SATELLITE-GUIDED BOMBS: GPS and the Next War

SCIENTIFIC AMERICAN

Alcohol and Health: How Much Is Good for You?

FEBRUARY 2003 $4.95
WWW.SCIAM.COM

Starquake on a MAGNETAR

The Inventing Machines
Computer Programs Evolve New Circuits

Suicide and Biology
How the Brain Betrays Itself

Freaky Frogs
A Jump in Deformities

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

february 2003

contents

features

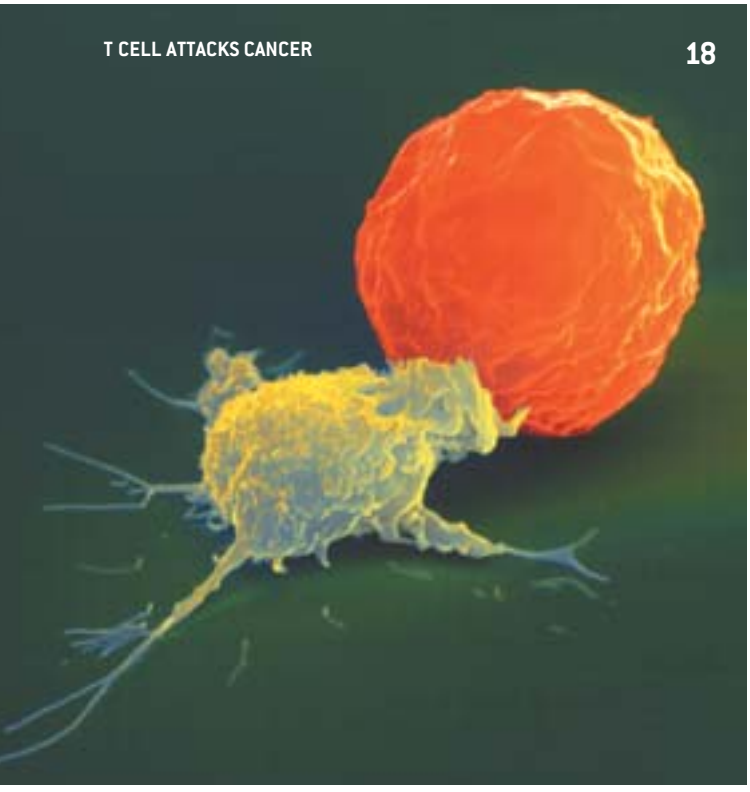
SCIENTIFIC AMERICAN Volume 288 Number 2

52 Better circuits through evolution.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SCIENTIFIC AMERICAN Volume 288 Number 2

departments



columns

Cover image by Don Dixon.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2003 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced by any mechanical, photographic or electronic process, or in the form of a phonographic recording, nor may it be stored in a retrieval system, transmitted or otherwise copied for public or private use without written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 242764. Canadian BN No. 127387652RT; QST No. Q1015332537. Subscription rates: one year $34.97, Canada $49, International $55. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408 or send e-mail to sacust@sciam.com **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 247-7631.** Printed in U.S.A.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

No Immunity to Pork

Critics may gripe about whether the new Homeland Security Act fights terrorism well, but no one can say it doesn't do a great job of protecting drug companies from autistic children.

A short provision at the end of the act, added quietly just days before its passage, exempts Eli Lilly and other firms from direct civil litigation over whether vaccine additives cause autism. Parents suing on behalf of their autistic children are shunted to a federal "vaccine court," where damages are capped. Conveniently, in late November 2002 the Justice Department also requested that the court seal documents relating to hundreds of the lawsuits, complicating the cases for plaintiffs.

VACCINATION FEARS are fed by bad legislation.

Ever since these shameful developments became public, they have drawn bipartisan scorn. Beyond the provision's offensiveness as political pork, it is harmful to lifesaving vaccination efforts.

Worries about childhood vaccines and autism stretch back for years. Studies suggest that rates of autism may have as much as tripled in the past decade. Autism's first symptoms often emerge around age two, shortly after most infants start to receive vaccinations against measles, whooping cough and other illnesses. Because the number of vaccinations that children receive has also skyrocketed, concerned parents sought a linkage, and they found one in thimerosal, a mercury compound used as a preservative in many vaccines. Some symptoms of autism resemble those of mercury poisoning.

As a precaution, in 1999 the Food and Drug Administration ordered the elimination of thimerosal from children's vaccines, although medical authorities generally maintain that the mercury exposure was too low to cause autism's neurological defects. Studies have repeatedly failed to find an epidemiological tie between vaccines and autism, but an Institute of Medicine review in 2001 concluded that the thimerosal theory was "biologically plausible," and so investigation continues.

The U.S. needs a better, comprehensive strategy for vaccines. Vaccines are the most effective public health measure ever devised, but drug companies are reluctant to work on them because the profitability is low and the liability risks are high. If we want new vaccines against bioweapons such as smallpox, we will probably need to give the pharmaceutical industry more incentives and protection. Senator Bill Frist of Tennessee outlined one such scheme in 2002, but his proposal caught legislative flu and died.

Then, presto: language crafted as a shield against thimerosal torts suddenly materialized at the end of the nearly 500-page Homeland Security Bill. No one—not Eli Lilly, not administration officials, not committee members who oversaw the bill—will admit to having inserted the vaccine rider. It just appeared, a Thanksgiving miracle for drugmakers.

The provision does nothing to promote new vaccine development. By lending support to the impression that the industry has something to hide, it fuels distrust of vaccines—exactly when better data absolving the drugs are emerging. Consequently, too many parents are denying their children vaccinations that could save them from potentially fatal diseases.

Here's a suggestion: If no one will accept responsibility for the mysterious legislation, would any of its beneficiaries like to repudiate it? To ask for the repeal of the rider so that vaccine policies can be debated intelligently, as they deserve? Anyone?

THE EDITORS editors@sciam.com

SATURN STILLS/SPL/PHOTO RESEARCHERS, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138
Los Angeles
310-234-2699
fax: 310-234-2670
San Francisco
415-403-9030
fax: 415-403-9033
Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457
Southeast/Southwest
MancheeMedia
972-662-2503
fax: 972-662-2577
Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138
Canada
Fenn Company, Inc.
905-833-6200
fax: 905-833-2116
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922
France and Switzerland
PEM-PEMA
+33-1-46-37-2117
fax: +33-1-47-38-6329
Germany
Publicitas Germany GmbH
+49-211-862-092-0
fax: +49-211-862-092-21
Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +49-8-442-7059
Belgium
Publicitas Media S.A.
+32-(0)2-639-8420
fax: +32-(0)2-639-8430
Middle East and India
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320
Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139
Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662
Hong Kong
Hutton Media Limited
+852-2528-9135
fax: +852-2528-9281

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

Growing Gas Giants

Current models of solar system evolution hold that a planet of Jupiter's size would need more than a million years to form. According to the results of a new study, however, such gas giants may take shape much more quickly than that—perhaps in just hundreds of years.

Mouse Genome Sequenced

In the name of science, investigators have fashioned numerous kinds of mice: fat, thin and hairless, to name a few. The first draft sequence of the mouse genome should make the rodents even more helpful for future research into a variety of human disorders.

Researchers Refine Musical Map of the Brain

A wrong note in a piano concerto can stick out like a proverbial sore thumb. That's because the relations among pitches in a piece of music prime us to hear certain sounds together. Scientists have now identified the brain region involved in tracking tones.

Sound Waves Chill in Novel Freezer Design

Most existing methods for cooling things down require the use of chemical refrigerants, many of which are potent greenhouse gases. But the chemicals in your freezer may one day be replaced by harmless sound waves.

Ask the Experts

Why do men have nipples?

Biologist Andrew M. Simons of Carleton University in Ottawa, Ontario, explains.
www.sciam.com/askexpert_directory.cfm

SCIENTIFIC AMERICAN MARKETPLACE

Your one-stop destination for buying and selling almost anything, including business and professional services, general merchandise and collectibles.

View our listings or place an ad today.
www.sciam.com/marketplace

NASA

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SCIENTIFIC AMERICAN®
Established 1845

EDITOR IN CHIEF: John Rennie
EXECUTIVE EDITOR: Mariette DiChristina
MANAGING EDITOR: Ricki L. Rusting
NEWS EDITOR: Philip M. Yam
SPECIAL PROJECTS EDITOR: Gary Stix
REVIEWS EDITOR: Michelle Press
SENIOR WRITER: W. Wayt Gibbs
EDITORS: Mark Alpert, Steven Ashley, Graham P. Collins, Carol Ezzell, Steve Mirsky, George Musser
CONTRIBUTING EDITORS: Mark Fischetti, Marguerite Holloway, Michael Shermer, Sarah Simpson, Paul Wallich

EDITORIAL DIRECTOR, ONLINE: Kristin Leutwyler
SENIOR EDITOR, ONLINE: Kate Wong
ASSOCIATE EDITOR, ONLINE: Sarah Graham
WEB DESIGN MANAGER: Ryan Reid

ART DIRECTOR: Edward Bell
SENIOR ASSOCIATE ART DIRECTOR: Jana Brenning
ASSISTANT ART DIRECTORS: Johnny Johnson, Mark Clemens
PHOTOGRAPHY EDITOR: Bridget Gerety
PRODUCTION EDITOR: Richard Hunt

COPY DIRECTOR: Maria-Christina Keller
COPY CHIEF: Molly K. Frances
COPY AND RESEARCH: Daniel C. Schlenoff, Rina Bander, Shea Dean, Emily Harrison

EDITORIAL ADMINISTRATOR: Jacob Lasky
SENIOR SECRETARY: Maya Harty

ASSOCIATE PUBLISHER, PRODUCTION: William Sherman
MANUFACTURING MANAGER: Janet Cermak
ADVERTISING PRODUCTION MANAGER: Carl Cherebin
PREPRESS AND QUALITY MANAGER: Silvia Di Placido
PRINT PRODUCTION MANAGER: Georgina Franco
PRODUCTION MANAGER: Christina Hippeli
CUSTOM PUBLISHING MANAGER: Madelyn Keyes-Milch

ASSOCIATE PUBLISHER/VICE PRESIDENT, CIRCULATION: Lorraine Leib Terlecki
CIRCULATION MANAGER: Katherine Corvino
CIRCULATION PROMOTION MANAGER: Joanne Guralnick
FULFILLMENT AND DISTRIBUTION MANAGER: Rosa Davis

PUBLISHER: Bruce Brandfon
ASSOCIATE PUBLISHER: Gail Delott
SALES DEVELOPMENT MANAGER: David Tirpack
SALES REPRESENTATIVES: Stephen Dudley, Hunter Millington, Stan Schmidt, Debra Silver

ASSOCIATE PUBLISHER, STRATEGIC PLANNING: Laura Salant
PROMOTION MANAGER: Diane Schube
RESEARCH MANAGER: Aida Dadurian
PROMOTION DESIGN MANAGER: Nancy Mongelli

GENERAL MANAGER: Michael Florek
BUSINESS MANAGER: Marie Maher
MANAGER, ADVERTISING ACCOUNTING AND COORDINATION: Constance Holmes

DIRECTOR, SPECIAL PROJECTS: Barth David Schwartz

MANAGING DIRECTOR, SCIENTIFICAMERICAN.COM: Mina C. Lux

DIRECTOR, ANCILLARY PRODUCTS: Diane McGarvey
PERMISSIONS MANAGER: Linda Hertz
MANAGER OF CUSTOM PUBLISHING: Jeremy A. Abbate

CHAIRMAN EMERITUS: John J. Hanley
CHAIRMAN: Rolf Grisebach
PRESIDENT AND CHIEF EXECUTIVE OFFICER: Gretchen G. Teichgraeber
VICE PRESIDENT AND MANAGING DIRECTOR, INTERNATIONAL: Charles McCullagh
VICE PRESIDENT: Frances Newburg

Letters

EDITORS@SCIAM.COM

LETTER WRITERS OFTEN comment on the perceived substance—or lack thereof—of *Scientific American*'s articles. But one correspondent takes the concept to an admirable level. "Graham P. Collins seems to be taking an overly skeptical, even facetious, view of perpetual-motion research ['There's No Stopping Them,' Staking Claims, October 2002]. Clearly, he has not made a serious effort to investigate the matter fully," writes Stephen Palmer of Plainfield, N.J. "For example, I have recently applied for a patent of a perpetual-motion device that has been proven to work perfectly and, indeed, perpetually. This amazing invention sets into motion an infinite number of virtual particles, which flicker in and out of existence every instant. I have decided to call it 'nothing.' Like all entrepreneurs, I intend to make my fortune from royalties as soon as nothing is patented. I will follow the path of many wealthy dot-com pioneers, except that I have a firm business plan: when I receive investment capital, I will promptly send nothing in return." There's nothing more we can add about this topic, but others weigh in on the substance of the rest of the October issue below.

FUEL-CELL FOLLIES?

"You would still have to cut down the trees and pave everything over for roads." This was an answer given by a fourth-grade student when I asked what environmental effects cars would have if they were powered by a nonpolluting source of energy, such as hydrogen fuel cells ["Vehicle of Change," by Lawrence D. Burns, J. Byron McCormick and Christopher E. Borroni-Bird].

The biggest impact of private motor vehicles is the creation of sprawling land use, which in turn causes forced dependency on cars. Fuel-cell cars would also still injure millions of Americans in collisions, another problem with personal transportation, and would still leave stranded the one third of the U.S. population that doesn't drive. Cars would still sit in traffic jams and average a lower effective speed than bicycles. We can do much better with transportation and land use.

Robert Bernstein
Transportation chair
Sierra Club–Santa Barbara Group
Goleta, Calif.

"Vehicle of Change" fails to discuss the challenges facing fuel cells. For one, the authors state: "The hydrogen fuel-cell vehicle is nearly twice as efficient as an internal-combustion engine, so it will require only half the fuel energy." In fact, the efficiency depends on electrical load. Although proton-exchange-membrane fuel-cell systems can achieve an efficiency of 50 percent under low loads, it is unlikely that they would be operated in this manner in a production vehicle.

The article also neglects to account for losses associated with deriving hydrogen from other energy sources. Hydrogen will initially be obtained by reforming natural gas, a process with, at best, an efficiency of 85 percent.

The authors list problems with storing hydrogen, yet they fail to note how serious these are. A tank with hydrogen at the suggested 350 bar would be about 10 times as large in volume as one holding gasoline with the same energy content. In addition, the energy required to compress and transport hydrogen by pipeline or truck to the point of use is three to four times as great for hydrogen as for natural gas on a per-unit energy basis.

The transformation to a vehicle fleet powered by hydrogen fuel cells would require an extensive and expensive change in the fuel-supply infrastructure but result in only marginal efficiency gains. From an environmental standpoint, there is minimal reduction in greenhouse-gas production when hydrogen comes from reforming carbon-based fuels, because

GENERAL MOTORS

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

CAR BODIES would sit atop a "skateboard" chassis in the General Motors fuel-cell concept vehicle.

carbon dioxide is a by-product. Further problems relating to safety, fuel-cell stack life and refueling methods are significant. It would be far more productive to focus on hybrid-electric internal-combustion vehicles, mass-transportation concepts and smaller, lighter vehicles.

S. A. Klein and D. T. Reindl
Department of Mechanical Engineering
University of Wisconsin–Madison

Perspectives describes how automakers use fantastic future technology programs to obscure their more immediate and less lofty motives. You chide the automakers for this greenwashing, but you spare the government officials who coconspire in the charade. Their abdication of leadership deserves most of the blame for a failing American energy policy and our appalling consumption of petroleum.

The U.S. consumes 45 percent of the world's gasoline but has 5 percent of the planet's population. Still, our lawmakers can't pass a five-cent gas tax or close fuel-economy loopholes big enough for millions of pickups and SUVs. We need an energy policy that reduces petroleum consumption through conservation and substitution starting now. Instead we get "the hydrogen economy," a far-fetched scheme that is well into the future and will probably stay there. Greenwashing won't hide the ugly truth of armed conflict as energy policy.

Tom Gage
Sunnyvale, Calif.

JOE ZEFF

The auto industry is not interested in making fuel-efficient vehicles because the public is not interested in purchasing such vehicles. Just look at the top-selling cars in the 1990s—SUVs and pickup trucks. This is called supply and demand. Until we have fuel-cell cars, let's try buying already available efficient vehicles. As for me, I like my motorcycle.

Mark Baker
Cuddebackville, N.Y.

HOPE FOR SPINAL INJURIES

I'd like to point out a misconception about spinal-cord injuries raised by "Controlling Robots with the Mind," by Miguel A. L. Nicolelis and John K. Chapin. The authors state that scientists may be able to repair spinal-cord breaks in the distant future. Although this may be true, most people with spinal-cord injuries (including myself) have contusion injuries: the cord is not cut. Because this is a simpler problem, there are already promising thera-

Letters

SCIENTIFIC AMERICAN
Help Desk

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

Scientific American Frontiers

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

www.sciam.com

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

Scientific American subscription rates

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

Subscription inquiries

Call us: In the U.S. and Canada:(800) 333-1199. In other countries: (515) 247-7631. E-mail: subscriptions@sciam.com. Or write: Subscription Manager; Scientific American; PO Box 3187; Harlan, IA 51593. The date of the last issue of your subscription appears on the mailing label. For change of address, please notify us at least eight weeks in advance, and include both old and new addresses.

Back issues

In U.S. $9.95 each. Elsewhere $12.95 each. Many issues are available. Fax your order with your Visa, MasterCard or AMEX information to: (212) 355-0408.

Reprints

$4 each, minimum order 10 copies, prepaid. Limited availability. Write to: Reprint Department; Scientific American; 415 Madison Avenue; New York, NY 10017-1111; fax: (212) 355-0408 or tel. (212) 451-8877.

Photocopying rights

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

Subscriber alert!

Scientific American has been made aware that some subscribers have received notifications/subscription offers from companies such as Publishers Services Exchange, United Publishers Network, Global Publication Services, Associated Publishers Subscription Services and Publishers Access Service. These are not authorized representatives/agents of Scientific American and are selling subscriptions at a much higher rate than the regular subscription or renewal price. Please forward any correspondence you may receive from these companies to: L. Terlecki; Scientific American; 415 Madison Ave.; NY, NY 10017

pies for repairing damaged spinal cords that are either nearing or in clinical trials.

Bruce Hanson
Bellevue, Wash.

THE NAKED, AQUATIC APE?

"**Skin Deep,**" by Nina G. Jablonski and George Chaplin, makes a good case for the evolution of melanin as a strategy for human reproductive success. But it glosses over the reason for such an adaptation: the loss of hair. The hypothesis presented, that our ancestors lost their hair to adapt to savanna life, is untenable on several grounds. First, other savanna- and desert-dwelling mammals have hair, which shades their skin and reduces heat stress. Second, humans did not lose the hair that covers our most heat-sensitive organ, the brain.

It seems likely that another evolutionary force besides heat protection was responsible for human hairlessness. Although fossil evidence may be thin, the "aquatic ape" hypothesis makes sense. If our ancestors had taken to foraging for food along seacoasts, loss of hair and an increase in subcutaneous fat would have been adaptive as protection from the chilling effects of water. These adaptations are observed in most modern aquatic animals as well as in humans.

Michael DeWeert
Kailua, Hawaii

SHOCKED BY ELECTROSTATICS

"**Lightning Rods** for Nanoelectronics," by Steven H. Voldman, asserts that "people who like to tinker with their computers know that when they open up their machines, they should 'ground' themselves—perhaps by touching the metal radiator panel or attaching a wire from their fingers to a metal fixture." Without a bit more detail, this information could be deadly. Obviously, unless someone is properly trained, any tinkering with electrical devices should be done with the power disconnected.

Robert E. Fields
Los Alamos National Laboratory
Los Alamos, N.M.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Old Fish ▪ New Cars ▪ Blue Light

FEBRUARY 1953

LIVING FOSSIL—"In the Indian Ocean off Madagascar, fishermen last month netted a five-foot, 100-pound fish which evolutionists promptly hailed as the 'most important zoological discovery of this century.' J.L.B. Smith, South African ichthyologist, flew 3,000 miles in a government-supplied plane to reach the fish in time to preserve it. When he arrived, and found it smelling somewhat strong but largely intact, he broke down and wept. The object of his emotion was a coelacanth, the earliest type of bony fish. Until a few years ago it was believed that such fish had been extinct for 75 million years, but in 1938 one was pulled out of the water by a South African trawler. By the time Smith got hold of it, only its skeleton and skin were left. Since that time he has been on a constant lookout for another specimen."

BEFORE WATSON AND CRICK—"An intact molecule of desoxyribonucleic [*sic*] acid, called DNA for short, is a very large, complicated structure: it may contain as many as 3,000 molecules of a 5-carbon sugar. DNA is an example of what is nowadays called a high polymer. A familiar example of a high polymer is nylon. The characteristic of a high polymer is that some chemical unit is linked together repeatedly to form a big structure. In nylon the unit is relatively simple, there being but one type of submolecule. In DNA the units are far more complex. To learn how they are polymerized to form a giant molecule is a formidable task which has not yet been accomplished. When it is, we shall understand better how DNA functions in the chromosome. —Alfred Ezra Mirsky"

FEBRUARY 1903

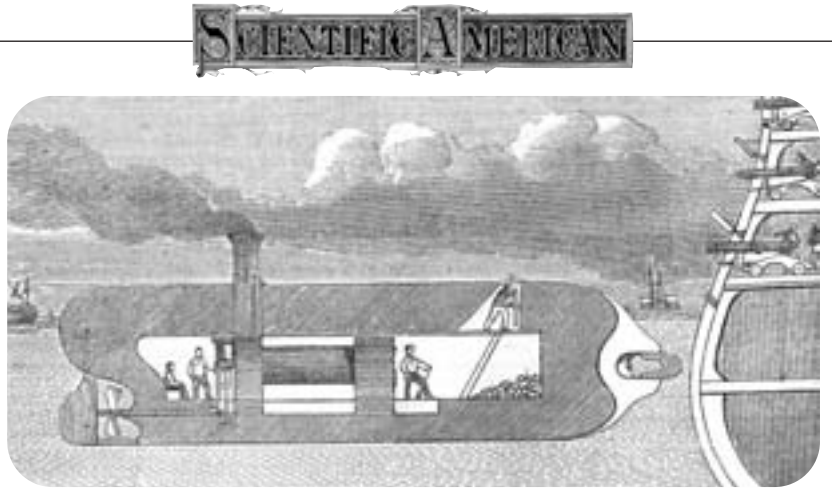
NEW CARS—"Three-quarters of the vehicles at the New York automobile show were of the internal-combustion cylinder type, the rest being steam or electric carriages. Prices ranged from $500 to $8,000. The entire absence of racing monsters was a sign of the tendency to build for comfort, economy, and efficiency, with moderate speed for touring purposes. If touring over the country is not popular this coming season, it never will be."

MERCURY VAPOR LAMP—"Mr. George Westinghouse, during his recent stay in London, exhibited the new lamp invented by Mr. Peter Cooper Hewitt. The lamp consists of a glass tube filled with the vapor of mercury. On passing a direct current through the lamp, the vapor which fills the tube is rendered incandescent and gives off a steady, blue-white light. Owing to the great resistance at the negative electrode to the initial flow of current, it is necessary to use a high voltage to start the lamp. The light given off by the incandescent vapor is entirely lacking in red rays, but on account of its wonderfully low cost, the Cooper Hewitt light should be found very useful, without the addition of any rectifying light, for illuminating factories, yards, etc., where the differentiating of colors is unimportant. Another promising field for the new light is that of photography."

FEBRUARY 1853

INTERIOR OF THE EARTH—"Prof. Silliman, of Yale College, says, 'Heat in the earth increases about one degree for every fifty feet of descent; so that, if we were to go down two miles, we should find boiling water. Is all then beneath us on fire? There is strong evidence to justify such a theory. Witness the hot springs of Bath in England. These are the more remarkable as there are no volcanoes in the British Islands. We know that from the time of the Romans these waters have never ceased to gush up in vast abundance.' "

PAGING CAPTAIN NEMO—"Our engraving is a view of a partly submerged Propeller Torpedo Vessel, proposed by James Nasmyth, of Patricroft, England, for destroying large ships of an invading fleet. The entire mass of the vessel (mortar and all) is brought into play, and the great brass mortar and shell explodes the instant it is crushed against the side of the enemy vessel. We must say that England seems afraid now of trusting in her wooden walls, and instead of terrifying her foes by keeping watch on their coasts, as she once did, she is keeping a sharp look-out for the defence of her own coasts by such water hogs as this of Mr. Nasmyth."



DUBIOUS COAST DEFENSE—the submarine mortar frigate, 1853

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

CLIMATE POLICY

Greenhouse Suits

LITIGATION BECOMES A TOOL AGAINST GLOBAL WARMING **BY MADHUSREE MUKERJEE**

A low-key case filed in a San Francisco court last August promises to be just the first ripple. The suit, now with the Friends of the Earth, Greenpeace and the cities of Boulder, Colo., and Oakland, Calif., as plaintiffs, seeks to force two government agencies to assess the total impact on climate of the projects they finance. Rather than treaties and regulations, litigation may soon be the weapon of choice for those concerned about human-induced global warming.

SEASIDE ESCAPE: The tiny Alaskan town of Shishmaref has voted to move inland to avoid the rising waters caused by climate warming.

LAURA RAUCH *AP Photo*

In the San Francisco case, the plaintiffs charge that in the past decade, the Overseas Private Investment Corporation (OPIC) and the Export-Import Bank of the United States (ExIm) have provided $32 billion in loans, insurance and loan guarantees for oil pipelines, oil drilling and other fossil-fuel endeavors that will ultimately result in the emission of 32 billion tons of carbon dioxide over the life of the projects. (All human activity currently emits about 24 billion tons of CO_2 a year.) In contrast, the agencies provided only $1.3 billion for renewable-energy projects during the same period. (A spokesperson for OPIC states in the agency's defense that OPIC-supported efforts are not "major contributors to global greenhouse gas emissions or climate change.")

The lawsuit does not attempt to cancel ongoing projects but asks only that OPIC and ExIm determine the "cumulative impact" on the climate of every future project. Such a review, asserts Jon Sohn of Friends of the Earth, is required by the National Environmental Policy Act.

The plaintiffs are confronted with many hurdles. To begin with, they will have to demonstrate that they face harm from global warming and, in particular, from the agencies' actions. The cities contend that their water supplies are in jeopardy. Boulder depends on runoff from mountain snow, but the snowpack at lower elevations has evaporated. Oakland fears that rising seas will salinate its underground aquifer. Other litigants include a coral-reef scientist who finds that his object of study is vanishing, and a couple who fear that their island home will be washed away.

Scientific uncertainties over such claims can be partly overcome by aggregating harm done over a large span of space and time, contends David Grossman, a recent graduate of Yale Law School and now a law clerk in Anchorage. In a paper to be published in the *Columbia Journal of Environmental Law*, Grossman argues that tort litigation over global warming—in which communities or states seek damages from oil companies, electric utilities and automobile manufacturers—

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

is entirely feasible. The main problem is causation—that is, proving that the defendant caused harm to the plaintiff. Statistics can help, he says, as when a town's residents can attribute an enhanced frequency of cancer to a nearby pesticide plant. Thus, a homeowner will probably not be able to show that the hurricane that destroyed his house was spawned by global warming, but the state of Florida may well prove that increased damage to coastal property over several years has a lot to do with climate change.

In truth, sea-level rise and greater frequency of storms are higher-order results of global warming, in that they would require several links in a causal chain to be proved. An easier case to make, notes Donald Goldberg of the Center for International Environmental Law in Washington, D.C., will simply be warming. In Alaska, for example, average temperatures have risen by about two degrees Celsius since 1970. Two coastal villages, Kivalina and Shishmaref, have suffered from erosion that Gunter Weller of the University of Alaska–Fairbanks attributes to three factors, all directly deriving from warming. Permafrost has thawed, causing houses to slide off suddenly muddy cliffs; sea ice has thinned, creating expanses of open water that rise up in ever higher storm surges; and glaciers are melting, leading local sea levels to climb (albeit very slightly). The townships must be relocated (at an estimated cost of more than $100 million), so they should stand a good chance of a court upholding a claim that they suffered damages because of global warming.

A plaintiff's next task would be to show that the defendants are meaningfully responsible. The issue will be vigorously fought, Grossman predicts. Environmentalists can estimate the quantity of greenhouse gases for which, say, a large oil producer is responsible. But calculating the fraction of warming is a far more contentious task, points out climatologist Stephen H. Schneider of Stanford University, because of the inherent uncertainty and variability of climate models. Even so, Goldberg holds that U.S. courts can solve the problem of apportioning blame: "It may take a few cases, but ultimately the courts will figure out a formula for assigning responsibility."

Shifting the cost of global warming to those who are disproportionately the perpetrators, Grossman argues, could make fossil fuels more expensive and thus force corporations to pay more attention to renewable energy. Environmental groups have been frustrated by the Bush administration's rejection of the Kyoto treaty and what Sohn describes as its tendency to "deny, deflect blame and delay" when it comes to issues involving global warming. So don't be surprised if "See you in court" becomes the environmentalist's new rallying cry.

Madhusree Mukerjee is based in Darien, Ill.

WHOSE FAULT IS IT, ANYWAY?

A question one can ask is why energy producers or automakers should be liable for emitting greenhouse gases, as opposed to consumers. Three answers:

- **There is no legal means of fixing responsibility on consumers, whose individual emissions are very small.**
- **Consumers arguably have little choice in the matter, given that infrastructure and product availability in most of the U.S. makes high use of fossil fuels unavoidable.**
- **Energy producers and other fossil-fuel corporations are in a better position than consumers to internalize the costs of climate change and to implement less damaging technology. The consumer might ultimately have to pay anyway, through higher fossil-fuel prices.**

HEALTH

Bad Medicine

WHY DATA FROM DRUG COMPANIES MAY BE HARD TO SWALLOW

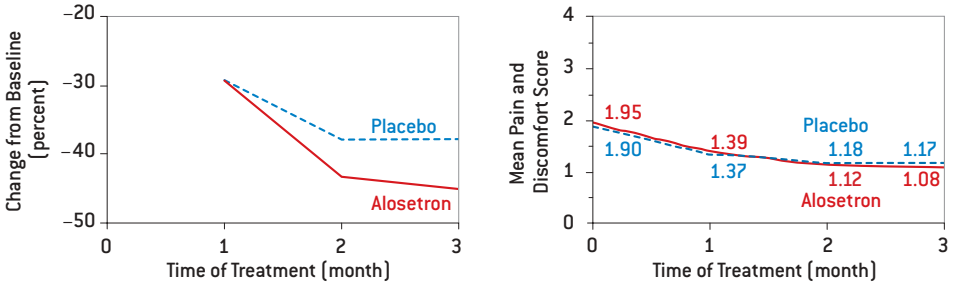
BY GUNJAN SINHA

Getting drugs on the market means playing games. So says Peter Lurie of Public Citizen, an interest group founded by Ralph Nader and based in Washington, D.C. Of course, it's the agency's mission to be leery. But lately pharmaceutical companies are giving groups like Lurie's more to be leery about. Drug firms now wield a great deal of control over their research, Lurie charges, and they are frequently manipulating their data or withholding unfavorable results entirely.

One of Public Citizen's latest battles is over a drug for irritable bowel syndrome (IBS). Three years ago the Food and Drug Administration approved Lotronex (alosetron hydrochloride), the first agent to treat the disorder specifically. As published in the *Lancet*, clinical trials in women revealed that 41 percent taking the drug felt some relief, as did 29 percent taking a placebo.

The data, Lurie insists, "are incredibly misleading." One figure, for example, plots percent change on one axis and time on the other. First, plotting percent change instead of absolute change makes the effectiveness of the drug appear large. Second, the graph omits data from the first month, during which the drug and placebo worked almost identically.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



BATTLE LINES: Lotronex (alosetron) appears to work much better than a placebo when percent change in pain severity from a baseline is plotted (*left*). Data replotted with absolute figures, done by the advocacy group Public Citizen, show much less of a difference (*right*). Both graphs appeared in the *Lancet*.

Public Citizen replotted the data using absolute values. The graph, which the *Lancet* published in a letter, better represents the drug's "marginal" efficacy, the group argues.

"I don't understand the accusation," responds Michael Camilleri of the Mayo Clinic in Rochester, Minn., who led the study. "Such presentation is standard and accepted in peer-reviewed scientific journals. The data clearly show that the drug was better than the placebo for months two and three."

But many observers believe that drug companies go too far. "It shouldn't happen in the scientific literature," insists Bob Goodman, founder of New York City–based No Free Lunch, which is focused on reining in the marketing ploys of drug companies. "Doctors should be able to decide the appropriateness of a drug. But how can they when drug companies leave out crucial information?" he asks. Goodman is referring to another common practice: excluding data.

Illustrating the point is the ongoing controversy over Cox-2 inhibitors, touted as a safer alternative to nonsteroidal anti-inflammatories such as ibuprofen. Sales of one, Celebrex, reached a whopping $3 billion in 2001. But last year the *Washington Post* revealed that Pharmacia, the drug's maker, had published just six months of results. Data for the next six months indicated that patients on Celebrex suffered complications such as stomach ulcers at the same rate as those taking older medications. This information became public only because one of the paper's reviewers happened to be on the drug's FDA review committee. Pharmacia says that the data for the last six months were too flawed to include them.

There are also rumblings that even though the FDA is aware of such practices, the agency is increasingly acting more favorably toward drug companies. The Lotronex story again provides the spark for that charge. After the drug hit the market in February 2000, the FDA assigned Paul D. Stolley of the University of Maryland to review the drug's side-effects profile. Stolley noticed a distressing pattern. Day after day he would see reports of patients being hospitalized, presumably because of Lotronex. "This for a disease that never leads to hospitalization, never perforates your colon and is not life-threatening," Stolley points out.

GlaxoSmithKline, the drug's maker, pulled Lotronex off the shelves in November 2000 after 49 reports of ischemic colitis and three deaths. A few months later, responding in part to requests from IBS advocacy groups, the company appealed to the FDA to bring the drug back. That move alarmed Stolley, who felt that the risks far outweighed the drug's marginal benefit. But when he spoke up, he was shut out. "FDA personnel were told not to discuss the case with me," complained Stolley, who had consulted for the FDA for the past 30 years. Others were opposed to the drug, but "they were intimidated," says Stolley, who now works for Public Citizen.

Some scientists argue that the FDA has become so chummy with the drug industry partly because of the Prescription Drug User Fee Act, passed in 1992. The act requires firms to pay the FDA almost $500,000 in total fees for each approved drug. Such fees account for almost half the agency's cost of reviewing drugs.

"I was shocked the FDA buckled even after they'd seen the obfuscation and the attempts to hide data. They seemed more comfortable working with the company than with their own staff," Stolley grouses. Lotronex is now back on the market. But only authorized doctors can prescribe it, and patients must sign an agreement stating that they fully understand the hazards. Here's hoping that for them, it is truly a risk worth taking.

Gunjan Sinha is based in Frankfurt, Germany.

JOHNNY JOHNSON

BAD NEWS AS NO NEWS

Academic researchers who carry out drug investigations may not always be aware of data manipulation. A recent study in the *New England Journal of Medicine* surveyed 108 U.S. medical schools and found that only 1 percent of contracts between industry and academic institutions required that every researcher of a multicenter study have access to all data. And less than 1 percent of contracts guaranteed that results would be published at all, ensuring that negative results are not published.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

DEFENSE

Storm before the Calm

CAN KNOCKOUT GASES REALLY BE NONLETHAL? BY DANIEL G. DUPONT

Last November 4 the Naval Studies Board of the National Research Council issued a report calling on the U.S. to increase its research into "calmatives," drugs that could be used to control and sedate unruly or hostile groups of people. Whereas most of the board's research had been finished a year earlier, the report was especially timely: nine days before, Russian troops had used a gas to subdue Chechen rebels in an attempt to rescue the 700 hostages they were holding in a Moscow theater. The gas—actually a nebulized aerosol said to contain fentanyl, an opiate used as an anesthetic—killed more than 100 hostages.

The U.S. looked into calmatives in the 1980s and 1990s, but the development of many types of chemical agents slowed or stopped in the wake of the Chemical Weapons Convention, ratified in 1997. The rise of terrorist activity throughout the world has led many military experts to believe that some kind of knockout gas would be helpful. Andrew Mazzara, a retired U.S. Marine colonel who heads the Institute for Emerging Defense Technologies at Pennsylvania State University's Applied Research Laboratory, states that the Russian example highlights a need for "more research rather than less" into nonlethal means of incapacitating hostage takers.

Even before the Naval Studies Board, the Penn State lab had investigated nonlethal weapons and concluded that such calmative gases could work safely. Researchers led by Joan Lakoski, now at the University of Pittsburgh, reviewed the medical literature on pharmaceutical agents that produce "a calm state." Ideally, according to the investigators, an effective calmative would be easy to administer and be adaptable for use in a variety of forms, fast-acting but short-lived, and reversible. After examining more than 7,800 articles and other references, the Penn State team declared in an October 2000 report that "the wide variety of drug classes and specific agents" that they studied "serve to underscore that the development and use of nonlethal calmative techniques is achievable and desirable."

The Penn State authors identified many compounds that have a "high potential for consideration" as nonlethal agents: sedative-hypnotic agents, anesthetic agents, muscle relaxants, opioid analgesics, anxiolytics, antipsychotics and antidepressants. But they singled out several major classes, two of which are convulsants and "selected drugs of abuse," including certain "club drugs." They also pointed to two drugs deserving immediate attention: diazepam (Valium) and dexmedetomidine.

