GENETIC CODE: EVOLVED TO EVOLVE • CHOICE AND MISERY

SCIENTIFIC AMERICAN

SPACESHIPS, INC.
The Race to Build a Low-Cost Launch Industry

APRIL 2004 $4.95
WWW.SCIAM.COM

HAS SCIENCE MISSED HALF THE BRAIN?

Neglected Cells Hold Keys to Thought and Learning

The First Nanochips Have Arrived

Dusty Clues to Hidden Planets

april 2004

contents

SCIENTIFIC AMERICAN Volume 290 Number 4

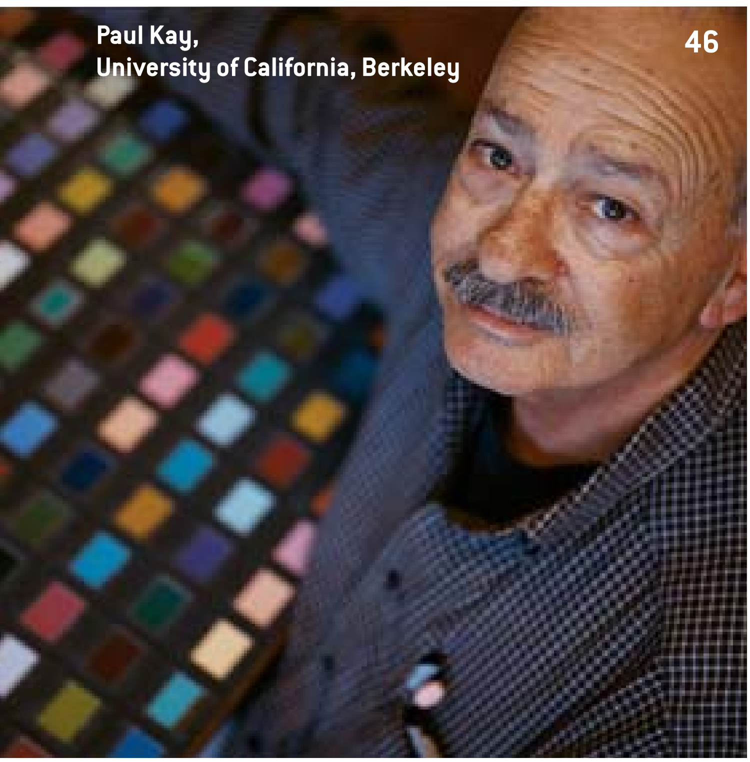
features

54 Glial cells regulate neurons in the brain

departments


100


111


Paul Kay, University of California, Berkeley 46

columns

Cover illustration by Jeff Johnson, Hybrid Medical Animation.

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

Breaking Out of Orbit

Shortly after 11:30 P.M. Houston time on December 13, 1972, the commander of *Apollo 17*, Gene Cernan, took one final look across Mare Serenitatis, climbed into the lunar module and closed the hatch. It was the last time anyone has had his boots planted in alien soil. Since then, the human space program has been adrift. Lacking an overarching mission, astronauts putter around in orbit doing make-work.

ASTRONAUTS on Mars?

This past January 14, President George W. Bush gave them something big to shoot for: a return to the moon by 2020 and a human mission to Mars sometime after that. His plan phases out the shuttle by 2010, replaces it by 2014 and abandons the space station in 2016. A presidential commission headed by aerospace veteran Edward C. "Pete" Aldridge, Jr., has started to flesh out the details, and NASA is already ramping up a technology development effort. Meanwhile the European Space Agency has laid out similar goals with a similar timetable and initial budget. Plenty of blanks need to be filled in, but that is natural in the early stages of a multigenerational project.

Does the budget add up? Many editorial writers and bloggers complain that Bush's sums don't match his lofty goals, but the numbers are more plausible than they might sound. Through 2020 the shuttle and station cancellations free up roughly $65 billion after inflation, and NASA gets another $18 billion or so in new money. The Apollo program, starting from scratch, cost $100 billion or so in today's dollars; a decade ago a General Dynamics report put the price tag of a revived moon program at less than a fifth of that. For a combined moon-Mars program, many critics cite a figure of $500-plus billion, but that estimate derives from NASA's notoriously extravagant 90-Day Study in 1989. Newer proposals cut the price in half or even a tenth—for example, by manufacturing fuel and water on the Martian surface rather than hauling everything from Earth. As with many long-term projects, no one knows exactly what the real costs will turn out to be, but the administration's funding plans are not unrealistic.

