SUDOKU SCIENCE • NEW ANTI-HIV DRUGS • SAVE BEACHES FROM GERMS

SCIENTIFIC AMERICAN

JUNE 2006 WWW.SCIAM.COM

Engineering Life

Genetic Circuits Will Revolutionize Medicine, Energy and Biotech

Toward Better **Pain Control**

The Secrets of **Supervolcanoes**

How to Prevent **Software Disasters**

june 2006

contents

features

SCIENTIFIC AMERICAN Volume 294 Number 6

44 Genetic circuitry

departments

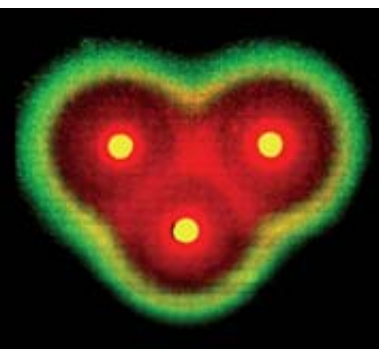
columns

Cover image by Kenn Brown; photograph at left by Kathleen Dooher.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40012504. Canadian BN No. 127387652RT; QST No. Q1015332537. Publication Mail Agreement #40012504. Return undeliverable mail to Scientific American, P.O. Box 819, Stn Main, Markham, ON L3P 8A2. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 248-7684. Send e-mail to sacust@sciam.com** Printed in U.S.A.

How to Kill Synthetic Biology

Thirty-three years after the invention of gene-splicing, the reality of biotechnology is still far short of what many once dreamed it would be, partly because the tools for manipulating genes have been crude. That is about to change. As nine scientists explain in "Engineering Life," beginning on page 44, new "bio fab" approaches to assembling complete genetic circuits promise to advance biotechnology in much the same way that the invention of integrated circuits transfigured electronics. They should enable workers to reengineer cells more ambitiously, to create organisms programmed at the genetic level to behave as desired. Colossal payoffs could accrue to medicine, agriculture, manufacturing, energy production and other fields. It is the birth of synthetic biology.



PROGRAMMED CELLS obey chemical commands.

COURTESY OF RON WEISS *Princeton University*

But like every newborn, synthetic biology is still intensely vulnerable. There are many ways to kill this young science; here are just two:

Underestimate safety concerns. Around the globe, people continue to worry that unnatural organisms containing recombinant DNA will become environmental headaches, if not pathogenic blights. For them, the news that scientists could soon genetically tinker more easily and more extensively is anything but good.

Fortunately, the research community is already undertaking activities, including the Synthetic Biology 2.0 meeting in May and an Alfred P. Sloan Foundation–funded study to be completed this summer, aimed at ensuring safety and instilling confidence. Following through on those recommendations and continuing the discussion will be imperative. But the burden of making synthetic biology secure does not fall just on the scientists; policymakers and opinion leaders need to endorse those plans, too, for the public to believe in the governance.

Patent indiscriminately. As the "Engineering Life" authors explain, underpinning the synthetic biology endeavor is a willingness for researchers to share their new genetic "devices" through a communal repository, the Registry of Standard Biological Parts. Investigators working on new systems could then simply look up the designs of any available needed components rather than reinventing everything from scratch.

That cooperative spirit is regrettably out of sync with biotechnology's modus operandi these days. As Gary Stix reviewed in "Owning the Stuff of Life," in the February issue, companies and universities have been on a spree of patenting not only whole genes but also genetic fragments of unknown utility. Some of those DNA patents seem overly broad or otherwise fail to meet the traditional standards for awarding intellectual property, and critics fret that their abusive exercise could hamper biomedical research. Stix concluded that few problems had emerged so far but that conflicts could intensify as more discoveries were put to work. To be clear: the problem is not patents as such but *bad* patents. Overly restrictive licensing and smotheringly broad patent interpretations could make a shambles of synthetic biology.

Half a century ago if recklessness, greed and unreasonable fear had somehow handicapped the development of integrated circuits, then the computing and communications revolutions would have been snuffed out. Now is an equally pivotal moment for the future of biotechnology.

THE EDITORS editors@sciam.com

How to Contact Us

EDITORIAL

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

For general inquiries:
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8200
fax: 212-755-1976
or
editors@sciam.com

SUBSCRIPTIONS

For new subscriptions, renewals, gifts, payments, and changes of address:
U.S. and Canada
800-333-1199
Outside North America
515-248-7684
or
www.sciam.com
or
Scientific American
Box 3187
Harlan, IA 51537

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

On the Web

WWW.SCIAM.COM

UPDATED EVERY WEEKDAY

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

TED DAESCHLER

Newfound Fossil Is Transitional between Fish and Landlubbers

Paleontologists working in the Canadian Arctic have discovered the fossilized remains of an animal that elucidates one of evolution's most dramatic transformations: that which produced land-going vertebrates from fish. Dubbed *Tiktaalik roseae*, the large predatory fish bears a number of features found in four-limbed creatures, a group known as tetrapods.

The Value of the Insect Labor Force

New research shows that bumblebees and other insects provide $57 billion in pollination services as well as other free labor in the U.S. alone.

The Scientific American Podcast

Tune in each week as host Steve Mirsky explores such topics as animal intelligence, avian flu, sustainable development, and combat stress through interviews with leading scientists and journalists.

Ask the Experts

How do space probes navigate large distances with such accuracy?

Jeremy Jones, chief of the navigation team for the Cassini Project at the NASA Jet Propulsion Laboratory, enlightens.

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

TOPICS RANGING FROM tabletop reactors to forensic investigations of the first president helped to pep up the humdrum month of February for readers. Two articles that caught letter writers' attention were about rethinking established technology: In "Plasma Accelerators," Chandrashekhar Joshi showed how small, inexpensive particle accelerators promise a wealth of data in lower-energy research. "Protecting New Orleans," by Mark Fischetti, assessed various flood-control system designs for the Mississippi Delta able to withstand major storms and rising sea levels.

In "Putting a Face on the First President," forensic anthropologist Jeffrey H. Schwartz intriguingly chronicled how he solved the mystery of what George Washington looked like as a young man. A number of readers noticed a curious resemblance. Harry Melkonian wrote via e-mail, "From what I can see from the digital reconstruction, the 19-year-old Washington looked like a young Jack Benny. Well! Now cut that out!"

CATCHING A PLASMA WAVE

In "Plasma Accelerators," Chandrashekhar Joshi creates the erroneous impression that in 1993 his group was the first to unambiguously demonstrate acceleration by plasma waves. In fact, in 1986 then University of Wisconsin–Madison graduate student Jamie Rosenzweig showed the acceleration of electrons by plasma waves using a facility at Argonne National Laboratory that was constructed by my group specifically for wakefield experiments. Those proof-of-concept experiments, whose results were published in 1988, were the subject of his Ph.D. thesis. Rosenzweig and his collaborators have continued to make important contributions to plasma-based accelerator physics and technology since then.

I should also point out that the equipment required to power a useful "tabletop" wakefield accelerator most likely would not fit on a table. It might require much more room: perhaps the whole kitchen, dining room and a good-size backyard combined.

James Simpson
Fountain Hills, Ariz.

In a laser wakefield accelerator, could one fire the laser several times in sequence to create many wave fields and accelerate the target particles faster with each wave? Wouldn't the first wave accelerate them a little and then the next wave a little more until the particles are in sync (at the same speed as the wave)? Just a thought.

Paul Turner
Allen, Tex.

JOSHI REPLIES: *Simpson is correct that the Argonne National Laboratory researchers carried out a plasma wakefield experiment. That experiment showed acceleration gradients of 1.6 MeV (million electron volts) per meter. The history of electron acceleration by plasma waves, of course, begins even earlier. In 1981 my group demonstrated gains of 1.4-MeV energy in less than one millimeter by a process known today as self-modulated laser wakefield acceleration.*

Turner's suggestion amounts to "staging" many laser accelerators to gain very high energy particles. Demonstrating such staging is one of the principal goals of the laser-accelerator community.

ENDING PAP FEARS

Thank you for highlighting in "To Banish a Cancer" [SA Perspectives] the achievements in the fight against human papillomavirus (HPV), which causes all the cervical cancers. I must point out, however, a few facts that your editorial overlooked, because understanding the virus's biology removes the sexual connotation.

HPV can live only in the squamous skin layer, which is about 400 nanometers deep. No other sexually transmitted infection (STI) lives in the skin. HPV is rapidly transmitted from squamous skin to squamous skin cells but is *not* transmitted by bodily fluids, as are all other STIs. Most important is that penetrative intercourse, though the most efficient method for transmission, is not necessary for this rapid transfer from skin to skin. Fingernails and the skin around them contain the same types of cervical cancer–causing HPV as the anogenital skin does, allowing autoinoculation from simple habits such as toilet hygiene or tampon insertion. This virus is ancient and omnipresent in all environments, and 80 percent of women (and probably more men, but this has not been well documented yet) develop at least one HPV infection in their lifetime.

Thus, all references, surveys and propagation of "educational" materials that assert cervical cancer is transmitted solely through sexual contact are deceiving and do the public a disservice. The potential for improved health with the new vaccine in both men and women is tremendous—and it is the biggest health care advance in 50 years for women. Therefore, it is important to emphasize the truth.

Diane M. Harper
Director, Gynecologic Cancer Prevention Research Group
Dartmouth Medical School

SYNAPTIC SMARTS

In "Intrigue at the Immune Synapse," Daniel M. Davis shows that much of the information needed for immune cell communication is stored or transmitted by the spatial arrangement of the proteins and other molecular units of the cells. This seems to be a very efficient system for encoding data at a high density. It would suggest that computer architectures should be developed that also use spatial arrangement among interacting units to encode more information than just the bit that each unit alone can encode. Perhaps this is one of the reasons why organic brains are more compact and efficient than the best computers.

Will Kastens
Madang, Papua New Guinea

GOING DUTCH

When I saw the pictures of the broken levees in New Orleans, as an immigrant from the Netherlands, I could not believe my eyes: How could you have hoped to hold back water with those little walls of cement? I wondered if Katrina's devastation and media exposure would have one positive effect: to make government officials look for a better solution. Glancing at the comprehensive plan in "Protecting New Orleans," by Mark Fischetti, I was disappointed. How can a design that seems like it was devised by 19th-century planners be considered? If Dutch engineers had designed the proposed line of more substantial levees, the lines would be much less zigzaggy. Building delta works with slightly curved lines, which would result in levees half the length of those in the plan, would provide enormous cost savings. It would also avoid corners, which create weak spots along the defense line.

BE-FOREFATHER: Digital forensic reconstruction of George Washington at age 19 might also recall comedian Jack Benny at "age 39."

The U.S. is an enormous nation where one can expect to find the best professionals in any field. Although New Orleans is neither Venice nor London, for the sake of the people the plan must protect, you should seek advice from those who know how to protect their own lands. How about running a contest for all the engineers in the world to tap their knowledge about what to do in Louisiana and then picking the plan that is best for its price?

Fredrick Schermer
San Francisco

THE GOOD OLD HAZE

The article excerpt from 1856 in "50, 100 & 150 Years Ago" caught my eye: "The fuel required to cook dinner in Paris costs nearly as much as the dinner itself." For the price of cooking fuel to reach equivalent levels today in the U.S., natural gas would have to be about $944 per MCF (1,000 cubic feet). Compare that with the U.S. Energy Information Agency's estimate of retail price of natural gas of about $16 per MCF for the winter of 2005–2006. So an American family today is paying less than 2 percent of what a Parisian family was paying in 1856 for cooking fuel. (Did I mention what all that burning of wood and coal in the homes of Parisians meant for their indoor and outdoor air quality?) While we have all felt the bite of recent increases in the price of fuel, the 1856 article demonstrates we may still be better off than in times past.

Branko Terzic
Energy & Resources Department
Deloitte Services,
McLean, Va.

ERRATUM Ask the Experts stated that adding salt to ice creates a hydration process that gives off heat. Rather it absorbs heat. For a complete explanation of how salt melts ice, see www.sciam.com/ontheweb

TOM DRAPER DESIGN

50, 100 & 150 Years Ago

FROM *SCIENTIFIC AMERICAN*

Lysenko Retreats ▪ Consumers Benefit ▪ Neptune's Realm Photographed

JUNE 1956

THE ANTIPROTON—"Since it was apparent that creation of the antiproton required tremendous energy, the most likely place to look for it was in cosmic rays. On a few occasions investigators found events which seemed to signal the generation of an antiproton, but there was never sufficient information to identify it with certainty. When the Bevatron at the University of California began to bombard a target made of copper with six-Bev (billion electron volts) protons, the next problem was to detect and identify any antiprotons created. A plan for the search was devised by Owen Chamberlain, Thomas Ypsilantis and the authors of this article. Tracks of about 20 antiprotons have now been detected in emulsions by observers in Berkeley. —Emilio Segrè and Clyde E. Wiegand" [*Editors' note: Emilio Segrè won a Nobel Prize in Physics in 1959.*]

POLITICALLY CORRECT CORN—"Trofim Lysenko's personal rule over U.S.S.R. genetics came to an end in April with his resignation as head of the All-Union Academy of Agricultural Sciences. The event is interpreted to mean that after 15 years during which 'Mendelian' genetics was a politically discredited branch of science in the Soviet Union, it is once again acceptable. Hybrid corn, which is incompatible with Lysenko's theories of acquired characteristics, is now being planted on Soviet farms. The week before Lysenko retired from prominence there was an announcement from Moscow that the scientific works of Nikolai Vavilov were about to be published. Vavilov, a world-famous geneticist, opposed Lysenko's rise, and in 1942 died in exile in Siberia."

JUNE 1906

PURE FOOD AGENDA—"The disclosures of the unsanitary methods in use at Western packing houses, recently brought to the attention of the President [Theodore Roosevelt] and now sought to be controlled and remedied by special governmental inspection, as provided in the Beveridge amendment to the Agricultural Appropriation bill, will bring about a wholesome check on the preparation of improper products for human consumption. The fact that the proposed law covers the preparation of only a small part of our foods seems to have been wholly lost sight of. What is true of such meat foods as are derived from cattle, sheep, goats, and swine (chiefly canned and prepared meats or lard) is equally true of a great range of other foods, such as fowls, game, fish, and eggs." [*Editors' note: This effort was greatly influenced by Upton Sinclair's book* The Jungle, *published in 1906.*]

MOTOR CAR NO. 7, Union Pacific Railroad, 1906

ENGINE NO. 7—"The accompanying illustration is of the latest of the Union Pacific gasoline motor cars, constructed for suburban passenger traffic and inspection service on the lines of that road. Number 7 was built in the Omaha, Neb., shops of the Union Pacific Railroad, and has recently undergone a series of successful tests between that city and Grand Island. Among the conspicuous features of the design are the round, porthole-like windows, the sharp forward end tapered to a knife-like edge, and the convenient side entrances. The engine is a six-cylinder, Standard gasoline motor located over the forward trucks."

JUNE 1856

FIRST UNDERWATER PHOTOGRAPH—"In the *Journal of the Society of Arts,* W. Thompson, of Weymouth, England, gives an account of the means he adopted for taking a photograph of the bottom of the sea, in Weymouth Bay, at a depth of three fathoms. It appears that the camera was placed in a box, with a plate-glass front, and a movable shutter to be drawn up when the camera was sunk to the bottom. When at the bottom the plate was exposed for about ten minutes. The box was then drawn into the boat, and the image developed in the usual manner. A view was thus taken of the rocks and weeds lying at the bottom of the bay. Mr. Thompson anticipates that it will be a ready and inexpensive means of arriving at a knowledge of the condition of piers, bridges, piles, structures, and rocks under water."

TEXAS CAMELS—"The Arabian camels for which a government expedition was sent out to the Mediterranean arrived in the Bay of Matagorda, Texas. They are to be employed as transports on the western deserts of our country."

POLICY

The End of the Everglades?

SUPREME COURT CASE JEOPARDIZES 90 PERCENT OF U.S. WETLANDS **BY SARA BEARDSLEY**

On February 21, his first day on the job, Justice Samuel Alito settled into one of the nine high-backed chairs at the Supreme Court to hear *Rapanos v. United States* and *Carabell v. the U.S. Army Corps of Engineers*—a pair of cases that, though not as well publicized as Jose Padilla's anti-government petition nor as high profile as federal wiretapping, will probably eclipse their importance. Bundled together, the cases ask the justices (and Alito in particular, a projected swing vote) to declare whether national agencies can patrol the soggy patches of earth between dry, developable land and federally protected wetlands. A decision in favor of the plaintiffs could put most of the nation's watery habitats at risk.

AT RISK: U.S. wetlands, such as this one in Medfield, Mass., could lose federal protection if the Supreme Court rules in favor of plaintiffs who say that the Clean Water Act goes too far.

EMILY HARRISON

"The tension," summarizes Margaret Strand, a wetlands lawyer in Washington, D.C., "is that protecting biodiversity translates into regulating private property"—pitting two American precepts, federalism and environmentalism, against each other. That tension originated with the 1972 Clean Water Act, which tasked the Environmental Protection Agency with preserving all "navigable waters," later defined as "waters of the U.S.," against unpermitted discharges. At the time, wetlands protection was just an appendage to the broader water safety effort, but recognition of the biome's ecological value inspired courts to uphold broad-based government power over the next 20 years. They could do so because "almost anything a little bit squishy might count as waters of the U.S.," says Don Carr, another lawyer in Washington.

Enter John Rapanos and June Carabell, two Michigan landowners who were denied development permits despite a 20-mile distance from the nearest navigable water and a berm blocking drainage, respectively. Over decades of court battles, the petitioners have

marshaled a crusade against EPA jurisdiction they claim goes too far. One brief on their behalf traces government power to "remote desert washes hundreds of miles from the nearest navigable waters"—far beyond what they think Congress intended with the act. Instead *Rapanos* lobbies for federal jurisdiction limited to "navigable-in-fact" waters (commercial channels) and their adjacent wetlands. But such an interpretation would leave 90 percent of protected wetlands up for grabs, according to Richard Lazarus, a Georgetown University law professor.

States are unlikely to replace federal enforcement because only half have wetlands programs, explains Jon Kusler, president of the Association for State Wetland Managers. And "there are pressures in local communities for land development" that keep district guidelines from being as rigorous, Strand adds. Carabell's experience is a good example: whereas Michigan authorized a permit for condominiums on the land in question, the federal government asserted jurisdiction and denied one.

Relying on state laws becomes an even bigger problem with downstream pollution: discharge between differently regulated states presages a domino effect on the entire watershed. "All [polluters] have to do is [dump] far enough upstream," reasoned Justice David Souter during oral arguments; "it will eventually get in the navigable water."

The petitioners, however, claim that their lands are just too detached to cause that kind of trouble. And a 2001 Supreme Court decision offers precedent for their case, finding that "hydrologically isolated" wetlands belong under state jurisdiction. Federalists such as Robert Pierce, a former Army Corps of Engineers official who once enforced the EPA's laws, believe that the government's permit denials are often off target. According to Pierce, states may be able to take up the slack, but perpetuating government programs is "a major economic drain" and a "waste" of money that could be used elsewhere.

Yet with natural wetlands in constant decline, government inefficiencies may be a price worth paying. "You can't protect the nation's biological integrity without protecting these waters," notes Sandra Postel, director of the Global Water Policy Project. And one only has to recall Hurricane Katrina—gathering strength where wetlands once buffered the levee system—to measure the costs to society.

Nevertheless, federalist watchdogs cling to *Rapanos* (which may be decided this month) as an opportunity to curb Washington's power. "This case isn't about losing wetlands or saving wetlands," says Pierce, who has conceded that a government rollback might cause environmental harm. "This is about putting the federal government where the Constitution says it should be. In my mind, that's more important."

news SCAN

FOR WETLANDS, BAD ACCOUNTING

Although wetlands are "the richest areas on earth in terms of biodiversity," according to Jon Kusler of the Association for State Wetland Managers, about half have been lost over the past century. Land developers are supposed to compensate by creating wetlands elsewhere, but these replacements are often far from the original site, both breaking up the natural ecosystem and sacrificing original habitat conditions.

Many of the seeming advances in wetlands protection over the past few years—such as a 2006 Fish & Wildlife Services report trumpeting "no net loss" of wetlands for the first time—have relied on these nascent wetlands for their accounting. In fact, 12 percent of the freshwater wetlands increase noted in that report (a margin that tipped the balance of gain and loss) came from new freshwater ponds that "would not be expected to provide the same range of wetland values," according to its authors. As Kusler puts it: "Digging a farm pond is not restoring a wetland."

DEFENSE

Disruptive Threats

THE PENTAGON TRIES TO KNOW THE UNKNOWN **BY DANIEL G. DUPONT**

In the seventh century, during the Arab siege of Constantinople, the Byzantines introduced "Greek fire." Consisting of a thick, flaming substance launched from cannons, it proved devastating during naval battles because the fire could not be extinguished and was said to burn Arab ships even below the water line.

In today's Pentagon parlance, that weapon would be known as a "disruptive threat"—something that comes out of left field to tilt the balance of power. Ryan Henry, the principal deputy undersecretary of defense for policy, says that this kind of capability "puts at risk military power as a key instrument of the broader U.S. national power."

Henry was a key architect of the Pentagon's Quadrennial Defense Review, completed in February. In it, defense officials analyzed the military and categorized threats into four groups: traditional, large militaries such as China's; "irregular" forces such as those fighting U.S. troops in Iraq; catastrophic weapons; and disruptive threats.

ORBITAL FEAR: Sputnik 1 upset the balance of power.

Past disruptive actions include the Soviet detonation of an atomic bomb in 1949 (earlier than expected) and the launch of Sputnik 1. But today, what could offer an adversary the shock-and-awe equivalent of Greek fire? Henry offers one scenario: the ability to "counter ballistic effects"—to shoot down U.S. missiles and other projectiles.

Beyond broad outlines, though, officials cannot offer much that is concrete. In its 2006 National Security Strategy report, the White House said that such threats could take the form of "biotechnology, cyber and space operations, or directed energy weapons" used in new ways. Other policy statements emphasize "unanticipated" challenges, noting the difficulties of planning for the unknown.

Some experts argue that the military authorities do not have a good, systematic way to recognize disruptive threats. The Defense Science Board, an influential group of outside advisers to the Pentagon, said in a March report that, although the military has undertaken "extensive activity in the area of disruptive challenges," it lacks "a comprehensive, coherent effort to identify and address these challenges." Instead the science board said that "much of the current effort addresses excursions to traditional challenges or approaches that the U.S. can use to disrupt adversary operations rather than the reverse."

The science board and others note that recent operations in Iraq and the larger war on terrorism should give the U.S. pause in trying to anticipate the unknown in the information age. According to an internal report commissioned last year by the Office of Force Transformation, a Pentagon think tank, today's battlefields "extend into cyberspace" and involve ubiquitous modern products such as cell phones, the Internet and garage-door openers. Such technology means that disruptive threats could be deployed quickly. As the analysis warns: "We no longer have the luxury of years of warning" as in a previous era.

Henry acknowledges that the quadrennial review showed that the Pentagon's bureaucracy was too rooted in the industrial age; its modernization, he believes, will enable the armed forces to better predict and prevent unexpected challenges. He also adds that U.S. intelligence agencies are "scanning the science and technology horizon" for future disruptive threats.

Perhaps the most effective answer is to push the technological envelope. Adversaries might then be too busy playing catch-up. According to the Office of Force Transformation report, "the new solutions to disruptive challenges lie less in trying to detect and catch them … and more in making the U.S. military a more elusive target by increasing the rate at which it transforms" itself. As Ryan notes, during the cold war the U.S. enjoyed decades of dominance in one crucial area: stealthy aircraft, its own disruptive capability. "If the other side would have had it, and we didn't," he remarks, "we don't know what the outcome would have been, but we would have expended a significant amount of blood and treasure trying to defeat it."

Daniel G. Dupont edits InsideDefense.com, an online news service.

TO RUSSIA, WITH SUSPICION

So which countries are likeliest to develop weapons that can surprise and shake up the U.S. military? The Pentagon makes no secret of its concern over North Korea, Iran and, to a lesser extent, China. But in the Quadrennial Defense Review, only one country is linked directly to disruptive threats: Russia. "Internationally," the report states, "the United States welcomes Russia as a constructive partner but views with increasing concern its sales of disruptive weapons technologies abroad." The Pentagon will not publicly say more about Russia and disruptive threats.

DMITRI KESSEL, © TIME INC./TIME LIFE PICTURES/GETTY IMAGES

GENETICS

Trace Elements

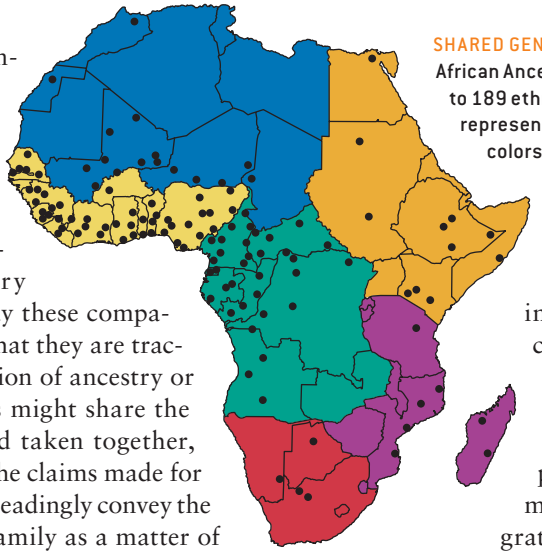
RECONNECTING AFRICAN-AMERICANS TO AN ANCESTRAL PAST **BY SALLY LEHRMAN**

Even as population geneticists battle over the meaning of race, cline and "biogeographical ancestry," a small industry has emerged out of the quest to understand human migration and identity. At least 11 companies offer individuals the ability to trace their African, Native American, Asian or European roots through DNA markers. One, Oxford Ancestors, will even help men learn if they carry the "heroic Y chromosome that flowed through the veins of the High Kings of Ireland" or inherited it

RACE, GENES AND MEDICINE

Rick Kittles of Ohio State University has used statistical estimates of biogeographic ancestry to show the shortcomings of hunting for disease susceptibility genes assumed to be shared by all African-Americans. On a genetic level, he finds, "there's no definitive line between black and white." One gene variant believed to be associated with prostate cancer risk in African-Americans, *CYP3A4-V,* was no more frequent in that group when he controlled for patients' ancestries. "Race is a bad proxy for shared ancestry and environment," Kittles concludes.

Consumers and physicians seem to agree, judging by the market failure of the first race-based medicine, BiDil, used for heart failure. Its maker, NitroMed in Lexington, Mass., has had trouble with insurance reimbursement and has been criticized for testing its drug only in black patients for no purpose other than patent technicalities. (Many cardiologists think the drug works just as well in white people.) Sales never took off—perhaps more for social reasons than scientific ones. In March, NitroMed's chief executive and chief financial officer called it quits.



SHARED GENES: A firm called African Ancestry can match DNA to 189 ethnic groups. The dots represent sampling areas; colors indicate broad geographic regions.

"directly from Genghis Khan."

Hyperbolic promises aside, many population geneticists urge caution in interpreting these ancestry tests. Detractors say these companies rarely clarify that they are tracing just a tiny fraction of ancestry or that many lineages might share the same markers. And taken together, critics emphasize, the claims made for this technology misleadingly convey the sense of race and family as a matter of biological precision.

Rick Kittles, a geneticist at Ohio State University, acknowledges these pitfalls but has pressed forward with his firm, African Ancestry, because of what he considers a critical social need. The Washington, D.C.–based company compares a client's DNA with that of ethnic groups now living mainly in West and Central Africa, aiming to recover the family ties destroyed by the slave trade. Like the other firms, its technicians examine either a section of mitochondrial DNA or markers on the Y chromosome and then try to match them up with databases of maternal and paternal lineages.

With 13,690 maternal and 11,747 paternal lineages, African Ancestry offers unusually rich detail. Even so, Kittles acknowledges that each of the tests follows only a single ancestral line out of as many as 1,024 over 250 to 300 years (10 generations). The analysis cannot detect all the historical groups that may have contributed to a person's ancestry or even trace a single line beyond the populations Kittles has sampled so far.

But unlike other firms, Kittles is careful to say he is helping his clients connect to modern Africans, not historical figures or tribes, and he is not trying to categorize ethnicity or race. Instead of boiling identity down to genetics, he aims to highlight the interweaving of biology, history and culture. One concrete link to a part of Africa can be deeply healing, he says. Otherwise, most African-Americans can follow their history back only to enslavement. "It creates this void in the psyche of African-Americans, this missing piece of their identity," Kittles explains.

African Ancestry is also adding 200 lineages every few months, broadening out geographically from the original population samples. Kittles hopes that this expanding database may illuminate migration patterns from east and south African regions into the west and central areas, which had been the heart of the transatlantic slave trade. By tapping into the continent's high genetic diversity, this approach might overcome the difficulty of interpreting DNA markers shared across populations on the continent.

Besides technical improvements, with anthropologist Mark Shriver, Kittles has called for a code of conduct that would require companies offering personalized genetic histories to explain both the promises and limitations of their science. "He's careful and responsive to the ethical concerns," says Duana Fullwiley, a medical anthropologist at Harvard University. Still, Fullwiley and others remain cautious. For one thing, they see Kittles's desire to bridge the injury of slavery as overly idealistic. Ancestry tracing "may heal certain wounds," she acknowledges. But "it doesn't give us lost history back." Moreover, rather than resolving questions about identity, the test may open up new questions about genetic pedigree.

With his extensive database, Kittles has pinpointed African-American ancestry in more detail than other scientists thought possible, according to Joanna Mountain, a Stanford University population geneticist. Africa "is not just one homogeneous pot of people," Mountain says. Furthermore, individual African-Americans, who have a very mixed ancestry in their genomes, may be more similar to one another than to any group within Africa. "It's a neat story that's arising," she observes.

Sally Lehrman wrote this month's Insights.

MELISSA THOMAS

PSYCHOLOGY

No Prayer Prescription

SEND GOOD VIBRATIONS, BUT KEEP IT TO YOURSELF **BY CHRISTINE SOARES**

Seeking to assess the effect of third-party prayer on patient outcomes, investigators found no evidence for divine intervention. They did, however, detect a possible proof for the power of negative thinking.

PRAYER showed no therapeutic effect in a large trial.

JORGE SAENZ *AP Photo*

The three-year Study of the Therapeutic Effects of Intercessory Prayer (STEP), published in the April 4 *American Heart Journal,* was the largest-ever attempt to apply scientific methods to measure the influence of prayer on the well-being of another. It examined 1,800 patients undergoing heart-bypass surgery. On the eve of the operations, church groups began two weeks of praying for one set of patients. Each recipient had a praying contingent of about 70, none of whom knew the patient personally. The study found no differences in survival or complication rates compared with those who did not receive prayers. The only statistically significant blip appeared in a subgroup of patients who were prayed for and knew it. They experienced a higher rate of postsurgical heart arrhythmias (59 versus 52 percent of unaware subjects).

The research team—a psychologist, clergy and doctors from six institutions, including Harvard Medical School and the Mayo Clinic—speculated that nerves might have been to blame. "We know that high levels of adrenaline from the anxiety response can make fibrillation worse," said Charles Bethea, a physician at Integris Baptist Heart Hospital, a study site in Oklahoma City, in an April press conference. "The patient might think, 'Am I so sick that they have to call in the prayer team?'" Dean Marek, chief chaplain at the Mayo Clinic, saw the problem as a possible flaw in the study design: "The sense of community was not there. You could call it impersonal prayer rather than intercessory prayer."

Stopping short of suggesting that the healing power of prayers by friends and family might reside in the personal connections rather than in the prayers, the authors stated that they have no plans for a follow-up study. This one, sponsored largely by the John Templeton Foundation, cost $2.4 million.

PHYSICS

Hard Landscape

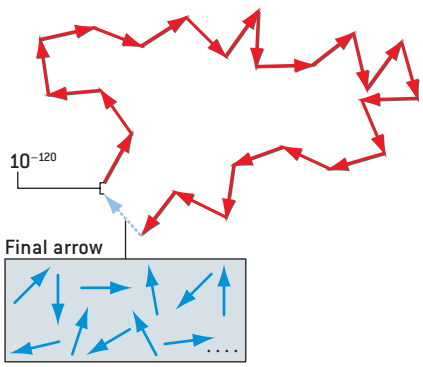
FINDING OUR UNIVERSE IN STRING THEORY APPEARS IMPOSSIBLE **BY JR MINKEL**

Cosmology was always going to be trouble for particle physics. Traditional quantum theory predicts that the vacuum of space should bubble over with short-lived "virtual" particles, whose combined energy, represented by the so-called cosmological constant, would have long ago blown galaxies far and wide like grease on water at the touch of detergent. Lately, trying to make sense of why the cosmological constant is tiny, physicists have toyed with a concept based on string theory. Called the landscape model, it supposes that many universes with different cosmological constants are realized in a larger multiverse. Despite long-standing hopes to the contrary, landscapers now have found that singling out a universe from this array is mathematically nigh impossible.

In 1998 astronomers discovered that the universe's expansion is accelerating at a rate consistent with a cosmological constant

10^{-120} times the value predicted by quantum theory. String theory, which unites gravity with quantum mechanics, offered the hope of explaining the attenuated cosmological constant. It recasts particles as one-dimensional strings, or filaments of energy, which play around in tiny tangles of extra spatial dimensions. The shape of the tangles influences the properties of strings and therefore the vacuum's energy. But no mathematical principle forces the extra dimensions to fold in a unique way.

One approach is to assume that many universes exist and that we necessarily find ourselves in one where the constant is hospitable to life. In 2000 Joseph Polchinski of the University of California, Santa Barbara, and Raphael Bousso, now at U.C. Berkeley, proposed that in the string picture, magnetic field–like quantities called fluxes, when coiled in various combinations inside the extra dimensions, could generate enough different cosmological constants to include the one we observe. This set of vacuum energies (or vacua, as physicists like to refer to universes) has been dubbed the anthropic landscape of string theory.



LANDSCAPE MODEL ANALOGY connects up inch-long arrows to create a loop with a gap of 10^{-120} inch at the end. The final arrow, picked at random from an infinite collection, is unlikely to close the loop properly.

In principle, physicists might like to examine one or a few vacua having the right cosmological constant to see if they offer testable predictions. The huge number of choices—at least 10^{120} vacua—seemed to demand vast computational power, perhaps even a quantum computer.

Unfortunately, the latest analysis suggests that no advanced machine or clever programming will cut much into the search time, according to Michael R. Douglas of Rutgers University and Frederik Denef, now at Catholic University in Louvain, Belgium. They have discovered that the task falls into a notorious category of problems called NP-complete. An example is the "subset sum" problem, which asks, given some group of integers (1, –17, 6,435, and so on), find a subset that adds up to 0. In these problems, the time needed to check the answer grows relatively slowly with the size of the subset, whereas the time typically needed to identify that subset grows exponentially—the twin hallmarks of NP-completeness.

Analogously, the problem in the landscape model is to add the energies of fluxes up to the desired value for the whole vacuum, 10^{-120}. The individual flux energies are represented by arrows, or vectors, all one inch in length and pointing in random directions. The game is to find a set of arrows that, when linked up end to end, leaves a gap between head and tail (the two ends) of 10^{-120} inch. A group of vectors very near to the desired result is enormously likely to be thrown off by the next randomly pointing vector. "The old ambition of exactly finding the precise vacuum in which we live may be intractable," Denef sums up. "The Douglas-Denef paper is surely a problem for drawing conclusions about what the landscape predicts," asserts Thomas Banks of U.C. Santa Cruz.

But perhaps scrutinizing all those vacua is unnecessary; Bousso notes that the vacua configurations may be immaterial—like atoms in a lump of iron. "It wouldn't really occur to you to attempt to figure out precisely where every atom is," he says. In that case, statistics derived from many vacua, rather than trying to make predictions from a single one, might be more useful. So for landscapers, getting around the intractable may all depend on asking the right questions.

JR Minkel is a frequent contributor.

LUCY READING-IKKANDA

SPACE SCIENCE

NASA's Reverse Thrust

CUTS MAY BE MORE THAN A MERE MONETARY SPEED BUMP **BY GEORGE MUSSER**

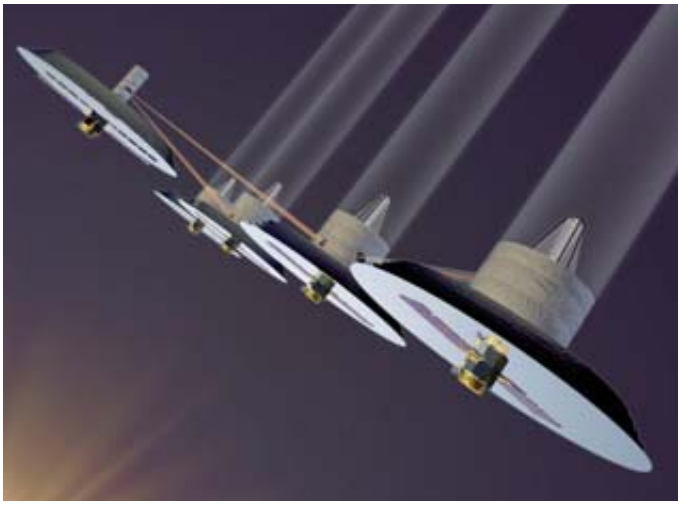
When President George W. Bush unveiled his plan for a new moon shot two years ago, a lot of people worried that it was long on rhetoric and short on cash—ultimately forcing NASA to raid its science budget to pay for it. On close examination, though, the trajectory seemed reasonable. The money freed up by phasing out the space shuttle and the International Space Station was not an implausible amount to build a postshuttle spacecraft (known as the Crew Exploration Vehicle, or CEV) and send it moonward by 2020. A "go as you can pay" strategy would extend the deadlines if money got tight, rather than pickpocketing other programs. A modest dollop of extra funds would help cover the transitional costs. NASA administrator Michael Griffin said at a press conference last September: "In our forward planning, we do not take one thin dime out of the science program in order to execute this architecture."