Despite advances, drug delivery "remains a key issue in the development of calmative agents as nonlethal techniques," the Penn researchers pointed out. The problem is one of dosage: when an incapacitating gas is pumped into the ventilation system of a building, as was the case in the Moscow theater, some recipients will inevitably receive more than others. An opiate such as fentanyl is particularly crude when used in this way because it has a small dosage window in which it is considered safe. Benzodiazepines, used to anesthetize and to treat anxiety and amnesia (Valium is one), are considered more promising but do not act as fast.

For these reasons, a nonlethal and effective knockout gas is a myth, maintains Elisa Harris, a researcher at the University of Maryland and a former National Security Council staff member. "I just can't see how [such a gas] is technologically feasible," she says. "In decades and decades of research, it's never materialized." Harris and other opponents argue that knockout gases cannot be described as nonlethal—they will kill some of the people they are intended to save. James Cottrell, president of the American Society of Anesthesiologists, believes it would be "almost impossible" to develop an anesthetic gas that won't kill.

One way to reduce casualties is to combine the use of a gas with postexposure treatment. Doctors in Moscow were reportedly not aware of what ailed the rescued hostages, which stymied their efforts to treat them. Russian authorities denied the charge,

PRESCRIPTION FOR PACIFICATION

According to the Naval Studies Board, the U.S. worked on calmatives in the 1980s and 1990s at the army's Edgewood Chemical and Biological Command in Maryland. Moreover, it states that the use of calmatives has been discussed numerous times during meetings held by the Office of the Secretary of Defense and the Joint Staff. In May 2000 the Pentagon reportedly started at least one effort to research chemical immobilizing agents.

Candidate compounds:

- **Benzodiazepines**
- **Alpha$_2$-adrenoreceptor agonists**
- **Dopamine D3 receptor agonists**
- **Selective serotonin reuptake inhibitors**
- **Serotonin 5-HT$_{1A}$ receptor agonists**
- **Opioid receptors and mu agonists**
- **Neurolept anesthetics**
- **Corticotropin-releasing-factor receptor antagonists**
- **Cholecystokinin B receptor antagonists**

DMITRY LOVETSKY *AP Photo*

GASSED VICTIM is carried by a Russian officer after a raid to free hostages in a Moscow theater on October 26, 2002.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

saying that antidotes were prepared and used.

In the end, determining whether a calmative gas can be made safe and effective depends on how those criteria are defined. Whereas the gas used in Moscow killed more than 100 of the hostages, it contributed to the rescue of six times that many. Alan Zelicoff, a senior scientist at Sandia National Laboratories, remarks that "it might be nice to have something other than high-speed lead, chemical explosives and other lethal means to quell riots or even deny terrorists their targets."

Hostage negotiations should be tried first, although in the case of the Moscow incident, a peaceful end seemed unlikely. As Penn State's Mazzara notes, without the use of calmatives, such no-win situations might "very possibly lead to more tragic results."

Daniel G. Dupont, a frequent contributor, edits InsideDefense.com, an online news service, from Washington, D.C.

CANCER

T Cell Triumph

IMMUNOTHERAPY MAY HAVE FINALLY TURNED A CORNER **BY DIANE MARTINDALE**

Immunotherapy for cancer is a targeted treatment that uses a patient's own immune cells to attack and destroy tumors. Highly touted when it was conceived in the early 1980s, the approach has met with little success. Now researchers think they may have gotten over the hump: they have successfully treated several cases of a deadly skin cancer with immune cells taken from the patients, grown in large numbers in the laboratory and then given back to them. "We can now repopulate the body's immune system with cells that fight the cancer," says Steven A. Rosenberg of the National Cancer Institute, who pioneered immunotherapy.

The idea is to exploit a subset of T cells, the so-called tumor-infiltrating lymphocytes (TILs), found deep inside cancerous tissue. These killer T cells attack the rapidly dividing cells and provide a natural protection against cancer. But the body seldom makes enough to keep the disease in check.

Rosenberg first isolated and grew TILs and gave them to patients in the 1980s, in a process called adoptive T cell therapy. Although the T cells retained their antitumor properties, they did not proliferate or survive long enough in patients to kill their tumor cells. The recent success came when Rosenberg's team altered its method in two crucial ways. First, the scientists improved the way antitumor T cells are generated. TILs were isolated from multiple samples of each patient's tumor and grown in the lab. The group then tested up to 50 different samples against each patient's cancer cells and chose the most reactive T cells to expand and reinfuse into the patients. Previously, cells were simply extracted from the tumors without any type of selection.

Second, the researchers changed the way patients are prepared before the treatment. This time subjects underwent robust chemotherapy to wipe out their immune systems temporarily and thereby make room for the incoming tumor-killing T cells. The procedure may have removed suppressor cells (made by the immune system or the tumor), which prevent T cells from proliferating, Rosenberg says. After the reinfusion, patients received repeated doses of interleukin 2, a potent immune system hormone that stimulates the growth of T cells.

T CELL (*yellow*) attacks a cancer cell.

The study relied on 13 individuals with advanced metastatic melanoma, a skin cancer that eventually spreads to other organs. The patients, who had exhausted all other treatments, including surgery, received on average 80 billion of their own TILs—enough to give

EYE OF SCIENCE *Photo Researchers, Inc.*

A TREATMENT IN WAITING

Despite the recent success, immune cell therapy is still highly experimental. Side effects were serious in some cases: they included vitiligo (white patches of skin where normal pigment cells were attacked by the tumor-infiltrating lymphocytes) and opportunistic infections. This is not like a drug you can just pull off the shelf. "Every cell we give is basically a different drug because it's unique to that patient. And every patient has a different kind of tumor," says Steven A. Rosenberg of the National Cancer Institute, who is still trying to understand why the therapy works in some and not in others. Rosenberg thinks it will be at least two years before the therapy is ready for other types of cancer patients.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

them a new immune system. As of December 2002, 10 of those subjects were still alive: six had major remissions of their cancer, and four had some of their tumors shrink.

Analysis of patients' blood and tumor samples showed that the TILs multiplied and then attacked the tumor tissue. "In the past when we transferred cells, maybe 1 or 2 percent survived," Rosenberg explains. "Now we have 80 percent that survive for months, and when that happens the cancer disappears."

"The good news about Rosenberg's work is that as a proof of principle, it's extraordinary," says Robert A. Figlin, an oncologist at the University of California at Los Angeles School of Medicine. "The bad news is that it's not easily extrapolated to a large group of patients." Moreover, "we are asking a lot of these T cells to treat patients with very large tumor burdens," says Cassian Yee, an immunologist who has developed a similar T cell transfer therapy at the Fred Hutchinson Cancer Research Center in Seattle. "The T cell therapy might be more effective with smaller tumors and with repeated treatments over time."

According to Figlin, the key to immunotherapy is selecting the right patients. "There will be a smaller number of patients that have a higher response, and not the other way around," he explains. "That's the reality until we understand the subtleties of the immune response."

Diane Martindale is based in Toronto.

PHYSICS

Scaled-Up Superposition

SUPERSIZING SCHRÖDINGER'S CAT—BY A BILLION TIMES **BY CHARLES CHOI**

Cats may have nine lives, but only Schrödinger's can be both alive and dead at the same time. The quirky laws of quantum mechanics suggest that objects can literally exist in two states or places simultaneously until perturbed in some way, after which they collapse out of this "superposition" to just one outcome. Physicists have created such Schrödinger's cats before, usually in the guise of a single particle—a photon or electron. That's because the bigger the "cat," the harder it is to keep it undisturbed, a necessary condition for preserving the superposed state.

Physicists have come up with a scheme they think will produce a Schrödinger's cat billions of times larger than before. That would make it about the size of a feline cell—still a speck to human eyes but gigantic on the quantum scale. Roger Penrose of the University of Oxford originally conceived an experiment in space involving satellites, but collaborator Dik Bouwmeester of the University of California at Santa Barbara realized that a copycat version could be done on a tabletop, perhaps in three to five years as technology improves.

The setup, a kind of interferometer, monitors two paths that a photon of light can take. A photon is directed toward a beam-splitting crystal, which gives the light an equal chance of going down one of two paths, both capped with reflective cavities. The photon travels into either cavity and bounces around inside for a while. It then eventually leaks out to head back to the beam splitter, where it is reconstituted for detection. A photon will enter a superposition of traversing both paths simultaneously.

But one of the cavities is crucially different—one of its mirrors is mounted on an oscillating arm. Similar to a cantilever in atomic force microscopes, it would be sensitive enough to detect the push felt by the mirror. The quivering mirror would end up being in a superposition for about a millisecond because it was coupled to the photon. This superposition would appear as an interference pattern formed by the photon traveling two paths.

The requirements for this experiment, developed with physicists William Marshall and Christoph Simon, both at Oxford, are exquisitely sensitive. The mirror has to be minuscule to be jostled by a photon—maybe 10 microns thick (about a tenth the width of a human hair) and five billionths of a gram in weight. Temperature must be kept a few millionths of a degree from absolute zero, to keep all vibrations to a minimum. Ultrahigh vacuum must be maintained to make sure a stray atom doesn't knock the arm askew. To-

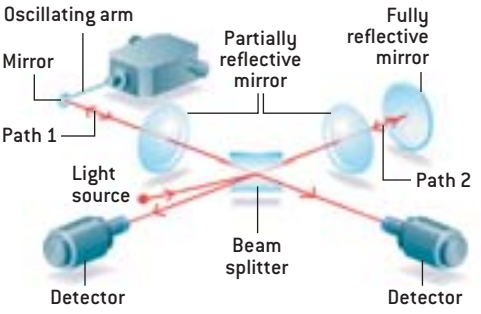
STAKING A SUPERPOSITION

Roger Penrose of the University of Oxford conceived the experiment described in the main text because he thinks that the very fabric of existence forbids large objects from remaining in superposition for long. If something exists in two places at once, it would result in two different structures of space and time, he says. Such a blister in reality represents an energy uncertainty; the larger the blister, the shorter the amount of time each can stay apart. Whereas electrons could exist in superposition for millions of years, something the size of a dust mote would exist for just a second or so.

The proposed experiment with mirrors won't settle the question: it would need 100,000 times more mass to reach the regime in which Penrose expects to see this cutoff in Schrödinger's cat size. An experiment involving long distances—such as orbiting satellites—may be needed.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

day's technology can meet both temperature and vacuum conditions, but such a tiny mirror on an equally small arm challenges existing fabrication techniques. Bouwmeester suggests that in the future one could make this mirror on a carbon nanotube, a small but incredibly strong rod that researchers are still trying to perfect.



GIANT QUANTUM CAT could be made if a photon is directed to a beam splitter, giving it two paths to follow. The photon enters a superposition of traversing both paths—and takes the mirror on the oscillating arm with it. The detector records the superposition as an interference signal.

"I would be quite surprised if a decade or so from now the experiment had not been done," comments quantum physicist Paul Kwiat of the University of Illinois at Urbana-Champaign. "Technology has a wonderful tendency to improve, despite the aphorism 'They don't make 'em like they used to.'"

Bouwmeester says that creating a large superposition could improve quantum computers, which rely on particles in superposition to represent 0's and 1's simultaneously. The proposed experiment, if successful, could help solve the problem of keeping these quantum cats trapped in superposition—and without the scratches, either.

Charles Choi is based in New York City.

SECURITY

Nothing but Net

HOW NOT TO BREAK THE SAFETY BARRIER

BY PHIL SCOTT

World War II footage has that familiar black-and-white scene: a heavily damaged war bird lands out of control on the stern of a straight-deck carrier and crashes into a steel cable net, which prevents it from ramming into aircraft waiting to take off. Volunteer firefighter Matthew Gelfand was watching one such documentary in 1993 when a lightbulb went on above his head. He had heard about an accident in which a car struck a train, then another vehicle whizzed past the crossing and hit a firefighter. "If a carrier could catch a plane with a net, why not a car with a net?" Gelfand wondered.

BETTER THAN A SPEED BUMP: A retractable net can stop vehicles without damaging them.

The result is GRAB, for ground retractable automobile barrier. Essentially, it is a tennis net made from Kevlar strips, with two metal stanchions on either side. Remote sensors or a manual push button shoots the net up from a two-inch-wide recess in the ground in as little as three seconds. As the vehicle hits the net, the energy is absorbed by pistons in the stanchions and the net—not unlike the barriers on the WWII aircraft carriers, whose nets had cables that folded down onto the deck and were connected to energy-absorbing stanchions.

Gelfand, who received $650,000 from the state of New York to develop GRAB through his new company, Universal Safety Response, envisions the system installed not only on railroad crossings but also at tunnels, bridges and security gates on government buildings. During tests, the net could stop a 1,800-pound automobile traveling at 45 miles an hour in just 13 feet. The quick stop did not inflict much damage to the vehicle.

Last December the first GRAB was installed at a fitting location: the security entrance to the USS *Intrepid*, a WWII-era aircraft carrier converted to a floating aviation museum docked on New York City's Hudson River.

Phil Scott is based in New York City.

SAMUEL VELASCO (*top*); UNIVERSAL SAFETY RESPONSE, INC. (*bottom*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COMPUTERS

Taking the Heat

That burning sensation on the thighs may become a thing of the past for laptop computer users. Sandia National Laboratories researcher Michael Rightley has devised a way to pipe computer heat out the side. He developed "smart" heat pipes, made from 60-micron-deep channels etched in copper. The self-contained system relies on methanol in the tiny tubes. Heat from a chip or circuit board turns the liquid to gas, which moves warmth to the laptop edge, away from the lap. Once the gas cools, it condenses and travels back to its start point via capillary action. Rightley expects the method to replace laptop heat sinks, which are chunks of metal, affixed next to the source, that can handle up to 100 watts of heat per square centimeter. Today's circuits throw off about half as much, but future chips will run hotter and may require liquid cooling. The research will appear in *Microelectronics Journal*.

—*Tariq Malik*

ASTRONOMY

Water or Not?

The debate over the likelihood of liquid Martian water flows on. Researchers at the University of Colorado at Boulder and NASA conclude that Mars has generally been cold, dry and inhospitable. Water flowed only briefly in the past, they posit, when asteroids crashing into Mars billions of years ago unleashed scalding rains for decades at a time. They calculate that the impact of a body 250 kilometers wide would have delivered energy equivalent to 100 billion megatons of TNT to the planet, melting exposed polar ice and injecting enough water into the atmosphere to rain out 16 meters of precipitation. Life probably wouldn't have had time to evolve under such brief deluges, according to their report in the December 6, 2002, *Science*. A more optimistic argument for the presence of running water on Mars came in a presentation at the American Geophysical Union, also in December. Scientists from the University of Arizona argued that some of the dark streaks on the planet's surface might be caused by current hydrological activity. Very briny water, they say, could exist as a liquid at the low temperatures and pressures on Mars's surface and flow down slopes, leaving streaks with telltale features in its wake.

—*Sarah Graham*

GEOCHEMISTRY

Fire and Ice

Ice may seem an unlikely fire starter, but John Maclennan of the Paris Geophysical Institute and his colleagues beg to differ. They say that ancient volcanoes in Iceland became suddenly more active because of the abrupt meltdown of kilometer-thick ice sheets that covered the island until about 10,000 years ago. Free of the ice's weight, the land popped up and relaxed pressure on the hot mantle rocks below. The team's analysis of massive lava flows from that period provides the first solid evidence that this pressure drop could cause mantle rocks to melt and rise to the surface. The flows—whose compositions indicate that they came from the mantle rather than the shallower crust—reveal a 30- to 100-fold jump in eruption rates for the 1,500 years following the deglaciation. The new report appears in the November 5 G^3 (*Geochemistry, Geophysics, Geosystems*).

—*Sarah Simpson*

VOLCANOES ROAR as glaciers recede. This particular subglacial eruption occurred in Grimsvoetn, Iceland.

HRAFNSSON GISLI EGILL *Corbis Sygma*; ILLUSTRATION BY MATT COLLINS

DATA POINTS: OIL IN WATER

The fuel released last November off the coast of Spain by the tanker *Prestige* could have long-lasting effects. In 2000 Christopher M. Reddy of the Woods Hole Oceanographic Institution and his colleagues drilled a 36-centimeter-long core in a West Falmouth, Mass., marsh, near where the barge *Florida* ran aground in 1969 and spilled its oil. The team found that petroleum contamination still persists there, more than 30 years later.

Contaminated surface sediment detected, in milligrams per gram of soil, in 1976: 5.7

In 2000, at the surface: 0
In 2000, between 12 and 16 centimeters down: 4 to 8

Number of liters spilled by *Florida*: 700,000
Number of liters spilled by *Prestige*: 10 million

SOURCE: Environmental Science and Technology, *November 15, 2002. Figures are approximate; contamination data from 2000 may be slightly higher because of improved detection sensitivity.*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

ANTHROPOLOGY

The Olmec's Write Stuff

Recently discovered artifacts—plaque fragments and a seal—contain intriguing scripts that may be remnants of the first written language in the New World. The pieces, found near the Gulf coast of Tabasco, Mexico, belonged to the Olmec people and date to 650 B.C. The cylindrical seal shows a bird with symbols coming out of its beak, suggesting that the glyphs were spoken. The artifacts' discoverers think the bird is saying "King 3 Ajaw": the Olmec used "3 Ajaw" to refer both to a day of the sacred, 260-day calendar and to the king born on that day. The script predates other Mesoamerican writing by at least 250 years and is the basis for subsequent Mesoamerican writing, including that of the Maya, the researchers say. Other anthropologists, however, argue that the symbols could simply be drawings, rather than representations of speech. The artifacts are described in the December 6, 2002, *Science*. —*Philip Yam*

WRITTEN EVIDENCE: Olmec seal (*left*) has an etching of a bird apparently saying "King 3 Ajaw."

BIOLOGY

Regenerating the Heart

Scarring prevents human hearts from repairing themselves, but a common aquarium dweller now appears to hold a secret remedy. Howard Hughes Medical Institute investigators Kenneth D. Poss, Lindsay G. Wilson and Mark T. Keating found that the zebrafish can naturally regenerate its own heart. Two months after the surgical removal of 20 percent of the hearts of adult fish, the vital organs had recovered their natural size and were beating properly. Under a microscope, the researchers could see that scar tissue clotted the wound initially, but proliferating muscle cells soon took over the healing process. Future exploration of the fish's regeneration-promoting genes—many of which are shared by humans—could lead to strategies for the scar-free repair of human hearts. This work appears in the December 13, 2002, *Science*. —*Sarah Simpson*

QUICK FIX: Zebrafish can mend their broken hearts. The inset shows DNA (*green*) that signals the successful regeneration of muscle cells (*red*).

BRIEF POINTS

- **The first high-quality draft of the mouse genome is now available. The rodent has about 30,000 genes, 99 percent of which have counterparts in human DNA.**

  Nature, December 5, 2002

- ***Helicobacter pylori*, the bacterium that causes stomach ulcers, uses hydrogen as an energy source, rather than carbohydrates, as most bacteria do.**

  Science, November 29, 2002

- **Freshwater flows from Arctic rivers have increased with global warming and may affect the oceans' deep circulation, resulting in a cooler northern Europe.**

  Science, December 13, 2002

- **Hazards of modern security: A man being treated with radioactive iodine was strip-searched twice after setting off radiation detectors. Such patients should carry a letter and a 24-hour telephone number of the physician in charge.**

  Journal of the American Medical Association, December 4, 2002

CHRISTOPHER VON NAGY (*top left and right*); MARK SMITH *Photo Researchers, Inc.* (*bottom*); KENNETH D. POSS, LINDSAY G. WILSON AND MARK T. KEATING (*bottom inset*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

BY THE NUMBERS

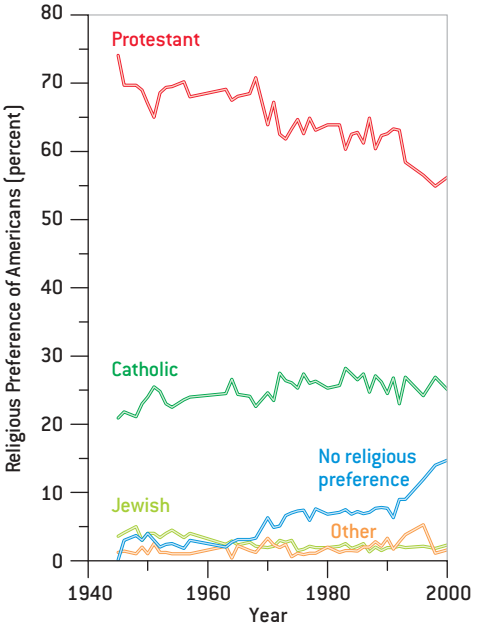
Religion in America

CHURCH ATTENDANCE HAS DIPPED, BUT FAITH REMAINS STRONG

BY RODGER DOYLE

Not long ago many believed that the spread of science and education would cause religion to wither, but although churchgoing has diminished, Americans generally retain their religious affiliations. Church attendance in the U.S. is higher than in any European country except Ireland and Poland [see By the Numbers, July 1999].

Since at least the end of World War II, Protestantism has declined, reflecting a weakening of mainline denominations. A likely cause may be the lower fertility seen since the early 20th century, when women from these denominations became active in the family-planning movement. In comparison with evangelicals, who emphasize saving souls, mainline Protestants have been less active in recruiting new members. Despite the decline, members of the "Protestant establishment" churches—Episcopalians, Congregationalists, Presbyterians, Quakers and Unitarians—continue to hold positions of power in business, government, white-collar professions and the arts far out of keeping with their numbers. Although their importance, as measured by listings in *Who's Who*, fell during the 20th century, in the early 1990s they still had more entries than Catholics and Jews combined.



SOURCE: National Opinion Research Center, General Social Survey

Despite a long-standing schism between church doctrine and lay practice, particularly on abortion and contraception, Catholicism has managed to maintain the allegiance of about a quarter of Americans over the past five decades. That is in part a result of higher levels of natural increase and the reinforcing effect of Catholic education. According to social scientist Father Andrew M. Greeley of the University of Chicago, Catholics remain loyal because they are powerfully attracted by the experiences, images and traditions of the Church. The pedophile priest scandal, however, has taxed that loyalty: a Gallup poll in June 2002 reported that 22 percent of Catholics said that they questioned whether they would remain in the fold.

The proportion of those adhering to Judaism has declined since World War II, in part because of low fertility and because marriages outside the faith (aided in part by a shift from Orthodox toward Reform synagogues) frequently result in disaffiliation. Nevertheless, Judaism, at an estimated six million affiliates, remains the largest of the non-Christian religions, followed by Islam at 1.9 million, Buddhism at 1.5 million and Hinduism at about one million.

The 1990s saw a substantial increase in the proportion of Americans with no religious preference, mostly because of a shift in demographics, not a rise in religious skepticism. Young adults frequently disengage from religion when leaving the parental home but reengage after forming a family, but as a result of the recent trend toward marrying later in life, for many that reengagement hasn't happened yet. The percentage of adults raised with no religion rose from 3 to 6 percent over the past 30 years, but only about one third of those without a religious preference can be counted as nonbelievers.

Next month: Fundamentalism.

Rodger Doyle can be reached at rdoyle2@adelphia.net

RODGER DOYLE

SUNDAY SERMONS

Percent of U.S. adults surveyed saying they attend church:

At least weekly: 24 to 30

Less than once a week: 54 to 58

As a group, respondents typically overstate their attendance by up to 70 percent.

FURTHER READING

The Demographic Imperative in Religious Change in the United States. Michael Hout, Andrew Greeley and Melissa J. Wilde in *American Journal of Sociology*, Vol. 107, No. 2; 2001.

Persistence and Change in the Protestant Establishment, 1930–1992. James D. Davidson, Ralph E. Pyle and David V. Reyes in *Social Forces*, Vol. 74, No. 1; 1995.

Why More Americans Have No Religious Preference: Politics and Generations. Michael Hout and Claude S. Fischer in *American Sociological Review*, Vol. 67, No. 2; 2002.

Yearbook of American and Canadian Churches 2002. Edited by Eileen W. Lindner. Abingdon Press, 2002.

Handbook of Denominations in the United States. 11th edition. Frank S. Mead and Samuel S. Hill. Abingdon Press, 2001.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Take a Number

Toilet reservations afford a glimpse of the world of business-method patents By GARY STIX

U.S. Patent 6,329,919 covers "an apparatus, system and method for providing reservations for restroom use." In 2001 the Patent and Trademark Office (PTO) deemed IBM's electronic toilet queue worthy of a patent, thus fulfilling the office's constitutional edict to "promote the progress of science and useful arts." You don't have to be a legal scholar to wonder whether IBM deserved to be given exclusive rights for nearly 20 years to stop others from determining who gets to go next. The IBM restroom patent joined Amazon.com's patent on one-click ordering and countless lesser-known issuances from the PTO as members of the infamous subclass of intellectual property known as business-method patents.

JENNIFER KANE

A 1998 decision by the U.S. Court of Appeals for the Federal Circuit, *State Street v. Signature Financial Group,* opened the floodgates by throwing out a long-standing judicial rule against business-method patents and giving a boost to the gold-rush environment of the Internet boom—a dot-com company that sells dog food on the Web might now decide to apply for a patent on its method of doing business. After *State Street,* the number of applications for class 705 patents (patents on business-related data-processing methods and technologies) soared from 1,340 in the 1998 federal fiscal year to a peak of 8,700 in fiscal 2001. Ascertaining what a business-method patent is remains part of the problem. Among the difficulties: not all business-method patents fit into class 705, and some of them predate the *State Street* decision.

The controversy surrounding intellectual property on business methods has rivaled or exceeded disputes on software patents (a related area) and on gene patents. Critics contend that many business methods fail to meet the standard that something being patented should not be obvious. They also assert that extending property rights to broad areas beyond the sphere of technological invention can unfairly restrict economic activity. "Should any one company be permitted to own the concept of frequent-flyer miles for 20 years?" asks Brian Kahin, director of the University of Maryland Center for Information Policy. Kahin is one of a number of intellectual-property scholars who think these patents should be eliminated or severely restricted.

Business methods, which the appeals court held should be patentable as long as they are "useful, concrete and tangible," have made even the PTO squirm. In 2000 the office started requiring that each application evaluated in class 705 be reviewed by a second experienced examiner. Around 45 percent of applications filed in class 705 are granted, compared with about 70 percent for patents in all classes. Because of the hurdles imposed by the PTO, Chicago intellectual-property attorney Stephen Lesavich now counsels his clients to file applications with the office in a way that avoids having them classified as business-method patents. Applications for these patents actually dropped by an estimated 43 percent in fiscal 2002, fallout in part from the dot-com bust.

Even IBM, the company that has garnered more patents of all types than any other for the past decade or so, had second thoughts about the one covering the toilet queue. After Patent Commissioner James Rogan ordered a reexamination last year, IBM relieved itself of the patent. "The company known as Big Blue does not also want to be known as Big Loo," noted the English *Guardian*. According to an IBM spokesperson, the corporation found that the patent did not meet its quality standards and decided to abandon it. A similar review might be counseled elsewhere for other business-method patents, such as those for cutting hair, conducting an auction or privatizing government. SA

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Reverse-Engineering Clinical Biology

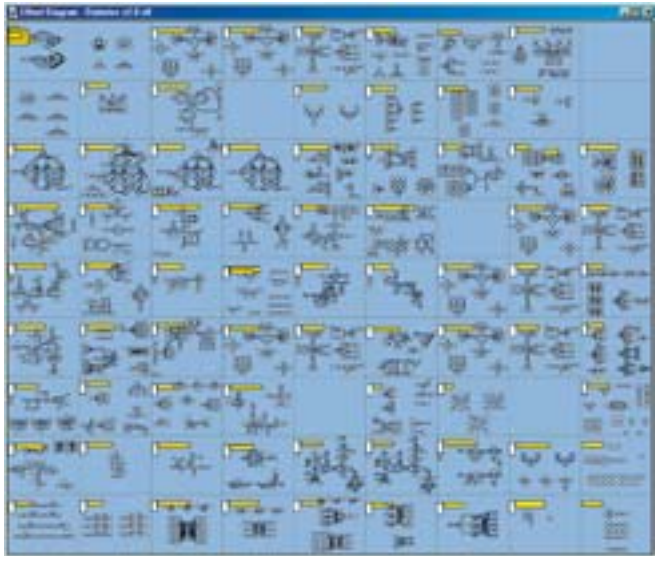
A peacetime dividend yields drug trials on virtual patients

By GARY STIX

During the Star Wars years of the 1980s, Tom Paterson worked at a defense think tank creating elaborate mathematical models to help military commanders quickly decide which weapons to deploy to counter incoming missiles. Inputs from hundreds of sensors had to be combined to generate a consummate picture of events that would be unfolding in a matter of minutes, enabling the fateful choice about when to launch.

When the cold war ended, Paterson, like many defense engineers, tried to find a way to apply his skills elsewhere. He ultimately took on a task that made shooting down missiles seem pedestrian. A challenge faced by engineers in the Star Wars program—designing software to pick out critical targets despite an overload of data—carried over to simulations of how drugs work in the metabolic and immune systems that drive the most complex machine we know.

A few years later, after obtaining a master's degree in decision analysis from Stanford University, a mecca for modelers, Paterson went on to Strategic Decisions Group (SDG), a California consultancy that did studies for pharmaceutical companies about how to balance risks and payoffs in their overall drug-development portfolio. SDG decided to extend its expertise to help a corporate customer perform "data mining" on complex relational databases that tracked a patient's illness over an extended period. "When our clients went in to mine the data, they were able to pull out things that they already knew, but they weren't able to pull out anything that was particularly novel," Paterson recounts. A connection between smoking and severe periodontal disease, a link that Procter & Gamble had turned up in its database, wasn't exactly a revelation.



METABOLIC PATHWAYS, represented by the rectangles that make up the grid, form the basis for Entelos's model of a virtual diabetic patient.

COURTESY OF ENTELOS, INC.

For the answers, SDG turned to modeling. But models of biology are so complex that Paterson and his collaborators could have spent the rest of their careers on a single cell. So the team—led by engineers, not biologists—did not begin at the cell nucleus and work toward a computerized rendition of Einstein. Rather the group adopted a reverse-engineering strategy, disassembling a disease in the way that a Ford engineer might take apart a Toyota to find out what the competition was doing. The model would identify manifestations of the ailment and work back to the known biological pathways involved, while looking for new ones that had yet to be characterized. In the Procter & Gamble case, it started, for instance, with symptoms such as inflammation of the gums and then identified components of the immune system that contribute to periodontal disease. The modeling analysis suggested that the company should focus its search for fruitful drug targets on relatively overlooked areas of the innate immune system, which serves as an initial line of defense against a tide of invading pathogens.

In the mid-1990s Paterson and his team at SDG found that the continuing refinement of the technology was stymied by the nature of its relationship with clients.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

To produce the models, the engineers at SDG relied on the biological expertise provided by clients, but it was impossible to get the biologists they were working with to devote sufficient time to these projects. "We really wanted [to have] life scientists on the team who would do nothing else," Paterson remarks. "Not people who were trying to squirrel away a couple hours a week from their normal jobs." As a consulting firm, however, SDG was not about to hire a staff of biologists.

By 1996 Paterson, software maven Alexander Bangs and a few others came up with a solution. They broke away from SDG to form Entelos, a corporation whose name means "completely" in Greek. The work on periodontal disease and other models at SDG gave the start-up's engineers principles to apply to maladies ranging from asthma to diabetes. The elemental constructs of the model, synthesized from thousands of journal articles, are different immunological or metabolic states in the body: for example, the concentration of insulin in muscle tissue represents one of 700 such variables in the diabetes model. To ensure that it is not the collective delusion of a bunch of overcaffeinated engineers and biologists, the model is put through 350 tests comparing the outputs with actual clinical data to validate its accuracy.

Pharmaceutical companies enter collaborative research contracts with Entelos to run their drugs through the models. A virtual patient pops a pill by "swallowing" a detailed technical description of a drug candidate, which then gets "metabolized" (subjected to a series of differential equations) in digital organs ranging from the pancreas to the liver to the brain. The testing can demonstrate the effect of a drug based on weight, age, sex and degree of disease severity. If a drug makes it to market, Entelos often will receive royalties, which is somewhat unusual for an information provider.

The company has demonstrated that its PhysioLab system is more than a video game. The diabetes simulation, for one, began to emerge more than three years ago, when Johnson & Johnson signed up with Entelos to test the obesity model and wanted it extended to cover diabetes. Entelos did so by adding parameters such as hemoglobin A1c (a measure of average blood sugar over the past three months). Johnson & Johnson then tried to mimic the dosing of patients involved in a clinical trial for a still proprietary antidiabetes drug candidate that worked in a novel way and for which it had no clinical data in humans. When the dose of the

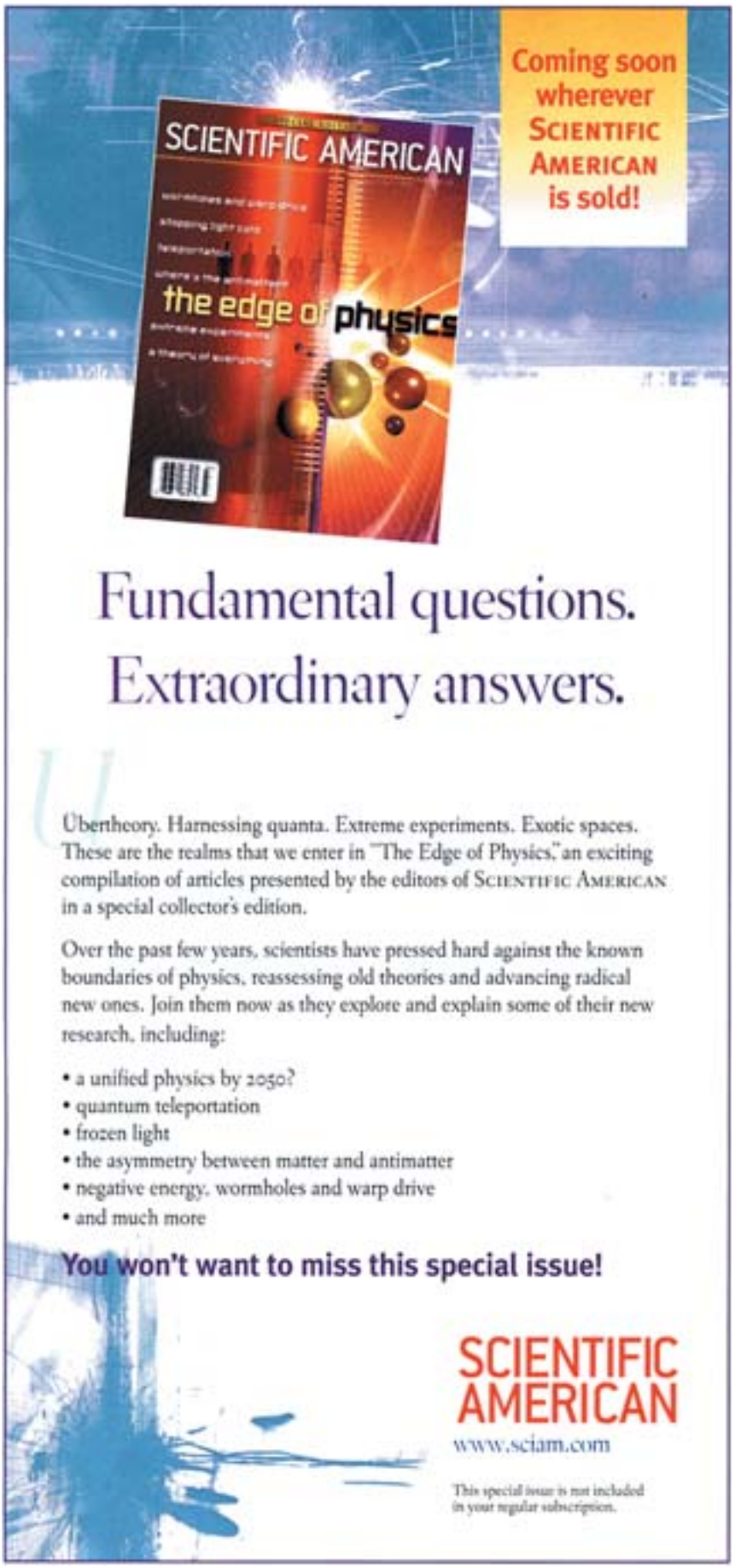
A virtual patient pops a pill by "swallowing" a detailed technical description of a drug candidate, which then gets "metabolized" in digital organs ranging from the pancreas to the liver to the brain.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

compound was administered in virtual patients in five gradual steps, very few differences in therapeutic or toxic effects were registered from the lowest to the highest dose. So, in human trials, the company went immediately to the highest dose, avoiding clinical trials for each lesser dose that had been simulated and cutting the number of patients it needed by two thirds. The models cannot do everything, though. There were not enough data in the scientific literature to build a satisfactory model of fat cell growth and differentiation, for instance.

Still, keen interest remains. Michael Jackson, a senior vice president at Johnson & Johnson Pharmaceutical Research and Development, imagines that the PhysioLab might eventually be coupled with wireless monitoring devices in clinical trial patients. Every day or two the devices would broadcast information, such as blood pressure and glucose levels, to update the models constantly. The progress in simulating diabetes prompted the American Diabetes Association to form a partnership last year with Entelos, an attempt to draw in drug manufacturers to use modeling to speed new treatments for diabetes.

