could "play forward" all the e-mail messages created after the infection, returning the system to normal operation. The newly installed antivirus software would filter all subsequent e-mail traffic. In this way, the operator could undo the damage without losing important mes-

THE AUTHORS

ARMANDO FOX and *DAVID PATTERSON* have studied how to improve the reliability of computing systems for many years. Fox has been assistant professor at Stanford University since 1999. As a Ph.D. student at the University of California, Berkeley, he joined with Professor Eric A. Brewer in constructing prototypes of today's clustered Internet services and showing how they can support mobile computing applications, including the first graphical Web browser for personal digital assistants. Fox also helped to design microprocessors at Intel and later founded a small mobile computing company. He received his other degrees in electrical engineering and computer science from the Massachusetts Institute of Technology and the University of Illinois. Patterson has spent the past quarter of a century as a professor at Berkeley. He received all his advanced degrees in computer science from the University of California, Los Angeles. Patterson is best known for work on the simplification of microprocessor architecture, for building reliable digital storage systems and for co-authoring the classic books *Computer Architecture: A Quantitative Approach* and *Computer Organization and Design: The Hardware/Software Interface*. He first published in the pages of *Scientific American* two decades ago.

STRIVING FOR DEPENDABILITY

COMPUTER SYSTEMS and their "organs"—microprocessors, applications and communications networks—are becoming ever more powerful. But they are also becoming ever more complex and therefore more susceptible to failure. As the costs of administration, oversight and downtime expand in response, scientists and engineers in the computer industry are working to enhance the dependability of their products. Significantly, many of their efforts aim to take humans (and the errors they inevitably engender) out of the loop.

Concerned about security holes, bugs and other weaknesses in its present product line, Microsoft's management took the unusual step recently of halting software development for an entire month to focus on what it calls Trustworthy Computing. The issue of dependability has grown in importance as more administrators adopt the company's Windows operating system to run Web servers. Operating-system developers at Microsoft attended classes to learn techniques that improve security and reliability for desktop systems. They are now refining Windows for the next version, called Palladium. Engineers plan to cull potential weak points from current products while developing new features that boost defenses against hackers.

Little research exists on ways to reduce the lifetime cost of computer ownership—the price individuals and companies pay for running their systems. Programmers at Hewlett-Packard Laboratories and IBM Research are working to cut those expenses by adding new capabilities and developing products that can manage themselves. Hewlett-Packard officials envision a globally networked system of computational and storage resources that monitor, heal and adapt themselves without operator intervention. HP's Planetary Computing project concentrates on developing corporate data centers that could contain as many as 50,000 individual office computers, a collection 10 times as large as today's counterparts.

IBM's scheme borrows ideas from control theory (the use of feedback to stabilize closed-loop systems) and artificial intelligence (mimicking or otherwise capturing expert human skills or intelligence to solve complex problems). These concepts will help create data centers that can diagnose problems on their own, adjust their configurations to match changes in demand, repair themselves and defend against hacker attacks. Drawing an analogy with the body's autonomic nervous system, IBM's management calls this goal Autonomic Computing.

When designers of other engineering systems have discovered a propensity for operator error, they often have attempted to remove the need for human input. Removing human operators can lead to a well-established pitfall known as the Automation Irony. Because designers can typically reduce but not eliminate the need for human intervention, such efforts frequently make things worse. That's because engineers generally automate the tasks that are easy, leaving the hard jobs for people. These measures mean that administrators must carry out difficult tasks intermittently on unfamiliar systems—a sure recipe for failure.

Will the path toward truly dependable computing be additional automation, leading to hands-off computing, or will it be streamlined design combined with tools that dramatically improve the effectiveness and productivity of human operators? Only time will tell. *—A.F. and D.P.*

sages. To prevent potential confusion among users—who might notice that some e-mails had been eradicated—the system could send a message saying that notes were deleted during an attempt to stop the spread of a virus.

Injecting Test Errors

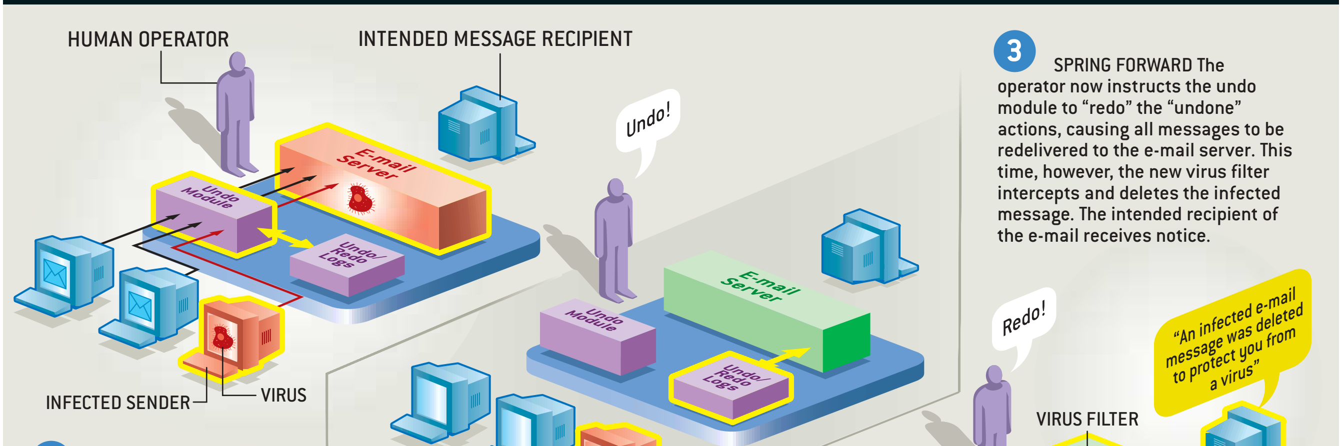
LAST AMONG OUR ROC principles is the idea of harnessing errors to do good: we advocate testing the system periodically by inserting artificial faults. This practice would aid in evaluating the recovery performance of a system and also in coming up with new methods to make it more robust. In an industry analogy, designers of microprocessor chips have regularly added circuits to simplify the testing of chips, even though these additions increase chip size and remain unused after the microprocessors leave the factory. Manufacturers consider these test circuits worth the effort; they lower the cost of ensuring that the completed chips work as planned. Part of this benefit arises because the test circuits allow designers to inject "failures" artificially to verify that the chip detects and recovers from them correctly.

Our group proposes a software equivalent for computer systems. When operators employed the selective-rebooting strategy, for example, test errors would help determine which software components to reinitialize to counter a particular kind of failure. If the problem propagated to only one or two other elements, operators could reboot only those. If the flaw came to involve a great many components, it might be more sensible to reboot the entire system. We have started using error injection to characterize the fault-propagation behavior of Internet sites built using Java 2 Enterprise Edition.

Another version of this software could permit potential buyers to see how a computer system handled failures—a benchmark program that could inform their purchase decisions. Developed by Berkeley graduate students Pete Broadwell, Naveen Sastry and Jonathan Traupman, the Fig application tests the ability of programs to cope correctly with unexpected errors in the standard C library, a part of the operating system used by nearly all software programs. "Fig"

SAMUEL VELASCO

TURNING THE CLOCK BACK ON PROBLEMS



1 MARK TIME An e-mail user sends out a virus-laden message to a friend. The human operator of the e-mail system cannot protect the server from an undetected virus. But because the "undo" module monitors and logs all messages traveling in and out (including the infected one), it can fix these kinds of problems once they become apparent.

2 FALL BACK The operator hears that a new virus is circulating and suspects that the server may be infected. Fortunately, the infected message has not yet been forwarded to its intended recipient. The operator tells the undo module to use its undo/redo log data to "undeliver" all messages that arrived since the virus appeared. The operator installs a virus filter.

3 SPRING FORWARD The operator now instructs the undo module to "redo" the "undone" actions, causing all messages to be redelivered to the e-mail server. This time, however, the new virus filter intercepts and deletes the infected message. The intended recipient of the e-mail receives notice.

SAMUEL VELASCO

stands for "Fault Injection in *glibc*" (a version of the standard C library used by numerous programmers).

Error injection would also permit computer programmers to test their repair mechanisms, which is difficult to do. Fig would allow operators to try their hands at diagnosing and repairing failures, perhaps on a small experimental system if not on the real thing. The program has been used several times and is available for no charge on our ROC Web site [*see More to Explore below*].

Benchmarking Recovery

THE HISTORY OF the computer industry makes us strong believers in the importance of measuring technical progress and publicizing it. When computer firms finally adopted standard benchmarking programs to compare performance (after many years of delay), customers could at last see the relative merits of each product clearly. Companies that trailed in technology were forced to spend more on engineering and subsequently could gauge the effect of their innovations using standard evaluation metrics. The resulting test data led to a continuing cascade of performance improvements.

By focusing on evaluating recovery time, Brown and fellow Berkeley graduate student David Oppenheimer, together with Patterson, are working to re-create that kind of competition for computer system dependability. Products that were shown to come back from crashes quickest would win greater sales. We envision a test suite that would incorporate common failures, including errors caused by humans, software and hardware. Prospective buyers could insert these faults into systems and then monitor the time to recovery. It is noteworthy that current computer-marketing efforts more typically quote availability (percentage downtime) metrics, which are much more difficult to measure than recovery time.

When scientists and engineers focus their efforts, they can achieve amazing progress in a relatively short time. The meteoric 30-year rise of both computer performance and cost-effectiveness proves it. If the industry continues traveling blindly down the current path, computers in 2023 may end up another 10,000 times faster yet no more dependable than today's machines. But with dependability-enhancing software tools and benchmarks to inspire us, computing may one day become as reliable as users expect it to be. SA

MORE TO EXPLORE

To Engineer Is Human: The Role of Failure in Successful Design. H. Petroski. Vintage Books, 1992.

The Mythical Man-Month: Essays on Software Engineering. Second edition. Frederick P. Brooks. Addison-Wesley, 1995.

Managing the Risks of Organizational Accidents. James Reason. Ashgate Publishing Company, 1997.

Autonomic Computing. IBM Research, 2001. Available at **www.research.ibm.com/autonomic/manifesto**

Building a Secure Platform for Trustworthy Computing. Microsoft, December 2002. Available at **www.microsoft.com/security/whitepapers/secure_platform.asp**

Planetary Computing Web site at HP Labs: **www.hpl.hp.com/org/issl/current_research_themes.htm**

Recovery Oriented Computing Web site at University of California at Berkeley and Stanford University: **http://roc.cs.berkeley.edu**

Turing. Christos Papadimitriou. MIT Press (in press).

PANDORA'S BABY

In vitro fertilization was once considered by some to be a threat to our very humanity. Cloning inspires similar fears

BY ROBIN MARANTZ HENIG

On July 25, a once unique person will turn 25. This nursery school aide in the west of England seems like an average young woman, a quiet, shy blonde who enjoys an occasional round of darts at the neighborhood pub. But Louise Brown's birth was greeted by newspaper headlines calling her the "baby of the century." Brown was the world's first test tube baby.

Today people may remember Brown's name, or that she was British, or that her doctors, Steptoe and Edwards, sounded vaguely like a vaudeville act. But the past quarter of a century has dimmed the memory of one of the most important aspects of her arrival: many people were horrified by it. Even some scientists feared that Patrick Steptoe and Robert Edwards might have brewed pestilence in a petri dish. Would the child be normal, or would the laboratory manipulations leave dreadful genetic derangements? Would she be psychologically scarred by the knowledge of how bizarrely she had been created? And was she a harbinger of a race of unnatural beings who might eventually be fashioned specifically as a means to nefarious ends?

Now that in vitro fertilization (IVF) has led to the birth of an estimated one million babies worldwide, these fears and speculations may seem quaint and even absurd. But the same concerns once raised about IVF are being voiced, sometimes almost verbatim, about human cloning. Will cloning go the way of IVF, morphing from the monstrous to the mundane? And if human cloning, as well as other genetic interventions on the embryo, does someday become as commonplace as test tube baby–making, is that to be feared—or embraced? The lessons that have been

RYSZARD HOROWITZ *Corbis*

MICRONEEDLE INJECTS a sperm's package of DNA directly into a human egg, thereby achieving in vitro fertilization (*left*). The first human being born as a result of IVF, Louise Brown was 14 months old when she frolicked on the set of the *Donahue* television program (*right*). With her was Vanderbilt University IVF researcher Pierre Soupart, who predicted that "by the time Louise is 15, there will be so many others it won't be remarkable anymore."

PASCAL GOETGHELUCK *Science Photo Library* (*left*); RAY FOLI *Corbis-Bettmann* (*right*)

learned from the IVF experience can illuminate the next decisions to be made.

Then and Now

AS IVF MOVED FROM the hypothetical to the actual, some considered it to be nothing more than scientists showing off: "The development of test tube babies," one critic remarked, "can be compared to the perfecting of wing transplants so that pigs might fly." But others thought of IVF as a perilous insult to nature. The British magazine *Nova* ran a cover story in the spring of 1972 suggesting that test tube babies were "the biggest threat since the atom bomb" and demanding that the public rein in the unpredictable scientists. "If today we do not accept the responsibility for directing the biologist," the *Nova* editors wrote, "tomorrow we may pay a bitter price—the loss of free choice and, with it, our humanity. We don't have much time left."

A prominent early enemy of IVF was Leon Kass, a biologist at the University of Chicago who took a professional interest in the emerging field of bioethics. If society allowed IVF to proceed, he wrote shortly after Louise Brown's birth, some enormous issues were at stake: "the idea of the humanness of our human life and the meaning of our embodiment, our sexual being, and our relation to ancestors and descendants."

Now read Kass, a leading detractor of every new form of reproductive technology for the past 30 years, in 2003: "[Cloning] threatens the dignity of human procreation, giving one generation unprecedented genetic control over the next," he wrote in the *New York Times*. "It is the first step toward a eugenic world in which children become objects of manipulation and products of will." Such commentary coming from Kass is particularly noteworthy because of his unique position: for the past two years he has been the head of President George W. Bush's Council on Bioethics, whose first task was to offer advice on how to regulate human cloning.

Of course, IVF did not wind up creating legions of less than human children, nor did it play a role in the disintegration of the nuclear family, consequences that people like Kass feared. And so many newer, more advanced methods of assisted reproduction have been introduced in the past decade that the "basic IVF" that produced Louise Brown now seems positively routine. One early prediction, however, did turn out to contain more than a kernel of truth. In the 1970s critics cautioned that IVF would set us tumbling down the proverbial slippery slope toward more sophisticated and, to some, objectionable forms of reproductive technology—and that once we opened the floodgates by allowing human eggs to be fertilized in the laboratory, there would be no stopping our descent.

If you consider all the techniques that might soon be available to manipulate a developing embryo, it could appear that the IVF naysayers were correct in their assessment of the slipperiness of the slope. After all, none of the genetic interventions now being debated—prenatal genetic diagnosis, gene insertions in sex cells or embryos to correct disease, the creation of new embryonic stem cell lines and, the elephant in the living room, cloning—would even be potentialities had scientists not first learned how to fertilize human eggs in a laboratory dish.

But does the existence of a such a slippery slope mean that present reproductive technology research will lead inevitably to developments that some find odious, such as embryos for tissue harvesting, or the even more abhorrent manufacture of human-nonhuman hybrids and human clones? Many people clearly fear so, which explains the current U.S. efforts to

Overview/*In Vitro Veritas*

- Many arguments against in vitro fertilization in the past and cloning today emphasize a vague threat to the very nature of humanity.
- Critics of IVF attempted to keep the federal government from supporting the research and thus ironically allowed it to flourish with little oversight.
- Because of the lack of oversight, it is only in the past few years that the increased rate of birth defects and low birth weight related to IVF have come to light.

MEMBERS of the Christian Defense Coalition and the National Clergy Council protest Advanced Cell Technologies's human cloning research outside the biotechnology firm's headquarters in Worcester, Mass., on November 30, 2001. Similar protests against IVF occurred in the 1970s.

WILLIAM PLOWMAN *Getty*

curtail scientists' ability to manipulate embryos even before the work gets under way. But those efforts raise the question of whether science that has profound moral and ethical implications should simply never be done. Or should such science proceed, with careful attention paid to the early evolution of certain areas of research so that society can make informed decisions about whether regulation is needed?

IVF Unbound

THE FRENZY TO TRY to regulate or even outlaw cloning is in part a deliberate attempt not to let it go the way of IVF, which has been a hodgepodge of unregulated activities with no governmental or ethical oversight and no scientific coordination. Ironically, the reason IVF became so ubiquitous and uncontrolled in the U.S. was that its opponents, particularly antiabortion activists, were trying to stop it completely. Antiabortion activists' primary objection to IVF was that it involved the creation of extra embryos that would ultimately be unceremoniously destroyed—a genocide worse than at any abortion clinic, they believed. Accordingly, they thought that their best strategy would be to keep the federal government from financing IVF research.