Now it looks like the skeptics were right. The NASA budget announced in February mows down a scarily long list of science missions, from a Europa orbiter to a space-based gravitational-wave observatory. Research grants to individual scientists, traditionally kept safe from high-level budget machinations, have taken a 15 percent hit, retroactive to last fall; hundreds have already received "termination letters" canceling their projects. Griffin went before Congress in February as the bearer of bad news: "Fulfilling our commitments on the International Space Station and bringing the Crew Exploration Vehicle online in a timely manner, not later than 2014 and possibly much sooner, is a higher priority than these science missions during this period."

The countdown to the crisis actually began a year ago, when the Bush administration lopped off the dollop of bridging funds it had promised. Then came Hurricane Katrina, which damaged shuttle facilities in Mississippi and Louisiana, and an across-the-board federal budget cut, largely to raise money for the Iraq War. Worst of all, a new analysis of the shuttle and space station found them at least $2 billion in the hole. Griffin went cap in hand to the administration but was told to make up the difference from the agency's own wherewithal.

Compared with the plan of two years ago, science gives up a total of $6.4 billion (in 2005 dollars) over the five years from 2007 through 2011—a 20 percent cut. Planetary exploration is the worst-hit area—40 percent. Human spaceflight gains $5.2 billion, but its situation is hardly to be envied either. The shuttle fleet will make 16 rather than 28 trips to the space station before retiring in 2010, and from then until the CEV debuts, the country will have no capability to launch astronauts into orbit at all.

Griffin has described the shift of money as a "speed bump," a temporary measure to get human spaceflight back on course. Veteran observ-



"DEFERRED INDEFINITELY": Terrestrial Planet Finder, one part of which might consist of heat-shielded orbiting telescopes (*shown here collecting beams of light*), is one of many science missions that NASA's new budget zeroes out. Designed to look for life-related chemistry on Earth-like planets, the $1.7-billion project had been planned for launch in the next decade.

NASA/JPL

SLICED AWAY

NASA's redirection of funds to the space shuttle and space station will eliminate or postpone many of its science programs.

Eliminated:

- Deep Space Climate Observatory
- Geosynchronous Imaging Fourier Transform Spectrometer (GIFTS)
- Hydros
- Jupiter Icy Moons Orbiter (JIMO)
- Keck Outrigger Telescopes
- Mars Scout series (after 2011)
- Mars Telecommunications Orbiter
- Nuclear Spectroscopic Telescope Array (NuSTAR)

Delayed "indefinitely":

- Constellation-X
- Laser Interferometer Space Antenna (LISA)
- Mars Sample Return
- Stratospheric Observatory for Infrared Astronomy (SOFIA)
- Terrestrial Planet Finder (TPF)

Delayed one to three years:

- Global Precipitation Measurement Mission (GPM)
- Landsat Data Continuity Mission
- National Polar-Orbiting Operational Environmental Satellite System (NPOESS)
- Space Interferometry Mission (SIM)
- Wide Field Survey Explorer (WISE)

ers express sympathy for his dilemma. "It's a knotty problem," says John M. Logsdon, director of the Space Policy Institute at George Washington University. "There's no clear answer."

Nevertheless, many complain that he has been heavy-handed. Multiyear projects require some consistency in their funding. By making such an abrupt budget change, NASA will mothball or abandon half-built (in some cases, fully built) hardware, lose expertise developed at great effort, and leave gaps in data coverage, notably of the earth's climate. NASA has had budget crunches before, but seldom have they been so wasteful.

"It's the sudden change in slope: that's why this is more difficult than it was in previous years," says Lennard Fisk, chair of the National Research Council's Space Studies Board and himself a former NASA official. The unprecedented targeting of research grants strikes scientists as particularly gratuitous: for a small savings, only about $80 million, NASA is causing a huge disruption.

The Space Studies Board is investigating how to hold on to the grants and smaller missions by delaying or downgrading the bigger fry. Several flagship missions, such as the James Webb Space Telescope, have run over budget and need housecleaning anyway. Wesley Huntress, director of the Geophysical Laboratory of the Carnegie Institution of Washington and another past NASA official, says scientists need to take responsibility for making the necessary trade-offs, rather than leaving it up to NASA headquarters and Capitol Hill.

Some hope the crisis may finally force some out-of-the-capsule thinking. Should NASA jettison the shuttle and station right away? Should it do the opposite and stretch out the station's construction to reduce its annual cost? Should NASA be split into separate science and astronautics agencies? If it were, would that really be good for science? Unless some helpful reform can be salvaged from the situation, what seemed like such a grand vision two years ago may fail in the execution.

BIOCHEMISTRY

Body Blazes

HOW NICOTINE STOPS INFLAMMATION COULD LEAD TO NEW DRUGS **BY LISA MELTON**

Nicotine has undergone an image overhaul, at least biomedically. In the past few years researchers have found that the substance can alleviate symptoms of ailments such as Alzheimer's disease and ulcerative colitis. Just how nicotine battles these foes, however, has remained unclear. Now, by studying sepsis, Luis Ulloa of North Shore University Hospital in Manhasset, N.Y., has evidence elucidating nicotine's biochemical pathways that could lead to more potent anti-inflammatory drugs.

NICOTINE is a powerful anti-inflammatory but too dangerous to use as a therapeutic.

Sepsis, the most lethal of inflammatory conditions, is a bacterial invasion of the bloodstream. The third leading cause of death in the developed world, it accounts for nearly 10 percent of overall deaths in the U.S. every year. Infection causes part of the damage, but what makes patients critically ill is their own fiercely aggressive immune response. Macrophages churn out huge quantities of proinflammatory cytokines. This exaggerated immune response leads to tissue damage, and eventually the patient dies of cardiovascular dysfunction and multiorgan failure.

Ulloa and his collaborators have found something remarkable: nicotine can shut down this overshooting inflammatory response, to the point of reversing sepsis in mice. As far as anti-inflammatory treatments go, this is powerful stuff. "Nicotine taps into the body's own potent anti-inflammatory mechanisms," Ulloa explained in February at a Novartis Foundation meeting in London. "That is the beauty of our approach. By using nicotine, we are copying physiological mechanisms that have been selected by evolution to modulate the immune system."

Specifically, nicotine mimics acetylcholine, the Cinderella of neurotransmitters. Largely ignored over the years, acetylcholine has been catapulted into a starring role, linking the nervous and the immune systems. Through acetylcholine the nervous system controls the inflammatory fires that constantly crop up in our bodies. Receptors for acetylcholine reside not only on nerve cell endings but also on immune cells. Nicotine binds and activates these receptors, allowing cross talk between the brain and immune system.

"This is something quite phenomenal," comments Wouter de Jonge of the Academic Medical Center Amsterdam, who studies how macrophages respond to acetylcholine. "Smokers suffering from ulcerative colitis seemed to benefit from their habit, so there were hints that nicotine could ameliorate inflammatory diseases, but nobody could get a handle on it," he notes.

Now Ulloa's group may have provided an explanation for the positive effects that nicotine has on illnesses as diverse as schizophrenia, Alzheimer's, Parkinson's disease, Tourette's syndrome and ulcerative colitis. In laboratory experiments, Ulloa demonstrated that nicotine latches onto the nicotinic receptors on macrophages and stops them from spewing out inflammatory cytokines. This clampdown is brutally effective. The researchers also identified the specific receptor subtype, the alpha-7 acetylcholine receptor, that nicotine binds in macrophages to stop cytokine production.

But as a drug, nicotine is fraught with toxicity issues. Apart from its addictive nature, it can lead to cardiovascular problems and contribute to cancer. "No one is looking to use nicotine to treat inflammation," Ulloa says. "We want to design specific compounds that will target this receptor to take advantage of nicotine's anti-inflammatory effects while eluding its collateral toxicity."

"This is one of the great stories in immunology in the past few years—no question about it," remarks Mitchell Fink, an expert in critical care medicine at the University of Pittsburgh. A selective nicotinelike compound may be a promising therapy not only for sepsis but for a whole slew of chronic conditions, including heart disease, cancer and diabetes. The task at hand is to find the best surrogate for nicotine. Ulloa's petri dishes are the ones to watch.

Lisa Melton is based in London.

A NICK OF NICOTINE

As a potent anti-inflammatory, nicotine can damp down a dangerous immune response. But it is too risky as a treatment. Fortunately, substitutes may exist. Pharmaceutical firms have developed nicotinelike drugs, such as GTS-21, that were designed to stimulate the alpha-7 acetylcholine receptors in the brains of patients with Alzheimer's disease. But the clinical trials failed to show a clear benefit, and the drugs were dropped. The compounds may have been unable to cross the blood-brain barrier—which would actually be a plus for an anti-inflammatory, because then it could target the periphery and avoid the brain. Researchers have begun testing such substitutes to combat inflammation.

T & L/IMAGE POINT FR/CORBIS

DATA POINTS: GIRDLED BY GARBAGE

Discarded rockets, exploded satellites, paint flecks and even human waste contribute to the earth's orbital litter. The U.S. Space Surveillance Network tracks all man-made objects down to baseball-size fragments (10 centimeters wide). They pose a risk of serious collision with future space missions. According to a NASA analysis, even if humans stopped launching satellites now, debris will increase after 2055. That is when the formation of new pieces, resulting when larger ones break up, will exceed the rate of destruction by reentry burn-up.

Percent of orbital objects that are debris: 93

Number of fragments at least 10 centimeters wide: 9,000

Combined mass, in kilograms: 5 million

Number of known orbital collisions, 1991–2005: 3

Number of collisions expected in the next 200 years: 18

Number of collisions expected to be catastrophic: 11

SOURCES: U.S. Strategic Command; Science, January 20, 2006

BEHAVIOR

The Lure of Punishment

Penalizing others for their selfishness might not be fun, but it wins out against the alternative, suggesting how social sanctions and the cooperation they entail might have emerged over time. German researchers instructed volunteers to play a game in which they could contribute to a pot of cash-redeemable tokens, which paid dividends to everyone regardless of who contributed. In one group, subjects could pay a token to punish free riders who did not ante up, making the freeloaders lose tokens; the other group forbade punishment. Before each round, everyone saw each group's dividends and chose one to participate in. Although most players started out in the nonpunishment crowd, almost all ended up with the punishers, who saw their dividends grow because free riding was less common. Check out the April 7 *Science* for more—or else. —*JR Minkel*

HEALTH

Bottoms Up—or Maybe Down

For the health-conscious, conflicting studies reporting the supposed benefits of different foods must create confusion, if not hypertension. Now researchers have taken a second look at 54 studies that seemingly document the role of moderate drinking in staving off death from heart disease. They found that most of the investigations lumped nondrinkers with former drinkers, who might have stopped drinking for health reasons and pulled down the average health of that group. Of seven studies that split the two categories, none found a protective effect for alcohol. More telling, when the team regrouped the other published data, combining drinkers and former drinkers, the positive effects evaporated. Paradoxically, physiological data suggest drinking is probably still somewhat helpful, remarks Kaye Fillmore of the University of California, San Francisco, co-author of the May *Addiction Research and Theory* report. To bear that out, future analyses would have to disentangle drinking history, she says. —*JR Minkel*

MODERATE DRINKING may not protect the cardiovascular system.

BIOENGINEERING

Viral Wire

Traditional lithium-ion batteries have carbon electrodes, which are relatively bulky for the energy they provide. To slim electrodes down, researchers have turned to that paragon of self-assembly, the virus. A team at the Massachusetts Institute of Technology engineered a long tubular virus called M13 to sheathe itself in proteins that glom onto cobalt and gold atoms. The scientists dipped a polymer electrolyte film into a solution containing the virus, then into a solution of metal atoms, yielding a thin, transparent sheet coated in cobalt oxide and gold. The result, an electrode, stores almost three times as much energy as the carbon-based kind, the researchers reported online April 6 in *Science*. Next, the goal is to build a counter electrode and, ultimately, a self-assembling battery.

—*JR Minkel*

PHOTOCUISINE/CORBIS; ILLUSTRATION BY MATT COLLINS

EVOLUTION

Fish Out of Water

MISSING LINK: Reconstruction of *Tiktaalik*, which bridges the gap between water and land animals.

New fossils of fish with limblike fins excavated from arctic Canada now serve as a missing link in the evolution of animals from water onto land. Three nearly complete specimens of a flattened, alligator-like fish species dubbed *Tiktaalik roseae* possess the scales, fins, snout and lower jaw of a fish, but they have the ribs, neck, skull, wrists and fingerlike bones of a land animal. Scientists had to dig the up to three-meter-long fossils out from icy rock in polar bear country, but they say the tundra resembled a subtropical version of today's Mississippi River delta when the fish was alive some 375 million years ago. The sharp-toothed predator's overlapping ribs would have produced a stiff trunk not required by fish buoyed up by water, suggesting it lived in the shallows, perhaps with excursions onto land, the researchers write in the April 6 *Nature*. —*Charles Q. Choi*

NEUROBIOLOGY

Peaking Intelligence

Brainiess in kids may not depend on the amount of gray matter so much as how fast it both thickens and thins as they grow up. Researchers at the National Institute of Mental Health and at McGill University used magnetic resonance imaging to scan the brains of 307 children from ages five to 19. They focused on the cerebral cortex, or gray matter, which is responsible for higher thinking. In kids who scored highest on the most commonly used IQ tests, the cortex was actually relatively thinner in early childhood, thickening rapidly between ages seven to 11, and peaking later than in their average-IQ peers. The delayed growth perhaps reflects a longer developmental window for high-level thinking circuitry. The cortex of the smart children also thins faster during the late teens, which could reflect withering of unused neural connections as the brain streamlines its operations. Scan the March 30 *Nature* for more details. —*Charles Q. Choi*

TRANSGENICS

The Other, Other White Meat

People talk about waiting until pigs fly, but those less patient may only have to wait until swine swim with the fishes: scientists have now made them more piscine. Omega-3 fatty acids can reduce the risk of heart attack and stroke. Fish oil is the best-known source for these fats, but seafood may also contain high levels of mercury. To grow their fishy pigs, a team of scientists inserted a gene into fetal pig cells that gives rise to connective tissue. These cells were cloned to generate piglets with omega-3 levels that are three times higher than normal. Future tests must assess whether the levels remain high into adulthood. Not only could these pigs lead to healthier bacon, but such oinkers could themselves have healthier hearts and therefore live longer, which would limit the loss of livestock for farmers. Digest the findings in the April *Nature Biotechnology*. —*Charles Q. Choi*

HEALTHY BACON coming soon?

UNIVERSITY OF CHICAGO (*top*); FOODPIX (*bottom*)

BRIEF POINTS

- **The world's first optical telescope dedicated to finding extraterrestrial intelligence opened April 11 in Harvard, Mass. The 72-inch-wide mirror is designed to glimpse any otherworldly laser flash lasting at least one nanosecond.**

  The Planetary Society announcement, April 11

- **Flower power: the diversification of ants 100 million years ago depended on the coincidental emergence of flowering plants, which presented ants with new ecological opportunities.**

  Science, April 7

- **Jesus may well have walked on water—in its frozen state. The Sea of Galilee (Lake Kinneret in modern Israel) could have chilled enough for parts of it to ice over during protracted cold spells 1,500 to 2,500 years ago.**

  Journal of Paleolimnology, April

- **When hit with the right laser frequency, gold particles 50 nanometers wide can heat up enough to melt ice over an area 1,000 times as big. The technology may offer a precise way to burn targets such as tumor cells.**

  Nano Letters, April 12

The Flipping Point

How the evidence for anthropogenic global warming has converged to cause this environmental skeptic to make a cognitive flip By MICHAEL SHERMER

In 2001 Cambridge University Press published Bjørn Lomborg's book *The Skeptical Environmentalist,* which I thought was a perfect debate topic for the Skeptics Society public lecture series at the California Institute of Technology. The problem was that all the top environmental organizations refused to participate. "There is no debate," one spokesperson told me. "We don't want to dignify that book," another said. One leading environmentalist warned me that my reputation would be irreparably harmed if I went through with it. So of course I did.

My experience is symptomatic of deep problems that have long plagued the environmental movement. Activists who vandalize Hummer dealerships and destroy logging equipment are criminal ecoterrorists. Environmental groups who cry doom and gloom to keep donations flowing only hurt their credibility. As an undergraduate in the 1970s, I learned (and believed) that by the 1990s overpopulation would lead to worldwide starvation and the exhaustion of key minerals, metals and oil, predictions that failed utterly. Politics polluted the science and made me an environmental skeptic.

Reducing our CO_2 emissions by 70 percent by 2050 will not be enough.

Nevertheless, data trump politics, and a convergence of evidence from numerous sources has led me to make a cognitive switch on the subject of anthropogenic global warming. My attention was piqued on February 8 when 86 leading evangelical Christians—the last cohort I expected to get on the environmental bandwagon—issued the Evangelical Climate Initiative calling for "national legislation requiring sufficient economy-wide reductions" in carbon emissions.

Then I attended the TED (Technology, Entertainment, Design) conference in Monterey, Calif., where former vice president Al Gore delivered the single finest summation of the evidence for global warming I have ever heard, based on the recent documentary film about his work in this area, *An Inconvenient Truth*. The striking before-and-after photographs showing the disappearance of glaciers around the world shocked me out of my doubting stance.

Four books eventually brought me to the flipping point. Archaeologist Brian Fagan's *The Long Summer* (Basic, 2004) explicates how civilization is the gift of a temporary period of mild climate. Geographer Jared Diamond's *Collapse* (Penguin Group, 2005) demonstrates how natural and human-caused environmental catastrophes led to the collapse of civilizations. Journalist Elizabeth Kolbert's *Field Notes from a Catastrophe* (Bloomsbury Publishing, 2006) is a page-turning account of her journeys around the world with environmental scientists who are documenting species extinction and climate change unmistakably linked to human action. And biologist Tim Flannery's *The Weather Makers* (Atlantic Monthly Press, 2006) reveals how he went from being a skeptical environmentalist to a believing activist as incontrovertible data linking the increase of carbon dioxide to global warming accumulated in the past decade.

It is a matter of the Goldilocks phenomenon. In the last ice age, CO_2 levels were 180 parts per million (ppm)—too cold. Between the agricultural revolution and the industrial revolution, levels rose to 280 ppm—just right. Today levels are at 380 ppm and are projected to reach 450 to 550 by the end of the century—too warm. Like a kettle of water that transforms from liquid to steam when it changes from 99 to 100 degrees Celsius, the environment itself is about to make a CO_2–driven flip.

According to Flannery, even if we reduce our carbon dioxide emissions by 70 percent by 2050, average global temperatures will increase between two and nine degrees by 2100. This rise could lead to the melting of the Greenland Ice Sheet, which the March 24 issue of *Science* reports is already shrinking at a rate of 224 ±41 cubic kilometers a year, double the rate measured in 1996 (Los Angeles uses one cubic kilometer of water a year). If it and the West Antarctic Ice Sheet melt, sea levels will rise five to 10 meters, displacing half a billion inhabitants.

Because of the complexity of the problem, environmental skepticism was once tenable. No longer. It is time to flip from skepticism to activism. SA

Michael Shermer is hosting an international conference on the science and politics of the environment at Caltech from June 2 to 4 (www.environmentalwars.org).

BRAD HINES

The New Geopolitics

Preventing wars and other strife will increasingly depend on facing the ecological consequences of our economic activities By JEFFREY SACHS

Each era has its own dominating themes of global politics. The 19th century had the politics of industrialization and empire. The first half of the 20th century bowed to world wars and economic depression. The second half was overshadowed by the cold war. Our era, I believe, will be dominated by the geopolitics of sustainability.

Economic development has become a generalized global phenomenon, except in sub-Saharan Africa and a few other poverty hot spots. Even those impoverished areas will probably achieve economic takeoff with a little international help and the application of "best option" technologies. The world's total economic throughput every year, adjusted for differences in countries' purchasing power and measured as the gross world product (GWP), now stands at approximately $60 trillion. Over the past century, the GWP has grown roughly 18-fold in price-adjusted terms.

With that increase in economic output have come some phenomenal benefits, such as rising life expectancy and improved overall public health, and some planet-threatening adverse effects, such as massive tropical deforestation, ocean fisheries depletion, man-made climate change, violent competition over limited hydrocarbon resources, and newly emerging diseases such as SARS and avian flu (H5N1). Until now, the favorable outcomes have outweighed the bad. Yet because many of the environmental consequences are hidden from view and from our national income accounts, we sit atop ticking ecological time bombs.

Every major ecosystem, whether marine or terrestrial, is under stress. The world economy is depleting the earth's biodiversity, ocean fisheries, grasslands, tropical forests, and oil and gas reserves. We are massively and quickly changing the climate. These trends are occurring on a planet of 6.5 billion people and with economic activities that are already unsustainable as practiced. Yet with the economic successes now propelling India and China and the momentum of global population growth, we are on a trajectory to some nine billion people and a GWP of perhaps $275 trillion by mid-century.

Without new technologies and a new kind of geopolitics, we won't get there. Hydrocarbon scarcities can easily lead to war unless we develop energy alternatives, including much safer ways to use the vast remaining coal reserves. The misguided Iraq War, in my view, counts as just such an oil war, and its disastrous results demonstrate the urgency of new global approaches to energy. Irreplaceable marine ecosystems will be destroyed unless we learn to develop environmentally sound aquaculture and slow the man-made acidification of the oceans. More hurricanes like Katrina will ravage our coastal-based economies as extreme weather events continue to mount in frequency, intensity and human impact. More humanitarian disasters such as in Darfur, Sudan, will result from inadequate water in arid regions. More global demands on food production will lead to massive deforestation in the Amazon and beyond unless diets and agricultural processes change markedly.

We sit atop ticking ecological time bombs.

Our global politics is not yet adapted to the challenges of sustainability. The superpowers spend far more time angling for short-term military and economic advantage than they do honoring international agreements on biodiversity, climate, oceans, desertification and other fundamental issues that will count much more for our well-being in the decades to come. Indeed, unless we face these problems honestly, terrorism and war are likely to be more frequent outcomes.

This new monthly column will be about the emerging geopolitics of sustainability and the search for genuine solutions. It will show that topics usually treated through a political lens—war, terror, corruption—more and more frequently have an ecological underpinning. Global market forces can be "re-engineered" to channel economic activity in a sustainable manner. Better technologies can square the circle of economic growth with sustainability. And perhaps most important, new approaches to global politics and governance itself, based firmly on the budding science of sustainability, can provide a vital bridge to future prosperity and peace. Stay tuned. SA

Jeffrey Sachs is director of the Earth Institute at Columbia University and of the U.N. Millennium Project.

ARNOLD ADLER

The Implicit Prejudice

Mahzarin Banaji can show how we connect "good" and "bad" with biased attitudes we hold, even if we say we don't. Especially when we say we don't By SALLY LEHRMAN

Mahzarin Banaji wrestled with a slide projector while senior executives filed grumpily into the screening room at New Line Cinema studios in Los Angeles. They anticipated a pointless November afternoon in which they would be lectured on diversity, including their shortcomings in portraying characters on-screen. "My expectations were of total boredom," admitted Camela Galano, president of New Line International.

By the break, though, executives for New Line and its fellow Time Warner subsidiary HBO were crowding around Banaji, eager for more. The 50-year-old experimental social psychologist from Harvard University had started with a series of images that showed the tricks our minds play. In one video clip, a team passed around a basketball. Of the 45 executives watching, just one noticed the woman who walked slowly right through the game, carrying an open white umbrella. After a few more examples, Banaji had convinced the audience that these kinds of mistakes in perception, or "mind bugs," operate all the time, especially in our unconscious responses to other people.

KATHLEEN DOOHER

MAHZARIN BANAJI: TESTING BIAS

- As a young Zoroastrian in southern India, Banaji says she had "greater latitude than other Indian girls in seeking the life of the mind." The religion's central notion now resonates in the "good-bad" distinctions she asks study participants to make.
- Battles her own implicit bias with screensaver images, such as of black intellectuals and women athletes, that counter social stereotypes.

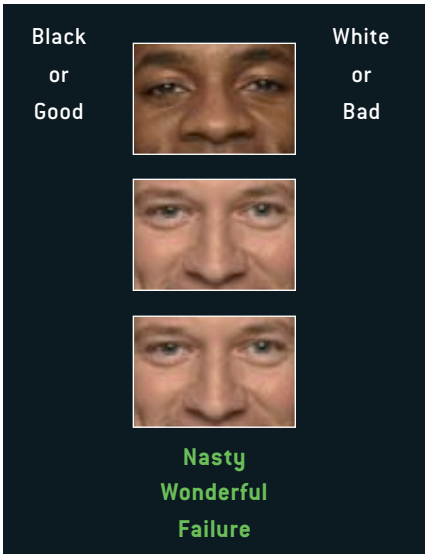
"It's reasonable and rational," Banaji told them. "And it's an error." We may intend to be fair, she explained, but underneath our awareness, our minds automatically make connections and ignore contradictory information. Sure enough, in a paper quiz, the executives readily associated positive words with their parent firm, Time Warner, but they found it harder to link them to their top competitor, the Walt Disney Company. To their chagrin, they discovered the same tendency to pair positive terms with faces that have European features and negative ones with faces that have African features.

Banaji has been studying these implicit attitudes and their unintended social consequences since the late 1980s, when she first teamed up with Anthony Greenwald of the University of Washington. Greenwald created the very first implicit association test (IAT). He measured how quickly people tapped keys on a computer keyboard in response to prompts on the screen. Would they more easily associate positive words such as "happy" or "peace" with pictures of flowers and negative words such as "rotten" or "ugly" with insects? Predictably, they did. Then he began testing responses to words and images associated with ethnicity and race. Participants' automatic reactions did not match the attitudes they said they held. Among social psychologists seeking investigative instruments, "the IAT just took off in a flash," Greenwald recalls.

In the decades since, Banaji, Greenwald and a third

collaborator, Brian Nosek of the University of Virginia, have continued to find fresh ways to use the IAT and other tools to probe bias: its nature, where it comes from and how it works. With neuroscientists, for instance, Banaji combined classical fear conditioning, implicit attitude measures and people's own descriptions of interracial dating to study how social groups come to fear one another. Banaji hopes next to work with primatologists to learn about our predisposition as a species to build bias into our perceptions.

Even in people with genuinely egalitarian views, Banaji and her colleagues find that bias is ordinary and ingrained and remains active outside our awareness. When the team realized the power of unconscious attitudes in everyday decision making, she says, "we knew the right thing was to take this to the public." On an IAT Web site (implicit.harvard.edu/implicit/), users can try 14 measures—to find out whether they automatically favor young over old, for instance, or prefer thin to overweight. Ten new sections include country-specific IATs, such as Muslim-Hindu and Pakistan-India associations.



IMPLICIT ATTITUDE TEST means rapidly putting images (here, of black and white faces) and words in green in the correct columns.

At least two million people have tried the tests online so far, and many have offered suggestions. "Once you put it out there, you have to listen to what people are saying—and their ideas are brilliant," Banaji finds. She has begun venturing from the lab to teach people about prejudice, employing humor, intellect and kindness as she alerts investment bankers, media executives and lawyers to the buried biases that lead to mistakes.

As a research tool, the IAT has fed close to 300 papers in fields ranging from neuroscience to marketing. It has also fueled academic challenge and debate, with a few social psychologists accusing the team of liberal bias and overinterpretation of the results. Some critics insist that the test does not really measure unconscious prejudice, only harmless cultural knowledge that differs from true racism. Psychologists argue over the underlying cognitive mechanism. One project found that some people will show bias just because they fear they will.

After finishing a meta-analysis across 61 studies, however, Greenwald and Banaji decided that the validity of the IAT holds. The test predicted judgments, behavior, and physiological reactions linked to stereotyping and prejudice better than expressed attitudes could. "In my own field, subtle prejudice, the IAT has helped crystallize ideas that we've been talking about for years," observes Jack Dovidio of the University of Connecticut. And it is an excellent teaching tool, he adds. When users experience their own discomfort and slowness in making associations, it is hard to ignore the message, agrees Princeton University social psychologist Susan Fiske. "Part of Mahzarin's genius was to see the IAT's potential impact on real-world issues," she points out.

Most recently, Banaji has been trying to discern when race attitudes first form and when conscious beliefs begin to diverge from those below the surface. In child-friendly tests, Banaji discovered that Japanese and white New England children as young as six both openly and implicitly preferred people like themselves. By age 10, their unconscious and conscious attitudes started to split. Despite expressing more egalitarian views as they grew older, people in the two societies continued to show automatic bias against black faces. For Japanese participants, both implicit and explicit attitudes toward European faces became more positive.

Banaji now suspects that if she could test for prejudice in babies, she would find it. But that does not mean that we are born with bias. Certainly we have the mental machinery to generalize and rank across social categories, she says, but culture fills in the necessary information. And humans absorb ideas about racial status early. In a study of 234 Hispanic-Americans, for instance, children compared themselves favorably with African-Americans. But when they used the IAT to compare themselves with white children, the natural preference for their own group fell away. "This work suggests that what we value, what we think is good, is in the air," Banaji remarks. It might develop through things like the warnings that a parent conveys to a child, in a tightening grip on a little hand. As adults, we continue to observe our environment and unintentionally adapt the stereotypes we hold to match.

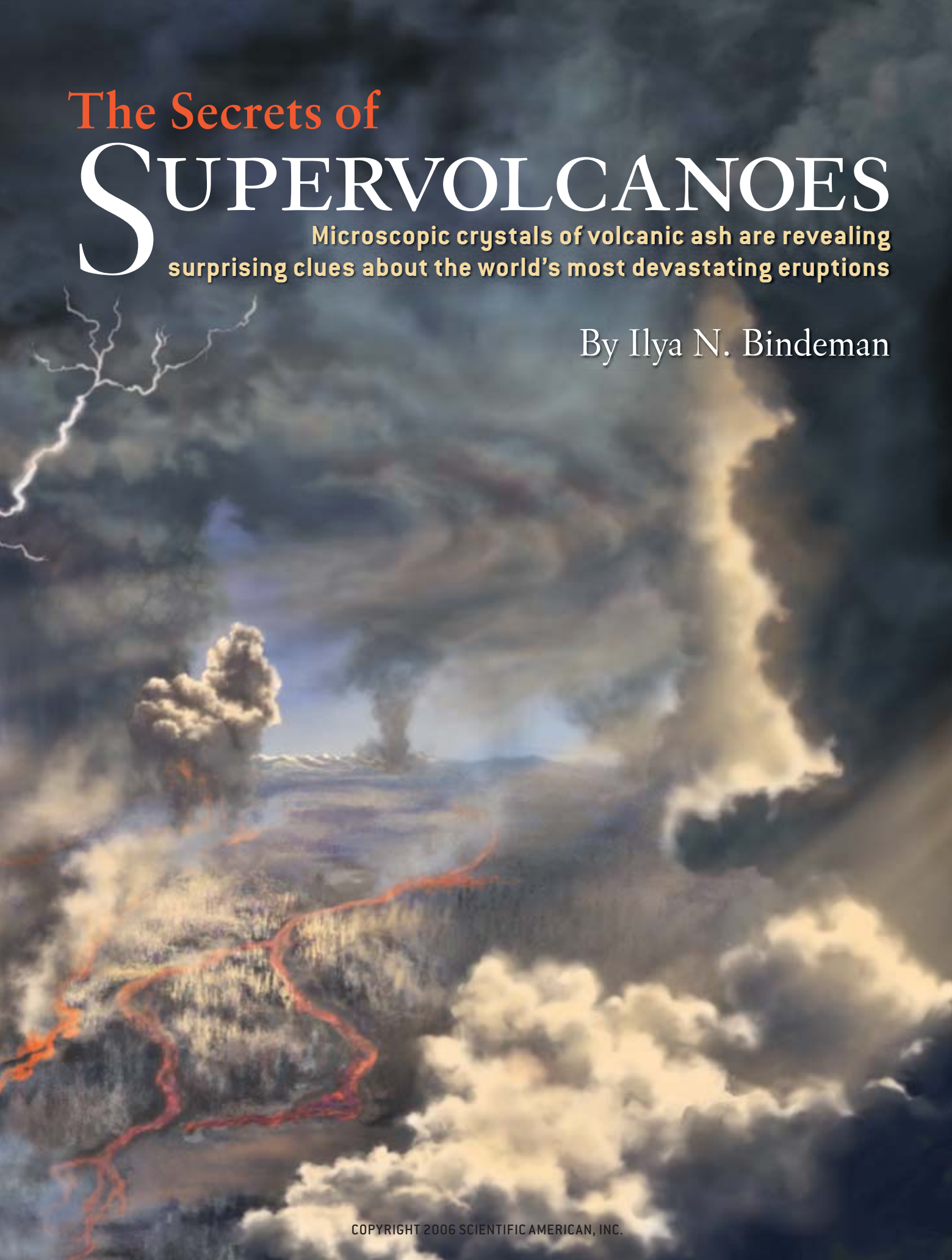
Fortunately, our brains do not seem permanently stuck on bias. Powerful cultural signals push in one direction, but awareness, close relationships and experience can push back. Banaji, Greenwald and Nosek are starting a nonprofit to help people apply their research. They envision seminars and lectures, followed by "booster shots" of online exercises.

By weaving awareness into our day, Banaji states, we can help our conscious attitudes take charge. It is like exercising regularly and eating healthfully, she explained to the filmmakers. And she suggested that they could build protective measures into their lives and work, much like fluoride in drinking water. "In every movie where you can do things counter to stereotype," she told them, "you are likely to produce change." SA

Sally Lehrman writes from the San Francisco Bay Area.

CARY WOLINSKY *Aurora Photos*

RING OF FIRE: Mountain-size vents exploding around the outer edge of an active supervolcano smother the landscape in clouds of hot gas and ash.



The Secrets of SUPERVOLCANOES

Microscopic crystals of volcanic ash are revealing surprising clues about the world's most devastating eruptions

By Ilya N. Bindeman

Lurking deep below the surface in California and Wyoming are two hibernating volcanoes of almost unimaginable fury. Were they to go critical, they would blanket the western U.S. with many centimeters of ash in a matter of hours. Between them, they have done so at least four times in the past two million years. Similar supervolcanoes smolder underneath Indonesia and New Zealand.

A supervolcano eruption packs the devastating force of a small asteroid colliding with the earth and occurs 10 times more often—making such an explosion one of the most dramatic natural catastrophes humanity should expect to undergo. Beyond causing immediate destruction from scalding ash flows, active supervolcanoes spew gases that severely disrupt global climate for years afterward.

Needless to say, researchers are eager to understand what causes these giants to erupt, how to predict when they might wreak havoc again, and exactly what challenges their aftermath might entail. Recent analysis of the microscopic crystals in ash deposits from old eruptions has pointed to some answers. These insights, along with improved technologies for monitoring potential disaster sites, are making scientists more confident that it will be possible to spot warning signs well before the next big one blows. Ongoing work hints, however, that supervolcano emissions could trigger alarming chemical reactions in the atmosphere, making the months following such an event more hazardous than previously suspected.

Almost all volcano experts agree that those of us living on the earth today are exceedingly unlikely to experience an active supervolcano. Catastrophic eruptions tend to occur only once every few hundred thousand years. Yet the sheer size and global effects of such episodes have commanded scientific attention since the 1950s.

Early Awe

ONE OF GEOLOGISTS' first discoveries was the existence of enormous circular valleys—some 30 to 60 kilometers across and several kilometers deep—that looked remarkably similar to the bowl-shaped calderas located atop many of the planet's most well-known volcanoes. Calderas typically form when the chamber of molten rock, or magma, lying under a volcanic vent empties out, causing the ground above it to collapse. Noting that these calderalike valleys sit close to some of the earth's largest deposits of volcanic rocks laid down during a single event, those early investigators realized they were seeing the remnants of volcanoes hundreds or even thousands of times larger than the familiar Mount St. Helens in Washington State. From the extreme scale of the calderas and the estimated volume of erupted material, researchers knew that the magma chambers below them had to be similarly monstrous.

Because the thick continental crust and heat sources needed to create such massive magma chambers are rare, supervolcanoes themselves are also uncommon. In the past two million years, a minimum of 750 cubic kilometers of magma has exploded all at once in only four regions: Yellowstone National Park in Wyoming, Long Valley in California, Toba in Sumatra and Taupo in New Zealand. The search for similarly large eruptions continues in other areas of thick continental crust, including in western South America and far eastern Russia.

By the mid-1970s, investigations of past events revealed some ways that the chambers can form and become dangerous. Under the surface of Yellowstone, the North American tectonic plate is moving over a buoyant plume of warm, viscous rock rising through the mantle, the 2,900-kilometer-thick layer of the earth's interior that is sandwiched between the molten core and the relatively thin veneer of outer crust. Functioning like a colossal Bunsen burner, this so-called hot spot has melted enough overlying crust to fuel catastrophic eruptions for the past 16 million years. In Toba, the source of the chamber is different. That region lies above a subduction zone, an area where one tectonic plate is slipping under another; the convergence produces widespread heating, mainly through partial melting of the mantle above the sinking plate.