The attention garnered by virtual patients has justified the early vision of Entelos's founders. Paterson remembers that when he visited pharmaceutical companies in the late 1990s, officials would comment that this type of technology would soon become obsolete; they thought the Human Genome Project would reveal all disease genes and lead directly to new therapies. But the current vogue for "systems biology," an attempt to go beyond the study of isolated genes and proteins, has lent support to the Entelos approach. This year the company will even supplement its simulations with real-life bench tops. It plans to construct a laboratory for cell-based testing at its new headquarters in Foster City, Calif. Thus, in silico biologists will adopt in vitro experimentation to get needed answers to questions that cannot be resolved by stringing together binary digits. SA



COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Psychic Drift

Why most scientists do not believe in ESP and psi phenomena

By MICHAEL SHERMER

In the first half of the 19th century the theory of evolution was mired in conjecture until Charles Darwin and Alfred Russel Wallace compiled a body of evidence and posited a mechanism—natural selection—for powering the evolutionary machine.

The theory of continental drift, proposed in 1915 by Alfred Wegener, was not accepted by most scientists until the 1960s, with the discovery of midoceanic ridges, geomagnetic patterns corresponding to continental plate movement, and plate tectonics as the driving motor.

Data and theory. Evidence and mechanism. These are the twin pillars of sound science. Without data and evidence, there is nothing for a theory or mechanism to explain. Without a theory and mechanism, data and evidence drift aimlessly on a boundless sea.

Until psi finds its Darwin, it will continue to drift on the fringes of science.

For more than a century, claims have been made for the existence of psi, or psychic phenomena. In the late 19th century organizations such as the Society for Psychical Research were begun to employ rigorous scientific methods in the study of psi, and they had world-class scientists in support, including none other than Wallace (Darwin was skeptical). In the 20th century psi periodically appeared in serious academic research programs, from Joseph B. Rhine's experiments at Duke University in the 1930s to Daryl J. Bem's research at Cornell University in the 1990s.

In January 1994, for example, Bem and his late University of Edinburgh parapsychologist colleague Charles Honorton published "Does Psi Exist? Replicable Evidence for an Anomalous Process of Information Transfer" in the prestigious review journal *Psychological Bulletin*. Conducting a meta-analysis of dozens of published experiments, the authors concluded that "the replication rates and effect sizes achieved by one particular experimental method, the ganzfeld procedure, are now sufficient to warrant bringing this body of data to the attention of the wider psychological community." (A meta-analysis is a statistical technique that combines the results from studies to look for an overall effect, even if the results from the individual studies are insignificant; the ganzfeld procedure places the "receiver" in a room with Ping-Pong ball halves over the eyes and headphones over the ears playing white noise and the "sender" in another room psychically transmitting visual images.)

Despite the evidence for psi (subjects had a hit rate of 35 percent, when 25 percent was predicted by chance), Bem and Honorton lamented that "most academic psychologists do not yet accept the existence of psi, anomalous processes of information or energy transfer (such as telepathy or other forms of extrasensory perception) that are currently unexplained in terms of known physical or biological mechanisms."

Why don't scientists accept psi? Bem has a stellar reputation as a rigorous experimentalist and has presented statistically significant results. Aren't scientists supposed to be open to changing their minds when presented with new data and evidence? The reason for skepticism is that we need replicable data and a viable theory, both of which are missing in psi research.

Data. The meta-analysis and ganzfeld techniques have been challenged. Ray Hyman of the University of Oregon determined that there were inconsistencies in the experimental procedures used in different ganzfeld experiments (which were lumped together in Bem's meta-analysis as if they used the same procedures). He also pointed out flaws in the target randomization process (the sequence in which the visual targets were sent to the receiver), resulting in a target-selection bias. Richard Wiseman of the University of Hertfordshire in England conducted a meta-analysis of 30 more ganzfeld experiments and found no evidence for psi, concluding that psi data are nonreplicable.

Theory. The deeper reason scientists remain unconvinced of psi is that there is no theory for how psi works. Until psi proponents can elucidate how thoughts generated by neurons in the sender's brain can pass through the skull and into the brain of the receiver, skepticism is the appropriate response, as it was for continental drift sans plate tectonics.

Until psi finds its Darwin, it will continue to drift on the fringes of science. SA

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

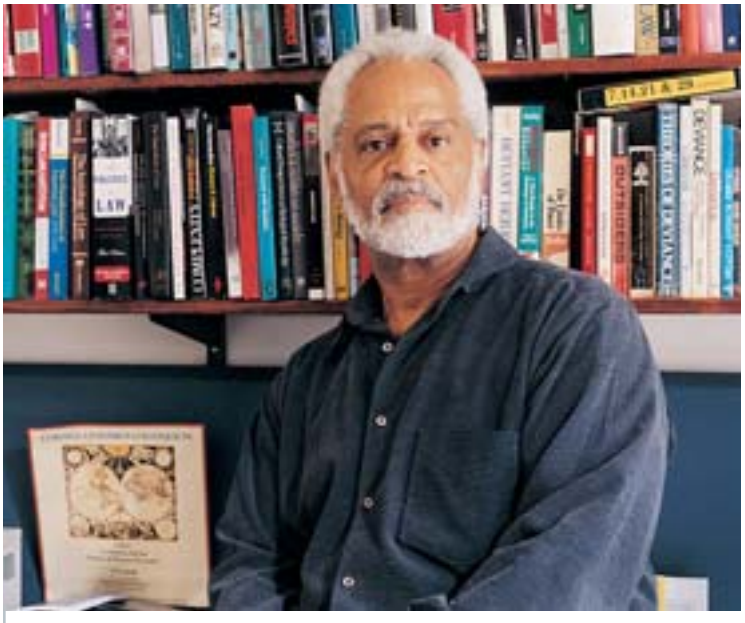
BRAD HINES

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The Reality of Race

There's hardly any difference in the DNA of human races. That doesn't mean, argues sociologist Troy Duster, that genomics research can ignore the concept By SALLY LEHRMAN

Race doesn't exist, the mantra went. The DNA inside people with different complexions and hair textures is 99.9 percent alike, so the notion of race had no meaning in science. At a National Human Genome Research Institute (NHGRI) meeting five years ago, geneticists were all nodding in agreement. Then sociologist Troy Duster pulled a forensics paper out of his briefcase. It claimed that criminologists could find out whether a suspect was Caucasian, Afro-Caribbean or Asian Indian merely by analyzing three sections of DNA.



ADAM FRIEDBERG

TROY DUSTER: THINKING ABOUT GENES

- **Grandson of Ida B. Wells-Barnett, newspaper publisher, muckraker and antilynching crusader.**
- **"The King of Coolocity," says Harry G. Levine of Queens College, City University of New York, because like a disciplined musician Duster combines seriousness, virtuoso skill, grace, balance and a relaxed playfulness in his work (he is a jazz aficionado).**
- **Current worry: "It is almost inevitable that a research agenda will surface to try to find patterns of allele frequencies and then create computer-generated profiles of different types of criminals."**

"It was chilling," recalls Francis S. Collins, director of the institute. He had not been aware of DNA sequences that could identify race, and it shocked him that the information can be used to investigate crimes. "It stopped the conversation in its tracks."

In large part thanks to Duster, Collins and other geneticists have begun grappling with forensic, epidemiological and pharmacogenomic data that raise the question of race at the DNA level. The NHGRI now routinely includes experts from the social disciplines to assist in guiding research priorities and framing the results for the public. "The complexities of the DNA sequence require not just simplistic statements about similarities between groups but a full appreciation of history, anthropology, social science and politics," Collins has realized. "Duster is a person that rather regularly gets tapped on the shoulder and asked for help."

The urbane 66-year-old Duster, who splits his time between appointments at the University of California at Berkeley and New York University, examines how the public absorbs news about genetics into existing beliefs and how those perceptions also shape the use of genetic sequencing, DNA probes and other molecular techniques.

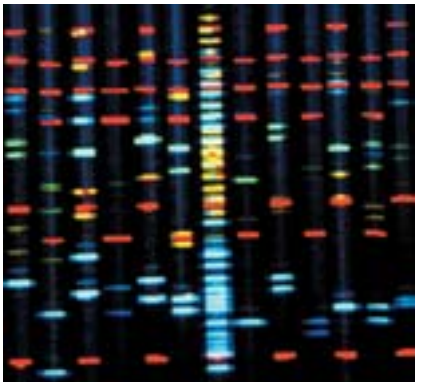
Those techniques have revealed that race is minor at the DNA level. The genetic differences between any two randomly selected individuals in one socially recognized population account for 85 percent of the variation one might find between people of separate populations. Put another way, the genetic difference between two individuals of the same race can be greater than those between individuals of different races—table sugar may look like salt, but it has more similarities with corn syrup.

But genetics cannot prove that race doesn't exist, Duster explains. No amount of logic will erase the concept or destroy the disparities that arise from it, because people use race to sort their social groupings and to de-

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

fine their social and economic interactions. Moreover, they do so in ways that have significant biological consequences. Duster recently helped to draft a 15-page statement for the American Sociological Association showing how race persists as a factor in disparities in health and other areas of life. "You cannot just get rid of the concept without doing tremendous damage to the epidemiologic research done so far," Duster says. African-Americans are three times as likely to die from heart disease, for example. "Blacks are redlined by banks, followed by department store security, pulled over by the police. This can produce hypertension," he points out. "It can give you a heart attack."

A new approach, gene clustering, avoids race by dividing according to medically important markers, such as genes for the enzymes necessary to metabolize drugs. But society will very likely re-create racial categories and rankings under the new terms, Duster predicts. And by failing to name the social context, this strategy gives base-pair differences undue emphasis at the expense of environmental influences. Race is a social reality, Duster observes, and he warns that science itself is a social institution susceptible to essentialist perceptions of race.



DNA PROFILES raise issues about race that sociologists such as Troy Duster must ponder.

Raised in poverty during the Great Depression by a mother from an upper-class family, Duster, whose father died when he was nine, grew up navigating between Chicago's tough streets and its privileged intellectual and civic parlors. He witnessed firsthand the complexities of social categories and learned to "code-switch" from one to another, much as he capably moves among sociology, anthropology and genetics now.

Duster started out as a journalist but quit in moral indignation when chided for failing to interview a trapped subway motorman waiting for a leg amputation. He turned to sociology and joined Berkeley in 1967, quickly developing a reputation for thought-provoking work on drugs and social policy. In the 1970s Duster was a familiar voice in National Institutes of Health committees reviewing grants for research on mental health and drug abuse. While sitting on a panel for President Jimmy Carter's Commission on Mental Health, he began to hear researchers speculate that drug addiction and mental illness were linked to genetic susceptibilities.

Duster found the conversations alarming. His book, *Backdoor to Eugenics,* aimed to stimulate public debate by showing how genetic-screening policies tended to reinforce the power structures already within society. Since then, he has pressed geneticists and molecular biologists to consider the social meaning that emerges from what they perceive as unbiased fact.

At first they resisted. As a member of the Ethical, Legal and Social Implications Working Group advising the agencies on human genome research, Duster urged the NIH and the Department of Energy to challenge *The Bell Curve,* the 1994 best-seller that argued that race correlated with intelligence. Government officials held up a response for eight months, convinced that the nonexistence of race at the genome level spoke for itself.

Duster, along with fellow committee member Dorothy Nelkin of New York University, highlighted the ways in which cultural context influences the application of medical and behavioral genetics. Now Collins is relying on Duster and other collaborators, such as University of Wisconsin molecular biologist Pilar Ossorio, to help explain why race must be acknowledged even if it is biologically inconsequential. "It's a tightrope between trying to rescue the importance and meaning of research on race without giving it a false reality," Duster says.

Indeed, although he maintains that race is significant in genetics, Duster insists it is misleading to reinscribe race as a definitive system to group people who share geographic origins and thus some genes. For one, concepts of race vary geographically as well as historically. The ethnic status of South Asians, for example, has changed over the past century in the U.S. and more often serves to define a political and cultural "other" than something biological. In 1920 Oregon granted citizenship to Bhagat Singh Thind of India during a ban on Asian immigration. But the U.S. Supreme Court disagreed, stating that even though Thind should be considered "Caucasian," he still wasn't "white." (Thind, who had joined the U.S. Army during World War I, managed to stay in the country, earn a Ph.D. and publish 15 books on metaphysics.)

Researchers have also advocated assessing health risks within ethnic groups based on inherited variations in just one DNA base pair. But such single-nucleotide polymorphism (SNP) profiles can be deceptive, Duster warns. Ethnic differences in drug metabolism or response to tobacco exist, but they appear to be minimal and depend strongly on the environment. The emphasis on DNA, he remarks, transforms health status into a biological inevitability, and it is tempting to use the same tools to profile criminality or intelligence at the genome level.

Specific variations in DNA can be linked to ancestral geographic origins, but those differences only occasionally offer a medically important clue. They fail to define any essential characteristics of a whole group. Race, itself a fluid idea, is part of the environmental context of the genome, Duster suggests. "Race is a relationship," he says. "When you talk about race as a relationship, it prevents anyone from giving it false meaning." SA

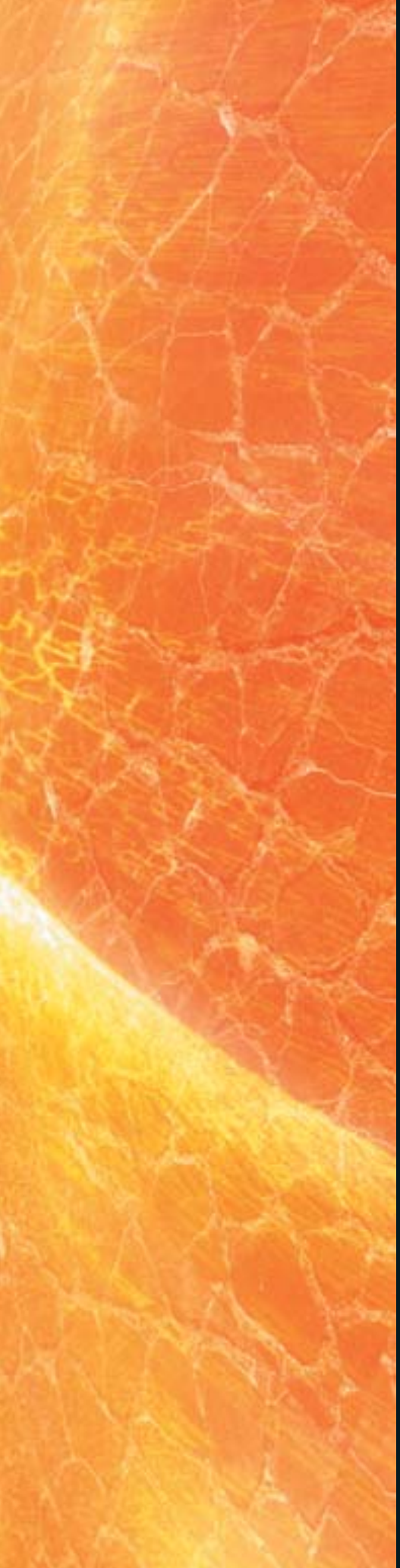
Sally Lehrman is a medical technology and health policy journalist based in San Francisco.

TEK IMAGE *Photo Researchers, Inc.*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

STARQUAKE ON A MAGNETAR releases a vast amount of magnetic energy—equivalent to the seismic energy of a magnitude 21 earthquake—and unleashes a fireball of plasma. The fireball gets trapped by the magnetic field.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



MAGNETARS

Some stars are magnetized so intensely that they emit huge bursts of magnetic energy and alter the very nature of the quantum vacuum

By Chryssa Kouveliotou, Robert C. Duncan and Christopher Thompson

On March 5, 1979, several months after dropping probes into the toxic atmosphere of Venus, two Soviet spacecraft, Venera 11 and 12, were drifting through the inner solar system on an elliptical orbit. It had been an uneventful cruise. The radiation readings on board both probes hovered around a nominal 100 counts per second. But at 10:51 A.M. EST, a pulse of gamma radiation hit them. Within a fraction of a millisecond, the radiation level shot above 200,000 counts per second and quickly went off scale.

Eleven seconds later gamma rays swamped the NASA space probe Helios 2, also orbiting the sun. A plane wave front of high-energy radiation was evidently sweeping through the solar system. It soon reached Venus and saturated the Pioneer Venus Orbiter's detector. Within seconds

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

the gamma rays reached Earth. They flooded detectors on three U.S. Department of Defense Vela satellites, the Soviet Prognoz 7 satellite, and the Einstein Observatory. Finally, on its way out of the solar system, the wave also blitzed the International Sun-Earth Explorer.

The pulse of highly energetic, or "hard," gamma rays was 100 times as intense as any previous burst of gamma rays detected from beyond the solar system, and it lasted just two tenths of a second. At the time, nobody noticed; life continued calmly beneath our planet's protective atmosphere. Fortunately, all 10 spacecraft survived the trauma without permanent damage. The hard pulse was followed by a fainter glow of lower-energy, or "soft," gamma rays, as well as x-rays, which steadily faded over the subsequent three minutes. As it faded away, the signal oscillated gently, with a period of eight seconds. Fourteen and a half hours later, at 1:17 A.M. on March 6, another, fainter burst of x-rays came from the same spot on the sky. Over the ensuing four years, Evgeny P. Mazets of the Ioffe Institute in St. Petersburg, Russia, and his collaborators detected 16 bursts coming from the same direction. They varied in intensity, but all were fainter and shorter than the March 5 burst.

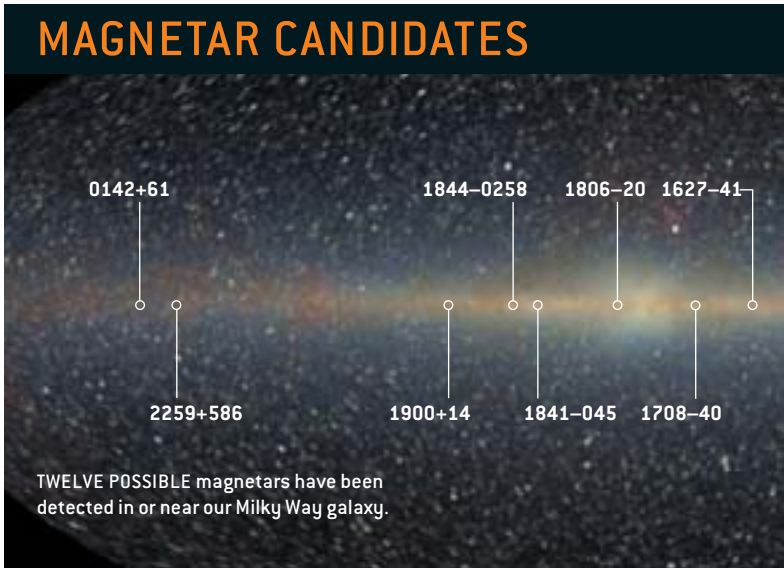
Astronomers had never seen anything like this. For want of a better idea, they initially listed these bursts in catalogues alongside the better-known gamma-ray bursts (GRBs), even though they clearly differed in several ways. In the mid-1980s Kevin C. Hurley of the University of California at Berkeley realized that similar outbursts were coming from two other areas of the sky. Evidently these sources were all repeating—unlike GRBs, which are one-shot events [see "The Brightest Explosions in the Universe," by Neil Gehrels, Luigi Piro and Peter J. T. Leonard; SCIENTIFIC AMERICAN, December 2002]. At a July 1986 meeting in Toulouse, France, astronomers agreed on the approximate locations of the three sources and dubbed them "soft gamma repeaters" (SGRs). The alphabet soup of astronomy had gained a new ingredient.

Another seven years passed before two of us (Duncan and Thompson) devised an explanation for these strange objects, and only in 1998 did one of us (Kouveliotou) and her team find compelling evidence for that explanation. Recent observations connect our theory to yet another class of celestial enigmas, known as anomalous x-ray pulsars (AXPs). These developments have led to a breakthrough in our understanding of one of the most exotic members of the celestial bestiary, the neutron star.

Neutron stars are the densest material objects known, packing slightly more than the sun's mass inside a ball just 20 kilometers across. Based on the study of SGRs, it seems that some neutron stars have magnetic fields so intense that they radically alter the material within them and the state of the quantum vacuum surrounding them, leading to physical effects observed nowhere else in the universe.

Overview/*Ultramagnetic Stars*

- Astronomers have seen a handful of stars that put out flares of gamma and x-radiation, which can be millions of times as bright as any other repeating outburst known. The enormous energies and pulsing signals implicate the second most extreme type of body in the universe (after the black hole): the neutron star.
- These neutron stars have the strongest magnetic fields ever measured—hence their name, magnetars. Magnetic instabilities analogous to earthquakes can account for the flares.
- Magnetars remain active for only about 10,000 years, implying that millions of them are drifting undetected through our galaxy.



Not Supposed to Do That

BECAUSE THE MARCH 1979 BURST was so bright, theorists at the time reckoned that its source was in our galactic neighborhood, hundreds of light-years from Earth at most. If that had been true, the intensity of the x-rays and gamma rays would have been just below the theoretical maximum steady luminosity that can be emitted by a star. That maximum, first derived in 1926 by English astrophysicist Arthur Eddington, is set by the force of radiation flowing through the hot outer layers of a star. If the radiation is any more intense, it will overpower gravity, blow away ionized matter and destabilize the star. Emission below the Eddington limit would have been fairly straightforward to explain. For example, various theorists proposed that the outburst was triggered by the impact of a chunk of matter, such as an asteroid or a comet, onto a nearby neutron star.

But observations soon confounded that hypothesis. Each spacecraft had recorded the time of arrival of the hard initial pulse. These data allowed astronomers, led by Thomas Lytton Cline of the NASA Goddard Space Flight Center, to triangulate the burst source. The researchers found that the position coincided with the Large Magellanic Cloud, a small galaxy about 170,000 light-years away. More specifically, the event's position matched that of a young supernova remnant, the glowing

DON DIXON (*preceding pages*); E. L. WRIGHT *University of California at Los Angeles*, THE COBE PROJECT, DIRBE AND NASA (*above*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

remains of a star that exploded 5,000 years ago. Unless this overlap was pure coincidence, it put the source 1,000 times as far away as theorists had thought—and thus made it a million times brighter than the Eddington limit. In 0.2 second the March 1979 event released as much energy as the sun radiates in roughly 10,000 years, and it concentrated that energy in gamma rays rather than spreading it across the electromagnetic spectrum.

No ordinary star could account for such energy, so the source was almost certainly something out of the ordinary—either a black hole or a neutron star. The former was ruled out by the eight-second modulation: a black hole is a featureless object, lacking the structure needed to produce regular pulses. The association with the supernova remnant further strengthened the case for a neutron star. Neutron stars are widely believed to form when the core of a massive but otherwise ordinary star exhausts its nuclear fuel and abruptly collapses under its own weight, thereby triggering a supernova explosion.

Identifying the source as a neutron star did not solve the puzzle; on the contrary, it merely heightened the mystery. Astronomers knew several examples of neutron stars that lie within supernova remnants. These stars were radio pulsars, objects that are observed to blink on and off in radio waves. Yet the March 1979 burster, with an apparent rotation period of eight seconds, was spinning much more slowly than any radio pulsar then known. Even when not bursting, the object emitted a steady glow of x-rays with more radiant power than could be supplied by the rotation of a neutron star. Oddly, the star was significantly displaced from the center of the supernova remnant. If it was born at the center, as is likely, then it must have recoiled with a velocity of about 1,000 kilometers per second at birth. Such high speed was considered unusual for a neutron star.

Finally, the outbursts themselves seemed inexplicable. X-ray flashes had previously been detected from some neutron stars, but they never exceeded the Eddington limit by very much. Astronomers ascribed them to thermonuclear fusion of hydrogen or helium or to the sudden accretion of matter onto the star. But the brightness of the SGR bursts was unprecedented, so a new physical mechanism seemed to be required.

Spin Forever Down

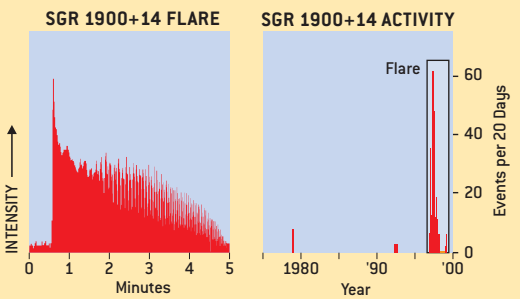
THE FINAL BURST FROM the March 1979 source was detected in May 1983; none has been seen in the 19 years since. Two other SGRs, both within our Milky Way galaxy, went off in 1979 and have remained active, emitting hundreds of bursts in the years since. A fourth SGR was located in 1998. Three of these four objects have possible, but unproved, associations with young supernova remnants. Two also lie near very dense clusters of massive young stars, intimating that SGRs tend to form from such stars. A fifth candidate SGR has gone off only twice; its precise location is still unknown.

As Los Alamos National Laboratory scientists Baolian L. Cheng, Richard I. Epstein, Robert A. Guyer and C. Alex Young pointed out in 1996, SGR bursts are statistically similar to earthquakes. The energies have very similar mathematical distributions, with less energetic events being more common. Our graduate student Ersin Gögüs of the University of Alabama at Huntsville verified this behavior for a large sample of bursts from various sources. This and other statistical properties are a hallmark of self-organized criticality, whereby a composite system attains a critical state in which a small perturbation can trigger a chain reaction. Such behavior occurs in systems as diverse as avalanches on sandpiles and magnetic flares on the sun.

But why would a neutron star behave like this? The solution emerged from an entirely separate line of work, on radio pulsars. Pulsars are widely thought to be rapidly rotating, magnetized neutron stars. The magnetic field, which is supported by electric currents flowing deep inside the star, rotates with the star. Beams of radio waves shine outward from the star's magnetic poles and sweep through space as it rotates, like lighthouse beacons—hence the observed pulsing. The pulsar also blows out a wind of charged particles and low-frequency electromag-

NAME	YEAR OF DISCOVERY	ROTATION PERIOD (seconds)
SGR 0526–66	1979	8.0
SGR 1900+14	1979	5.16
SGR 1806–20	1979	7.47
SGR 1801–23*	1997	?
SGR 1627–41	1998	?
AXP 1E 2259+586	1981	6.98
AXP 1E 1048–59†	1985	6.45
AXP 4U 0142+61	1993	8.69
AXP 1RXS 1708–40†	1997	11.0
AXP 1E 1841–045	1997	11.8
AXP AXJ1844–0258	1998	6.97
AXP CXJ0110–7211†	2002	5.44

* Not shown on map; location not known precisely
† Abbreviated name



GIANT X-RAY FLARE in August 1998 confirmed the existence of magnetars. It started with a spike of radiation lasting less than a second (*left*). Then came an extended train of pulses with a period of 5.16 seconds. This event was the most powerful outburst to come from the object, designated SGR 1900+14, since its discovery in 1979 (*right*).

SOURCE FOR TABLE: CHRYSSA KOUVELIOTOU, ROBERT C. DUNCAN AND CHRISTOPHER THOMPSON

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

netic waves, which carry away energy and angular momentum, causing its rate of spin to decrease gradually.

Perhaps the most famous pulsar lies within the Crab Nebula, the remnant of a supernova explosion that was observed in 1054. The pulsar rotates once every 33 milliseconds and is currently slowing at a rate of about 1.3 millisecond every century. Extrapolating backward, it was born rotating once every 20 milliseconds. Astronomers expect it to continue to spin down, eventually reaching a point when its rotation will be too slow to power the radio pulses. The spin-down rate has been measured for almost every radio pulsar, and theory indicates that it depends on the strength of the star's magnetic field. From this, most young radio pulsars are inferred to have magnetic fields between 10^{12} and 10^{13} gauss. For comparison, a refrigerator magnet has a strength of about 100 gauss.

The Ultimate Convection Oven

THIS PICTURE LEAVES a basic question unanswered: Where did the magnetic field come from in the first place? The traditional assumption was: it is as it is, because it was as it was. That is, most astronomers supposed that the magnetic field is a relic of the time before the star went supernova. All stars have weak magnetic fields, and those fields can be amplified simply by the act of compression. According to Maxwell's equations of electromagnetism, as a magnetized object shrinks by a factor of two, its magnetic field strengthens by a factor of four. The core of a massive star collapses by a factor of 10^5 from its birth through neutron star formation, so its magnetic field should become 10^{10} times stronger.

If the core magnetic field started with sufficient strength, this compression could explain pulsar magnetism. Unfortunately, the magnetic field deep inside a star cannot be measured, so this simple hypothesis cannot be tested. There are also good reasons to believe that compression is only part of the story.

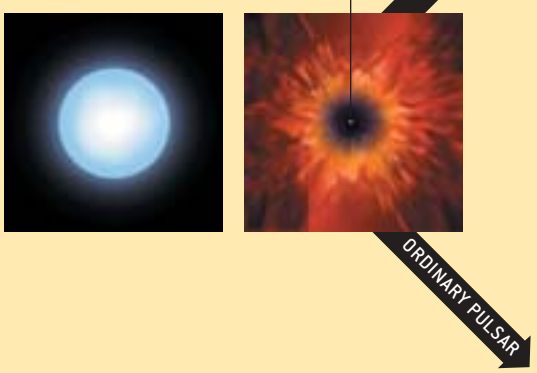
Within a star, gas can circulate by convection. Warm parcels of ionized gas rise, and cold ones sink. Because ionized gas conducts electricity well, any magnetic field lines threading the gas are dragged with it as it moves. The field can thus be reworked and sometimes amplified. This phenomenon, known as dynamo action, is thought to generate the magnetic fields of stars and planets. A dynamo might operate during each phase of the life of a massive star, as long as the turbulent core is rotating rapidly enough. Moreover, during a brief period after the core of the star turns into a neutron star, convection is especially violent.

This was first shown in computer simulations in 1986 by Adam Burrows of the University of Arizona and James M. Lattimer of the State University of New York at Stony Brook. They found that temperatures in a newborn neutron star exceed 30

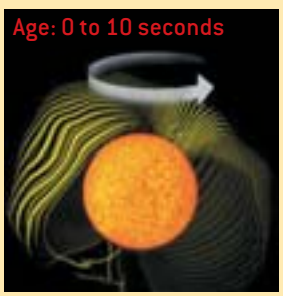
TWO TYPES OF NEUTRON STARS

1 Most neutron stars are thought to begin as massive but otherwise ordinary stars, between eight and 20 times as heavy as the sun.

2 Massive stars die in a type II supernova explosion, as the stellar core implodes into a dense ball of subatomic particles.



3 A: If the newborn neutron star spins fast enough, it generates an intense magnetic field. Field lines inside the star get twisted.



4 A: The magnetar settles into neat layers, with twisted field lines inside and smooth lines outside. It might emit a narrow radio beam.

5 A: The old magnetar has cooled off, and much of its magnetism has decayed away. It emits very little energy.

3 B: If the newborn neutron star spins slowly, its magnetic field, though strong by everyday standards, does not reach magnetar levels.

4 B: The mature pulsar is cooler than a magnetar of equal age. It emits a broad radio beam, which radio telescopes can readily detect.

5 B: The old pulsar has cooled off and no longer emits a radio beam.

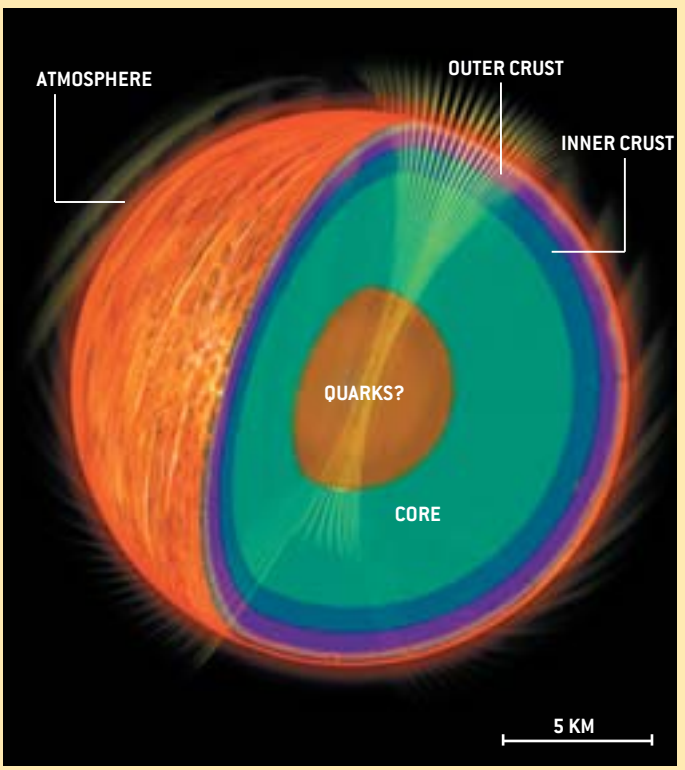
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

billion kelvins. Hot nuclear fluid circulates in 10 milliseconds or less, carrying enormous kinetic energy. After about 10 seconds, the convection ceases.

Not long after Burrows and Lattimer conducted their first simulations, Duncan and Thompson, then at Princeton University, estimated what this furious convection means for neutron-star magnetism. The sun, which undergoes a sedate version of the same process, can be used as a reference point. As solar fluid circulates, it drags along magnetic field lines and gives up about 10 percent of its kinetic energy to the field. If the moving fluid in a newborn neutron star also transfers a tenth of its kinetic energy to the magnetic field, then the field would grow stronger than 10^{15} gauss, which is more than 1,000 times as strong as the fields of most radio pulsars.

Whether the dynamo operates globally (rather than in limited regions) would depend on whether the star's rate of rotation was comparable to its rate of convection. Deep inside the sun, these two rates are similar, and the magnetic field is able to organize itself on large scales. By analogy, a neutron star born rotating as fast as or faster than the convective period of 10 milliseconds could develop a widespread, ultrastrong magnetic field. In 1992 we named these hypothetical neutron stars "magnetars."

An upper limit to neutron-star magnetism is about 10^{17} gauss; beyond this limit, the fluid inside the star would tend to mix and the field would dissipate. No known objects in the universe can generate and maintain fields stronger than this level. One ramification of our calculations is that radio pulsars are neutron stars in which the large-scale dynamo has *failed* to operate. In the case of the Crab pulsar, the newborn neutron star rotated once every 20 milliseconds, much slower than the rate of convection, so the dynamo never got going.



STRUCTURE OF A NEUTRON STAR can be inferred from theories of nuclear matter. Starquakes can occur in the crust, a lattice of atomic nuclei and electrons. The core consists mainly of neutrons and perhaps quarks. An atmosphere of hot plasma might extend a grand total of a few centimeters.

Crinkle Twinkle Little Magnetar

ALTHOUGH WE DID NOT develop the magnetar concept to explain SGRs, its implications soon became apparent to us. The magnetic field should act as a strong brake on a magnetar's rotation. Within 5,000 years a field of 10^{15} gauss would slow the spin rate to once every eight seconds—neatly explaining the oscillations observed during the March 1979 outburst.

As the field evolves, it changes shape, driving electric currents along the field lines outside the star. These currents, in turn, generate x-rays. Meanwhile, as the magnetic field moves through the solid crust of a magnetar, it bends and stretches the crust. This process heats the interior of the star and occasionally breaks the crust in a powerful "starquake." The accompanying release of magnetic energy creates a dense cloud of electrons and positrons, as well as a sudden burst of soft gamma rays—accounting for the fainter bursts that give SGRs their name.

More infrequently, the magnetic field becomes unstable and undergoes a large-scale rearrangement. Similar (but smaller) upheavals sometimes happen on the sun, leading to solar flares. A magnetar easily has enough energy to power a giant flare such as the March 1979 event. Theory indicates that the first half-second of that tremendous outburst came from an expanding fireball. In 1995 we suggested that part of the fireball was trapped by the magnetic field lines and held close to the star. This trapped fireball gradually shrank and then evaporated, emitting x-rays all the while. Based on the amount of energy released, we calculated the strength of the magnetic field needed to confine the enormous fireball pressure: greater than 10^{14} gauss, which agrees with the field strength inferred from the spin-down rate.

A separate estimate of the field had been given in 1992 by Bohdan Paczyński of Princeton. He noted that x-rays can slip

THE AUTHORS

CHRYSSA KOUVELIOTOU, ROBERT C. DUNCAN and *CHRISTOPHER THOMPSON* have studied magnetars for a collective 40 years and have collaborated for the past five. Kouveliotou, an observer, works at the National Space Science and Technology Center in Huntsville, Ala. Besides soft-gamma repeaters, her pets include gamma-ray bursts, x-ray binaries and her cat, Felix; her interests range from jazz to archaeology to linguistics. Duncan and Thompson are theorists, the former at the University of Texas at Austin, the latter at the Canadian Institute for Theoretical Astrophysics in Toronto. Duncan has studied supernovae, quark matter and intergalactic gas clouds. In his younger days he ran a 2:19 marathon in the 1980 U.S. Olympic trials. Thompson has worked on topics from cosmic strings to giant impacts in the early solar system. He, too, is an avid runner as well as a backpacker.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

through a cloud of electrons more easily if the charged particles are immersed in a very intense magnetic field. For the x-rays during the burst to have been so bright, the magnetic field must have been stronger than 10^{14} gauss.

What makes the theory so tricky is that the fields are stronger than the quantum electrodynamic threshold of 4×10^{13} gauss. In such strong fields, bizarre things happen. X-ray photons readily split in two or merge together. The vacuum itself is polarized, becoming strongly birefringent, like a calcite crystal. Atoms are deformed into long cylinders thinner than the quantum-relativistic wavelength of an electron [*see box on opposite page*]. All these strange phenomena have observable effects on magnetars. Because this physics was so exotic, the theory attracted few researchers at the time.