Can we afford it? In answering that question, one must keep the costs in perspective. Bush proposes to increase NASA's budget by 3 to 4 percent (after inflation) for the next three years and then hold it nearly steady. Even then, the agency will soak up just 0.6 percent of the total federal budget. By 2020, Americans will have spent more on potato chips than on the moon shot.

Will science get squeezed out? Many researchers think their fields will benefit from the new initiative [see "Fly Me to the Moon," by Mark Alpert, on page 20], but others fear their areas will take a hit when the human program runs over budget, as surely it will. Indeed, on its graph of the long-term budget, NASA lumped certain of its robotic missions together with human spaceflight. The two should be kept separate so that one program would not suffer from the mismanagement of the other.

Is NASA up to the task? Will it have the stomach to close bases, fire people and switch contractors if that's what it takes? Will its institutional culture be open to innovative ideas? If not, the country should consider founding a new agency or a public-private partnership or even multiple organizations to stir up competition. The private sector can't do it on its own, at least not yet [see "Blastoffs on a Budget," by Joan C. Horvath, on page 92].

The human space program has reached a go/no-go decision. Either give astronauts something meaningful to do or stop sending them into space. Muddle is no longer an option.

THE EDITORS editors@sciam.com

PAT RAWLINGS

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

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

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

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

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

Southeast
Publicitas North America, Inc.
404-262-9218
fax: 404-262-3746

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

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

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

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

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

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

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

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

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

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

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

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

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

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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

Digital Revolutionary: Interview with Leonardo Chiariglione

An electronics engineer and former vice president of multimedia at the corporate research laboratories of Italian Telecom, Leonardo Chiariglione is founder and chair of the Moving Picture Experts Group (MPEG), which has established such ubiquitous digital multimedia formats as MP3 and MPEG-2. Chiariglione would have good reason to rest on his laurels: in 1999 Time Digital ranked him among the top 50 innovators in the digital world, and his résumé lists an impressive series of awards, including an Emmy in 1996. Instead he has just called fellow experts to arms against "the stalemate" that he believes is crippling the development of digital media. Earlier this year Chiariglione established the Digital Media Project (DMP), a not-for-profit organization of individuals and companies—among them giants British BT and Japanese Matsushita Electric Works—with the ambitious goal of formulating a new standard for digital audio and video. If things proceed according to plan, the media world will never be the same.

Researchers Unveil New Form of Matter

Scientists have manufactured a new form of matter, a so-called fermionic condensate, which is composed of pairs of atoms in a gas at temperatures close to absolute zero. The achievement could help pave the way for room-temperature superconductors.

Ask the Experts

How can a poll of only 1,004 Americans represent 260 million people with just a 3 percent margin of error?
Andrew Gelman, professor of statistics at Columbia University, explains.

The Astronomy Channel

www.sciam.com/astronomy

Zoom in on Mars or travel to the farthest reaches of space—ScientificAmerican.com brings you the latest developments from the solar system and beyond. Explore the cosmos with our space news, a calendar of upcoming astronomical events and historical facts, plus new interactive features that put the universe at your fingertips.

SERGIO PISTOI

Letters

IN THE DECEMBER 2003 issue's cover story, "Does Race Exist?" authors Michael J. Bamshad and Steve E. Olson predicted that new genetic studies of matters related to race will lead to "a much deeper understanding of both our biological nature and our human connectedness." In the same issue, *Scientific American*'s Board of Editors recognized 50 visionaries whose work in research, technology and policy left the world a bit better at the close of the year. While geneticists work on the microscopic explanations of our human connectedness, the particularly international character of the responses to the "Scientific American 50" was heartening macroscopic evidence of our interrelations. The letters come together on the following pages.

RACE TRACKING

I was disappointed to learn from "Racing to Conclusions" [SA Perspectives] that the Food and Drug Administration is proposing that "racial" data be collected as part of clinical trials. Your article did not state strongly enough that the current racial/ethnic classifications promoted by the Census Bureau are archaic, inaccurate and confounding. Data derived from such classification are of extremely limited value, the main result being the perpetuation of outdated concepts about the human race.