No matter the heat source, pressure

Overview/*Mighty Eruptions*

- Recent analysis of the composition of tiny crystals inside ash deposits from prehistoric eruptions is overturning old beliefs about supervolcano behavior—and revealing new surprises about the aftermath.
- The inner workings of the magma chambers that fuel supervolcanoes can evolve in ways that strongly influence the style of future eruptions.
- The volcanic winter that grips the planet in the wake of a supereruption is probably shorter than once suspected, although chemical reactions in the atmosphere may be much more dangerous.

JULIA GREEN (*preceding pages*)

BIG, BIGGER, BIGGEST

Supervolcanoes (*orange* and *blue*) spread ash much farther than even large versions of what most people think of as "normal" volcanoes (*yellow* and *purple*), because the behemoths, with their massive magma chambers, eject so much more material.

in the magma chambers builds over time as more magma collects under the enormous weight of overlying rock. A supereruption occurs after the pressurized magma raises overlying crust enough to create vertical fractures that extend to the planet's surface. Magma surges upward along these new cracks one by one, eventually forming a ring of erupting vents. When the vents merge with one another, the massive cylinder of land inside the ring has nothing to support it. This "roof" of solid rock plunges down—either as a single piston or as piecemeal blocks—into the remaining magma below, like the roof of a house falling down when the walls give way. This collapse forces additional lava and gas out violently around the edges of the ring [*see box on next two pages*].

Fingerprinting Eruptions

YET MYSTERIES REMAINED. Notably, as researchers soon realized, not every large magma chamber will necessarily erupt catastrophically. Yellowstone, for example, is home to three of the world's youngest supervolcano calderas—they formed 2.1 million, 1.3 million and 640,000 years ago, one nearly on top of the other—but in the gaps between these explosive events, the underlying chamber released similar volumes of magma slowly and quietly. Why magma sometimes oozes slowly to the surface is still uncertain.

A look at the composition of tiny crystals trapped inside erupted lava and ash at Yellowstone has suggested a partial answer, by providing new insight into how magma forms. For decades, geologists assumed that magma sits as a pool of liquefied rock for millions of years at a time and that each time some of it pours out onto the earth's surface, new liquid rises up from below to refill the chamber immediately. If that conception were correct, one would expect many more catastrophic, voluminous eruptions, because it is mechanically and thermally infeasible to keep monster magma bodies in the crust without emptying them frequently.

The old idea was based largely on so-called whole-rock analyses in which researchers would obtain a single set of chemical measurements for each fist-size piece of volcanic rock they collected. Those data provided important general patterns of magma evolution, but they were insufficient for determining the age of the ejected magma and the depth at which it formed.

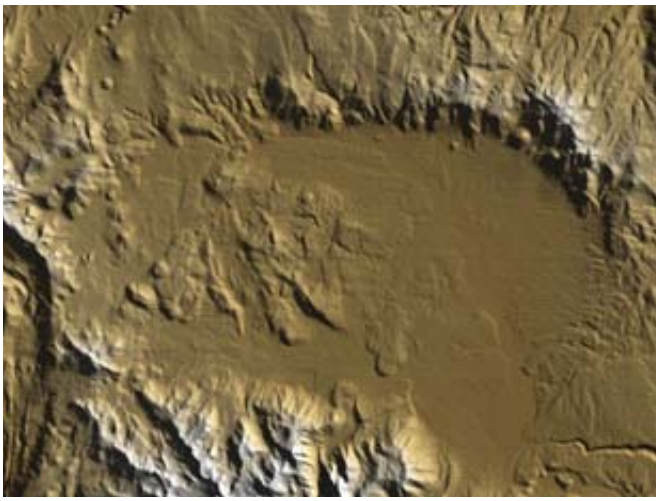
Every chunk of rock is actually made up of thousands of tiny crystals, each with its own unique age, composition and history. So when technological advances made it possible in the late 1980s to analyze individual crystals with good precision, it was like being able to read individual chapters in a book rather than relying on the jacket blurb to explain the story. Investigators began to see that some crystals—and thus the magmas in which they originally formed—arose much earlier than others, for instance, and that some formed deep underground, whereas others formed near the earth's surface.

During the past 10 years, geochemists have been paying particular attention to an especially durable type of volcanic crystal called zircon. Knowing that zircons can withstand extreme changes

THE AUTHOR

ILYA N. BINDEMAN is a geochemist and assistant professor in the department of geological sciences at the University of Oregon. Born in Moscow, Bindeman first became interested in volcanology while studying the remote volcanoes of Kamchatka in far eastern Russia. After completing his Ph.D. at the University of Chicago in 1998, he began investigating nearly microscopic crystals of ash for clues about the origin and effects of the world's largest eruptions. He worked at the University of Wisconsin–Madison and at the California Institute of Technology before joining the Oregon faculty in December 2004 and setting up his own geochemistry laboratory.

JEN CHRISTIANSEN; SOURCE: U.S. GEOLOGICAL SURVEY



SLEEPING SUPERVOLCANOES, such as that in Long Valley in California (*left*), are not obvious cone-shaped peaks like Washington State's Mount St. Helens (*above*). Instead they are marked by enormous calderas, depressions in the earth's surface that formed when the land collapsed into the magma chambers that fed the most recent supereruptions.

COURTESY OF J. S. LACKEY *College of Wooster*; DATA FROM U.S. GEOLOGICAL SURVEY (*left*); TODD CULLINGS *National Park Service* (*right*)

in heat and pressure without compromising their original composition, a few researchers—among them John W. Valley of the University of Wisconsin–Madison—have been using them to study the early evolution of the earth's crust [see "A Cool Early Earth?" by John W. Valley; SCIENTIFIC AMERICAN, October 2005]. When I joined Valley's team as a postdoctoral fellow in 1998, we used Yellowstone zircons to trace the history of their parent magma—which in turn revealed important clues about how the volcano may behave in the future.

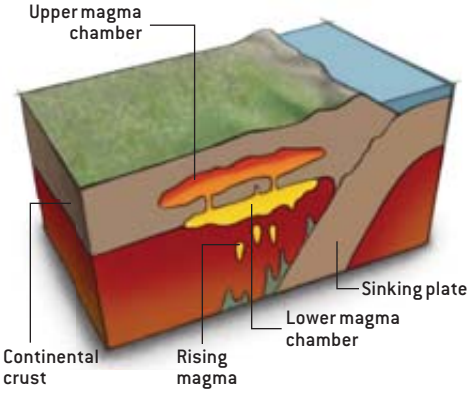
The first step was to measure the ratios of different forms of oxygen in zircons from the youngest Yellowstone supereruption—which after exploding 640,000 years ago gave rise to the Lava Creek tuff, a fossilized ash deposit 400 meters thick in some places—as well as younger deposits that were expelled during milder eruptions since then. When I finished my initial analyses, Valley and I were both surprised to see that oxygen composition of those zircons did not match that of deep, hot mantle, as would be expected if drained chambers always filled from below. Zircons born of mantle-derived magmas would have had a distinctive signature: as elements that are dissolved in magmas come together to form a zircon, that crystal takes on a notably high proportion of oxygen 18, a heavy isotope of oxygen that has 10 neutrons in its nucleus instead of the usual eight.

Valley and I saw immediately that the magma must have originated in rock once near the earth's surface. The zircons we studied were depleted in oxygen 18 relative to the mantle, and such depletion occurs only if the crystals formed from rocks that interacted with rain or snow. We thus suspected that the collapsed roof rock from one of the two oldest Yellowstone supereruptions must have melted to form the bulk of the magma that was ejected during the younger Lava Creek catastrophe and smaller eruptions since. This hypothesis gained strength when we learned that the ages of the zircons from post–Lava Creek eruptions span the entire two-million-year duration of Yellowstone volcanism. Such old zircons could exist in the youngest ash only if they originated in material that was ejected during the oldest eruptions and if that material later collapsed back into the magma chamber and remelted to help fuel the youngest eruptions.

Our findings mean that scientists can now expect to make certain predictions about how the Yellowstone supervolcano, and possibly those elsewhere, will behave in the future. If a new round of small, precursor eruptions begins in Yellowstone—and they usually do so weeks to hundreds of years before a catastrophic explosion—testing the oxygen fingerprint of those lavas and the ages of their zircons should reveal what type of magma is abundant in the chamber below. If the next eruption is depleted in oxygen 18, then it is most likely still being fed by stagnant remnants of the original magma, which by now is probably more of a thick crystal mush than an explosive liquid. On the other hand, if the new lava carries the fingerprint of fresh magma from the mantle and does not contain old zircons, then it very likely came from a large volume of new magma that has filled the chamber from below. Such findings would imply that a new cycle of volcanism had commenced—and that the newly engorged

SUPERCYCLES

The vast chambers of molten magma that feed supervolcanoes form above hot spots (buoyant plumes of rock rising from deep



1 Partial melting of the mantle rock above the sinking plate of oceanic crust produces magma that works its way up toward the base of the continental crust and pools there. This lower magma chamber acts as a colossal Bunsen burner that eventually melts parts of the continental crust, which has a lower melting point than the rock below. Some magma also rises via small vertical conduits between the two chambers.

magma chamber had more potential to explode catastrophically.

Immediate Aftermath

TINY CRYSTALS and their isotopic signatures have also revealed surprises—good and bad—about the aftermath of supereruptions. One of the best-studied examples of supervolcano aftermath is the Bishop tuff, a volcanic layer tens to hundreds of meters thick that is exposed at the earth's surface as the Volcanic Tablelands in eastern California. This massive deposit represents what is left of the estimated 750 cubic kilometers of magma ejected during the formation of the Long Valley supervolcano caldera some 760,000 years ago.

For decades, many geologists assumed that a series of distinct eruptions over millions of years must have occurred to produce the extensive Bishop tuff. But careful studies of microscopic, magma-filled bubbles trapped inside tiny crystals of quartz tell a different story. The rate at which magma leaves a chamber depends primarily on two factors: the magma's viscosity, or ability to flow, and the pressure difference between the chamber and the earth's surface. Because the pressure inside a bubble matches that of the chamber where the magma formed, the bubble acts like a mini version of the chamber itself.

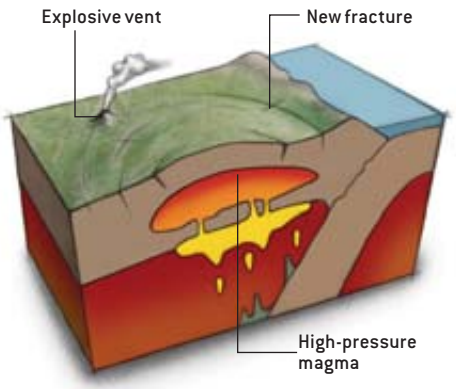
Aware of this correspondence, Alfred Anderson of the University of Chicago and his colleagues studied the size of the bubbles under a microscope to estimate how long it took the magma to leak out. Based on these and other experiments and field observations from the 1990s, geologists now think that the Bishop tuff—and probably most other supererupted debris—was expelled in a single event lasting a mere 10 to 100 hours.

Since that discovery, investigators have had to modify their reconstructions of supervolcano eruptions. Here is what they now generally expect from an event the scale of those that struck Long Valley and Yellowstone: Instead of a slow leak of red-hot lava as is seen creeping down the sides of Kilauea Volcano in Hawaii, these eruptions feature supersonic blasts of superheated, foamlike gas and ash that rise buoyantly all the way into the earth's stratosphere, 50 kilometers high. As the land above the magma chamber collapses, immense gray clouds called pyroclastic flows burst out horizontally all around the caldera. These flows are an intermediate stage between lava and ash, so they move extremely rapidly—up to 400 kilometers an hour, some sources say; cars and even small airplanes would have no chance of outrunning them. These flows are also intensely hot—600 to 700 degrees Celsius—so they burn and bury everything for tens of kilometers in every direction.

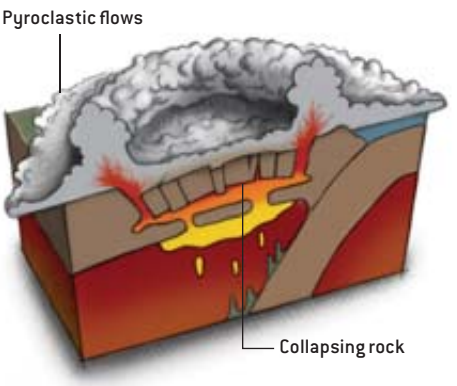
As bad as the pyroclastic flows are, the ash injected into the atmosphere can have even more far-reaching consequences. For hundreds of kilometers around the eruption and for perhaps days or weeks, pale-gray ash would fall like clumps of snow. Within 200 kilometers of the caldera, most sunlight would

JEN CHRISTIANSEN

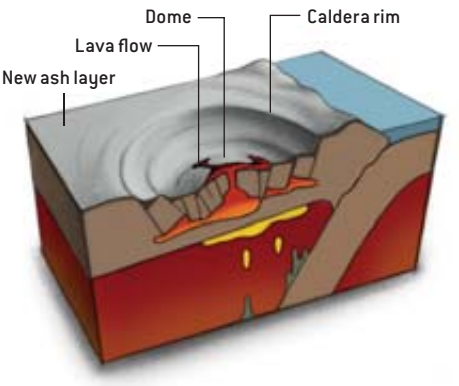
within the earth) and subduction zones (regions where one tectonic plate is slipping underneath another). In both cases, the giant volcanoes tend to follow an eruption cycle that is now far better understood than it once was. Here are the four basic steps, beginning with initial formation of the magma chamber, depicted for a subduction zone:



2 As the upper magma chamber grows, the land above bulges and cracks. The silica-rich composition and low temperature of this magma relative to those that form in the mantle make it particularly flow-resistant, so water and gas have trouble rising through it. Thus, when a plug of the sticky magma suddenly works its way to the surface along a vertical crack, the high-pressure material underneath tends to explode violently rather than oozing slowly.



3 The earth's strained surface eventually shatters as new explosive vents form a ring as wide across as the magma chamber. The fractured pieces of rock plunge down into the chamber, forcing additional magma up the outside edges of the ring. The sudden release of this magma transforms it into vast, scalding clouds of ash, gas and rock known as pyroclastic flows, which destroy the landscape for tens of kilometers in all directions.



4 After the volcano's eruption, a craterlike depression known as a caldera sits above the partially drained magma chamber. Over time, the collapsed land that is within the chamber begins to melt from the inside, thereby creating a smaller batch of magma that, along with other forces, forms a dome in the center of the caldera. Slow-moving lava may leak from this region many times before enough magma accumulates to fuel a new supereruption.

EXTENSIVE VOLCANIC DEPOSITS make up a steep slope of Yucca Mountain in Nevada. They are the remains of scalding ash flows resulting from supereruptions that struck nearby approximately 12.8 million (*lower layer*) and 12.7 (*upper layer*) million years ago.

MASSIVE WALL of gray rock in western Nebraska originated as a suffocating pile of ash left by a supereruption at an unknown site 28 million years ago. Elements in the ash imply that such explosions can alter the chemistry of the stratosphere.

be blocked out, so the sky at noon would look like that at dusk. Homes, people and animals would be buried, sometimes crushed. Even 300 kilometers away, the ash could be half a meter thick; mixed with rain, the weight would be plenty sufficient to collapse roofs. Less ash than that would knock out electrical power and relay stations. As little as a millimeter, which could well dust the ground halfway around the globe, would shut down airports and dramatically reduce agricultural production.

Only gradually would rain (made acidic by volcanic gases) wash away the thick blanket of ash. And because volcanic rock and ash float, it would clog major waterways. Transportation along big rivers could grind to a halt. Indeed, recent oil drilling in the Gulf of Mexico struck a surprisingly thick layer of supervolcanic debris near the Mississippi Delta—more than 1,000 miles from its source in Yellowstone. Only by floating downriver and then sticking to sediment that sank to the ocean bottom could that amount of debris have accumulated from a volcano so far away.

The Long Haul

INVESTIGATORS HAVE reason to believe that other consequences, arising from the great volumes of problematic gas expelled into the upper atmosphere, would also transpire and could persist for many years. New work suggests that some of these outcomes may not be as bad as once feared but that others may be worse. Once again, looking at the composition of small by-products from past eruptions has been illuminating.

Of the varied gases that make up any volcanic eruption, sulfur dioxide (SO_2) causes the strongest effect on the environment; it reacts with oxygen and water to produce tiny droplets of sulfuric acid (H_2SO_4). These droplets are the main sun-blocking source of the dramatic climatic cooling that would grip the planet in the wake of a supereruption. Knowing that the planet's hydrological cycle takes months or years to fully wash away the acid droplets, many researchers made apocalyptic estimates of "volcanic winters" lasting decades, if not centuries. But in recent years other investigators have uncovered evidence that drastically reduces that calculation.

Almost always, traces of the sulfuric acid produced after large volcanic eruptions are trapped in snow and ice as the acid precipitates out of the contaminated atmosphere. In 1996 investigators studying ice cores from Greenland and Antarctica found the sulfuric acid peak that followed the supereruption of Toba 74,000 years ago. That eruption ejected 2,800 cubic kilometers of lava and ash and reduced average global temperatures by five to 15 degrees C. The conse-

SKELETONS OF ANIMALS buried in ash from a catastrophic eruption in Idaho 12 million years ago are now exposed in northeastern Nebraska's Ashfall Fossil Beds State Historical Park. Most of the animals probably died slowly as the falling ash—which is essentially powdered glass—filled their lungs and abraded their teeth. Toxic chemicals in the ash may have poisoned their drinking water as well.

U.S. DEPARTMENT OF ENERGY (*top left*); ILYA N. BINDEMAN (*top right*); RICK OTTO *University of Nebraska State Museum* (*bottom*)

quences of such a chill were undoubtedly severe but did not last as long as once thought: sulfuric acid in the ice record disappeared after only six years; some researchers suggest that it vanished even earlier.

That volcanic winters are probably shorter than expected is the good news. But a new method developed in the past five years for studying the composition of the oxygen atoms in the volcanic acid rain is revealing an entirely different, alarming sign about the long-term effects of sulfur dioxide in the atmosphere. For SO_2 to become H_2SO_4, it must be oxidized—in other words, it must acquire two oxygen atoms from other compounds already existing in the atmosphere. Exactly which compounds play the key role is a hotly debated topic of current research, so when I started working with John M. Eiler as a staff scientist at the California Institute of Technology in 2003, he and I looked for evidence in my samples of ashes from the prehistoric Yellowstone and Long Valley eruptions.

We began our analyses with a focus on a particularly efficient oxidant, ozone. Ozone is a gas molecule made up of three oxygen atoms best known for shielding the earth from the sun's dangerous ultraviolet rays. Because of rare chemical transformations that certain gases undergo in the presence of that intense solar radiation, ozone is characterized by an anomaly in its so-called mass-independent oxygen isotope signature, which in simple terms can be thought of as an excess of oxygen 17.

When ozone or any other oxygen-rich molecule in the stratosphere interacts with SO_2, it transfers its oxygen isotope signature to the resulting acid—that is, the oxygen 17 anomaly persists in the new acid. In 2003 geochemists at the University of California, San Diego, found the first evidence that this signature is also preserved in the oxygen atoms of the acid that later falls as rain and in the sulfate compounds that form as the acid rain reacts with ash on the ground.

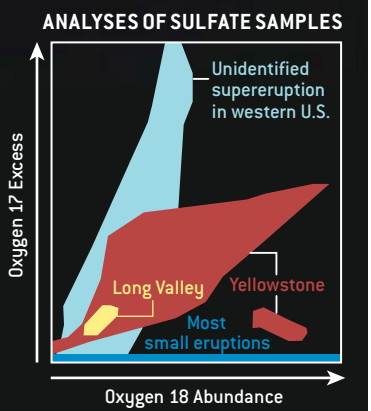
The oxygen 17 excess and other chemical patterns that we found in sulfate from the Yellowstone and Long Valley ash samples thus implied that significant amounts of stratospheric ozone were used up in reactions with gas from the supereruptions in those regions. Other researchers studying the acid layers in Antarctica have demonstrated that those events, too, probably eroded stratospheric ozone. It begins to look as if supervolcano emissions eat holes in the ozone layer for an even longer period than they take to cool the climate.

This loss of protective ozone would be expected to result in an increased amount of dangerous ultraviolet radiation reaching the earth's surface and thus in a rise in genetic damage caused by rays. The magnitude and length of the potential ozone destruction are still being debated. Space observations have revealed a 3 to 8 percent depletion of the ozone layer following the 1991 eruption of Mount Pinatubo in the Philippines. But what would happen after an event 100 times larger? Simple arithmetic does not solve the problem, because the details of atmospheric oxidation reactions are extremely complex and not fully understood.

Scientific techniques for studying and monitoring volcanoes of all sizes are developing with all deliberate speed. But no matter how much we learn, we cannot prevent an eruption. And what can be said about the aftermath of the most catastrophic occurrences is still speculative at best. The good news, though, is that researchers now know enough about the sites of possible eruptions to predict with reasonable assurance that no such catastrophe will happen anytime soon. SA

OZONE DESTRUCTION

Dangerous gases spewed from Mount Pinatubo in the Philippines in 1991 appeared as colors in satellite images of the earth's upper atmosphere (*background*). New evidence suggests that such gases emitted from future supervolcanoes may significantly deplete the planet's protective ozone layer before falling as acid rain and mixing with ash to form sulfate. Sulfate samples from four supervolcano deposits carry an unusual excess of oxygen 17 (*irregular colored areas in graph represent collections of measurements*); such abundance occurs only in compounds that have acquired the rare atoms during reactions with special gases, very likely ozone, in the earth's upper atmosphere. Materials that originate on the ground and stay there, such as the products of most small eruptions, show no such anomaly (*blue line*).



MORE TO EXPLORE

Low-$\delta^{18}O$ Rhyolites from Yellowstone: Magmatic Evolution Based on Analyses of Zircons and Individual Phenocrysts. Ilya N. Bindeman and John W. Valley in *Journal of Petrology,* Vol. 42, pages 1491–1517; 2001.

Sulfate Oxygen-17 Anomaly in an Oligocene Ash Bed in Mid-North America: Was It the Dry Fogs? Bao Huiming, Mark H. Thiemens, David B. Loope and Xun-Lai Yuan in *Geophysical Research Letters,* Vol. 30, pages 1843–1848; 2003.

Rare Sulfur and Triple-Oxygen Isotope Geochemistry of Volcanogenic Sulfate Aerosols. Ilya N. Bindeman, John M. Eiler, Boswell Wing and James Farquhar in *Earth and Planetary Science Letters* (in preparation, 2006).

NASA JOHNSON SPACE CENTER (*atmosphere spectrum*); GRAPH SOURCE: ILYA N. BINDEMAN

ENGINEERING LIFE: Building a FAB for Biology

BY THE BIO FAB GROUP*

*David Baker, George Church, Jim Collins, Drew Endy, Joseph Jacobson, Jay Keasling, Paul Modrich, Christina Smolke and Ron Weiss

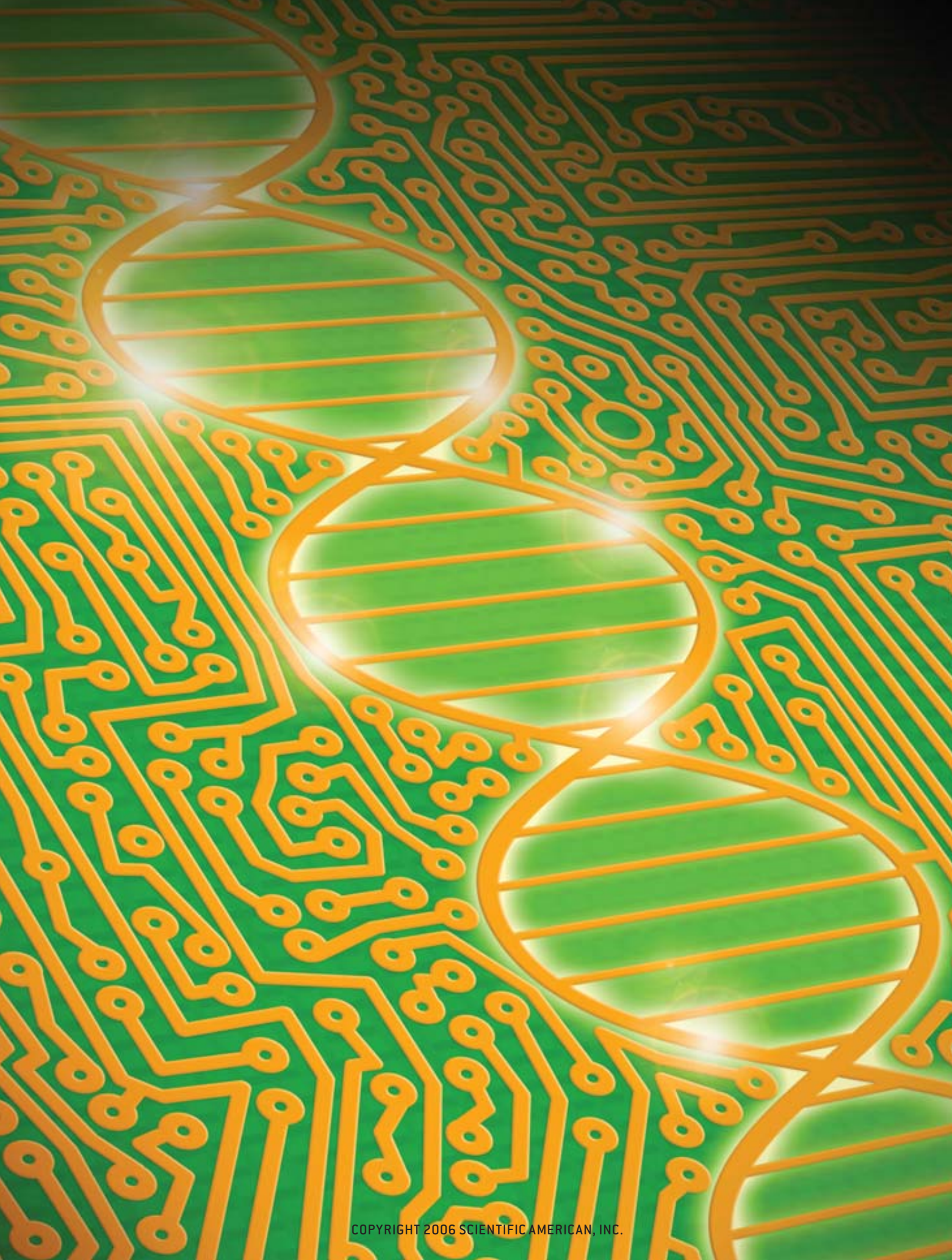
Principles and practices learned from engineering successes can help transform biotechnology from a specialized craft into a mature industry

Although the term "genetic engineering" has been in use for at least three decades, and recombinant DNA methods are now mainstays of modern research, most biotechnologists' work with living things has little in common with engineering. One reason is that the tools available for building with biological "parts" have yet to reach a level of standardization and utility equal to that in other engineering fields. Another has to do with methods and mind-sets in biology, although these, too, can be powerfully influenced by technology.

Electronic engineering, for example, was transformed beginning in 1957, when Jean Hoerni of Fairchild Semiconductor, a small company in what would later be known as Silicon Valley, invented planar technology. It was a system for layering and etching metals and chemicals within silicon wafers using templates called photomasks. This new approach allowed engineers to produce integrated circuits cleanly and consistently and to create a wide variety of circuit types just by changing the pattern on the photomask. Soon engineers could draw from libraries of simple circuits made by others

BIOLOGICAL COMPONENTS are the basis of an approach to biotechnology modeled on electronics engineering.

SLIM FILMS



and combine them in increasingly complex designs with a widening range of applications.

Until then, standard practice for manufacturing an electronic circuit had been to wire together its individual transistors one by one. It was an artisanal process, with uneven results, and a recognized bottleneck in the fledgling electronics industry. In contrast, planar technology kept steadily improving, enabling amazing advances at a rate famously commemorated by Moore's Law.

This combination of technology and methodology for designing and fabricating semiconductor chips—the "chip fab"—constitutes one of the most successful engineering paradigms of all time, and it is a valuable model for another nascent technology sector: fabrication of biological systems.

In effect, today's genetic engineers are still hand-wiring every circuit. As our colleague Tom Knight of the Massachusetts Institute of Technology artificial intelligence laboratory has observed, "The lack of standardization in assembly techniques for DNA sequences forces each DNA assembly reaction to be both an experimental tool for addressing the current research topic, and an experiment in and of itself."

Standardization of methods and components in biological engineering could give rise to design libraries of compatible parts and make outsourcing of fabrication possible. That uncoupling of concept and manufacture would free biological engineers to imagine increasingly complex devices and to use powerful engineering tools, such as computer-aided design, to manage that complexity. Toward these ends, members of our group have begun to identify and develop the equipment and techniques that could become the basis of a "bio fab." We are also trying to encourage a community that applies the best principles and practices of engineering to biotechnology.

It is a way of thinking about existing biological machines and of constructing new ones.

Overview/*Fab for Life Science*

- Flexible, reliable fabrication technology along with standardized methods and design libraries gave rise to the semiconductor chip "fab" system. It enabled engineers to create extraordinarily complex and powerful electronic devices with broad applications.
- A fab approach could similarly empower biological engineers to conceive and build sophisticated devices from biological parts.
- Bio fab technologies and techniques are already being developed and used. Addressing safety issues and encouraging biologists to think more like engineers are ongoing efforts.

Quality Parts

IF INDIVIDUAL TRANSISTORS are the basic components of electronic circuits, then their biological equivalents are genes: long, carefully ordered stretches of DNA. To construct genetic circuits for advanced biological devices, therefore, we need a way to manufacture long pieces of DNA quickly, reliably and at a reasonable price.

Twenty years ago Marvin H. Caruthers of the University of Colorado at Boulder built on earlier work by others to develop a system for synthesizing single DNA strands by exploiting their natural chemistry. DNA is composed of nucleotides, which are distinguished by the type of subunit, called a base, they contain: adenine (A), cytosine (C), guanine (G) or thymine (T). Affinities between the bases cause them to pair with one another—A with T and C with G—to form the rungs of the ladderlike double-stranded DNA molecule. Chemical groups create the bonds between base pairs as well as between adjoining nucleotides along either strand.

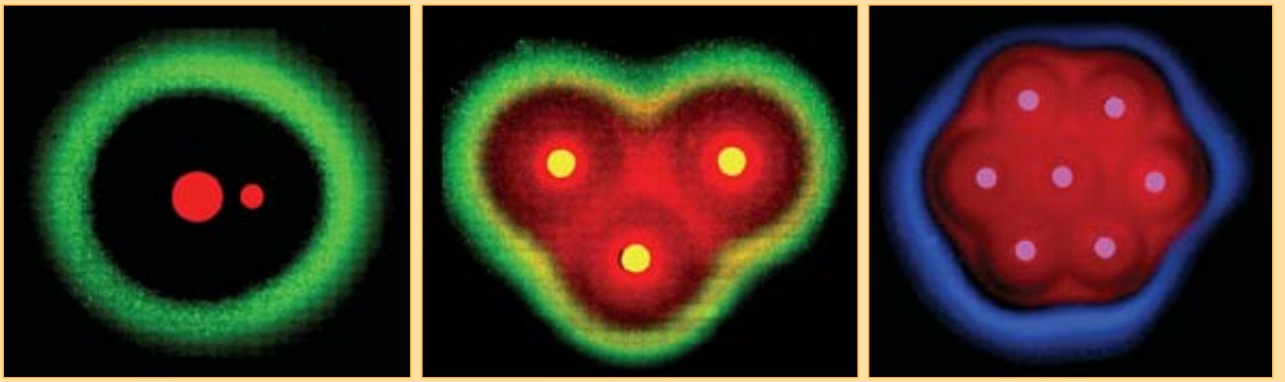
Caruthers's method, known as solid phase phosphoramidite chemistry, is still the basis of most commercial DNA synthesis. It begins with a single nucleotide attached to a solid support, such as a polystyrene bead, suspended in liquid. When exposed to an acid, the nucleotide's base becomes open to forming a bond with a new nucleotide added to the solution. That second nucleotide is then exposed to acid, and another nucleotide is joined to it, contributing to the growing chain. Repeating this cycle makes it possible to synthesize any desired nucleotide sequence with an error rate of approximately one base in 100.

Unfortunately, many of the genetic constructs that biological engineers wish to build are far longer than that. A simple network of genes may be several thousand bases long; the genome of even a small organism such as a bacterium can run to several million bases. Several of us working on finding synthesis methods with higher yields and lower error rates have therefore looked to nature for clues.

In living organisms, biological machinery composed of enzymes such as polymerase is able to manufacture and repair DNA molecules at speeds of up to 500 bases a second, with error rates of about one base in a billion. That represents a trillionfold performance improvement in yield throughput (output divided by error rate) over the best DNA synthesis machines, which add a base every 300 seconds. Moreover, multiple polymerases work in parallel when copying a long piece of DNA, such as a bacterial genome, so they are able to churn out about five million bases in 20 minutes.

One of us (Church) set out to emulate that parallelism by adapting the existing technology of microarrays. These are

Living Devices



Programmed bacteria form ring patterns in response to signals from colonies of "sender" cells that appear red (*left*), yellow (*center*) or pink (*right*). Lawns of engineered *Escherichia coli* "receiver" cells detect chemicals emitted by senders placed in their midst. The receivers then produce different-colored fluorescent proteins depending on their distance from the senders, which they judge by the chemicals' concentration. Increasingly complex patterns can be created by varying the initial placement of sender colonies. This example of an artificial multicellular system can be used in research to understand signaling and pattern formation in natural cell systems, such as during an organism's development. The same technology could also be applied in sensors, three-dimensional tissue engineering, and biofabrication of materials by programmed organisms.

large slides dotted with short, single DNA strands known as oligonucleotides, or oligos, about 50 to 70 bases in length. They are manufactured simultaneously right on the microarray surface using phosphoramidite chemistry, anchored in a grid pattern that approaches densities of one million dots per square centimeter. To the traditional technology, we added cuttable linkers that allow specific oligos to be released from the microarray. Each dot in our experimental microarray is about 30 microns wide and contains some 10 million oligo molecules.

We call these strands construction oligos because they are designed to overlap with one another in sequence so that they can later be assembled to form longer DNA constructs, such as a whole gene. But any oligos containing sequence errors must be weeded out. For that purpose, we have pursued two different error-correction systems.

The first uses the same microarray synthesis method to produce what we call selection oligos, with sequences complementary to the construction oligos. We then release the selection oligos from their slide and wash them across the construction oligo array. The selection oligos will follow base-pairing rules and bind, or hybridize, with their complementary construction oligos to form double DNA strands. We can then identify any unmatched construction oligo strands or gross imperfections in bound pairs as containing errors and release the bad oligos from the array. Interestingly, although the selection oligos are just as likely to bear some mistakes—having been made in the same manner as the construction oligos—the probability that erroneous sequences in either set will find a perfect complement is very low. Thus, using one set of oligos to proofread the other is an effective approach that allows us to create oligos with an average of only one error in every 1,300 bases.

As one might expect, biological systems have an interest in copying themselves accurately, and our second method of error correction is borrowed directly from nature. One of us (Modrich) first worked out the details of the process 10 years ago and dubbed it "MutS, L, H." When two DNA strands hybridize but their A-T and C-G base pairing is not perfect, the double-stranded molecule will not assume a helix shape at the location of the mismatch. MutS is a naturally occurring protein that recognizes and binds to such imperfections and eventually recruits other proteins, MutL and MutH, to correct the error. One of us (Jacobson), with Peter Carr of M.I.T., has employed this system to achieve error rates of just one in 10,000 bases of synthetic DNA, which is sufficient fidelity to produce small networks of genes.

These technologies—releasable parallel synthesis and error correction—permit us to assemble long, relatively error-free DNA constructs far more rapidly and inexpensively than has been possible to date. They can therefore constitute the

GREETINGS FROM BACTERIA glow in this device entered in the 2004 iGEM competition by a team from the University of Texas at Austin. The group incorporated multiple light-receiving and color-producing genetic parts into *Escherichia coli* to turn a biofilm into biological film that displays images inscribed by light on its surface. In a nod to computer programming tradition, the machine's first message was "Hello World."

COURTESY OF RON WEISS *Princeton University (top)*; AARON CHEVALIER/ UNIVERSITY OF TEXAS AT AUSTIN/UNIVERSITY OF CALIFORNIA, SAN FRANCISCO *(bottom)*

THE ABSTRACTION ADVANTAGE

Biological engineers can benefit from methods that made very large scale integrated (VLSI) electronics practical for the semiconductor industry. Standardization of technologies allowed chip engineers to specialize in circuit design or fabrication and to thereby manage complex problems at different levels of abstraction. Bio fab engineers can also cope with complexity by using abstraction hierarchies to hide unnecessary information. Thus, a bio fab designer working at the level of whole systems need worry only about which devices to include and how to connect them to perform the desired function without having to manufacture each device from scratch. Similarly, a device-level designer should know the functions and compatibility of individual parts within a device, whereas a parts-level engineer should understand how each part works internally but need not be able to synthesize its DNA raw material.