A succession of presidential commissions starting in 1973 debated the ethics of IVF but failed to clarify matters. Some of the commissions got so bogged down in abortion politics that they never managed to hold a single meeting. Others concluded that IVF research was ethically acceptable as long as scientists honored the embryo's unique status as a "potential human life," a statement rather than a practical guideline. In 1974 the government banned federal funding for fetal research. It also forbade funding for research on the human embryo (defined as a fetus less than eight weeks old), which includes IVF. In 1993 President Bill Clinton signed the NIH Revitalization Act, which allowed federal funding of IVF research. (In 1996, however, Congress again banned embryo research.) The bottom line is that despite a series of recommendations from federal bioethics panels stating that taxpayer support of IVF research would be acceptable with certain safeguards in place, the government has never sponsored a single research grant for human IVF.

This lack of government involvement—which would also have served to direct the course of IVF research—led to a funding vacuum, into which rushed entrepreneurial scientists supported by private money. These free agents did essentially whatever they wanted and whatever the market would bear, turning IVF into a cowboy science driven by the marketplace and undertaken without guidance. The profession attempted to regulate itself—in 1986, for example, the American Fertility Society issued ethical and clinical guidelines for its members—but voluntary oversight was only sporadically effective. The quality of clinics, of which there were more than 160 by 1990, remained spotty, and those seeking IVF had little in the way of objective information to help them choose the best ones.

Today, in what appears to be an effort to avoid the mistakes made with IVF, the federal government is actively involved in regulating cloning. With the announcement in 1997 of the birth of Dolly, the first mammal cloned from an adult cell, President Clinton established mechanisms, which remain in place, to prohibit such activities in humans. Congress has made several attempts to outlaw human cloning, most recently with a bill that would make any form of human cloning punishable by a $1-million fine and up to 10 years in prison. (The House of Representatives passed this bill this past winter, but the Senate has yet to debate it.) Politicians thus lumped together two types of cloning that scientists have tried to keep separate: "therapeutic," or "research," cloning, designed to produce embryonic stem cells that might eventually mature into specialized human tissues to treat degenerative diseases; and "reproductive" cloning, undertaken specifically to bring forth a cloned human being. A second bill now

THE AUTHOR

ROBIN MARANTZ HENIG has written seven books, most recently *The Monk in the Garden: The Lost and Found Genius of Gregor Mendel*. Her articles have appeared in the *New York Times Magazine, Civilization* and *Discover,* among other publications. Her honors include an Alicia Patterson Foundation fellowship and a nomination for a National Book Critics Circle Award. She lives in New York City with her husband, Jeffrey R. Henig, a political science professor at Columbia University; they have two nearly grown daughters. Her next book, entitled *Pandora's Baby,* is about the early days of in vitro fertilization research.

From Outrage to Approval

THE STORY of Doris Del-Zio demonstrates the ironies resulting from society's changing attitude toward IVF in the 1970s. After years of failure to conceive a child, Del-Zio and her husband turned to Landrum Shettles of what is now known as the Columbia Presbyterian Medical Center. In the fall of 1973 Shettles prepared to attempt a hasty IVF procedure on the couple. The operation was abruptly terminated by Shettles's superior, Raymond Vande Wiele, who was outraged at Shettles's audacity and who questioned the medical ethics of IVF. Vande Wiele confiscated and froze the container holding the Del-Zios' eggs and sperm. As far as the Del-Zios were concerned, Vande Wiele had committed murder: they sued him and his employers for $1.5 million.

By coincidence, the Del-Zios' case against Vande Wiele was finally brought to trial in July 1978, the same month that Louise Brown was born. The birth of the world's first test tube baby put Shettles's early IVF attempt in a different light. After Brown's appearance, most people—including the two men and four women on the Del-Zio jury—seemed much more inclined to think of IVF as a medical miracle than as a threat to civilized society.

The trial lasted six weeks, each side making its case about the wisdom, safety and propriety of IVF. In the end, Vande Wiele was found to be at fault for "arbitrary and malicious" behavior, and he and his co-defendants were ordered to pay Doris Del-Zio $50,000.

IVF developed rapidly after the trial, and 200 more test tube babies—including Louise Brown's sister, Natalie—were born over the next five years. (Natalie is now a mother, having conceived naturally, and is the first IVF baby to have a child.) Seeing so many healthy-looking test tube babies worldwide changed Vande Wiele's opinion, a change that paralleled the transformation in feeling about IVF that was occurring in the public at large. When Columbia University opened the first IVF clinic in New York City in 1983, its co-director was Raymond Vande Wiele. *—R.M.H.*

COURTING JUSTICE: Doris Del-Zio and her attorney, Michael Dennis, outside U.S. district court in New York City on July 17, 1978, after a session of jury selection. Del-Zio and her husband, John, sued physician Raymond Vande Wiele for derailing their early attempt at in vitro fertilization.

AP PHOTO

before the Senate would explicitly protect research cloning while making reproductive cloning a federal offense.

IVF Risks Revealed

ONE RESULT OF the unregulated nature of IVF is that it took nearly 25 years to recognize that IVF children *are* at increased medical risk. For most of the 1980s and 1990s, IVF was thought to have no effect on birth outcomes, with the exception of problems associated with multiple births: one third of all IVF pregnancies resulted in twins or triplets, the unintended consequence of the widespread practice of implanting six or eight or even 10 embryos into the womb during each IVF cycle, in the hope that at least one of them would "take." (This brute-force method also leads to the occasional set of quadruplets.) When early studies raised concerns about the safety of IVF—showing a doubling of the miscarriage rate, a tripling of the rate of stillbirths and neonatal deaths, and a fivefold increase in ectopic pregnancies—many people attributed the problems not to IVF itself but to its association with multiple pregnancies.

By last year, however, IVF's medical dark side became undeniable. In March 2002 the *New England Journal of Medicine* published two studies that controlled for the increased rate of multiple births among IVF babies and still found problems. One study compared the birth weights of more than 42,000 babies conceived through assisted reproductive technology, including IVF, in the U.S. in 1996 and 1997 with the weights of more than three million babies conceived naturally. Excluding both premature births and multiple births, the test tube babies were still two and a half times as likely to have low birth weights, defined as less than 2,500 grams, or about five and a half pounds. The other study looked at more than 5,000 babies born in Australia between 1993 and 1997, including 22 percent born as a result of IVF. It found that IVF babies were twice as likely as naturally conceived infants to have multiple major birth defects, in particular chromosomal and musculoskeletal abnormalities. The Australian researchers speculate that these problems may be a consequence

THE DAY AFTER her 20th birthday, Louise Brown poses at home with her parents.

ADRIAN ARBIB Corbis

of the drugs used to induce ovulation or to maintain pregnancy in its early stages. In addition, factors contributing to infertility may increase the risk of birth defects. The technique of IVF itself also might be to blame. A flawed sperm injected into an egg, as it is in one IVF variation, may have been unable to penetrate the egg on its own and is thus given a chance it would otherwise not have to produce a baby with a developmental abnormality.

Clearly, these risks could remain hidden during more than two decades of experience with IVF only because no system was ever put in place to track results. "If the government had supported IVF, the field would have made much more rapid progress," says Duane Alexander, director of the National Institute of Child Health and Human Development. "But as it is, the institute has never funded human IVF research of any form"—a record that Alexander calls both incredible and embarrassing.

Although the medical downsides of IVF are finally coming to light, many of the more alarmist predictions about where IVF would lead never came to pass. For example, one scenario was that it would bring us "wombs for hire," an oppressed underclass of women paid to bear the children of the infertile rich. But surrogate motherhood turned out to be expensive and emotionally complex for all parties, and it never became widespread.

Human cloning, too, might turn out to be less frightening than we currently imagine. Market forces might make reproductive cloning impractical, and scientific advancement might make it unnecessary. For example, people unable to produce eggs or sperm might ponder cloning to produce offspring. But the technology developed for cloning could make it possible to create artificial eggs or sperm containing the woman's or man's own DNA, which could then be combined with the sperm or egg of a partner. In the future, "cloning" might refer only to what is now being called therapeutic cloning, and it might eventually be truly therapeutic: a laboratory technique for making cells for the regeneration of damaged organs, for example. And some observers believe that the most common use of cloning technology will ultimately not involve human cells at all: the creature most likely to be cloned may wind up being a favorite family dog or cat.

The history of IVF reveals the pitfalls facing cloning if decision making is simply avoided. But despite similarities in societal reactions to IVF and cloning, the two technologies are philosophically quite different. The goal of IVF is to enable sexual reproduction in order to produce a genetically unique human being. Only the site of conception changes, after which events proceed much the way they normally do. Cloning disregards sexual reproduction, its goal being to mimic not the process but the already existing living entity. Perhaps the biggest difference between IVF and cloning, however, is the focus of our anxieties. In the 1970s the greatest fear related to in vitro fertilization was that it would fail, leading to sorrow, disappointment and possibly the birth of grotesquely abnormal babies. Today the greatest fear about human cloning is that it may succeed. SA

MORE TO EXPLORE

Moving toward Clonal Man: Is This What We Want? James D. Watson in *Atlantic Monthly,* Vol. 227, No. 5, pages 50–53; May 1971.

The Frankenstein Myth Becomes a Reality: We Have the Awful Knowledge to Make Exact Copies of Human Beings. Willard Gaylin in *New York Times Magazine,* pages 12–13, 41–46; March 5, 1972.

The Embryo Sweepstakes. David Rorvik in *New York Times Magazine,* pages 17, 50–62; September 15, 1974.

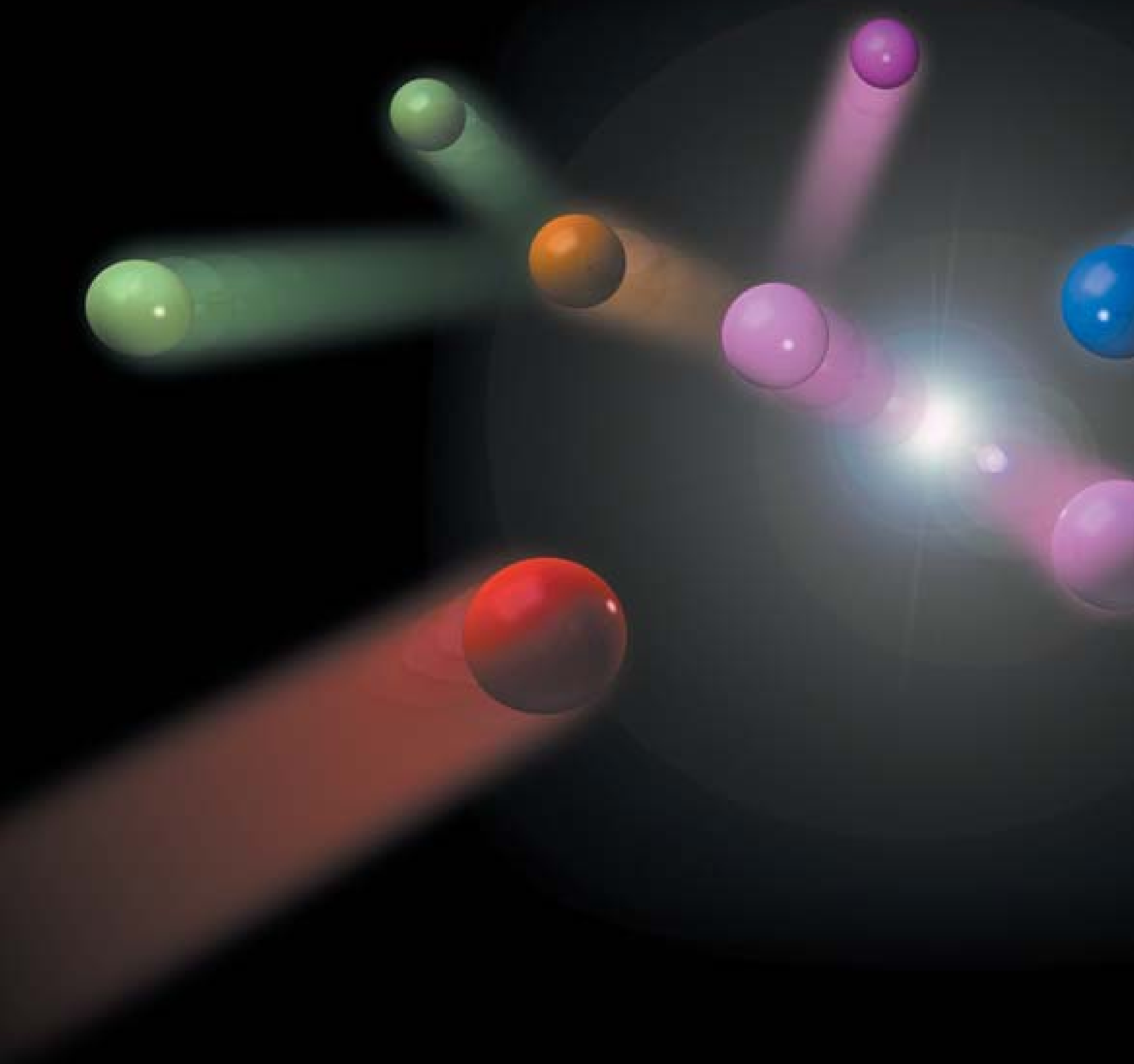
Remaking Eden: How Genetic Engineering and Cloning Will Transform the American Family. Lee M. Silver. Avon Books, 1998.

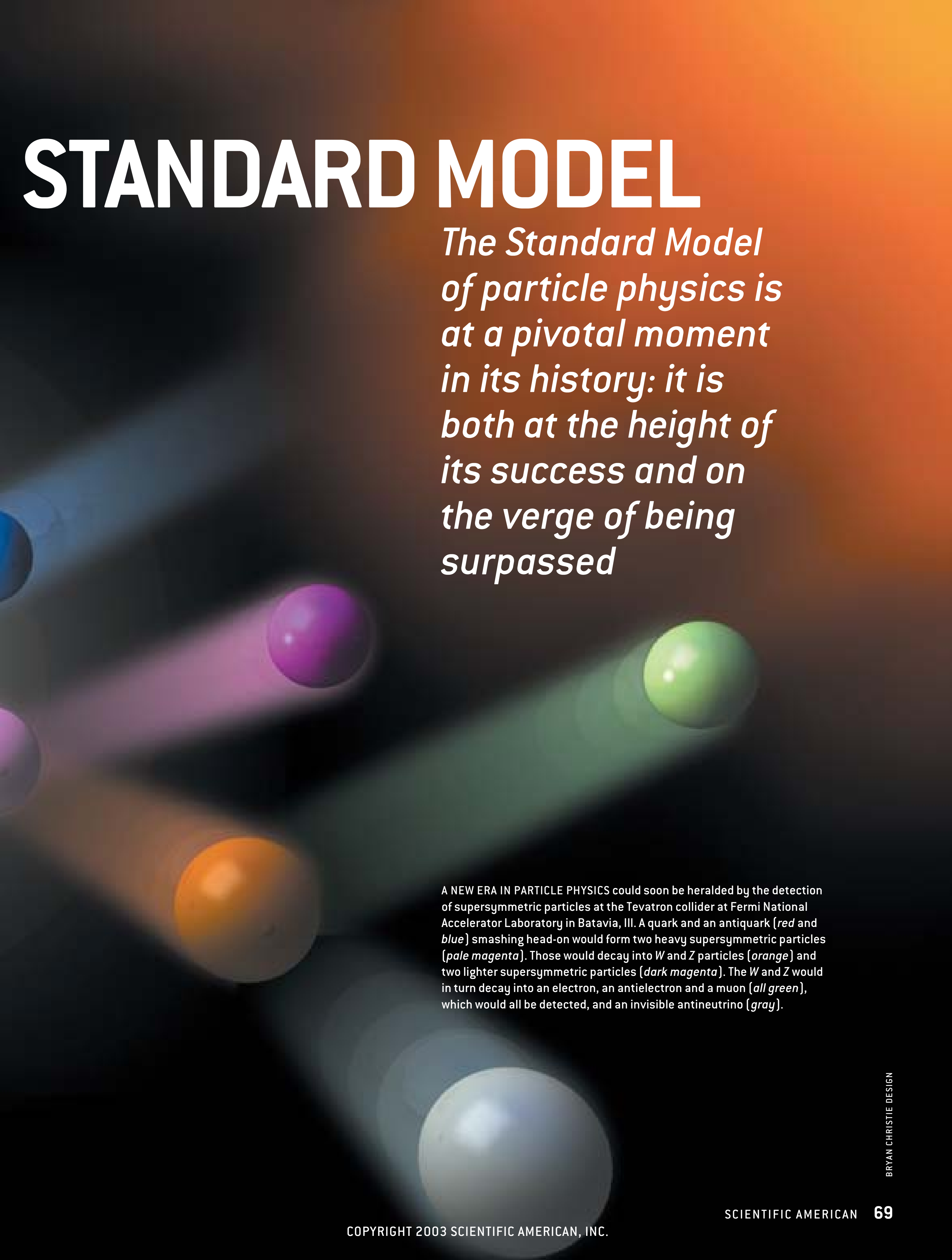
The Clone Age: Adventures in the New World of Reproductive Technology. Lori B. Andrews. Henry Holt and Company, 1999.

Free to Be Me: Would-Be Cloners Pushing the Debate. Rick Weiss in *Washington Post,* page A1; May 12, 2002.