Part of the trouble lies in the classification of peoples according to phenotype, given that phenotype can heavily misrepresent genotype. If we are really interested in "population group" frequencies, we should look at Northern European versus Southern European versus Central Asian versus West African, and so forth. The genes for skin color, facial features and hair texture are not necessarily linked to gene frequencies for disease states, and many medical diagnoses may be missed with this version of stereotyping. Unfortunately, much of the medical community continues to endorse the Census Bureau's racial/ethnic designations.

Marie F. Weston
President, Physician Consultant Services
Davie, Fla.

FEEDBACK ON THE 50

"The Scientific American 50" recognized Ken Livingstone, mayor of London, for championing the congestion charge in that city. His scheme has cut traffic but has also created a cash crisis.

Retail businesses in central London have had to lay off staff or close altogether as shoppers have turned to areas outside the capital, resulting in an ever shrinking tax base. Because the scheme is falling short of its revenue projections, the city must subsidize Capita, the private company hired to run the scheme. The biggest miscalculation is the hundreds of newly purchased buses. They stand idle around the city, spewing diesel fumes while they await the all-clear to travel already overserviced routes. The mayor plans to add even more buses, even though forecasters are projecting a loss of at least US$920 million by 2005. On top of this disaster, the mayor has said that Tube [subway] fares will rise by 25 percent in 2004.

In two years' time, the city will be facing a wrenching fiscal crisis and Londoners will be left footing the bill.

Stephen Previs
London

Anthony S. Fauci deserves his lauds as a Policy Leader for convincing the Bush administration to commit $15 billion to combat AIDS in Africa and the Caribbean. Unfortunately, related funding flaws can't be giving Fauci much satisfaction.

Because of conservative social policy and political considerations, one third of the funds intended for AIDS prevention focus on abstinence-only messages that are forbidden to mention condoms. Funds funneled through USAID may not be spent on syringe exchange, although such exchanges are proven to prevent disease transmission. Religious organizations get special consideration for grants rather than more capable nonsectarian family-planning groups, which have been starved under Bush's requirement that aid not be provided to any agency that mentions abortion. Moreover, the administration's restrictions regarding cheaper generic HIV medicines mean the costs of treatment will escalate for hard-pressed nations.

These are just the most obvious problems that will ultimately result in more AIDS deaths in Africa and the Caribbean. So keep hammering away, Dr. Fauci, there is so much more to be done.

Don Bay
Froson, Sweden

I disagree with your choice to honor Steve Jobs as a Business Leader on the basis of his "low cost" (99 cents per song) iTunes service. iTunes is actually exorbitantly expensive. A CD retails for around $16, and most of its cost is associated with distribution and retail markup. Internet downloading should enable a CD's worth of music to be sold directly by the record label for around $4, with no reduction in revenue to the label or the artist. At $4 per CD, music consumption (in unit volume) would skyrocket, enriching both artists and labels.

I believe that few people would pirate music were it available in the format they want at a reasonable cost. The recording industry has shot itself in both feet by not making the music available. Apple is helping on that front by removing $12 in cost from the chain and not passing a dime on to consumers. The company should get your Greedy Capitalist award.

Kirk Palmer
San Francisco

THE EDITORS REPLY: Awards of any kind always invite disagreement, and the SA 50 is no exception. We stand by our selections for the 2003 list, for the reasons cited in the original entries. Previs notes serious fiscal problems arising from Livingstone's congestion charge, but he also acknowledges that, as intended, the plan did accomplish the popular goal of reducing traffic. Continued work needs to improve the policy and reduce its negative effects (whether all the problems Previs indicates are the fault of congestion charging is controversial). We nonetheless salute Livingstone for providing a lesson from which all cities can learn. Regarding the letter from Bay, it should be noted that Fauci credits the president with wanting to commit heavily to fighting AIDS and that his own role was to work out the details. Palmer complains that iTunes is unfairly expensive, but a true market in music services will in time arrive at a fair price. The crucial difference between iTunes and earlier commercial services was that iTunes offered users the services they really wanted, such as the ability to buy individual songs and to store them on multiple devices.

SMASHING INTO a shallow sea, the asteroid that hit the Chicxulub region ignited a global holocaust.