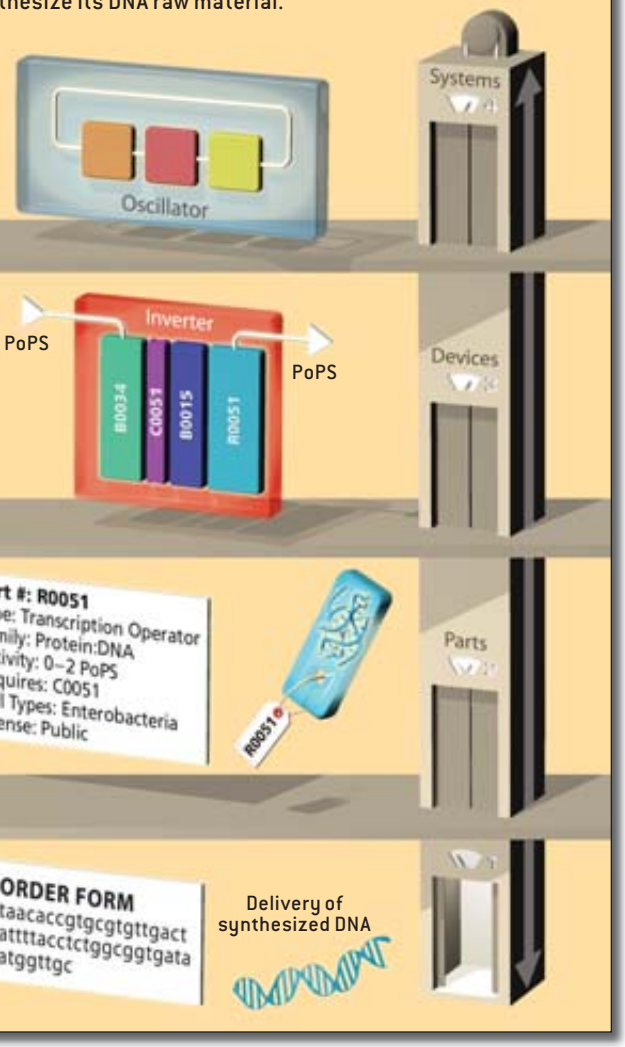
ABSTRACTION HIERARCHY

Systems
Combinations of biological devices that perform functions encoded by humans. A system of three inverters, for example, can operate as an oscillator.

Devices
Combinations of parts that perform discrete tasks. One inverter can take an input signal—for example, "HIGH"—and convert it to the opposite output signal, "LOW." A common signal carrier standard, polymerase per second (PoPS), allows devices to more easily be combined into systems.

Parts
Genetic material encoding biological functions. A transcription operator such as part #R0051, for example, is a piece of DNA that works with a matching binding protein (#C0051 in this case) to regulate gene activity. Off-the-shelf parts with clear specifications can be combined in a variety of devices.

DNA
Sequences for genetic parts. These can be specified by parts designers, manufactured off-site, then delivered. Fast synthesis technologies with low error rates make fabrication of custom DNA quick and reliable.



basis of a bio fab, and much like semiconductor chip lithography, these processes can be expected to keep steadily improving over time. That frees us to think about what we will build in the fab.

Nature Revised and Improved

AMONG OUR EARLIEST OBJECTIVES is using the bio fab platform to explore new ways of combating disease. Two of us (Keasling and Baker) run laboratories involved in creating cures for two of the most insidious diseases plaguing humanity: malaria and HIV. Although we are pursuing different types of therapeutics, the work of both our groups relies heavily on the ability to synthesize long, accurate pieces of DNA. Our projects thus offer examples of how the fab approach will considerably change the way biomedical scientists can go about developing new cures.

In the case of malaria, a treatment already exists that can eradicate the causative parasite from the body of an infected person. It is the small molecule C-15 sesquiterpene, commonly known as artemisinin, a natural compound made by the sweet wormwood plant, found mostly in northern China. The trees produce too little of the substance for the drug to be widely deployed at an affordable cost, however. That is why Keasling's group has been working for the past five years to copy the collection of genes, known as a genetic pathway, responsible for manufacturing artemisinin in the tree and

SLIM FILMS

inserting them into yeast to mass-produce the compound.

Once inside the yeast, this pathway can also be modified to operate much more efficiently than it does in the native plant. So far we have been able to redesign key subsets of the genes, known collectively as the mevalonate pathway, to produce an artemisinin precursor called amorphadiene at yields 100,000 times greater than the original pathway produces in bacteria. Increasing yields still further, to the point of making the drug widely available, will require us to reengineer the entire artemisinin pathway in an integrated way.

The full pathway consists of nine genes, each with an average length of about 1,500 DNA bases. Every new version of the pathway that we construct therefore contains approximately 13,000 bases. It would also be useful for us to be able to make variants of each gene in the pathway so that we could see which combinations worked most effectively. Manufacturing just two variants of each gene would mean synthesizing 2^9, or 512, constructs, for a total of about six million nucleotide bases. That is an extremely challenging goal using conventional DNA synthesis techniques, but that amount of DNA can fit on a single microarray chip.

The same technology that makes wholesale synthesis of gene networks possible can also be employed to generate novel proteins, such as new catalysts for synthetic chemistry reactions or environmental waste remediation and highly specific enzymes for gene therapy or pathogen destruction. Baker's group is developing computational methods for designing such new protein structures, including two that mimic essential features on the surface of HIV, which are already being tested as potential vaccines.

The trouble is that computer models are not sufficiently advanced to guarantee that each newly designed protein will have the desired function, but the computers can generate tens or hundreds of promising candidate structures to try. Turning all those into genetic sequences would require synthesis of hundreds of thousands of DNA bases—a difficult and expensive proposition using current technology but one well within the reach of the first generation of bio fab techniques.

These DNA and protein synthesis projects targeting malaria and HIV illustrate an approach, enabled by bio fab technology, that could be applied to a wider range of diseases, including newly emerging threats. For example, by combining high-speed, low-cost DNA sequencing methods [see "Genomes for All," by George M. Church; Scientific American, January] with the synthetic capabilities of the fab, a novel virus such as SARS or a new flu strain could be characterized, and protein-based vaccines against them could be readied far faster than is currently possible.

Of course, a bio fab is more than a collection of speedier synthesis technologies. It is a way of thinking about existing biological machines and of constructing new ones, which borrows both language and methodology from engineering.

BioBricks

In 2000 Michael Elowitz and Stanislas Leibler, then at Princeton University, as well as one of us (Collins), with colleagues Tim Gardner and Charles Cantor of Boston University, built the first basic circuit elements—a ring oscillator and a toggle switch—from biological parts. Scientists had known for some 25 years that natural organisms employ this type of circuitry to regulate their own genes, but the separate efforts of our two teams represented the first successes in manufacturing functional artificial biological circuitry.

Much of the future of biological technology will require many different groups to contribute subsystems.

What we mean by that term is well illustrated by Elowitz and Leibler's ring oscillator, which they began as an attempt to build a synthetic biological clock, hoping that it would provide insight into the clocks that exist naturally in biological systems. Their basic circuit consisted of a DNA ring called a plasmid containing three genes: *tetR*, *lacI* and λ *cI*, which encode the proteins TetR, LacI and λ cI, respectively. For any gene to be translated into a protein, the enzyme polymerase must first bind to a region of the DNA strand called a promoter that lies upstream of the gene. Polymerase then transcribes the gene into messenger RNA, which in turn is translated into a protein. If polymerase cannot bind the promoter, the gene is not translated and the protein is not made.

Elowitz and Leibler arranged for the protein products of the three genes in their circuit to selectively bind to one an-

THE AUTHORS

THE BIO FAB GROUP members are David Baker of the University of Washington, George Church of Harvard Medical School, Jim Collins of Boston University, Drew Endy and Joseph Jacobson of the Massachusetts Institute of Technology, Jay Keasling of the University of California, Berkeley, Paul Modrich of Duke University, Christina Smolke of the California Institute of Technology and Ron Weiss of Princeton University. They are friends, colleagues and sometime collaborators who wrote this article as a group because the diversity of their expertise, and hence of their contributions to the bio fab effort, embodies the interdisciplinary nature of biological engineering. All the authors are also scientific advisers to Codon Devices in Cambridge, Mass., the first commercial enterprise launched to apply engineering principles to synthetic biology. Church, Endy, Jacobson and Keasling are among its founders. Endy is also a founder of the not-for-profit BioBricks Foundation, and Keasling founded Amyris Biotechnologies.

Better Safety through Synthesis

Exploration of the many new opportunities a bio fab would offer for medicine, manufacture of new materials, sensors, waste remediation and energy production is just beginning. But like any worthwhile undertaking, it involves risk. A hallmark of biological systems is their ability to evolve and replicate, prompting understandable concerns that biological "devices" might cause unintentional or deliberate harm.

Thirty-one years ago a conference convened at Asilomar in California to address similar worries about the then new technology of recombinant DNA. For the first time, scientists could extract an individual gene from one organism and insert it into another, producing genetic combinations that might not exist naturally. That ability is now an essential tool in virtually every molecular biology laboratory in the world, in part because the governance that came from Asilomar eased fears about the use of recombinant DNA.

In a sense, then, the issues surrounding the new technologies that make up a fab for biology are not themselves new, but our community is committed to keep discussing them. A panel of scientists and ethicists tackled the implications of synthetic genomics at the Synthetic Biology 2.0 conference held in May. Its conclusions as well as results from a 15-month study of risks, benefits and potentially necessary safeguards, funded by the Alfred P. Sloan Foundation, will be available online at www.syntheticbiology.org in the coming months.

Right now scientists can certainly take the same precautions, such as working in secure biosafety laboratories, and observe the same ethical codes that have served us well for 30 years. Of course, ensuring that responsible investigators behave responsibly is easy. But the possibility that one day widespread access to DNA synthesis capability could allow malefactors to create deadly new pathogens, for example, is also a concern. It has prompted one of us (Church) to propose a compliance monitoring system that might include registration of synthetic biology workers—much as researchers working with so-called select agents are currently registered with the U.S. government—as well as surveillance of purchases of designer organisms, equipment and precursor materials for synthetic biology.

BIOBRICK VIALS contain DNA parts for biological engineering. These are safe for use in the lowest-level biosafety laboratories.

Another intriguing prospect is that the bio fab itself could represent the ultimate safe system because of the exquisite control it will permit. Most of the applications we have described would not require synthetic organisms ever to be exposed to the environment, but, just in case, those organisms could be created with genetic encoding different from any in nature, making it impossible for them to exchange genes with other life-forms. A synthetic biological device might be designed to self-destruct after a certain number of cell divisions or to be dependent for life on chemicals not present in normal environments. Genetic watermarks could be inscribed in every BioBrick to identify and track fabricated organisms. In other engineering disciplines, the ability to construct devices with higher precision affords greater safety—for example, in triply redundant flight-control systems in aircraft. We think the same may prove true of synthetic biological systems built in the bio fab.

—The Bio Fab Group

other's promoter regions. Thus, the LacI protein would bind the *tetR* promoter, whereas the λ cI protein would bind the *lacI* gene's promoter, and TetR would bind the promoter of the λ *cI* gene. These interrelations enable the protein product of one gene to block polymerase from binding to the promoter of another gene. Manufacture of the three proteins consequently happens in an oscillatory cycle: an abundance of LacI protein represses *tetR* gene activity; the absence of TetR protein then allows the λ *cI* gene to be turned on, which has the effect of repressing LacI production, and so on.

When one of the protein products in this cycle is also linked to a gene for making a green fluorescent protein and the entire circuit is inserted into bacteria, the oscillation of this device can be observed as the bacteria blink on and off like holiday lights. Similarly, the latest version of the Collins group's genetic toggle switch can be used to program bacteria to detect cellular DNA damage and then report their findings by arranging themselves into a green fluorescent "lawn" known as a biofilm.

Perhaps the most striking thing about these synthetic biological circuits is that they are identical in function to the first types of circuits that electrical engineers build when they want to test a new process for manufacturing semiconductor chips. Engineers know that basic components, such as an oscillator or a switch, are logically complete. Being able to build these simple parts reliably and accurately makes it possible to design and fabricate much more complex circuitry. Once biological engineers, too, can take such basic building blocks for granted, they can move on to more complicated projects such as multicellular systems, two- and three-dimensional designs, and devices whose function is not biological.

One of us (Weiss) recently produced a prototype for a multicellular system that could be used, for example, to detect

SAM OGDEN

explosives or other chemicals and then report them with a visible signal [*see box on page 47*]. This biological machine allows us to program millions of bacterial cells with instructions and protocols both for communicating with one another while carrying out their orders and for outputting light signals in a variety of patterns.

Inspired by these early examples, one of us (Endy), along with Knight and our M.I.T. colleague Randy Rettberg, is developing a library of biological components similar to the libraries available to chip designers. This Registry of Standard Biological Parts should facilitate a wide range of biological building projects, and our hope is that others will contribute new entries. So far the registry contains more than 1,000 individual BioBricks, as we call them, including many parts analogous to electronics, such as inverters, switches, counters, amplifiers, and components that can receive input or output a display. We have also defined a standard signal carrier—polymerase per second, or PoPS—akin to the current in a wire connecting two electronic components, so that bio fab engineers can more easily combine and reuse genetic devices.

To demonstrate the power of the fab approach and to seed this new field, the M.I.T. group offered the first course in fab-style engineering with biological parts in 2003. That class quickly evolved into an annual competition, which will draw teams from more than 30 universities this summer. In its short existence, the International Genetically Engineered Machine (iGEM) contest has already generated a number of amazing cellular devices, including biofilms that can record and display a photograph as well as programmed cells able to sense and respond like switches to small-molecule inputs, such as caffeine.

Still another iGEM entry that three of us (Smolke, Collins and Church) developed was a device capable of digital counting using a series of DNA segments. Twenty such DNA bits would be enough to count and report up to one million (2^{20}) cell states. This technology can be incorporated into sensors, which in turn might be connected to engineered metabolic pathways, such as the Keasling group's optimized version for artemisinin production. That would allow increased manufacture of the desired drug at literally the flip of a switch.

Constructing Synthetic Biology

WHEN WE AUTHORS of this article began our efforts to build a bio fab, no clear approach existed to making long DNA constructs accurately, rapidly and inexpensively. Today that is one among several technologies in an expanding toolbox for biological engineering. We are progressing toward first designing and modeling biological devices in computers, then "cutting" them into biological form as the final step—much as silicon chips are planned, then etched.

As with semiconductor circuitry, this approach has the added benefit of allowing us to optimize interactions between parts and to anticipate bugs. This ability grows increasingly useful as the constructed systems become increasingly complex. Yet another advantage of designing in the abstract is that a biological engineer does not need to actually build every part from scratch or even know how every one of them works internally—only that they do so reliably.

The students participating in iGEM may represent the first generation of biologist-engineers trained from the beginning of their careers to think of themselves as both. An important challenge going forward, however, will be to get more biologists to think like silicon engineers (and lure more engineers into biology)—particularly when it comes to sharing parts. Until now, biotechnology has been characterized by self-contained teams working to develop single-purpose applications, such as one drug compound. Much of the future of biological technology will require many different groups to contribute subsystems. Our hope is that building a fab for biology will facilitate that progression and help to spur advances as revolutionary as those achieved in the semiconductor industry. SA

Bio Fab Beginnings

A handful of companies and organizations are already applying engineering principles and tools to commercial biological manufacturing, bringing the fab closer to reality.

COMPANY	FOCUS
BioBricks Foundation Cambridge, Mass.	Promoting open tools, standards and parts for biological engineering
Blue Heron Biotechnology Bothell, Wash.	DNA synthesis
Amyris Biotechnologies Emeryville, Calif.	Engineering metabolic pathways for drug production in microbes
Codon Devices Cambridge, Mass.	Building biological devices
Foundation for Applied Molecular Evolution Gainesville, Fla.	Generating novel proteins and materials
Synthetic Genomics Rockville, Md.	Engineering microbes to produce fuels

MORE TO EXPLORE

Synthetic Life. W. Wayt Gibbs in *Scientific American,* Vol. 290, No. 5, pages 74–81; May 2004.

Foundations for Engineering Biology. Drew Endy in *Nature*, Vol. 438, pages 449–453; November 24, 2005.

Let Us Go Forth and Safely Multiply. George Church in *Nature,* Vol. 438, page 423; November 24, 2005.

Adventures in Synthetic Biology. Drew Endy, Isadore Deese, the M.I.T. Synthetic Biology Working Group and Chuck Wadey. A comic book available online at **http://openwetware.org/wiki/Adventures**

Registry of Standard Biological Parts: **http://parts.mit.edu/**

FECAL BACTERIA from animal and human wastes are threatening the health of beachgoers and shellfishers in the U.S. Disease-causing microbes such as *Escherichia coli* are carried to coastal waters by storm water runoff from commercial and residential developments and by seepage from improperly sited septic systems.

WADING IN WASTE

Thanks to unchecked development along America's coasts, disease-causing microbes are increasingly fouling beaches and shellfish beds

By Michael A. Mallin

America's stunning, sinuous coastlines have long exerted an almost mystical pull on the imaginations of the country's citizens. The irresistible attraction is perhaps best described by Herman Melville in the opening pages of *Moby Dick:* "Nothing will content them but the extremest limit of the land.... They must get just as nigh the water as they possibly can without falling in." In recent years, millions of Americans have moved to coastal areas, particularly in the Southeast, to take advantage of their balmy climate, recreational opportunities and natural beauty. Unfortunately, rapid and poorly planned development is spoiling this beauty in a shocking way: a growing number of beaches and shellfish beds along the coast have been contaminated by disease-causing microorganisms coming from animal and human wastes.

According to a recent report by the Natural Resources Defense Council, in 2004 coastal states ordered 19,950 days of closures and pollution advisories affecting 1,234 ocean and freshwater beaches, or about one third of all the beaches regularly monitored by health officials. The total number of beach days covered by the regulatory actions was 9 percent higher than the total for 2003 (which, in turn, was 50 percent higher than the 2002 total, although that jump was partly caused by changes in federal monitoring rules). The reason for 85 percent of the closures and advisories was the detection of excessive counts of fecal bacteria in the beach waters.

Moving downstream with animal feces in storm water runoff or with human waste in sewage overflows and septic-tank leaks, the waterborne microbes can cause liver disease, respiratory infections and potentially fatal gastrointestinal disorders. Such illnesses are common in Third World countries with poor sanitation, but in the U.S. the problem stems from unwise growth, not poverty. The construction of so many homes, roads, shopping centers and parking lots has disrupted the natural drainage systems in coastal areas, and wastes that were once filtered by forests or wetlands are now regularly fouling marinas and beaches.

How can coastal states and communities reduce microbial pollution? The issue has led to conflicts pitting developers and pro-growth politicians against regulatory authorities, commercial and recreational shellfishers, surfers, swimmers, divers and conservationists. Fortunately, some innovative solutions are available. "Smart growth" strategies can restore polluted coastlines and pro-

EMILY HARRISON; SOURCES: EMILY HARRISON, STEF OKASAKI, BEN HERSTON, EYE OF SCIENCE/PHOTO RESEARCHERS, INC., KEVIN FLEMING *Corbis* AND MARY GODLESKI *AP Photo*

vide economic benefits as well. Because frequent beach closures can dampen tourism and depress real estate values, the adoption of reasonable controls on coastal development would safeguard the shoreline economy as much as it would protect the public's health.

The Coastal Boom

IN ITS 2004 Coastal Trends Report, the National Oceanic and Atmospheric Administration noted that 153 million Americans—or 53 percent of the country's population—lived in the counties bordering the seacoasts and the Great Lakes, which make up only 17 percent of the continental U.S. land area [*see box on page 57*]. What is more, another seven million people are expected to join the coastal population by 2008, and the total is seasonally expanded by hordes of vacationers. Large areas that used to be forests or farmland are being turned into resorts, residential subdivisions, strip malls, restaurants, office complexes and industrial parks. In the process, construction companies are draining wetlands and covering formerly vegetated soils with asphalt, concrete and housing materials.

The resulting landscape is dominated by impervious surfaces—parking lots, roads, sidewalks, rooftops and construction sites compacted by heavy equipment—that do not let water soak through. When it rains, the water flows over these surfaces, picking up animal feces and other pollutants and washing them into drainage ditches or storm drains, many of which lead directly to urban lakes, coastal creeks or beach areas. Whereas sewage treatment plants remove harmful bacteria and other contaminants from their effluent, storm water runoff is usually untreated. Because this runoff emanates from a wide area rather than a single source, it is classified as nonpoint source pollution. The Environmental Protection Agency has stated that this type of pollution is the leading remaining cause of water-quality problems in the U.S.

Storm water runoff carries fertilizers, pesticides, heavy metals and petrochemicals, but it is the disease-causing microbes—the bacteria, viruses and protozoa derived from feces—that pose the principal threat to human health. A single gram of dog feces, for example, contains an estimated 23 million bacteria. (And for certain particularly hazardous bacteria, as few as 10 organisms can cause an infection.) Urban and suburban watersheds receive a steady influx of manure from domestic animals such as dogs and cats and from wildlife such as raccoons and squirrels. In vegetated areas, rainwater and runoff trickle through the soil, which cleanses the water of fecal bacteria and viruses, as well as many other pollutants. In contrast, impervious surfaces accumulate pollutants during dry periods and funnel high concentrations of the contaminants downstream when it rains.

The problem is particularly worrisome along the coast because the microbes pollute shellfish beds and areas used for recreation. Shellfish are primarily filter feeders, meaning that they strain large volumes of seawater through their bodies to concentrate food material such as microscopic algae. But they also concentrate the potentially harmful organisms that are present in the water. If humans consume raw or poorly cooked shellfish from waters contaminated with fecal microbes, they run the risk of contracting gastroenteritis (a disease characterized by vomiting, diarrhea and stomachache) as well as more severe, life-threatening illnesses.

To protect shellfish consumers, state agencies are required to post signs in polluted shellfish beds notifying the public that harvesting clams, mussels or oysters there is illegal. The U.S. Public Health Service has set a nationwide safety standard for shellfish beds using measurements of fecal coliform bacteria, a broad category of microorganisms found in the intestines of humans and animals. Shellfish cannot be harvested from the area if the geometric mean of the bacterial counts in 30 sets of samples is higher than 14 colony-forming units (CFU) per 100 milliliters of seawater. (A geometric mean is a type of average that minimizes the effects of outlying values.) In 1995, which was the last time NOAA compiled a national shellfish register, harvesting was restricted or prohibited in 31 percent of the country's shellfish-growing areas. The agency reported that urban runoff was the most commonly cited source of the pollution invading shellfish beds.

The reason for 85 percent of BEACH CLOSURES and advisories is the detection of excessive fecal bacteria.

Overview/*Microbial Pollution*

- Because of booming growth in U.S. coastal regions, much of the area near the shoreline is covered with impervious surfaces such as parking lots, roads and sidewalks. When it rains, storm water runoff from these surfaces can carry animal feces and their accompanying microbes into drainage ditches that lead directly to lakes, streams and beaches.
- Contamination by fecal bacteria is the leading cause of beach closures and advisories, which now affect one third of the country's monitored beaches. Hazardous microbes also plague marinas, tidal creeks and shellfish beds.
- To fight microbial pollution, coastal communities can encourage the preservation of green spaces, install filters in storm drains and prevent the placement of septic tanks in areas with porous soils.

FECAL MICROBES ON THE MOVE

Poorly planned development is the primary cause of microbial pollution in coastal areas. Most shopping centers, for example, are surrounded by enormous parking lots that funnel waste-laden runoff to drainage ditches. In many coastal residential developments, the effluent from septic tanks flows into fissured limestone or sandy soils that allow fecal microbes to seep into the groundwater. Industrial-style livestock operations typically spread their manure over fields or store it in lagoons; heavy rains can wash the wastes into nearby streams. And new construction often destroys the wetlands that filter disease-causing organisms from the water before they can reach beaches and shellfish beds.

DAVID FIERSTEIN

More recently, my laboratory at the University of North Carolina, Wilmington, analyzed data from five coastal North Carolina counties and found a strong correlation between human population growth and the closure of shellfish beds. In 1984, when 352,125 people lived in the five counties, 35,275 acres of shellfish waters were closed; by 2003, the combined population had risen to 501,596 and the closed acreage had grown to 42,304.

Microbial pollution also poses a serious danger to people involved in common recreational activities such as swimming, surfing, wading, diving, snorkeling, waterskiing and boating. If fecal organisms contaminate a lake, stream or seashore, anyone in the water risks infection by microbes entering through the mouth, nose, eyes or open wounds. Some of the illnesses caused by water contact include gastroenteritis, conjunctivitis (eye infections), cellulitis (skin irritations such as swimmer's itch), ear infections, respiratory infections and more serious diseases such as hepatitis and Guillain-Barré syndrome, an inflammatory disorder of the peripheral nerves that can induce paralysis. Some of the waterborne bacteria that can cause these health problems are *Escherichia coli*, *Clostridium perfringens* and various species of *Enterococcus*, *Aeromonas*, *Campylobacter*, *Salmonella*, *Shigella* and *Yersinia*. Among the many waterborne disease-causing viruses are hepatitis A and Norwalk; the pathogenic protozoa include *Cryptosporidium*, *Entamoeba* and *Giardia*.

Don't Drink the Water

IN RECENT YEARS, several of these microbes have triggered severe disease outbreaks in the U.S. and Canada. In

1993 *Cryptosporidium* protozoa infected the Milwaukee drinking water system, leading to more than 100 deaths and 400,000 illnesses. In 1999 an outbreak of *Escherichia coli* and *Campylobacter* resulted in two deaths and 116 illnesses among fairgoers in New York's Washington County after they drank water contaminated by runoff from a cattle barn. In 2000 residents of Walkerton, Ontario, fell victim to infections by *E. coli* and *Campylobacter* that caused 2,300 illnesses and seven deaths, mostly among the elderly and infants. Again, the contamination was traced to microbe-laden runoff from a cattle feedlot that entered the town's water supply.

Health officials measure the concentrations of various indicator bacteria to assess the danger from waterborne pathogens, especially along beaches. When the bacterial counts get too high, the authorities issue polluted-water advisories or close beaches to swimming and other recreational activities. The EPA recommends that states use *Enterococcus* bacteria as the safety indicator for ocean and bay water. Under the EPA's standards, seawater is considered unsafe if the geometric mean of five *Enterococcus* samples collected within 30 days is greater than 35 CFU per 100 milliliters or if any individual sample is above 104 CFU per 100 milliliters. In freshwater, the EPA limits on mean concentrations are 33 CFU per 100 milliliters for *Enterococcus* and 126 CFU per 100 milliliters for *E. coli*. These standards, however, may be too lax. The EPA has estimated that swimming in seawater with the maximum acceptable counts will cause 2 percent of bathers to fall ill.

Up the Creek

RESEARCHERS HAVE BEEN studying the environmental damage caused by impervious surface coverage since the late 1980s, but my laboratory was the first to examine the effects on fecal bacterial counts. We focused on New Hanover County, a rapidly growing area in North Carolina. From 1990 to 2000 the county's population increased by 25 percent, and it is expected to rise another 31 percent by 2020. My research team has studied the water quality of six urbanizing tidal creeks in the county for the past decade. We collected and analyzed more than 1,000 samples of fecal coliform bacteria and *E. coli* taken throughout the creeks and looked for correlations between the bacterial counts and various demographic and landscape attributes of the creek watersheds.

We found that the average fecal coliform counts were generally higher in the creeks with more people and with a larger percentage of developed land in their watersheds. But the bacterial counts were most strongly correlated with the prevalence of impervious surfaces. In Futch Creek—where impervious surfaces covered only 7 percent of the surrounding land—the average coliform count was 12 CFU per 100 milliliters, but in Bradley Creek—where asphalt and concrete blanketed 22 percent of the watershed—the count was more than seven times higher [*see box on page 58*]. The correlation between *E. coli* abundance and the percentage of impervious surface in the watershed was also very strong. Ours was not an isolated finding: A. Fred Holland, Denise M. Sanger and their colleagues at the South Carolina Department of Natural Resources later reported a significant correlation between fecal coliform counts and impervious surface coverage in the watersheds of 22 tidal creeks in the Charleston metropolitan area.

The results suggest that storm water runoff from developed areas may have a multiplier effect on bacterial concentrations downstream. The abnormally high flows coming from large parking lots or subdivisions may erode the drainage ditches and stream banks, bringing suspended sediments into the water. These sediments are also readily washed off construction sites where the soils have been stripped of vegetation. The suspended sediments and other particles cloud the receiving waters. (The degree of cloudiness is called turbidity.) What is more, the sediments, particularly clays, can physically and chemically bind with pollutants such as ammonium, phosphate, metals, and fecal bacteria and viruses.

Binding to soils protects bacteria from ultraviolet radiation that would normally kill the organisms. Bacteria can also obtain carbon, nitrogen and phosphorus nutrients from the soil particles, and the microbes can hitch long-distance rides as they travel downstream with the sediments. In our study of the tidal creeks in New Hanover County, we found a highly significant correlation between turbidity and the abundance of fecal coliform bacteria. Other studies performed in the Chesapeake Bay, western Florida, the Mediterranean coast and Australia have yielded similar findings.

The sediments lying at the bottom of shallow coastal waters are also a reservoir for fecal bacteria and other microbes. A research team led by Lawrence B. Cahoon, my collaborator at U.N.C. Wilmington, has found high concentrations of disease-causing organisms—including fecal coliform bacteria, enterococci and streptococci—in tidal creek sediments. The microbes can

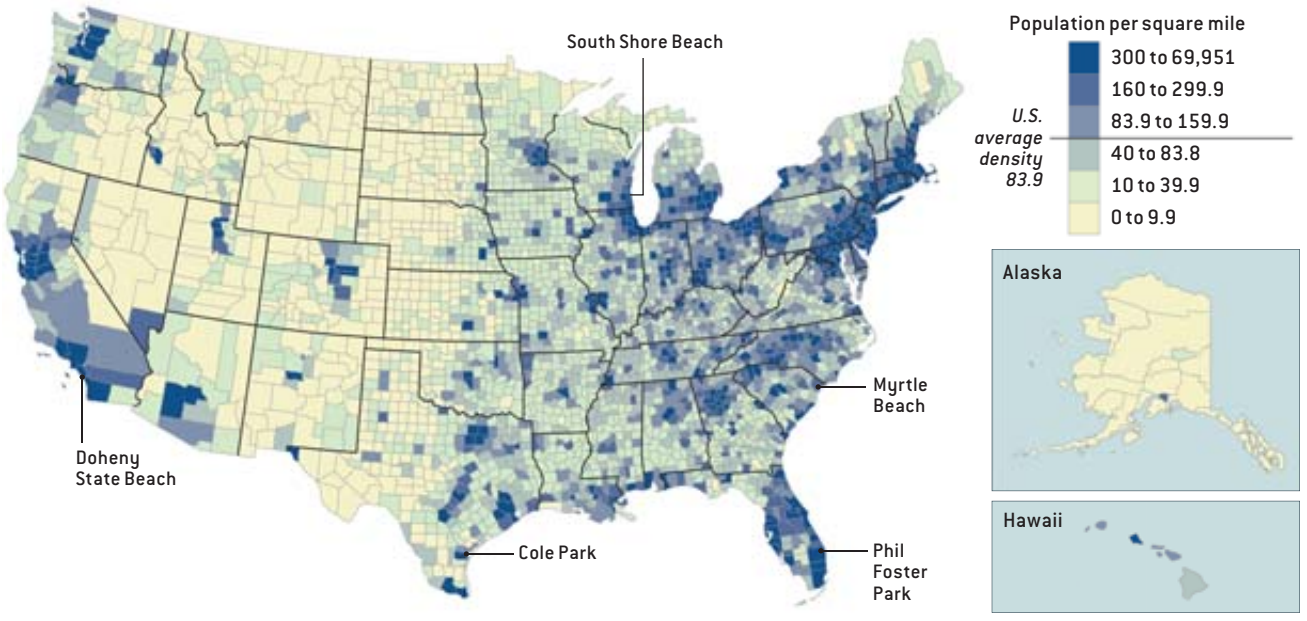
Fecal microbes from SEPTIC SYSTEMS in the Florida Keys can enter coastal waters within hours.

THE AUTHOR

MICHAEL A. MALLIN is an aquatic ecologist who has extensively studied pollution in freshwater, estuarine and coastal marine systems. Currently a research professor at the Center for Marine Science at the University of North Carolina, Wilmington, Mallin received his Ph.D. in marine and estuarine biology from U.N.C. Chapel Hill. His research interests include analyzing the impact of land-use practices on water quality as well as studying the effects of nutrient loading and catastrophic weather events.

AMERICA'S SULLIED BEACHES

The most polluted beaches and bays in the U.S. are generally located in densely populated coastal counties. Bacterial contamination prompted the closures and advisories at the areas listed below in 2004 (the last year for which nationwide records are available). The highest counts measured at each beach were far above the safety standards for individual samples of *Enterococcus* and *Escherichia coli,* 104 and 235 colony-forming units (CFU) per 100 milliliters, respectively.



BEACH LOCATION	TYPE OF BACTERIA MEASURED	HIGHEST COUNT (CFU PER 100 MILLILITERS)	NUMBER OF CLOSURE OR ADVISORY DAYS IN 2004
Doheny State Beach Orange County, California	*Enterococcus*	38,800	312
Phil Foster Park Palm Beach County, Florida	*Enterococcus*	600	108
South Shore Beach Milwaukee County, Wisconsin	*Escherichia coli*	2,419	72
Myrtle Beach Horry County, South Carolina	*Enterococcus*	1,130	54
Cole Park Nueces County, Texas	*Enterococcus*	14,400	53

SOURCES: U.S. CENSUS BUREAU (*map*); U.S. ENVIRONMENTAL PROTECTION AGENCY, NATURAL RESOURCES DEFENSE COUNCIL, ORANGE COUNTY OCEAN WATER PROTECTION PROGRAM, WISCONSIN BEACH HEALTH AND TEXAS GENERAL LAND OFFICE (*table*)

survive for extended periods in the sediments because they are protected from ultraviolet radiation and have ready access to nutrients. And because the tidal creeks are generally shallow, agitating the sediments at the bottom can suspend enough bacteria in the water to exceed the safety standards for human contact. Windy conditions or wading children and pets can easily pollute the water just by kicking up the creek's muddy bottom. Furthermore, our students have found high concentrations of fecal microbes in the sediments near several public boat ramps; such ramps are ubiquitous and heavily used in the coastal regions of the Southeast.

Sand and Sewage Don't Mix

POORLY DESIGNED sanitation systems in coastal regions can also contribute to microbial pollution. In communities where storm drains feed into sewer pipes, heavy rains can cause overflows that dump untreated human waste into rivers, lakes and bays. Many municipalities are now tackling this problem by separating their sewer and storm water systems. A new dilemma is emerging, however, in coastal areas where residents do not have sewage hookups and must put their wastes in septic tanks instead.

One such area is the Florida Keys, which is home to more than 25,000 septic systems. The predominant terrain in the Keys is karst topography—the underlying soils consist primarily of limestone, with many cracks and sinkholes caused by erosion. This geologic forma-

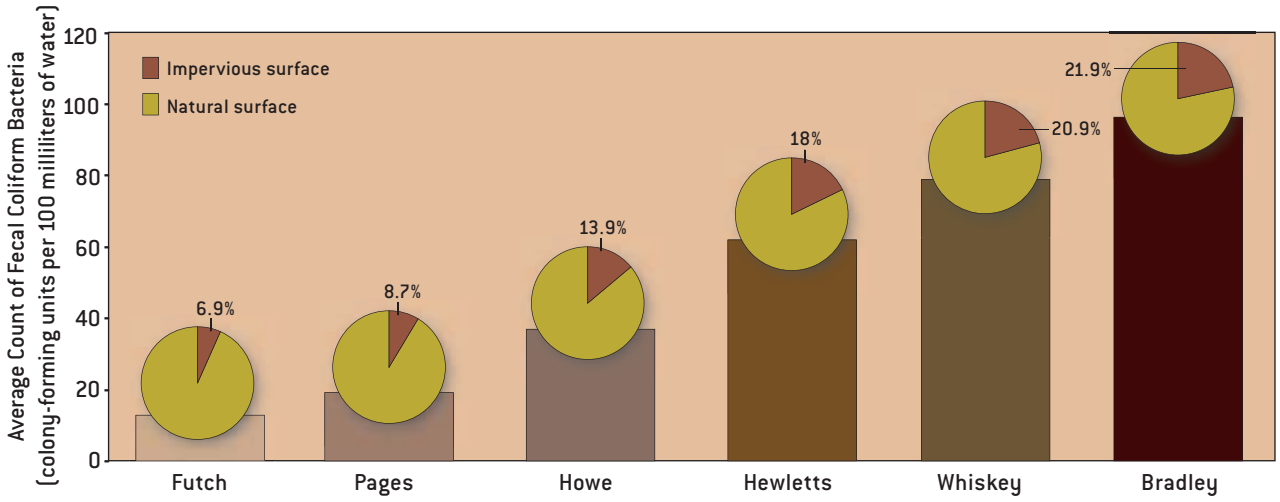
tion is very porous and hence cannot effectively filter the bacteria-rich effluent from septic tanks. In 1995 a research team led by John H. Paul and Joan B. Rose of the University of South Florida found that fecal microbes from septic systems in the Keys readily pass through the soil and can enter coastal waters near the shore within hours.