Zapped Again

AS THESE THEORETICAL developments were slowly unfolding, observers were still struggling to see the objects that were the sources of the bursts. The first opportunity came when NASA's orbiting Compton Gamma Ray Observatory recorded a burst of gamma rays late one evening in October 1993. This was the break Kouveliotou had been looking for when she joined the Compton team in Huntsville. The instrument that registered the burst could determine its position only to within a fairly broad swath of sky. Kouveliotou turned for help to the Japanese ASCA satellite. Toshio Murakami of the Institute of Space and Astronautical Science in Japan and his collaborators soon found an x-ray source from the same swath of sky. The source held steady, then gave off another burst—proving beyond all doubt that it was an SGR. The same object had first been seen in 1979 and, based on its approximate celestial coordinates, was identified as SGR 1806–20. Now its position was fixed much more precisely, and it could be monitored across the electromagnetic spectrum.

The next leap forward came in 1995, when NASA launched the Rossi X-ray Timing Explorer (RXTE), a satellite designed to be highly sensitive to variations in x-ray intensity. Using this instrument, Kouveliotou found that the emission from SGR 1806–20 was oscillating with a period of 7.47 seconds—amazingly close to the 8.0-second periodicity observed in the March 1979 burst (from SGR 0526–66). Over the course of five years, the SGR slowed by nearly two parts in 1,000. Although the slowdown may seem small, it is faster than that of any radio pulsar known, and it implies a magnetic field approaching 10^{15} gauss.

More thorough tests of the magnetar model would require a second giant flare. Luckily, the heavens soon complied. In the early morning of August 27, 1998, some 19 years after the giant flare that began SGR astronomy was observed, an even more intense wave of gamma rays and x-rays reached Earth from the depths of space. It drove detectors on seven scientific spacecraft to their maximum or off scale. One interplanetary probe, NASA's Comet Rendezvous Asteroid Flyby, was forced into a protective shutdown mode. The gamma rays hit Earth on its nightside, with the source in the zenith over the mid-Pacific Ocean.

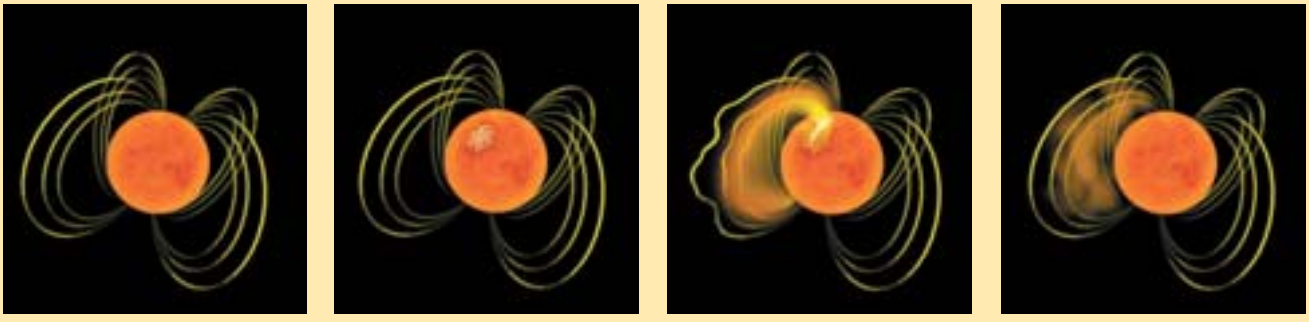
Fortuitously, in those early morning hours electrical engineer Umran S. Inan and his colleagues from Stanford University were gathering data on the propagation of very low frequency radio waves around Earth. At 3:22 A.M. PDT, they noticed an abrupt change in the ionized upper atmosphere. The inner edge of the ionosphere plunged down from 85 to 60 kilometers for five minutes. It was astonishing. This effect on our planet was caused by a neutron star far across the galaxy, 20,000 light-years away.

Another Magneto Marvel

THE AUGUST 27 FLARE was almost a carbon copy of the March 1979 event. Intrinsically, it was only one tenth as powerful, but because the source was closer to Earth it remains the most intense burst of gamma rays from beyond our solar system ever detected. The last few hundred seconds of the flare showed conspicuous pulsations, with a 5.16-second period. Kouveliotou

HOW MAGNETAR BURSTS HAPPEN

THE MAGNETIC FIELD OF THE STAR is so strong that the rigid crust sometimes breaks and crumbles, releasing a huge surge of energy.



1 Most of the time the magnetar is quiet. But magnetic stresses are slowly building up.

2 At some point the solid crust is stressed beyond its limit. It fractures, probably into many small pieces.

3 This "starquake" creates a surging electric current, which decays and leaves behind a hot fireball.

4 The fireball cools by releasing x-rays from its surface. It evaporates in minutes or less.

DON DIXON

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

and her team measured the spin-down rate of the star with RXTE; sure enough, it was slowing down at a rate comparable to that of SGR 1806–20, implying a similarly strong magnetic field. Another SGR was placed into the magnetar hall of fame.

The precise localizations of SGRs in x-rays have allowed them to be studied using radio and infrared telescopes (though not in visible light, which is blocked by interstellar dust). This work has been pioneered by many astronomers, notably Dale Frail of the National Radio Astronomy Observatory and Shri Kulkarni of the California Institute of Technology. Other observations have shown that all four confirmed SGRs continue to release energy, albeit faintly, even between outbursts. "Faintly" is a relative term: this x-ray glow represents 10 to 100 times as much power as the sun radiates in visible light.

By now one can say that magnetar magnetic fields are better measured than pulsar magnetic fields. In isolated pulsars, almost the only evidence for magnetic fields as strong as 10^{12} gauss comes from their measured spin-down. In contrast, the combination of rapid spin-down and bright x-ray flares provides several independent arguments for 10^{14}- to 10^{15}-gauss fields in magnetars. As this article goes to press, Alaa Ibrahim of the NASA Goddard Space Flight Center and his collaborators have reported yet another line of evidence for strong magnetic fields in magnetars: x-ray spectral lines that seem to be generated by protons gyrating in a 10^{15}-gauss field.

One intriguing question is whether magnetars are related to cosmic phenomena besides SGRs. The shortest-duration gamma-ray bursts, for example, have yet to be convincingly explained, and at least a handful of them could be flares from magnetars in other galaxies. If seen from a great distance, even a giant flare would be near the limit of telescope sensitivity. Only the brief, hard, intense pulse of gamma rays at the onset of the flare would be detected, so telescopes would register it as a GRB.

Thompson and Duncan suggested in the mid-1990s that magnetars might also explain anomalous x-ray pulsars, a class of objects that resemble SGRs in many ways. The one difficulty with this idea was that AXPs had not been observed to burst. Recently, however, Victoria M. Kaspi and Fotis P. Gavriil of McGill University and Peter M. Woods of the National Space and Technology Center in Huntsville detected bursts from two of the seven known AXPs. One of these objects is associated with a young supernova remnant in the constellation Cassiopeia.

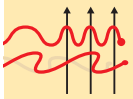
Another AXP in Cassiopeia is the first magnetar candidate to have been detected in visible light. Ferdi Hulleman and Marten van Kerkwijk of Utrecht University in the Netherlands, working with Kulkarni, spotted it three years ago, and Brian Kern and Christopher Martin of Caltech have since monitored its brightness in visible light. Though exceedingly faint, the AXP fades in and out with the x-ray period of the neutron star. These observations lend support to the idea that it is indeed a magnetar. The main alternative—that AXPs are ordinary neutron stars surrounded by disks of matter—predicts too much visible and infrared emission with too little pulsation.

In view of these recent discoveries, and the apparent silence of the Large Magellanic Cloud burster for nearly 20 years, it appears that magnetars can change their clothes. They can remain quiescent for years, even decades, before undergoing sudden periods of extreme activity. Some astronomers argue that AXPs are younger on average than SGRs, but this is still a matter of debate. If both SGRs and AXPs are magnetars, then magnetars plausibly constitute a substantial fraction of all neutron stars.

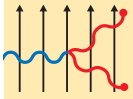
The story of magnetars is a sobering reminder of how much we have yet to understand about our universe. Thus far, we have discerned at most a dozen magnetars among the countless stars. They reveal themselves for a split second, in light that only the most sophisticated telescopes can detect. Within 10,000 years, their magnetic fields freeze and they stop emitting bright x-rays. So those dozen magnetars betray the presence of more than a million, and perhaps as many as 100 million, other objects—old magnetars that long ago went dark. Dim and dead, these strange worlds wander through interstellar space. What other phenomena, so rare and fleeting that we have not recognized them, lurk out there? SA

EXTREME MAGNETISM

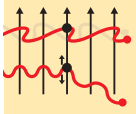
MAGNETAR FIELDS wreak havoc with radiation and matter.



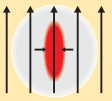
VACUUM BIREFRINGENCE
Polarized light waves (*orange*) change speed and hence wavelength when they enter a very strong magnetic field (*black lines*).



PHOTON SPLITTING
In a related effect, x-rays freely split in two or merge together. This process is important in fields stronger than 10^{14} gauss.



SCATTERING SUPPRESSION
A light wave can glide past an electron (*black circle*) with little hindrance if the field prevents the electron from vibrating with the wave.



DISTORTION OF ATOMS
Fields above 10^9 gauss squeeze electron orbitals into cigar shapes. In a 10^{14}-gauss field, a hydrogen atom becomes 200 times narrower.

MORE TO EXPLORE

Formation of Very Strongly Magnetized Neutron Stars: Implications for Gamma-Ray Bursts. Robert C. Duncan and Christopher Thompson in *Astronomical Journal,* Vol. 392, No. 1, pages L9–L13; June 10, 1992. Available at **makeashorterlink.com/?B16A425A2**

An X-ray Pulsar with a Superstrong Magnetic Field in the Soft Gamma-Ray Repeater SGR1806–20. C. Kouveliotou, S. Dieters, T. Strohmayer, J. Von Paradijs, G. J. Fishman, C. A. Meegan, K. Hurley, J. Kommers, I. Smith, D. Frail and T. Murakami in *Nature,* Vol. 393, pages 235–237; May 21, 1998.

The Life of a Neutron Star. Joshua N. Winn in *Sky & Telescope,* Vol. 98, No. 1, pages 30–38; July 1999.

Physics in Ultra-strong Magnetic Fields. Robert C. Duncan. Fifth Huntsville Gamma-Ray Burst Symposium, February 23, 2002. Available at **arXiv.org/abs/astro-ph/0002442**

Flash! The Hunt for the Biggest Explosions in the Universe. Govert Schilling. Cambridge University Press, 2002.

More information can be found at Robert C. Duncan's Web site: **solomon.as.utexas.edu/magnetar.html**

BRYAN CHRISTIE DESIGN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

New research addresses the wrenching question left when someone ends his or her own life

WHY?!?

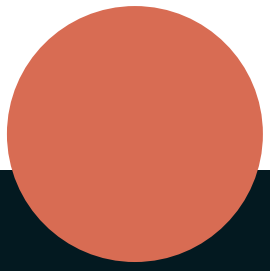
The neuroscience of suicide

By Carol Ezzell


MATT MAHURIN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

In 1994, two days after returning from a happy family vacation, my 57-year-old mother put the muzzle of a handgun to her left breast and fired, drilling a neat and lethal hole through her heart—and, metaphorically, through our family's as well.



SUICIDE IS THE **11TH LEADING CAUSE OF DEATH** IN THE U.S., ACCOUNTING FOR 1.2 PERCENT OF ALL FATALITIES.

A PERSON DIES BY SUICIDE ROUGHLY **EVERY 18 MINUTES** IN THE U.S. SOMEONE ATTEMPTS SUICIDE EVERY MINUTE.

FOUR MALES DIE BY SUICIDE FOR EVERY FEMALE, BUT AT LEAST **TWICE AS MANY WOMEN** AS MEN ATTEMPT SUICIDE.

APPROXIMATELY **80 AMERICANS** TAKE THEIR OWN LIVES **EVERY DAY**.

THE SUICIDE RATE FOR **WHITE MALES AGED 15 TO 24** HAS TRIPLED SINCE 1950.

It was around midnight on a Saturday night in July, the time of year, I was later surprised to learn, that has the highest incidence of suicide in the Northern Hemisphere. My stepfather was at home but didn't hear the single shot because he was taking a shower in a bathroom at the other end of the house. When he returned to their bedroom, she was crumpled on the carpet in her pajamas, almost gone. She tried to say something to him before she died, but he couldn't make out what it was. The emergency medical technicians arrived to find a patient, but not the one they expected: my stepfather nearly died himself that night after hyperventilating from the shock, which all but overwhelmed lungs already compromised by emphysema.

Through it all, I was asleep in my apartment 200 miles away. I was awakened at 2 A.M. by a call from my building's front desk, telling me that my sister-in-law was downstairs and wanted to come up. My first words to her when I opened my door were, "It's Mother, isn't it?"

Our family has too much company in suffering the agony of having a loved one die by suicide: annually, 30,000 people in the U.S. take their own lives. That is roughly half again the number who died of AIDS last year. Why do they do it?

Like an estimated 60 to 90 percent of U.S. suicides, my mother had a mental illness. In her case, it was manic-depression, also called bipolar disorder. Unless they are taking—and responding well to—the appropriate medication, manic-depressives oscillate between troughs of despair and peaks of elation or agitation. Most who end their lives have a history of depression or manic-depression, but people with severe depression differ in their propensity for suicide.

Scientists have begun uncovering behavioral tip-offs and are also exploring clues to anatomical and chemical differences between the brains of suicides and of those who die of other causes. If such changes could be detected in medical imaging scans or through blood tests, doctors might one day be able to identify those at highest risk of dying by suicide—and therefore attempt to prevent the tragedy from occurring. Sadly, that goal is not immediately in sight: many who have suicidal tendencies still end up taking their own lives, despite intensive intervention.

My Mother's Legacy

THE QUESTION of what drove my mother to her desperate act that humid night nearly nine years ago is the second most difficult thing I live with. Scarcely a day has gone by that I haven't been pierced by the anguish of wanting to know exactly what prompted her suicide on that particular night as well as the crushing guilt over what I could have done—should have done, would have done—to stop her. The hardest thing I have to live with is the realization that I will never know the answer for sure.

In the future, some parts of her story should become less mysterious, because researchers are studying those very issues. One age-old question, whether a tendency to commit suicide is inborn or the result of an accumulation of bad experiences, is at least closer to resolution.

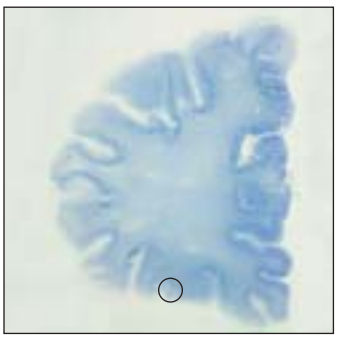
Although the nature-versus-nurture debate still rages in some psychiatric circles, most researchers who study suicide fall somewhere in the middle. "You need several things to go wrong at once," explains Victoria Arango of the New York State Psychiatric Institute, which is affiliated with Columbia-Presbyterian Medical Center. "I'm not saying that suicide is purely biological, but it starts with having

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

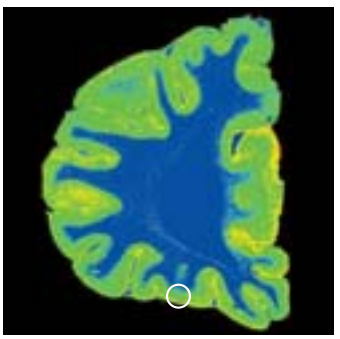
PHYSICAL CLUES TO SUICIDE

CHANGES IN THE ORBITAL PREFRONTAL CORTEX

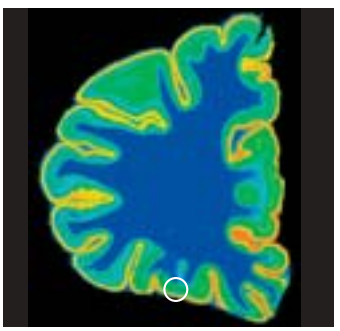
Slices from the brain of a suicide victim contain fewer neurons in a subsection (*circled*) of the orbital prefrontal cortex.



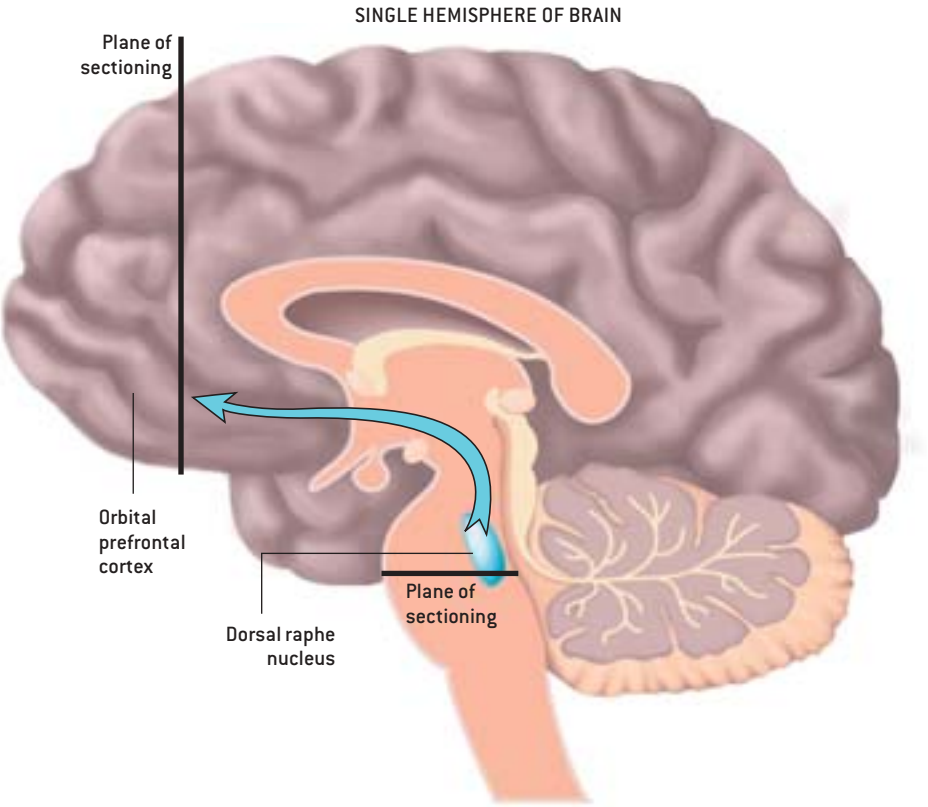
Throughout the cortex, serotonin transporters (*gold*) absorb serotonin. In the marked subsection, the number of these transporters is reduced.



The analyzed area also exhibits more binding of serotonin (*orange*) per neuron. Together the analyses indicate that the brain tried to make the most of the serotonin it had.

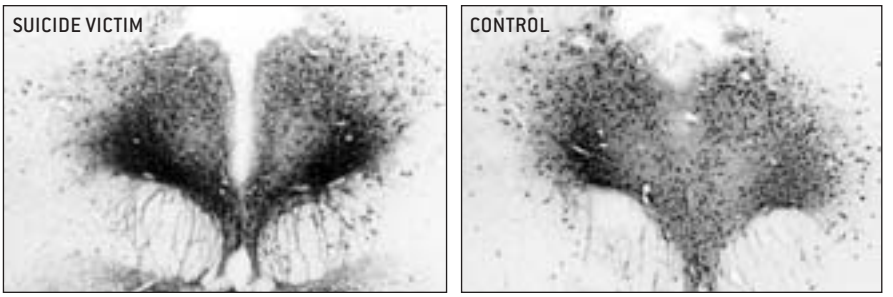


IN PEOPLE WHO DIE BY SUICIDE, anatomical and chemical changes occur in two brain regions: the orbital prefrontal cortex, which lies just above the eyes, and the dorsal raphe nucleus of the brain stem. The alterations are evidence of a reduced ability to make and use serotonin, a key neurotransmitter known to be lacking in the brains of impulsive people and in those suffering from depression. Neurons in the dorsal raphe nucleus produce serotonin; they have long projections (*blue arrow*) that carry the neurotransmitter to the orbital prefrontal cortex. In suicide victims, the dorsal raphe nucleus sends less than normal amounts of serotonin to the orbital prefrontal cortex. *—C.E.*



CHANGES IN THE DORSAL RAPHE NUCLEUS

Neurons in the dorsal raphe nucleus of the brain stem of someone who died by suicide contain more of the enzyme that synthesizes serotonin (*dark staining*) than the corresponding brain region of someone who died of another cause. The difference indicates that the brains of suicides are attempting to produce more serotonin.



VICTORIA ARANGO ET AL. *New York State Psychiatric Institute (brain slices)*; REPRODUCED WITH PERMISSION FROM *BIOLOGICAL PSYCHIATRY*, VOL. 46, NO. 4; 1999 *(neuron micrographs)*; CLEO VILETT *(brain diagram)*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

BETWEEN 1980 AND 1996 THE SUICIDE RATE FOR **AFRICAN-AMERICAN MALES** AGED 15 TO 19 INCREASED 105 PERCENT.

SUICIDE IS THE **THIRD-RANKING CAUSE OF DEATH FOR TEENS** AGED 10 TO 19.

WHITE MEN 85 AND OLDER DIE BY SUICIDE AT SIX TIMES THE OVERALL NATIONAL RATE.

SUICIDE RATES FOR WOMEN PEAK BETWEEN THE AGES OF 45 AND 54 AND SURGE AGAIN AFTER AGE 85.

ALCOHOLISM IS A FACTOR IN ROUGHLY 30 PERCENT OF ALL COMPLETED SUICIDES.

APPROXIMATELY 7 PERCENT OF **PEOPLE WITH ALCOHOL DEPENDENCE** WILL DIE BY SUICIDE.

EIGHTY-THREE PERCENT OF **GUN-RELATED DEATHS** IN THE HOME ARE THE RESULT OF SUICIDE.

DEATH BY FIREARMS IS THE **FASTEST-GROWING METHOD** OF SUICIDE.

an underlying biological risk." Life experience, acute stress and psychological factors each play a part, she asserts. At the root of the mystery of suicide, however, lies a nervous system whose lines of communication have become tangled into unbearably painful knots.

Arango and her Columbia colleague J. John Mann are leading the effort to pick apart those knots and discern the neuropathology of suicide. They have assembled what is generally acknowledged to be the country's best collection of brain specimens from suicide victims. Twenty-five deep freezers in their laboratories hold a total of 200 such brains, which the researchers are examining for neuroanatomical, chemical or genetic alterations that might be unique to those compelled to end their lives. Each brain is accompanied by a "psychological autopsy," a compendium of interviews with family members and intimates probing the deceased's state of mind and behavior leading up to his or her final act. "We try to get a complete picture," Mann says, "and come up with an aggregate explanation for that person." A suicide brain is matched against a control brain from a person of the same sex without a psychiatric disorder who died at approximately the same age of a cause other than suicide.

Contained within the three-pound gelatinous mass of the human brain are the cells and molecules that were inextricably linked to what that person once thought—and, indeed, once was. Mann's and Arango's research concentrates in part on the prefrontal cortex, the portion of the brain encased in the bone of the forehead. The prefrontal cortex is the seat of the so-called executive functions of the brain, including the internal censor that keeps individuals from blurting out what they really think in awkward social situations or acting on potentially dangerous impulses.

The impulse-dampening role played by the prefrontal cortex particularly interests Mann and Arango. Scientists have looked to impulsivity as a predictor for suicide for decades. Although some people plan their deaths carefully—leaving notes, wills and even funeral plans—for many, including my mother, suicide appears to be spontaneous: a very bad decision on a very bad day. So Arango and Mann search in these brains for clues to the biological basis for that impulsivity. One focus is on differences in the availability of the brain chemical serotonin—previous research on the basis of impulsivity has indicated a dearth of it.

Serotonin is a neurotransmitter, one of the molecules that jumps the tiny gaps known as synapses between neurons to relay a signal from one such brain cell to another. Tiny membranous bubbles called vesicles erupt from each signal-sending, or presynaptic, neuron, releasing serotonin into the synapse. Receptors on the receiving, or postsynaptic, neurons bind to the neurotransmitter and register biochemical changes in the cell that can change its ability to respond to other stimuli or to turn genes on or off. After a short while, the presynaptic cells reabsorb the serotonin using molecular sponges termed serotonin transporters.

Serotonin somehow exerts a calming influence on the mind. Prozac and similar antidepressant drugs work by binding to serotonin transporters and preventing presynaptic neurons from soaking up the secreted serotonin too quickly, allowing it to linger a bit longer in the synapse and continue to transmit its soothing effect.

Traces of Pain

MORE THAN two decades of reports have linked low serotonin levels in the brain to depression, aggressive behavior and a tendency toward impulsiveness, but the evidence has been particularly confusing with regard to suicide. A number of studies have found reductions in serotonin in the brains of suicides, whereas others have not. Some have observed a lack of serotonin in one part of the brain but not elsewhere. Still others have described increases in the number of receptors for serotonin or deficits in the chain of chemical events that convey the serotonin signal from those receptors to the inside of a neuron.

Despite the inconsistencies, the bulk of evidence points strongly to a problem in the brains of suicides involving the serotonin system. That line of thinking has been bolstered by the recent findings of Arango and Mann.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

THE "MAGIC" OF LITHIUM

Lithium appears to prevent suicide. Why do so few suicidal people take it?

"Lithium ... is the lightest of the solid elements, and it is perhaps not surprising that it should in consequence possess certain modest magical qualities." —G. P. Hartigan, psychiatrist

"Only crazy people take lithium!" my mother shouted during one of our many arguments over her not receiving the best treatment for her manic-depression. She accused me and my stepfather of wanting to medicate her so she would "just shut up." To be honest, she was partially right: it is very trying to be around someone in the grip of a mania, which often brings on incessant, stream-of-consciousness talking.

Many people find lithium—which generally comes in capsules of lithium carbonate or lithium citrate—difficult to take. It can cause hand tremors, constant thirst, frequent urination, weight gain, lethargy, reduced muscle coordination, blurred thinking and short-term memory deficits. People on it must also have its concentration in their blood assessed regularly to ensure that it is within the therapeutic range: the drug is usually ineffective below 0.6 millimole per liter of blood serum and can cause life-threatening toxic reactions if the level becomes higher than two millimoles per liter.

Lithium is used routinely to even out the extreme mood swings of patients with manic-depressive illness, or bipolar disorder. Increasingly, however, it is also offered to people with depression. But a growing body of evidence indicates that this compound can literally keep people who are at risk of suicide alive. In 1998 lithium pioneer Mogens Schou of the Psychiatric Hospital in Risskov, Denmark, pulled together the results of various studies of lithium as a suicide preventive and observed that people not taking the drug were three to 17 times as likely to end their own lives as depressed people who took the medication. Likewise, Schou determined that lithium reduced suicide attempts by a factor of between six and 15.

How does it exert its salutary effects? Despite a number of tantalizing leads, researchers are still not certain. "It's hard to say at this time," says Ghanshyam N. Pandey of the University of Illinois. "There are so many modes of action." Lithium is thought to affect tiny ports called ion channels on the surfaces of nerve cells, or neurons. As they open and close, ion channels admit or bar charged atoms that determine the electrical potential within the cells, thereby dictating their activity and ability to communicate with other neurons. Scientists posit that the drug stabilizes the excitability of the neurons by influencing the ion channels or by skewing the chain reaction of biochemical events that occur within an excited cell.

A drug only works, though, if someone takes it properly. In the May 2002 issue of the *Journal of Clinical Psychiatry,* Jan Scott and Marie Pope of the University of Glasgow reported that half of a group of 98 patients who were taking a mood-stabilizing drug such as lithium failed to stick with their drug regimen. Yet, the researchers noted, just 1 percent of scientific publications on the subject of mood stabilizers looked at why patients did not take their lithium as prescribed.

LITHIUM is the lightest of the solid elements and, in its pure form, floats (*left*). When compounded in pill form as lithium carbonate or lithium citrate (*above*), it can be taken to stabilize moods.

J. John Mann of the New York State Psychiatric Institute says that a major factor in noncompliance is the human desire not to want to think of oneself as ill. "There's a natural reluctance to take any medicine long-term," Mann explains. "When a person is depressed, they have a problem imagining ever getting better. When they're well, they can't imagine getting sick again."

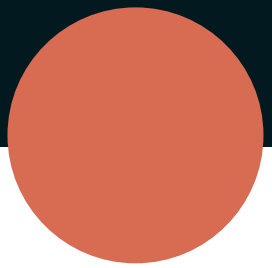
The side effects of lithium also play a role. Kay Redfield Jamison, a psychiatrist at Johns Hopkins University who studies manic-depression and suicide—and who is a manic-depressive herself—has found that the most common reasons patients stop taking the drug are cognitive side effects, weight gain and impaired coordination. In her moving memoir, *An Unquiet Mind,* she recounts her own struggle to come to terms with the fact that she will probably be coping with lithium's side effects for the rest of her life. Perhaps if my mother had lived to read it, she would have been heartened by Jamison's example and motivated to begin lithium therapy. *—C.E.*

THOMAS SIELNACHT (*container*); CUSTOM MEDICAL STOCK PHOTO (*pill*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SUICIDES **OUTNUMBER HOMICIDES TWO TO ONE** EVERY YEAR IN THE U.S.

SUICIDE ACCOUNTS FOR **NEARLY 57 PERCENT OF ALL FIREARM DEATHS** IN THE U.S.; 60 PERCENT OF ALL SUICIDES INVOLVE FIREARMS.



In a second-floor laboratory at the upper tip of Manhattan, Arango's technician leans into an open freezer to use a machine called a microtome to pare a feather-light slice from a frozen brain donated by grieving relatives anxious to help science address the mystery of suicide. Using a chilled brush, she delicately coaxes the rime of icy tissue onto a glass slide the size of a snapshot. With the body heat from her own gloved hands, she then melts the brain sliver onto the glass; observing the process is reminiscent of watching bright sunlight on a frigid winter day dissolve frost on a window.

The scientists working with the Columbia collection divide the brains into left and right hemispheres and then carefully section each hemisphere into 10 or 12 blocks from front to back. Once frozen and put through the microtome, every block yields roughly 160 slices that are thinner than a human hair.

The chief benefit of this approach is that Arango's and Mann's groups can perform several different biochemical tests on the same brain slice and know the exact anatomical locations of the variations they find. By reassembling the slices virtually, they can compile an overall model of how those abnormalities might work in concert to affect a complex behavior.

At a conference of the American College of Neuropsychopharmacology in 2001, Arango reported that the brains of people who were depressed and died by suicide contained fewer neurons in the orbital prefrontal cortex, a patch of brain just above each eye. What is more, in suicide brains, that area had one third the number of presynaptic serotonin transporters that control brains had but roughly 30 percent more postsynaptic serotonin receptors.

Together the results suggest that the brains of suicides are trying to make the most of every molecule of serotonin they have, by increasing the molecular equipment for sensing the neurotransmitter while decreasing the number of transporters that absorb it back again. "We believe there is a deficiency in the serotonergic system in people who commit suicide," Arango concludes. "They can be so sick Prozac can't help them." Inhibiting the reuptake of serotonin isn't always enough to prevent suicide: it wasn't for my mother, who died despite taking 40 milligrams of Prozac a day.

Mann and his colleagues are now trying to devise a positron emission tomography (PET) test that might one day aid doctors in determining which among their depressed patients have the most skewed serotonin circuitry—and are therefore at highest risk of suicide. PET scans mirror brain activity by monitoring which brain regions consume the most blood glucose; administering drugs, such as fenfluramine, that cause the release of serotonin can help scientists zero in on active brain areas using serotonin.

In the January *Archives of General Psychiatry,* Mann and his co-workers reported a relation between activity in the prefrontal cortex of people who had attempted suicide and the potential deadliness of the attempt. Those who had used the most dangerous means—for example, by taking the most pills or jumping from the highest point—had the least serotonin-based activity in the prefrontal cortex. "The more lethal the suicide attempt, the bigger the abnormality," Mann observes.

Ghanshyam N. Pandey of the University of Illinois agrees that the brain's serotonin system is key to understanding suicide. "There is a lot of evidence to suggest serotonin defects in suicide, but these defects do not exist in isolation but in concert with other deficits," he says. "The whole system appears to be altered."

The serotonin hypothesis does not rule out important contributions by other neurotransmitters, however. Serotonin is only one molecule in the intricate biochemical network named the hypothalamic-pituitary-adrenal (HPA) axis, in which the hypothalamus and pituitary glands in the brain communicate with the adrenal glands atop the kidneys. The HPA is responsible for the so-called fight-or-flight response exemplified by the racing heartbeat and sweaty palms you get after a close scrape while driving, say. In particular, corticotrophin-releasing factor, which the hypothalamus releases in times of stress, causes the anterior pituitary to make adrenocorticotropic hormone, which in turn causes the adrenal cortex to produce glucocorticoids such as cortisol. Cortisol prepares the body for stress by raising blood sugar concentrations, increasing heart rate and inhibiting the overreaction of the immune response.

Serotonin fits into the HPA because it modulates the threshold of stimulation. Researchers such as Charles B. Nemeroff of the Emory University School of Medicine and his colleagues are finding that extremely adverse early life experiences, such as child abuse, can throw the HPA axis off kilter, literally leaving biochemical imprints on the brain that make it vulnerable to depression as a result of overreacting to stress later on.

In 1995 Pandey's group reported indications that the abnormalities in serotonin circuitry present in those at risk for suicide could be detectable using a relatively simple blood test. When he and his co-workers compared the number of serotonin receptors on platelets (clotting cells) in the blood of suicidal people with those of nonsuicidal people, they observed that individuals considering suicide had many more serotonin receptors. (Platelets just happen to have receptors for serotonin, although it is unclear why.)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

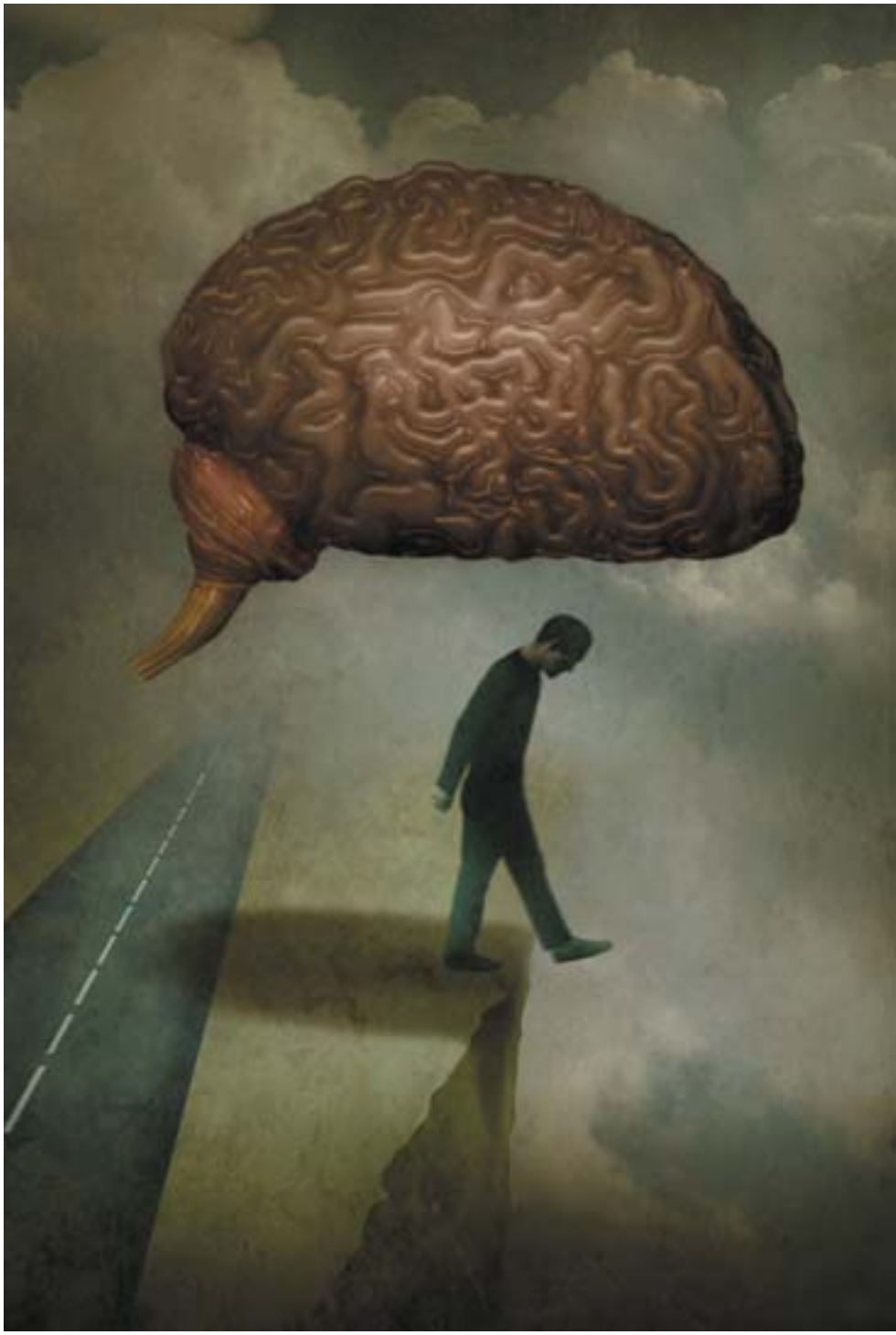
Pandey says that his group concluded that the boost in receptors reflects a similar increase in the suicide-prone brains—a vain attempt to garner as much serotonin as possible. To prove the link, Pandey would like to determine whether the association holds up in people who go on to take their own lives. "We want to know if platelets can be used as markers for identifying suicidal patients," Pandey says. "We are making progress, but it's slow."