The Dawn of PHYSICS BEYOND THE

By Gordon Kane





STANDARD MODEL

The Standard Model of particle physics is at a pivotal moment in its history: it is both at the height of its success and on the verge of being surpassed

A NEW ERA IN PARTICLE PHYSICS could soon be heralded by the detection of supersymmetric particles at the Tevatron collider at Fermi National Accelerator Laboratory in Batavia, Ill. A quark and an antiquark (*red* and *blue*) smashing head-on would form two heavy supersymmetric particles (*pale magenta*). Those would decay into *W* and *Z* particles (*orange*) and two lighter supersymmetric particles (*dark magenta*). The *W* and *Z* would in turn decay into an electron, an antielectron and a muon (*all green*), which would all be detected, and an invisible antineutrino (*gray*).

BRYAN CHRISTIE DESIGN

Today, centuries after the search began for the fundamental constituents that make up all the complexity and beauty of the everyday world, we have an astonishingly simple answer—it takes just six particles: the electron, the up and the down quarks, the gluon, the photon and the Higgs boson. Eleven additional particles suffice to describe all the esoteric phenomena studied by particle physicists [*see box at right*]. This is not speculation akin to the ancient Greeks' four elements of earth, air, water and fire. Rather it is a conclusion embodied in the most sophisticated mathematical theory of nature in history, the Standard Model of particle physics. Despite the word "model" in its name, the Standard Model is a comprehensive theory that identifies the basic particles and specifies how they interact. Everything that happens in our world (except for the effects of gravity) results from Standard Model particles interacting according to its rules and equations.

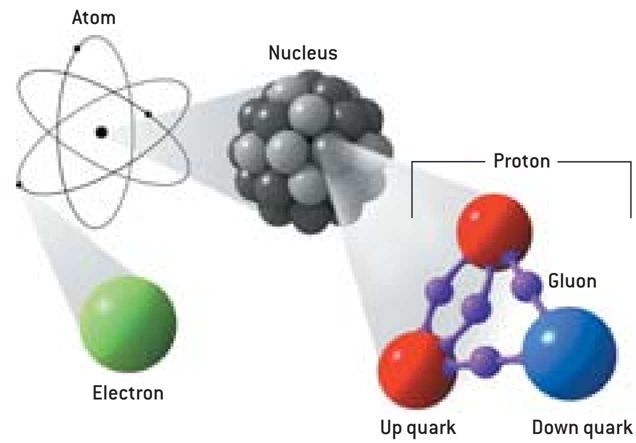
The Standard Model was formulated in the 1970s and tentatively established by experiments in the early 1980s. Nearly three decades of exacting experiments have tested and verified the theory in meticulous detail, confirming all of its predictions. In one respect, this success is rewarding because it confirms that we really understand, at a deeper level than ever before, how nature works. Paradoxically, the success has also been frustrating. Before the advent of the Standard Model, physicists had become used to experiments producing unexpected new particles or other signposts to a new theory almost before the chalk dust had settled on the old one. They have been waiting 30 years for that to happen with the Standard Model.

Their wait should soon be over. Experiments that achieve collisions that are higher in energy than ever before or that study cer-

Overview/*A New Era*

- The Standard Model of particle physics is the most successful theory of nature in history, but increasingly there are signs that it must be extended by adding new particles that play roles in high-energy reactions.
- Major experiments are on the verge of providing direct evidence of these new particles. After 30 years of consolidation, particle physics is entering a new era of discovery. Many profound mysteries could be resolved by post–Standard Model physics.
- One element of the Standard Model—a particle called the Higgs boson—also remains to be observed. The Tevatron collider at Fermilab could detect Higgs bosons within the next few years.

THE STANDARD MODEL



The Particles

ALTHOUGH THE STANDARD MODEL needs to be extended, its particles suffice to describe the everyday world (except for gravity) and almost all data collected by particle physicists.

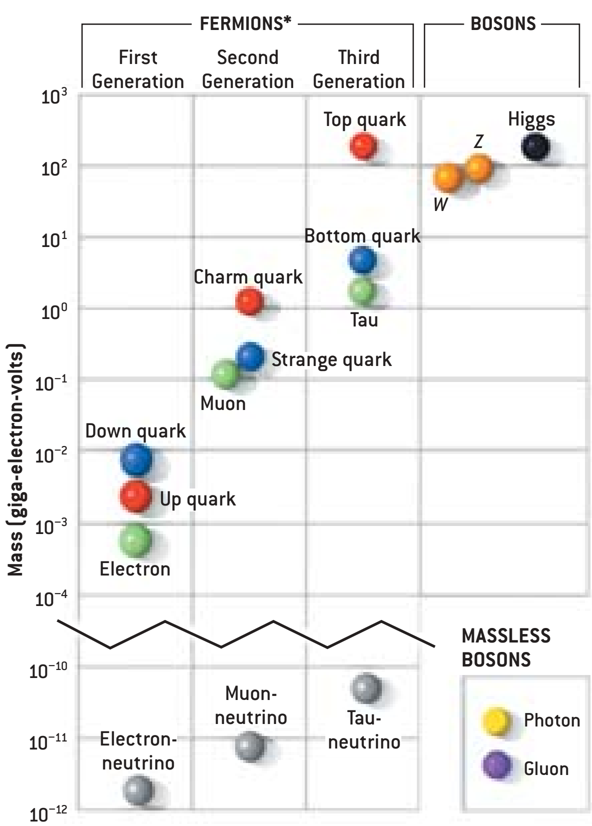
MATTER PARTICLES (FERMIONS) In the Standard Model, the fundamental particles of ordinary matter are the electron, the up quark (u) and the down quark (d). Triplets of quarks bind together to form protons (uud) and neutrons (udd), which in turn make up atomic nuclei (*above*). The electron and the up and the down quarks, together with the electron-neutrino, form the first of three groups of particles called generations. Each generation is identical in every respect except for the masses of the particles (*grid at right*). The values of the neutrino masses in the chart are speculative but chosen to be consistent with observations.

FORCE CARRIERS (BOSONS) The Standard Model describes three of the four known forces: electromagnetism, the weak force (which is involved in the formation of the chemical elements) and the strong force (which holds protons, neutrons and nuclei together). The forces are mediated by force particles: photons for electromagnetism, the *W* and *Z* bosons for the weak force, and gluons for the strong force. For gravity, gravitons are postulated, but the Standard Model does not include gravity. The Standard Model partially unifies the electromagnetic and weak forces—they are facets of one "electroweak" force at high energies or, equivalently, at distances smaller than the diameter of protons.

One of the greatest successes of the Standard Model is that the forms of the forces—the detailed structure of the equations describing them—are largely determined by general principles embodied in the theory rather than being chosen in an ad hoc fashion to match a collection of empirical data. For electromagnetism, for example, the validity of relativistic quantum field theory (on which the Standard Model is based) and the existence of the electron imply that the photon must also exist and interact in the way that it does—we finally understand light. Similar arguments predicted the existence and properties, later confirmed, of gluons and the *W* and *Z* particles.

THE SOURCE OF MASS In addition to the particles described above, the Standard Model predicts the existence of the Higgs boson, which has not yet been directly detected by experiment. The Higgs interacts with the other particles in a special manner that gives them mass.

DEEPER LEVELS? Might the Standard Model be superseded by a theory in which quarks and electrons are made up of more fundamental particles? Almost certainly not. Experiments have probed much more deeply than ever before without finding a hint of additional structure. More important, the Standard Model is a consistent theory that makes sense if electrons and quarks are fundamental. There are no loose ends hinting at a deeper underlying structure. Further, all the forces become similar at high energies, particularly if supersymmetry is true [*see box on next page*]. If electrons and quarks are composite, this unification fails: the forces do not become equal. Relativistic quantum field theory views electrons and quarks as being pointlike—they are structureless. In the future, they might be thought of as tiny strings or membranes (as in string theory), but they will still be electrons and quarks, with all the known Standard Model properties of these objects at low energies.



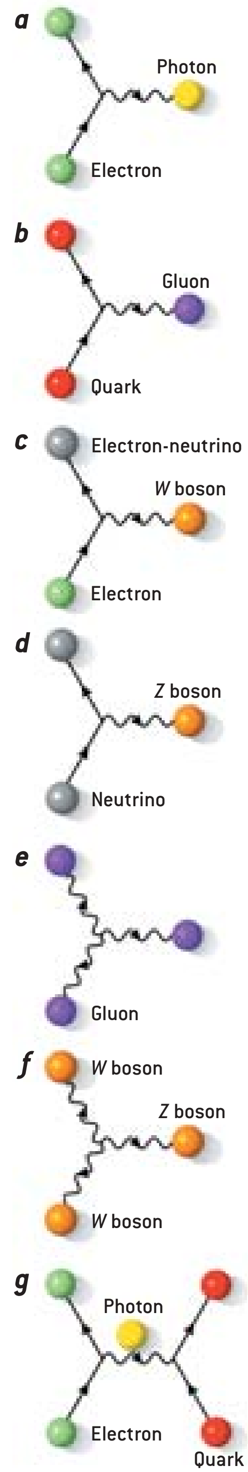
The fermions are subdivided into quarks and leptons, with leptons including electrons, muons, taus and three forms of neutrino.

The Rules of the Game

THE STANDARD MODEL describes the fundamental particles and how they interact. For a full understanding of nature, we also need to know what rules to use to calculate the results of the interactions. An example that helps to elucidate this point is Newton's law, $F = ma$. F is any force, m is the mass of any particle, and a is the acceleration of the particle induced by the force. Even if you know the particles and the forces acting on them, you cannot calculate how the particles behave unless you also know the rule $F = ma$. The modern version of the rules is relativistic quantum field theory, which was invented in the first half of the 20th century. In the second half of the 20th century the development of the Standard Model taught researchers about the nature of the particles and forces that were playing by the rules of quantum field theory. The classical concept of a force is also extended by the Standard Model: in addition to pushing and pulling on one another, when particles interact they can change their identity and be created or destroyed.

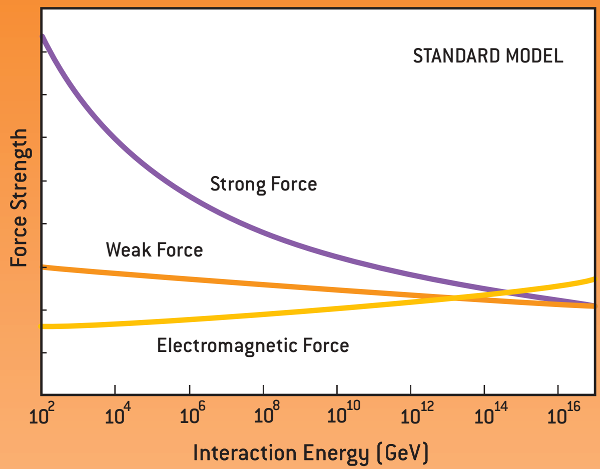
Feynman diagrams (*a–g, at right*), first devised by physicist Richard P. Feynman, serve as useful shorthand to describe interactions in quantum field theory. The straight lines represent the trajectories of matter particles; the wavy lines represent those of force particles. Electromagnetism is produced by the emission or absorption of photons by any charged particle, such as an electron or a quark. In *a*, the incoming electron emits a photon and travels off in a new direction. The strong force involves gluons emitted (*b*) or absorbed by quarks. The weak force involves *W* and *Z* particles (*c, d*), which are emitted or absorbed by both quarks and leptons (electrons, muons, taus and neutrinos). Notice how the *W* causes the electron to change identity. Gluons (*e*) and *W*s and *Z*s (*f*) also self-interact, but photons do not.

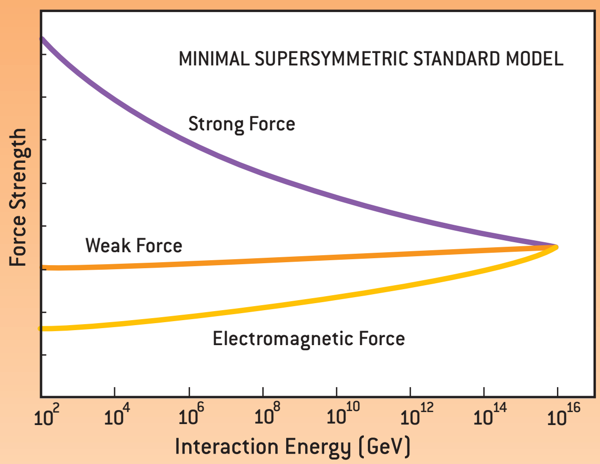
Diagrams *a* through *f* are called interaction vertices. Forces are produced by combining two or more vertices. For example, the electromagnetic force between an electron and a quark is largely generated by the transfer of a photon (*g*). *Everything* that happens in our world, except for gravity, is the result of combinations of these vertices. —*G.K.*



BRYAN CHRISTIE DESIGN

Evidence for Supersymmetry





THE MOST WIDELY FAVORED THEORY to supersede the Standard Model is the Minimal Supersymmetric Standard Model. In this model, every known particle species has a superpartner particle that is related to it by supersymmetry. Particles come in two broad classes: bosons (such as the force particles), which can gather en masse in a single state, and fermions (such as quarks and leptons), which avoid having identical states. The superpartner of a fermion is always a boson and vice versa.

Indirect evidence for supersymmetry comes from the extrapolation of interactions to high energies. In the Standard Model, the three forces become similar but not equal in strength (*top*). The existence of superpartners changes the extrapolation so that the forces all coincide at one energy (*bottom*)—a clue that they become unified if supersymmetry is true.

THE AUTHOR

GORDON KANE, a particle theorist, is Victor Weisskopf Collegiate Professor of Physics at the University of Michigan at Ann Arbor. His work explores ways to test and extend the Standard Model of particle physics. In particular he studies Higgs physics and the Standard Model's supersymmetric extension, with a focus on relating theory and experiment and on the implications of supersymmetry for particle physics and cosmology. His hobbies include playing squash, exploring the history of ideas, and seeking to understand why science flourishes in some cultures but not others.

tain key phenomena with greater precision are on the verge of going beyond the Standard Model. These results will not overturn the Standard Model. Instead they will extend it by uncovering particles and forces not described by it. The most important experiment is occurring at the upgraded Tevatron collider at Fermi National Accelerator Laboratory in Batavia, Ill., which began taking data in 2001. It might produce directly the still elusive particles that complete the Standard Model (Higgs bosons) and those predicted by the most compelling extensions of the theory (the so-called superpartners of the known particles).

Significant information is also beginning to come from "B factories," particle colliders running in California and Japan configured to create billions of b quarks (one of the 11 additional particles) and their antimatter equivalents to study a phenomenon called CP violation. CP (charge-parity) is the symmetry relating matter to antimatter, and CP violation means that antimatter does not exactly mirror matter in its behavior. The amount of CP violation observed so far in particle decays can be accommodated by the Standard Model, but we have reasons to expect much more CP violation than it can produce. Physics that goes beyond the Standard Model can generate additional CP violation.

Physicists are also studying the precise electric and magnetic properties of particles. The Standard Model predicts that electrons and quarks behave as microscopic magnets with a specific strength and that their behavior in an electric field is determined purely by their electric charge. Most extensions of the Standard Model predict a slightly different magnetic strength and electrical behavior. Experiments are beginning to collect data with enough sensitivity to see the tiny effects predicted.

Looking beyond the earth, scientists studying solar neutrinos and cosmic-ray neutrinos, ghostly particles that barely interact at all, have recently established that neutrinos have masses, a result long expected by theorists studying extensions of the Standard Model [see "Solving the Solar Neutrino Problem," by Arthur B. McDonald, Joshua R. Klein and David L. Wark; SCIENTIFIC AMERICAN, April]. The next round of experiments will clarify the form of theory needed to explain the observed neutrino masses.

In addition, experiments are under way to detect mysterious particles that form the cold dark matter of the universe and to examine protons at higher levels of sensitivity to learn whether they decay. Success in either project would be a landmark of post–Standard Model physics.

As all this research proceeds, it is ushering in a new, data-rich era in particle physics. Joining the fray by about 2007 will be the Large Hadron Collider (LHC), a machine 27 kilometers in circumference now under construction at CERN, the European laboratory for particle physics near Geneva [see "The Large Hadron Collider," by Chris Llewellyn Smith; SCIENTIFIC AMERICAN, July 2000]. A 30-kilometer-long linear electron-positron collider that will complement the LHC's results is in the design stages.

As the first hints of post–Standard Model physics are glimpsed, news reports often make it sound as if the Standard Model has been found to be wrong, as if it were broken and ready to be discarded, but that is not the right way to think about it. Take the example of Maxwell's equations, written

NINA FINKEL

down in the late 19th century to describe the electromagnetic force. In the early 20th century we learned that at atomic sizes a quantum version of Maxwell's equations is needed. Later the Standard Model included these quantum Maxwell's equations as a subset of its equations. In neither case do we say Maxwell's equations are wrong. They are extended. (And they are still used to design innumerable electronic technologies.)