The problem is not limited to the Keys. Sandy soils along the coast have relatively large spaces between the sand grains. When these soils become saturated with water, bacteria and viruses can easily move through them. Areas containing sandy soils and a high water table are thus unsuitable for septic systems, yet poor planning has allowed their presence in many rapidly growing coastal regions, including some of the sandy barrier islands along the Atlantic and Gulf coasts. In a study of water quality in Brunswick County, North Carolina, which has many septic tanks in the sandy soil (up to 20 per hectare), Cahoon found high fecal coliform counts in the freshwater and seawater downstream from the dense residential sites. What is more, the ditching and drainage systems serving the developed areas appeared to facilitate the discharge of fecal bacteria into nearby waters, including shellfish beds.

Other investigators have identified distinctive patterns in the severity of microbial pollution. Erin K. Lipp, now at the University of Georgia, and her colleagues determined that in Florida Gulf Coast communities such as Charlotte Harbor and Sarasota Bay, fecal bacteria counts in bays and tributaries increase sharply with the outgoing tide. The microbes move readily through the sandy, saturated soils surrounding the crowded fields of septic tanks into nearby creeks draining into the bays. The pattern not only is tidal but varies according to weather conditions as well. The researchers discovered that in wet years corresponding to El Niño events, the water in Tampa Bay is significantly more polluted by fecal bacteria and viruses than it is in dry years. This effect is again the result of increased runoff and subsurface movement through saturated soils around improperly sited septic tanks.

THE EFFECTS OF CONCRETE AND ASPHALT

A study of six tidal creeks in New Hanover County, North Carolina, found a strong correlation between microbial pollution and the prevalence of impervious surfaces such as parking lots, roads and sidewalks. In the relatively pristine Futch Creek (*left*), where impervious surfaces covered less than 7 percent of the watershed, the average counts of fecal coliform bacteria were much lower than in Bradley Creek (*right*), where concrete and asphalt blanketed more than one fifth of the surrounding area.



LUCY READING-IKKANDA; MICHAEL A. MALLIN (*left*); NEW HANOVER COUNTY PLANNING DEPARTMENT (*right*)

In southern California, where large amounts of urban runoff flow into the Pacific Ocean, Rachel Noble of U.N.C. Chapel Hill and her colleagues found that the extent of shoreline water that failed to meet safety standards was 10 times greater following rainfall than it was during dry periods. Nationwide, many shellfish beds are automatically closed for a few days or weeks after a rainfall, because these areas are subject to bacterial pollution from storm water runoff.

But urban runoff and septic seepage are not always the main culprits behind microbial pollution. In coastal streams draining rural areas, the driving factor is often the waste from livestock farms. On the coastal plain that extends from Maryland to Florida and in certain Gulf Coast areas as well, traditional livestock farms have been largely supplanted by giant, industrial-style facilities where huge numbers of swine, poultry and cattle are raised in close confinement. The facilities dispose of the vast amounts of manure by either spraying it as a liquid or spreading it as litter on nearby fields. If the spraying or spreading occurs shortly before or during a rainstorm, fecal microbes from the waste can enter nearby streams via overland runoff.

A Cleaner Future

TO PROTECT America's coastal waters, developers and builders clearly need to move away from their current destructive practices—including clear-cutting, wetlands drainage and extensive use of pavement—and switch to smart-growth strategies. When planning new resorts, shopping centers, office complexes and residential subdivisions, the designers must minimize the use of impervious surfaces and maximize the amount of vegetated areas. A site with plenty of green spaces among the paved areas will have less runoff, and percolation through the soil will remove many of the contaminants swept from the asphalt and concrete.

In particular, wetlands need to be preserved and, if possible, enlarged to maintain the natural filtering of storm water runoff. In a study of 11 streams in the coastal plain of eastern North Carolina, my laboratory found that in the watersheds where the wetland coverage was relatively large—in this case, greater than 13.5 percent—periods of rainfall did not substantially increase the fecal coliform bacteria counts. The results indicate that preserving wetlands (and presumably expanding them) is an effective way to guard downstream waters from suspended sediments and microbial pollution. Builders can help the effort by curtailing the runoff of sediments from construction sites.

Developers should take advantage of new technologies that can reduce the amount of storm water runoff and even treat it on-site. For example, parking lots can now be paved with porous concrete, a semipervious substance that allows water to pass into the soil below and yet provides enough structural support for automobiles. And new collection systems can funnel storm water runoff from parking lots to filters that use layers of absorbent minerals and organic material to cleanse the polluted water. (Large parking lots can also be downsized, because most were built to accommodate holiday shopping rather than day-to-day traffic.) These technologies could be applied to existing developments as well as new ones. If a community wishes to reduce runoff pollution to revitalize its beaches or save the local shellfish industry, it can install filters in all its parking lots, create vegetated buffer zones along streams and ditches, and reconstruct wetlands in selected locations.

The poor performance of septic systems in sandy soils and in areas of karst topography is a clear indication that coastal regions require more advanced sewage treatment systems. Some conservation groups have pointed out, however, that the construction of centralized sewage systems will lead to even denser development and more water pollution. Thus, when coastal communities build sewage systems, the municipal authorities should put restrictions on impervious surface coverage in new developments, limiting it to, say, 10 to 15 percent of the total area (especially near shellfish beds).

America's coasts can be a wonderful destination, both for vacation travel and for relocating homes and businesses. But without careful planning and the political will to protect coastal resources, the gorgeous beaches, sparkling bays and tranquil tidal creeks will become nothing but hazardous receptacles of our waste. We must ensure that unchecked development does not ruin the very qualities that brought people to the coast in the first place. SA

PRESERVING WETLANDS is an effective way to guard downstream waters from microbial pollution.

MORE TO EXPLORE

Effect of Human Development on Bacteriological Water Quality in Coastal Watersheds. Michael A. Mallin, Kathleen E. Williams, E. Cartier Esham and R. Patrick Lowe in *Ecological Applications,* Vol. 10, pages 1047–1056; 2000.

Demographic, Landscape, and Meteorological Factors Controlling the Microbial Pollution of Coastal Waters. Michael A. Mallin, Scott H. Ensign, Matthew R. McIver, G. Christopher Shank and Patricia K. Fowler in *Hydrobiologia,* Vol. 460, pages 185–193; 2001.

Linkages between Tidal Creek Ecosystems and the Landscape and Demographic Attributes of Their Watersheds. A. F. Holland, D. M. Sanger, C. P. Gawle, S. B. Lerberg, M. S. Santiago, G.H.M. Riekerk, L. E. Zimmerman and G. I. Scott in *Journal of Experimental Marine Biology and Ecology,* Vol. 298, pages 151–178; 2004.

More information can be found online at **www.nrdc.org/water/oceans/gttw.asp** and **www.epa.gov/beaches/**

Advances in understanding the cells and molecules that transmit pain signals are providing new targets for drugs that could relieve various kinds of pain—including those poorly controlled by existing therapies

CONTROL

By Allan I. Basbaum and David Julius

Throbbing, itching, aching, stabbing, stinging, pounding, piercing.

Pain comes in a range of unpleasant flavors. But all pain has one thing in common: those who endure it want it to stop.

Yet the most widely used analgesics today are essentially folk remedies that have served for centuries: morphine and other opiates derive from the opium poppy, and aspirin comes from willow bark. Although these treatments can give relief, each has its limitations. Aspirin and other nonsteroidal anti-inflammatory drugs (NSAIDs), such as ibuprofen, cannot ease the most severe types of discomfort. And even opiates, generally the strongest medicines, do not work for everyone. Moreover, they can have serious side effects, and patients tend to become tolerant to them, requiring escalating doses to get any relief at all.

Over the past 20 years neurobiologists have learned a great deal about the cellular circuits and the specialized molecules that carry pain signals. Today this knowledge is being exploited to devise new strategies for managing pain better and causing fewer side effects. Indeed, more approaches than we have room to discuss are now under study.

Particles of Fire

IN THE 17TH CENTURY French philosopher René Descartes enumerated a theory to explain how people sense pain. In his view, a pinch, a whack or a poke essentially tugged on a neural rope that then rang a pain alarm bell in the brain. Imagine, for example, burning a foot. "Fast moving particles of fire," Descartes thought, would create a disturbance that "passes along the nerve filament until it reaches the brain."

Descartes was not too far off. Pain generally begins at the periphery: in the skin, an internal organ or any other site outside the central nervous system (CNS)—that is, outside the brain and spinal cord. Stubbing a toe or leaning against a hot stove activates neurons (nerve cells) called nociceptors that respond specifically to hurtful stimuli, such as extreme temperature or mechanical pressure, or to chemicals generated in response to injury or inflammation.

Nociceptors have two arms: a sensation-detecting branch that projects out to the periphery, where it innervates small patches of tissue, and a second branch that extends into the spinal cord [*see box on page 63*]. The neuron's cell body, which resides in a structure outside the spine, sits between the two. When specialized detector molecules on the peripheral branch encounter a noxious agent in the skin or an organ, they trigger an impulse that travels up the line, along the central branch and on to an area of the spinal cord known as the dorsal horn. There the nociceptor releases signaling molecules called neurotransmitters that activate neurons in the dorsal horn, prompting

them to transmit the alarm message up to the brain. Although nociceptors are often called pain-sensing neurons, they merely indicate the presence of potentially harmful stimuli; it is the brain that interprets the signal as painful and prompts us to say "ouch."

Not all pain is worrisome. For example, the acute kind that accompanies a minor tissue injury such as a sprain or abrasion is protective: it encourages an organism to avoid further injury. This kind tends to be temporary and to subside over time.

The pain that most troubles patients—and doctors—fails to disappear and is difficult to treat. In many cases, the problem arises because the injury or the inflammation that triggers the discomfort persists. The aches of arthritis result from ongoing inflammation, and the agony that can accompany invasive cancer stems to a large extent from tissue injury and inflammation.

In other cases, persistent pain is neuropathic, resulting from damage to nerve cells themselves. It can develop when neurons in the CNS sustain damage from multiple sclerosis, a stroke or spinal cord injury, for example. Or it can derive from injury to peripheral neurons. Amputees who endure aching in a limb that is no longer there (phantom limb pain) and people who feel burning in their skin for years after a herpes infection has subsided (postherpetic neuralgia) are all suffering from neuropathic pain. When this kind of hurt continues, it is not symptomatic of some ongoing injury or another disease; it is itself a disease of the nervous system and requires the attention of a pain specialist.

Pain without End

A MAJOR COMMON denominator in those who suffer from hard-to-manage pain is abnormal sensitivity to stimuli. This sensitivity can take the form of hyperalgesia (an excessive reaction to typically painful inputs) or allodynia (pain in response to normally innocuous inputs). In those affected by allodynia, even the gentle pressure of clothing against one's skin or bending a joint can become unbearable.

Biologists now understand that such heightened sensitivity—or sensitization—stems from molecular or structural changes in nerve cells. In the periphery, for instance, molecules that promote inflammation may cause the nociceptors that detect noxious stimuli to become overly reactive to those inputs. Inflammatory molecules can even cause nociceptors to begin generating signals in the absence of any environmental input.

Sensitization can also result from CNS changes that lead to hyperactivity of pain-transmission pathways. The changes, which may persist for a long time, can include display of increased numbers of the receptors that respond to the neurotransmitters released by nociceptors and might even include rewiring of connections and a loss of nerve cells that normally inhibit pain signaling. When the CNS is involved, the condition is termed central sensitization.

Regardless of which specific processes are at fault, ongoing pain, it turns out, can lead to sensitization and thus exacerbate and prolong discomfort. Many researchers, therefore, have amelioration of hyperalgesia and allodynia foremost in their minds as they hunt for new analgesics. Meanwhile patients need to realize that persistent pain should not be borne stoically; it requires aggressive treatment to prevent further sensitization.

Start at the Beginning

IN THE SEARCH for new analgesics, much effort has been directed toward the place where hurtful signals typically originate: the periphery. Certain of the specialized molecules that nociceptors use to detect noxious stimuli rarely occur elsewhere in the body. Blocking these molecules would presumably shut off pain signaling without disrupting other physiological processes and, thus, without causing unwanted side effects.

Today's most popular remedies—aspirin and other NSAIDs—largely work their magic in the periphery. When a tissue is injured, a variety of cells in the area pump out chemicals called prostaglandins, which act on the pain-sensing branches of nociceptors, lowering their activation threshold. Aspirin and NSAIDs inhibit the activity of a family of enzymes (cyclooxygenases) that cells use to generate the pain-inducing prostaglandins. These over-the-counter compounds relieve everyday aches and pains. But they also inhibit prostaglandin production elsewhere in the body, often causing such side effects as stomach pain, diarrhea and ulcers. These problems can prevent the drugs' long-term use and limit the doses that can be given.

To reduce the gastrointestinal consequences, pharmaceutical companies developed a family of drugs that target the

Patients need to realize that persistent pain should not be borne stoically; it needs AGGRESSIVE TREATMENT.

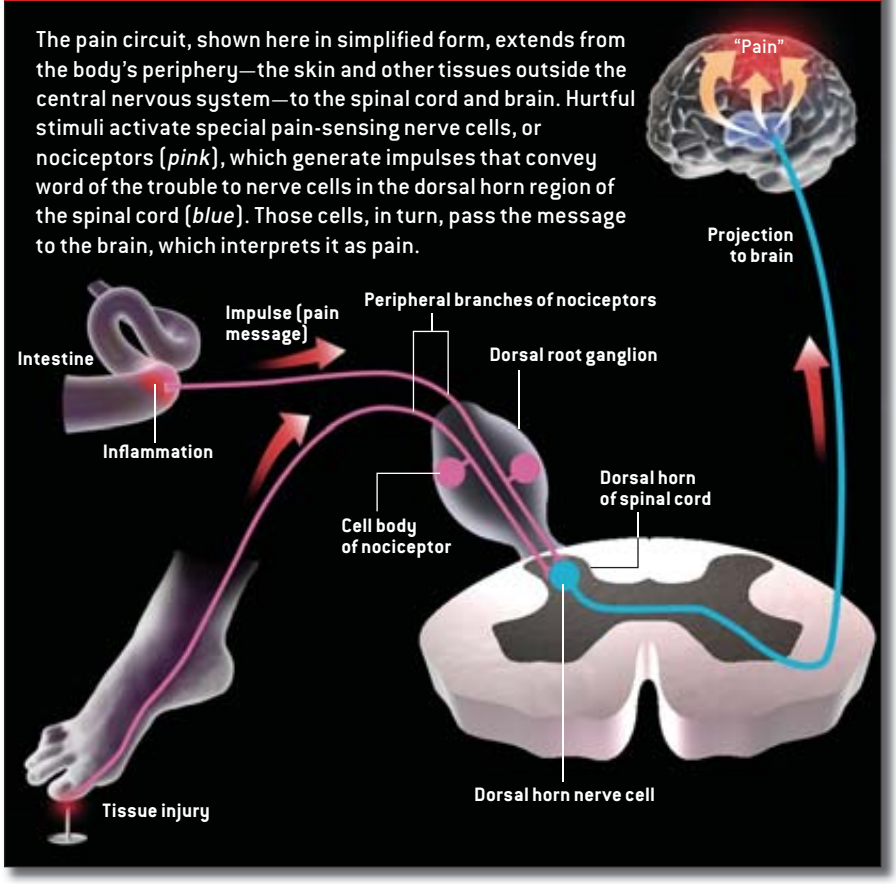
Overview/*Easing Pain*

- Specialized nerve cells called nociceptors respond to noxious stimuli. These cells transmit a message to nerve cells in the spinal cord, which then carry the signal to the brain.
- Nociceptors and other nerve cells in pain circuits possess specialized molecules for detecting pain-causing stimuli. These molecules can serve as targets for the development of medicines that may alleviate pain with fewer side effects than are caused by existing drugs.

JANA BRENNING (*preceding pages*)

FEELING THE PAIN

The pain circuit, shown here in simplified form, extends from the body's periphery—the skin and other tissues outside the central nervous system—to the spinal cord and brain. Hurtful stimuli activate special pain-sensing nerve cells, or nociceptors (*pink*), which generate impulses that convey word of the trouble to nerve cells in the dorsal horn region of the spinal cord (*blue*). Those cells, in turn, pass the message to the brain, which interprets it as pain.



enzyme cyclooxygenase-2 (COX-2). Because COX-2 does not normally operate in the stomach or intestinal tract, blocking its activity should not cause the same disruptions as traditional NSAIDs do. Whether they are in fact gentler on the stomach remains to be established. In the meantime, the drugs have problems of their own. Rofecoxib (Vioxx), a COX-2 inhibitor that had been prescribed for relief of arthritis pain, was removed from the market when it was found to boost the risk of heart attack and stroke. Other COX-2 inhibitors are also being scrutinized for ill effects.

Send in the Salsa

DISCOVERY OF TARGETS that reside almost exclusively on nociceptors provided an opportunity to develop drugs that act selectively to relieve pain. A particularly appealing one is the capsaicin receptor. This ion channel, present in the membrane of many nociceptors, responds not only to capsaicin, the pungent ingredient in chili peppers, but also to distressful heat and to protons (the hydrogen ions that make substances acidic); protons are unusually abundant in inflamed tissue. In the presence of these chemicals or of temperatures above 43 degrees Celsius, the channel allows sodium and calcium ions to flood into the nociceptor, stimulating it to generate a signal that translates into the burning sensation induced by heat, inflammation or spicy food.

Substances that inhibit the capsaicin receptors should therefore dampen inflammatory pain. Indeed, in laboratory animals, such "antagonists" have been able to relieve the very severe pain caused by the acidic environment around tumors that have metastasized to and damaged bone tissue. Today many pharmaceutical companies are competing to develop capsaicin receptor antagonists.

The possibilities for manipulating the receptor do not end there. Ironically, in some instances, purposely stimulating capsaicin receptors can alleviate pain. Topical creams containing capsaicin are being prescribed to relieve the itching, prickling and stinging sensations that can accompany postoperative wound healing or nerve impairments stemming from HIV infection, bouts of herpes and diabetes. Exactly how the ointments work is unclear, although small doses over time might ultimately make the receptor less responsive to the usual stimuli or might cause depletion of the neurotransmitters emitted by nociceptors.

Block Other Channels

A DIFFERENT KIND of molecule found on the peripheral terminals of nociceptors is also attracting interest as a drug target. All neurons possess sodium channels that open in response to changes in the voltage across the nerve cell membrane, generating the impulses that relay messages from one neuron to the next. Local anesthetics that temporarily inactivate such voltage-gated sodium channels currently treat a variety of different pains, particularly those arising from a trip to the dentist. The problem, though, is that those anesthetics have to be applied at the site of the discomfort: disabling sodium channels throughout the nervous system could be fatal.

Pain-sensing neurons, however, possess a subclass of voltage-gated sodium channels, known as the TTX-resistant type, that do not occur in the CNS. Investigators therefore hope that drugs able to block this subclass could be administered systemically (throughout the body) without ill effects. Moreover, studies suggest that such agents could well dampen inappropriate hyperactivity by injured peripheral nerves and thus might relieve some neuropathic pain. Unfortunately, the pharmaceutical industry has so far been unable to successfully develop selective inhibitors for such channels, in part because they closely resemble TTX-sensitive sodium channels, which appear widely throughout the nervous system.

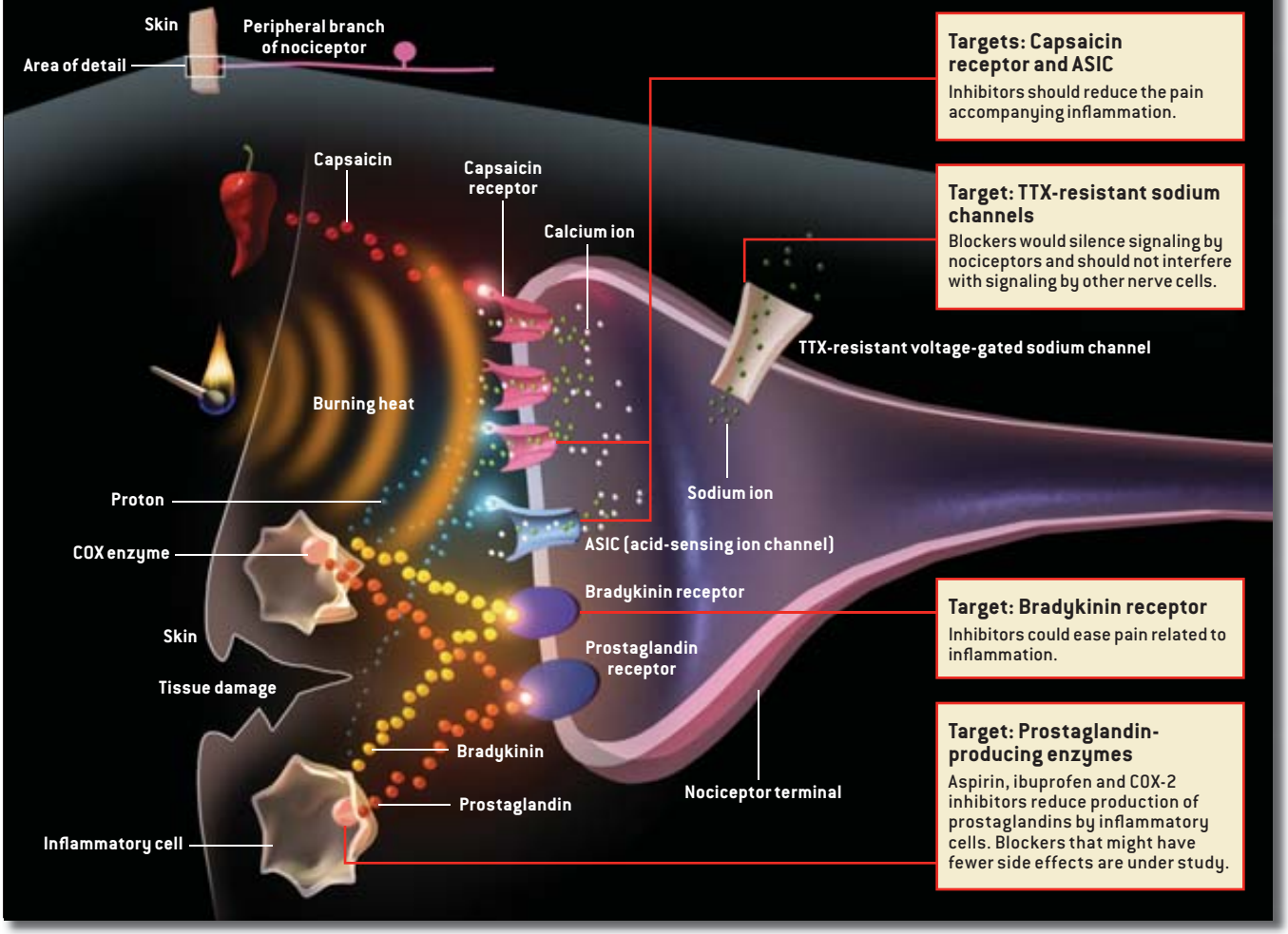
The channels could perhaps be selectively removed, however, with a new technique called RNA interference. The method relies on introducing into an organism tiny molecules known as small interfering RNAs (siRNAs). These

AMADEO BACHAR

DRUG TARGETS IN THE PERIPHERY

The small branches of nociceptors that innervate the skin and internal organs possess specialized molecules—receptors—that detect noxious stimuli. Those stimuli can include the chemical capsaicin in hot peppers, intense heat or substances released by the inflammatory cells that respond to an injury.

Recognition by the detectors causes some of them to usher sodium and calcium ions into the cells. This inflow or the activation of other receptors induces nociceptors to emit pain signals and can make the cells responsive to innocuous stimuli. Signal propagation also requires activation by voltage-gated ion channels. Inhibiting the activity of the detector molecules or of the voltage-gated channels should be therapeutic, in the ways indicated in the boxes. For clarity, only some of the drug targets being explored are noted.



AMADEO BACHAR

siRNAs prevent the production of an unwanted protein by inducing the degradation of the molecules (messenger RNAs) that direct the protein's synthesis. The technique is being studied in humans for certain retinal conditions, but turning RNA interference into a pharmacological intervention for pain will be challenging. As is true of gene therapy, a virus will most likely be needed to deliver siRNA, and this aspect has raised safety concerns. Time will tell whether the approach will be practical as a pain therapy, but it remains an exciting possibility.

Suppose drug companies do develop a so-called magic bullet analgesic: a compound that specifically and effectively eliminates the activity of one of the pain-transducing molecules on nociceptors. Would this intervention provide relief from intractable pain? Maybe not, because closing off a single entrance to the pain pathway might not be enough.

Imagine, for example, a pharmaceutical that knocks out the receptor for bradykinin—a small protein, or peptide, that is produced during inflammation in the periphery. Bradykinin powerfully stimulates nociceptors, and an antagonist that blocks its receptors would certainly prevent those receptors from activating nociceptors. But it would not stop the neurons from recognizing and responding to other pain-inducing molecules generated by injury or inflammation—protons, prostaglandins, and a protein called nerve growth factor, for example. Similarly, hobbling only the capsaicin receptors might not mitigate all proton-mediated pain, because under certain circumstances, protons activate a separate population of detectors, called ASICs (acid-sensing ion channels), on nociceptors.

Focus on the Cord

ONE WAY AROUND this redundancy problem would be to administer a cocktail of inhibitory molecules that targets multiple pain mechanisms. Another approach, though, would target molecules that act more centrally, blocking the ability of all nociceptors—no matter what stimuli initially activated them—to pass their pain signals to spinal cord neurons.

Morphine and other opiates, which bind to opioid receptors on the nociceptor endings that reach into the spinal cord, employ this latter tactic. In activating these receptors, opiates prevent neurotransmitter release, thus blocking the transmission of the pain message to spinal cord neurons. They also render dorsal horn neurons less able to respond to pain signals. Because these drugs act in the spinal cord, they should in theory be able to treat all types of pain, although they tend to work best against those related to inflammation.

Unfortunately, opioid receptors are present on neurons throughout the body, including in the brain and the gastrointestinal system. This ubiquity explains why morphine and its cousins can generate a broad set of undesirable side effects, including severe constipation and respiratory shutdown. These problems can restrict the amount of drug a patient can take safely or that a doctor will prescribe. And many physicians are reluctant to prescribe opiates for fear patients will become addicted. Addiction, however, is not common in those who take opiates only for pain. In part to avoid some of the undesirable effects, opiates are often delivered directly into the fluid-filled space around the spinal cord (intrathecally). The medicines may also be administered by injection (for postoperative pain) or via an indwelling pump (for chronic pain).

Alternatives to opiates are available as well. Medicines that interfere with calcium channels can prevent the release of neurotransmitters from nociceptor endings in the spinal cord. Gabapentin (Neurontin), an anticonvulsant, is believed to relieve some forms of pain by interacting with a specific subunit of certain calcium channels. And a relatively new drug called ziconotide (Prialt)—derived from the venom of a Pacific Ocean cone snail—inhibits a different variety of calcium channel known as the N-type.

Like opioid receptors, N-type calcium channels occur throughout the nervous system. If ziconotide were delivered systemically, blood pressure would decline precipitously. So the compound is administered intrathecally. Although the toxin blocks pain, its action within the CNS can still generate unpleasant side effects, including dizziness, nausea, headache and confusion. Ziconotide, therefore, is given mostly to patients with late-stage cancer who cannot get relief another way [see "A Toxin against Pain," by Gary Stix; SCIENTIFIC AMERICAN, April 2005].

Recently drugs that act on cannabinoid receptors—the ones that mediate marijuana's effects—have been advancing through clinical trials. These agents seem to ease pain in several ways, including by interfering with signal transmission between nociceptors and their target cells and by reducing the activity of inflammatory cells.

Batten Down the Hatches

SOME INVESTIGATORS are concentrating on preventing spinal neurons from responding to neurotransmitters released by nociceptors—particularly to the amino acid glutamate, the primary carrier of the pain message. Glutamate activates various receptors in the dorsal horn of the spinal cord. Of these, the NMDA class participates in central sensitization, which makes it a logical target for new analgesics.

Every neuron in the body possesses some type of NMDA receptor. Consequently, inhibiting all types at once would elicit catastrophic effects, including memory loss, seizures and paralysis. To avoid such reactions, researchers are attempting to hobble the receptor by acting on versions found mostly in the dorsal horn. Compounds that bind to a form containing what is called the NR2B subunit have yielded encouraging results in animal studies. For example, mice that had an NR2B inhibitor delivered directly into the spinal fluid were less sensitive to pain than were untreated animals. The drug also reversed allodynia in mice that had a peripheral nerve injury.

A number of nociceptors also release peptide neurotransmitters, such as substance P and calcitonin gene–related peptide (CGRP). These peptides activate pain-transmission neurons in the spinal cord by acting on discrete receptors, so drugs that bar interaction with those receptors would be expected to be helpful. Regrettably, selective blockade of the receptor used by substance P—the neurokinin-1, or NK-1, receptor—has failed in clinical trials for pain, perhaps because blocking that receptor by itself is insufficient. Whether quieting CGRP activity in the spinal cord will shut down pain is unknown, although the pharmaceutical industry is developing antago-

Investigators might be able to develop better COGNITIVE THERAPIES for altering pain perception.

THE AUTHORS

ALLAN I. BASBAUM and *DAVID JULIUS* often collaborate on studying the cellular and molecular mechanisms that underlie pain. Basbaum, who received his Ph.D. in neuroscience from the University of Pennsylvania, is professor and chair of the department of anatomy at the University of California, San Francisco. Julius, who received his Ph.D. in biochemistry from U.C., Berkeley, is professor of cellular and molecular pharmacology at U.C.S.F. Both authors consult for companies pursuing therapies for pain—among them GlaxoSmithKline, Hydra Biosciences, NeurogesX and Rinat Neuroscience.

nists that aim to ease the agony of migraines by interfering with the release of CGRP onto blood vessels on the surface of the brain.

Kill the Messenger?

IF ALL ATTEMPTS to modulate pain signaling fail, one can consider killing the messenger. Cutting nociceptive nerves, though, generally backfires because, as we have noted, nerve injury can promote the onset of even more stubborn, persistent pain. Severing pathways in the spinal cord that convey information to the brain (cordotomy) was once common but now is reserved for terminal cancer patients unresponsive to all pain treatments. The problem with this last procedure is that the surgeon cannot selectively cut the "pain" pathways.

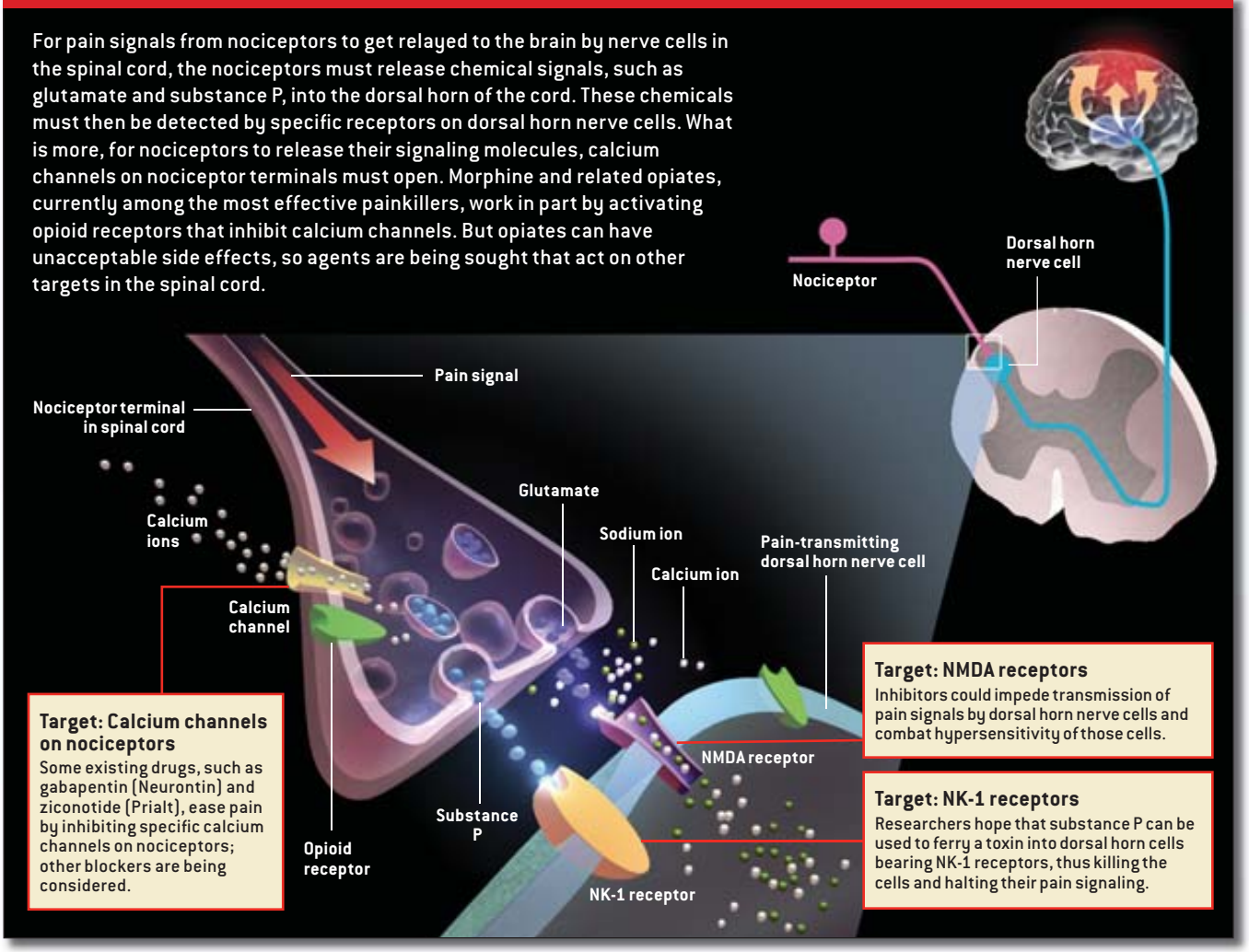
A possible solution, now drawing considerable attention because of its success in animals, is a molecular therapy that takes out a subset of the spinal cord neurons receiving input from nociceptors. This cell-killing therapy couples saporin, a toxin, to substance P. The substance P in the conjugate binds to NK-1 receptors, leading to internalization of the whole compound, after which the saporin is freed to kill the neuron. Because the conjugate can enter only cells having an NK-1 receptor, researchers hope that side effects will be limited.

Ablation of neurons in the spinal cord, however, should be considered a method of last resort: neurons in the CNS do not grow back, so the resulting changes—for better or worse—will be permanent. The same permanence does not hold in the peripheral nervous system, where cut fibers can regenerate. Ideally, therapies that trim back the signal-detecting branches of nociceptors—such as high doses of capsaicin—would halt pain but allow the branches to grow back eventually, restoring normal pain detection to the patch of tissue innervated before.

Targeting neurons may not be the sole way to overcome pain. Studies indicate that glia, the cells that nurture neurons in the CNS, swing into action in response to damage to peripheral nerves. The glia migrate to the region of the dorsal horn associated with the injured nerves. Then the glia discharge a bevy of chemicals that prod nociceptor terminals to release neurotransmitters in the cord, thus perpetuating a pain signal. Some of these substances, including growth factors and molecules known as cytokines, also make dorsal horn neurons overly excitable, and drugs blocking that hyperac-

DRUG TARGETS IN THE SPINAL CORD

For pain signals from nociceptors to get relayed to the brain by nerve cells in the spinal cord, the nociceptors must release chemical signals, such as glutamate and substance P, into the dorsal horn of the cord. These chemicals must then be detected by specific receptors on dorsal horn nerve cells. What is more, for nociceptors to release their signaling molecules, calcium channels on nociceptor terminals must open. Morphine and related opiates, currently among the most effective painkillers, work in part by activating opioid receptors that inhibit calcium channels. But opiates can have unacceptable side effects, so agents are being sought that act on other targets in the spinal cord.



AMADEO BACHAR

PAIN FIGHTERS IN DEVELOPMENT

The table below highlights some compounds with novel mechanisms of action that are being tested in people; it thus omits new variants of well-established pharmaceutical classes, such as opiates and COX inhibitors. Clinical trials advance in stages, with phase I focusing on safety, phase II consisting of early trials of efficacy and phase III involving much larger tests.

COMPOUND (DEVELOPER)	MECHANISM OF ACTION	TRIAL STAGE	COMPANIES STUDYING RELATED AGENT
AMG-517 (Amgen)	Blocks the capsaicin receptor	Phase I	GlaxoSmithKline; Neurogen
EVT-101 (Evotec)	Blocks NMDA receptors bearing the NR2B subunit	Phase I	Roche; Merck & Co.
Icatibant (Sanofi-Aventis)	Blocks a bradykinin receptor	Phase II	Merck & Co.
NGX-4010 (NeurogesX)	Overstimulates the capsaicin receptor	Phase III	
NMED-160 (Neuromed Pharmaceuticals)	Blocks N-type calcium channels	Phase II	
Ralfinamide (Newron Pharmaceuticals)	Blocks sodium channels	Phase II	
RN624 (Rinat Neuroscience)	Stops nerve growth factor from stimulating nociceptors	Phase II	Amgen
SAB-378 (Novartis)	Activates a cannabinoid receptor	Phase II	GW Pharmaceuticals; GlaxoSmithKline

SOURCE: FRANZ F. HEFTI *Rinat Neuroscience Corporation*

tivity should help undercut excessive sensitivity. Various groups are working to identify—and find ways to inhibit—the molecules that recruit and activate glia when nerves are damaged.