A Curse of Generations

UNTIL RESEARCHERS can develop tests to forecast those at highest risk for suicide, doctors might concentrate their efforts on the biological relatives of suicide victims. In the September 2002 issue of *Archives of General Psychiatry*, Mann, David A. Brent of the Western Psychiatric Institute and Clinic in Pittsburgh and their colleagues reported that the offspring of suicide attempters have six times the risk of people whose parents never attempted suicide. The link appears in part to be genetic, but efforts to pin down a predisposing gene or genes have not yet yielded any easy answers. In studies in the early 1990s Alec Roy of the Department of Veterans Affairs Medical Center in East Orange, N.J., observed that 13 percent of the identical twins of people who died by suicide also eventually took their own lives, whereas only 0.7 percent of fraternal twins traveled the same path as their suicidal siblings.

These statistics serve as warnings to me and to others with biological ties to suicide. In a small jar in my bedroom I keep a bullet from the same box that contained the one that killed my mother. The police took the gun after her death, and I myself threw away the remaining bullets while cleaning out her bedroom closet. But I like to think that I hold on to that single, cold pellet of metal as a reminder of how tenuous life is and how one impulsive act can have immense and rippling consequences. Perhaps someday science will better understand the basis for such harrowing acts so that families like mine will be spared. SA

Carol Ezzell is a staff editor and writer.



MATT MAHURIN

MORE TO EXPLORE

Night Falls Fast: Understanding Suicide. Kay Redfield Jamison. Vintage Books, 2000.

Reducing Suicide: A National Imperative. Institute of Medicine. Edited by Sarah K. Goldsmith, Terry C. Pellmar, Arthur M. Kleinman and William E. Bunney. National Academies Press, 2002.

Information and education materials on preventing suicide can be obtained from the National Mental Health Association (**www.nmha.org**), the American Foundation for Suicide Prevention (**www.afsp.org**) and the American Association of Suicidology (**www.suicidology.org**). The groups also have support materials for the survivors of loved ones who died by suicide.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

evolving Inventions

By John R. Koza, Martin A. Keane and Matthew J. Streeter

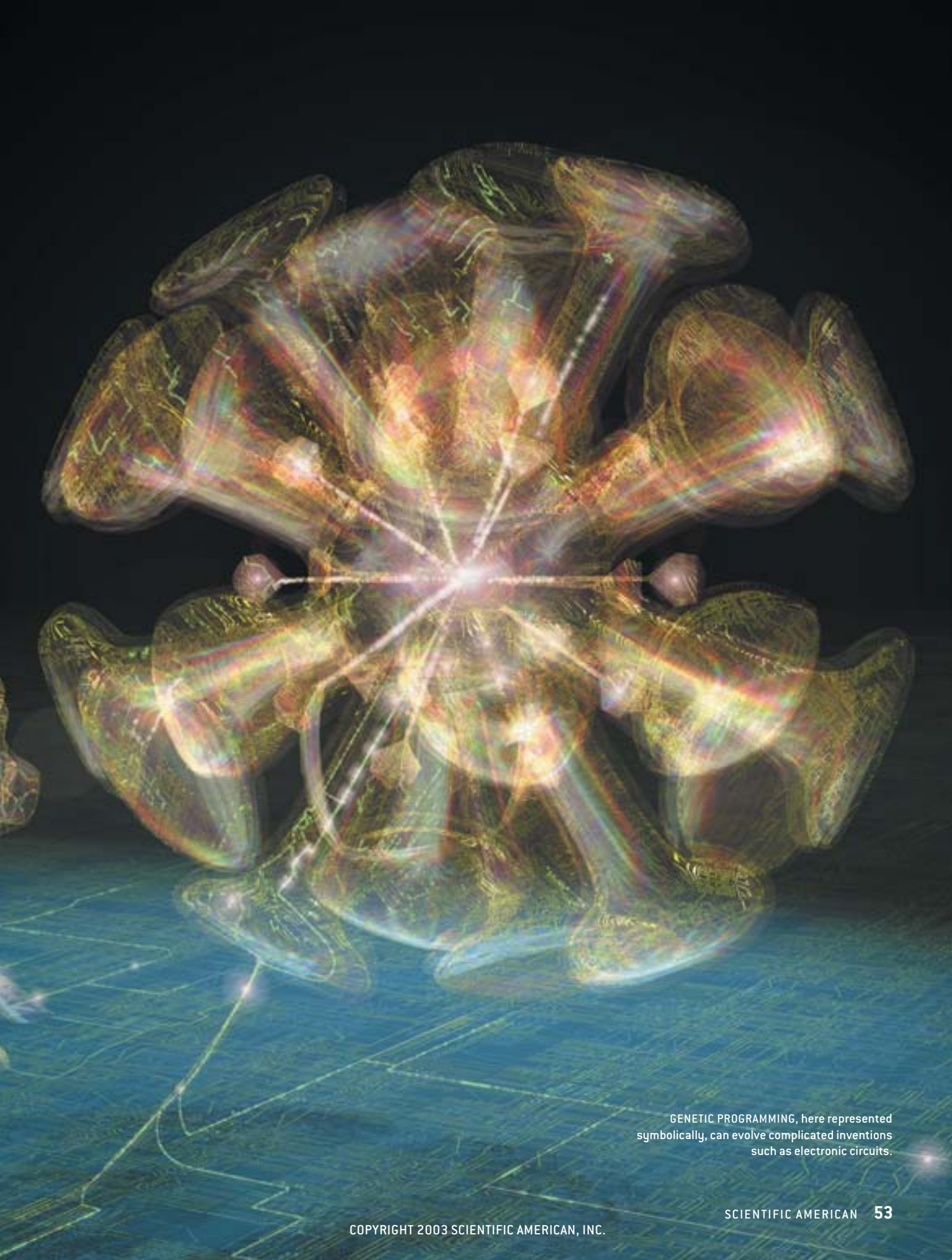
Evolution is an immensely powerful creative process. From the intricate biochemistry of individual cells to the elaborate structure of the human brain, it has produced wonders of unimaginable complexity. Evolution achieves these feats with a few simple processes—mutation, sexual recombination and natural selection—which it iterates for many generations. Now computer programmers are harnessing software versions of these same processes to achieve machine intelligence. Called genetic programming, this technique has designed computer programs and electronic circuits that perform specified functions.

In the field of electronics, genetic programming has duplicated 15 previously patented inventions, including several that were hailed as seminal in their respective fields when they were first announced [*see box on page 57*]. Six of these 15 existing inventions were patented after January 2000 by major research institutions, which indicates that they represent current frontiers of research in domains of scientific and practical importance. Some of the automatically produced inventions infringe squarely on the exact claims of the previously patented inventions. Others represent new inventions by duplicating the functionality of the earlier device in a novel way. One of these inventions is

Illustrations by Bryan Christie Design

Computer programs that function via Darwinian evolution are creating inventions that are novel and useful enough to be patented

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.



GENETIC PROGRAMMING, here represented symbolically, can evolve complicated inventions such as electronic circuits.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Genetic programming begins with a primordial ooze of randomly generated "organisms."

a clear improvement over its predecessor.

Genetic programming has also classified protein sequences and produced human-competitive results in a variety of areas, such as the design of antennas, mathematical algorithms and general-purpose controllers [*see box on page 59*]. We have recently filed for a patent for a genetically evolved general-purpose controller that is superior to mathematically derived controllers commonly used in industry.

The first practical commercial area for genetic programming will probably be design. In essence, design is what engineers do eight hours a day and is what evolution does. Design is especially well suited to genetic programming because it presents tough problems for which people seek solutions that are very good but not mathematically perfect. Generally there are complex trade-offs between competing considerations, and the best balance among the various factors is difficult to foresee. Finally, design usually involves discovering topological arrangements of things (as opposed to merely optimizing a set of numbers), a task that genetic programming is very good at.

Human engineers tend to look at problems in particular ways, often based on ideal mathematical models. Genetic programming offers the advantage of not being channeled down narrow paths of thinking. Evolution does not know anything about the underlying math; it simply tries to produce a sequence of improved results. Thus, we frequently see creative things come out of the evolutionary process that would never occur to human designers.

Out of the Primordial Ooze

WHATEVER THE FIELD of endeavor, genetic programming begins with a primordial ooze of randomly generated trial "organisms" and a high-level description of what function the organisms are meant to accomplish—the criteria for scoring their fitness. As an example, consider a case in which the organisms are elementary mathematical functions and we are endeavoring to find a function whose graph matches a given curve. The organisms in this case are composed of numerical constants and primitive operations such as addition, subtraction, multiplication and division. The fitness of a function is determined by how closely its graph follows the target curve.

The genetic program evaluates the fitness of each mathematical function in the population. The initial, randomly created functions will, of course, match the target curve quite poorly, but some will be better than others. The genetic program tends to discard the worst functions in the population, and it applies genetic operations to the surviving functions to create offspring. The most important genetic operation is sexual reproduction, or crossover, which mates pairs of the better organisms to sire offspring composed of genetic material from the two parents [*see top illustration on opposite page*]. For instance, mating the functions $(a + 1) - 2$ and $1 + (a \times a)$ might result in the $(a + 1)$ part of the first function substituting for one a of the second function, producing offspring $1 + ((a + 1) \times a)$. Recombining the traits of two relatively fit organisms in this fashion sometimes produces superior offspring.

In addition to sexual reproduction, genetic programming copies about 9 percent of the fittest individuals unaltered into the next generation, which generally ensures that the best organisms in each generation are at least as fit as those of the previous generation. Finally, about 1 percent of the programs undergo mutation—for instance, $a + 2$ might mutate into $(3 \times a) + 2$—in the hope that a random modification of a relatively fit program will lead to improvement.

These genetic operations progressively produce an improved population of mathematical functions. The exploitation of small differences in fitness yields major improvements over many generations in much the same way that a small interest rate yields large growth when compounded over decades.

One can visualize the evolutionary process as being a search in the space of all possible organisms. The crossover operation conducts the most creative kind of search, which is why we use it to produce around 90 percent of the offspring in each generation [*see bottom illustration on opposite page*]. Mutation, in contrast, tends to conduct a local search for advantage near the existing good individuals. We believe that too great a mutation rate results in less efficient evolution except in the case of particularly simple problems.

A more sophisticated example than a mathematical function is the evolution of

Overview/*Darwinian Invention*

- Genetic programming harnesses a computerized version of evolution to create new inventions. Starting from thousands of randomly generated test objects, the method selects the better individuals and applies processes such as mutation and sexual recombination to generate successive generations.
- Over the course of dozens of generations, the population of individuals gradually fulfills the target criteria to a greater degree. At the end of the run, the best individual is harvested as the solution to the posed problem.
- In electronics, the technique has reproduced patented inventions, some of which lie at the forefront of current research and development. Other inventions include antennas, computer algorithms for recognizing proteins, and general-purpose controllers. Some of these computer-evolved inventions should themselves be patentable.
- By the end of the decade, we envision that increased computer power will enable genetic programming to be used as a routine desktop invention machine competing on equal terms with human inventors.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

a computer program, such as one employing iterations and memory for classifying protein sequences. In this case, genetic programming can carry out analogues of the biological processes of gene duplication and gene deletion, to create or delete subroutines, iterations, loops and recursions in the evolving population of programs. The evolutionary process itself determines the character and content of the computer program needed to solve the problem.

A low-pass filter circuit provides a good illustration of how genetic programming designs analog electronic circuits. A low-pass filter is used in a hi-fi system to send only the lowest frequencies to the woofer speaker. To create a low-pass filter by using genetic programming, the human user specifies which components are available for building the circuit (say, resistors, capacitors and inductors) and defines the fitness of each candidate circuit to be the degree to which it passes frequencies up to 1,000 hertz at full power while filtering out all higher frequencies.

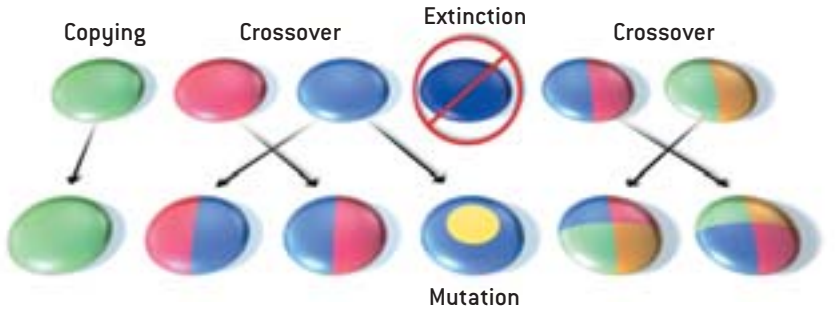
The circuits are generated in a way that borrows mechanisms from developmental biology. Each circuit begins as an elementary "embryo" consisting of a single wire running from the input to the output. The embryonic circuit grows by progressive application of circuit-constructing functions. Some of the circuit-constructing functions insert particular components. Others modify the pattern of connections between components: they might duplicate an existing wire or component in series or parallel, or they might create a connection from a particular point to a power supply, the ground or a distant point in the growing circuit. This developmental process yields both the circuit topology and the numerical component sizes. The system automatically synthesizes circuits without using any advanced know-how from the field of electrical engineering concerning circuit synthesis.

Most of the initial population of rudimentary circuits generated randomly in this way will behave nothing like a low-pass filter. A few, however, will contain an inductor between the circuit's input and output, thereby slightly imped-

UNNATURAL SELECTION

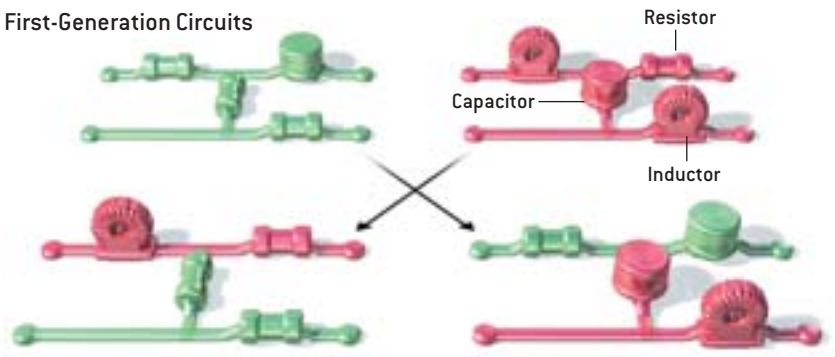
Evolutionary Processes

THREE PROCESSES propagate "organisms" (*represented here by colored disks*) from one generation to the next in a genetic programming run. Some of the better organisms are copied unaltered. Others are paired up for sexual reproduction, or crossover, in which parts are swapped between the organisms to produce offspring. A small percentage are changed randomly by mutation. Organisms not chosen for propagation become extinct. The crossover operation is applied more frequently than copying and mutation because of its ability to bring together new combinations of favorable properties in individual organisms.



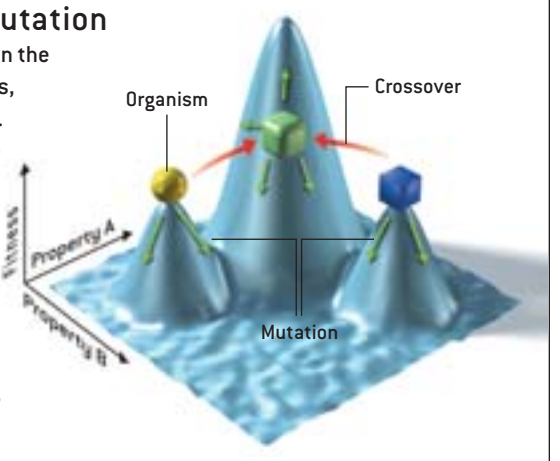
Crossover of Electronics

ACTING ON electronic circuits, the crossover operation takes two circuits and swaps some of their components, producing two new circuits.



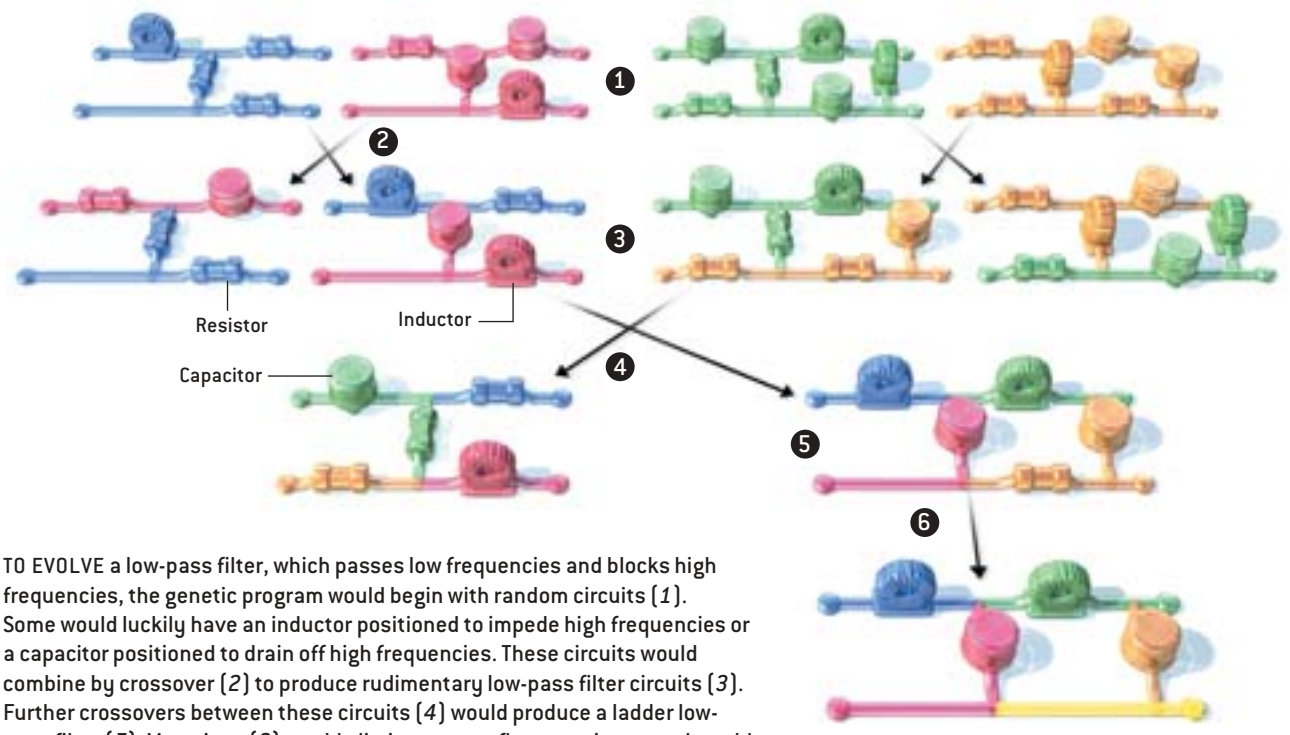
Crossover versus Mutation

EVOLUTION ACTS like a search in the space of all possible organisms, represented here by the plane. Crossover searches this space creatively, occasionally combining disparate good features, leaping to a new region of organism space where much fitter individuals reside (*red arrows*). Mutation, in contrast, tends to find the best organism that is "nearby" (*green arrows*).



COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

EVOLVING A LOW-PASS FILTER



TO EVOLVE a low-pass filter, which passes low frequencies and blocks high frequencies, the genetic program would begin with random circuits (*1*). Some would luckily have an inductor positioned to impede high frequencies or a capacitor positioned to drain off high frequencies. These circuits would combine by crossover (*2*) to produce rudimentary low-pass filter circuits (*3*). Further crossovers between these circuits (*4*) would produce a ladder low-pass filter (*5*). Mutations (*6*) would eliminate superfluous resistors and would fine-tune the values of the components.

ing higher frequencies. Others will have a capacitor running from the input to the ground, thereby slightly draining the power of higher frequencies [*see illustration above*]. Such circuits will be selected to mate more frequently than others, and eventually later generations will contain offspring incorporating both features. The crossover and mutation operations acting on numerical expressions will adjust component values so that the cutoff frequency approaches the desired 1,000 Hz. Other crossovers and mutations will delete resistors that dissipate power. Additional crossovers will double or triple the inductor-capacitor combination, yielding the ladder structure patented in 1917 by George A. Campbell of AT&T.

Other devices are designed with similar combinations of evolutionary and developmental processes. Antennas, for instance, are automatically designed with a "turtle" that deposits (or does not deposit) metal on a plane as it moves and turns under the control of various operations (similar to those in the LOGO programming language).

The primitive ingredients used to create controllers automatically consist of differentiators, integrators and amplifiers. An example of a basic controller is a cruise control on a car, which must reduce fuel intake if the speed rises too high or increase it if the speed falls too low. A good controller will allow for the delayed response to fuel changes and will continuously monitor how the speed is varying to avoid excessive overshooting of the target speed. Of great importance are general-purpose controllers, which can be customized to a variety of specific tasks—such as the control of a home furnace, manufacturing processes in factories or the reading arm of a computer's disk storage device. Small improvements in the "tuning rules" used in customizing a controller can result in large economic savings.

A commonplace controller is the PID controller invented in 1939 by Albert Callender and Allan Stevenson of Imperial

THE AUTHORS

JOHN R. KOZA, MARTIN A. KEANE and *MATTHEW J. STREETER* work closely with one another studying genetic programming using a home-built, 1,000-Pentium parallel computer. Koza received his Ph.D. in computer science from the University of Michigan in 1972. He co-founded Scientific Games, Inc., in Atlanta in 1973, where he co-invented the rub-off instant lottery ticket used by state lotteries. In 1987 Koza invented genetic programming. He is currently consulting professor in the Stanford Biomedical Informatics program in the department of medicine and consulting professor in the university's department of electrical engineering. Keane received a Ph.D. in mathematics from Northwestern University in 1969. From 1976 to 1986 he was vice president for engineering at Bally Manufacturing Corporation in Chicago. He is now chief scientist at Econometrics, Inc., also in Chicago. Streeter received a master's degree from Worcester Polytechnic Institute in 2001. His primary research interest is applying genetic programming to problems of real-world scientific or practical importance. He works at Genetic Programming, Inc., in Los Altos, Calif., as a systems programmer and researcher.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The first patent for an invention created by genetic programming may soon be granted.

Chemical Limited in Northwich, England. PID controllers (the initials stand for the controller's *p*roportional, *i*ntegrative and *d*erivative parts) are used in myriad applications. Our genetic programs have evolved two distinct improvements in this field. First, they have developed a new set of tuning rules for PID controllers. A relatively simple and effective set of PID tuning rules has been in general use since 1942 and was improved on in 1995; our rules outperform the 1995 rules. Second, we evolved three new controller circuit topologies that also outperform PID controllers that use the old tuning rules. We have filed a patent application that covers both the new rules and the new controller topologies. If (as we expect) the patent is granted, we believe that it will be the first one granted for an invention created by genetic programming.

Evolvable Hardware

DURING THE EVOLUTIONARY process, we must efficiently evaluate the fitness of thousands or millions of offspring in each generation. For electronic circuits, we usually use standard circuit-simulator software to predict the behavior of each circuit in the population. In an important emerging area of technology called evolvable hardware, however, microchips can be instantaneously configured to physically implement each circuit of a genetic programming run.

Known as rapidly reconfigurable field-programmable gate arrays, these chips consist of thousands of identical cells, each of which can perform numerous different logical functions, depending on how it is programmed. Sets of memory bits in the "basement" of the chip customize each cell so that it performs a particular logical function. Other configuration bits program interconnection routes on the chip, permitting many different ways of connecting the cells to one another and to the chip's input and output pins. The "personality" of the chip (its logical functions and interconnections) can be changed dynamically in nanoseconds merely by changing its configuration bits.

These rapidly reconfigurable chips are

Patented Inventions Re-created by Computer

TO DATE, genetic programming has re-created 15 inventions that were previously patented by the inventors listed here.

INVENTION	YEAR	INVENTOR	INSTITUTION
LADDER FILTER	1917	George A. Campbell	AT&T, *New York City*
CROSSOVER FILTER	1925	Otto Julius Zobel	AT&T
NEGATIVE FEEDBACK AMPLIFIER	1927	Harold S. Black	AT&T
ELLIPTIC FILTER	1934–36	Wilhelm Cauer	University of Göttingen, *Germany*
PID (proportional, integrative and derivative) CONTROLLER	1939	Albert Callender and Allan Stevenson	Imperial Chemical Limited, *Northwich, England*
SECOND-DERIVATIVE CONTROLLER	1942	Harry Jones	Brown Instrument Company, *Philadelphia*
DARLINGTON EMITTER-FOLLOWER SECTION	1953	Sidney Darlington	Bell Telephone Laboratories, *New York City*
PHILBRICK CIRCUIT	1956	George A. Philbrick	George A. Philbrick Researches, *Boston*
SORTING NETWORK	1962	Daniel G. O'Connor and Raymond J. Nelson	General Precision, *Los Angeles*
MIXED ANALOG-DIGITAL INTEGRATED CIRCUIT for producing variable capacitance	2000	Turgut Sefket Aytur	Lucent Technologies, *Murray Hill, N.J.*
VOLTAGE-CURRENT CONVERTER	2000	Akira Ikeuchi and Naoshi Tokuda	Mitsumi Electric, *Tokyo*
CUBIC FUNCTION GENERATOR	2000	Stefano Cipriani and Anthony A. Takeshian	Conexant Systems, *Newport Beach, Calif.*
LOW-VOLTAGE, HIGH-CURRENT TRANSISTOR CIRCUIT for testing a voltage source	2001	Timothy Daun-Lindberg and Michael Miller	IBM, *Armonk, N.Y.*
LOW-VOLTAGE BALUN CIRCUIT	2001	Sang Gug Lee	Information and Communications University, *Taejon, Korea*
TUNABLE INTEGRATED ACTIVE FILTER	2001	Robert Irvine and Bernd Kolb	Infineon Technologies, *Munich, Germany*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

sold by about a dozen companies, but they are primarily of use for digital circuits. Commercially available analog chips are extremely limited in their abilities. We used a reconfigurable digital chip to create a sorting network with fewer steps than the originally patented version.

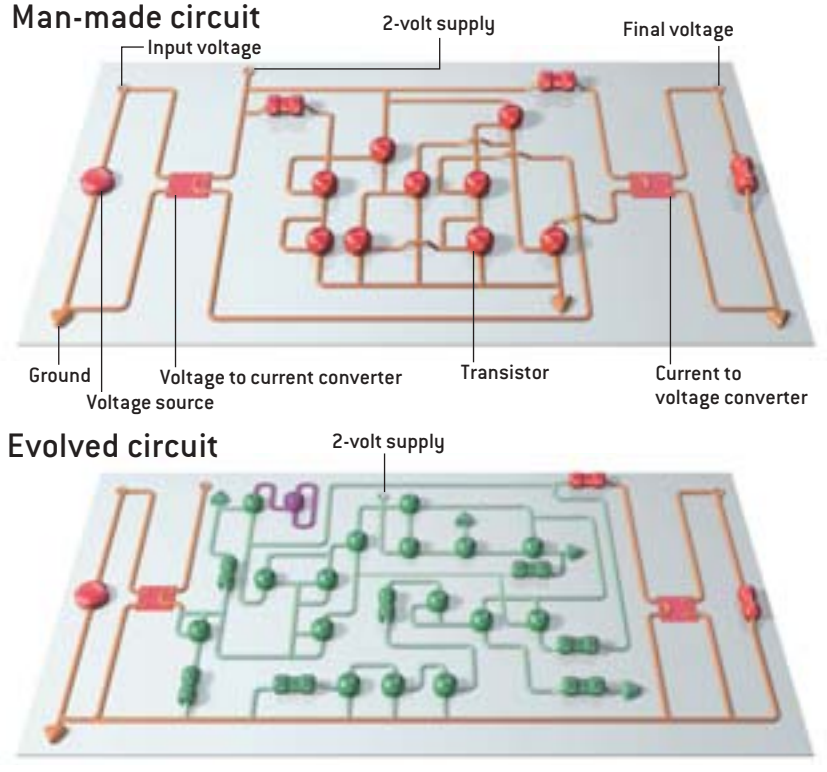
Run Times

NATURAL EVOLUTION has had billions of years of "run time" to produce its wonders. Artificial genetic programming would not be of much use if it took as long. A genetic programming run typically spawns a population of tens or hundreds of thousands of individuals that evolve over dozens or hundreds of generations. A weeklong run on a laptop computer is sufficient to produce half of the human-competitive results listed in the box on the preceding page; however, all six of the inventions patented after 2000 required more horsepower than that.

Evolution in nature thrives when organisms are distributed in semi-isolated subpopulations. The same seems to be true of genetic programming run on a loosely connected network of computers. Each computer can perform the time-consuming step of evaluating the fitness of individuals in its subpopulation. Then, at the end of each generation, a small percentage of individuals (selected based on fitness) migrates to adjacent computers in the network so that each semi-isolated subpopulation gets the benefit of the evolutionary improvement that has occurred elsewhere.

HUMAN VERSUS COMPUTER

THE TWO CIRCUITS shown below are both cubic signal generators. The upper circuit is a patented circuit designed by a human; the green and purple parts of the lower circuit were evolved by genetic programming (the other parts are standard input and output stages). The evolved circuit performs with better accuracy than the human-designed one, but *how* it functions is not understood. The evolved circuit is clearly more complicated but also contains redundant parts, such as the purple transistor, that contribute nothing to its functioning.



We have built a Beowulf-style computer cluster consisting of 1,000 somewhat outdated 350-megahertz Pentium computers [see "The Do-It-Yourself Supercomputer," by William W. Hargrove, Forrest M. Hoffman and Thomas Sterling; SCIENTIFIC AMERICAN, August 2001]. For our most time-consuming problems, evaluation of the fitness of a single candidate individual takes about a minute of computer time. A run involving a population of 100,000 individuals for 100 generations can be completed in about seven days on our cluster.

The 1,000 computers together perform about 350 billion cycles a second. Although this amount of computer time may, at first blush, sound like a lot, it pales in comparison to the amount of computation performed by the trillion cells of the human brain (each of which is thought to have about 10,000 connections and operate at a rate of 1,000 operations a second).

We expect that 50-gigahertz computers (performing 50 billion cycles a second) will be commonly available toward the end of this decade, putting the power to evolve patent-worthy inventions using genetic programming in the hands of anyone owning a moderately priced desktop workstation. We envision that genetic programming will be regularly used as an invention machine.

Passing an Intelligence Test

GENETIC PROGRAMMING is now routinely reproducing human inventions, just half a century after computer pioneer Alan M. Turing predicted that human-competitive machine intelligence would be achieved in about 50 years. During those 50 years, the two main academically fashionable approaches taken by researchers striving to vindicate Turing's prediction have used logical deduction or databases containing accumulated human knowledge and expertise (so-called expert systems). Those two approaches roughly correspond to two broad methods outlined by Turing in 1950. The first (not surprising in light of Turing's work in the 1930s on the logical foundations of computing) was the construction of programs designed to an-

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

More Human-Competitive Creations

AS WELL AS re-creating patented inventions, genetic programming has generated these results that a human would be proud of.

SOCCER-PLAYING PROGRAM that ranked in the middle of the field of 34 human-written programs in the RoboCup 1998 competition
REAL-TIME ANALOG CIRCUIT for time-optimal control of a robot
FOUR DIFFERENT ALGORITHMS for identifying transmembrane segments of proteins
DERIVING MOTIFS (highly conserved sequences of amino acids) to identify certain families of proteins
ALGORITHMS FOR QUANTUM COMPUTERS that in some cases solve problems better than any previously published result
NAND CIRCUIT for carrying out the NOT AND logical operation on two inputs
ANALOG COMPUTATIONAL CIRCUITS for the square, cube, square root, cube root, logarithm and Gaussian functions
DIGITAL-TO-ANALOG CONVERTER CIRCUIT
ANALOG-TO-DIGITAL CONVERTER CIRCUIT

alyze situations and problems logically and to respond accordingly. The second, which Turing called a cultural search, applied knowledge and expertise gathered from experts.

The goal of artificial intelligence and machine learning is to get computers to solve problems from a high-level statement of what needs to be done. Genetic programming is delivering human-competitive machine intelligence with a minimum of human involvement for each new problem and without using either logical deduction or a database of human knowledge.

Turing also proposed a famous test for machine intelligence. In one widely used restatement of the Turing test, a judge receives messages "over a wall" and tries to decide whether the messages came from a human or a machine. We do not claim that genetic programming has achieved the kind of general imitation of human cognition associated with the Turing test. But it *has* passed a test of creativity and ingenuity that only a relatively small number of humans pass. The U.S. patent office has been administering this test for more than 200 years.

The patent office receives written descriptions of inventions and then judges whether they are unobvious to a person having ordinary skill in the relevant field. Whenever an automated method duplicates a previously patented human-designed invention, the automated method has passed the patent office's intelligence test. The fact that the original, human-designed version satisfied the patent office's criteria of patent-worthiness means that the computer-created duplicate would also have satisfied the patent office.

This intelligence test does not deal with inconsequential chitchat or the playing of a game. When an institution or individual allocates time and money to invent something and to embark on the time-consuming and expensive process of obtaining a patent, it has made a judgment that the work is of scientific or practical importance. Moreover, the patent office requires that the proposed invention be useful. Patented inventions represent nontrivial work by exceptionally creative humans.

Although some people may be surprised that routine human-competitive machine intelligence has been achieved with a nondeterministic method and without resorting to either logic or knowledge, Alan Turing would not be. In his 1950 paper, Turing also identified this third approach to machine intelligence: "the genetical or evolutionary search by which a combination of genes is looked for, the criterion being the survival value."

Turing did not specify how to conduct a "genetical or evolutionary search" to achieve machine intelligence, but he did point out that:

> We cannot expect to find a good child-machine at the first attempt. One must experiment with teaching one such machine and see how well it learns. One can then try another and see if it is better or worse. There is an obvious connection between this process and evolution, by the identifications
>
> Structure of the child machine
> = Hereditary material
> Changes of the child machine
> = Mutations
> Natural selection
> = Judgment of the experimenter

Genetic programming has in many ways fulfilled the promise of Turing's third way to achieve machine intelligence. SA

MORE TO EXPLORE

Computing Machinery and Intelligence. Alan M. Turing in *Mind,* Vol. 59, No. 236, pages 433–460; October 1950. Available at **www.abelard.org/turpap/turpap.htm** by permission of Oxford University Press.

Genetic Programming: On the Programming of Computers by Means of Natural Selection. John R. Koza. MIT Press, 1992.

Genetic Programming: The Movie. John R. Koza and James P. Rice. MIT Press, 1992.

Genetic Programming III: Darwinian Invention and Problem Solving. John R. Koza, Forrest H Bennett III, David Andre and Martin A. Keane. Morgan Kaufmann, 1999.

Genetic Programming III: Videotape: Human-Competitive Machine Intelligence. John R. Koza, Forrest H Bennett III, David Andre, Martin A. Keane and Scott Brave. Morgan Kaufmann, 1999.

Genetic Programming IV: Routine Human-Competitive Machine Intelligence. John R. Koza, Martin A. Keane, Matthew J. Streeter, William Mydlowec, Jessen Yu and Guido Lanza. Kluwer Academic Publishers (in press).

More information can be obtained from Genetic Programming, Inc. (**www.genetic-programming.com**), and the Genetic Programming Conference organization (**www.genetic-programming.org**)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Explaining Frog Deformities

An eight-year investigation into the cause of a shocking increase in deformed amphibians has sorted out the roles of three prime suspects

ONE HOT SUMMER DAY in 1995 eight middle school children planning a simple study of wetland ecology began collecting leopard frogs from a small pond near Henderson, Minn. To their astonishment, one captured frog after another had five or more hind legs, some twisted in macabre contortions. Of the 22 animals they caught that day, half were severely deformed. A follow-up search by pollution-control officials added to the gruesome inventory. Occasional frogs in the pond had no hind limbs at all or had mere nubbins where legs should be; others had one or two legs sprouting from the stomach. A few lacked an eye.

The story seized national media attention and raised many questions—among them, was this an isolated occurrence or one facet of a widespread trend? And what caused the deformities? As researchers elsewhere in the country began investigating their local amphibian populations, it became clear that this bizarre collection of ailments was not confined to Minnesota. Since 1995, malformations have been reported in more than 60 species, including salamanders and toads, in 46 states. In some local populations 80 percent of the animals are afflicted. International reports show that this phenomenon extends beyond the U.S. Surprising numbers of deformed amphibians have been found in Asia, Europe and Australia as well. Worldwide, extra legs and missing legs are most common.

The aberrations cannot be discounted as being a normal part of amphibian life. Research dating back to the early 1900s indicates that a few individuals in every population have defects resulting naturally from genetic mutation, injury or developmental problems. In healthy populations, however, usually no more than 5 percent of animals have missing limbs or digits; extreme deformities, such as extra hind legs, are even less common. Moreover, fresh reviews of historical records by one of us (Johnson) and new field studies indicate that deformities have become more prevalent in recent times.