ASTEROID FALLOUT

In "The Day the World Burned," David A. Kring and Daniel D. Durda write that following the collision of the asteroid with Earth, the two worst places to be were the impact site and "ironically, the place farthest away: India." It is not intuitive to me why the antipode of the impact site would be a focus point for debris.

Barry Goldstein
Newtonville, Mass.

KRING AND DURDA REPLY: As the plume material travels along its various ballistic trajectories around the planet, the focusing of reaccreting material near the antipode is a simple geometric effect of "polar crowding."

Think of the trajectories emanating from the impact site as being rather like lines of longitude emanating from the north pole of a globe. Not all ejected material travels as far as halfway around the globe, and some travels farther than halfway. But for those trajectories that travel even roughly halfway around the planet, the landing locations are all near the same point near the antipode of the impact, just like the lines of longitude reconverging at the south pole of a globe. The ejecta are launched at a variety of azimuths away from the impact site and then appear to converge from all directions from the point of view of the antipode. It is the one special place where the density of reaccreting debris is greatest. On the real Earth, the rotation of the planet offsets the region of greatest debris deposition to the west of the actual antipode. Our model accounted for this.

WRONGING THE WRIGHTS?

In "The Equivocal Success of the Wright Brothers," Daniel C. Schlenoff writes: "Not surprisingly, customers balked at buying so novel a device without seeing whether it worked." This is not correct. The written record of the Wright brothers' correspondence does not leave the impression that they expected money to change hands without a demonstration of what their airplane could do.

For example, in a letter dated October 9, 1905, to the Board of Ordnance and Fortification of the U.S. Government, Wilbur Wright writes: "We are prepared to furnish a machine on contract, to be accepted only after trial trips in which the conditions of the contract have been fulfilled ... the minimum performance to be a flight of at least twenty-five miles at a speed of not less than thirty miles an hour."

The Wrights were trying to secure contracts from good-faith potential customers, not from people more interested in a demonstration to get ideas that they could then "borrow" or modify. Suggesting that the Wright brothers were so unreasonable does not serve well either their memory or the historical record.

Donald DuBois
Portola Valley, Calif.

CHRIS BUTLER

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Tracking Tritium ▪ Chasing Dengue ▪ Questioning Evidence

APRIL 1954

SEX FOR PLEASURE—"Social, political and public health leaders in many countries are now seriously concerned with the population question and are taking active steps to disseminate family planning information in an effort to bring about a better balance between resources and populations. In attempting to introduce family planning measures, however, they are confronted with a major problem: the need for a contraceptive method which is simple, practical and within economic reach of everyone."

TRITIUM—"Until less than a decade ago men did not know tritium existed. It was discovered first as a synthetic product of nuclear transformation in a reactor; then it was detected in nature. The finding of tritium in nature was not easy. The total amount on our planet is about two pounds, and most of that is in the oceans, so diluted as to be beyond detection. Why bother to hunt down this infinitesimal substance? The answer is that tritium (radiohydrogen), like radiocarbon, may be an excellent tracer for studying natural processes. With it we can date plant products, and tritium in the earth's precipitation may tell us a good deal about the great movements of air and moisture over the face of the globe. — Willard F. Libby" [*Editors' note: Libby won the 1960 Nobel Prize in Chemistry for his work on carbon 14.*]

APRIL 1904

DENGUE VECTOR—"According to Dr. Graham, of Beirut, another disease is to be set down against the mosquito, namely, dengue fever, variously called African fever, break-bone fever, giraffe fever, dandy fever, etc. The disease is rarely fatal, but leaves various disagreeable sequelae: paralysis, insomnia, marked mental and physical prostration, etc. It occurs in hot climates and in the Southern States. In one experiment Dr. Graham carried dengue-infected mosquitoes to a mountain town 3,000 feet in altitude, where there were no mosquitoes and no dengue. One of the natives was shut up in the room with the mosquitoes, and on the fourth day came down with a sharp attack of dengue. The mosquitoes were immediately destroyed, and no further cases occurred."