A Permanent Edifice

SIMILARLY, THE STANDARD MODEL is here to stay. It is a full mathematical theory—a multiply connected and highly stable edifice. It will turn out to be one piece of a larger such edifice, but it cannot be "wrong." No part of the theory can fail without a collapse of the entire structure. If the theory were wrong, many successful tests would be accidents. It will continue to describe strong, weak and electromagnetic interactions at low energies.

The Standard Model is very well tested. It predicted the existence of the *W* and *Z* bosons, the gluon and two of the heavier quarks (the charm and the top quark). All these particles were subsequently found, with precisely the predicted properties.

A second major test involves the electroweak mixing angle, a parameter that plays a role in describing the weak and electromagnetic interactions. That mixing angle must have the same value for every electroweak process. If the Standard Model were wrong, the mixing angle could have one value for one process, a different value for another and so on. It is observed to have the same value everywhere, to an accuracy of about 1 percent.

Third, the Large Electron-Positron (LEP) collider at CERN, which ran from 1989 to 2000, looked at about 20 million *Z* bosons. Essentially every one of them decayed in the manner expected by the Standard Model, which predicted the number of instances of each kind of decay as well as details of the energies and directions of the outgoing particles. These tests are but a few of the many that have solidly confirmed the Standard Model.

In its full glory, the Standard Model has 17 particles and about as many free parameters—quantities such as particle masses and strengths of interactions [*see box on pages 70 and 71*]. These quantities can in principle take any value, and we learn the correct values only by making measurements. Armchair critics sometimes compare the Standard Model's many parameters with the epicycles on epicycles that medieval theorists used to describe planetary orbits. They imagine that the Standard Model has limited predictive power, or that its content is arbitrary, or that it can explain anything by adjusting of some parameter.

The opposite is actually true: once the masses and interaction strengths are measured in any process, they are fixed for the whole theory and for any other experiment, leaving no freedom at all. Moreover, the detailed forms of all the Standard Model's equations are determined by the theory. Every parameter but the Higgs boson mass has been measured. Until we go beyond the Standard Model, the only thing that can change with new results is the precision of our knowledge of the parameters, and as that improves it becomes harder, not easier, for all the experimental data to remain consistent, because measured quantities must agree to higher levels of precision.

Adding further particles and interactions to extend the Standard Model might seem to introduce a lot more freedom, but this is not necessarily the case. The most widely favored extension is the Minimal Supersymmetric Standard Model (MSSM). Supersymmetry assigns a superpartner particle to every particle species. We know little about the masses of those superpartners, but their interactions are constrained by the supersymmetry. Once the masses are measured, the predictions of the MSSM will be even more tightly constrained than the Standard Model because of the mathematical relations of supersymmetry.

Ten Mysteries

IF THE STANDARD MODEL works so well, why must it be extended? A big hint arises when we pursue the long-standing goal of unifying the forces of nature. In the Standard Model, we can extrapolate the forces and ask how they would behave at much higher energies. For example, what were the forces like in the extremely high temperatures extant soon after the big bang? At low energies the strong force is about 30 times as powerful as the weak force and more than 100 times as powerful as electromagnetism. When we extrapolate, we find that the strengths of these three forces become very similar but are never all exactly the same. If we extend the Standard Model to the MSSM, the forces become essentially identical at a specific high energy [*see box on opposite page*]. Even better, the gravitational force approaches the same strength at a slightly higher energy, suggesting a connection between the Standard Model forces and gravity. These results seem like strong clues in favor of the MSSM.

Other reasons for extending the Standard Model arise from phenomena it cannot explain or cannot even accommodate:

1. All our theories today seem to imply that the universe should contain a tremendous concentration of energy, even in the emptiest regions of space. The gravitational effects of this so-called vacuum energy would have either quickly curled up the universe long ago or expanded it to much greater size. The Standard Model cannot help us understand this puzzle, called the cosmological constant problem.

2. The expansion of the universe was long believed to be slowing down because of the mutual gravitational attraction of all the matter in the universe. We now know that the expansion is accelerating and that whatever causes the acceleration (dubbed "dark energy") cannot be Standard Model physics.

3. There is very good evidence that in the first fraction of a second of the big bang the universe went through a stage of extremely rapid expansion called inflation. The fields responsible for inflation cannot be Standard Model ones.

4. If the universe began in the big bang as a huge burst of energy, it should have evolved into equal parts matter and antimatter (CP symmetry). But instead the stars and nebulae are made of protons, neutrons and electrons and not their antiparticles (their antimatter equivalents). This matter asymmetry cannot be explained by the Standard Model.

5. About a quarter of the universe is invisible cold dark matter that cannot be particles of the Standard Model.

6. In the Standard Model, interactions with the Higgs field (which is associated with the Higgs boson) cause particles to have mass. The Standard Model cannot explain the very special forms that the Higgs interactions must take.

7. Quantum corrections apparently make the calculated Higgs boson mass huge, which in turn would make all particle masses huge. That result cannot be avoided in the Standard Model and thus causes a serious conceptual problem.

8. The Standard Model cannot include gravity, because it does not have the same structure as the other three forces.

9. The values of the masses of the quarks and leptons (such as the electron and neutrinos) cannot be explained by the Standard Model.

10. The Standard Model has three "generations" of particles. The everyday world is made up entirely of first-generation particles, and that generation appears to form a consistent theory on its own. The Standard Model describes all three generations, but it cannot explain why more than one exists.

In expressing these mysteries, when I say the Standard Model *cannot* explain a given phenomenon, I do not mean that the theory has not yet explained it but might do so one day. The Standard Model is a highly constrained theory, and it cannot *ever* explain the phenomena listed above. Possible explanations do exist. One reason the supersymmetric extension is attractive to many physicists is that it can address all but the second and the last three of these mysteries. String theory (in which particles are represented by tiny, one-dimensional entities instead of point objects) addresses the last three [see "The Theory Formerly Known as Strings," by Michael J. Duff; Scientific American, February 1998]. The phenomena that the Standard Model cannot explain are clues to how it will be extended.

It is not surprising that there are questions that the Standard Model cannot answer—every successful theory in science has increased the number of answered questions but has left some unanswered. And even though improved understanding has led to new questions that could not be formulated earlier, the number of unanswered fundamental questions has continued to decrease.

Some of these 10 mysteries demonstrate another reason why particle physics today is entering a new era. It has become clear that many of the deepest problems in cosmology have their solutions in particle physics, so the fields have merged into "particle cosmology." Only from cosmological studies could we learn that the universe is matter (and not antimatter) or that the universe is about a quarter cold dark matter. Any theoretical understanding of these phenomena must explain how they arise as part of the evolution of the universe after the big bang. But cosmology alone cannot tell us what particles make up cold dark matter, or how the matter asymmetry is actually generated, or how inflation originates. Understanding of the largest and the smallest phenomena must come together.

The Higgs

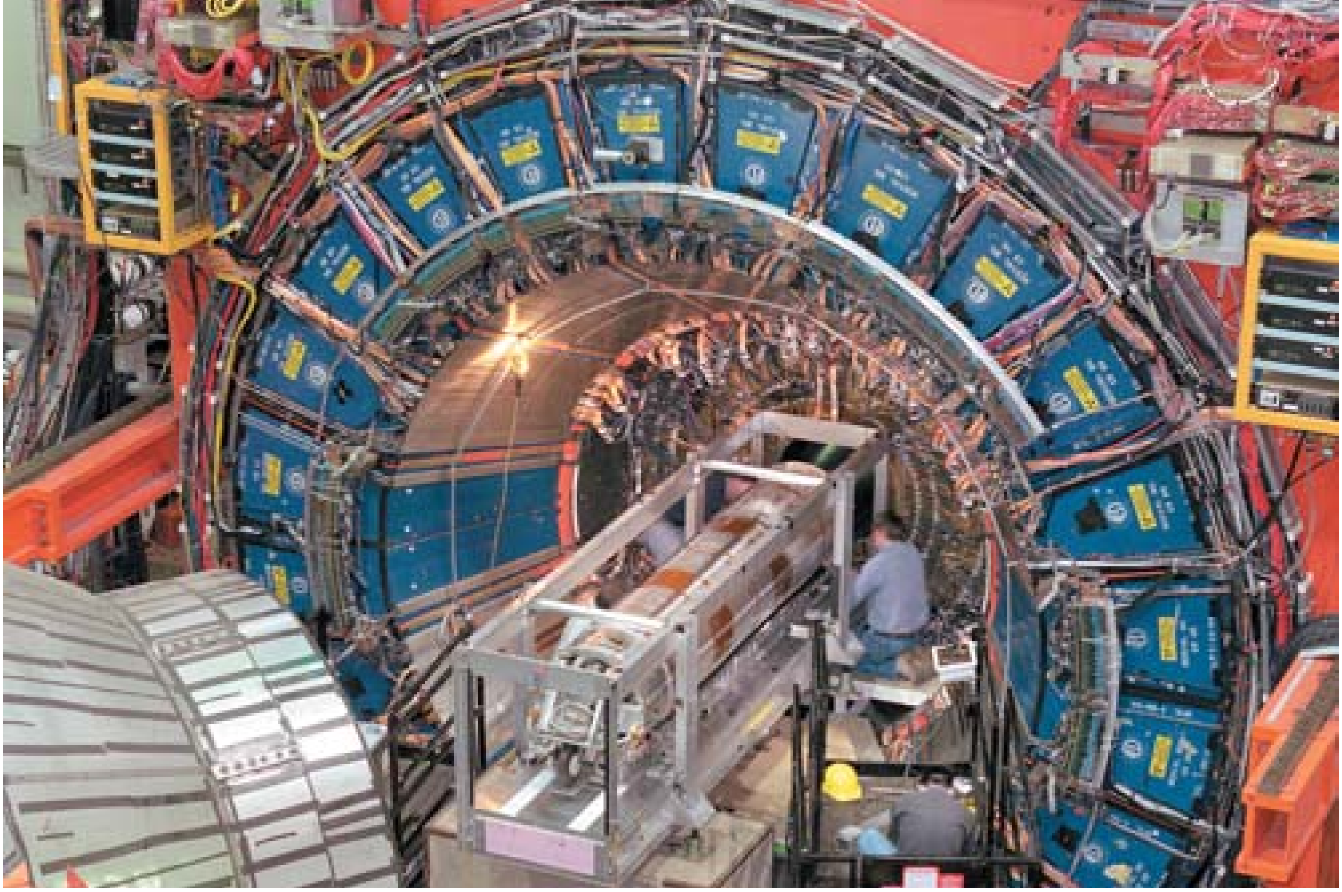
PHYSICISTS ARE TACKLING all these post–Standard Model mysteries, but one essential aspect of the Standard Model also remains to be completed. To give mass to leptons, quarks, and *W* and *Z* bosons, the theory relies on the Higgs field, which has not yet been directly detected.

The Higgs is fundamentally unlike any other field. To understand how it is different, consider the electromagnetic field. Electric charges give rise to electromagnetic fields such as those all around us (just turn on a radio to sense them). Electromagnetic fields carry energy. A region of space has its lowest possible energy when the electromagnetic field vanishes throughout it. Zero field is the natural state in the absence of charged particles. Surprisingly, the Standard Model requires that the lowest energy occur when the Higgs field has a specific nonzero value. Consequently, a nonzero Higgs field permeates the universe, and particles always interact with this field, traveling through it like people wading through water. The interaction gives them their mass, their inertia.

Associated with the Higgs field is the Higgs boson. In the Standard Model, we cannot predict any particle masses from first principles, including the mass of the Higgs boson itself. One can, however, use other measured quantities to calculate some masses, such as those of the *W* and *Z* bosons and the top quark. Those predictions are confirmed, giving assurance to the underlying Higgs physics.

Physicists do already know something about the Higgs mass. Experimenters at the LEP collider measured about 20 quantities that are related to one another by the Standard Model. All the parameters needed to calculate predictions for those quantities are already measured—except for the Higgs boson mass. One can therefore work backward from the data and ask which Higgs mass gives the best fit to the 20 quantities. The answer is that the Higgs mass is less than about 200 giga-electron-volts (GeV). (The proton mass is about 0.9 GeV; the top quark 174 GeV.) That there is an answer at all is strong evidence that the Higgs exists. If the Higgs did not exist and the Standard Model were wrong, it would take a remarkable coincidence for the 20 quantities to be related in the right way to be consistent with a specific Higgs mass. Our confidence in this procedure is bolstered because a similar approach accurately predicted the top quark mass before any top quarks had been detected directly.

LEP also conducted a direct search for Higgs particles, but it could search only up to a mass of about 115 GeV. At that very upper limit of LEP's reach, a small number of events involved particles that behaved as Higgs bosons should. But there were not enough data to be sure a Higgs boson was actually discov-



UPGRADING of the Tevatron's huge particle detectors, which was carried out by physicists at Fermilab from 1996 to 2000, has primed the facility for observing Higgs bosons and supersymmetry.

ered. Together the results suggest the Higgs mass lies between 115 and 200 GeV.

LEP is now dismantled to make way for the construction of the LHC, which is scheduled to begin taking data in four years. In the meantime the search for the Higgs continues at the Tevatron at Fermilab [*see illustration above*]. If the Tevatron operates at its design intensity and energy and does not lose running time because of technical or funding difficulties, it could confirm the 115-GeV Higgs boson in about two to three years. If the Higgs is heavier, it will take longer for a clear signal to emerge from the background. The Tevatron will produce more than 10,000 Higgs bosons altogether if it runs as planned, and it could test whether the Higgs boson behaves as predicted. The LHC will be a "factory" for Higgs bosons, producing millions of them and allowing extensive studies.

There are also good arguments that some of the lighter superpartner particles predicted by the MSSM have masses small enough so that they could be produced at the Tevatron as well. Direct confirmation of supersymmetry could come in the next few years. The lightest superpartner is a prime candidate to make up the cold dark matter of the universe—it could be directly observed for the first time by the Tevatron. The LHC will produce large numbers of superpartners if they exist, definitively testing whether supersymmetry is part of nature.

Effective Theories

TO FULLY GRASP the relation of the Standard Model to the rest of physics, and its strengths and limitations, it is useful to think in terms of effective theories. An effective theory is a description of an aspect of nature that has inputs that are, in principle at least, calculable using a deeper theory. For example, in nuclear physics one takes the mass, charge and spin of the proton as inputs. In the Standard Model, one can calculate those quantities, using properties of quarks and gluons as inputs. Nuclear physics is an effective theory of nuclei, whereas the Standard Model is the effective theory of quarks and gluons.

From this point of view, every effective theory is open-ended and equally fundamental—that is, not truly fundamental at all. Will the ladder of effective theories continue? The MSSM solves a number of problems the Standard Model does not solve, but it is also an effective theory because it has inputs as well. Its inputs might be calculable in string theory.

Even from the perspective of effective theories, particle physics may have special status. Particle physics might increase our understanding of nature to the point where the theory can be formulated with no inputs. String theory or one of its cousins might allow the calculation of all inputs—not only the electron mass and such quantities but also the existence of spacetime and the rules of quantum theory. But we are still an effective theory or two away from achieving that goal. SA

MORE TO EXPLORE

The Particle Garden. Gordon Kane. Perseus Publishing, 1996.

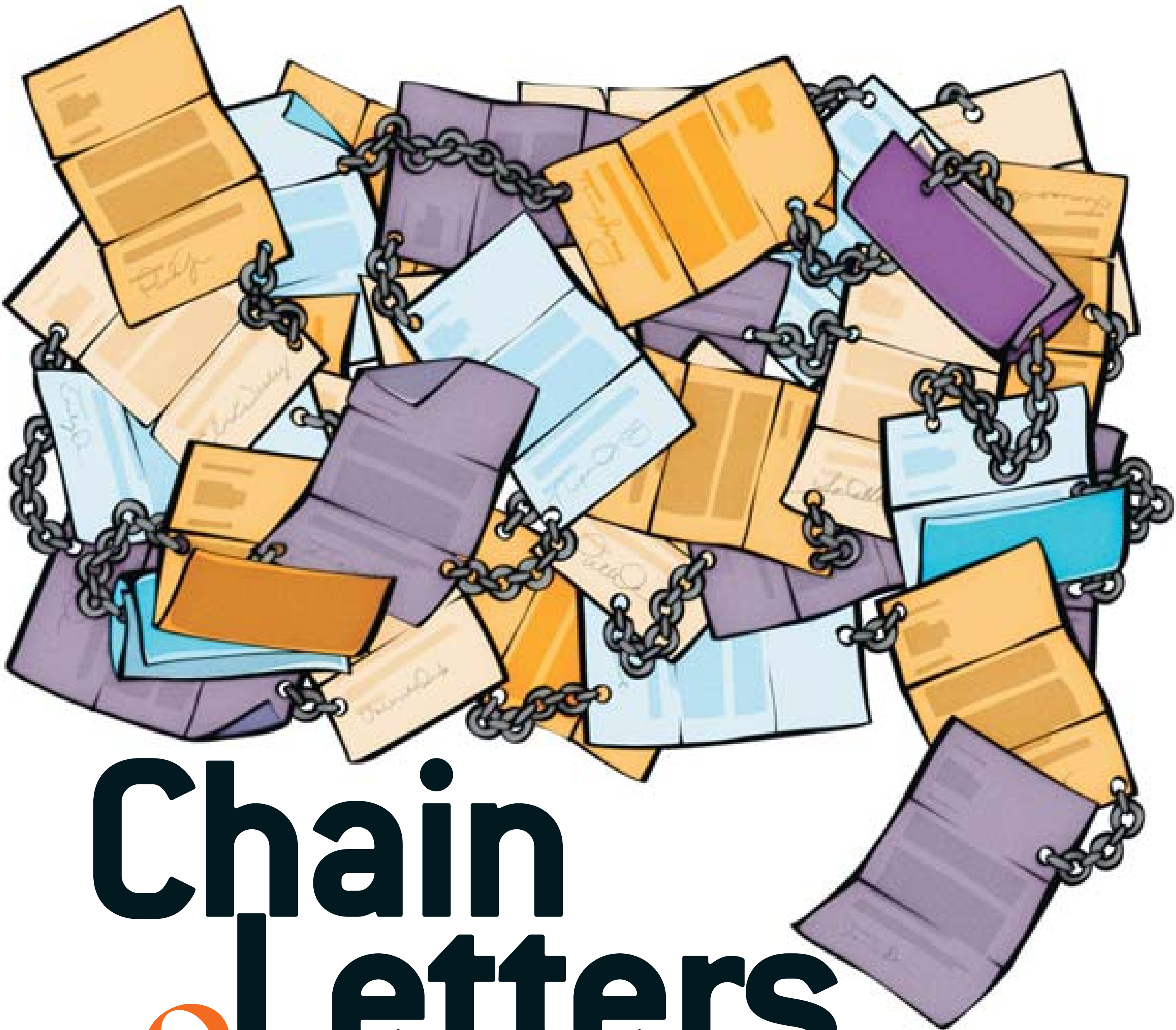
The Rise of the Standard Model: A History of Particle Physics from 1964 to 1979. Edited by Lillian Hoddeson, Laurie M. Brown, Michael Riordan and Max Dresden. Cambridge University Press, 1997.