Over the past eight years, dozens of investigators have blamed the increase on the amphibians' greater exposure to ultraviolet radiation, on chemical contamination of water or on a parasite epidemic. Not surprisingly, every time another report appeared, media outlets touted the new view, thus providing a misleading picture of the situation. It turns out that all these factors probably operate to varying extents, each causing particular disfigurements, and that all three may at times act in concert. Moreover, all stem in part from human activities such as habitat alteration.

Deformities undoubtedly impair amphibian survival and most likely contribute to the dramatic declines in populations that have been recognized as a global concern since 1989 [see "The Puzzle of Declining Amphibian Populations," by Andrew R. Blaustein and David B. Wake; SCIENTIFIC AMERICAN, April 1995]. Both trends are disturbing in their own right and are also a warning for the planet [*see box on page 63*]. Amphibians have long been regarded as important indicators of the earth's health because their unshelled eggs and permeable skin make them extremely sensitive to perturbations in the environment. Chances are good that factors affecting these animals harshly today are also beginning to take a toll on other species.

An Early Suspect

ONE PUTATIVE CAUSE of the deformities, excess exposure to ultraviolet radiation, came under suspicion almost as soon as the malformations were discovered, because it had already been implicated in declines of amphibian populations and because laboratory work had shown it to be capable of disrupting amphibian development. This form of radiation—which can damage immune systems and cause genetic mutations, among other effects—has been reaching the earth in record doses since chlorofluorocarbons and other human-made

ALARMING LEGS: Ecologist Andrew R. Blaustein eyes malformed Pacific tree frogs collected from farm ponds in northwestern Oregon.

FRANS LANTING *Minden Pictures*

BY ANDREW R. BLAUSTEIN AND PIETER T. J. JOHNSON

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Chances are good
that the factors affecting amphibians are also taking a toll on other species.

chemicals began thinning the protective layer of ozone in the stratosphere, a problem first measured in the 1970s. Between 1994 and 1998 one of us (Blaustein) and his colleagues demonstrated that exposure to ultraviolet rays can kill amphibian embryos and larvae, cause serious eye damage in adult frogs, and induce various types of bodily deformities in frogs and salamanders.

Whether exposure to ultraviolet radiation could disrupt leg development remained uncertain until the late 1990s, when Gary Ankley and his co-workers at the Environmental Protection Agency in Minnesota carried out the most focused experimental research on this question to date. When the investigators shielded developing frogs from ultraviolet rays, the animals grew normal limbs, whereas tadpoles exposed to full doses of natural levels of ultraviolet radiation developed with parts of their legs missing or without digits. These deformities resembled some of those found in wild frogs from several sites around the country.

The EPA team was quick to point out, however, that ultraviolet radiation does not explain all types of leg deformities seen in nature. Most notably, it does not lead to the growth of *extra* legs, one of the deformities reported most frequently since 1995. Many laboratory and field experiments, several performed by Blaustein and his colleagues, have come to the same conclusion. Other biologists have also pointed out that many wild amphibians can avoid the continuous exposure to radiation studied in the EPA experiments. Juvenile and adult amphibians alike can move in and out of sunlight, often live in muddy water, or may be nocturnal.

Pollution's Part

AS SOME RESEARCHERS were examining the link between ultraviolet radiation and deformities, others were pursuing the influence of water pollution, such as pesticide runoff. They focused on pollution because so many of the early reports of amphibian ailments came from areas where large amounts of insecticides and fertilizers are applied every year. By the mid-1990s numerous laboratory studies had shown that myriad contaminants can kill amphibians, but it was unclear whether they could induce extra or incomplete limb formation.

A major challenge for toxicologists is isolating a single chemical or even a group of chemicals as a likely candidate. Millions of tons of hundreds of different pollutants are applied annually in regions where deformed amphibians have been found. Yet one chemical rose immediately to the top of the list: methoprene. First approved for commercial use in 1975, methoprene was promoted as a safer replacement for the banned pesticide DDT.

Initial concern over methoprene came from its chemical similarity to compounds called retinoids. These substances, especially retinoic acid, play an integral role in vertebrate development; too little or too much can lead to deformities in embryos. Indeed, numerous miscarriages and birth defects in humans have resulted from pregnant women's use of acne medicines that contain a retinoic acid derivative.

Some biologists suspected methoprene might have a similar effect on frogs. In a series of experiments in the late 1990s, the EPA in Minnesota did show that high amounts of retinoic acid could trigger poor formation of hind limbs in frogs, but comparable tests with methoprene caused no malformations at all. Separate field measurements also indicated that the pesticide could not be the sole cause. Methoprene breaks down quickly in the environment, and investigators found little evidence that it persists where deformities are abundant. The same is true for 61 other agricultural chemicals and their breakdown products that have been measured in locations that harbor malformed animals throughout the western U.S. Pesticides are not off the hook, however. Hundreds remain untested, and some evidence implies that certain pesticides can cause bodily damage (albeit not the formation of extra limbs).

At the moment, then, laboratory research tentatively suggests that water pollutants and ultraviolet radiation are capable of causing disfigurement. But a more potent threat appears to have a much broader impact in nature.

Prolific Parasites

THE EARLIEST HINTS of this threat—apparently the cause of the widespread hind-leg anomalies—turned up long before the disturbing findings in Minnesota garnered nationwide attention. In the mid-1980s, Stephen B. Ruth, then at Monterey Peninsula College, was exploring ponds in northern California when he found hundreds of Pacific tree frogs and long-toed salamanders with missing legs, extra legs and other deformities. He sounded no alarm, however, because he assumed he was seeing an isolated oddity.

In 1986 Ruth asked Stanley K. Ses-

Overview/*Amphibian Ailments*

- Since the mid-1990s striking deformities have turned up in more than 60 species of frogs, toads and salamanders in 46 states and on four continents. The number of disfigured animals in some populations averages around 25 percent—significantly higher than in previous decades.
- Contradictory reports have blamed the deformities on increasing exposure to ultraviolet radiation, contaminated water or a parasite epidemic.
- New evidence indicates that the parasite epidemic accounts for one of the most prevalent deformities—extra hind legs—and strongly suggests that human activities such as habitat alteration are exacerbating the problem.

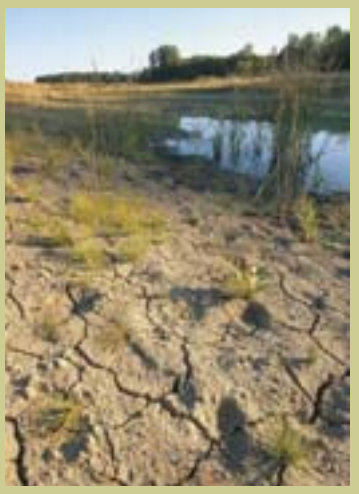
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Disfigured and Dwindling

Do the deformities explain recent declines in amphibian populations?

FROGS, TOADS AND SALAMANDERS have been climbing up the long list of creatures in danger of disappearing from the earth entirely ever since the first reports of dwindling populations were made 20 years ago. An obvious question for biologists is to what degree physical deformities are contributing to overall population declines.

Most malformed amphibians eventually vanish from a population because they can neither escape their predators nor hunt for food efficiently. Events known to increase the number of animals that mature into disabled adults—such as the parasite epidemic that is currently afflicting dozens of sites across North America—could cause a whole population to crash, particularly if the incidence of deformities continues to increase. Although such crashes may be occurring at some sites, numerous amphibian populations have declined severely in the absence of any deformities, leaving researchers to conclude that deformities are far from the sole basis for the declines. Environmental hazards seem to be a more significant cause.



CRACKED EARTH signals the temporary demise of a frog habitat in the U.S. West.

Amphibian species inhabit a wide variety of ecosystems, including deserts, forests and grasslands, from sea level to high mountains. But as diverse as their niches are, few are shielded completely from a nearly equal variety of insults that humans inflict on them. Some important amphibian habitats have been totally destroyed or are polluted to an intolerable degree. In other cases, people have introduced foreign animals that either devoured or pushed out the native amphibians.

Some of the most widespread alterations may lead to both population declines and deformities. Many studies have shown, for instance, that excess ultraviolet radiation—resulting from human-induced ozone loss in the upper atmosphere—can inhibit limb formation in amphibian juveniles or even kill embryos inside their vulnerable, unshelled eggs. In the future, global warming is expected to dry out certain suitable aquatic habitats while elsewhere encouraging the emergence of infections that produce abnormal development. When it comes to problems as pressing as these, tackling declines will most likely help alleviate the deformities as well. *—A.R.B. and P.T.J.J.*

sions, now at Hartwick College in New York State, to inspect his bizarre amphibians. Sessions agreed and quickly realized that they were all infected with a parasitic trematode, known commonly as a flatworm or a fluke. The California trematodes—whose specific identity was unknown at the time—did not appear to kill their hosts outright, but Sessions suspected that their presence in a tadpole mechanically disturbed natural development wherever the parasites formed cysts in the body, most frequently near the hind legs. To test this hypothesis, he simulated trematode cysts by implanting small glass beads in developing limb buds of African clawed frogs and a salamander known as an axolotl. These two species, which serve as the "white rats" of amphibian biology because they are easy to breed in captivity, developed extra legs and other abnormalities—much as if they were parasitized.

As intriguing as those experimental results were, though, they could not prove that trematodes were responsible for the deformities in Ruth's specimens: African clawed frogs and axolotls are not known to have limb deformities in nature. The research that would eventually connect the dots between trematodes and extra or missing legs in frogs was conducted after the stir of 1995. At that time, Johnson pored over the scientific literature for clues to the cause and came across the discoveries made by Ruth and Sessions. Johnson and his colleagues then conducted broader surveys of California wetlands between 1996 and 1998 and discovered that ponds where tree frogs had abnormal limbs also had an abundance of the aquatic snail *Planorbella tenuis*, one in a series of hosts colonized by Sessions's trematode, now known to be *Ribeiroia ondatrae*.

Thinking that they may well have uncovered a direct correlation between a parasite epidemic and amphibian deformities in the wild, Johnson and his team immediately collected deformed frogs from the same ponds and dissected them. In every case, they found cysts of the parasite densely clustered just below the skin around the base of the hind legs. To test the idea that the trematodes were triggering the growth of the extra limbs, the researchers then exposed Pacific tree frog tadpoles to *R. ondatrae* parasites in the lab. As expected, the infected tree frogs developed deformities identical to those

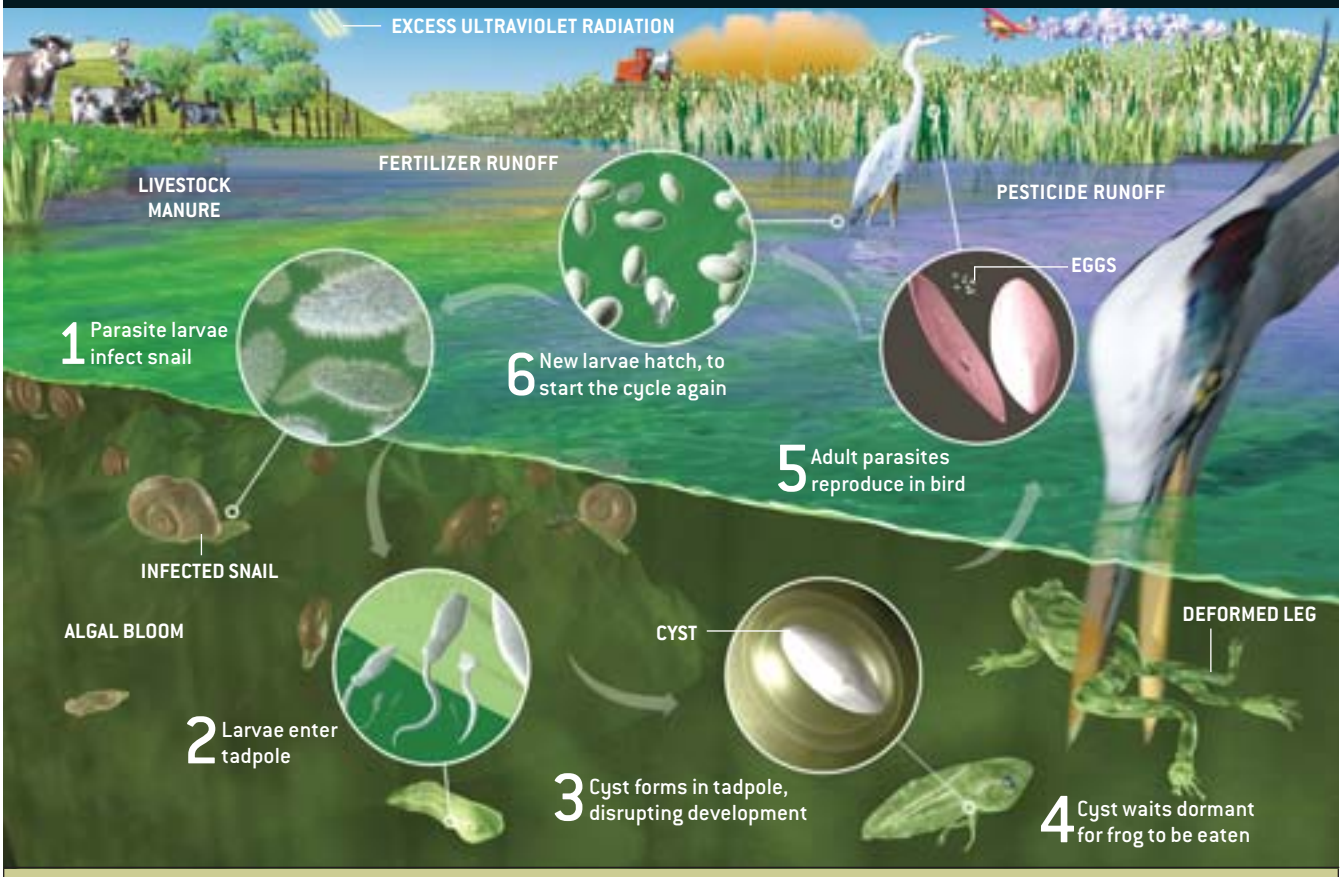
THE AUTHORS

ANDREW R. BLAUSTEIN and *PIETER T. J. JOHNSON* began exploring the potential causes of amphibian deformities as a team in 1998. Blaustein, who earned a Ph.D. in 1978 from the University of California, Santa Barbara, is professor in the zoology department at Oregon State University. A behavioral and population ecologist by training, he has spent the past several years investigating the dynamics of worldwide declines in amphibian populations, specifically addressing the effects of ultraviolet radiation, pollutants, pathogens and nonnative species. Johnson, a doctoral candidate at the Center for Limnology at the University of Wisconsin–Madison, studies human influences on emerging diseases in aquatic environments.

FRANS LANTING *Minden Pictures*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

HOW PARASITES CAN CRIPPLE FROGS



LIFE CYCLE of the trematode *Ribeiroia ondatrae* enables the parasite to induce deformities—including extra hind legs—in generation after generation of frogs. In its first larval form the trematode infects snails (*1*). After transforming into a second free-swimming form inside a snail, the parasite embeds itself near a tadpole's future hind leg (*2*). There it forms a cyst that disrupts normal limb development and can cause the tadpole to sprout extra legs as it grows into a frog (*3*). The disabled frog then becomes easy prey for the parasite's final host, often a heron or egret (*4*). The parasite matures and reproduces inside the bird, which releases trematode eggs into the water with its feces (*5*). When larvae hatch (*6*), they begin the cycle again. Human activities can exacerbate this process, especially where livestock manure or fertilizers enter a pond and trigger algal blooms that nourish, and thus increase, snail populations. Excess ultraviolet radiation and pesticide runoff—which might cause other types of deformities when acting alone—may facilitate the cycle by weakening a tadpole's immune system and making the animal more vulnerable to parasitic infection.

—A.R.B. and P.T.J.J.

found in nature, including extra limbs and missing limbs. Higher levels of infection led directly to more malformations, whereas uninfected frogs developed normally.

This study turned out to be a key breakthrough in solving the mystery of deformed amphibians. Subsequent experiments, including one we conducted in 2001 on western toads, provided evidence of *Ribeiroia*'s major role in disfiguring amphibians other than Pacific tree frogs. Two studies reported last summer by Joseph M. Kiesecker of Pennsylvania State University and by a team made up of Sessions, Geffrey Stopper of Yale University and their colleagues showed that *Ribeiroia* can cause limb deformities in wood frogs and leopard frogs as well.

Other evidence indicates that *Ribeiroia* is almost always found where deformed amphibians are present, whereas chemical pollutants are found much less frequently. What is more, the parasitic infection seems to have skyrocketed in recent years, possibly reaching epidemic levels. An exhaustive literature search we conducted early in 2001 identified only seven records prior to 1990 of amphibian populations that exhibited both significant malformations and *Ribeiroia* infection.

In contrast, a field study that we published last year turned up 25 such habitats in the western U.S. alone. Among those sites, six species displayed deformities, and the proportion of affected individuals in each population ranged from 5 to 90 percent. Over the past two years, other investigators have identified *Ribeiroia*-triggered deformities in Wisconsin, Illinois, Pennsylvania, New York and Minnesota, including the pond where the eight schoolchildren made headlines. Although heightened surveillance could account for some of this increase in report-

DAVID FIERSTEIN

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Parasitic infection seems
to have skyrocketed in recent years, possibly reaching epidemic levels.

ing, the vast majority of deformed frogs have been found by people, often children, who were looking for frogs for reasons unrelated to monitoring abnormalities.

Not Working Alone

SCIENTISTS NOW UNDERSTAND how the life cycle of *Ribeiroia* helps to perpetuate the development of deformities in generation after generation of amphibians that are unlucky enough to share a habitat with infected snails [*see illustration on opposite page*]. After the parasite leaves its snail host and enters a tadpole, it embeds itself near the tadpole's hind leg. Infected tadpoles then sprout extra legs or fail to develop both limbs. In either case, the young amphibian becomes unable to move properly and thus becomes easy prey for the parasite's final host, often a heron or egret. The parasite matures inside the bird and becomes reproductively active. Through the bird's feces, trematode eggs enter the water. When the larvae hatch, they find a snail and begin the cycle again.

If a spreading epidemic of *Ribeiroia* accounts for much, or even most, of the increase in frog deformities seen in recent years, what accounts for the epidemic? Current environmental trends suggest that human alteration of habitats is at fault. In human as well as wildlife populations, infectious diseases emerge or become more prevalent as features of the landscape change in ways that favor the proliferation of disease-causing organisms. Reforestation of the northeastern U.S., for example, has led to the emergence of Lyme disease by encouraging the proliferation of white-tailed deer, which transport ticks that harbor the Lyme bacterium. On the other side of the Atlantic, the damming of African rivers has led to the spread of human blood flukes that depend on snails as a host and cause human schistosomiasis. During the past several decades, alteration of habitats has also encouraged the expansion of such diseases as hantavirus, Ebola, West Nile virus, dengue fever and AIDS.

DEFORMITIES DIFFER in their most likely causes. Trematode parasites can trigger the growth of extra hind legs (*left*). They can also lead to poorly formed or missing hind legs, although excess ultraviolet radiation, chemical pollutants or injury from predators may also be to blame. Ultraviolet radiation probably accounts more often for abnormalities of the eyes (*right*) and skin.

We recently showed a direct relation between human habitat alteration and sites where *Ribeiroia* parasites are especially abundant. Indeed, our survey of the western U.S., reported in 2002, revealed that 44 of the 59 wetlands in which amphibians were infected by *Ribeiroia* were reservoirs, farm ponds or other artificial bodies of water. Fertilizer runoff and cattle manure near these habitats often encourage overwhelming blooms of algae, which means more food for the snails that host *Ribeiroia* parasites. Larger populations of snails infected with *Ribeiroia* lead directly to more deformed frogs. Wading birds, the other necessary parasite hosts, are usually found in abundance at such human-made locales.

Although parasitism by trematodes is the likeliest explanation for most outbreaks of amphibian deformities, it is certainly not the only cause and may often be abetted by additional factors. At times, water pollutants or excess ultraviolet radiation may act alone to cause specific problems, such as disfigured bodies and eye or skin abnormalities. At other times, pollutants or radiation may set the stage for infection by weakening an amphibian's immune system and thus leaving the animal more vulnerable to a parasitic invasion. In yet another scenario, an increase in amphibian predators, such as fish, leeches or turtles, may create more deformities by biting off tadpole limbs.

Clearly, amphibians are subjected to a cocktail of agents that stress individual animals and then, perhaps, entire populations. The challenge to scientists becomes teasing apart these agents to understand their interactions. Humans and other animals may be affected by the same environmental insults harming amphibians. We should heed their warning. SA

MORE TO EXPLORE

The Effect of Trematode Infection on Amphibian Limb Development and Survivorship. Pieter T. J. Johnson, Kevin B. Lunde, Euan G. Ritchie and Alan E. Launer in *Science*, Vol. 284, pages 802–804; April 30, 1999.

A Plague of Frogs. William Souder. Hyperion Press, 2000.

Ecosystems, Evolution and UV Radiation. Edited by Charles S. Cockell and Andrew R. Blaustein. Springer-Verlag, 2001.

Visit the North American Reporting Center for Amphibian Deformities at **www.npwrc.usgs.gov/narcam/**

STEVEN HOLT *Stockpix.com*

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Satellite-Guided

Highly accurate yet affordable strike weapons, proved in Afghanistan, are the latest upgrades to America's arsenal

HORNET'S STING: A U.S. Navy F/A-18 Hornet fighter-bomber armed with America's most common GPS-aided precision bomb—the JDAM, or Joint Direct Attack Munition—patrols the skies over Afghanistan (*above*). Crews onboard the aircraft carrier USS *John C. Stennis* prepare JDAMs and laser-guided bombs for loading onto Hornets (*far right*).

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

MUNITIONS

By Michael Puttré

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

A U.S. Special Forces team and 1,000 anti-Taliban Afghan fighters

were advancing on the Kandahar airport one cold November evening in 2001 as Air Force Lt. Col. Tom Lawhead peered from the cockpit of his F-16 fighter-bomber. Even with night-vision goggles, the commander of the 389th Expeditionary Fighter Squadron found it difficult to determine whether the coalition force 15,000 feet below was making progress toward its objective—to cut the city's main highway to the Pakistan border. To complicate matters, one of Lawhead's flight leaders had spotted a convoy moving behind a nearby ridge. The anti-Taliban Afghans had already lost radio contact with a scouting party they had sent ahead, so no one knew if the approaching vehicles were the returning scouts or an enemy ambush.

The answer came within moments. As the convoy's vehicles broached the ridgeline, they doused their lights and launched a rocket-propelled grenade. A fierce firefight erupted. Taliban elements were counterattacking.

"The Special Forces's forward air controller was busy trying to figure out where the good guys stopped and the bad guys started," Lawhead recalls. Only 200 to 300 yards separated the opponents. "We waited, watching the shooting from three miles up. Finally, [the forward air controller] was able to give us the coordinates of where he wanted the first bombs to go." The air squadron swooped in and dropped precision munitions that stopped the enemy attack cold. "Had that ambush succeeded," the squadron leader explains, "it could have forestalled our plans to take over southern Afghanistan."

The American strike fighters could hit an important target in the dark only a few hundred meters from friendly forces because of new, highly accurate weapons technology. Rather than "dumb," or iron, bombs, the airplanes were carrying the latest in low-cost "smart" bombs. The forward air controller on the scene had radioed a set of Global Positioning System (GPS) coordinates to the patrolling aircraft. The F-16 pilots entered the number sequences into their fire-control computers, which, in turn, downloaded them to microcomputers in the weapons. When each bomb was released, the onboard inertial navigation system (INS) used the GPS positioning information to steer it to an area inside a radius of 85 feet of the specified map point half of the time.

Afghanistan marked the first use of GPS-aided weapons to support ground services. Although the U.S. deployed limited numbers of these arms during the Kosovo intervention of 1999 and in subsequent strikes against Iraqi air defense sites in the northern and southern no-fly zones, the vast majority of bombs dropped in the Afghan conflict were satellite-guided. Pentagon planners expect to use many more of these smart weapons in the intense assaults that would mark any attack of Iraq.

Unlike the previous generation of laser- and television-guided munitions, a GPS-aided bomb does not require that the launch aircraft remain in the vicinity to illuminate the target for guidance. Once released, this true "fire and forget" weapon operates in an autonomous, self-correcting fashion, which makes it accurate even in bad weather. Further, current GPS-aided arms are designed to work with the latest infrared- and radar-targeting systems on American strike aircraft. They also link directly to the nation's world-spanning navigation and communications infrastructure, which connects control bases with orbital spies such as the KH-11 photo and Lacrosse radar satellites and an array of aerial battlefield sensor platforms—all of which can help ascertain the GPS coordinates of a potential target. Among these intelligence systems are the ever useful U-2 spy planes, the far-seeing E-8A J-STARS and RC-135 Rivet Joint reconnaissance aircraft, the sturdy and reliable U.S. Navy EP-3 Aries and Army RC-12 patrol planes, the familiar E-3 AWACS planes, and the newly capable Predator and Global Hawk surveillance (and increasingly attack) drones [*see illustration on opposite page*].

With such weapons, fewer sorties need to be flown to meet mission goals, easing wear and tear on equipment and personnel. Aircraft can in addition attack targets from beyond the reach of air defense systems, saving lives and reducing the chances of crew capture. Accidental civilian deaths and damage to property and infrastructure can be avoided as well. The major weakness of this technology is its susceptibility to sophisticated signal jamming efforts.

Overview/*GPS-Aided Weapons*

- The Pentagon has acquired a family of "smart" air-to-ground weapons that use inertial guidance technology updated by Global Positioning System (GPS) satellite coordinates to target enemy troops and emplacements accurately from afar. GPS-aided systems help to guide the most common of these precision munitions to within a radius of 40 feet of a designated aim point 50 percent of the time.
- The relative low cost of many of these smart armaments allows the U.S. to use them in mass quantities against opposing forces. Thus, aircraft and crews can attack hazardous targets from a safe distance; fewer strike sorties are necessary to achieve mission objectives; and inadvertent—so-called collateral—damage to civilian lives and property can be greatly reduced (compared with the use of conventional "dumb," or iron, bombs).

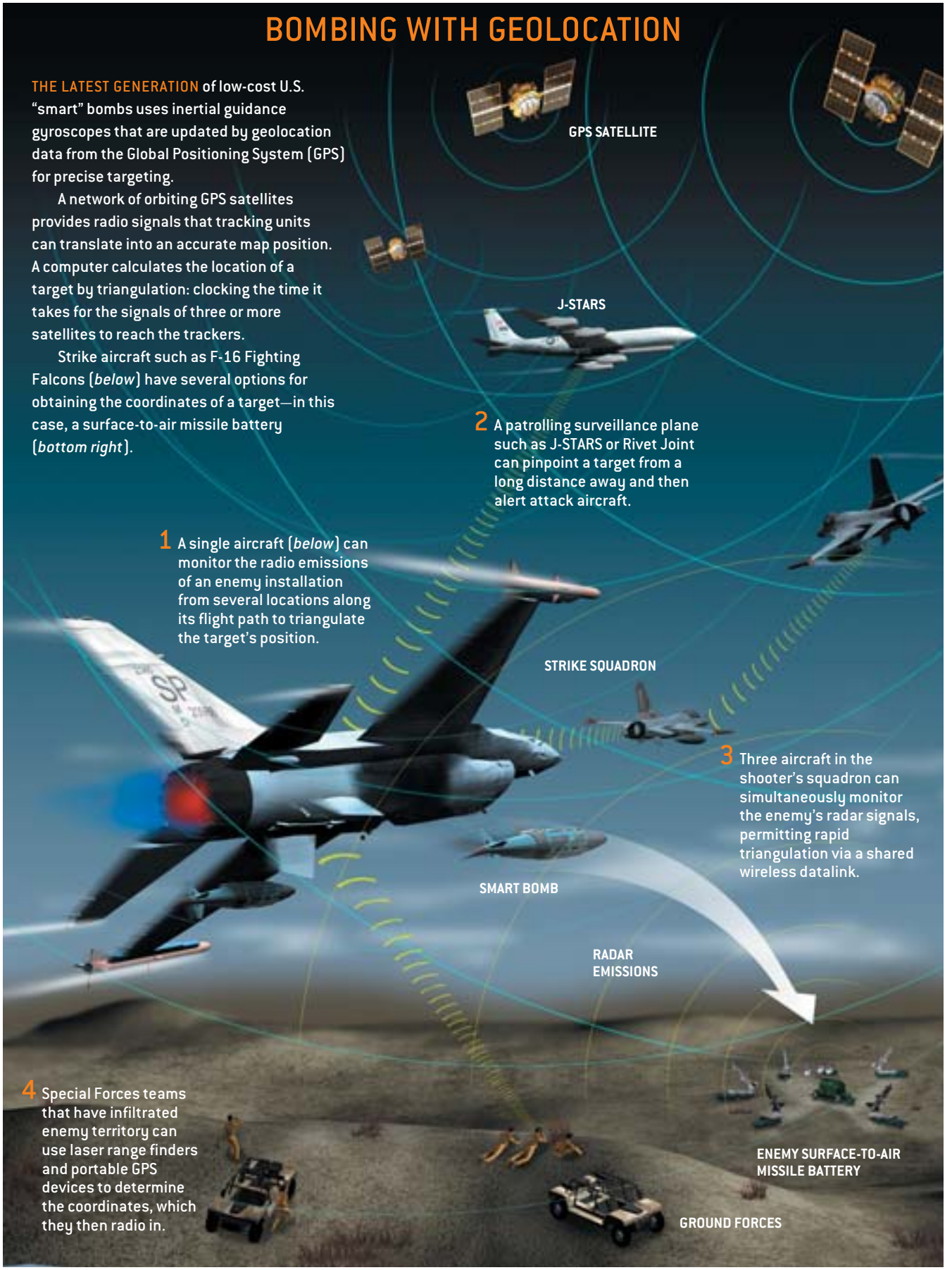
A Decade's Difference

GPS-AIDED WEAPONS provide a flexibility in combat that was almost inconceivable just 10 years ago. "In Desert Storm, we generally knew ahead of time what the aim point [target] of the mission was going to be," Lawhead notes. Unfortunately, that knowledge could not be readily updated. Nowadays mission leaders can respond faster to changing conditions, which permits much more flexible

U.S. NAVY (*preceding pages*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

BOMBING WITH GEOLOCATION



THE LATEST GENERATION of low-cost U.S. "smart" bombs uses inertial guidance gyroscopes that are updated by geolocation data from the Global Positioning System (GPS) for precise targeting.

A network of orbiting GPS satellites provides radio signals that tracking units can translate into an accurate map position. A computer calculates the location of a target by triangulation: clocking the time it takes for the signals of three or more satellites to reach the trackers.

Strike aircraft such as F-16 Fighting Falcons (*below*) have several options for obtaining the coordinates of a target—in this case, a surface-to-air missile battery (*bottom right*).

1 A single aircraft (*below*) can monitor the radio emissions of an enemy installation from several locations along its flight path to triangulate the target's position.

2 A patrolling surveillance plane such as J-STARS or Rivet Joint can pinpoint a target from a long distance away and then alert attack aircraft.

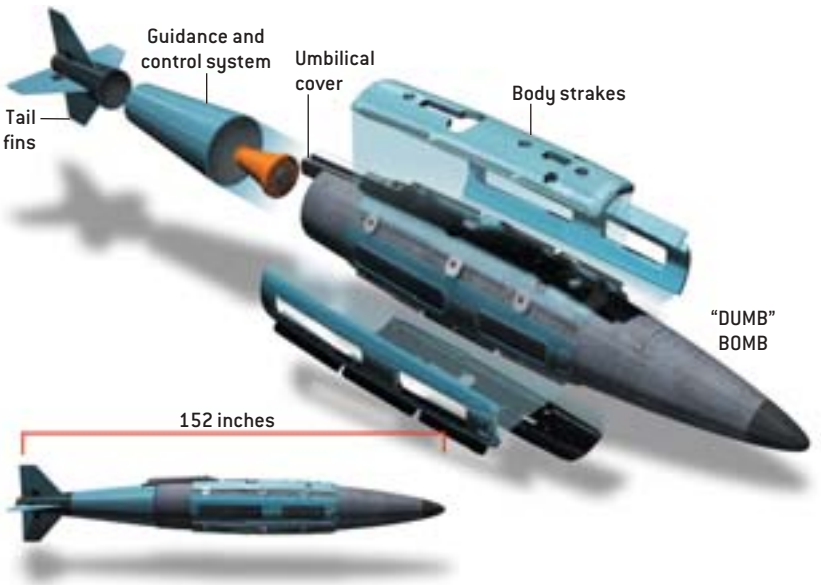
3 Three aircraft in the shooter's squadron can simultaneously monitor the enemy's radar signals, permitting rapid triangulation via a shared wireless datalink.

4 Special Forces teams that have infiltrated enemy territory can use laser range finders and portable GPS devices to determine the coordinates, which they then radio in.

DON FOLEY

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

GPS-AIDED BOMB



INTELLIGENT ADD-ON: The Joint Direct Attack Munition, or JDAM (*above*), is the most common of the Pentagon's GPS-aided aerial weapons. JDAM comes as a $20,000 kit that technicians fit to "dumb" bombs costing several thousand dollars, converting them to "smart" bombs capable of pinpoint accuracy. The conversion kit comprises tail fins and body strakes, a GPS receiver and its antenna, an inertial guidance system, a mission computer (with specialized software), electric actuators to move the tail fins, and a power supply, as well as a cabling harness and connectors. Navy bomb crews on the aircraft carrier USS *John C. Stennis* load a JDAM onto an aircraft on the flight deck (*top right*). The Air Force's B-52 bomber can fit multiple JDAMs on wing pylon racks (*bottom right*).

use of available forces. "Of the 178 sorties the 389th Squadron flew into Afghanistan, there was only one time we actually attacked the target that we had planned to hit ahead of time," he says.

Better efficiency has other strategic implications. The conventional wisdom before the 1990–91 Persian Gulf War was that six iron bombs—about the entire load of a typical strike fighter—were required to destroy a particular target. "In Desert Storm, when I was dropping dumb bombs, we would run a package of 20 to 24 F-16s to get a single target with several impact points," Lawhead says. (A target air base might have many hangars that require individual aim points.) But with precision weapons aimed at the same target, "rather than 24 jets, now you're talking two to four jets."

The technology also takes relatively inexpensive iron bombs—the most common item in the Pentagon's aerial arsenal—and makes them "smart." That means they can be used on critical targets that had previously been the sole province of expensive, low-inventory cruise and antiship missiles such as the Tomahawk and the Harpoon. To date, GPS weapons account for approximately nine tenths of all the tonnage dropped on Afghanistan. Only about 6 percent of the ordnance launched during the Persian Gulf War was precision-guided—and those were predominantly laser-directed bombs, television-aimed missiles and inertially guided cruise missiles, all of which were then state-of-the-art. At that time, GPS was just becoming available for navigation purposes, and it had no weapons-guidance role.

"We are moving away from dumb bombs as much as possible," says Capt. Robert Wirt, program manager for conventional strike weapons at the U.S. Naval Air Systems Command (NAVAIR) at the Patuxent River Naval Air Station in Maryland. "Given concerns over collateral damage, the need for close-air support, and the tight quarters in which we operate, there is good reason to want the lion's share of our weapons to be precision-guided."

Production is increasing accordingly. Before the Afghanistan campaign, the Pentagon's inventory contained approximately 10,000 JDAMs, the predominant type of GPS-aided bomb. Boeing is now building about 1,500 of the systems a month at its St. Charles, Mo., plant to fill a combined U.S. Air Force and Navy order for 230,000 JDAMs; that rate will rise to 2,800 a month by August.

Beating the Weather

THE PENTAGON developed GPS munitions to overcome the limitations of previous-generation precision weapons, mainly Vietnam-era laser-guided bombs, which are still in use. These earlier systems are often equipped with seekers that home in on laser energy reflected off the target. The laser designators can be mounted on the attacking aircraft or another plane or carried by an observer team on the ground. Such weapons remain highly valued because of their accu-

U.S. NAVY AND U.S. AIR FORCE (*photographs*); DON FOLEY (*illustrations*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

racy. Still, heavy cloud cover, smoke, dust and other concealments can make them ineffective. Analysis of the air war over Iraq and Kuwait showed that many strike missions were scrubbed because pilots could not see their targets. "The one variable we can't control is weather," NAVAIR's Capt. Wirt says. "All it takes is a cloud getting between the path of the designator and the target, and a normal laser-guided weapon would lose track."

This shortcoming prompted then Air Force Chief of Staff Lt. Gen. Merrill McPeak to create a Pentagon requirement on May 1, 1991, for an autonomous weapon that pilots could drop through clouds and other obscuring conditions. That led to a call for proposals for GPS-aided munitions that would serve both the U.S. Air Force and the Navy. The winning idea had to be cheap, accurate and compatible with existing weapons, aircraft platforms, and infrastructure. Because the technical tasks were relatively straightforward, cost was the key issue. Boeing won a competition among 12 contractors in 1995. The result was the Joint Direct Attack Munition, or JDAM, and its kin. Each type of these so-called J-weapons has a specific range. "How each is used will be driven by the threat," Capt. Wirt says. "How far do I have to stand back from the target to shoot and not go into harm's way?" (The designation "joint" means that these systems can be deployed by aircraft from the Navy and the Air Force.)