Interestingly, prostaglandins are among the key substances released from activated glia in the spinal cord. There they enhance pain by blocking receptors for glycine on dorsal horn neurons. Glycine, an inhibitory neurotransmitter, normally quiets these neurons. NSAIDs may therefore work not only by interfering with the production of prostaglandins in the periphery (the familiar way) but also by inhibiting COX enzymes in glia. In that case, direct delivery of COX inhibitors into the spinal fluid might minimize the side effects caused by systemic delivery. A pharmaceutical that enhanced glycine receptor activity could also help tamp the transmission of pain messages to the brain.

A Question of Perception

IN THIS ARTICLE we have discussed a subset of the experimental approaches to treating pain, all of which have shown promise in animal studies. Those evoking the greatest excitement leave normal sensation intact while diminishing the heightened sensitization characteristic of the difficult-to-treat inflammatory and neuropathic pains and have an acceptable side-effects profile. But will these therapies help patients? And will they work on all types of pain? These questions remain unanswered.

One approach that deserves further exploration is the use of behavioral, nondrug therapies for intractable pain—particularly that associated with conditions such as fibromyalgia and irritable bowel syndrome, for which no one has conclusively established an organic cause. Roughly a decade ago researchers at McGill University demonstrated that hypnosis could alter brain activity along with a person's perception of pain. The scientists hypnotized volunteers and suggested to them that the hot water bath in which they had immersed their hands was either more unpleasant or less unpleasant than it really was.

Using positron-emission tomographic scanning to monitor brain activity, the investigators found that the somatosensory cortex, which responds to the magnitude of the physical stimulus, was equally active in both situations. But a second brain region, the cingulate cortex, was more active when subjects believed that the stimulus was more unpleasant, suggesting that hypnosis changed the way these individuals perceived sensations. By learning more about how the brain modulates the pain experience, investigators might be able to develop better cognitive therapies for moderating pain perception.

Poet Emily Dickinson often contemplated pain. In one work, she noted:

> Pain has an element of blank;
> It cannot recollect
> When it began, or if there were
> A day when it was not.
> It has no future but itself.

We can only hope continued research into the mechanisms of pain sensation will lead to safe, effective treatments that will alter pain's future, such that it reverts to a time when it was not. SA

MORE TO EXPLORE

The Perception of Pain. A. I. Basbaum and T. Jessel in *Principles of Neural Science.* Edited by Eric R. Kandel et al. McGraw-Hill, 2000.

Molecular Mechanisms of Nociception. David Julius and Allan I. Basbaum in *Nature,* Vol. 413, pages 203–210; September 13, 2001.

Immune and Glial Cell Factors as Pain Mediators and Modulators. S. B. McMahon, W. B. Cafferty and F. Marchand in *Experimental Neurology,* Vol. 192, No. 2, pages 444–462; 2005.

Pain Collection in *Nature Reviews Neuroscience,* July 2005. Available online at **www.nature.com/nrn/focus/pain**

Emerging Strategies for the Treatment of Neuropathic Pain. Edited by James N. Campbell et al. IASP Press, 2006.

"ONLY AS STRONG AS THE WEAKEST LINK" describes the construction of a bridge as well as that of a computer program. Like bridges, software programs are key components of the critical infrastructure of modern society—but researchers only recently have invented effective ways to pretest the soundness of software designs.

Dependable Software by Design

Computers fly our airliners and run most of the world's banking, communications, retail and manufacturing systems. Now powerful analysis tools will at last help software engineers ensure the reliability of their designs

By Daniel Jackson

GEORGE RETSECK

Almost all grave software problems can be traced to conceptual mistakes made before programming started.

Failed automated baggage system at Denver International Airport.

An architectural marvel when it opened 11 years ago, the new Denver International Airport's high-tech jewel was to be its automated baggage handler. It would autonomously route luggage around 26 miles of conveyors for rapid, seamless delivery to planes and passengers. But software problems dogged the system, delaying the airport's opening by 16 months and adding hundreds of millions of dollars in cost overruns. Despite years of tweaking, it never ran reliably. Last summer airport managers finally pulled the plug—reverting to traditional manually loaded baggage carts and tugs with human drivers. The mechanized handler's designer, BAE Automated Systems, was liquidated, and United Airlines, its principal user, slipped into bankruptcy, in part because of the mess.

The high price of poor software design is paid daily by millions of frustrated users. Other notorious cases include costly debacles at the U.S. Internal Revenue Service (a failed $4-billion modernization effort in 1997, followed by an equally troubled $8-billion updating project); the Federal Bureau of Investigation (a $170-million virtual case-file management system was scrapped in 2005); and the Federal Aviation Administration (a lingering and still unsuccessful attempt to renovate its aging air-traffic control system).

Such massive failures occur because crucial design flaws are discovered too late. Only after programmers began building the code—the instructions a computer uses to execute a program—do they discover the inadequacy of their designs. Sometimes a fatal inconsistency or omission is at fault, but more often the overall design is vague and poorly thought out. As the code grows with the addition of piecemeal fixes, a detailed design structure indeed emerges—but it is a design full of special cases and loopholes, without coherent principles. As in a building, when the software's foundation is unsound, the resulting structure is unstable.

Managers involved in high-profile software blowouts could claim in their defense that they followed standard industry practices, and unfortunately they would be right. Developers rarely articulate their designs precisely and analyze them to check that they embody the desired properties. But with computers now flying airplanes, driving trains and cars, and running most of the financial, communications, trading and production machinery of the world, society has an urgent need to improve software dependability.

Now a new generation of software design tools is emerging [*see box on page 74*]. Their analysis engines are similar in principle to tools that engineers increasingly use to check computer hardware designs. A developer models a software design using a high-level (summary) coding notation and then applies a tool that explores billions of possible executions of the system, looking for unusual conditions that would cause it to behave in an unexpected way. This process catches subtle flaws in the design before it is even coded, but more important, it results in a design that is precise, robust and thoroughly exercised. One example of such a tool is Alloy, which my research group and I constructed. Alloy (which is freely available on the Web) has proved useful in applications as varied as avionics software, telephony, cryptographic systems and the design of machines used in cancer therapy [*see box on page 73*].

Alloy and related design-checking tools build on a quarter of a century of existing research into ways to prove mathematically whether programs are correct. But rather than requiring proofs to be done by hand, they employ automated reasoning techniques that treat a software design problem as a giant puzzle to be solved. These analyzers operate on designs, not program code, so they

Overview/*Software Design Checkers*

- Despite the ever increasing importance of computer software in our daily lives, software engineers rarely analyze their designs to ensure reliability. That situation is starting to change with the recent development of software design checking tools such as Alloy.
- Alloy combines a language that eases the modeling of complex software designs with an analysis engine that checks extensively for conceptual and structural flaws in an automated fashion, treating designs as huge puzzles to be solved.
- In the relatively near future, tools similar to Alloy will greatly improve the dependability of software by basing program development on more robust and constructive design practices.

CHRIS CARTER *Denver International Airport*

ALLOY IN ACTION

Alloy helps software designers find and fix design flaws by providing both a language that clarifies a program's structure and an automated analyzer that searches the vast numbers of possible executions of a system for a "counterexample" that shows how it could fail to behave as desired. In the simplified example below, an engineer uses Alloy to evaluate the design of a file system—the software that organizes your computer files into folders and stores them on a disk. A crucial task for Alloy is to work out the effects various operations would have on the file structure. Here is how a designer might model and check the operation that moves a folder, or "directory," from one location in the file hierarchy to another.

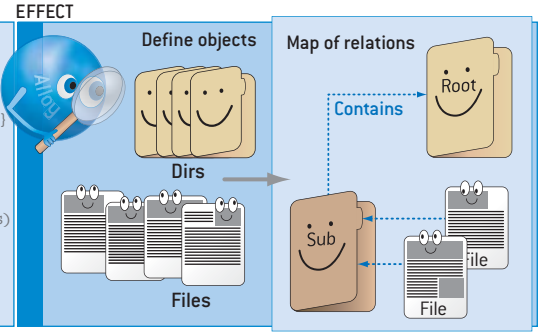
STEP 1: DEFINE THE OBJECTS

The designer identifies the system's objects—files, directories and the file system as a whole—and their relations with one another. The Alloy model says the file system (FS) has three components: "files" (its set of files), "dirs" (its set of directories) and "contains" (a mapping that gives, for each directory, the set of files and directories it contains).

ALLOY CODE

```
module filesystem

abstract sig Object {}
sig File, Dir extends Object {}

sig FS {
  dirs: set Dir,
  files: set File,
  contains: dirs -> (dirs + files)
  }
```

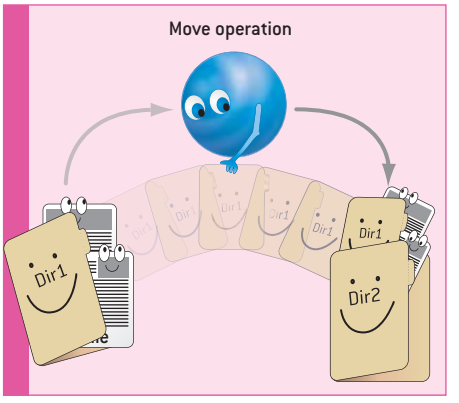
EFFECT



STEP 2: MODEL THE OPERATION

Next, the designer models the move ("move_dir") of the file system before ("fs") to a file system after ("fs'"). The operation involves two directories: "d," the directory being moved, and "to," the place it is being moved to—its new parent. Three constraints follow, which describe the intended effect, on three separate lines: First, both the moved object and its new location are directories of the file system. Second comes the essence of the operation: it says that the new containment mapping is the old one, with every mapping from a directory to "d" removed, and the mapping from "to" to "d" added. The third line says that nothing else changes.

```
pred move_dir (fs, fs': FS, d, to: Dir) {
d + to in fs.dirs
fs'.contains = fs.contains - Dir->d + to->d
fs'.files = fs.files and fs'.dirs = fs.dirs
  }
```



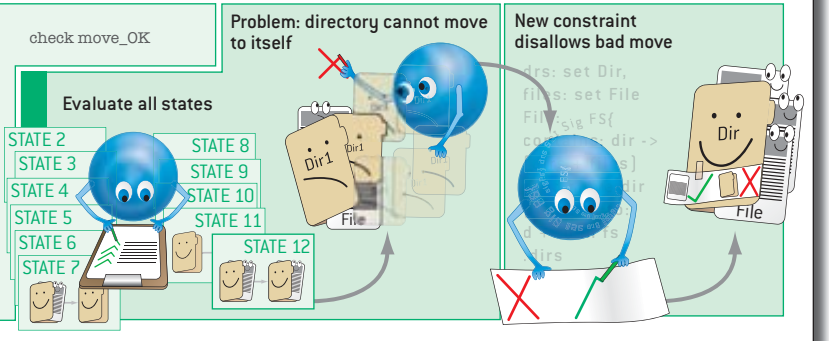
STEP 3: SPECIFY REQUIREMENTS

The designer then formulates a crucial requirement: every file and directory should be "reachable" (have a pathway) from some root. This is recorded in the Alloy model as an "assertion" (called "move_OK"), which says that executing the move operation does not make a file or directory unreachable from a root.

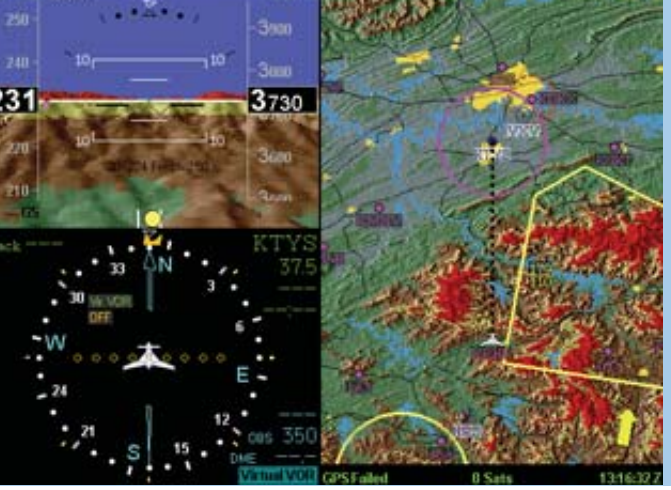
```
pred reachable (fs: FS) {
  some root: fs.dirs | fs.(dirs+files) in root.*(fs.contains)
  }

assert move_OK {
  all fs, fs': FS, d, to: Dir |
    reachable (fs) and move_dir (fs, fs', d, to) implies reachable (fs')
  }
```

STEP 4: FIND AND FIX THE FLAW

Alloy executes "check move_OK" by generating all possible states of the system (up to a certain size) and checking the assertion for each—thus simulating possible moves as they might occur when the software is run. Alloy finds a counterexample to the assertion—a directory that can be moved to itself. The action would disconnect the directory from a root, making it unreachable. As a remedy, a designer could add a new constraint disallowing a directory to move to itself or any of its descendants.



LUCY READING-IKKANDA



The idea is to simulate every state that the software can take to determine that none leads to a failure.

Alloy helped to make an avionics system hacker-proof.

cannot guarantee that a program will not crash. But they potentially offer software engineers the first practical tools to ensure that designs are robust and free from conceptual flaws and thus provide a firm foundation on which to build reliable software systems.

Evaluating Designs

BAD SOFTWARE is not a new problem. Warnings of a software crisis go back to the 1960s and have only intensified as computers have been woven into the fabric of society [see "Software's Chronic Crisis," by W. Wayt Gibbs; SCIENTIFIC AMERICAN, September 1994].

Today most software typically is debugged and refined by testing. Human engineers run the program using a wide range of starting conditions (or inputs) to see if it operates as expected. Although the practice catches a raft of small flaws, it often overlooks faults in the basic design of the software. In some sense, these test procedures miss the (diseased) forest for the (rotting) trees.

What is worse, bugs "fixed" during the testing process often exacerbate design problems. As programmers debug the code and insert new features, the software invariably grows barnacles of complexity, creating more opportunities for errors and inefficient operation. This situation is reminiscent of the (incorrect) Ptolemaic theory of planetary motion first developed by the ancient Greeks. In the Middle Ages, as observations showed the predictions to be inaccurate, astronomers adjusted Ptolemy's system, which relied on epicycles. When that proved insufficient, they resorted to adding epicycles to the epicycles. Further fine-tuning over the centuries never solved the problem, because the initial concept was fatally flawed.

Similarly, bad software tends to get more and more complicated and less and less reliable, however much time and money are poured into improving it. It is well known that serious problems with software systems rarely arise from programming errors; almost all grave difficulties can be traced back to conceptual mistakes made before programming even started. In contrast, a small amount of modeling and analysis during the initial determination of requirements, specifications, or program design costs only a tiny fraction of the price tag of checking all the code but provides a large part of the benefit gained from an exhaustive analysis. Focusing on design early saves costly headaches down the road.

Design tools for software have been slow in coming because software does not obey physical laws. Because computer programs are in essence mathematical objects whose values are constructed from bits, software programs are discrete (particlelike) rather than continuous. A mechanical engineer can stress a component with a large force and assume that if it survives it will not fail when subjected to a slightly smaller force. When an object is subject to the (mostly continuous) principles of the physical world, a small change in one quantity generally produces a small change in another. Unfortunately, no such generalities apply to software: one cannot extrapolate between test cases. If one chunk of software works, that fact says nothing about the operations of a similar chunk of code; they are discrete and separate.

In the early days of computer science, researchers hoped that programmers might prove their codings were correct in the same way that mathematicians prove their theorems. With no way to automate the many steps involved, however, a human expert had to do much of the work. These so-called heavy-duty formal methods were impractical except for relatively modest but especially critical pieces of software, such as an algorithm for controlling railroad intersections.

More recently, researchers have adopted a very different approach, one that harnesses the power of today's faster processors to test every possible scenario. This method, known as model checking, is now used extensively to verify integrated-circuit designs. The idea is to simulate every possible sequence of states (the conditions of the system at specific times) that might arise in practice and to determine that none leads to a failure. For a microchip design, the number of states to evaluate is often huge: 10^{100} or more. The challenge is far more stringent for software. But clever encoding techniques (by which large sets

THE AUTHOR

DANIEL JACKSON leads the Computer Science and Artificial Intelligence Laboratory's Software Design Group at the Massachusetts Institute of Technology. His main research interest is software engineering, with a focus on software design, specification and analysis, particularly of critical systems. Jackson received an M.A. from the University of Oxford in physics, and his S.M. and Ph.D. from M.I.T. in computer science. Before his professorship at M.I.T., he taught at Carnegie Mellon University. An avid photographer, Jackson recently exhibited his work at the Newton Free Library outside of Boston.

GREG RICHTER *Blue Mountain Avionics*

of software states can be represented very compactly) make it possible to check every state by considering these large sets simultaneously.

Model checking alone regrettably cannot handle states with complex structures, which is characteristic of most software designs. My research colleagues and I have developed an approach that shares the same spirit yet employs a different mechanism. Like model checking, it considers all possible scenarios (although in truth, some bounds need to be introduced to keep the problem finite, because software is not restricted by the physical limitations imposed by hardware). Unlike model checking, however, our technique does not examine scenarios in their entirety, one at a time. Instead it searches for a bad scenario—one that results in failure—by filling in each state in an automated fashion, one bit at a time, in no particular order.

The process is in some sense comparable to a robotic arm fitting each piece of a jigsaw puzzle into place one by one until the completed image finally emerges. If that image corresponds to a bad scenario, Alloy would have done its job. Alloy thus treats design analysis as if it were a puzzle to be solved. Some other recently developed software model checkers work this way as well.

The Solution Is a Puzzle

TO UNDERSTAND HOW Alloy solves software design puzzles, it helps to consider an old riddle: A farmer goes to market where he buys a fox, a goose and a bag of corn. On his way home, he has to carry his goods across a river by boat. The skiff will hold only the man and one purchase at a time, however. Herein lies a problem: if left unsupervised, the fox would eat the goose and the goose would eat the corn. So how does the farmer get all of his goods to the far bank intact?

This variety of puzzle involves finding scenarios that satisfy a collection of constraints. Mentally we do this task by imagining a series of steps: The farmer transports the goose first; on the next trip, he takes the fox, whereupon he brings back the goose and then, leaving it behind, crosses with the corn; he then returns to fetch the goose. By checking whether each step satisfies the constraints, we ensure that each item remains safe.

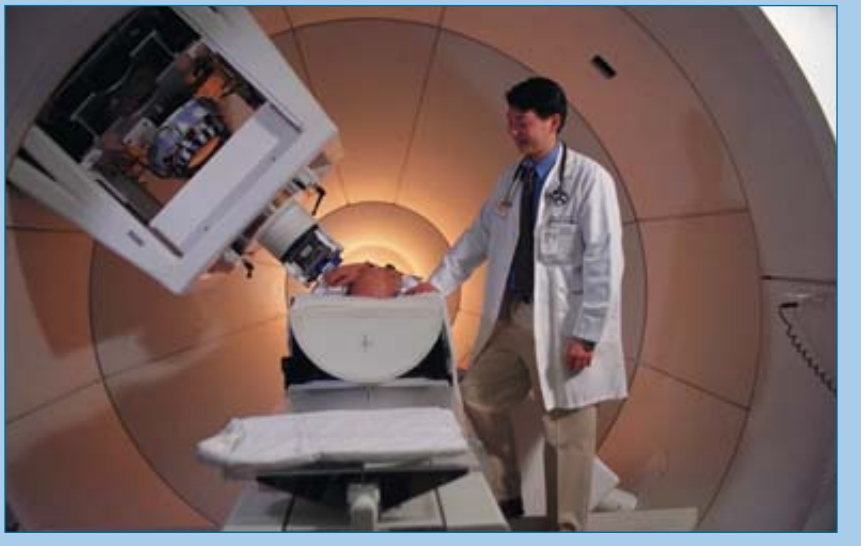
A successful software design imposes a similar, though much more complicated, array of rules. To be useful, a design-checking tool must be able to find counterexamples: solutions to the puzzle that meet all the "good" constraints (and thus could occur when the program is run) and an additional "bad" constraint (and thus yield an unacceptable outcome). If any such counterexamples turn up, they will reveal flaws in the design. So whereas the puzzle solver is happy to find a solution to the "farmer's dilemma," a solution to a software

Debugging Cancer Therapy Machines

Modern medical devices rely on software for almost every aspect of their operation. In a machine used for cancer therapy, even the "emergency stop" button is not an actual electrical switch but a software program: hitting it causes about 15,000 lines of code to execute and shut the system down—unless, of course, there is a bug or design flaw in the software. That is where Alloy comes in—it analyzes programs to find the design problems.

Working with the developers of a cancer-therapy system, for example, we have used Alloy to explore the design of some of its features. In one case, we took a design for a new scheduling system that determines the treatment room to which the beam is sent. We set Alloy to look for scenarios in which interactions between the operator in the main control room and the therapists in the treatment rooms would produce unexpected results. Alloy found various scenarios that had not been anticipated originally.

In another case, we applied Alloy to the design of an elaborate protocol for positioning the patient under the proton beam, which turned out to have a subtle and unexpected consequence: the angle of the gantry crept around over time, even when it was not being intentionally adjusted. With a small Alloy model we showed how, by choosing the right abstractions, this problem could be reduced to the same, rather simple problem as that for designing a car accessory system that remembers driver-seat positions. In fact, the therapy system has many safeguards and the gantry movement was not a dangerous problem. But if the correct abstractions had been used from the start, the design would have been much simpler and operating the software considerably easier. —*D.J.*



CORRECT POSITION of a patient—controlled by software—is critical to control radiation dosage in a cancer therapy machine. Alloy helped to improve the software design for a similar machine.

COURTESY OF LOMA LINDA UNIVERSITY MEDICAL CENTER

design puzzle is bad news: it means that an undesirable scenario exists and the design is defective. In practice, the counterexample might not itself lead to any problems. It may instead reveal a discrepancy in how the designer originally characterized the unacceptable outcomes. Either way something needs to be fixed—the design or the designer's expectations.

The great difficulty in searching for counterexamples is that the number of potential scenarios in a software design of even moderate complexity is typically vast, but only a tiny proportion correspond to counterexamples. Imagine trying to plan who sits next to whom at a wedding reception. If all attendees get along, the solution is trivial. Throw in a few ex-spouses who require separation, and the problem gets trickier. Now consider the seating chart for Romeo and Juliet's reception. If there are 20 seats and any of 10 guests can sit in each, that makes 10^{20} possible combinations. Even checking a billion scenarios per second, a computer would take 3,000 years to explore them all.

In the 1980s, researchers identified problems of this form as a special class of problems that, in the worst case, can be solved only by enumerating all possible scenarios. But in the past decade, with new search strategies and algorithms and by building on ever increasing computational power, researchers have developed tools called SAT (*sat*isfiability) solvers that can handle these problems fairly easily. Many are now freely available and can often solve problems with millions of constraints.

Importance of Abstraction

AS ITS NAME SUGGESTS, Alloy melds two elements that help make software designs more robust. One is a new language that helps to elucidate the structure and behavior of the software design. The other is an automated analyzer (which incorporates a SAT solver) to hash through a multitude of possible scenarios.

The first step in applying Alloy is to create a model of the design: not the rough sketch or flowchart typical in software engineering but a precise model that spells out the "moving parts" and specific behaviors, both desired and undesired, of the system and its components. A software engineer first writes down definitions of the various kinds of objects in the design, then groups those objects into mathematical sets: collections of things that are alike in their structure and behavior (for example, the set of all Capulets) and linked by mathematical relations (such as the relation that associates guests sitting next to one another).

Next come facts that constrain these sets and relations. In a software design, the facts include the mechanism of the software system and assumptions about other components (say, statements about how human users are expected to behave). Some of these facts are simple assumptions—for example, that nobody is both a Capulet and a Montague and that every guest sits next to exactly two other guests. Some of them reflect the design itself: in our seating planner, for instance, the rule that each table, with the exception of the top table, is assigned either to one family or the other.

Finally, there are assertions, which are constraints that are expected to follow from the facts. In our example, with

Tools for Checking Software Designs

Computer scientists have developed a new generation of software design checking tools (in addition to Alloy) that programmers can use to analyze and test their codings for structural and conceptual inconsistencies that could lead to system failure. In general, these commercial and open-source design-evaluation tools are based on specialized high-level languages (notations that summarize blocks of code) that researchers have developed to ease the specification, modeling and simulation of different types of software schemes.

Such tools incorporate automated analysis engines that explore the huge number of potential executions of systems for subtle design flaws that would cause them to behave in undesirable ways (an instance of which is called a counterexample). These software design tools often include facilities that can help designers visualize counterexamples or relations between blocks of code.

LANGUAGE	TOOL	SOURCE	WEB SITE
B	B-Toolkit	B-Core	www.b-core.com
	Atelier-B	Steria	www.atelierb.societe.com
	Pro-B	University of Southampton	www.ecs.soton.ac.uk/~mal/systems/prob.html
CSP	FDR	Formal Systems Europe	www.fsel.com
FSP	LTSA	Imperial College London	www.doc.ic.ac.uk/~jnm/book/ltsa/LTSA.html
Lotos	CADP	INRIA Research Institute	www.inrialpes.fr/vasy/cadp/
OCL	USE	University of Bremen	www.db.informatik.uni-bremen.de/projects/USE/
PROMELA	Spin	Bell Laboratories	spinroot.com/
Statecharts	Statemate	I-Logix	www.ilogix.com
VDM	VDMTools	CSK Corp.	www.csk.com/support_e/vdm/
			www.vdmbook.com/tools.php
Z	Jaza	University of Waikato	www.cs.waikato.ac.nz/~marku/jaza/
Zing	Zing	Microsoft Research	research.microsoft.com/zing/

Alloy has uncovered serious deficiencies in published software designs.

Alloy checked a software program that finds printers on wireless networks.

the exception of Romeo and Juliet, no Capulet should be seated next to a Montague. The assertions say that the system can never get into certain undesirable states and that specific bad sequences of events can never occur.

The analyzer component of Alloy harnesses a SAT solver to search for counterexamples—possible scenarios of the software system that are permitted by its design but that fail a sanity check (which is accomplished by writing assertions that must be true if the model is correctly designed). In other words, the tool attempts to construct situations that satisfy the facts but violate a stated assertion. In our case, it would generate a seating plan in which a Capulet (other than Juliet) sits next to a Montague (other than Romeo) at the top table. To fix the seating rule, we can add a new fact: that Romeo and Juliet occupy the top table alone. Now Alloy would find no counterexample.

Together the declarations of the sets and relations, the facts, and the assertions make up an abstraction that captures the essence of the software design. Writing all this out makes the limitations of the design explicit and forces engineers to think hard about exactly which abstractions will work best. Bad abstraction choices lie at the root of many unnecessarily complicated or unreliable systems.

Systems that rely on software built on simple and robust abstractions should also be easier to use. Consider how e-ticketing simplified air travel, how universal product codes made shopping easier or how 800-number-based conference calls made teleconferencing more feasible. Each of these innovations stemmed from a transformation in the basic abstractions embodied in the underlying software.

The Road to Reliability

TOOLS AKIN TO ALLOY are currently used primarily in research and in cutting-edge industrial settings. The technology has been employed to explore new architectures for telephone switching systems, to design avionics processors that are secure against hackers and to describe access-control policies for communications networks. We have used it to check widely used and robust software devices, such as protocols for finding printers on networks and tools for synchronizing files across machines.

In addition, Alloy has uncovered serious deficiencies in published software designs—such as a key management protocol that was supposed to enforce special-access rules based on membership in a group but turned out to grant access to former members who should have been rejected. It is noteworthy that many programmers who have used Alloy have been surprised by the number of flaws that the tool turns up in the designs for even their simplest applications.

It is most likely only a matter of time until tools resembling Alloy are adopted more widely in industry. Improvements in the underlying SAT solvers will make analysis tools faster and better able to handle very large systems. Meanwhile a new generation of software designers, educated in these methods, will incorporate them into their work. Modeling is growing in popularity, particularly among managers desperate to see some description of a software system's design beyond the code itself.

At some point, there may come a time when software becomes so essential to our day-to-day infrastructure that society will no longer tolerate bad software. As a result, governments may even establish inspection and licensing regulations that enforce high-quality program construction techniques. One day, perhaps, software systems will be truly robust, predictable and easy to use—by design. SA

MORE TO EXPLORE

Exploring the Design of an Intentional Naming Scheme with an Automatic Constraint Analyzer. Sarfraz Khurshid and Daniel Jackson in *Proceedings of the 15th IEEE International Conference on Automated Software Engineering, Grenoble, France.* IEEE, September 2000. (Describes application of Alloy to the design of a system for finding resources on a network.)

Automating First-Order Relational Logic. Daniel Jackson in *Proceedings of the 8th ACM SIGSOFT International Symposium on Foundations of Software Engineering: Twenty-First Century Applications.* ACM Press, 2000. (Explains Alloy's analysis.)

A Micromodularity Mechanism. Daniel Jackson, Ilya Shlyakhter and Manu Sridharan in *Proceedings of the Joint 8th European Software Engineering Conference (ESEC) and 9th ACM SIGSOFT Symposium on the Foundations of Software Engineering.* ACM Press, 2001. (Explains key concept in the latest version of Alloy language.)

Alloy: A Lightweight Object Modeling Notation. Daniel Jackson in *ACM Transactions on Software Engineering and Methodology,* Vol. 11, Issue 2, pages 256–290; April 2002. (Original description of Alloy.)

Software Abstractions: Logic, Language, and Analysis. Daniel Jackson. MIT Press, 2006.

Daniel Jackson's Web site: **http://people.csail.mit.edu/dnj/**

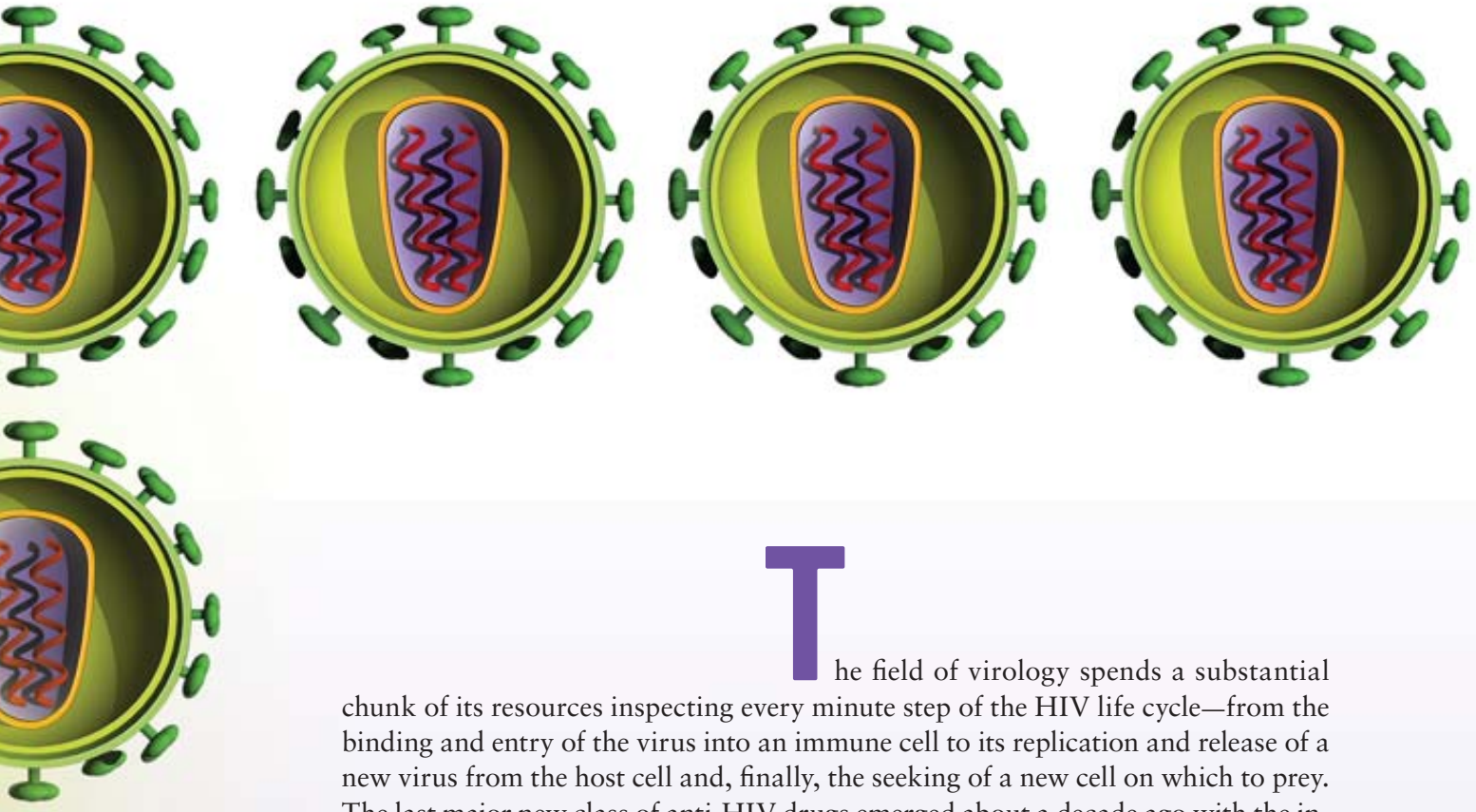
Alloy Web site: **http://alloy.mit.edu**

EMILY HARRISON

A New ASSAULT on HIV

The constant search for weak points in the virus yields ideas for

By Gary Stix



The field of virology spends a substantial chunk of its resources inspecting every minute step of the HIV life cycle—from the binding and entry of the virus into an immune cell to its replication and release of a new virus from the host cell and, finally, the seeking of a new cell on which to prey. The last major new class of anti-HIV drugs emerged about a decade ago with the introduction of the protease inhibitors, which curb the action of an enzyme that is critical to a late stage of viral replication.

At the time, a few members of the HIV research community wondered whether protease inhibitors could provide the basis for a cure. The ingenuity of the virus has proved the hollowness of that hope. As many as half of HIV-positive patients under treatment in the U.S. were found in one study to be infected with viruses that have developed resistance to at least one of the drugs in their regimen. Clinicians can choose from more than 20 pharmaceuticals among protease inhibitors and two classes of drug that prevent the invading virus from copying its RNA into DNA, thereby sabotaging viral replication. Combinations of these agents are administered to counteract the virus's inherent mutability, but that strategy does not always ward off resistance to the medicines, including the protease inhibitors. "Given increasing resistance to protease inhibitors, it's of paramount importance to identify new ways to interfere with the virus replication cycle," asserts Eric Freed, a researcher in the HIV drug resistance program at the National Institutes of Health.

a wholly new class of drug

Drugs that interrupt the beginning, middle and end of viral processing within the host are now in various stages of development. Academic researchers and Panacos, a small biotechnology outfit based in Watertown, Mass., are taking inspiration from the success of protease inhibitors by developing drug candidates known as maturation inhibitors that block protease activity in a novel way. Protease inhibitors mount a direct attack on the HIV protease, preventing the enzyme from processing a viral protein called GAG. When GAG proteins are cut properly, pieces spliced out of it form the conical protective core, or capsid, that encloses RNA. In contrast, the Panacos maturation inhibitor blocks a site on the GAG protein where the protease normally binds, keeping the protease from clipping GAG correctly. As a result, the capsid does

not form appropriately and the virus cannot infect another cell.

Looking for Leads

THE PATH TO THE PANACOS drug candidate began in the mid-1990s, when the company Boston Biomedica undertook a collaboration with a professor from the University of North Carolina at Chapel Hill to screen compounds from a collection of traditional Chinese herbs for biochemical activity against HIV. Kuo-Hsiung Lee's laboratory turned up a potential drug lead in a Taiwanese herb.

The substance, betulinic acid, had weak activity against HIV. After the lab separated the compound into its chemical constituents, the investigators found that one of these components, when chemically modified, exhibited a much stronger effect. "Betulinic acid had activity against HIV at the micromolar level," says Graham Allaway, Panacos's chief operating officer. "This derivative had activity at the nanomolar level."