SAFETY SUIT for firefighters, 1904

SMOKE SUIT—"The type of fire which is most dreaded by firemen is that in which volumes of stifling smoke and noxious gases are emitted. To enable firemen to successfully cope with fires of this kind a Colorado inventor has designed a garment resembling a diving suit which we illustrate herewith. This garment is composed of gas-tight material which hangs from the helmet and is strapped about the man's waist. The air within the garment is kept pure by means of proper chemicals stored in a box on the man's back."

APRIL 1854

EXPERT WITNESS—"One of the most important poisoning cases ever tried in our country was that of John Hendrickson, Jr., in June and July, 1853, for the murder of his wife, Maria. It was charged that he poisoned his wife with aconitine [wolfbane], and it was the scientific evidence which went to convict the prisoner. The whole testimony of the trial having been published, a copy of it fell into the hands of Prof. Wells, of Boston, who being deeply impressed by the utter want of soundness in the scientific testimony on which the prisoner was condemned, has submitted a petition signed by a number of the first-rate Chemists in our country, endeavoring to avert the execution." [*Editors' note: Hendrickson was hanged on May 5, 1854.*]

A LOVELY PLACE—"Dr. Hooker, in his 'Himalayan Journals,' just published, gives the following sketch of a pleasant excursion on the Nepaulese Himalaya: 'Leeches swarmed in incredible profusion in the streams and damp grass, and among the bushes; they got into my hair, hung on my eyelids, and crawled up my legs and down by my back. I repeatedly took upwards of a hundred from my legs where they collected in clusters on the instep; the sores which they produced were not healed for five months, and I retain the scars to the present day.' "

news SCAN

SPACEFLIGHT

Fly Me to the Moon

GOING TO THE MOON MEANS WINNERS AND LOSERS IN SCIENCE **BY MARK ALPERT**

When President George W. Bush declared in January that NASA would set its sights on returning astronauts to the moon by 2020, scientists quickly lined up on opposing sides. Although Bush's plan promises more funding for researchers studying the moon and Mars, other branches of space science are already feeling the pinch. The most prominent loser by far is the Hubble Space Telescope. Just two days after the president presented his initiative, NASA announced that it would cancel a shuttle flight to install new gyroscopes, batteries and scientific instruments to the Hubble. If NASA does not reverse the decision, its premier space observatory will cease operating when its current equipment fails in the next few years.

ONLY 12 ASTRONAUTS set foot on the moon in half a dozen landings between 1969 and 1972—here *Apollo 11*'s Buzz Aldrin shows off his boot. A new NASA plan calls for sending astronauts back to the moon by 2020, but some critics doubt the feasibility of the scheme.

NASA

The problem arises from the Bush administration's strategy of financing the moon effort through the early retirement of the space shuttle. During the phaseout, targeted for 2010, much of the shuttle's $4-billion annual budget will be shifted toward designing a crew exploration vehicle that could take astronauts to the moon. In the meantime, shuttle missions will focus on assembling the International Space Station.

NASA officials insist that they canceled the Hubble mission strictly because of safety concerns. To prevent a repeat of last year's *Columbia* catastrophe, NASA will require all shuttles to dock with the space station, where astronauts can inspect and repair damage to the vehicles or, if necessary, await a rescue effort. A shuttle bound for the space telescope would not be able to rendezvous with the station. But two reports written by a dissenting NASA engineer, who declined to be identified for fear of losing his job, claim that the agency could perform the Hubble mission safely by developing alternative repair methods and preparing a rescue mission in advance.

Although ground telescopes equipped with adaptive optics can match Hubble's resolution, they cannot duplicate all of the space telescope's abilities. For example, Adam G. Riess, an astronomer at the Space Telescope Science Institute, notes that ground telescopes cannot accurately measure the brightness of distant type Ia supernovae, which are used to gauge the expansion history of the universe [see "From Slowdown to Speedup," by Adam

G. Riess and Michael S. Turner; Scientific American, February]. "It's frustrating," Riess says. "It will be a long while before we have a way of doing this science again."

The biggest winners are the lunar geologists, who argue that the Apollo missions left many questions unanswered and that continued exploration of the moon could reveal much about the evolution of the solar system. The Bush plan earmarks $1.3 billion for unmanned missions to the moon over the next five years, including a lunar orbiter to be launched by 2008 and a robotic lander scheduled for 2009. Although both craft would pave the way for manned missions—by investigating potential landing sites, for instance—they would also provide researchers with a treasure trove of new data. "The moon is still mostly unexplored," says Alan Binder, the principal investigator for the Lunar Prospector orbiter that studied the moon in the late 1990s. "So lunar science can make a giant leap forward."