The Little Book of the Big Bang: A Cosmic Primer. Craig J. Hogan. Copernicus Books, 1998.

Supersymmetry: Unveiling the Ultimate Laws of Nature. Gordon Kane. Perseus Publishing, 2001.

An excellent collection of particle physics Web sites is listed at **particleadventure.org/particleadventure/other/othersites.html**

FERMILAB PHOTO



Chain Letters & Evolutionary Histories

A study of chain letters shows how to infer the family tree of anything that evolves over time, from biological genomes to languages to plagiarized schoolwork

BY CHARLES H. BENNETT, MING LI AND BIN MA

IN OUR HANDS ARE 33 VERSIONS OF A CHAIN LETTER,

collected between 1980 and 1995, when photocopiers, but not e-mail, were in widespread use by the general public. These letters have passed from host to host, mutating and evolving. Like a gene, their average length is about 2,000 characters. Like a potent virus, the letter threatens to kill you and induces you to pass it on to your "friends and associates"—some variation of this letter has probably reached millions of people. Like an inheritable trait, it promises benefits for you and the people you pass it on to. Like genomes, chain letters undergo natural selection and sometimes parts even get transferred between coexisting "species." Unlike DNA, however, these letters are easy to read. Indeed, their readability makes them especially suitable for classroom teaching of phylogeny (evolutionary history) free from the arcana of molecular biology.

The letters are an intriguing social phenomenon, but we are also interested in them because they provide a test bed for the algorithms used in molecular biology to infer phylogenetic trees from the genomes of existing organisms. We believe that if these algorithms are to be trusted, they should produce good results when applied to chain letters. Using a new algorithm that is general enough to have wide applicability to such problems, we have reconstructed the evolutionary history of our 33 letters [*see illustration on page 79*]. The standard methods do not work as well on these letters. Originally developed for genomes, our algorithm has also been applied to languages and used to detect plagiarism in student assignments: anything involving a sequence of symbols is grist for its mill.

A Mind Virus

THE 33 CHAIN LETTERS are a fascinating collection. We labeled them arbitrarily from L1 to L33. The letters differ significantly. There are 15 titles, 23 names for "an office employee," and 25 names for the original author of the letter. Misspellings, swapped sentences, and missing or added phrases, sentences and paragraphs are common. (A typical letter is shown on the next page, along with some of the many variations.) Nearly all are more or less faded photocopies of typescripts, leading us to surmise that the mutations arose by an intermittent process, whereby a letter would be photocopied for several generations until its legibility was so reduced that the next recipient decided to retype it, introducing new errors and variations.

All but three of the letters we received were unique; for L4, L6 and L22, a second copy arrived within a few months of the first. Besides the 33 English-language letters, we received (but did not include in our study) four in French and one each in Dutch and German, all clearly sharing a common ancestry with the ones in English.

To analyze the letters, we retyped them into computer files entirely in lowercase, ignoring extra information such as dates and marginal notes as well as the division of the text into lines and paragraphs. Each letter became one continuous string of characters.

Before we applied our new algorithm, we tried analyzing the letters with a procedure called multiple alignment, which is widely used for examining genes to infer phylogeny. This method attempts to line up as many matching sections of all the letters as possible. The amount of matching between any pair of letters defines their similarity, and from that data another algorithm constructs an evolutionary tree. Unfortunately, multiple alignment only finds matches with everything remaining in the same order, so it gets confused by L12 and L26, in which the order of sentences has been rearranged. For the same reason, the technique is known to work better within individual genes than for whole genomes, in which such translocations are more common.

We tried omitting L12 and L26 and then performing multiple alignment on the remaining 31 letters. Even with this truncated set, the resulting tree seemed wrong, classifying L6, L7 and L13 as closely related. This error occurred because those three letters are all relatively short, giving them a correspondingly small number of differences. This problem can arise in genetics as well: merely counting differences can overestimate the similarity of short genomes while underestimating that of long ones. A proper measure should give more weight to a small difference in a small genome than to a small difference in a big genome.

We turned to devising our own similarity measure, one that could be applied to genomes, chain letters or any other type of data that might be stored as a computer file. We wanted to make our new similarity measure insensitive to minor mix-ups such as translocations, which represent only a small loss of informational similarity. To cope with differences in length, we wanted our measure to assign two completely dissimilar data files a score of 0 and two identical ones a score of 1, regardless of their sizes.

The natural measure of the information content in a data file is not its raw size in bits but rather the smallest size it can be compressed to by a file compression program such as zip or StuffIt. These programs are designed to save space on hard disks by finding and squeezing out the most common kinds of redundancy (for instance, repeated phrases), resulting in a smaller file from which the original can be perfectly reconstructed when needed.

Something interesting happens if we compress two files together so that both can be regenerated from the compressed file. If the two files share no information at all, the joint compressed file will be as big as the two individual compressed files combined. But if the two files contain some of the same information,

SARA CHEN

THE CHAIN LETTERS ARE ESPECIALLY SUITABLE FOR CLASSROOM TEACHING **OF PHYLOGENY.**

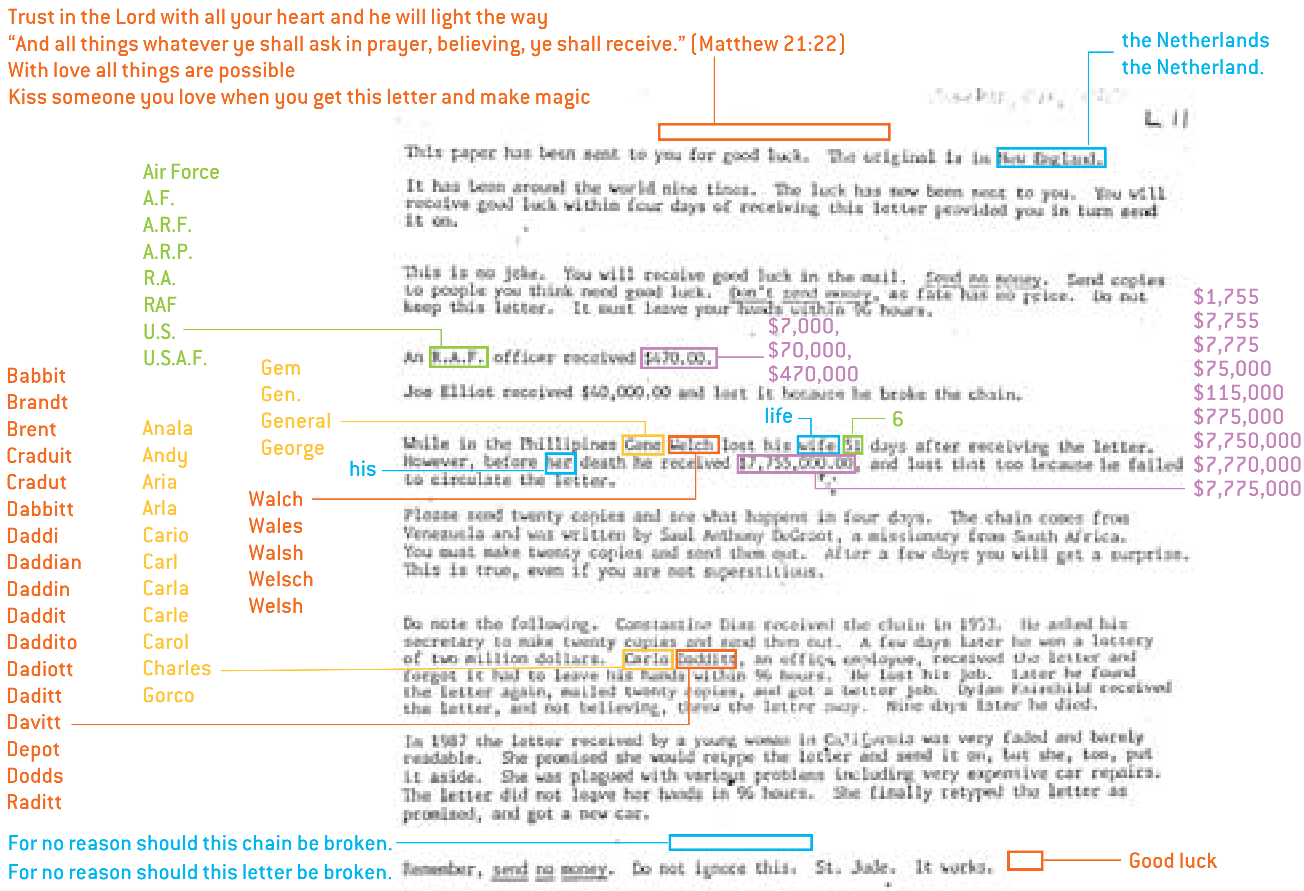
that repetition will be detected by a good compression program, and the joint compressed file will be smaller. In this way, the size of the joint compressed file compared with the sum of the individual compressed files provides a measure of the files' similarity.

That measure is not yet a good one for our purposes, because two large files will tend to have greater similarity than two small files. To correct this problem, we define our "relatedness" measure to be the *proportion* of shared information—that is, the percentage by which the sum of the separately compressed files exceeds the size of the jointly compressed file. This makes the relatedness range from 0 for unrelated files to 1 (or 100 percent) for identical files, regardless of length.

Which compression program should we use? Obviously, our relatedness measure will depend on that. Ideally, we would want to use a program that compresses every file to the smallest possible size. The study of information measures defined in terms of such ultimate compressibility forms an elegant branch of information theory known as algorithmic information theory or Kolmogorov complexity (after the late Russian mathematician Andrei N. Kolmogorov, one of its founders). Unfortunately, information theorists have proved that such an ideal zip program would take essentially an infinitely long time to perform its task. For our purposes, then, we decided to use a particular compression algorithm called GenCompress, created by Xin Chen of the University of California at Santa Barbara. GenCompress was designed for genomes and works well on them. As we shall see, it also works well on chain letters.

THEME AND VARIATIONS

A SAMPLE CHAIN LETTER, labeled L11, illustrates some of the ways that related letters changed when they were retyped (presumably after photocopies became illegible). The greatest variations occurred with unfamiliar names and quantities of money—errors in those elements are easily overlooked because they do not change the meaning of the letter.



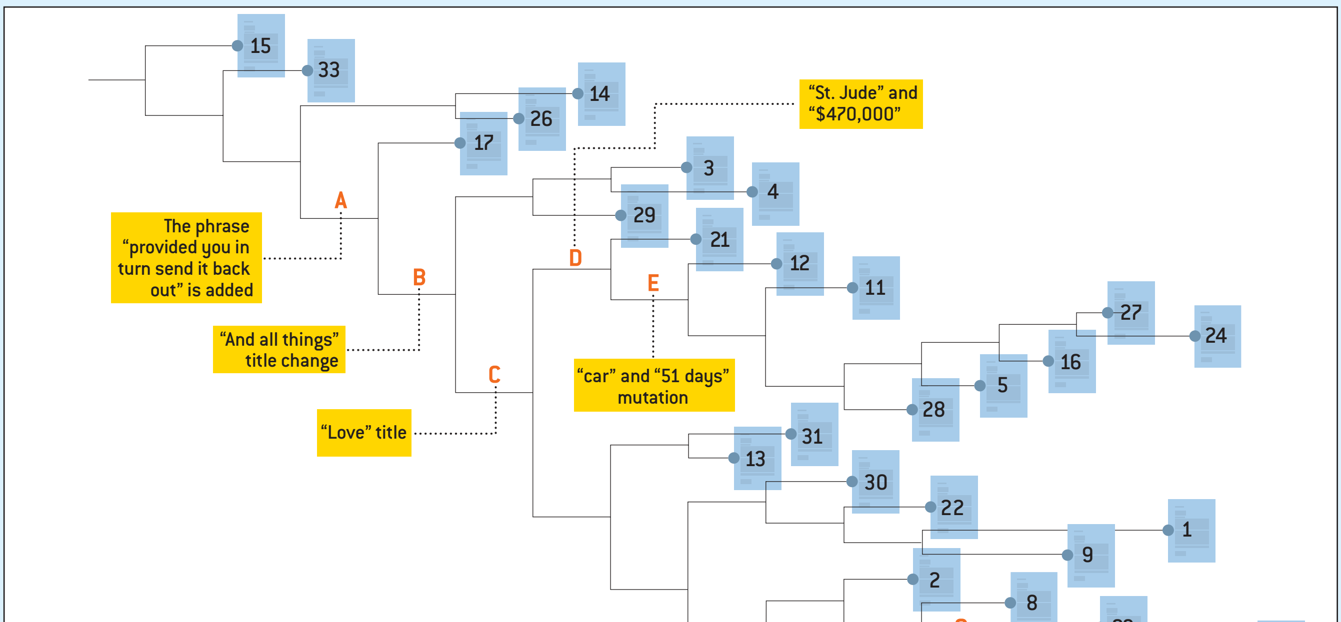
How the Letters Evolved

THE EVOLUTIONARY TREE of the chain letters deduced by the authors' automated "relatedness" measure shows a number of interesting features. (Some other features are discussed in the main text.) At point A, a phrase, "provided you in turn send it back out," was adopted. At point B, a new title evolved: "And all things whatever ye shall ask in prayer, believing, ye shall receive." At point C, the title further mutated to "With love all things are possible." Also at point C, "The Netherlands" changed to "New England," and "General Welch" became "Gene Welch." The sentence "For no reason should it be broken" got lost.

At point F, the title mutated to "Kiss someone you love when you get this letter and make magic." After we had finished our analysis, we found a very interesting and comprehensive study of more than 460 chain letters of many varieties by mathematician Daniel W. VanArsdale. His study raised the question, Which title came first: the "Kiss" or the "Love" one? We have not yet applied our algorithm to the many chain letters in his collection, but judging from our phylogeny, the "Love" title came first. This conclusion is further supported by the amount of money Gene Welch received: in all the "Kiss" letters (except for the mutation at G), he received \$7,755, whereas in the "Love" group he got \$7,755,000. In the group before C, the figure was \$775,000. The mutation sequence \$775,000 → \$7,755,000 → \$7,755 is clearly more parsimonious than \$775,000 → \$7,755 → \$7,755,000.

At point D, \$70,000 grew to \$470,000, and "St. Jude" was adopted. Neither \$470,000 nor St. Jude appears outside of this group. At point E, two interesting alterations happened concurrently: the California woman's car story was added and the time between Gene Welch's receiving the letter and losing his wife changes from six days to 51 days. This is consistent across all letters in the group rooted at E, except for L28, which requires a special explanation. The "car and 51" mutation doesn't appear anywhere outside of the E group.

L28 provides evidence of horizontal transfer—that is, the transfer of information from one "organism" to another in addition to simple inheritance. In the group rooted at D, every letter has each R.A.F. officer receive \$470,000 or \$470, except for L28, in which the amount is \$70,000. L28 also has Gen. Welch, who otherwise features only in letters before point C. All the letters in the D group except L21 have the car story, and all but L21 and L28 have the "51 days" mutation. It seems unbelievable to assume that L28 gained "\$70,000" and "Gen. Welch" by mutation independently of the other instances of these mutations elsewhere in the tree. One might try placing the "car and 51" mutation before the "\$470,000 and St. Jude" mutation, but then L21 must undergo a very implausible genesis: either it must lose the whole car story and mutate "51 days" back to "six days," or it must gain "\$470,000" and "St. Jude" independently. Apparently somebody had two letters in his or her possession while composing L28 (or L21) and introduced a foreign gene from a letter before C. *—C.H.B., M.L. and B.M.*



THE AUTHORS' ANALYSIS of the similarities among 33 randomly numbered chain letters produced this tree that tracks how they evolved over time. A few of the more significant mutations, or changes, are indicated at the points in the tree that they occurred.