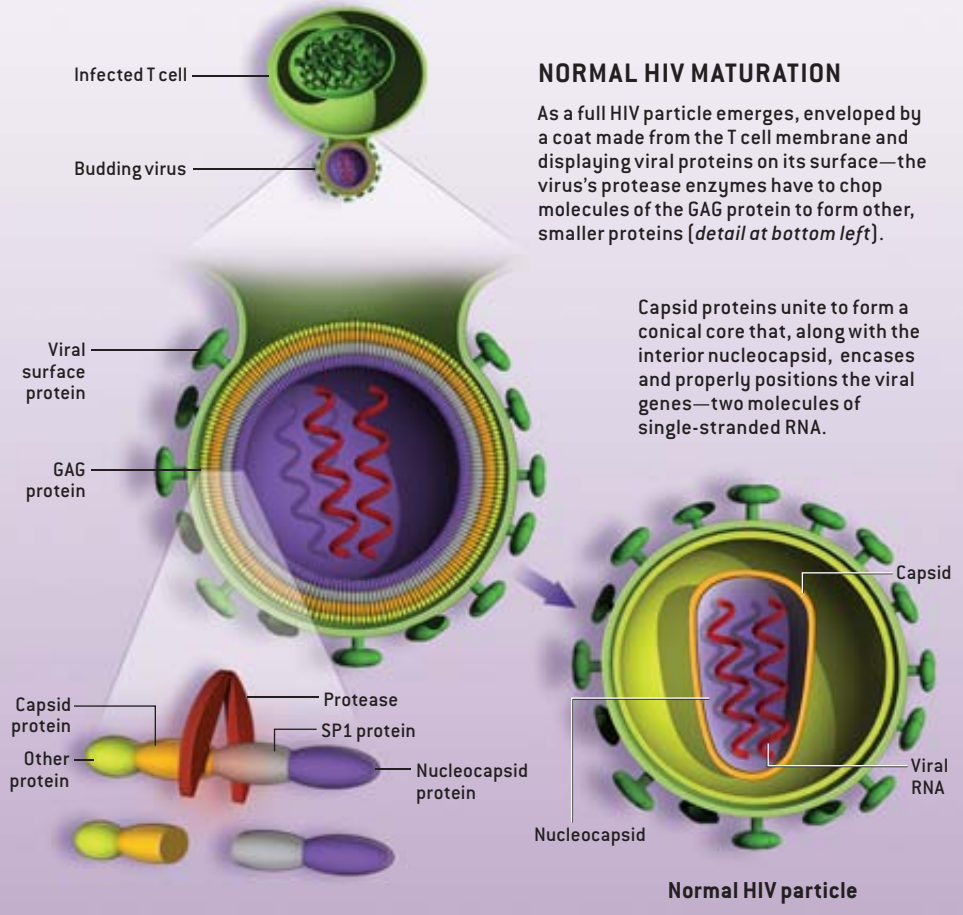
Six years ago Boston Biomedica spun off its HIV research unit into Panacos, which began to investigate the compound, by then named PA-457. PA-457 was not just another Taxol, the anticancer drug that required the controversial felling of rare yew trees until a semisynthetic substitute was found. Panacos did not need a steady source of Taiwanese herbs. Betulinic acid could be extracted from ubiquitous plane and birch trees, and a subsequent processing step yielded the desired molecule.

Even though researchers understood that PA-457 seemed to have activity against all strains of HIV, they needed to find out how the betulinic acid derivative worked against the virus on the molecular level. The company wanted a new class of drug, not just another protease inhibitor. It contacted Freed's laboratory at the NIH, which studies the virus life cycle.

Freed's group and Panacos determined that the drug worked late in the viral replication process, apparently at the stage of capsid formation. The researchers already knew that the HIV capsid forms when newly made GAG attaches from inside the host T cell to the cell's membrane and is then chopped by the HIV protease into smaller pieces. They knew as well, from the development of protease inhibitors, that any disruption to the processing of GAG would cause the virus to become noninfectious. So they began to study PA-457's interaction with GAG to see exactly how it bollixed up the cutting of GAG into its requisite parts.

A NOVEL TREATMENT STRATEGY FOR HIV

Maturation inhibitors constitute a new class of HIV drugs under study. They attack the virus at a late stage in its life cycle—when freshly made components of the virus are coming together



Cultivation of Resistance

TO UNDERSTAND HOW a compound works, scientists often begin by creating resistance, which lets them pin down the exact spot where the drug interacts with its target. To nurture resistance, Freed and his colleagues administered low doses of PA-457 to HIV-infected T cells in culture. The genome of the resistant viruses was sequenced and compared with that of viruses that succumbed to the drug. That analysis located the site on the newly produced viruses that changed in the resistant versions. It turned out to be a site on GAG where the protease binds, and this alteration prevented PA-457 from blocking the enzyme's activity.

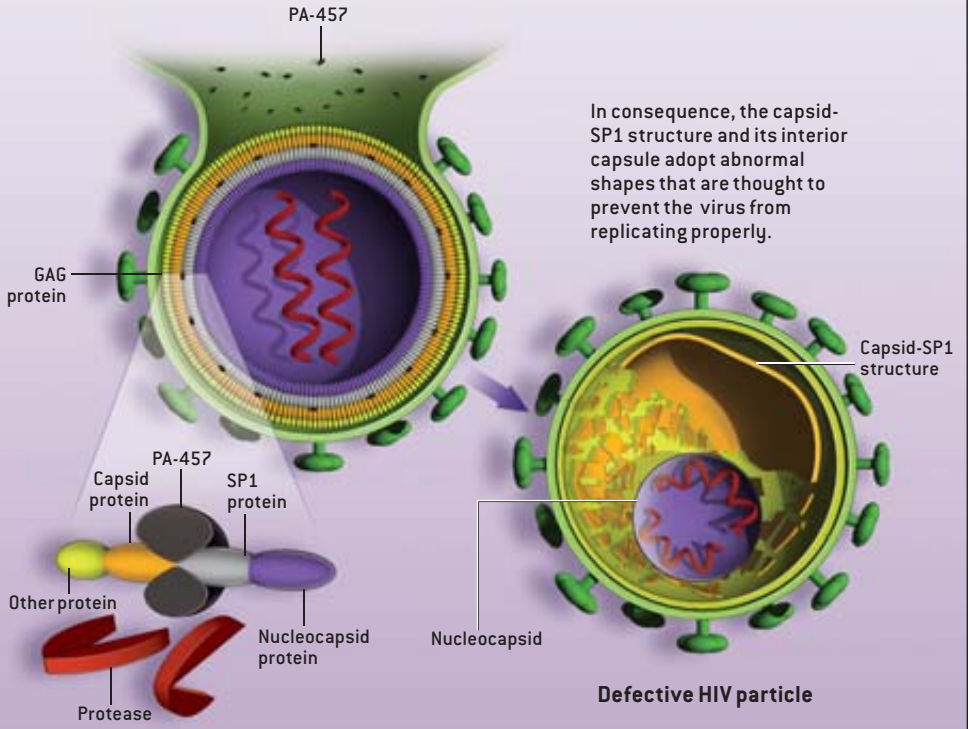
Analyzing the resistant strains allowed the researchers to ascertain that PA-457 was not simply another protease inhibitor. Most drugs, not just HIV inhibitors, work by tinkering with enzymes. "Targeting the substrate [instead of the enzyme] was unknown and surprising," Allaway comments. "As a result, we believe we will have a fairly strong patent position."

TAMI TOLPA (*preceding pages and above*)

into new infectious particles that are beginning to "bud" from one infected T cell so that they can move on to infect another cell.

TREATED VIRUS

The drug candidate PA-457 works by attaching to the GAG protein and preventing the protease from separating the capsid protein from its neighbor in GAG—the SP1 protein (*detail*).



Cultivating resistant strains does not necessarily mean that the drug will have a limited therapeutic life span. In fact, resistance to PA-457 may not develop quickly, because the site where it binds on the GAG protein does not readily change from one HIV strain to another through mutations.

PA-457 has already passed through a midphase clinical trial that checked for drug activity in patients who took it for 10 days while another group received a placebo. HIV replicates so rapidly that a short trial can be used to determine whether a drug is attacking the pathogen in the body. Viral levels averaged a drop of 92 percent at the highest dose of 200 milligrams. The study looked for a decrease in so-called viral load of at least 70 percent as a preliminary sign of the drug's effectiveness. Some patients, however, did not respond—and the company will determine in the next phase of testing whether it can give higher doses. "The main message is that this is an active drug and research should go forward," says Jeffrey M. Jacobson, chief of infectious diseases at Drexel University College of Medicine, who is the lead researcher in the clinical trials.

During the next round, investigators will be looking for interactions with other drugs, an essential test of any HIV drug prospect, because no treatment consists of a single drug therapy, given the threat of resistance. The Food and Drug Administration is encouraging tests earlier during clinical trials these days. In developing new HIV drugs, researchers have at times detected these interactions only much later in the clinical trial process. If all goes according to plan, Panacos could file its final FDA approval application by 2008.

Other Immaturity Preservers

PA-457 IS NOT THE ONLY example of a maturation inhibitor, although it has progressed the furthest toward commercialization. At the University of Alabama and the University of Maryland, researchers working independently have identified small organic molecules that prevent the multitude of capsid subunits from joining up to form the finished casing. "We're trying to jam the parts so they don't fit together," says Peter Prevelige, a professor in the department of microbiology at the University of Alabama.

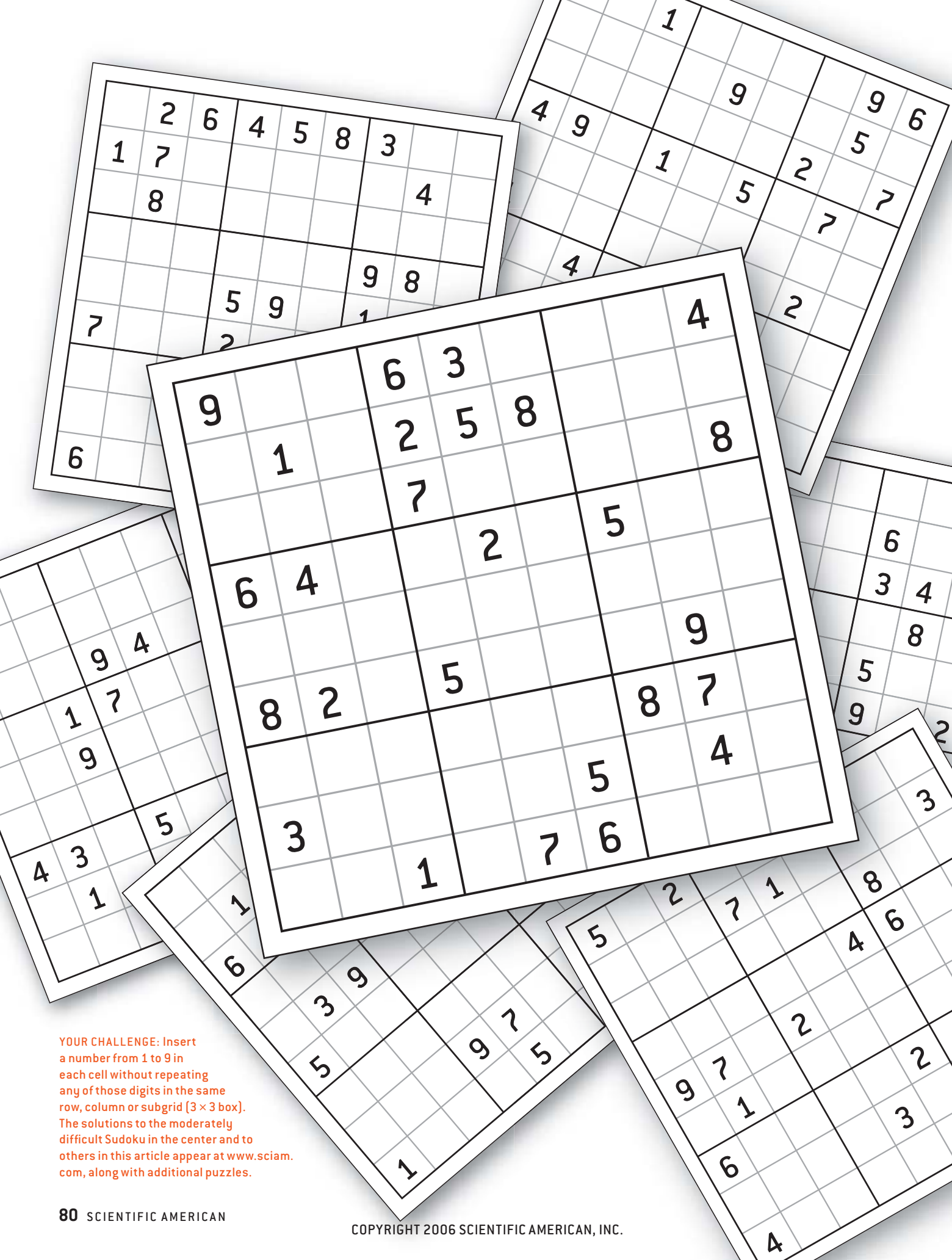
This strategy goes along with other approaches under development to sabotage the viral life cycle. Entry inhibitors, including one Panacos is working on, prevent the virus from entering the cell. (One injectable entry inhibitor has already received FDA approval, but the Panacos drug would be taken orally.) Among the other classes of drugs that have reached late-stage trials are integrase inhibitors, which undermine an enzyme that allows the viral-made DNA to integrate into the host DNA to produce new viral RNA. All these biological agents are needed—and more. Absent a vaccine—not a near-term prospect—the lowly virus, a nanometer-scale capsule of single-stranded RNA, will continue to outwit the best ideas that molecular biologists conjure. SA

MORE TO EXPLORE

PA-457: A Potent HIV Inhibitor That Disrupts Core Condensation by Targeting a Late Step in Gag Processing. F. Li et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 100, No. 23, pages 13555–13560; November 11, 2003.

The Prevalence of Antiretroviral Drug Resistance in the United States. Douglas D. Richman et al. in *AIDS,* Vol. 18, No. 10, pages 1393–1401; July 2, 2004.

The Discovery of a Class of Novel HIV-1 Maturation Inhibitors and Their Potential in the Therapy of HIV. Donglei Yu et al. in *Expert Opinion on Investigational Drugs,* Vol. 14, No. 6, pages 681–693; June 2005.

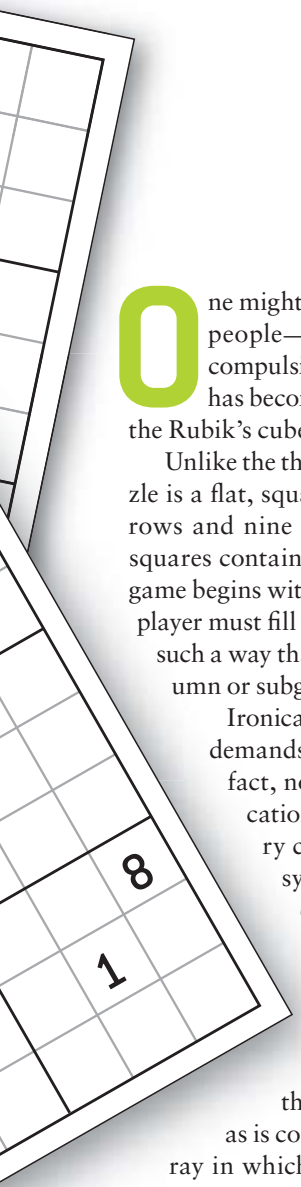


YOUR CHALLENGE: Insert a number from 1 to 9 in each cell without repeating any of those digits in the same row, column or subgrid (3 × 3 box). The solutions to the moderately difficult Sudoku in the center and to others in this article appear at www.sciam.com, along with additional puzzles.

The Science behind SUDOKU

Solving a Sudoku puzzle requires no math, not even arithmetic. Even so, the game poses a number of intriguing mathematical problems

BY JEAN-PAUL DELAHAYE



One might expect a game of logic to appeal to very few people—mathematicians, maybe, computer geeks, compulsive gamblers. Yet in a very short time, Sudoku has become extraordinarily popular, bringing to mind the Rubik's cube craze of the early 1980s.

Unlike the three-dimensional Rubik's cube, a Sudoku puzzle is a flat, square grid. Typically it contains 81 cells (nine rows and nine columns) and is divided into nine smaller squares containing nine cells each; call them subgrids. The game begins with numbers already printed in some cells. The player must fill in the empty cells with the numbers 1 to 9 in such a way that no digit appears twice in the same row, column or subgrid. Each puzzle has one unique solution.

Ironically, despite being a game of numbers, Sudoku demands not an iota of mathematics of its solvers. In fact, no operation—including addition or multiplication—helps in completing a grid, which in theory could be filled with any set of nine different symbols (letters, colors, icons and so on). Nevertheless, Sudoku presents mathematicians and computer scientists with a host of challenging issues.

Family Tree

ONE THING is not unresolved, however: the game's roots. The ancestor of Sudoku is not, as is commonly presumed, the magic square—an array in which the integers in all the rows, columns and diagonals add up to the same sum. Indeed, aside from the numbers and the grid, Sudoku has almost nothing to do with the magic square—but everything to do with the Latin square [*see box on next page*].

A Latin square of order n is a matrix of n^2 cells (n cells on a side), filled with n symbols such that the same symbol never appears twice in the same row or column (each of the n symbols is thus used precisely n times). The origin of those grids dates back to the Middle Ages; later, mathematician Leonhard Euler (1707–1783) named them Latin squares and studied them.

A standard Sudoku is like an order-9 Latin square, differing only in its added requirement that each subgrid contain the numbers 1 through 9. The first such puzzle appeared in the May 1979 edition of *Dell Pencil Puzzles and Word Games* and, according to research done by Will Shortz, the crossword editor of the *New York Times,* was apparently created by a retired architect named Howard Garns. Garns died in Indianapolis in 1989 (or 1981; accounts vary), too early to witness the global success of his invention.

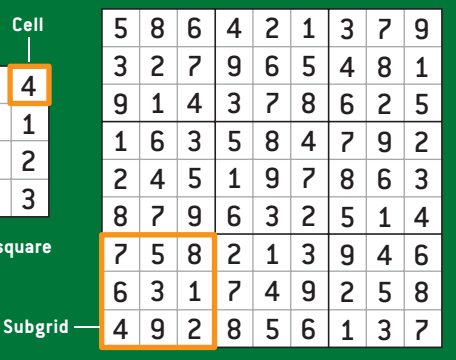
The game, published by Dell as "Number Place," jumped to a magazine in Japan in 1984, which ultimately named it "Sudoku," loosely translated as "single numbers." The magazine trademarked that moniker, and so copycats in Japan used the "Number Place" name. In yet another Sudoku-related irony, then, the Japanese call the puzzle by its English name, and English speakers call it by its Japanese name.

Sudoku owes its subsequent success to Wayne Gould, a peri-

SUDOKU'S PREDECESSORS

A Sudoku grid is a special kind of Latin square. Latin squares, which were so named by the 18th-century mathematician Leonhard Euler, are $n \times n$ matrices that are filled with n symbols in such a way that the same symbol never appears twice in the same row or column. Two examples are shown. The standard completed Sudoku grid (also known as a solution grid) is a 9×9 Latin square that meets the additional constraint of having each of its nine subgrids contain the digits 1 to 9.

Small Latin square ($n = 4$)



Latin square that is also a completed Sudoku grid ($n = 9$)

Leonhard Euler

patetic retired judge living in Hong Kong, who came across it while visiting Japan in 1997 and wrote a computer program that automatically generates Sudoku grids. At the end of 2004 the London *Times* accepted his proposal to publish the puzzles, and in January 2005 the *Daily Telegraph* followed suit. Since then, several dozen daily papers in countries all over the world have taken to printing the game, some even putting it on the cover page as a promotional come-on. Specialty magazines and entire books devoted to this diversion have sprung up, as have tournaments, Web sites and blogs.

As Many Grids as Humans

IT DID NOT TAKE LONG for mathematicians to begin playing "how many" games with Sudoku. For instance, they soon asked how many unique filled-in, or "solution," grids can be constructed. Clearly, the answer has to be smaller than the number of Latin squares because of the added constraints imposed by the subgrids.

There are only 12 Latin squares of order 3 and 576 of order 4, but 5,524,751,496,156,892,842,531,225,600 of order 9. Group theory, however, says that a grid that can be derived from another is equivalent to the original. For instance, if I systematically replaced each number with some other (say, 1 became 2, and 2 became 7, and so on), or if I swapped two rows or columns, the final results would all be essentially the same. If one counts only the reduced forms, then the number of Latin squares of order 9 is 377,597,570,964,258,816 (a result reported in *Discrete Mathematics* in 1975 by Stanley E. Bammel and Jerome Rothstein, then at Ohio State University).

Exactly how many Sudoku grids can exist has been rather difficult to establish. Today only the use of logic (to simplify the problem) and computers (to examine possibilities systematically) makes it possible to estimate the number of valid Sudoku solution grids: 6,670,903,752,021,072,936,960. That number includes all those derived from any particular grid by elementary operations. This result, from Bertram Felgenhauer of the Technical University of Dresden in Germany and Frazer Jarvis of the University of Sheffield in England, has now been verified several times. (Verification matters for results that are obtained this way.)

If we count only once those grids that can be reduced to equivalent configurations, then the number shrinks to 5,472,730,538—slightly lower than the population of humans on the earth. Despite this reduction, Sudoku devotees need not fear any shortage of puzzles.

Note that a complete Sudoku solution grid may be arrived at in more than one way from any starting, or clue, grid (that is, an incomplete grid whose solution is a given complete version). Nobody has yet succeeded in determining how many different starting grids there are. Moreover, a Sudoku starting grid is really only interesting to a mathematician if it is minimal—that is, if removing a single number will mean the solution is no longer unique. No one has figured out the number of possible minimal grids, which would amount to the ultimate count of distinct Sudoku puzzles. It is a challenge that is sure to be taken up in the near future.

Another problem of minimality also remains unsolved—to wit, what is the smallest number of digits a puzzle maker can place in a starting grid and still guarantee a unique solution? The answer seems to be 17. Gordon Royle of

Overview/*Scientific Sudoku*

- Sudoku is more than just an entertaining logic game for players; it also raises a host of deeper issues for mathematicians.
- Such problems include: How many Sudoku grids can be constructed? What is the minimal number of starting clues that will yield one unique solution? Does Sudoku belong to the class of hard problems known as NP-complete?
- Puzzle mavens have come up with an array of approaches to attacking Sudokus and with entertaining variations on the game.

JEN CHRISTIANSEN; SOURCE: JEAN-PAUL DELAHAYE (*preceding pages* and *this page*); EMANUEL HANDMANN, 1753, ORIGINAL PASTEL AT KUNSTMUSEUM BASEL

the University of Western Australia has collected more than 38,000 examples that fit this criterion and cannot be translated into one another by performing elementary operations.

Gary McGuire of the National University of Ireland, Maynooth, is conducting a search for a 16-clue puzzle with a unique solution but has so far come up empty-handed. It begins to look as if none exists. On the other hand, Royle and others working independently have managed to find one 16-clue puzzle that has just two solutions. Searchers have not yet uncovered any additional examples.

Is anyone near to proving that no valid Sudoku puzzle can have only 16 clues? McGuire says no. If we could analyze one grid per second, looking for a valid 16-clue puzzle within it, he notes, "we could search all the grids in 173 years. Unfortunately, we cannot yet do this, even on a fast computer." Soon, he says, searching a grid might be doable in one minute on a powerful computer, but at that rate the endeavor would take 10,380 years. "Even distributed on 10,000 computers, it would take about one year," he adds. "We really need a breakthrough in our understanding to make it feasible to search all the grids. We either need to reduce the search space or find a much better algorithm for searching."

Mathematicians do know the solution to the opposite of the minimum number of clues problem: What is the maximum number of givens that do not guarantee a unique solution? The answer is 77. It is very easy to see that with 80, 79 or 78 givens, if there is a solution, it is unique. But the same cannot be guaranteed for 77 givens [*see bottom box on page 86*].

Computer Solvers

BEYOND THE "HOW MANY" questions, mathematicians and computer scientists enjoy pondering what computers can and cannot do when it comes to solving and generating Sudoku puzzles. For standard Sudokus (9×9), it is relatively easy to write computer programs that solve all valid starting grids.

Another problem: What is the smallest number of digits that can be put in a starting grid and still guarantee a unique solution?

The solution programs can employ several methods, but the most common is backtracking, a systematic form of trial and error in which partial solutions are proposed and then modified slightly as soon as they are proved wrong.

The basic backtracking algorithm works like this: The program places the number 1 in the first empty cell. If the choice is compatible with the existing clues, it continues to the second empty cell, where it places a 1. When it encounters a conflict (which can happen very quickly), it erases the 1 just placed and inserts 2 or, if that is invalid, 3 or the next legal number. After placing the first legal number possible, it moves to the next cell and starts again with a 1.

If the number that has to be changed is a 9 (which cannot be raised by one in a standard Sudoku grid), the program backtracks and increases the number in the previous cell (the next-to-last number placed) by one. Then it moves for-

HOW LOW CAN YOU GO?

The minimum number of clues that a 9 × 9 Sudoku puzzle can start with and still yield a unique solution seems to be 17; an example is shown. One particular filled-in grid, known to Sudoku aficionados as "Strangely Familiar," or SF, hides 29 inequivalent 17-clue starting boards—an unusually high number. SF was once considered the grid most likely to harbor a 16-clue puzzle with a unique solution, but an exhaustive search has dashed that hope. The only known 16-clue Sudoku having just two solutions appears at bottom; the final grids interchange the 8's and 9's.

One 17-Clue Puzzle

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

"Strangely Familiar" Grid

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

16-Clue Puzzle . . .

5		2				4		
			7	1				3
					4	6		
	7		2					
	1							
6					2			
				3			1	
4								

→

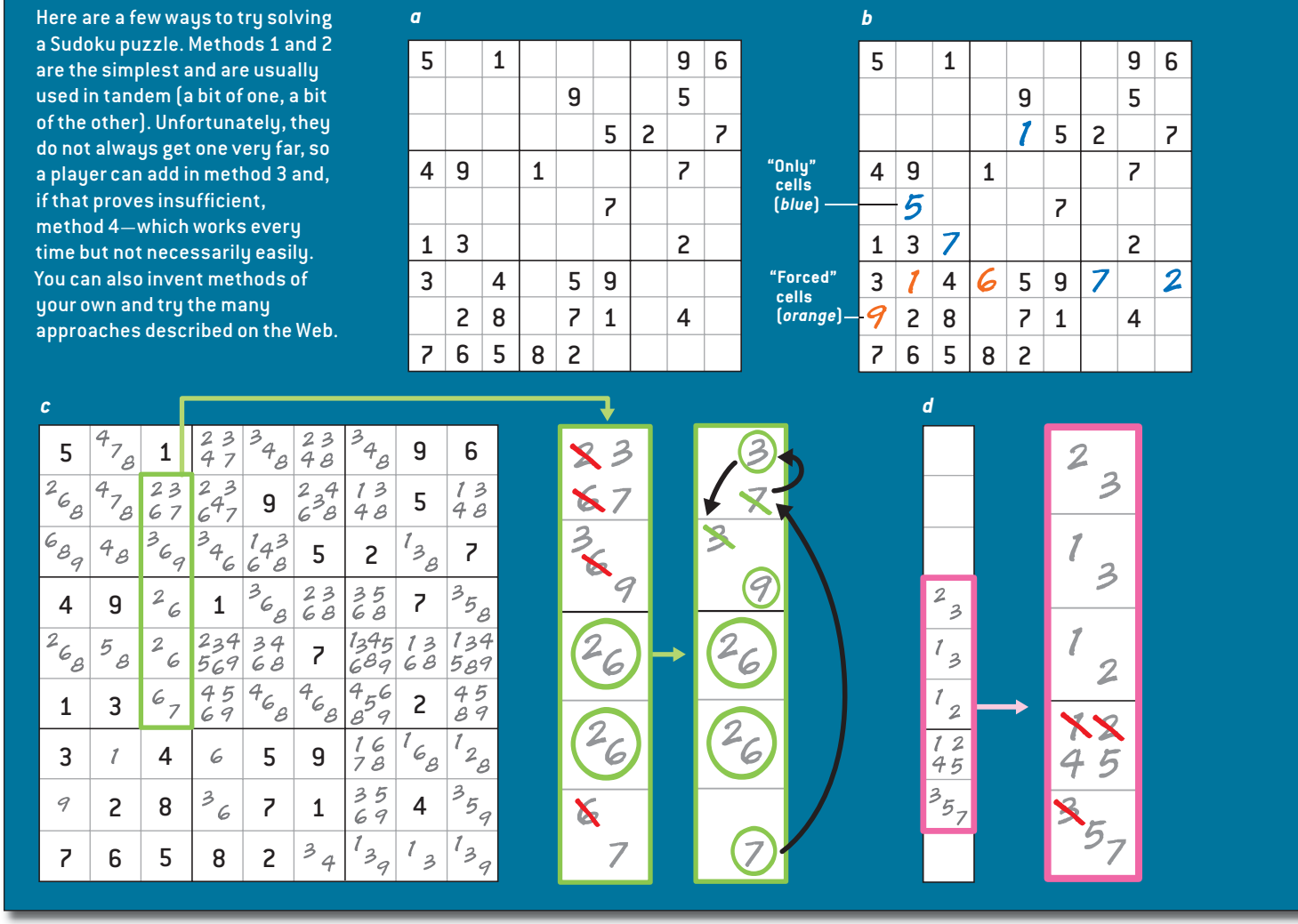
. . . with Two Solutions

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

JEN CHRISTIANSEN; SOURCE: GORDON ROYLE *University of Western Australia*

SOLUTION METHODS

Here are a few ways to try solving a Sudoku puzzle. Methods 1 and 2 are the simplest and are usually used in tandem (a bit of one, a bit of the other). Unfortunately, they do not always get one very far, so a player can add in method 3 and, if that proves insufficient, method 4—which works every time but not necessarily easily. You can also invent methods of your own and try the many approaches described on the Web.



ward until it hits a conflict. (The program sometimes backtracks several times before advancing.) In a well-written program, this method exhaustively explores all possible hypotheses and thus finishes by finding the solution, if one exists. And if multiple solutions exist, as would be the case with a flawed puzzle, the program finds all of them.

Of course, refinements are possible, and they speed up the discovery of the unique solution. One favorite is called constraint propagation: after each new number is placed, the program generates a table of the remaining possible numbers in each empty cell and considers only numbers from this table.

Backtracking techniques can be encoded by fairly short solution programs. Indeed, concise programs have been written for Sudoku in Prolog, a computer language incorporating a backtracking algorithm. Alain Colmerauer and Philippe Roussel of the University of Marseilles in France invented Prolog in the late 1970s.

For human players, the backtracking techniques employed by computer programs would be infeasible; they would require extraordinary patience. So mortals use more varied and smarter rules and generally turn to trial and error only as a last resort. Some programs try to mimic human methods to an extent: they are longer than the others but work just as well. The programs that simulate human reasoning are also useful for assessing the difficulty of starting grids, which are ranked from "easy" (requiring simple tactics) to what many call "devilish" or "diabolical" (because of the need to apply more mind-bending rules of logic).

One way computer scientists think about solving a Sudoku puzzle is to equate the task to a graph-coloring prob-

THE AUTHOR

JEAN-PAUL DELAHAYE is professor of computer science at the University of Sciences and Technologies of Lille in France and a researcher in the Computer Science Laboratory of Lille (LIFL) of the National Center for Scientific Research (CNRS) based at the university. His work focuses on computational game theory (such as the iterated prisoner's dilemma) and complexity theory (such as Kolmogorov complexity) and applications of these theories to genetic analysis and, recently, to economics. This article expands on one Delahaye published in the December 2005 *Pour la Science*, the French edition of *Scientific American*.

JEN CHRISTIANSEN; SOURCE: JEAN-PAUL DELAHAYE

METHOD 1

"FORCED" CELL

This approach considers a cell to be fixed. By eliminating as possibilities any other numbers in the same column, row and subgrid, you see whether only one possibility remains. Such analysis of grid *a* will reveal that the boxes containing orange numbers in grid *b* are "forced" cells.

METHOD 2

"ONLY" CELL

Here a given value is the focus—for example, the number 5. Columns one and three in grid *a* already have 5s, but column two so far has none. Where can that column's 5 be? Not in the three first cells of column two, because they sit in a subgrid that already has a 5. Not in the seventh cell of the column, because its subgrid, too, has a 5. Thus, the 5 of column two is either in the fourth, fifth or sixth cell of the column. Only the fifth cell is free, so the number goes there. The cells marked with blue numbers in grid *b* are "only" cells.

METHOD 3

SIMPLIFYING THE RANGE OF POSSIBILITIES

This technique is extremely powerful but requires a pencil and eraser. In each cell, you write all possible solutions very small or use dots whose positions represent the numbers 1 to 9. Then you apply logic to try to eliminate options.

For example, grid *c* shows how grid *a* would look if it were marked up by rote, without methods 1 and 2 being applied first. In the third column, the array of possibilities for the second, third, fourth, fifth and sixth cells are, respectively, {2, 3, 6, 7}, {3, 6, 9}, {2, 6}, {2, 6} and {6, 7}. The column must contain a 2 and a 6, so these numbers must be in the two cells whose sole possibilities are 2 and 6 (*circled in first detail*). Consequently, 2 and 6 cannot be anywhere else in this column and can be deleted from the column's other cells (*red*). The range of possibilities for the column is simplified to {3, 7}, {3, 9}, {2, 6}, {2, 6}, {7}. But that isn't all. Stipulating the position of 7 in turn dictates the positions of 3 and of 9 (*second detail*). The final possibilities are {3}, {9}, {2, 6}, {2, 6}, {7}. A single uncertainty remains: where 2 and 6 should go.

The general rule for simplifying possibilities is the following: if, among a set of possibilities (for a row, column or subgrid), you find *m* cells that contain a subset consisting only of *m* numbers (but not necessarily all of them in each cell), the digits in the subset can be eliminated as possibilities from the other cells in the larger set. For instance, in *d* the arrangement {2, 3}, {1, 3}, {1, 2}, {1, 2, 4, 5}, {3, 5, 7} can be simplified to {2, 3}, {1, 3}, {1, 2}, {4, 5}, {5, 7}, because the cells {2, 3}, {1, 3}, {1, 2} all come from the subset {1, 2, 3} and have no other numbers.

METHOD 4

TRIAL AND ERROR

By applying methods 1 through 3, you can solve many Sudoku grids. But diabolical-level Sudokus often require a phase of trial and error. When uncertainty persists, you make a random choice and apply all your strategies as if that were the correct decision. If you hit an impossibility (such as two identical numbers in the same column), you know you chose incorrectly. For instance, you might try 2 in the fourth cell of the third column in grid *c*. If that fails, you begin again from the same starting point, but this time with 6 in the box.

Unfortunately, sometimes you have to do several rounds of trial and error, and you have to be prepared to backtrack if you guess incorrectly. Indeed, the idea behind the trial-and-error method is the same as that used by backtracking algorithms, which computer programs can easily implement but which can sorely tax human brains. How remarkable it is that the method most effective for a machine is the least effective for a human being.

lem in which two adjacent cells (otherwise known as "two vertices joined by an edge") cannot take the same color and the available palette has nine colors. The graph contains 81 vertices, some of which are colored at the outset. The coloring problem is actually quite complex because each 9×9 grid has hundreds of edges. Each cell is part of a row including eight other cells, a column including eight other cells, and a subgrid including eight other cells (of which four have already been counted in the column or row). So each of the 81 cells is linked to 20 (8 + 8 + 4) other cells, which makes a grand total of 1,620 cells that share one edge with a neighbor—which, in turn, means that the total number of edges is 810 (1,620 divided by 2).

That Sudoku puzzles can be translated into a coloring problem is meaningful to scientists, because that property links Sudoku to a class of important problems. In particular, Takayuki Yato and Takahiro Seta of the University of Tokyo have recently demonstrated that Sudoku belongs to the category of NP-complete problems. Such problems are ones that probably cannot be solved in a realistic time frame. Well-known examples include the 3-colorability problem, which asks whether it is possible to shade each node in a graph with three colors in such a way that no two nodes joined by an edge are assigned the same color. In the case of Sudoku, the apparently impossible challenge is designing an efficient program that would solve Sudokus of all sizes—that is, every grid that takes the form $n^2 \times n^2$, not just the standard $3^2 \times 3^2$ (9×9) versions. No program for solving all puzzles would work efficiently because the time required to find a solution increases dramatically as n gets bigger.

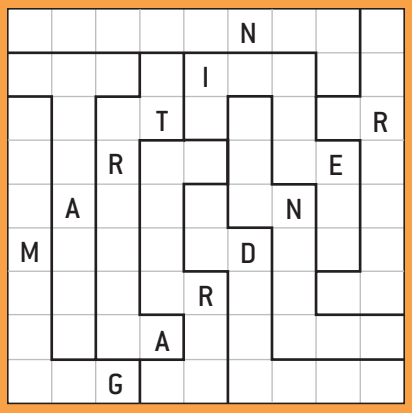
If you have an algorithm that solves classic Sudokus, you can use it to obtain an algorithm that designs them. Indeed,

That Sudoku puzzles can be translated into a coloring problem links the game to a class of important mathematical problems.

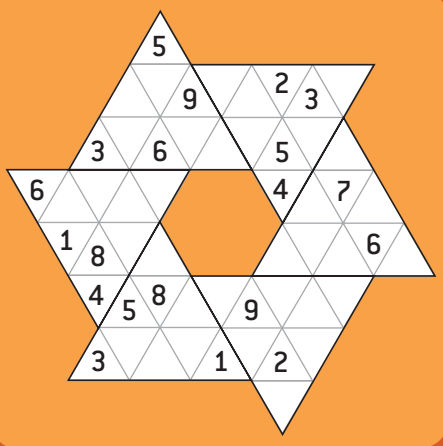
VARIATIONS ON A THEME

In need of something more than standard diabolical grids? In the puzzles here, the usual rules apply, with some twists. In *a*, the letters in the words GRAND TIME replace numbers, and geometric shapes replace the square subgrids. Its inventor calls it a Du-Sum-Oh puzzle. In *b*, which contains six triangular subgrids, the rows and the slanted columns may be interrupted in the center, and when a row or column has only eight cells, the nearby cell that forms a point of the "star" serves as the ninth cell. In *c*, the three-place numbers formed by the marked rows in the first two subgrids add up to the number in the third subgrid. In *d*, greater-than and less-than signs indicate where the digits belong. In *e*, the dominoes at the bottom need to be placed into the empty spaces. In *f*, three game boards overlap. Visit www.sciam.com for solutions and more games.