In some ways, planetary scientists know more about Mars than they do about the moon. The orbiters sent to the Red Planet in the past few years have thoroughly mapped its topography and mineralogy; in comparison, the moon maps obtained by Lunar Prospector and the earlier Clementine spacecraft are fuzzy and incomplete. The 2008 lunar orbiter could fill in the gaps by charting the moon's surface with radar imaging, laser altimetry and high-resolution spectroscopy. One probable goal of the mission will be to carefully delineate the permanently shadowed areas at the moon's poles, where some scientists believe that bits of water ice may be mixed in with the lunar dirt.

James Head, a planetary geologist at Brown University, hopes that the 2009 mission to the lunar surface will be the first in a series of unmanned landers. That craft may well carry a robotic rover similar to the Spirit and Opportunity vehicles that are now roaming the Martian surface. The moon mission, though, is more likely to be focused on applications that will aid human spaceflight—such as finding ice and learning how to extract it for life support or to produce rocket fuel by breaking the water into liquid hydrogen and oxygen.

"It's not really a science mission," says Paul D. Spudis, who was deputy leader of the science team for Clementine and is now a member of the space exploration panel advising the president. "The fundamental goal here is to expand the human presence in space." But given the uncertainty of the lunar initiative—critics in Congress doubt that NASA can send astronauts to the moon under the proposed budget—some researchers are wondering if the gains to science will outweigh the losses.

SHIFTING PRIORITIES

The White House's plan to send astronauts to the moon is already being incorporated into NASA's proposed budget. The projected outlays for the next five fiscal years show a decline in funding for the space shuttle and a rise in appropriations for unmanned lunar exploration, Mars missions and the development of new space transportation systems.

NASA Program	Budget Request (millions of dollars) 2005	2009
Space shuttle	4,319	3,030
Mars exploration	691	1,268
Transportation systems	689	1,863
Lunar exploration	70	420

GENETICS

Sobering Shift

GENE SEARCHES MOVE FROM ALCOHOLISM TO INTOXICATION BY SALLY LEHRMAN

Since the first "alcoholism gene," dubbed *DRD2,* was found in 1990, researchers have hunted for DNA sequences that might predispose someone to a drinking problem. But *DRD2*'s role in alcoholism has remained extremely controversial, and despite many efforts, no better candidates have emerged.

Many investigators are now taking a different tack. Instead of searching in families and populations of alcoholics for genes that might broadly confer a high risk for dependence, they are attempting to understand alcohol's effects and why they differ among people. In an explosion of studies, scientists have used rodents, fruit flies, zebra fish and roundworms to study characteristics such as sensitivity to intoxication and severity of withdrawal. By exploring alcohol's interaction with genes and the associated biological pathways, they hope to find clues to alcohol's addictive qualities.

Such studies are starting to yield intriguing results, including a recent report of a gene that some believe could have an important influence on dependence. Last December neurobiologist Steven McIntire of the University of California at San Francisco, who works with the worm *Caenorhabditis elegans* at the Ernest Gallo Clinic and Research Center in Emeryville, Calif., described a single gene that seemed to explain for the first time the mechanism of intoxication.

His team examined mutant worms resistant to alcohol's behavioral effects. They all had changes in a gene called *slo-1*. The gene ordinarily codes for a protein called the BK channel, found in nerve, muscle and gland cells. The channel operates as a gateway to control the flow of potassium ions. The researchers saw that alcohol makes the channel open more frequently, allowing more ions to pour out and slowing neuronal activity. In mouse and human cell cultures, alcohol similarly activated the BK channel, leading McIntire to believe that his group had uncovered the route for alcohol's diminishment of physical and mental control across species.

"*Slo-1* might determine sensitivity to alcohol as well as provide a mechanism for intoxication," McIntire says. The gene's influence on alcohol response probably makes it an important factor in dependency among people, he adds, pointing to studies by psychiatrist Marc Schuckit of the University of California at San Diego, who has been following for nearly 20 years 453 university alumni who are sons of alcoholic fathers. Among Schuckit's group, low sensitivity to alcohol's effects at age 20 has correlated with four times as high a risk of alcoholism later in life. Geneticist Raymond White, director of the Gallo Center, has already begun sequencing the genes of several hundred of Schuckit's subjects to investigate the role of *slo-1* in this population.