SARA CHEN

THE ACCUMULATION OF MUTATIONS IN A MITOCHONDRIAL GENOME (OR A CHAIN LETTER) ACTS AS A CLOCK.

Given a set of chain letters, it is a straightforward and entirely automatic process to calculate the relatedness of each pair using the GenCompress program. The next step, converting the relatedness data into an evolutionary tree, is also largely automatic (many software packages exist for this purpose). The result can either be a simple tree diagram with arbitrary branch lengths of the kind shown on the preceding page, indicating simply the qualitative pattern of descent, or a more detailed diagram, with branch lengths that represent relatedness quantitatively.

In either case, the main human input is deciding where to put the root of the tree, which represents the hypothesized common ancestor of all the letters (or species). In biological phylogenies, the root stands for an extinct species from millions of years ago, so it should not be too closely related to any of the branches denoting organisms alive today. In our study, the chain letters were collected over a 15-year period, and some of them were dated near the beginning of that time, so we chose to put the root near one of these (L15). Unfortunately, most of the letters were collected without recording their postmark or date of receipt, reflecting the fact that this project began as a hobby and only belatedly evolved into scientific research.

St. Jude's Phylogeny

THE EVOLUTIONARY TREE that was inferred for the chain letters appears to be almost a "perfect" phylogeny, in the sense that documents that share the same characteristic are always grouped together. After the tree was built, we were able to use it to help make numerous hypotheses about how the letters evolved.

First we judge that the letters before point C in our phylogeny are the oldest [*see illustration on preceding page*]. The chief evidence is that the name "Carlo Dadditt" and the title of the letter had the most mutations in this group of letters. We expect such errors to be more common in the oldest letters because photocopiers were less available at that time and retyping more frequent. In addition, among the 14 dated letters, the two that occur in the pre-C group (L4 and L15) are the oldest. These older letters are all titled with religious prayers, come from "The Netherlands" and contain the sentence "For no reason should it be broken."

Next we see an effect familiar from molecular biology, in which different parts of the genome have quite different mutation rates. Active sites of enzymes mutate scarcely at all, whereas parts far from the active site continually undergo random drift. Similarly, with chain letters the parts required for "viability" do not mutate at all, but more arbitrary parts, such as the types of mishaps that would befall those who did not propagate the chain, mutate more. Parts with little intrinsic meaning to help catch errors—for instance, unfamiliar names like "Gem Walsh" and "Carlo Craduit"—mutate the most.

Another biological phenomenon appearing in the chain letters is the occurrence of parallel, compensating mutations: two mutations that would be detrimental individually and that must occur together to be neutral or beneficial. Excluding L12 and L26, in which no one dies, all letters before point C (except nearby L29) read:

> General Welsh (or a variation)
> lost his life ... however before
> his death ...

On the other hand, letters after point C read:

> Gene Welch (or a variation) lost
> his wife ... however before her
> death ...

To preserve the sense, "his" mutates to "her" when "life" mutates to "wife." [*See box on preceding page for more of these observations.*]

Mammals and Plagiarism

BESIDES ANALYZING chain letters, our relatedness measure has been used in a wide range of settings. In bioinformatics itself, we used it to analyze the mitochondrial genomes of 18 mammals. Mitochondria are energy-producing organelles within cells whose genes are inherited solely from the mother (similar to how a chain letter inherits from a single "parent"). Because no reshuffling of maternal and paternal genes occurs, the accumulation of mutations in the mitochondrial genome acts as a clock measuring when an organism's ancestors diverged from related species.

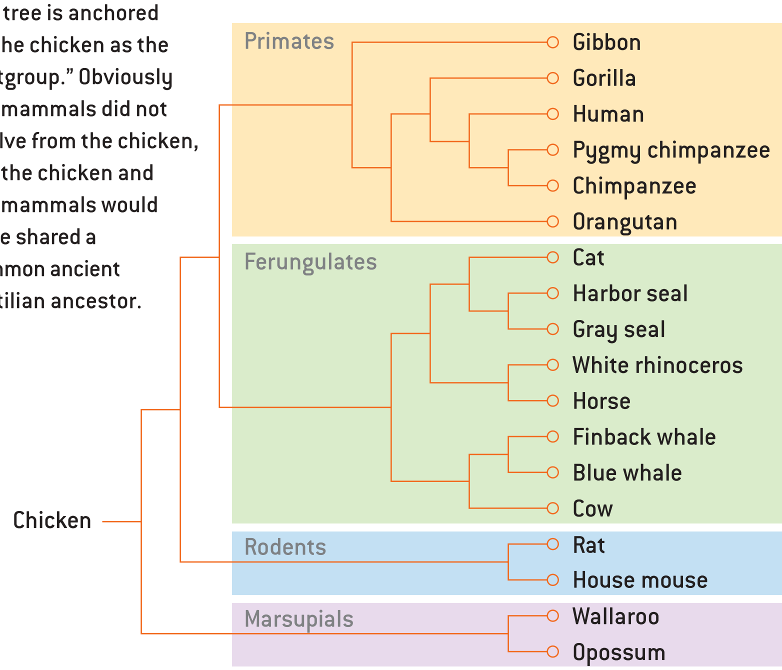
Traditional methods applied to different mitochondrial genes often give conflicting evolutionary trees, and many, unlike our new measure, cannot be applied successfully to an entire genome because of problems such as translocations. For example, using the traditional methods, about half a dozen mitochondrial genes imply that primates, such as ourselves, are more closely related to rodents than to ferungulates, a diverse group that includes cows, horses, whales, cats and dogs. Another half a dozen genes imply that primates and ferungulates are the more closely related pair, which is generally believed to be correct based on other multiple lines of evidence, such as nonmitochondrial genes and the fossil record.

THE AUTHORS

CHARLES H. BENNETT, MING LI and *BIN MA*'s joint work on chain letters began as a result of a hike in the Hong Kong mountains during which Bennett mentioned his collection to Li. Bennett is an IBM Fellow at the company's research division in Yorktown Heights, N.Y. He focuses on the physics of information processing and is a co-discoverer of quantum teleportation. Li studies bioinformatics and Kolmogorov complexity as professor of computer science at the University of Waterloo in Ontario. Ma works on bioinformatics and algorithm design as assistant professor of computer science at the University of Western Ontario.

RELATEDNESS OF MAMMALS

DIVERSE PROBLEMS can be analyzed with the authors' relatedness measure. Applied to whole mitochondrial genomes, it produced this phylogeny of mammals. Note how primates are more closely related to ferungulates than to rodents—which is believed to be true. This degree of kinship is ambiguous when using traditional techniques. The tree is anchored by the chicken as the "outgroup." Obviously the mammals did not evolve from the chicken, but the chicken and the mammals would have shared a common ancient reptilian ancestor.



When our method is applied to whole mitochondrial genomes, it produces this latter evolutionary tree without needing any ad hoc tinkering to resolve ambiguities or contradictions [*see illustration above*].

Taking the art of phylogenetic inference to an audacious extreme, Dario Benedetto, Emanuele Caglioti and Vittorio Loreto of La Sapienza University in Rome tried inferring a phylogeny of human languages not by analyzing the languages' known literatures or history but merely by applying a method similar to ours to 52 translations of the Universal Declaration of Human Rights. The result was surprisingly good, considering the tiny body of evidence on which it was based. One notable mistake was the classification of English as a romance language, closely related to French, whereas historically English evolved within the Germanic group. This error arises because of the great many French words that English acquired after the Norman Conquest (an example of parallel transfer).

Another application of our measure has been the detection of plagiarism in students' homework assignments. In one instance, two assignments in a computer programming class were flagged as being unusually alike, but the instructor could not see any obvious evidence of copying when he examined them himself. The two students were approached and, in the interest of research, given immunity to plagiarism charges in return for an honest account of whether they had collaborated. Apparently the two students had discussed the problem and how they planned to tackle it but had not worked together beyond that level. If that is what happened, our distance algorithm detected the subtle similarities engendered by their discussion!

The automatic nature of our procedure is both an advantage and a disadvantage. On the one hand, it yields objective answers, free from the need to weigh various lines of evidence (such as DNA versus the fossil record) or to take account of which parts of the genome mutate fastest. On the other hand, it does not benefit from the insights that might come from such additional data. All methods of phylogenetic inference are imperfect and will sometimes mistakenly infer a phylogeny that differs from what actually took place historically. Like historians and paleontologists, evolutionary molecular biologists have come to accept that no matter how many lines of evidence they consider, the full truth about the past can never be reconstructed. This is especially true with regard to extinct species. Many species that once existed will never be known, because they left neither fossils nor descendants. Likewise many languages have become extinct, vanishing without a trace even in the past century.

In the realm of chain letters, it is certain that many letters became extinct when too many recipients broke the chain. Like the lost plays of Sophocles, these letters' texts may never be recovered, and even their existence can only be surmised from circumstantial evidence, such as a rash of unemployment and expensive car repairs occurring in California for no apparent reason. SA

MORE TO EXPLORE

An Introduction to Kolmogorov Complexity and Its Applications. Second edition. Ming Li and Paul M. Vitányi. Springer-Verlag, 1997.

An Information-Based Sequence Distance and Its Application to Whole Mitochondrial Genome Phylogeny. Ming Li, Xin Chen, Jonathan H. Badger, Sam Kwong, Paul Kearney and Haoyong Zhang in *Bioinformatics,* Vol. 17, No. 2, pages 149–154; February 2001.

Language Trees and Zipping. Dario Benedetto, Emanuele Caglioti and Vittorio Loreto in *Physical Review Letters,* Vol. 88, No. 4, pages 048702-1–048702-4; January 28, 2002.

Chain Letter Evolution. Daniel W. VanArsdale: **www.silcom.com/~barnowl/chain-letter/evolution.html**

The chain letters used in this article and other data are at **www.math.uwaterloo.ca/~mli/chain.html**

A discussion of phylogenetic inference methods is at **helix.biology.mcmaster.ca/721/outline2/node47.html**

Kolmogorov complexity is discussed at **www.wikipedia.org/wiki/kolmogorov_complexity**

The program for plagiarism detection is at **http://dna.cs.ucsb.edu/SID/**

SARA CHEN

COCHLEAR IMPLANTS

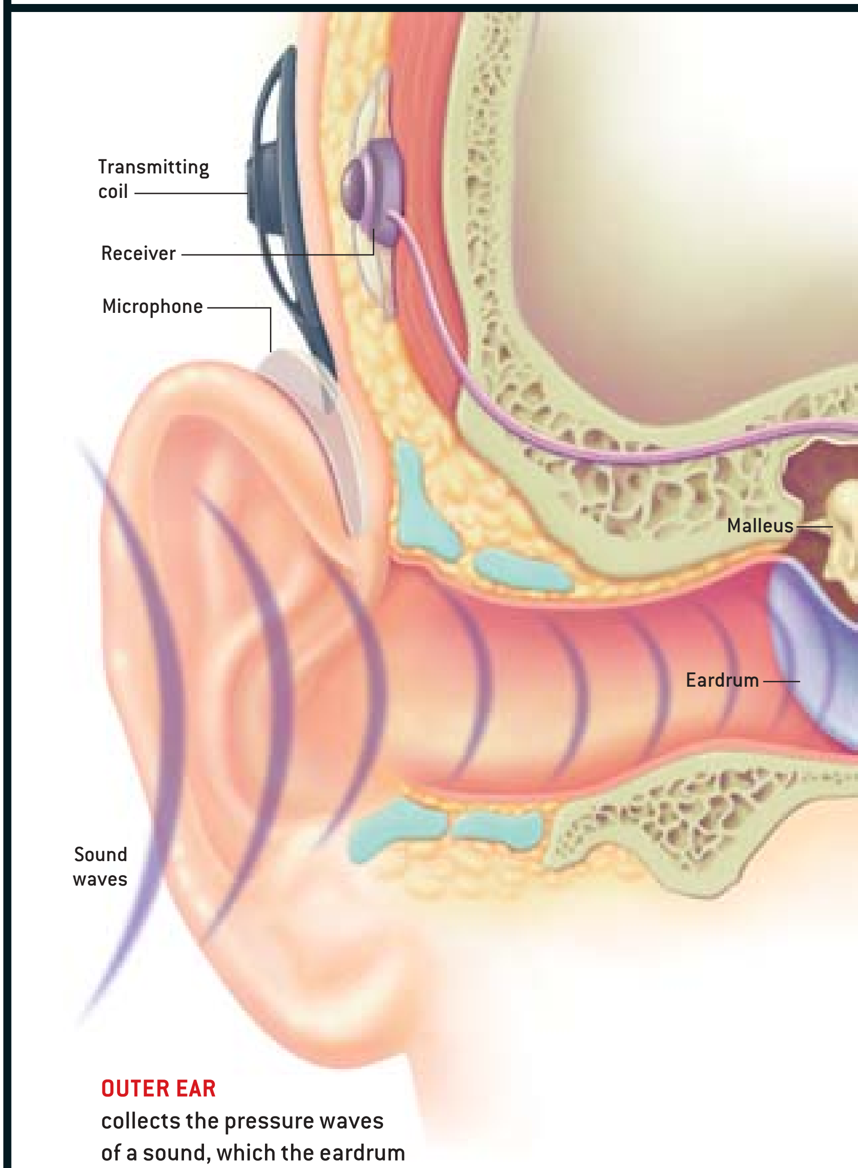
To Hear Again

We hear when the cochlea, in the inner ear, stimulates the auditory nerve. Deafness occurs most commonly when tiny hair cells inside the cochlea are damaged as a result of a genetic defect, infection, loud noise or aging. Cochlear implants bypass the damage by receiving and converting sound into signals sent along electrodes to cells adjacent to the auditory nerve.

Worldwide, more than 40,000 children and adults depend on cochlear implants. Certain hair cells lining the cochlear ducts alongside auditory neurons are tuned to respond to specific frequencies. Therefore, implants such as the Nucleus or the Clarion have from eight to 22 electrodes that surgeons place at different positions to maximize the range of frequency stimuli forwarded to the brain. Recent research indicates that more electrodes won't improve performance as much as optimizing their placement; most implant wearers perceive loudness properly but can still have trouble sensing pitch correctly, making speech comprehension difficult. "Something is preventing the brain from extracting or assimilating all the coding information," says Philip Loizou, a professor of electrical engineering at the University of Texas at Dallas. "We don't know what yet."

The sooner a person receives an implant after becoming deaf, the more likely he will adapt to the new sound input; people who have been deaf for years do not respond as well because of degeneration in the cochlea or auditory nerve. Typically, success means the wearer can hear moderate and perhaps soft sounds, can communicate without lip reading or signing, and may be able to converse over the telephone. Completely normal hearing is still uncommon.

To help, engineers are trying, among other things, to tailor signal-processing algorithms for particular situations. People with implants often have trouble perceiving speech clearly in noisy environments and appreciating music, which has complex waveforms. Today's implant processors come with a general algorithm, but perhaps they could store specialized ones as well. "That way," Loizou says, "an individual could select an algorithm depending on whether he was talking at home, eating in a noisy restaurant or sitting at a concert." *—Mark Fischetti*



OUTER EAR collects the pressure waves of a sound, which the eardrum converts into mechanical vibrations in tiny bones in the middle ear. The oscillating stapes sets off pressure waves within the fluid in the cochlea, which in turn stimulate nerve cells on the auditory nerve that leads to the brain.