These precision weapons achieve their unparalleled accuracy by using complementary targeting systems to correct the flight path charted by the munition's onboard autopilot. A GPS receiver triangulates a bomb's position using navigation signals beamed from orbiting satellites. The GPS technology combines with an inertial measurement unit (IMU), which plots a route using gimbal-mounted gyroscopic acceleration sensors. A GPS-IMU-guided bomb spends the first 25 to 30 seconds of its free fall acquiring the GPS signals. These coordinates then update the IMU, which provides steering data to the weapon's autopilot. This mission computer then modifies the flight path as required by deflecting electrically actuated tail fins. If the GPS signals are lost because of receiver failure or jamming, the inertial system will continue to direct the weapon toward the last updated coordinates.

The JDAM was the first of the GPS-aided air-to-ground arsenal to enter service, in 1998 [*see illustration on opposite page*]. It is not a complete weapon but a strap-on tail-kit assembly that converts standard-issue iron warheads (which cost up to a few thousand dollars each) into GPS smart bombs. The JDAM precision-guidance modules and steering fins are manufactured at relatively low cost, about $20,000 per weapon (compared with $1 million for a single cruise missile of the type used during the Gulf War). The kit includes a GPS receiver and antenna, an IMU, a mission computer, mechanical actuators for moving the tail fins, a power supply, a cabling harness, and connectors. The system employs mission-planning software and complementary software for the aircraft.

The JDAMs in U.S. service have a range of about eight miles when dropped from an altitude of 20,000 feet. Arms maker MBDA is developing a glider attachment for JDAMs called Diamond Back that will increase the range to about 24 miles.

JDAM's deadly cousin is the Wind Corrected Munitions Dispenser (WCMD), a $10,000 tail-kit assembly introduced in 2000 by Lockheed Martin that fits the existing family of standard cluster bombs [*see box on next two pages*]. The WCMD is not equipped with a GPS receiver to update its IMU continuously, so it has about half the range of a JDAM.

A more complex member of the J-weapons series is the Joint Standoff Weapon (JSOW) [*see box on next two pages*]. Built by Raytheon, JSOW is a glider bomb with a 500-pound warhead and wings that deploy en route to its destination. It features a maximum range of 15 to 40 miles, depending on the altitude of its release. JSOW is a totally new system, as opposed to a conversion kit. As the "standoff" in the name of the series suggests, each bomb can be released against targets at distances beyond the reach of short-range air defense systems. The intent was to produce an all-purpose carrier, a "bomb truck" with several variants. JSOW-A, the first one, entered service in 1999. It carries submunitions similar to WCMD's—and a price tag of $220,000. JSOW-B, which costs $375,000 and will become available this year, bears anti-tank submunitions.

Scheduled for 2005, the JSOW-C has a "unitary" penetrator warhead for destroying bunkers and other reinforced targets; its estimated cost will be $400,000. Its use of GPS and IMU for initial guidance is similar to that of JDAM, but the JSOW-C also features an imaging infrared seeker for terminal (last-second) guidance to the target. The seeker enables the onboard computer to perform a func-

AREA INTERDICTION: Joint Standoff Weapons, or JSOWs, often contain lethal bomblets that spread widely across a target area to destroy nearly everything—troops, vehicles, unreinforced buildings and emplacements—in their collective path.

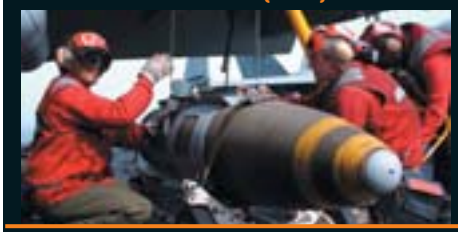
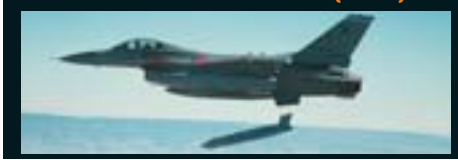
RAYTHEON

THE AUTHOR

MICHAEL PUTTRÉ is editor in chief of *JED, The Journal of Electronic Defense,* a monthly magazine that focuses on electronic warfare, battle command and control, reconnaissance, and precision-guided weapons. Puttré has been a military and technology writer for 15 years.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

GPS-AIDED WEAPONS ARSENAL

NAME	MANUFACTURER	DESIGNATION	MISSION	PAYLOAD	TOTAL WEIGHT (pounds)
Joint Direct Attack Munition (JDAM) 	Boeing	GBU-31(v)1	General purpose	Mk 84 bomb	2,000
		GBU-31(v)3B	Hardened targets	BLU-109 penetrator bomb	2,000
		GBU-32	General purpose	Mk 83 bomb	1,000
		GBU-38	General purpose	Mk 82 bomb	500
Wind Corrected Munitions Dispenser (WCMD) 	Lockheed Martin	CBU-103	Area soft targets	CBU-87 dispenser (202 BLU-97 combined-effects bomblets)	1,000
		CBU-104	Area denial	CBU-89 Gator dispenser (72 BLU-91 antitank mines and 22 BLU-92 antipersonnel mines)	1,000
		CBU-105	Antitank and vehicle	CBU-97 dispenser (10 BLU-108 infrared-detonated submunitions)	1,000
Joint Standoff Weapon (JSOW) 	Raytheon	AGM-154A (JSOW-A)	Area soft targets	145 BLU-97 combined-effects bomblets	1,065
		AGM-154B (JSOW-B)	Antitank and vehicle	6 BLU-108 IR-detonated submunitions	1,065
		AGM-154C (JSOW-C)	Hardened targets	500-pound-class unitary warhead	1,500
Joint Air-to-Surface Standoff Missile (JASSM) 	Lockheed Martin	AGM-158	High-value hardened targets	1,000-pound-class unitary warhead	2,250

tion called target-image or scene matching, in which an infrared image is compared with a stored shot of the target, acquired previously by other reconnaissance assets—aircraft, satellites or drones.

The longest-range and most sophisticated of the initial batch of J-weapons is the Joint Air-to-Surface Standoff Missile (JASSM), made by Lockheed Martin [*see box above*]. JASSM completed tests in late 2001 and is expected to enter service this year. A jet engine powers the cruise missile to a range of about 200 miles. The 1,000-pound penetrator warhead is designed to engage high-value, well-defended targets. Like JSOW-C, the estimated $700,000-plus JASSM has an imaging infrared terminal seeker in addition to its GPS-IMU midcourse guidance.

Passive Aggression

GPS-AIDED PRECISION weapons can also take advantage of transmissions made by the enemy. Almost every modern military unit or system emits signals of one kind or another in the course of its mission, and these can be intercepted and analyzed to determine its location and identity.

Lately, such direction-finding techniques have made it possible to locate a threat emitter on the surface of the earth within a radius tight enough for a GPS-aided weapon to be sent to those coordinates to destroy it. This geolocation capability supports a U.S. doctrine to replace active radar on attacking aircraft with newer, "passive" sensors, which do not emit telltale signals. Passive radar warning receiver systems on tactical aircraft measure Doppler shifts in an enemy radar resulting from the host aircraft's own motion. A special ranging algorithm compares the Doppler shift with the host aircraft's GPS and INS to derive precise direction and location readings. Also, three or more aircraft can triangulate on the emitter and share information via a wireless datalink at the same time.

The next-generation U.S. tactical aircraft, the F-22 Raptor air-superiority fighter and F-35 Joint Strike Fighter, will have greater capabilities than current aircraft because they are designed to incorporate passive sensors. They could acquire targets by geolocation at ranges similar to those achieved today only by specialized electronic-intelligence-gathering aircraft. The F-22 is expected to become operational in early 2006, with the F-35 following within a few years after that.

Jamming vs. Antijamming

THE SAME GPS SIGNALS that are so pivotal for precision bombing, however, are also their principal weak point. Because the signals are faint, they are highly susceptible to jamming. They are transmitted to ensure a minimum signal power level of between –160 and –166 dBW (160 to 166 decibels below one watt) at the earth's surface. This is comparable to the light from a 25-watt lightbulb as seen from 10,000 miles away. Put another way, GPS signals have one billionth the power of those received by a television set's antenna. Interfering with such weak signals is easy, so weapons makers incorporate one or more antijamming features into the receiver design. For example, GPS receivers with multiple channels can

U.S. NAVY (*top*); LOCKHEED MARTIN (*top middle and bottom*); RAYTHEON (*bottom middle*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

LENGTH (inches)	RANGE (miles)	ACCURACY (feet)	ESTIMATED UNIT COST	ENTERED SERVICE	NOTES
153	15	40	$20,000 (kit) + $3,000 (bomb)	1998	All JDAM tail-kit assemblies include GPS receiver
149	15	40	$20,000 (kit) + $5,000 (bomb)	1998	
120	15	40	$20,000 (kit) + $2,000 (bomb)	1999	GBU-35 version has a BLU-110 warhead payload stabilized for aircraft carriers
86	15	40	$20,000 (kit) + $1,000 (bomb)	2004	For B-2 Spirit bombers, which will carry up to 80 bombs
92	7	85*	$10,000 (kit) + $14,000 (dispenser)	1999	All WCMD tail-kit assemblies download GPS from aircraft
92	7	85*	$10,000 (kit) + $40,000 (dispenser)	1999	
92	7	85*	$10,000 (kit) + $300,000 (dispenser)	1999	
160	40+	50*	$220,000	1999	All JSOWs are unpowered glide bombs with low-altitude ranges of 15+ miles
160	40+	50*	$375,000	2003	Program has been scaled back significantly
160	40+	10	$400,000	2004	Has an imaging infrared terminal (last-second) seeker; turbojet-powered version with 120-mile range is in development
168	200+	10	$700,000	2003	Has an imaging infrared terminal seeker, a turbojet engine and enhanced GPS-jamming resistance technology

*Area-effects munition that strikes across a wide target

acquire signals from some satellites even if others are jammed from a given direction (generally each channel receives signals from one satellite). Another method is to cancel out, or null, incoming jamming signals using special antennas or software signal processing.

Precision systems are also not immune to human error. In October 2001 the Pentagon reported that a U.S. Navy F/A-18 Hornet fighter-bomber missed its intended target with a GPS-aided bomb. The aircraft dropped a JDAM into a residential area a mile from the Kabul airport target. At least four people died, and others were injured. An investigation attributed the cause to "targeting process error." In another incident, in December of that year, three Americans and five Allied Afghan soldiers were killed when the battery for a forward air controller's receiver was changed during a bombing run. The power interruption reset the calculated target coordinates to those of the controller himself, leading a misdirected JDAM to land on his position. These mistakes would seem to result mainly from operator inexperience, so better training and more practice in the field should help eliminate similar errors. Whatever the case, the proportion of misdirected GPS-aided munitions is low.

The mountains of eastern Afghanistan were an extreme proving ground for GPS-aided ordnance, demonstrating the capabilities—and the limitations—of the technology. Many times the enemy simply waited out sustained bombardment, no matter how accurate, in well-protected caves. This technology has also yet to pass the test that arises between jamming and antijamming contingents when both adversaries are technologically advanced. In the main, however, the widespread availability of GPS-IMU weapons is increasing the confidence of Pentagon planners that smaller groups of soldiers bearing lighter equipment can achieve victory over more numerous enemy troops with the aid of devastatingly precise firepower from the sky. SA

A broadcast version of this article will air January 21 on *National Geographic Today,* a program on the National Geographic Channel. Please check your local listings.

MORE TO EXPLORE

Air Warfare in the Missile Age. Lon O. Nordeen. Smithsonian Institution Press, 2002.

Code One. Issue on F-16 operations over Afghanistan. Vol. 17, No. 3; July 2002. Available at **www.codeonemagazine.com/archives/2002/jul_02.html**

Additional information can be obtained from *JED, The Journal of Electronic Defense* (**www.jedonline.com**), the U.S. Navy NAVAIR Integrated Battlespace Arena (**ibar.nawcwd.navy.mil**), the U.S. Air Force Combat Command (**www2.acc.af.mil**) and the GPS Joint Program Office (**gps.losangeles.af.mil**).

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Drink to Your Health?

Three decades of research shows that drinking small to moderate amounts of alcohol has cardiovascular benefits. A thorny issue for physicians is whether to recommend drinking to some patients

By Arthur L. Klatsky

Photographs by Tina West

Addressing an Illinois temperance society in 1842, Abraham Lincoln said something about "intoxicating liquor" that probably got a frosty reception. "It is true that … many were greatly injured by it," the future president noted. "But none seemed to think the injury arose from the use of a bad thing but from the abuse of a very good thing."

America has always had trouble deciding whether alcohol is a bad thing or a good thing. Millions who remember Prohibition, when all alcoholic beverages were illegal, now witness a constant stream of advertisements from producers of alcoholic beverages encouraging people to drink. Despite alcohol's popularity today, however, many still consider abstinence a virtue. Certainly, heavy drinking and alcoholism deserve deep concern for the terrible toll they take on alcohol abusers and society in general. But worry about the dangers of abuse often leads to emotional denials that alcohol could have any medical benefits. Such denials ignore a growing body of evidence indicating that moderate alcohol intake wards off certain cardiovascular (circulatory system) conditions, most notably heart attacks and ischemic strokes (those caused by blocked blood vessels). A few studies even show protection against dementia, which can be related to cardiovascular problems.

The Alcohol Effect

A DISCUSSION OF moderate drinking requires a working definition of "moderate." Simple definitions of light, moderate or heavy are somewhat arbitrary, but a consensus in the medical

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The large ARTERIES OF PEOPLE who died of alcoholic liver cirrhosis were remarkably free of atherosclerosis.

literature puts the upper limit for moderate drinking at two standard-size drinks a day [*see illustration on opposite page*]. Studies show that drinking above that level can be harmful to overall health, although sex, age and other factors lower and raise the boundary for individuals.

The main medical benefit of reasonable alcohol use seems to be a lowering of the risk for coronary heart disease (CHD), which results from the buildup of atherosclerosis (fatty plaque) in the arteries that feed blood to the heart. (The word "atherosclerosis" is in fact a descriptive union of two Greek words: *athera,* for "gruel" or "porridge," referring to the fatty deposits, and *sclera,* for "hard," pertaining to the loss of vessel flexibility.)

Atherosclerosis restricts blood flow to the heart and can promote the formation of vessel-blocking clots. It can thereby cause angina (chest discomfort resulting from low oxygen levels in the heart muscles), heart attack (the death of heart tissue that occurs when a blood clot or narrowing of the arteries prevents blood from reaching the heart) and death, often without warning. The condition usually starts at a young age but takes decades to blossom into overt CHD. The most common form of heart disease in developed countries, CHD causes about 60 percent of deaths from cardiovascular ills and about 25 percent of *all* deaths in those nations.

Pathologists uncovered the first clues to the value of alcohol in the early 1900s, noting that the large arteries of people who died of alcoholic liver cirrhosis seemed remarkably "clean"—that is, free of atherosclerosis. One explanatory hypothesis assumed that alcohol was a nebulous solvent, essentially dissolving the buildup in the arteries; another explanation held that heavier drinkers died before their atherosclerosis had a chance to develop. Neither idea truly explained drinkers' unblocked arteries, however.

A more telling hint emerged in the late 1960s, when Gary D. Friedman of the Kaiser Permanente Medical Center in Oakland, Calif., came up with a novel idea: use computers to unearth unknown predictors of heart attacks. The power of computing could first identify healthy people who had risk factors similar to heart attack victims. Such factors include cigarette smoking, high blood pressure, diabetes, elevated levels of low-density-lipoprotein (LDL, or "bad") cholesterol, low levels of high-density-lipoprotein (HDL, or "good") cholesterol, male gender, and a family history of CHD. Friedman then searched for predictors of heart attacks by comparing the patients and the newly found controls in hundreds of ways—for example, their exercise and dietary habits and their respective levels of various blood compounds. The computers spit out a surprising discovery: abstinence from alcohol was associated with a higher risk of heart attack.

Various studies had missed the connection because they neglected to examine alcohol use as a behavior separate from smoking. We now know that because drinkers often also use cigarettes, the negative impact of smoking was masking the beneficial effect of alcohol. In 1974 my Kaiser Permanente colleagues Friedman and Abraham B. Siegelaub and I were the first, to our knowledge, to publish an examination of moderate drinking in the absence of smoking. We saw a clear connection between alcohol consumption and a decreased risk of heart attack.

Since then, dozens of investigations in men and women of several racial groups in various countries have correlated previous alcohol use with current health. These studies have firmly established that nondrinkers develop both fatal and nonfatal CHD more often than do light to moderate drinkers. In addition, in 2000 Giovanni Corrao of the University of Milan-Bicocca in Italy, Kari Poikolainen of the Järvenpää Addiction Hospital in Finland and their colleagues combined the results of 28 previously published investigations on the relation between alcohol intake and CHD. In this meta-analysis, they found that the risk of developing CHD went down as the amount of alcohol consumed daily went up from zero to 25 grams. At 25 grams—the amount of alcohol in about two standard drinks—an individual's risk of a major CHD event, either heart attack or death—was 20 percent lower than it was for someone who did not drink at all.

New data about alcohol protecting against death from CHD are even more impressive. At a meeting of the American

Overview/*Alcohol and Heart Health*

- An assortment of studies from around the world indicates that drinking in small to moderate amounts decreases the risk of dying from coronary heart disease by almost one third.
- Some research points to red wine as being particularly protective against coronary heart disease. Other healthful habits of red wine drinkers, however, may be partly responsible for the apparent effect.
- A select group of people—those with CHD or at risk for CHD and without risks associated with alcohol itself—may wish to consult their physicians about moderate drinking as part of a heart-healthy diet.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Heart Association last November, my Kaiser Permanente colleagues Friedman, Mary Anne Armstrong and Harald Kipp and I discussed an updated analysis of 128,934 patients who had checkups between 1978 and 1985, with 16,539 of them dying between 1978 and 1998. CHD was responsible for 3,001 of those deaths. We discovered that those who had one or two alcoholic drinks a day had a 32 percent lower risk of dying from CHD than abstainers did.

The possible mechanisms by which alcohol has such an apparently profound effect on cardiovascular health primarily involve cholesterol levels and blood clotting. Blood lipids, or fats, play a central role in CHD. Numerous studies show that moderate drinkers have 10 to 20 percent higher levels of heart-protecting HDL cholesterol. And people with higher HDL levels, also known to be increased by exercise and some medications, have a lower risk of CHD.

That lower risk stems from HDL's ability to usher LDL cholesterol back to the liver for recycling or elimination, among other effects. Less cholesterol then builds up in the walls of blood vessels, and so less atherosclerotic plaque forms. Alcohol seems to have a greater influence on a different HDL subspecies (HDL_3) than on the type increased by exercise (HDL_2), although both types are protective. (The biochemical pathways in the liver that could account for alcohol's ability to raise HDL levels remain incompletely known; it is thought that alcohol probably affects liver enzymes involved in the production of HDL.) Three separate analyses aimed at determining specific contributions of alcohol all suggest that the higher HDL levels of drinkers are responsible for about half of the lowered CHD risk.

Alcohol may also disrupt the complex biochemical cascade behind blood clotting, which can cause heart attacks when it occurs inappropriately, such as over atherosclerotic regions in coronary arteries. Blood platelets, cellular components of clots, may become less "sticky" in the presence of alcohol and therefore less prone to clumping, although data on this question remain ambiguous. A 1984 study by Raffaele Landolfi and Manfred Steiner of Brown University's Memorial Hospital revealed that alcohol intake increases the level of prostacyclin, which interferes with clotting, relative to the level of thromboxane, which promotes clotting. Walter E. Laug of the University of Southern California Keck School of Medicine showed that alcohol raises levels of plasminogen activator, a clot-dissolving enzyme. Finally, several studies suggest that alcohol lowers levels of another promoter of blood clots, fibrinogen.

Overall, alcohol's anticlotting capacity is not as well established as its HDL effect, and some effects, such as platelet clumping, may be reversed by heavy or binge drinking. Nevertheless, anticlotting appears to have a role in the lower risk for heart attacks enjoyed by moderate drinkers. In addition, studies have shown a beneficial effect on CHD risk in people who have far fewer than two drinks a day—say, three or four drinks a week. Anticlotting could be a major factor in the protection accorded by alcohol in these small amounts, which seem insufficient to affect HDL levels greatly.

Although alcohol reduces heart disease risk mainly by raising HDL levels

"STANDARD" SERVINGS OF ALCOHOLIC BEVERAGES

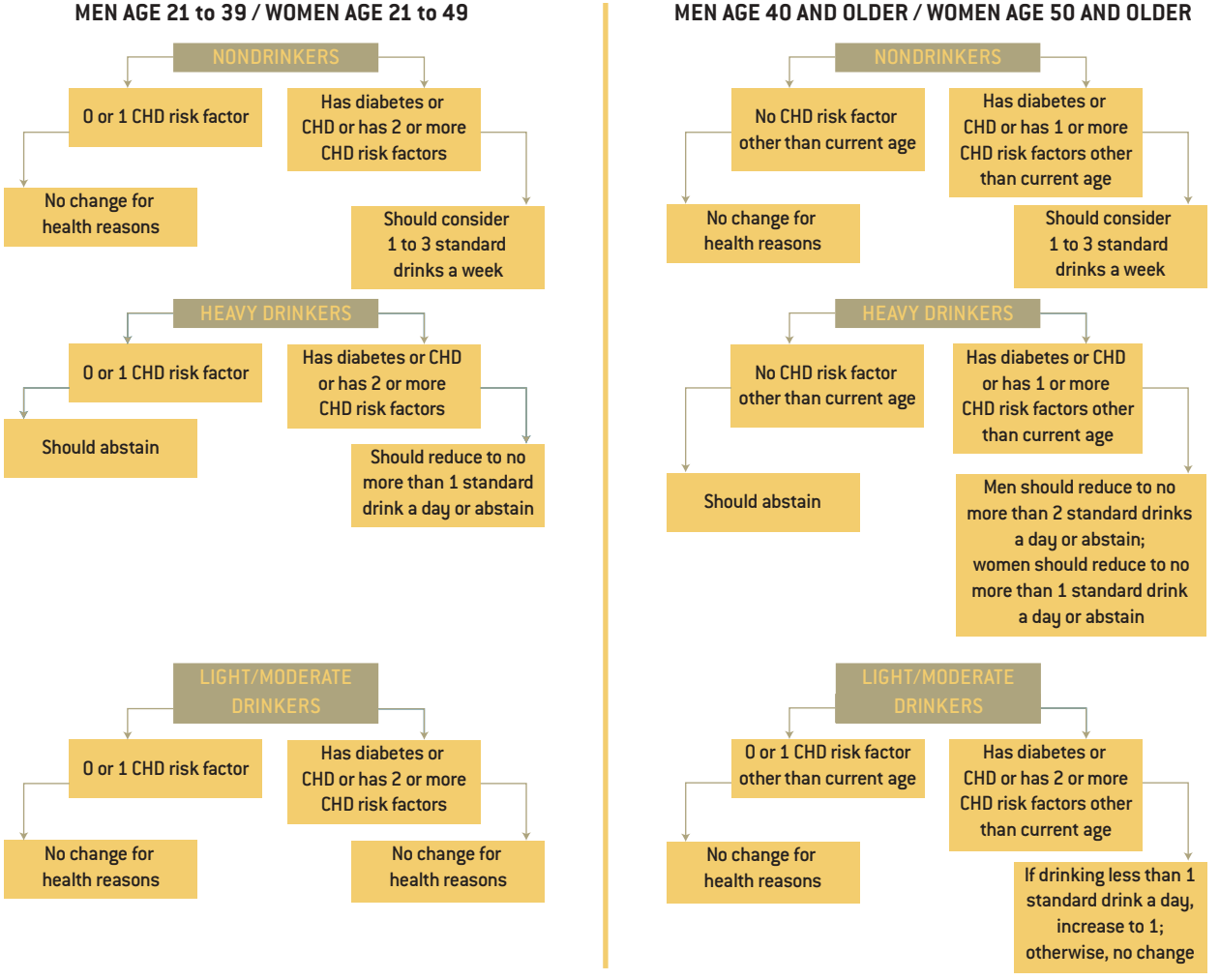
ALTHOUGH THERE IS NO formal definition of a standard-size drink, something of a consensus does exist. Beer is often sold in a 12-ounce bottle or can, which is a useful reference point as one standard drink. The amount of alcohol, about 0.6 ounce, in 12 ounces of beer is virtually the same as is found in a 5-ounce glass of wine or a 1.5-ounce glass of distilled spirits, such as vodka, gin, bourbon or scotch. Wine and distilled spirits in these amounts are thus also considered standard drinks.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

MAKING THE DRINKING DECISION

Roger R. Ecker, a cardiovascular surgeon at Summit Medical Center in Oakland, Calif., and I developed these charts to help individuals determine whether to include alcoholic beverages, and in what amounts, in their diets. The charts are designed to be used by physicians in consultation with patients. Coronary heart disease (CHD) risk factors are listed at the bottom. "Light/Moderate" is defined as up to one standard drink a day for women and up to two standard drinks a day for men. "Heavy" is three or more drinks a day for men and two or more drinks a day for women.

These charts do *not* apply to the following people, who should abstain from alcoholic beverages: anyone under the age of 21; pregnant women; nondrinkers with a family history of alcoholism, with moral or religious beliefs that preclude alcohol, with a personal history of alcohol abuse, with known organ damage from alcohol, with any chronic liver disease, or with a genetic risk of breast or ovarian cancer. —*A.L.K.*



Coronary heart disease (CHD) risk factors, according to National Cholesterol Education Program guidelines:

1. Family history of CHD (father or brother younger than 55 with CHD, mother or sister younger than 65 with CHD)
2. Smoking
3. High blood pressure
4. Total cholesterol higher than 200
5. HDL cholesterol lower than 35 (if HDL is higher than 60, subtract one risk factor)
6. Age 40 and older for men, 50 and older for women

and reducing clotting, it acts in other ways that could lower the risk more subtly. Moderate drinking may lessen CHD risk indirectly by decreasing the risk of type 2 (adult-onset) diabetes, which is a powerful predictor of CHD. This benefit appears to be related to enhanced insulin sensitivity, which promotes proper glucose usage. (Heavy drinking, however, has been connected to higher blood glucose levels, a marker for future diabetes.) Evidence is also growing that inflammation contributes to CHD, and alcohol's anti-CHD power may be related to an anti-inflammatory action on the endothelial tissue that lines blood vessels.

Before accepting alcohol's benefits, an epidemiologist attempts to locate hidden

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The collected data make a strong case for the CARDIAC BENEFITS of controlled drinking.

factors possibly at work. For instance, could lifelong abstainers differ from drinkers in psychological traits, dietary habits, physical exercise habits or other ways that might account for their higher CHD risk without the need to invoke the absence of alcohol? Were such traits to explain away alcohol's apparent protection, they would need to be present in both sexes, various countries and several racial groups. Considering that no such traits have been identified, the simpler and more plausible explanation is that light to moderate alcohol drinking does indeed enhance cardiovascular health.

In fact, the available evidence satisfies most standard epidemiological criteria for establishing a causal relation. The numerous studies examining light and moderate alcohol intake and health reach consistent conclusions. The prospective studies that exist have the correct temporal sequence—that is, individuals' habits of interest are identified, after which their health is monitored over the long term, and alcohol users have different health profiles than nondrinkers do. The positives associated with alcohol can be attributed to biologically plausible mechanisms. Alcohol offers specific enhancement of cardiovascular health, not general protection against all illness. And alcohol's effect can be identified independent of known "confounders," other alcohol-related factors that could be responsible for a subject's cardiovascular condition.

The 30 percent reduction in risk is, perhaps surprisingly to some, less convincing evidence than the arguments above, because a strong *unknown* confounder could still account for the connection. To take an extreme example, consider a hypothetical set of genes that confers on the possessor 60 percent less CHD risk *and* causes a strong predisposition toward liking moderate amounts of alcohol. The independent consequences of the genes could appear causally linked. (In fact, however, no such confounder is known or likely, and the 30 percent risk reduction appears to be a probable measure of alcohol's beneficial effect.)

Because heavy drinking is not more protective than lighter drinking, this absence of a clear dose-response relation is also a weakness. Nevertheless, the collected data make a strong case for the cardiac benefits of controlled drinking. I should note, however, that the kind of study considered to be the gold standard in human research—a prospective randomized blinded clinical trial—has not yet been done. Such a study might, for example, engage a large pool of nondrinkers, half of whom, chosen at random and without the knowledge of the researchers, would commence a moderate drinking regimen, while the other half remained abstainers. The two groups would be followed for years in a search for eventual differences in cardiovascular disease and heart-related deaths.

To Drink or Not to Drink

MOST PEOPLE DRINK for reasons other than alcohol's health benefits, and many of them are already using alcohol in amounts that appear to promote cardiovascular health. But the accumulated research on alcohol's positive effects pre-

HOW ALCOHOL MIGHT PROTECT AGAINST CHD

Alcohol Effect	*Probable Action*	*Evidence*
Raises blood HDL cholesterol	Removes and transports LDL cholesterol from vessel wall	Solid supporting evidence; effect explains at least half of alcohol's benefit
Lowers blood LDL cholesterol	Reduces level of one major CHD risk factor	Evidence weak; effect probably not independent of diet
Lowers the oxidation of LDL	Prevents the plaque formation associated with LDL oxidation	Largely hypothetical, although antioxidants are plentiful in red wine
Lowers levels of fibrinogen in blood	Lessens the risk of clot formation on atherosclerotic plaques	Moderate supporting data
Exerts other anticlotting actions: lessens platelet stickiness; raises levels of prostacylin; lowers levels of thromboxane	Lessens the risk of clot formation on atherosclerotic plaques	Inconsistent data; possible reversal of effect with heavy or binge drinking
Lessens insulin resistance	Lessens key risk factor for adult-onset diabetes and atherosclerosis	Evidence comes from a small number of studies
Lessens psychosocial stress	Unclear	No supporting data or likely mechanism
Improves conditioning of heart muscle	Imparts better resistance to damage from oxygen deprivation	Preliminary supporting evidence

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

WINE, BEER OR SPIRITS?

Beer, wine and liquor all seem to be related to a lower risk of coronary heart disease (CHD). A tantalizing question, however, is whether one kind of drink—wine, for example—is better than the others. The short answer: the jury is still out.

The death rate from CHD in France, where red wine consumption is common, is only about half that in the U.S., despite similar fat intake and sedentary lifestyles. That observation led to the catchphrase "the French paradox" and the idea that red wine is *the* beneficial alcoholic beverage. This belief has a hypothetical basis—red wine especially contains a number of ingredients with potential antioxidant and other atherosclerosis-fighting benefits.

An excellent 1995 Danish study, in which almost 13,000 people were followed during a 12-year period, suggested that wine drinkers have lower death rates from CHD than do other alcohol imbibers. My Kaiser Permanente colleagues Mary Anne Armstrong and Gary D. Friedman and I published on the risk of CHD death (in 1990) and the risk of CHD hospitalization (in 1997); in these investigations, which included almost 130,000 Californians, wine and beer drinkers had a lower CHD risk than did hard-liquor drinkers. At a meeting of the American Heart Association in November 2002, I presented new data that updated the 1990 study. We were surprised to find that those drinking wine daily had about a 25 percent lower risk of CHD death than did those who drank beer and wound up taking in the same amount of alcohol. And the wine drinkers had about a 35 percent lessened CHD death risk compared with the light to moderate hard-liquor drinkers. Significantly, there was no difference in apparent benefit between red wine and white wine.

A vexing complication of all these studies, however, is that the overall habits of wine drinkers, beer drinkers and hard-liquor drinkers tend to differ greatly. In Denmark, for example, wine drinking goes hand in hand with a healthful diet (high in fruits, vegetables, fish, salads and olive oil) and two other markers for better health in general: higher socioeconomic status and higher IQ. In our California studies, those who preferred wine also smoked less, had more education and had more temperate drinking habits than those who preferred beer or hard liquor.

Lifestyle differences among those who prefer one type of alcoholic beverage over another thus make it exceedingly difficult to determine whether the differences in apparent health effects are actually related to the beverage type itself (and therefore to wine constituents besides alcohol), to drinking pattern (imbibed slowly and with food, for wine) or to other factors.

—A.L.K.

sents a challenge to physicians. On one hand, mild to moderate drinking seems better for heart health than abstinence for select people. On the other hand, heavy drinking is clearly dangerous. It can contribute to noncardiovascular conditions such as liver cirrhosis, pancreatitis, certain cancers and degenerative neurological disorders, and it plays a part in great numbers of accidents, homicides and suicides, as well as in fetal alcohol syndrome. (No conclusive evidence links light to moderate drinking to any of these problems.)

Heavy drinking also contributes to cardiovascular disorders. Too much alcohol raises the risk of alcoholic cardiomyopathy, in which the heart muscle becomes too weak to pump efficiently; high blood pressure (itself a risk factor for CHD, stroke, heart failure and kidney failure); and hemorrhagic stroke, in which blood vessels rupture in or on the surface of the brain. Alcohol overindulgence is also related to "holiday heart syndrome," an electrical signal disturbance that disrupts the heart rhythm. The name refers to its increased frequency around particular holidays during which people engage in binge drinking.

Given the potential dangers of alcohol, how can individuals and their physicians make the decision as to whether to include alcoholic beverages in their lives and, if so, in what amounts? The ability to predict accurately an individual's risk of a drinking problem would be a great boon; the least disputed possible consequence of moderate drinking is problem drinking. Individual risk can be approxi-

THE AUTHOR

ARTHUR L. KLATSKY is a senior consultant in cardiology and an adjunct investigator at the division of research at the Kaiser Permanente Medical Center in Oakland, Calif. A graduate of Harvard Medical School, he headed the medical center's division of cardiology from 1978 to 1994 and directed its coronary care unit from 1968 to 1990. Since 1977 he has been principal investigator of a series of studies of the link between drinking alcoholic beverages and health. His 1974 *Annals of Internal Medicine* article [see More to Explore, on opposite page] was the first published epidemiological report of an inverse relation between alcohol drinking and coronary disease; it was cited in 1995 by the National Institute on Alcohol Abuse and Alcoholism as one of 16 seminal articles in alcohol research. His most recent honor was a Health Forum Cardiovascular Health Fellowship for 2000–2001. Klatsky has completed six marathons and in 1990 climbed Mount Kilimanjaro.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

DRINKING: RISKS AND BENEFITS

Light/Moderate Drinking		Heavy Drinking	
RISKS	BENEFITS	RISKS	BENEFITS
Established	***Probable***	***Noncardiovascular***	None
Heavy drinking	Decreased risk of CHD	Liver cirrhosis	
	Decreased risk of ischemic stroke	Pancreatitis	
Unresolved	Decreased risk of gallstones	Certain cancers	
Breast cancer		Accidents	
Fetal damage	***Possible***	Homicides	
	Decreased risk of diabetes	Suicides	
Unlikely	Decreased risk of peripheral vascular disease (narrowing or clogging of the arteries carrying blood to the arms and legs)	Fetal damage	
Bowel cancer		Degenerative disorders of the central nervous system	
Hemorrhagic stroke			
High blood pressure		***Cardiovascular***	
		High blood pressure	
		Arrhythmia	
		Hemorrhagic stroke	
		Cardiomyopathy (damaged heart muscle)	

mated using family and personal histories of alcohol-related problems or conditions, such as liver disease or, of course, alcoholism. Even when known factors are taken into account, however, unpredictable events late in life may result in deleterious drinking changes.

Exactly because of these dangers, public health concerns about alcohol until recently have been appropriately focused solely on the reduction of the terrible social and medical consequences of heavy drinking. And the correlation between total alcohol consumption in society and alcohol-related problems has been used to justify pushes for abstinence. Ultimately, however, a more complex message is necessary. Merely recommending abstinence is inappropriate health advice to people such as established light drinkers at high risk of CHD and at low risk of alcohol-related problems—which describes a large proportion of the population. Of course, the most important steps for this group are proper diet and exercise; effective treatment of obesity, diabetes, high blood pressure and high cholesterol; and avoidance of tobacco. But there is a place on that list of beneficial activities for light drinking. Most light to moderate drinkers are already imbibing the optimal amount of alcohol for cardiovascular benefit, and they should continue doing what they are doing.

Abstainers should never be indiscriminately advised to drink for health; most have excellent reasons for not drinking. Yet there are exceptions. One case is the person with CHD who "goes clean"—quits smoking, switches to a spartan diet, starts exercising and, with good intentions, gives up the habit of a nightly bottle of beer or glass of wine. This self-imposed prohibition should be repealed. In addition, a number of infrequent drinkers might think about increasing their alcohol intake to one standard drink daily, especially men older than 40 and women older than 50 at high risk of CHD and low risk of alcohol-related problems. But women also have to consider one possible drawback of alcohol: several studies link heavy drinking—and a few even link light drinking—to an increased risk of breast cancer, a less common condition than heart disease in postmenopausal women but certainly quite a serious one. For young women, who are generally at low short-term risk of CHD and therefore may not benefit greatly from alcohol's positive cardiovascular effects, this possible breast cancer link looms larger in estimating the overall risks and benefits of alcohol. And for all women, the upper limit on moderate drinking should be considered one drink a day.

The only clear-cut message regarding alcohol and health, then, is that all heavy drinkers should reduce or abstain, as should anyone with a special risk related to alcohol, such as a family or personal history of alcoholism or preexisting liver disease. Beyond that, however, the potential risks and benefits of alcohol are best evaluated on a case-by-case basis. Cardiovascular surgeon Roger R. Ecker and I constructed an algorithm that can help health practitioners and their patients decide how much—if any—alcohol is right for a given individual [*see box on page 78*].