David Goldman, chief of the neurogenetics laboratory at the National Institute on Alcohol Abuse and Alcoholism, praises the efficiency of a simple system such as *C. elegans* as a way to sift through the genome. His own lab is searching for susceptibility genes in a large-scale analysis of Native American families with both high and low levels of alcoholism, a far more painstaking approach. McIntire's work "is a decisive sort of experiment because it creates immediately testable hypotheses," Goldman explains.

Some researchers, however, wonder if the Gallo Center team may be expecting too much from *slo-1*. David W. Crabb, director of the Alcohol Research Center at Indiana University, compares the gene with another candidate, *cheap date,* which makes fruit flies more easily drunk. Discovered in 1998, "it hasn't broken the field open," Crabb remarks.

SENSITIVITY TO INTOXICATION may have genetic roots that influence alcohol dependence.

ROY MEHTA *Photonica*

Evan Balaban, a neuroscientist at McGill University who works on inborn behavioral differences in animals, cautions against using physiological mechanisms to stand in for a complex syndrome. "Social, developmental and personal things all go into our need to use a substance and whether alcohol becomes the substance of choice," he states. A follow-up to Schuckit's study that took into account sociocultural factors, for example, found that low sensitivity to alcohol along with a family history accounted for just 22 percent of later alcohol abuse or dependence.

Balaban says the mechanism for intoxication ought to be interesting in its own right. "It may not matter whether this happens in a person who abuses alcohol or not," he suggests. "If someone is going to go out and drive, it might be a good idea to understand how intoxication develops so you can develop a pharmacological agent that will cut down on accidents."

Sally Lehrman writes about medicine and health from the San Francisco Bay Area.

ALCOHOLISM GENES: HARD TO COME BY

In their quest for an alcoholism culprit, researchers follow genetic threads through families with multiple generations of dependency, study the brain-wave patterns of alcoholics, observe behavior in rats and mice, and identify environmental influences on risk. So far only two genes, which code for enzymes called ADH2 and ALDH2 and were found in East Asians more than 20 years ago, remain solidly identified as related to alcoholism—and these two genes *protect* against it. Both genes influence alcohol metabolism rather than the addictive potential of the brain.

Missing Movement

MAD COW REVEALS THE LIMITS OF ANIMAL TRACING **BY WENDY M. GROSSMAN**

In mid-February the U.S. government gave up on its search for the herd mates of the first known U.S. case of bovine spongiform encephalopathy (BSE), popularly known as mad cow disease. The end of the trace-back effort, which began after the sick animal was uncovered in December 2003, means that the whereabouts and disposition of 52 of the 81 cattle that entered the country with the infected cow from Canada will remain uncertain. Of those 52, 11 were born at about the same time as the BSE cow and may have eaten the same contaminated feed that is presumed to have been the vector for the sickness.

The problem lies with the antiquated method of keeping tabs on animals—important not just for BSE but for other illnesses among livestock, such as foot-and-mouth disease, and for food poisoning resulting from *Escherichia coli* or *Salmonella* contamination. Unlike Canada, the U.K., the European Union and Australia, the U.S. does not mandate livestock tracking nationally. Moreover, there are significant regional differences in how animals are handled. Reliance on paper records contributes to slowness and inefficiency. And because only sick animals and their herd mates are followed, success in wiping out some livestock diseases (such as brucellosis) has, ironically, acted in the past few years to reduce the number of animals being tracked.

Efforts to change this situation began in 2002, a year after the world watched the British outbreak of foot-and-mouth disease bring massive culling and closed export markets, among other dire consequences. Last October the U.S. National Animal Identification Development Team, a joint effort of federal and state animal health officials, presented a working draft for its animal identification plan. It sets an aggressive timetable that would see the necessary systems and infrastructure in place by 2006, so an animal's history could be traced within 48 hours. Because of the sheer volume of livestock—100 million cattle alone—and the one million farms in this country, the plan proposes to implement a system using radio-frequency identification, or RFID. Such a fully electronic system would allow information to be sent automatically into a national database.