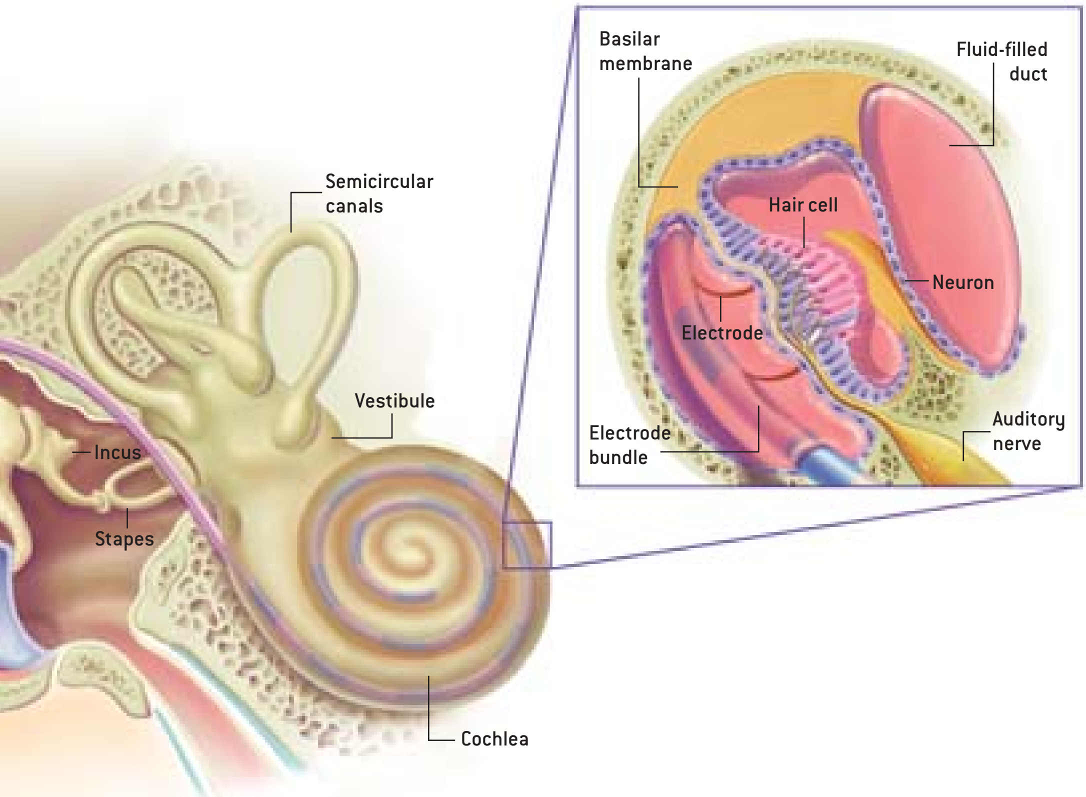
ALICE Y. CHEN

DID YOU KNOW . . .

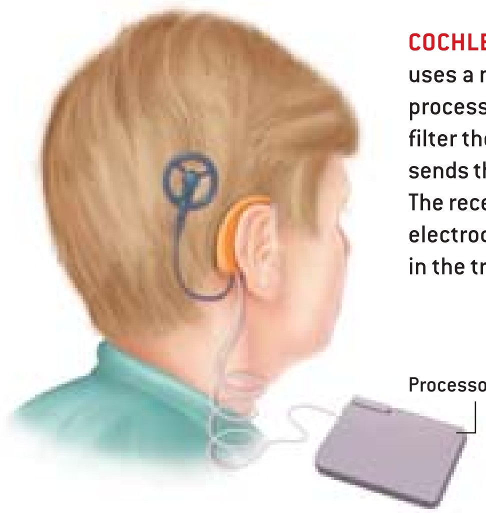
➤ LOUD! Audiologists measure how loud a test signal must become before a person can hear it. The normal threshold is anything up to 15 decibels. People who have "moderate" hearing loss require a signal of 41 to 55 dB; "severe" loss, 71 to 90 dB; and "profound" loss, 91 dB or greater. Typical conversation rings in at 40 to 50 dB; freeway traffic at 50 feet, 70 dB; and a blender, 90 dB. Only people with severe or profound loss in both ears qualify for cochlear implants.

➤ HAIRPIN TURN: Tipping the head forward, back or to the side tilts a gelatinous substance inside the vestibule, beside the cochlea. The shifting gel bends hair cells that tell attached nerve fibers which way the head is headed, so the brain can keep the body upright. Similarly, hair cells inside three fluid-filled semicircular canals, at right angles to one another just above the cochlea, respond to sudden changes in the head's velocity, to maintain balance.

➤ HARD ON HEARING: In its "Position Statement on Cochlear Implants," the U.S. National Association of the Deaf notes that the technology doesn't always work. It also discourages the use of implants in children who are born deaf or become deaf before learning language, because even with the technology, it is very hard for them to develop the cognition for spoken language. Meanwhile the children are often not taught visual, or sign, languages, which "can result in developmental delays that can be extremely difficult to reverse."



COCHLEA transmits pressure waves through its duct fluid, displacing the basilar membrane, which bends hair cells to various degrees. The cells release neurotransmitters that cause attached neurons to fire, telling the brain where along the duct the bending has occurred (which corresponds to the frequency of the original sound) as well as the amplitude of the bending (which indicates the loudness).



COCHLEAR IMPLANT uses a microphone worn behind the ear to pick up sound and send it to a processor, where integrated circuits and algorithms amplify, digitize and filter the sound into a coded signal sent to the transmitter coil. The coil sends the signal via radio waves through the skin to an implanted receiver. The receiver converts the waves into electrical impulses that travel along electrodes that end at cells at certain points along the cochlea. A magnet in the transmitter holds it against the implanted receiver.

This month's column was suggested by reader Thomas Boehm. Have a topic for a future column? Send it to workingknowledge@sciam.com

Name That Tune

MUSIC LOVERS USE CELL PHONES TO IDENTIFY SONGS BY FIONA HARVEY

The song is irritatingly familiar. You've heard it before, stacks of times. Its name is on the tip of your tongue, but you just can't place it. Hum along a little, maybe that'll help. No? Oh, what is the darn thing called?

Technology should be able to resolve this dilemma. Songs have been available in digital form for many years. Couldn't you identify a tune by matching it to a song stored in an enormous musical database? Alas, it's not that easy. "When I started looking at this question, I was told by all the experts in the field that this was a next-to-impossible task," says Avery Wang, chief scientist at Shazam Entertainment, an English company that specializes in music-recognition software.

The problem is the huge number of possible combinations of musical notes. The potential database is so large that even the best software algorithms would be hard put to search it. Some research institutions have tried to employ probability analysis to ease the task of distinguishing songs. "With every note that is played, the next note can be either higher, lower or the same," explains Mark Sandler of Queen Mary College at the University of London. "By analyzing the patterns of vast amounts of existing music, you can use the preceding notes to work out the probability of where that next note will be on the scale." Such analysis, however, takes time and painstaking effort. And one must also take into account the individual interpretations that composers, conductors and musicians make, the changes to tempo and style that can completely alter the texture of a piece, and the fact that the copying and sampling of certain passages of music within other pieces is commonplace.

THIS IS YOUR SONG: A new music-recognition system developed by Shazam Entertainment is now available in Britain. Just point your cell phone at the speakers, and you receive a text message giving the name of the song and the artist—as long as the background noise is not too loud.

JOHN FRASER

Nevertheless, the first breakthrough is here. Shazam Entertainment has created a commercial music-recognition system, now available in Britain, that allows people to identify tunes using their cell phones. When you hear a song—on the radio, over a public address system, in a bar or on television—you punch in the Shazam service's four-digit code number on the handset, point the phone at the source of the sound and hold it there for 15 seconds. Within a few minutes, the service should return a text message giving the name of the song and the artist.

An American, Chris Barton, co-founded Shazam three years ago after he dreamed up the idea while studying at the London Business School. Wang was working on sound analysis and manipulation at Stanford University when Barton caught up with him. "At first, I resisted," Wang says. "The general current of academic thought was that this was such a very hard problem that it would take years of work." But then Wang had a brainstorm. Instead of messing around with the probabilities of different notes following one another, he plotted graphs of musical frequency against time. Creating digital signatures of musical pieces had been done before—you can do it on a home computer with ordinary audio software. Such signatures are complex, though, and finding a match requires the entire graph to be compared with the graphs of all other tunes, which takes time.

Wang's innovation was to eliminate as much information from the signatures as he could while ensuring that they were still recognizable and unique to each piece of music. For each song, Wang recorded only the peak and trough of frequency in every 10-millisecond segment. Shazam's database currently contains the stripped-down digital signatures of 1.7 million songs. When a user with a mobile phone inputs 15 seconds of music, the system creates a digital signature for that snippet and then looks for a matching pattern in the database. Wang offers an analogy: "Imagine 306 million seconds of songs translated into constellations of dots and printed out in a long line. Next, imagine that the constellation of 15 seconds that you want to match is printed on a piece of transparent plastic. You slide the plastic along the whole line of dots until you find the same pattern."

Wang calculates that the probability of finding a false match is less than three in a million, which means that the system can work even when only 1 percent of the song is audible. To minimize the retrieval speed, all the data are kept in active memory (rather than on hard disks) on a distributed computer system consisting of about 70 PCs.

In practice, Shazam is easy and fun to use. I tried it out in cafés and bars, in busy shopping malls, in the backseats of automobiles while the radio was playing and in my own home. Most times, the service successfully recognized the music, returning the name of the song by text message within two or three minutes. Not surprisingly, I got the best results when the music was clearest—from the radio or CD player at home. But Shazam identified Blondie's "Heart of Glass" even when played on a taxicab's crackly radio and the Beatles's "Hey Jude" when broadcast over the public address system of a shop in London's West End. I found that the best way to use the service was to point the phone along an obstacle-free path toward the source of the sound, avoiding people engaged in loud conversations. (I got some funny looks from passersby as I earnestly held out my mobile phone in the middle of shopping malls.) Each 15-second call costs 50 pence, or about 75 cents, which is simply added to your cellphone bill.

Shazam's database focuses on newer music. But the system does not always work optimally in the noisy haunts of music lovers.

Shazam's database focuses on newer music. When I tried out a couple of recordings from the 1950s by blues singer Blind Willie McTell, the system was stumped. Shazam has not yet included any classical music in its database either, because the company believes that its main users will be young people seeking to identify popular songs. But the system does not always work optimally in the noisy haunts of music lovers. I got no response from Shazam when I attempted to get it to recognize Otis Redding's "Sitting on the Dock of the Bay" and Dusty Springfield's "I Only Want to Be with You" in a busy pub.

These failures were most likely the result of excessive background noise and distortions in the frequency of the songs. Background noise can be a problem because European mobile phones always focus on the loudest sounds and throw away the quieter signals. And when disc jockeys speed up or slow down songs, they alter the frequency profiles that Shazam relies on. Some tolerance is built into the system's search engine, though; Wang says that Shazam can recognize songs played within about 10 percent of their original speed.

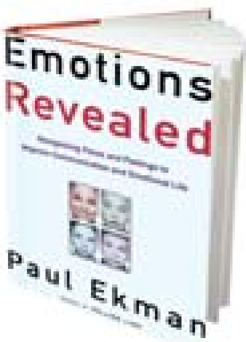
Despite the shortfalls, Shazam's technology is impressive. The company is planning to offer new services that will allow users to identify a musical piece simply by humming or maybe even whistling it into the handset. And Shazam is not alone. Philips, the Dutch electronics giant, is developing a prototype system designed to recognize hummed versions of songs. Philips could not say when such a product might be commercially available, however.

About 300,000 people in Britain have tried Shazam's service since it was introduced last August. The company is preparing to launch the service in Germany soon and to make it available in the U.S. and some Pacific Rim countries in the next 12 to 18 months. By then, Shazam's musical database should be even bigger—perhaps even big enough to recognize Blind Willie McTell. SA

Fiona Harvey is a journalist based in London who writes about all aspects of technology for the Financial Times.

A Polite Smile or the Real McCoy?

HOW FACIAL MUSCLES REVEAL—AND HIDE—EMOTION **BY CAROL TAVRIS**



EMOTIONS REVEALED: RECOGNIZING FACES AND FEELINGS TO IMPROVE COMMUNICATION AND EMOTIONAL LIFE
By Paul Ekman
Times Books, Henry Holt and Company, New York, 2003 ($25)

COURTESY OF HENRY HOLT AND COMPANY

Chelsea Thomas was born with Möbius syndrome, in which a nerve that transmits commands from the brain to the facial muscles is missing. As a result, for her first seven years Chelsea looked perpetually grumpy. Then surgeons transplanted nerves from Chelsea's leg to both sides of her mouth, and today Chelsea can do what most people in the world take for granted. She can smile.

Meanwhile, thousands of adults are botoxing the nerves that allow them to frown. Actors who do so cannot convey anger or fear, and some botoxed mothers complain that their children no longer take their admonitions seriously, accompanied as they are by the mothers' bland expressions.

Paul Ekman would not be surprised. He has been studying facial expression of emotions for some 30 years, in the noble tradition of Aristotle, who first observed the characteristic facial expressions of anger, fear "and all the other passions," and Charles Darwin, who added an evolutionary explanation. Darwin's theory of the universality of emotional expression was unpopular in the 1960s, when Ekman began his research. It was the era of the tabula rasa in social science; Ekman was to emotion what Harry Harlow was to love, swimming against the academic tides. As a graduate student at the time, I was in that tide up to my neck, and I remember how vehemently psychologists protested the idea that any aspect of human behavior might have a hardwired element. Facial expressions? Clearly cultural. Don't the Japanese coolly suppress any sign of emotion, and don't the Italians exuberantly reveal theirs?

Over the next decades, Ekman and his colleagues gathered evidence of the universality of seven facial expressions of emotion: anger, happiness, fear, surprise, disgust, sadness and contempt. In every culture they studied—in Japan, throughout Europe and the U.S., and among the nonliterate Fore of New Guinea—a large majority could recognize the basic emotional expressions portrayed by people in other cultures, and others could recognize theirs.

Yet, as Ekman also showed, cultures do differ widely in the "display rules" of emotional expression. Certain *emotions* are universal, hardwired into facial expressions and the brain; however, emotional *expressions* are culture-specific. People smile or display anger for many reasons, and they don't reveal these emotions when such displays would be considered rude or inappropriate.

Ekman and his collaborator Wallace Friesen created a coding system that identifies each of the nearly 80 muscles of the face, as well as the thousands of combinations of muscles associated with various emotions. (Ekman can do all of them himself.) When people try to hide their feelings or "put on" an emotion, Ekman found, they use different groups of muscles than they do for authentic feelings. For example, authentic smiles of joy involve the muscles surrounding the eyes; false or social smiles bypass the eyes completely.

In *Emotions Revealed*, Ekman, who is a professor of psychology at the University of California at San Francisco, beautifully interweaves his research with anecdotes, recommendations, and the behind-the-scenes flubs, accidental discoveries and debates that never make their way into published articles but that are the essence of scientific inquiry. He reviews what is known about the triggers, automatic and learned, that set off an emotion and how we might learn to manage or even get rid of them. He then

TRUE SMILE of enjoyment (*the one on the right, in these two photographs of author Paul Ekman*) involves the muscle that orbits the eye and is almost impossible to fake.

examines five emotions in detail: sadness, anger, fear, disgust and contempt, and the "enjoyable emotions." I was charmed to find *naches* on the list (the Yiddish word—it rhymes with "Loch Ness"—for the pleasure and pride that "parents feel when their child accomplishes something important"), along with "wonder," defined in terms of "its rarity and the feeling of being overwhelmed by something incomprehensible."

Because of Ekman's emphasis on the universality of emotions, especially those written on the face, readers will not learn much about the raging debate about emotions that do not necessarily have particular facial expressions, such as pride, envy, jealousy, compassion, and romantic or parental love (Ekman does not consider these to be "emotions," although other researchers do). Nor will readers learn much about the origins of emotion blends (such as *naches,* wonder, longing, the feeling of "bittersweet," and *schadenfreude*), which are more varied across cultures and individuals and which appear to be uniquely human, involving as they do higher cognitive processes.

Readers will enjoy seeing the many facial expressions of Ekman's favorite photographic subject, his daughter, Eve, who must have received ample compensation in fatherly *naches* for her ability to isolate and vary her facial muscles to reveal each basic emotion. These photographs serve brilliantly for scientific research, but whether they will help readers become better at accurately detecting another's emotion is doubtful. As research by others in this field has shown, when we read another's emotion, we do so through the filters and blinders of culture, the immediate situation, status, our own history, and degree of familiarity with the target. The face reveals, and the face lies. And as Ekman himself once observed, we wouldn't want it otherwise. SA

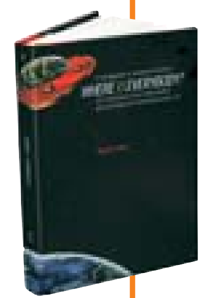
Carol Tavris, a social psychologist, is author of Anger: The Misunderstood Emotion *(Touchstone Books, 1989).*

THE EDITORS **RECOMMEND . . .**

. . . THREE BOOKS ON EXTRATERRESTRIAL LIFE

WHERE IS EVERYBODY?

by Stephen Webb. Copernicus Books, New York, 2002 ($27.50)



On the way to lunch at Los Alamos Scientific Laboratory one day in 1950, Enrico Fermi and three other physicists—Emil Konopinski, Edward Teller and Herbert York—chatted about flying saucers. At lunch, when the talk had turned to other matters, Fermi suddenly said, "Where *is* everybody?" His companions realized that the talk of flying saucers had turned his mind to the possibility that there is intelligent life elsewhere in the universe and that he was asking why, if there is, we have seen no sign of it. The question encapsulates what is now known as the Fermi paradox. Webb, lecturer in physics at the Open University in England, presents 49 solutions that have been proposed for the paradox, grouping them according to whether they hold that intelligent extraterrestrials are here, exist but have not communicated, or do not exist. He makes a splendid and enlightening story of it, concluding with his own solution, the 50th: "We are alone."