In short, health professionals should provide balanced, objective guidelines regarding their patients' use of alcohol, and such advice needs to be tailored to each person. I believe that it is possible to define a clear, safe limit for alcohol consumption that would offer a probable benefit to a select segment of the population. The ancient Greeks urged "moderation in all things." Three decades of research shows that this adage is particularly appropriate when it comes to alcohol. SA

MORE TO EXPLORE

Alcohol Consumption before Myocardial Infarction: Results from the Kaiser-Permanente Epidemiologic Study of Myocardial Infarction. Arthur L. Klatsky, Gary D. Friedman and Abraham B. Seigelaub in *Annals of Internal Medicine,* Vol. 81, No. 3, pages 294–301; September 1974.

Epidemiology of Coronary Heart Disease—Influence of Alcohol. Arthur L. Klatsky in *Alcoholism: Clinical and Experimental Research,* Vol. 18, No. 1, pages 88–96; January 1994.

Alcohol in the Western World. Bert L. Vallee in *Scientific American,* Vol. 278, No. 6, pages 80–85; June 1998.

Alcohol and Coronary Heart Disease. Giovanni Corrao, Luca Rubbiati, Vincenzo Bagnardi, Antonella Zambon and Kari Poikolainen in *Addiction,* Vol. 95, No. 10, pages 1505–1523; October 2000.

Alcohol in Health and Disease. Edited by Dharam P. Agarwal and Helmut K. Seitz. Marcel Dekker, 2001.

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

SYNTHETIC DIAMONDS

Carbon Copy

Consumers are finding it harder to distinguish synthetic diamonds from the real rocks. Since the 1950s scientists have grown imitations by depositing carbon atoms onto tiny natural or synthetic seed diamonds, under high pressure and temperature. But ever finer control of the temperature gradient around the seed is eliminating the telltale imperfections that jewelers could once detect with a magnifying glass.

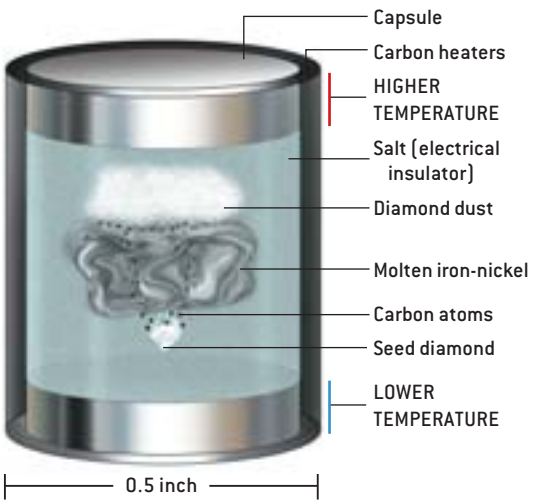
The synthetic approach builds up the seed slowly to prevent metal atoms needed for the process from being trapped in the growing stone, which causes noticeable inclusions. It takes up to a week to assemble a one-carat stone. Diamonds for industrial cutting tools can be grown faster, because inclusions do not reduce hardness. Polycrystalline diamond coatings for tools are laid down using chemical vapor deposition.

The original "belt" press, which squeezes the seed with pistons, is still employed today by the two leading makers of industrial synthetic diamonds: General Electric and natural-diamond king De Beers (neither retails synthetic gems). Russian physicists have perfected a spherical press that reportedly improves deposition control, spawning clearer gems faster, but details are proprietary. Gemesis, a start-up in Sarasota, Fla., recently hired several Russian alchemists and plans to produce thousands of synthetic gems a year.

The gems race has forced detection labs to invent more sophisticated instruments to prevent fraud. Adamas Gemological Laboratory in Brookline, Mass., has developed a spectrophotometer that shines full-spectrum light through a stone and detects specific wavelengths that are absorbed by occasional nitrogen atom clusters, common in natural stones but not in synthetics. Other instruments sense a stone's photoluminescence in laser light.

Diamond distributors worry that the spread of quality synthetics could erode consumer confidence in authentic gems. But if a synthetic stone is the same as a natural one, atom by atom, isn't it a real diamond? Readers may think twice about using that logic on Valentine's Day: "Honey, this diamond symbolizes my love. I bought it at Home Depot, right next to the power tools." —*Mark Fischetti*

HYDRAULIC PISTONS
of a diamond press squeeze a small capsule containing a seed diamond at pressures of 55,000 atmospheres (one million pounds per square inch). An electric current heats the capsule to 1,400 degrees Celsius or higher.



DIAMOND DUST
in the capsule diffuses down through a molten mix of iron and nickel, in one commercial process. The dust dissolves in the hot melt, then precipitates out one carbon atom at a time onto the cooler seed below, just the way water vapor precipitates as fog and clings to cooler grass as dew. The seed is a tiny natural or synthetic diamond fragment and grows about one carat a week. If pressure drops, the carbon atoms form graphite.

DANIELS & DANIELS (*illustrations*); ADAMAS GEMOLOGICAL LABORATORY (*graph*); COURTESY OF CHRIS WELBOURNE *De Beers DTC Research Center*

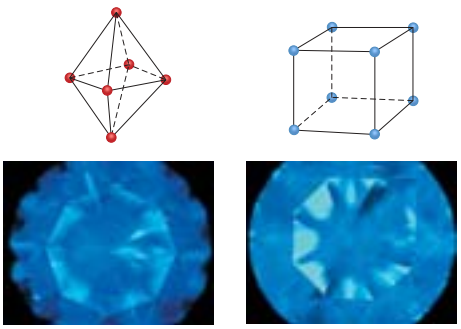
COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

DID YOU KNOW . . .

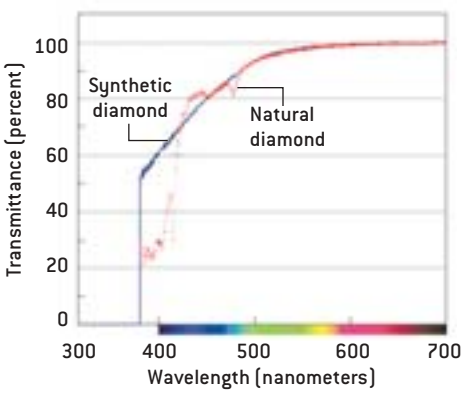
➤ CUBISM INDEED: Cubic zirconium "gems" were the first successful simulated diamonds, but they do not qualify as synthetics because they are not carbon. Zirconium oxide powder is melted at atmospheric pressures to about 2,400 degrees C with other compounds, such as calcium or magnesium oxide, and crystallizes out when cooled. Simple thermal detectors can identify these look-alikes.

➤ DEFORMED PRICES: Diamonds aren't quite pure carbon; other atoms occupy occasional crystal sites, giving stones their color, such as nitrogen (yellow) or boron (blue). About 1 percent are nitrogen-free, but they can have a brown cast because of crystal deformation. Pressing these stones in a synthetic-diamond machine beyond 60,000 atmospheres and up to 2,500 degrees C can essentially straighten the deformed crystal planes, creating a perfectly clear gem in a few minutes. "Browns" purchased for a few hundred dollars could be sold for $10,000 if clarified—perhaps a better payoff than growing synthetic diamonds from seed, which is expensive.

➤ FIRST TO PRESS: Baltazar von Platen of ASEA Laboratory in Stockholm is reputed to have made the first synthetic diamond in 1953, but he didn't publish his results. In 1955 a General Electric team led by H. Tracy Hall became the first on record to transform carbon into diamond. Hall and his son David R. Hall now oversee Novatek in Provo, Utah, which manufactures the presses that make synthetic stones.



NATURAL DIAMOND (*left*) has an octahedral crystal structure, visible during fluorescence after being illuminated with ultraviolet light. Synthetic diamond (*right*) contains cubic sectors, also visible during fluorescence.



SPECTROPHOTOMETER shines full-spectrum light through a synthetic diamond and a natural diamond of the same color and grade. Nitrogen atoms in the natural stone absorb blue light in a characteristic pattern from 430 to 480 nanometers; the man-made stone does not.

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

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Robots That Suck

HAVE THEY FINALLY COME OUT WITH A ROBOT FOR THE REST OF US? BY GEORGE MUSSER

For generations, tinkerers have been pointing out how much their projects will lighten the load of housework. For generations, spouses and parents have failed to be impressed by these claims. When I built my first robot seven years ago, people kept asking, "So what does it do?" I explained that it would eventually vacuum the floor. I should have just been honest: "Not much, but it sure is cool, isn't it?" All these years later I still have trouble getting my creations to do the most basic things, like move in a straight line. My professions of usefulness don't carry much weight around the house anymore.

At least I am not alone. Seldom in the history of technology has an industry been so eagerly anticipated, and so slow to emerge, as the consumer robot industry. Back in the early 1980s, when computers were turning from hobbyist playthings into mass-market appliances, it looked as though robots would soon follow. Heathkit's famous Hero I robot kit came out in 1982, not long after the original IBM PC. *Entrepreneur* magazine predicted a $2-billion home robot market by 1990. Today the original PC is a museum piece, and Hero I is still the state of the art.

Anyone who builds a robot appreciates what happened. When humans use a personal computer, we enter into the computer's world. If it can't do something, or if it crashes, too bad; we have to deal. But a robot enters into our world. If floors are uneven, if legs get in the way, if lighting conditions change, the robot has to deal. Extra computing power doesn't necessarily help; on the contrary, more sophistication typically means less resilience.

DON'T CRY OVER SPILLED DETERGENT if you have the Roomba floor cleaner, the first home robot that is both genuinely useful and reasonably priced. It won't completely relieve you of vacuuming duties, though. The robot is about 10 inches (25 centimeters) in diameter.

Through the school of hard knocks (lots of them), robot experimenters have learned to keep things simple. Massachusetts Institute of Technology professor and robo-guru Rodney A. Brooks led the way in the mid-1980s with a new style of robot programming, in which cheap sensors directly trigger elementary behaviors. Most robot kits these days, such as Lego Mindstorms, embrace this approach. And a similar design philosophy is reviving the fortunes of the home robot industry.

Some products, admittedly, achieve simplicity by giving up the pretense of doing anything useful at all. Robot dogs such as Sony's Aibo are the classic example. Others, such as robotic lawnmowers and pool cleaners, aim to do a single task in a highly controlled environment. The next step up is to do a single task in a highly uncontrolled environment, and the most obvious candidate for that is vacuuming. Over the past several years, a number of companies have promised to roll out floor-cleaning robots. A few of them have even delivered.

Apart from DustBot, a cheap but clever toy made by the Japanese company Tomy, the first consumer robot that could vacuum was Cye. Released in 1999 by Pittsburgh-based Probotics, Cye is the Apple II of robots: just pull it out of the box and plug it in. I tested one back in the fall of 2000. It's about the size of a toaster oven, with two wheels, a pair of wheel odometers to measure its movement, and

BETH PHILLIPS

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

a bump switch to sense when it hits something. To prove its usefulness, it can tow a small upright vacuum cleaner.

You control Cye from a PC via a wireless link, and the desktop software is where Cye really shines. As the bot blunders around, it relays back odometry readings, and the software estimates its position by dead reckoning. Crucially, the software keeps track of the uncertainty in its position; periodically the robot can reduce the error by reorienting against an object of known position, such as a wall. You can map a room, automatically calculate paths from A to B, and designate no-Cye zones—very handy in a home or office where not everyone shares your robotic enthusiasm.

For all its dummy-proofing, though, Cye still appeals mainly to gadget freaks. The price, which used to be $700 until the company lost its senses and raised it to $2,700, puts off the practical-minded. The mapping software tends to crash, and the vacuuming mode is primitive—the bot sweeps back and forth in a rectangular area and doesn't suffer obstacles gladly. Even I got bored. Lacking other sensors or the provision to run your own programs, Cye isn't capable of the richness of behavior that even entry-level kits can provide.

Last October, Brooks's own firm, iRobot, based in Somerville, Mass., brought out Roomba, a robot tailor-made for vacuuming. The lead designer, Joseph L. Jones, is co-author of the 1993 book *Mobile Robots: Inspiration to Implementation,* which remains the single best guide for beginning hobbyists (it got me started). The main subject of the book, the Rug Warrior project, grew out of a floor-cleaning bot that Jones had built for a contest at M.I.T. Four years ago he and mechanical engineer Paul Sandin finally got company backing to turn it into a product.

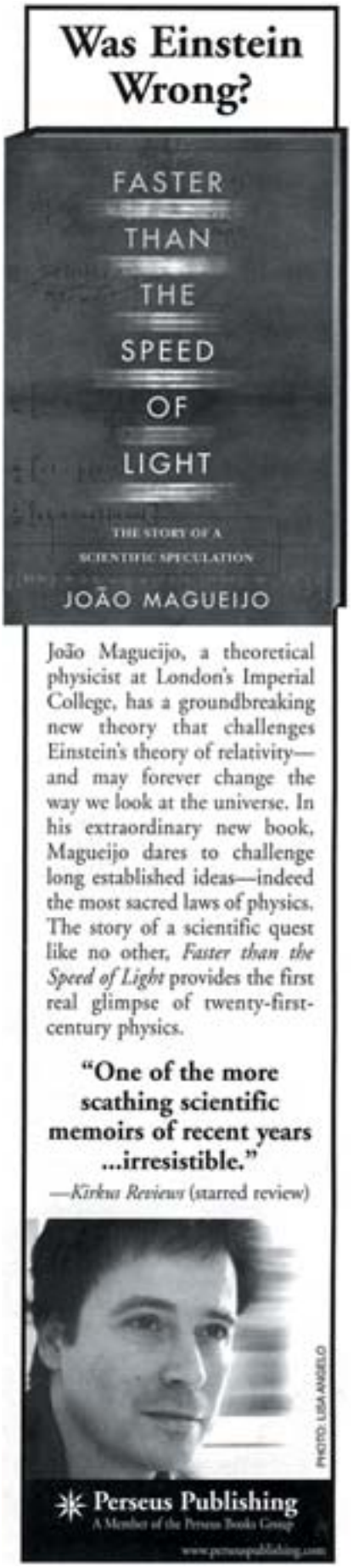
Roomba is roughly the size of a car hubcap and weighs about six pounds. The main cleaning mechanism is basically a Bissell carpet sweeper—one of those rug cleaners that is often found (and sometimes used) in college dorm rooms. A zigzagging wire forms a cage to keep the rotating brush from choking on the corners of rugs. A miniature weed whacker on the side flicks dust away from the base of walls. Behind the sweeper are two squeegee blades with a narrow slot between them—a "microvacuum" designed to suck up dust and hair. (Jones says the battery couldn't power a full-size vacuum.) The dirt ends up in a plastic cartridge.

The only controls are an "on" switch and three buttons to specify whether the room is small, medium or large. When you press one, Roomba starts moving in a spiral; after a while, it goes straight until it hits something, then turns, sometimes heading back toward the center of the room, other times executing a scallop-shaped path to try to follow a wall. The overall effect is a random walk. Half an hour later, give or take 10 minutes depending on room size, it declares victory and stops. You can also stop it by picking it up using the built-in handle. A battery charge lasts about an hour and a half.

Much as I tried, I couldn't entice Roomba to fall down a flight of stairs.

I tried Roomba on low-pile carpets and hardwood floors in rooms both empty and full. It didn't damage or topple anything, and it did remarkably well at extricating itself from power and phone cords, either by shifting direction or temporarily shutting off the brush. The edge detector—downward-pointing infrared sensors that watch for drop-offs—worked perfectly. Much as I tried, I couldn't entice Roomba to fall down a flight of stairs. I even put it on a table and let it clear off the crumbs.

Roomba slurped up most of the filth, but it didn't replace the need for manual vacuuming or sweeping, and iRobot is wise not to claim that it does. The real Achilles' heel of the robot, though, is the



COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

fulfilling job ✔
good friends ✔
hobbies and interests ✔
loving spouse

Do something about that last item by joining *Science Connection*, the network for single science professionals and others with science or nature interests.

Science Connection
(800) 667-5179
info@sciconnect.com
www.sciconnect.com

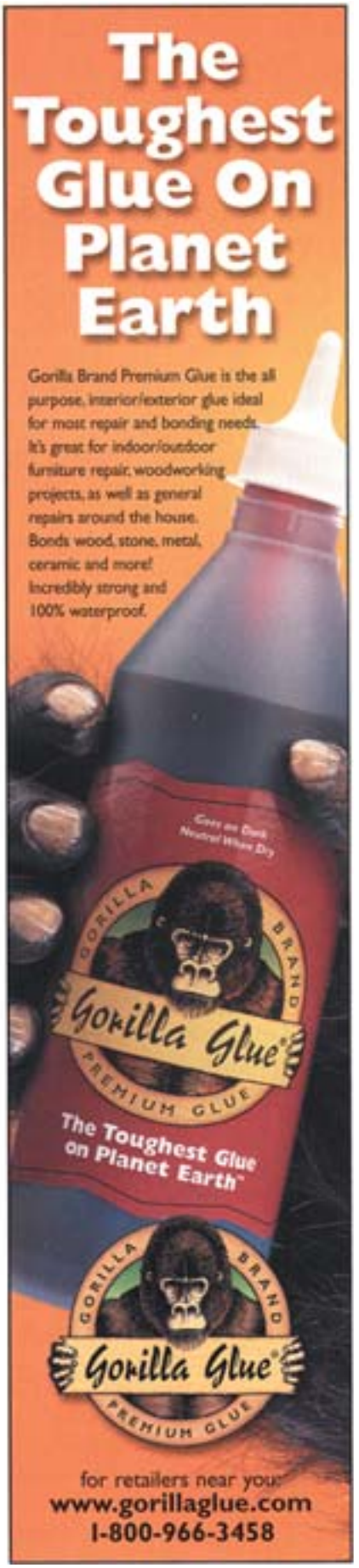
IT'S NOT THE MESSAGE.
IT'S NOT THE MEDIUM.
IT'S BOTH.

Only a sophisticated and innovative approach will reach a sophisticated and innovative audience. See what the Scientific American Custom Publishing Program can do for YOUR organization...

- EXTEND or enhance the reach of your conference/symposium
- DELIVER in-depth coverage of a product launch or a new corporate initiative
- BUILD your corporate image as a leader in cutting edge technology
- CONNECT and maintain an ongoing relationship with your customers

SCIENTIFIC AMERICAN
CUSTOM PUBLISHING

Contact us for a consultation
212.451.8859 or jabbate@sciam.com



wire that is supposed to keep rug corners from jamming the brush. It got yanked off within a couple runs, and the company had to send me a new one. Even with the wire, the bot didn't like the kilim in our living room one bit. And although it was usually able to free itself from cords, "usually" wasn't good enough: it got hung up at least once per run. You don't have to watch Roomba continuously, but you had better be nearby to help it. I think it's fair to say that Roomba rises above the level of mere gadget—but not by much. What makes it a breakthrough is the price, $200, which approaches the don't-need-spousal-preapproval range.

Roomba closely resembles a vacuum robot, Trilobite, that was introduced by Swedish appliance maker Electrolux in November 2001. Electrolux didn't respond to repeated requests for a demo model and doesn't sell Trilobite outside Sweden, but I tried it out in a shop on a visit there this past fall. Trilobite features a more powerful and rug-friendly vacuum; a sonar to detect obstacles, so it seldom actually makes contact with anything; and a position tracker, so it can return to its home base and plug itself in when it needs a charge. On the minus side, it lacks an edge detector, relying instead on magnetic strips that you lay around danger spots. Worse, its price, 14,000 kroner (about $1,500), is not likely to pass the spouse test.

Watching these robots bumble around gives you a new appreciation for how difficult housework really is. It takes agility, a tolerance for imprecision, and advanced pattern-matching skills—just the abilities with which evolution on the savanna endowed us. Rather than ever making a robot do all that, humans might, I suspect, find the tables turned on them: a future cyborg species could simply hire people to clean their homes for them. SA

For a review of another recently introduced robot, the ER1, see the online version of this article at www.sciam.com under "Technicalities."

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

The Next Big Thing?

USING METAPHOR INSTEAD OF MATHEMATICS, GEORGE JOHNSON BRINGS CLARITY TO THE STRANGE WORLD OF THE QUANTUM COMPUTER **BY JOSEPH F. TRAUB**

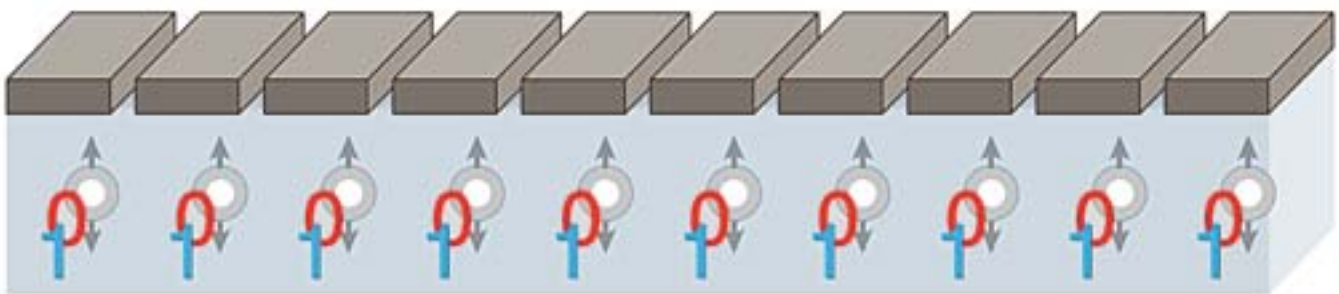
A SHORTCUT THROUGH TIME: THE PATH TO THE QUANTUM COMPUTER
by George Johnson
Alfred A. Knopf, New York, 2003 ($25)

In the 1960s Gordon Moore made the empirical observation that the density of components on a chip was doubling roughly every 18 months. Over the past 40 years, Moore's law has continued to hold. These doublings in chip density explain why today's personal computers are as powerful as those that only governments and large corporations possessed just a couple decades ago. But in 10 to 20 years each transistor will have shrunk to atomic size, and Moore's law, which is based on current silicon technology, is expected to end. This prospect drives the search for entirely new technologies, and one major candidate is a quantum computer—that is, a computer based on the principles of quantum mechanics. There is another motive for studying quantum computers. The functioning of such a device, which lies at the intersection of quantum mechanics, computer science and mathematics, has aroused great intellectual curiosity.

George Johnson, who writes about science for the *New York Times,* has set himself the task of deconstructing quantum computing at a level that readers of that newspaper—and this magazine—can understand. He has succeeded admirably. He explains the principles of quantum mechanics essential to quantum computing but tells no more than necessary. "We are operating here," he promises, "on a need-to-know basis."

One of the things readers really need to know about is superposition, a key principle of quantum mechanics, and Johnson proceeds to enlighten: "In the tiny spaces inside atoms, the ordinary rules of reality … no longer hold. Defying all common sense, a single particle can be in two places at the same time. And so, while a switch in a conventional computer can be either on or off, representing 1 or 0, a quantum switch can paradoxically be in both states at the same time, saying 1 *and* 0.… Therein lies the source of the power." Whereas three ordinary switches could store any one of eight patterns, three quantum switches can hold all eight *at once*. Thus, a quantum computer could process extraordinary amounts of information, and it could do so with such speed that it essentially takes "a shortcut through time."

In 1982 Richard Feynman conjectured that although simulations of the quantum world (needed for understanding the subatomic aspects of physics and chemistry) could never be done on a classical computer, they might be possible on a computer that worked quantum-mechanically. But interest in quantum computing didn't really take off until 1994, when Peter Shor, a mathematician at Bell Labs, showed that a quantum computer could be programmed to factor huge numbers—fast. There is a reason for the fascination with factoring large integers (breaking the large number into the smaller numbers that can be multiplied together to produce it). "Many of society's secrets, from classified military documents



IN A QUANTUM COMPUTER, each switch can be 1 and 0 at the same time, whereas a classical computer codes information as either 1 *or* 0.

COURTESY OF ALFRED A. KNOPF (COLORIZATION DONE BY *SCIENTIFIC AMERICAN*)

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

to the credit card numbers sent over the Internet, are protected using codes based on the near-impossibility of factoring large numbers.... Vulnerable codes are as disturbing to nations as vulnerable borders."

Despite such advances as Shor's algorithm and despite the importance for national security, serious impediments stand in the way of building a quantum computer. The superpositions from which quantum computing gets its power are lost when a measurement of the quantum state is made. And because the environment interacting with a quantum computer is akin to taking measurements, this presents a fundamental difficulty.

Another barrier is that although quantum computers with seven quantum bits (qubits) have been built, it is not clear whether the technology used—or any other technology—will scale up to handle enough qubits. Seth Lloyd of the Massachusetts Institute of Technology has estimated, for example, that interesting classically intractable problems from atomic physics can be solved by a quantum computer that has some 100 qubits (although error correction will require many additional qubits). Researchers are exploring a variety of technologies for building a quantum computer, including ion traps, nuclear magnetic resonance (NMR), quantum dots, and cavity quantum electrodynamics (QED). Which of these technologies, or technologies not yet conceived, will win out is not yet known. Of course, the unknown is part of the fun of science.

One of our most gifted science writers, Johnson is a master at bringing the reader along, giving increasingly better approximations to the truth. The book is lucid, elegant, brief—and imbued with the excitement of this rapidly evolving field. SA

Joseph F. Traub is Edwin Howard Armstrong Professor of Computer Science at Columbia University (homepage: www.cs.columbia.edu/~traub). His most recent book is Complexity and Information *(Cambridge University Press, 1998).*

THE EDITORS **RECOMMEND**

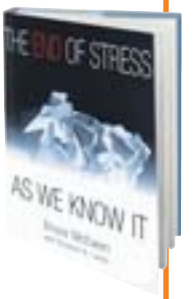
WINTER WORLD: THE INGENUITY OF ANIMAL SURVIVAL

by Bernd Heinrich. Ecco, New York, 2003 ($24.95)

There cannot be many people who have gone into a beavers' lodge. Heinrich, professor of biology at the University of Vermont, did that in his quest to see how animals survive winter. It was a summer when the pond had dried up and the beavers were not in residence, but with a flashlight and room enough to turn around, Heinrich was able to conclude that the accommodation would be quite cozy for a beaver family in winter. Similarly trying to see for himself as much as possible, he describes the winter survival strategies of many animals. He marvels in particular at the success of the golden-crowned kinglet (*Regulus satrapa*), a bird "scarcely larger than a ruby-throated hummingbird" that remains active all through the winters of Maine and Alaska, its life "played out on the anvil of ice and under the hammer of deprivation." The kinglet, he says, symbolizes the "astounding and ingenious strategies that animals have evolved for coping in the winter world."

THE END OF STRESS AS WE KNOW IT

by Bruce McEwen, with Elizabeth N. Lasley. Joseph Henry Press, Washington, D.C., 2002 ($27.95)

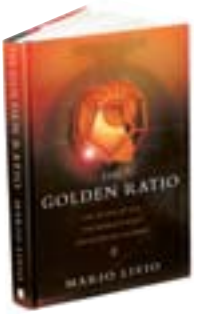


The stress response—fight or flee when confronting danger—has served humans well over the aeons, but under the multiple pressures of modern living it gets overworked. Chronic stress can cause health problems: cardiovascular diseases, disorders of the immune system, and afflictions of the mind "if normal feelings of distress and demoralization tilt towards clinical depression or anxiety." The message from McEwen, head of the Milliken Hatch Laboratory of Neuroendocrinology at the Rockefeller University, is that one does not have to fall victim to stress. "We cannot, and should not, eliminate the fight-or-flight response, for it is a powerful, highly sophisticated response. But we will be able to find ways of keeping the stress resource in balance, so that it works for us and not against us." Better-targeted medications will help. But "the best way to deal with stress is by maintaining our physical and emotional health."

THE GOLDEN RATIO: THE STORY OF PHI, THE WORLD'S MOST ASTONISHING NUMBER

by Mario Livio. Broadway Books, New York, 2002 ($24.95)

Most schoolchildren learn the importance of the number π, or pi (3.14159...), early on in their mathematical training. The area of a circle? $\pi(r^2)$. Circumference? $2\pi r$. But for astrophysicist Livio, the number ϕ, or phi (1.61803...), holds more wonder. Known as the Golden Ratio, Golden Number and other names related to gold or divinity, phi can be used to describe the characteristics of seashells, sunflowers, paintings and pentagrams. Livio traces the history of phi's discovery in tandem with the development of arithmetic, algebra and higher mathematics, examining along the way whether its value was factored into the construction of the Egyptian pyramids or the Parthenon of Greece. Practical examples placed throughout the story, requiring nothing more than a handheld calculator or some scratch paper, help to illustrate phi's relevance to mathematics and the world at large.



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

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Five Trusty Flares

BY DENNIS E. SHASHA

Imagine you have two piles of six flares each. One pile includes three useless flares. Call that the bad pile, although you don't at first know which group is bad. To test a flare, you must light it (because they all *look* good). Once you do that, you'll find out whether it works or not. But even if it is good, you won't be able to use it again.

Imagine, too, that you are heading for the Arctic. Can you find a testing method for selecting flares that will yield a $3/4$ or better probability of ending up with five working flares at your destination? Put another way, can you devise a selection approach that would be expected to provide 75 (or more) out of 100 Arctic visitors with five perfect flares? How would the probability change if the bad pile contained four nonworking flares instead of three?

Warm-up: How could you be sure to get two good flares and waste at most four, assuming the bad pile has three bad flares? *Solution:* Start testing flares from one pile. If you find a bad flare after lighting four, you will know you have chosen the bad pile; therefore, pick any two from the other (good) pile. Conversely, if you find four good flares in a pile, you will know you have the good pile; take its remaining two flares.

Second warm-up: How can you achieve a $1/4$ probability of getting seven good flares if the bad pile again contains three duds? *Solution:* Pick a pile at random to get six of the flares. Then choose the seventh flare from the other pile. You have a one-in-two chance of choosing the good pile at first. If you do, then you will be choosing the seventh flare from the bad pile; you have a one-in-two chance (three good out of six total) of selecting a good one. Those two choices are independent, so the probability of collecting seven good flares is $1/4$ ($1/2 \times 1/2$). In other words, according to the laws of probability, 50 of 100 people would be expected to choose the good pile, and half of those would then pick a good flare from the bad pile, for a total of 25 percent. Note that if one pile contained four bad flares, only two out of six would be functional, and the probability of getting seven good ones would fall to $1/6$ ($1/2 \times 1/3$).

Dennis E. Shasha's latest puzzle book is Dr. Ecco's Cyberpuzzles, *published by W. W. Norton (2002).*

Answer to Last Month's Puzzle

Force C4 for seven seconds, D4 for 10 seconds, E6 for seven seconds and F6 for 13 seconds. The T protein will then chime at 45 seconds and once every 70 seconds afterward (115, 185 and so on). For a full explanation, visit www.sciam.com

Web Solution

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

THE PROBLEM: Consider two piles of six flares. One pile contains three flares that look good but don't work (*left*). Find a testing method that would be expected to provide five working flares at least three quarters of the time, as, for example, above.

HEIDI NOLAND

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

Sheer Lunacy

WHICH IS NUTTIER: DENYING WE EVER WENT TO THE MOON OR TRYING TO CONVINCE THE TRUE NONBELIEVERS? **BY STEVE MIRSKY**

Once upon a time—July 20, 1969, to be specific—two earthlings got out of their little spaceship and wandered around on the moon for a while. Ten more earthlings walked on the moon over the next three and a half years. The End.

Unfortunately, not quite. A fair number of Americans think that this whole business of moon landings really *is* a fairy tale. They believe that the moon landings were a big hoax staged on a set in the Mojave Desert, perpetuated apparently to convince everyone that U.S. technology was the bestest in the whole wide world.

Time to shave with Occam's razor. Which is the harder thing to do: Send men to the moon or make believe we did? The fact is, the brute-force physics behind blasting people to the moon is simple. You can do it with slide rules and with computers whose entire memory capacities can now fit on RAM chips the size of postage stamps and that cost about as much as, well, a postage stamp. I know you can, because we *did*.

Nevertheless, last fall NASA considered spending $15,000 on what amounted to a public-relations campaign to convince the unimpressed that Americans had in fact gone to the moon. That idea was mostly a reaction to a Fox television program, first aired in February 2001, that claimed to expose the hoax [see "Fox's Flapdoodle," by Michael Shermer; Skeptic, SCIENTIFIC AMERICAN, June 2001]. The show's creator is a publicity hound who has lived up to the name in more ways than one by hounding Buzz Aldrin. Mr. X (as I will call him, thereby denying him the joyous sight of his name in print) recently followed the second man on the moon around and called him "a thief, liar and coward" until the 72-year-old astronaut finally lost it and—bang, zoom, to the moon, Alice—punched the 37-year-old Mr. X in the face.

Of course, the only Fox show that features good science is *The Simpsons,* on which Stephen Jay Gould, Stephen Hawking and, for that matter, Aldrin have all guest-starred, although there's some instructive physics in *The World's Most Hilarious Fatal Car Crashes*—or whatever they call this week's special in which large objects traveling at high speed smash into one another.

The moon-hoax show claimed that 20 percent of Americans have doubts about whether we ever really went (apparently up from the 6 percent that a 1999 Gallup poll identified). At first glance, that number looks alarming, but I would estimate that 20 percent of Americans probably think that the Fox show *Malcolm in the Middle* is a documentary about a family in crisis. (Sonoma State University astronomer Phil Plait supplies the details of the moon-landing polls and many other related items on his excellent Web site, www.badastronomy.com

Anyway, NASA's publicity campaign began to retrorocket as conspiracy theorists pointed to the effort as confirmation of something to hide and rational thinkers contended that $15,000 to convince people that the world was round—I mean, that we had gone to the moon—was simply a waste of money. (Actually, the $15,000 was supposed to pay for a rebuttal monograph by James E. Oberg, a serious astronomy writer who, with Aldrin, has contributed to *Scientific American.* As far as I'm concerned, paying a science writer is never a waste of money, but I'm severely prejudiced.)

If NASA's not paying Oberg, perhaps it could put the money to good use by hiring two big guys to drag Neil Armstrong out of the house. Armstrong is an extremely private man, but he is also *the first man on the moon,* so maybe he has an obligation to be a bit more vocal about the experience. Or NASA could just buy Aldrin a nice plaque commemorating his most recent moon shot, in which his fist slipped the surly bonds of decorum and touched the face of Mr. X. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

ASK THE EXPERTS

Why do some people get more cavities than others do?

Joel H. Berg, professor and chair of pediatric dentistry at the University of Washington School of Dentistry and president of the American Academy of Pediatric Dentistry Foundation, offers this answer:

Dental caries, the culprit behind the creation of cavities, is the most prevalent infectious disease in humans, affecting 97 percent of people at some point in their lifetime. Many factors are involved in the progression of tooth decay.

Caries is acid demineralization of the teeth caused by plaques of biofilms, complex communities of microorganisms that can coat surfaces in the mouth and reduce local pH levels. When tooth enamel is subjected to a pH lower than 5.5, it begins to demineralize, or break down; above this so-called critical pH, remineralization can occur. The success of this repair process depends on the presence of minerals in saliva, available fluoride ions and salivary flow rate. When the demineralization side wins this tug of war over time without compensatory remineralization, caries can progress to a visible cavity.

All bacterial biofilms are not alike, however. Although *Mutans streptococci* and other species have been implicated as primary culprits in causing caries, some people who are infected with these bacteria don't get cavities. So it is not simply the quantity of plaque biofilm present that leads to cavities.

Diet is another factor. Caries-causing organisms prefer sugars—specifically sucrose, or common table sugar—as the chief energy source. The metabolism of these sugars into lactic acid is what causes cavities. Controlling the number of sugar exposures—by limiting the consumption of sweets—aids in the remineralization side of the equation.

Salivary flow and composition also affect cavity production. In short, the more saliva there is in the mouth, the better it is at natural debridement—that is, scrubbing—of caries-causing organisms and the acids they generate off the teeth. Tooth morphology, or shape, makes a difference as well. Deep grooves on tooth surfaces (molars in particular) trap biofilms, making their removal by brushing and flossing more difficult.

Obviously, oral hygiene is key to keeping caries under control. Brushing and flossing must be performed religiously, preferably at least daily, to be effective.

Why are snowflakes symmetrical?

—V. Andersen, Santa Clara, Calif.

Miriam Rossi, associate professor of structural chemistry at Vassar College, explains:

Snowflakes reflect the internal order of water molecules as they arrange themselves in their solid forms—snow and ice. As water molecules begin to freeze, they form weak hydrogen bonds with one another. The growth of snowflakes (or any substance changing from a liquid to a solid) is known as crystallization. The molecules align themselves in their lowest-energy state, which maximizes the attractive forces among them and minimizes the repulsive ones. In the water ice found on the earth, each molecule is linked by hydrogen bonds to four other molecules, creating a lattice structure.

As a result, the water molecules move into prearranged spaces. The most basic shape is a hexagonal prism, with hexagons on top and bottom and six rectangular-shape sides. This ordering process is much like tiling a floor: once the pattern is chosen and the first tiles are placed, then all the other tiles must go in predetermined spaces to maintain the pattern. Water molecules settle themselves in low-energy locations that fit the spaces and maintain symmetry; in this way, the arms of the snowflake are made.

There are many types of snowflakes. The differentiation occurs because each snowflake forms in the atmosphere, which is complex and variable. A snow crystal may begin developing in one way and then change in response to alterations in temperature or humidity. The basic hexagonal symmetry is preserved, but the ice crystal branches off in new directions. SA

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

MATT COLLINS

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.

© 2003 ROZ CHAST, FROM CARTOONBANK.COM

COPYRIGHT 2003 SCIENTIFIC AMERICAN, INC.