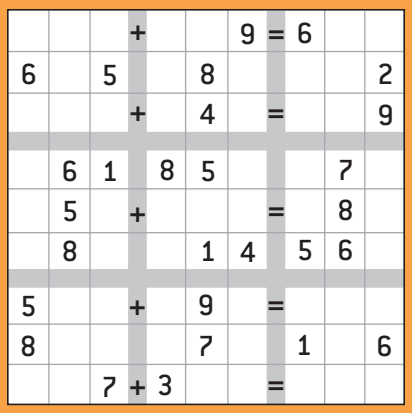
a



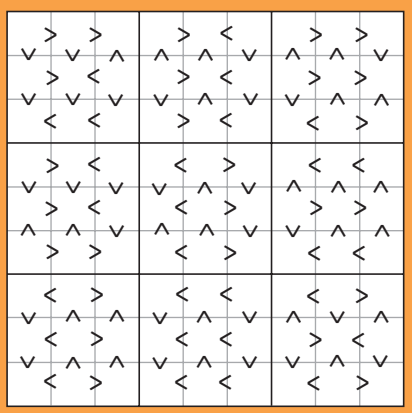
b



c

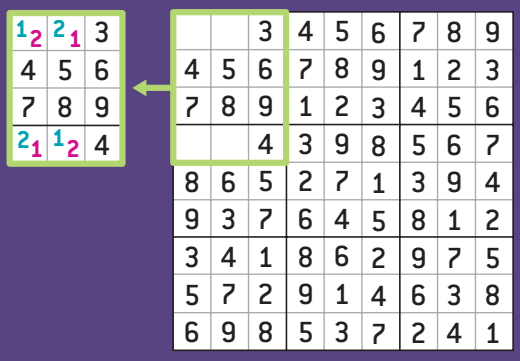


d



TOO FEW CLUES

77 clues are not necessarily enough to guarantee a unique solution. Despite having only four empty cells, the grid here has two solutions: in the first two columns the missing 1's and 2's (*inset*) are interchangeable.



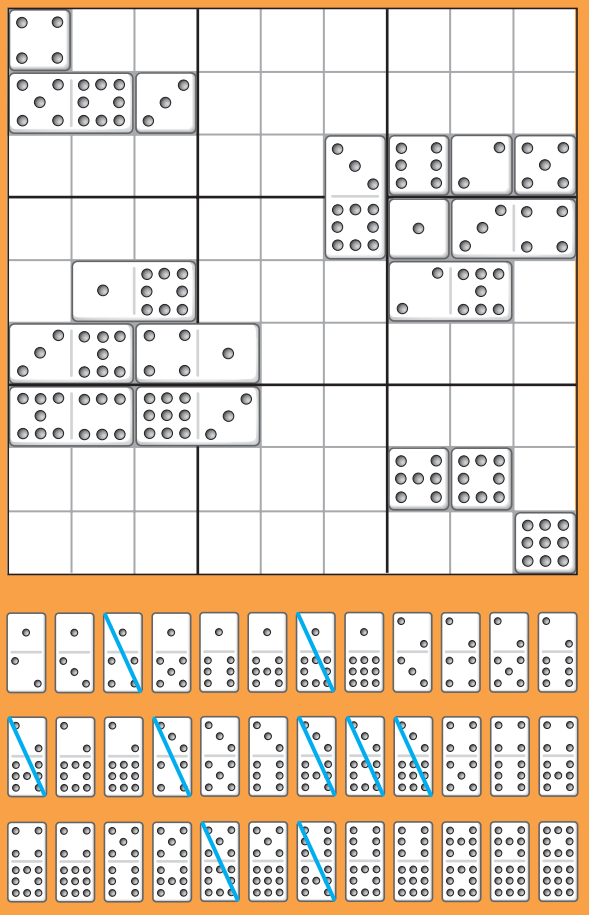
although early Sudokus were constructed by hand, today almost all are produced by computer programs based on the following approach or a similar one. Numbers are placed at random on a grid board, and a solution algorithm (for example, backtracking) is applied. If the puzzle possesses a unique solution, the program stops. If the partially completed problem has no solution, one number is taken away from the starting arrangement and the program begins again. If the puzzle has various solutions, one is chosen and the algorithm then adds as many numbers as needed to the starting clues to ensure that the chosen solution is unique.

Human Strategies

FANS WHO ENJOY solving Sudokus manually can choose among many tactics, but two basic approaches offer a decent starting point. First, search for the most constrained empty cells: those that belong to a row, column or subgrid that is already pretty well filled in. Sometimes eliminating impossibilities (the numbers already occupying cells in the same row, column or grid) will lead you to discover the only number that will work in a particular cell; in any case, the method should greatly narrow the options.

JEN CHRISTIANSEN; SOURCES: BOB HARRIS IN *THE GRAND TIME SUDOKU AND THE LAW OF LEFTOVERS*. A K PETERS LTD [IN PRESS] (*a*); © NONZERO/OLGA LEONTIEVA (*b*); ED PEGG, JR. (*c*); © NONZERO/CIHAN ALTAY (*d*)

e



f

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

Second, search for where a given value might be found in a particular column, row or subgrid (for example, look for the only places where a 3 might fit in row four). Sometimes the query will have only one possible response. Other times just knowing that the 3 can go only in two or three particular spots ends up being helpful. See the box on pages 84 and 85 for more details. Also, visit the Web sites listed in "More to Explore" to find a host of strategies, some of which have such creative names as "swordfish" and "golden chain."

A number of software programs easily found on the Internet will generate boards of specified difficulty and help you find solutions (without, of course, solving the puzzle for you!). For example, some allow you to put temporary marks in the cells and to erase them, thus making pencil and eraser unnecessary. Some even enable you to create links between cells. Do not overlook these software programs. In freeing you from such tedious tasks as erasing, they will actually spur you to greater subtlety and virtuosity in this game of logic.

Once you have gotten bored with traditional Sudokus, you can go looking for the innumerable variants: some overlap multiple grids; others replace square subgrids with other structures; still others introduce additional constraints. These alternatives appeal because they compel you to explore new logical strategies. Moreover, devotees who take only a quarter of an hour to do a traditional puzzle can immerse themselves the entire day in the delights of combining cells and numbers in giant versions of Sudoku. But enough of that. On to the next grid! SA

MORE TO EXPLORE

1st World Sudoku Championship: **www.wsc2006.com/eng/index.php**

Math Games. Ed Pegg, Jr.: **www.maa.org/editorial/mathgames/mathgames_09_05_05.html**

The Mathematics of Su Doku. Sourendu Gupta: **http://theory.tifr.res.in/~sgupta/sudoku/**

Mathematics of Sudoku. Tom Davis: **www.geometer.org/mathcircles**

SadMan Software Sudoku techniques: **www.simes.clara.co.uk/programs/sudokutechniques.htm**

Sudoku, an overview: **www.sudoku.com/howtosolve.htm**

Sudoku, from Wikipedia: **http://en.wikipedia.org/wiki/Sudoku**

A Variety of Sudoku Variants: **www.sudoku.com/forums/viewtopic.php?t=995**

ED PEGG, JR. (*e* and *f*)

OIL REFINERIES

Carbon Hooch

Heating oil, gasoline, jet fuel, kerosene and plastics. These products and more are derived from crude oil in one big fuming silo, siphoned off and fine-tuned through a bewildering maze of pipes.

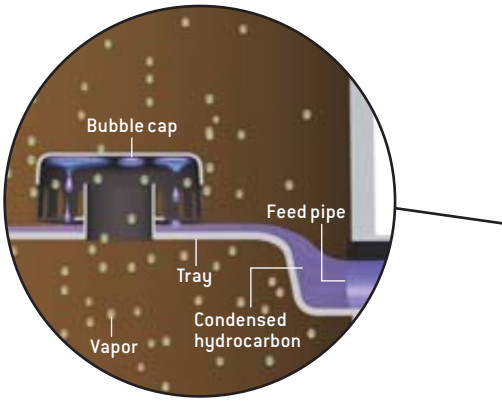
Crude oil contains hundreds of different hydrocarbons. Yet U.S. refineries convert half of all crude into gasoline—a blend of fuel stocks, particularly 2,2,4-trimethylpentane (eight carbon atoms chained together) and heptane (seven carbon atoms). The more complex the chain, the more the molecule can be compressed before it ignites spontaneously, allowing an engine to operate at a higher compression ratio—greater power output. The test mixture by which a gasoline's octane rating is judged combines 2,2,4-trimethylpentane and heptane (87 to 13 percent for "87 octane").

Refiners have tried additives over time to boost octane rating. Tetraethyl lead worked in "leaded" gasoline but was phased out because it spoiled catalytic converters. Producers switched to methyl tertiary butyl ether (MTBE), but it has been implicated in contaminating groundwater, and state governments are banning it. An alternative increasingly being used is ethanol, which has an octane rating of 108 or 110; gasoline with 10 percent ethanol is marketed as gasohol.

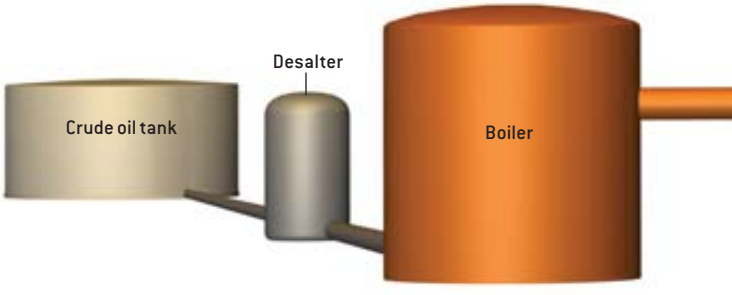
No new U.S. refineries have been built for 30 years because of "not in my backyard" public opposition and environmental permitting hurdles. Although some existing operations have been expanded, refineries nationally are running beyond 90 percent utilization, according to Harold Schobert, professor of fuel science and director of the Energy Institute at Pennsylvania State University. "So if one or two facilities go down, as they did during Hurricane Katrina," he says, "there will be refining shortages, and pump prices will go up."

Still, refineries cannot be blamed—as they have been—for the relentless rise in prices. According to the U.S. Energy Information Administration, the refining step accounts for only 18 percent of the final price (47 percent comes from crude oil, 23 percent from taxes, and 12 percent from distribution and retail dealers). Schobert says that refinery margins are only a few cents per gallon, too. Low profits, plus public resistance, give companies little incentive to invest the $2 billion or more needed to build a new plant. —*Mark Fischetti*

REFINERY boils crude oil, and the vapors enter a distillation column, where they rise and condense as they cool below their boiling points. Long, complex hydrocarbon chains, or fractions, condense at high temperatures near the bottom; shorter, simpler hydrocarbons condense at lower temperatures toward the top. Secondary processes such as cracking and reforming further refine products.



PERFORATED TRAYS allow vapors to bubble up into caps, where they condense, flowing into feed pipes.



GEORGE RETSECK

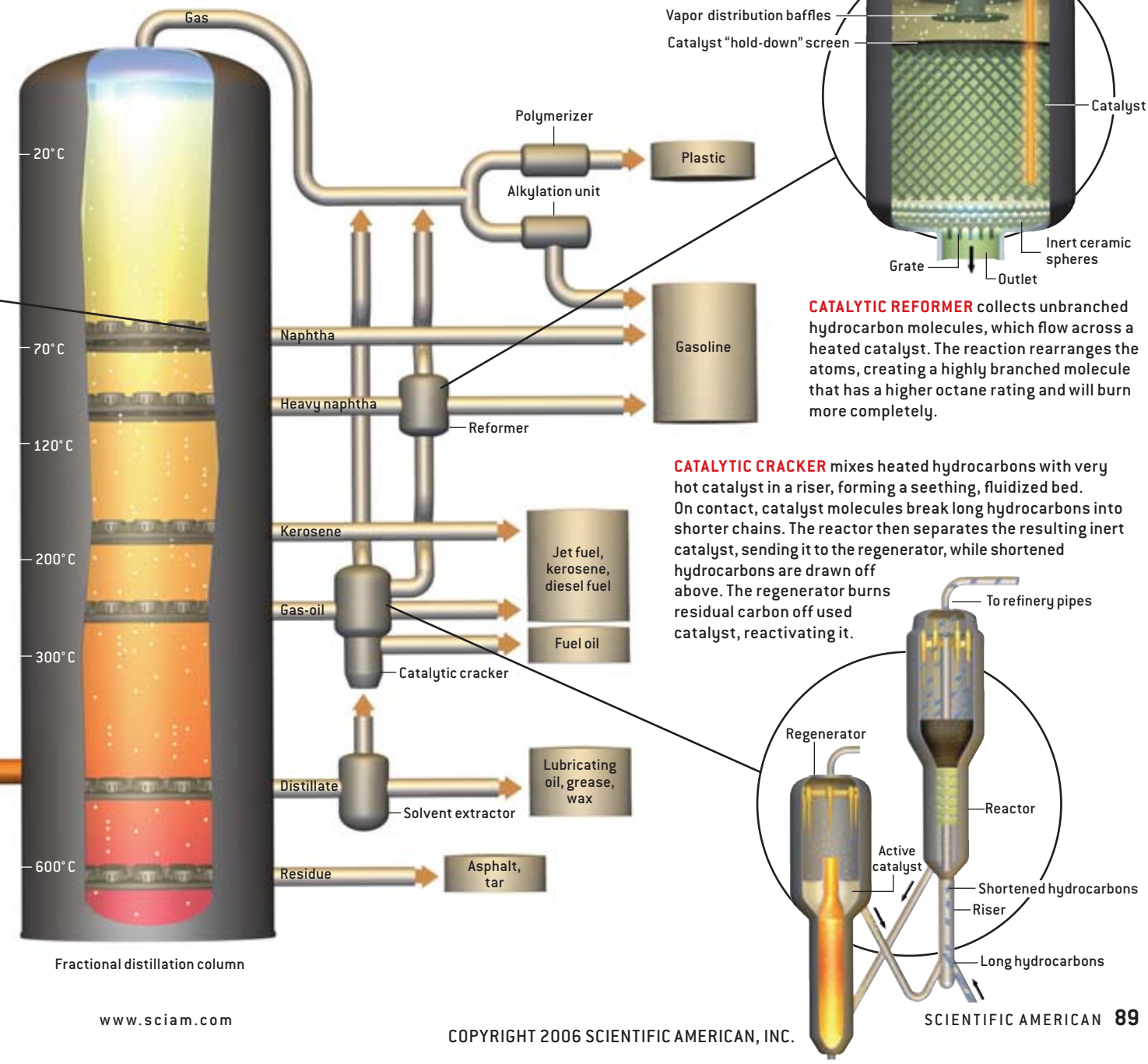
Have a topic idea? Send it to workingknowledge@sciam.com

DID YOU KNOW...

NUISANCE: As whale oil for lamps became scarce in the 1850s, Canadians converted seeping ground oil into kerosene. Russia had dug trial oil wells, but the industry ignited in 1859, when American industrialist George Bissell contracted Edwin Drake to drill a well near Oil Creek, Pa. It produced paraffin, readily distilled into kerosene. Gasoline, a by-product, was discarded as a nuisance.

ONE WORD—PLASTICS: Some crude oil contains dissolved propane and butane that is tapped at the top of the distillation column. Propane is later converted to propylene for use in textiles, food packaging and automotive components. Butane is converted to butadiene for synthetic rubber. Most ethylene—the base for polyethylenes found in plastic containers and packaging—is refined in a high-temperature, low-pressure "thermal" cracker. "The cracker unzips the hydrocarbons, one pair of carbon atoms at a time," explains Penn State's Harold Schobert. "Each double-bonded pair is an ethylene molecule."

OPEC AGAIN: Refining capacity is expanding substantially in OPEC countries. Plants under way include a 600,000-barrel-per-day (bpd) operation in Kuwait and a 450,000-bpd site in Saudi Arabia. India, China and South Korea also plan vast growth. Arizona Clean Fuels has proposed a 150,000-bpd refinery outside Yuma, but it has not yet completed funding or agreements.



Fractional distillation column

CATALYTIC REFORMER collects unbranched hydrocarbon molecules, which flow across a heated catalyst. The reaction rearranges the atoms, creating a highly branched molecule that has a higher octane rating and will burn more completely.

CATALYTIC CRACKER mixes heated hydrocarbons with very hot catalyst in a riser, forming a seething, fluidized bed. On contact, catalyst molecules break long hydrocarbons into shorter chains. The reactor then separates the resulting inert catalyst, sending it to the regenerator, while shortened hydrocarbons are drawn off above. The regenerator burns residual carbon off used catalyst, reactivating it.

The Ultimate Blood Test

A PRICEY WAY TO DETERMINE HEALTH RISKS: 250 TESTS AT ONCE **BY PHILIP YAM**

As the dizziness began to fade and nausea to subside, I kept thinking how two tablespoons did not sound like a lot of blood. During regular checkups, my physician draws only about half that amount. I suppose I might have guessed, especially after a 12-hour fast, I would sicken when my glucose levels dipped—I'm a terrible blood donor in that regard.

The nurse who drew my blood helpfully looked around my office for a sweet drink. "Do you have any soda or juice?" she asked. But the only thing I had was a can of Diet Coke. Which in a way is ironic: I used to drink regular Coke but switched to the sugar-free form after blood tests revealed that my triglycerides were too high.

Momentary ill feelings, though, were an acceptable physical price for 250 blood tests done at once—I was told that running them separately with conventional means would require a liter of blood. Imagine how dizzy and nauseated I'd feel then. So how could I not roll up my sleeve for Biophysical Corporation? The Austin, Tex.–based company promised to use the blood to screen for presymptomatic cancers, potential immune disorders, latent infections, undetected hormonal imbalances and unrecognized nutritional deficiencies. It seemed to mark a step toward that *Star Trek* future in which Dr. McCoy waves around a saltshaker device to determine a person's medical secrets. ("Heartbeat is all wrong. Body temperature is … Jim, this man is a Klingon!")

The Biophysical250 assessment, as the firm calls it, is more than just a battery of tests. It includes a medical-history interview; a personal visit to the home or office for the blood draw (I should have picked my home, where I actually keep sugar); and a follow-up physician consultation. All this attention does not come cheap. It costs $3,400, and it is not covered by health insurance. The company states that doing each test individually would cost 10 times more, so Biophysical250 is a bargain by comparison. Still, you need some disposable income, or you must be so indispensable to your employers that they will pay for it. I don't fall into either category. But because I was reviewing its product, Biophysical agreed to conduct the test on me for free.

FIVE VIALS, or about two tablespoons, is all the blood needed for the Biophysical250 assessment.

JOHN FRASER

The analysis focuses on blood biomarkers, which are chemicals whose presence or amount may indicate abnormal processes or reactions in the body. Among the most familiar are cardiovascular ones: high- and low-density lipoproteins (HDL and LDL, the good and bad cholesterols) and triglycerides.

Checking 250 biomarkers at once might seem like overkill. A routine exam screens for two or three dozen. Looking at one biomarker in isolation, however, is usually not especially informative—for instance, the ratio of LDL to HDL is more important than either alone. The Biophysical250 takes it much further: to assess the risk for heart disease and stroke, the firm analyzes 33 biomarkers.

And examining several biomarkers together improves the odds of finding problems early, especially malignancies.

Blood tests for cancers have been problematic, because healthy people may produce the same kinds and amounts of the biomarkers that cancer patients do. Moreover, the chemicals do not always show up in cancer patients, and they may result from unrelated conditions. The Biophysical250 screens for about four dozen blood chemicals tied to cancerous activity in general to increase the odds of detecting presymptomatic disease.

As an example, Biophysical points to ovarian cancer, which is usually diagnosed too late. Cancer antigen 125, the most commonly measured marker for the disease, shows up in only half of patients in stage 1, when treatment is most likely to succeed. The Biophysical250 tries to boost the chance of early detection by measuring other, biologically independent compounds, such as vascular endothelial growth factor, interleukin-6, and monocyte chemoattractant protein. "Just the fact that we stack so many biomarkers really minimizes false positives," comments Mark Chandler, CEO and founder of Biophysical.

The firm has basically miniaturized the standard blood tests, substituting the microliter wells with tinier polystyrene beads, each about half the size of a red blood cell. Each sphere is coated with a particular antibody. The serum from the blood sample mixes with the beads for 15 to 30 minutes, allowing the antibodies to grab onto the proteins they recognize. After the serum is washed away, another of the same set of antibodies goes in. This time, though, each antibody also has a fluorescent tag. The tagged antibodies sandwich the blood proteins already held by the first set of antibodies. Examining the fluorescent tags thus provides "an idea of how much of the chemical was pulled out of serum," explains Chandler, who began selling the Biophysical250 assessment last year.

My report arrived two weeks later via FedEx. It included a well-written summary plus a quantitative laboratory report. A second booklet defined all the biomarkers and the ailments with which they correlated. In terms of health, the most useful part is the summary of biomarkers organized by type: autoimmune, cancer, cardiovascular, cell signaling, diabetes, endocrine, hematology, immune/inflammation, infectious disease, nutritional, organ systems and osteoarthritic. Next to each biomarker was a color code, depending on whether the detected amount was out of range: green for "low risk," yellow for "caution" and red for "alert." The report also came with a copy to give to my personal physician; Biophysical will discuss the results

with a client's doctor. I also later had a telephone consultation about my results with George Rodgers, the company's president and cardiovascular specialist.

My results were mind-numbingly normal. Most everything came up green. The only surprise was my slightly out-of-range ferritin, a protein that stores iron. My report warned me that such iron overload might signal a genetic condition called hemochromatosis. The disease progresses silently and can cause toxic levels of iron to build up in organs. The treatment is simple—regular blood donations to drain off the excess iron. On the other hand, my ferritin level might reflect the fact that I had been taking a multivitamin with iron—a no-no for healthy men, I later discovered. It's something I'll check with my physician.

And therein lies a great strength of Biophysical250: it can uncover a presymptomatic, potentially fatal disease that physicians might not ordinarily test for. The firm reports that in an unpublished study of 120 clients, 15 turned out to have major health risks and another 27 indicated moderate risks; none showed outward signs of disease. The conditions included rheumatoid arthritis, scleroderma and hypothyroidism.

The company screens only for treatable ailments and avoids those that are invariably fatal. So for now, neurodegenerative conditions such as Alzheimer's disease are out. But, Chandler adds, the firm would consider testing for such illnesses "if there's a way to slow down progression of the disease."

My Biophysical250 results were limited in the sense that they reflect my health on January 10, 2006, at 9:30 A.M., when my blood was drawn. Biochemical changes over time, though, reveal more about the state of a person's health. But at the cost of a giant flat-screen plasma TV, the Biophysical250 is not exactly affordable, even if done every other year. Couldn't the company knock off a few of the tests? I mean, did I really need to find out that I have no African sleeping sickness parasites, considering that I've never been to Africa? Or to know that I am not pregnant?

Chandler says that plucking out a few of the beads would not be cost-effective, although perhaps a few dozen biomarkers might be enough to catch the most common afflictions and permit a less expensive assessment. He would like data from 10,000 clients before pruning the number of biomarkers. (He expects about 1,500 customers this year.) The company may head in the other direction and institute a Biophysical300 as research uncovers more biomarkers. I'd certainly be game for it if the price came down—and as long as it does not need more than two tablespoons of blood. SA

Favored by the Gods

HAPPINESS, ACCORDING TO CURRENT SCIENTIFIC THINKING, DEPENDS LESS ON OUR CIRCUMSTANCES THAN ON OUR GENETIC ENDOWMENT **BY GEORGE JOHNSON**

STUMBLING ON HAPPINESS
by Daniel Gilbert
Alfred A. Knopf, 2006 ($24.95)

THE HAPPINESS HYPOTHESIS
by Jonathan Haidt
Basic Books, 2006 ($26)

HAPPINESS: A HISTORY
by Darrin M. McMahon
Atlantic Monthly Press, 2006 ($27.50)

The sky was smeared with the lights from the midway, spinning, blinking, beckoning to risk takers, but I decided to go for a different kind of thrill: Ronnie and Donnie Galyon, the Siamese Twins, were at the Minnesota State Fair. Feeling some guilt, I bought my ticket and cautiously approached the window of the trailer they called home. Thirty years old, joined at the stomach, they were sitting on a sofa, craning their necks to watch television.

Twenty-five years later I am still struck by this dizzying conjunction of the grotesque and the mundane. Trying to project myself into their situation—a man with two heads, two men with one body—I felt only sickness, horror and a certainty that I would rather be dead. Yet there they were, traveling from town to town, leading some kind of life.

When we try to envision another's happiness, we suffer from arrogance and a poverty of imagination. In 1997, when science writer Natalie Angier interviewed Lori and Reba Schappell, connected at the back of the head and sporting different hairdos, each insisted that she was basically content.

"There are good days and bad days—so what?" Reba said. "This is what we know. We don't hate it. We live it every day." Lori was as emphatic: "People come up to me and say, 'You're such an inspiration. Now I realize how minor my own problems are compared to yours.' But they have no idea what problems I have or don't have, or what my life is like."

Three recent books—two by scientists and one by a historian—take on the quest for the good life, in which common sense and the received wisdom of the ages is increasingly confronted by findings from psychology, neuroscience and genetics.

In the fight for survival, evolution has naturally molded our nervous systems to respond more quickly to the negative than to the positive. Before the late-blooming forebrain can reason that the hiss we hear is from an espresso machine, the primitive amygdala whispers "snake." Depending on which part of the brain is paid more heed, a person will tend to be glum and anxious or cheerful and daring, but the choice may be beyond control.

Each of us, it seems, has an emotional "set point" or "cognitive style" that is genetically rooted and hard to budge. In *The Happiness Hypothesis,* Jonathan Haidt, a psychologist at the University of Virginia, observes that lottery winners and paralysis victims quickly return to equilibrium, regaining the level of joy—or sadness—they felt before their brush with fate. How Ronnie, Donnie, Lori, Reba and the rest of us feel about life may depend less on circumstance than on natural disposition, the shape of the brain.

Knowing nothing of neurology, the ancient Greeks equated happiness with being favored by the gods, something over which they were powerless. This fatalism is frozen in our language. Happiness, happenstance, haphazard, hapless—all derive from the same root. In *Happiness: A History,* Darrin M. McMahon, a historian at Florida State University, charts how this germ of an idea changed over time. With Christianity, happiness became something to aspire to in the afterlife, and with the Enlightenment something to pursue on

ROB HOWARD *Corbis*

SCHOOLGIRLS in Bhutan, where in the 1970s the king proclaimed gross national happiness more important than gross domestic product.

earth. From there it was a natural progression to the smiley face (invented in 1963, McMahon tells us, by one Harvey R. Ball) and the decision by the country of Bhutan to measure its economy according to "gross national happiness."

With evolutionary biology we have come, full circle, back to the Greeks: happiness is in the luck of the draw, how we fare in the genetic sweepstakes, the modern name for Fortuna's wheel. Not even geography or economic position is as influential a factor. Several years ago in the journal *Science & Spirit,* another psychologist, Robert Biswas-Diener, wrote about the remarkably high spirits he found among people in a Calcutta slum and on the harsh northern coast of Greenland. "Research shows that we are the fortunate inheritors of a highly evolved emotional system that leads us to be, for the most part, somewhat happy," he wrote. "We have a tendency to interpret things positively and to adjust quickly to most events."

The downside is that this reflexive optimism can keep us from making good guesses about what will or will not bring us joy. It is not just the hard lives of others that we have trouble imagining but also our own. In *Stumbling on Happiness,* Daniel Gilbert, a Harvard psychologist studying "affective forecasting," shows that people have inflated expectations about the joy they will derive from a vacation, a new car or child, or a second dessert. But our failure as futurists also cuts the other way. We overestimate how bad we will feel if we get fired or lose a tooth or even a friend or mate. Rationalization, our emotional immune system, insists on putting the best face possible on even the saddest events.

"We treat our future selves as though they were our children," Gilbert writes, "spending most of the hours of most of our days constructing tomorrows that we hope will make them happy." But the children turn out to be ingrates, complaining that we should have let them stay in the old house or study dentistry instead of law.

Taken together, all these findings may seem a little depressing. But true to our nature, we can see them in a sunnier way. A whole industry has sprung up—mass-market therapy, cosmetics, cheap luxury cruises—promoting a kind of gross national sappiness, an obligation to have fun. A little knowledge from the psych labs may take off some of the pressure, providing grist for the inverse of a self-help book—not a guide on how to achieve happiness but on understanding why, in the end, you probably won't. SA

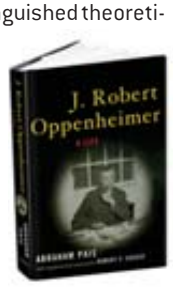
George Johnson's most recent book is Miss Leavitt's Stars: The Unknown Story of the Woman Who Discovered How to Measure the Universe *(W. W. Norton, 2005).*

THE EDITORS **RECOMMEND**

J. ROBERT OPPENHEIMER: A LIFE

by Abraham Pais, with supplemental material by Robert P. Crease. Oxford University Press, 2006 ($30)

Before you mutter "Not another book about Oppenheimer," recall that the author of this one—the late Abraham Pais—wrote what is arguably the most comprehensive biography of Albert Einstein, *Subtle Is the Lord,* which won a National Book Award. In the present book, Pais (who was a distinguished theoretical physicist) once again combines his sophisticated understanding of the science and his insider's knowledge of the man (he was Oppenheimer's next-door neighbor for many years) to produce a stunning portrait. Historian of science Crease completed the book after Pais's death.



THE GOSPEL OF THE FLYING SPAGHETTI MONSTER

by Bobby Henderson. Villard, 2006 ($13.95, paperbound)

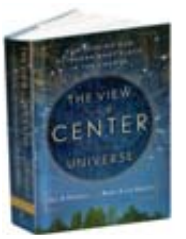
An elaborate spoof on Intelligent Design, *The Gospel of the Flying Spaghetti Monster* is neither too elaborate nor too spoofy to succeed in nailing the fallacies of ID. It's even wackier than Jonathan Swift's suggestion that the Irish eat their children as a way to keep them from being a burden, and it may offend just as many people, but Henderson, described elsewhere as a 25-year-old "out-of-work physics major," puts satire to the same serious use that Swift did.

Oh, yes, it is very funny.

THE VIEW FROM THE CENTER OF THE UNIVERSE

by Joel R. Primack and Nancy Ellen Abrams. Riverhead Books, 2006 ($26.95)

In this thoughtful and original book, a husband-and-wife team presents a science-based cosmology aimed at allowing us to understand the universe as a whole and our place in it. "Most of us have grown up thinking that there is no basis for our feeling central or even important to the cosmos," they write. "But with the new evidence it turns out that this perspective is nothing but a prejudice. There is no geographic center to an expanding universe, but we are central in several unexpected ways that derive directly from physics and cosmology." Primack is professor of cosmology at the University of California, Santa Cruz, and an originator of the theory of cold dark matter; Abrams is a lawyer and a writer.



Up the Lazy Creek

THE MORE THINGS CHANGE, THE MORE THEY'RE NOT THE SAME **BY STEVE MIRSKY**

I turned this column in very late. I just couldn't get started writing it. Low energy. You know how it is.

Fortunately, my editor can take no action against me, because my lateness, I was delighted to discover, was in fact brought about by a disease: I clearly suffer from "motivational deficiency disorder." The *British Medical Journal,* the praises of which were sung in this space in March, reported on this novel malaise in its April 1 issue. "Extreme laziness," the *BMJ* piece explains, "may have a medical basis, says a group of high-profile Australian scientists, describing a new condition called motivational deficiency disorder (MoDeD)." (MoDeD is most definitely not to be confused with Mos Def, who is clearly not a sufferer of MoDeD.)

The article went on to quote a Dr. Leth Argos as one of the discoverers of the disorder. The allusion to lethargy should have been the tip-off, even if you missed the date, that the article was an April Fools' joke. The MoDeD giveaway was the citing of a drug called Strivor, which was allegedly so successful in treating the disease that "one young man who could not leave his sofa is now working as an investment adviser." Talk about your dangerous side effects.

The prank, however, had a purpose: the piece was designed to bring attention to a conference on so-called disease mongering, the medicalization of ordinary conditions, which thereby opens markets for drugs to treat them as illnesses. The stunt, however, may have inadvertently demonstrated the existence of a true ailment. Because a quick search of the Internet reveals that numerous news outlets picked up the *BMJ* press release and ran it without a hint of skepticism. That's just motivationally deficient journalism.

In other factual news, hurricane season is here again, accompanied by reminders of the threat of terrorism so constant that the fear of a terrorist attack is probably itself a new malady. (Even though it's still heart disease, cancer or some driver on a bender or a cell phone that's probably going to get you.) But the obvious and dire need to be ready for emergencies notwithstanding, seven candidates for top jobs, including the directorship of the Federal Emergency Management Agency (FEMA), told the *New York Times* that they had removed themselves from consideration because they weren't sure the Bush administration was really serious about emergency management. (The acting director of FEMA, R. David Paulison, was eventually nominated for that top job.)

I bring this up only because Representative Harold Rogers of Kentucky, chair of the House subcommittee that controls the purse strings of the Department of Homeland Security and thus of FEMA, actually said, "Let the word go forth from this place that we want a permanent director of FEMA, and we want these regional directors and division directors to stop acting and be permanent. Because I want somebody responsible that we can turn to."

Rogers's use of the phrase "let the word go forth from this place" conjured up memories of another statement that began almost the same way. In his inaugural address, President John F. Kennedy said, "Let the word go forth from this time and place, to friend and foe alike, that the torch has been passed to a new generation of Americans—born in this century, tempered by war, disciplined by a hard and bitter peace, proud of our ancient heritage, and unwilling to witness or permit the slow undoing of those human rights to which this nation has always been committed, and to which we are committed today at home and around the world." In 45 years, "let the word go forth" has gone from introducing a sweeping summation of who we are and what we stand for to announcing frustration over the inability to find someone to run FEMA. Talk about motivational deficiency. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

ASK THE EXPERTS

What are the physical and chemical changes that occur in fireworks?

—S. Brealy, Chapel Hill, N.C.

Paul Nicholas Worsey, professor of mining engineering at the University of Missouri at Rolla, teaches a course in pyrotechnics. He provides the following answer:

Fireworks, or pyrotechnics, contain burning compounds. The most common is the aerial shell, which is fired from a mortar tube. It has four components: a lift charge, a time-delay fuse, a breaking charge and a light-effect generator.

The lift charge is a compound that burns rapidly and propels the shell from its tube. When the lift charge fires, it also lights the delay fuse, which in turn ignites the break charge when the shell reaches the appropriate height. The break charge explodes, scattering the shell's now fiery contents. The shell's payload usually comprises small pyrotechnic pellets designed to generate light. These capsules burn from the outside inward, and color changes result from layers of different compositions.

The metal within the pyrotechnic mixture and the burning temperature dictate the color and intensity of the display. When certain metals are heated, electrons jump between so-called electron shells, or energy levels, within the atoms. When they fall back to a lower state they emit a photon whose wavelength determines the color. The easiest hues to create are red (from strontium), green (barium), yellow (sodium) and white (titanium). Blue is more difficult because the reaction temperature has to be just right. Sparks, in contrast, come from slower-burning mixtures. Shapes such as rings, hearts and smiley faces require precise placement of the pellets inside the shell.

All pyrotechnic mixtures contain a fuel and an oxidizer. The mixture usually consists of a metal nitrate and a carbon-based fuel. When the solid blend ignites, it turns into predominantly gases and gives out a lot of heat—in excess of 2,000 degrees Celsius for brilliant colors. A number of factors control the speed of a firework reaction: the composition of the shell and other physical characteristics, such as the grain size (smaller means faster), the presence of accelerators (sulfur and sugars, for example) or retarders (salt, for instance), high pressure or confinement (which increases the reaction rate), lower packing density (which reduces the rate), and moisture content.

How do antibiotics kill bacterial cells but not human cells?

—M. Cantatore, New York City

Harry Mobley, chair, department of microbiology and immunology at the University of Michigan Medical School, replies:

Antibiotics can selectively target bacteria for eradication, leaving human cells unmolested, in several ways.

Most bacterial cell walls contain a macromolecule called peptidoglycan, which human cells do not make or need. Penicillin, for instance, prevents the final cross-linking step, or transpeptidation, in the macromolecule's assembly. The result is a fragile cell wall that bursts, killing the bacterium.

Certain drugs target bacterial metabolic pathways. Sulfonamide drugs are structurally similar to para-aminobenzoic acid, a compound needed for folic acid synthesis. All cells require folic acid. Whereas the vitamin diffuses easily into human cells, it cannot enter bacterial cells, and so bacteria must make their own. Sulfa drugs inhibit a critical enzyme in this process, and the bacteria can no longer grow.

Another antibiotic—tetracycline—interferes with bacterial growth by stopping protein synthesis. Both bacterial and human cells carry out protein synthesis on structures called ribosomes. Tetracycline binds to a site on the ribosome and blocks a key RNA interaction, shutting off the lengthening protein chain. In human cells, tetracycline does not accumulate enough to hinder protein synthesis.

Similarly, DNA replication must occur in both bacteria and human cells. Antibiotics such as ciprofloxacin can specifically target an essential enzyme called DNA gyrase in bacteria. But this antibiotic does not affect human DNA gyrases. SA

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

MATT COLLINS