SETI 2020: A ROADMAP FOR THE SEARCH FOR EXTRATERRESTRIAL INTELLIGENCE

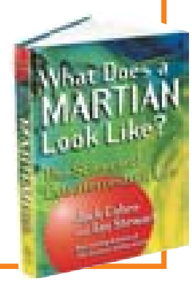
edited by Ronald D. Ekers, D. Kent Cullers, John Billingham and Louis K. Scheffer. SETI Press, Mountain View, Calif., 2002 ($25)

The point of view of the SETI project, of course, differs from Webb's. For more than 50 years, SETI researchers have been inspired by the following reasoning: "In a cosmos filled with billions of galaxies containing trillions of stars, is it possible that Earth, a world of inconsequential size and ordinary position, is alone in housing life that can discern the natural order? It is deeply incongruous to suppose that our enormous Universe is only sparsely occupied. To do so requires the belief that humans are exceedingly special. In view of astronomy's history, such a view is clearly suspect." This book reviews the history of the search and lays out a plan for SETI's next 18 years. The book presents sober science that holds out an awesome prospect should the search succeed: "We would come to view ourselves and our place in the Universe in a very different light. The changes would be as profound as those resulting from the revolutionary discoveries of Galileo and Darwin, which changed our understanding of our place in the solar system and of our biological evolutionary heritage."

WHAT DOES A MARTIAN LOOK LIKE? THE SCIENCE OF EXTRATERRESTRIAL LIFE

by Jack Cohen and Ian Stewart. John Wiley & Sons, Hoboken, N.J., 2002 ($27.95)

The authors of this book also see a possibility of extraterrestrial life. "We now know that there are many planets out there in the galaxy, and we have good grounds for supposing that a number of these will have life." It may be strange life, though, they say, nothing like what we know on Earth. On those principles, Cohen and Stewart (respectively, a reproductive biologist and a professor of mathematics at Warwick University in England) lay a basis for what they call xenoscience—knowledge of the strange. They draw on serious science—biology, chemistry, astronomy and physics—and also on science fiction, because the best of it has "made some useful contributions to the *scientific* understanding of possibilities for alien lifeforms."



All the books reviewed are available for purchase through www.sciam.com

Prime Squares

BY DENNIS E. SHASHA

The grid directly below is a prime square: the numbers in the rows (769, 953, 797) and columns (797, 659, 937) are all prime, none begins with zero, and no two row or column primes are the same. The grid is also "ambidextrous," because the numbers in the rows are still prime if read backward (from right to left). The square is not, however, "omnidextrous." To earn that label, its columns, too, would have to be prime when read backward (bottom to top), and its diagonals would have to be prime in both directions [*see squares at far right*]. In this case, 956 is even and therefore not prime.

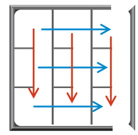
7	6	9
9	5	3
7	9	7

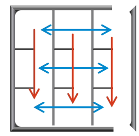
This puzzle asks you to create other prime squares. To warm up, build a prime 4-square (consisting of four-digit numbers) that uses nine distinct digits. This square need not be ambidextrous. (You can find a table of prime numbers at www.sciam.com/ontheweb.) See one solution at the right.

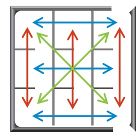
3	2	5	3
8	4	1	9
8	2	6	3
9	3	7	1

Now for the problems. Design an omnidextrous prime 3-square that uses as few distinct digits as possible. Then try to make a prime 5-square that uses all 10 digits; this one does not have to be omnidextrous or ambidextrous. Now it gets harder. Can you find the set of numbers n having prime n squares (whether plain, ambidextrous or omnidextrous)? I won't have an answer for that one; researchers are still working on it. But if you enjoy playing with primes, here's a new game to try: tic-tac-prime.

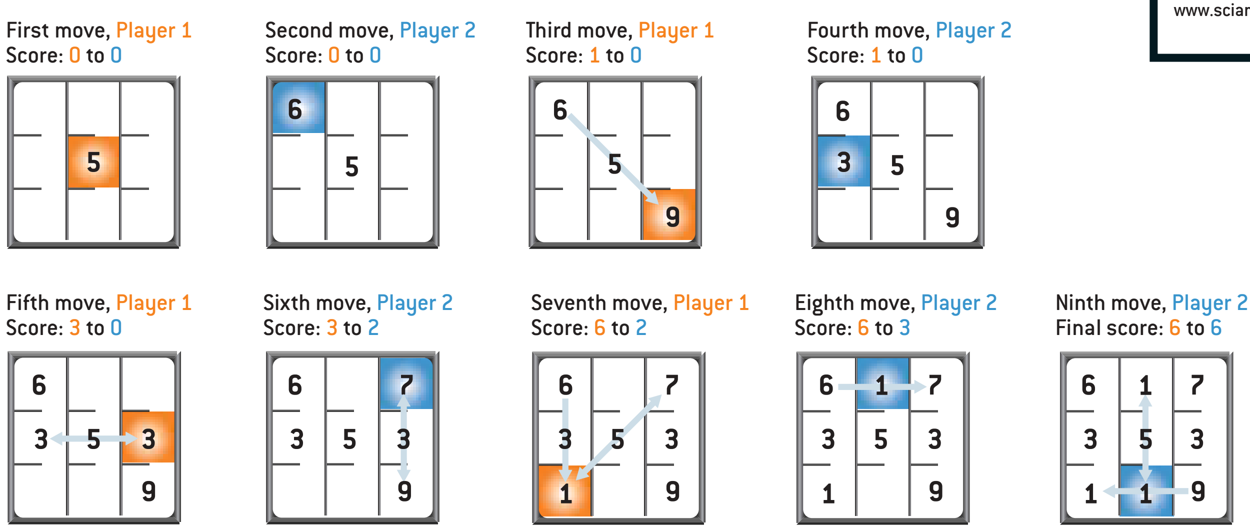
A move in tic-tac-prime consists of placing a single digit in an empty square of a three-by-three grid. The two players alternate moves except that the second player gets the ninth move. When one player's digit completes a segment of three digits in a line, that player gets a point for each three-digit prime (reading in any direction) that includes this last-placed digit (so two points for numbers that are prime backward and forward). Can you discover a winning strategy for either player? SA


PLAIN


AMBIDEXTROUS


OMNIDEXTROUS

A SAMPLE GAME OF TIC-TAC-PRIME



JOHNNY JOHNSON

Answer to Last Month's Puzzle

When the possible offers are $1 and $5, the best strategy is to take the first 50 offers no matter what and then take only $5 offers. If the ticket exchanger offers $1 for each ticket and then halts the trading, your regret ratio would be 1.8 ($90 divided by $50). If the exchanger initially offers $1 and then switches to $5, the regret ratio would still be 1.8 ($450 divided by $250). If the possible offers are $1 and $1 million, take the first 45 offers no matter what and then wait for $1-million offers. This strategy yields a regret ratio of 2.

Web Solution

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

Dropping By

SOMETIMES A NATURALIST CAN GET AHEAD BY ASSESSING WHAT'S BEEN LEFT BEHIND **BY STEVE MIRSKY**

Tennyson's well-known description of nature as "red in tooth and claw" conjures up dramatic moments in the prey-predator relationship. But nature's palette is rich, and Lord Alfred neglected to contemplate the grays, browns, blacks, greens and even blues to be found along the trailside. My interest in rectifying the poet's monochromatic view recently led me to the Loxahatchee National Wildlife Refuge in Boynton Beach, Fla., for an introductory session on the science of identifying scat.

If any of the 15 other attendees expected to be taught how to distinguish Ella Fitzgerald from Mel Tormé, they were sorely disappointed. There *was* singing, however. For the benefit of the six children in the room, workshop leader Serena Rinker warbled: "It begins with an 'S' and ends with a 'T,' it comes out of you and it comes out of me, I know what you're thinking, but don't say that, 'cause to be scientific we call it scat."

The ditty's silly but accurate: scat study truly is scientific. As author James Halfpenny notes in his book *A Field Guide to Mammal Tracking in North America,* "Scat, the feces of mammals, provides … an important window into the mammalian world. Information we gather from scat includes mammal identification and presence, location of activities, composition of diet, seasonal diet changes, and samples of prey species." Such species-by-feces analysis works for nonmammals, too: should I have an interest in, for example, the many large alligators at the Loxahatchee refuge, scat is a safe window through which to gaze.

Rinker wears the uniform of the U.S. Fish and Wildlife Service and describes herself as "only a part-time amateur scatologist." (Really, does anyone want to be a full-time professional scatologist?) She presented a brief primer on scat basics, replete with a table set with samples from

her collection, after which we attempted to identify the icky. Size and shape give obvious evidence, as big animals tend to do big things. Contents are likewise clear indicators of who's who by what's what: fish scales and crayfish parts suggest otter, says Halfpenny in his field guide. Cameras and cell phones might therefore suggest bear.

"Color also provides important clues," Rinker observed. A gator contribution was both large and gray, indicating a (possibly literally) heady mixture of hair and meat. Consuming grass can produce greenish scat; white scat may be old and dried-out or high in phosphates found in the bones of a prey animal that began its journey at the other end of the alimentary canal; brown indicates vegetable matter, even wood; blue means a critter passed the berries. Black scat has seen some fat, the result of a purely meat diet, although Rinker pointed out that a very dark sample on the table was in fact the result of coprophagy. I leave it to the more adventurous readers to look up the definition of coprophagy and then to wonder at nature's rich tapestry.

After the lecture, Rinker led us on a scat search along a boardwalk that snakes through a cypress swamp section of the refuge. I once encountered a playful otter wandering along the boardwalk, but on this day Rinker was delighted to find relatively fresh otter oeuvre, luminous with fish fins and crayfish shell shards. "Oh, my gosh!" she exclaimed. "I actually should come back and collect it, because this is one I don't have."

When the walk was over and the class dismissed, Rinker and I returned to the scene of the grime. She collected the scat and explained that the next step was to let it dry on newspaper, a plan that seemed quite sensible to me as a magazine writer. I then left the refuge, mulling over how a scientific outlook can make finding a pile of poop cause for celebration. And musing that the scatologist's motto must be "Good thing we didn't step in it." SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

Why do hangovers occur?

—P. Bouchard, Orange, Calif.

Sant P. Singh, a professor and chief of endocrinology, diabetes and metabolism at Chicago Medical School, offers this answer:

Several factors appear to be involved in getting a hangover—the unpleasant consequence visited on 75 percent of those who drink alcohol to intoxication. The effects include headache, nausea, vomiting, thirst, dryness of the mouth, tremors, dizziness, fatigue and muscle cramps. Often there is an accompanying slump in cognitive and visual-spatial skills.

A hangover has been suggested to be an early stage of alcohol withdrawal. Mild shakiness and sweats can occur; some people may even hallucinate. Acetaldehyde, a toxic breakdown product of alcohol metabolism, plays a role in producing symptoms. Chemicals known as congeners that are formed during alcohol processing and maturation also increase the likelihood and severity of a hangover; as a rule of thumb, the darker the liquor, the more congeners it contains. The toxins in congeners are distributed throughout the body as the liver breaks down the alcohol. Last, hangovers cause changes in the blood levels of various hormones, which are responsible for some symptoms. For example, alcohol inhibits antidiuretic hormone, which leads to excessive urination and dehydration. Blood aldosterone and renin levels also increase with a hangover—but unlike antidiuretic hormone, they do not correlate well with symptomatic severity, so their role is less clear.

MATT COLLINS

Individuals are more prone to develop a hangover if they drink alcohol rapidly, mix different types of drinks, and do not dilute the absorption of liquor by eating food or drinking nonalcoholic beverages. Sugar and fluids can help overcome the ensuing hypoglycemia and dehydration, and antacids can reduce nausea. To reduce headache, anti-inflammatory drugs should be used cautiously: aspirin may irritate the stomach, and the toxic effects of acetaminophen on the liver can be amplified by alcohol. Other drugs have been used to treat hangovers, but most have questionable value.

Why does shaking a can of coffee cause the larger grains to move to the surface?

—H. Kanchwala, Pune, India

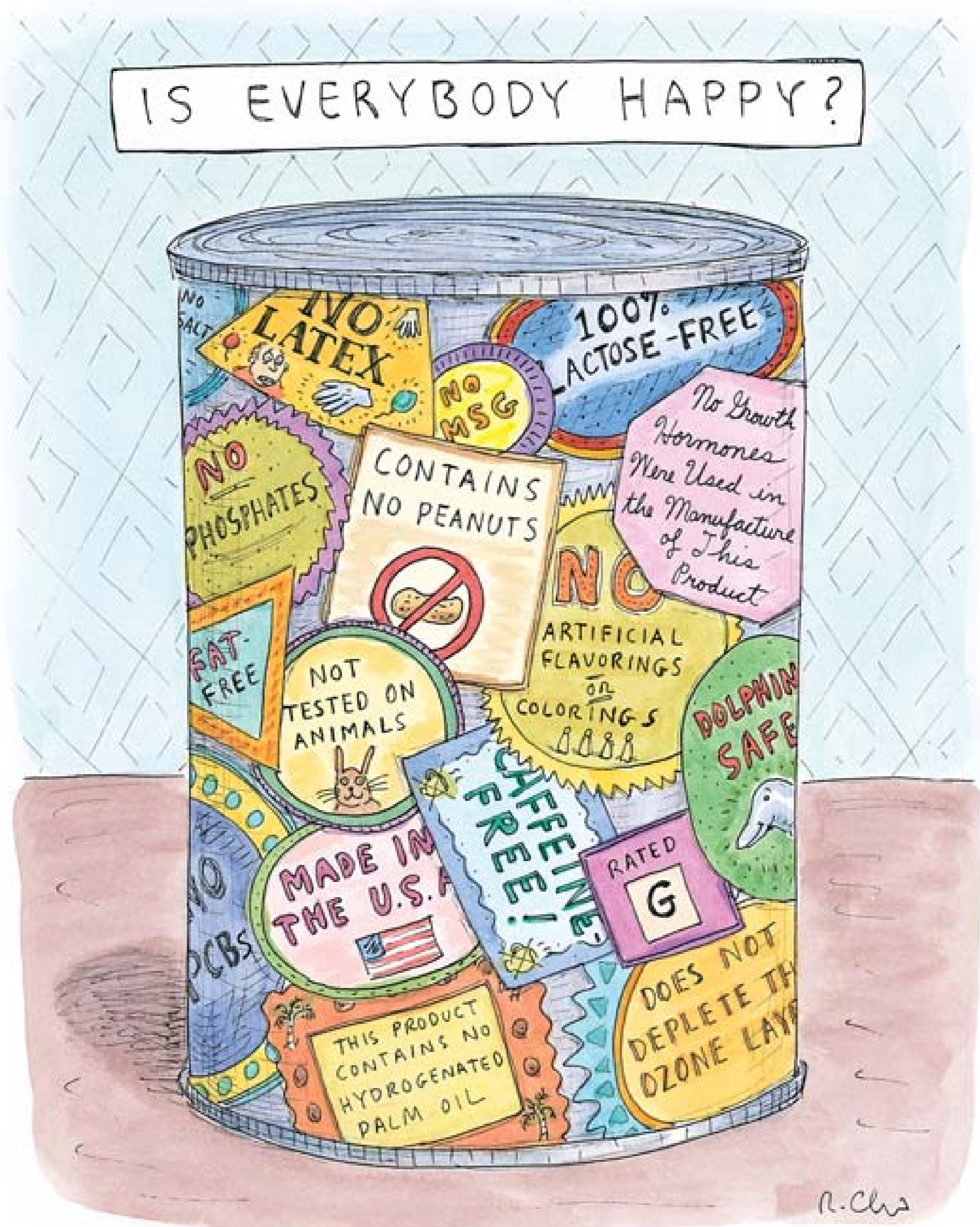
Heinrich M. Jaeger, a professor of physics at the University of Chicago, explains:

The phenomenon by which larger coffee grains move up and smaller ones travel down when shaking a can is called granular-size separation. It is often referred to as the Brazil nut effect, because the same thing will happen when you jiggle a can of mixed nuts. This occurs for two main reasons.

First, during a shaking cycle—as the material lifts off the bottom of the can and then collides with the base again—larger particles briefly separate from smaller ones, leaving gaps underneath. The tinier bits then slip into the gaps. When the shaking cycle finishes, the large particles cannot return to their original positions, and therefore the bigger particles slowly "ratchet" upward.

The second action at work is called a convective mechanism. When a can shakes, the coffee rubs against the sides. Friction causes a net downward motion of the grains along the walls, which is balanced by a net upward flow in the center—setting up a convection roll pattern. The downward flow is confined to a narrow region only a few (small) particle diameters in width. Once the large java grains reach the top, they move toward the side walls. If they are too large, they cannot fit into the region of downward flow and, after a few shakes, they aggregate near the top. Typically this mechanism dominates unless friction with the side walls is carefully minimized. SA

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


IS EVERYBODY HAPPY?
NO SALT
NO LATEX
100% LACTOSE-FREE
NO MSG
No Growth Hormones Were Used in the Manufacture of This Product
NO PHOSPHATES
CONTAINS NO PEANUTS
NO ARTIFICIAL FLAVORINGS or COLORINGS
FAT-FREE
NOT TESTED ON ANIMALS
DOLPHIN SAF
CAFFEINE-FREE!
MADE IN THE U.S.A.
RATED G
NO
CBs
DOES NOT DEPLETE TH OZONE LAY
THIS PRODUCT CONTAINS NO HYDROGENATED PALM OIL