The success of the scheme depends not so much on the system being electronic but on the presence of a central authority collecting the data. The U.K. (and Europe) relies on plastic ear tags for its national system, established in 1997 as part of its effort to control BSE. From birth, each cow must have two ear tags—in case one falls off—with the same unique number that is registered with the Department for Environment, Food and Rural Affairs, along with the animal's sex and breed. This tracking system builds on the most crucial weapon against BSE—namely, the ban on forced cannibalism. Mad cow disease spread in the U.K. because infected cows were turned into feed for healthy ones.

HEALTHY HOLSTEINS from Washington State, where the first U.S. mad cow was found, wear ear tags as part of a system that was unable to find all the sick cow's herd mates. Future systems will rely on a national database and electronic tags.

JEFF GREEN *Reuters Newmedia Inc./Corbis*

Electronic tracing, however, has practical advantages over a plastic-tag system. Steve Rawlings, a small farmer in Telford, in central England, reports that "the tags break, get cut and also just rip out of the ear. Not nice, and a real pain to replace." And replaced they must be: every cow must have both tags whenever it is moved. Electronic tags promise to eliminate a lot of paperwork and to aid in the accurate selection of animals through the use of handheld devices, which can capture and download the data quickly.

Richard Webber, managing director of the Somerset, England–based Shearwell Data, which supplies tracking systems, says that until very recently electronic systems simply have not been ready for commercial use. For example, only in the past six months has it become possible to have two readers in the same building without their signals canceling each other out.

Many other technical issues remain to be settled. Tests are exploring the best way to insert an electronic ID tag: injecting it subcutaneously, inserting it in an ear, or having the animal swallow it in a ceramic capsule that lodges in a stomach compartment. Which works best may depend on the species, according to Robert H. Fourdraine, CEO of the Wisconsin Livestock Identification Consortium and a member of the national animal identification team. For cattle, the U.S. is leaning toward RFID ear tags. Webber, however, states that in his company's trials the swallowing method has proved most reliable; moreover, it would prevent unscrupulous farmers from swapping animal identities.

Researchers are also exploring the best way to identify the premises through which cattle travel and the feasibility of incorporating DNA information. Genetic data would permit a particular cow to be traced even after it had been butchered and dispersed onto store shelves—and would give new meaning to the old slogan "from farm to fork."

Wendy M. Grossman writes about communications and information technology from London.

SCIENTIFIC AMERICAN Help Desk

Scientific American is the world's premiere magazine of scientific and technological information for the general public. A 12-issue subscription includes the annual September single-topic issue. A year's subscription costs $34.97 in the U.S., $49 USD in Canada and $55 USD elsewhere.

www.scientificamerican.com includes exclusive online stories, articles from current and past print issues, the Ask the Experts Q&A, our online store, subscription center, digital archive, and more.

Scientific American Digital [www.sciam.digital.com] is more than just a digital magazine. Subscribers also access all issues of Scientific American from 1993 to the present. Scientific American Digital costs $34.95 per year. Special Online Issues and digital versions of the newsstand Special Editions are available for purchase separately.

Subscription inquiries Contact us online at www.scientificamerican.com, or send an e-mail to subscriptions@sciam.com. From the U.S. or Canada, call 1-800-333-1199; from elsewhere, +515-247-7631. Or write to Subscription Manager, Scientific American, Box 3187, Harlan, IA 51537. For changes of address, please notify us eight weeks in advance and include old and new addresses.

Back issues cost $10.95 each if mailed to a U.S. address; $13.95 elsewhere. Visit www.sciam.com/orderissues. Or fax your order with your Visa, MasterCard or AmEx information to 1-212-355-0408.

Reprints of individual articles cost $4 apiece, for a minimum order of 25 copies, prepaid. Send your order to Reprint Department, Scientific American, 415 Madison Avenue, New York, NY 10017, or e-mail to reprints@ sciam.com. The fax is 1-212-355-0408. Or call 1-212-451-8877.

Photocopying rights Scientific American, Inc., grants to libraries and others registered with the Copyright Clearance Center [CCC] the right 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 purposes. Send payment directly to the CCC. For further details [including address information], see www.copyright.com

Subscriber alert! Some subscribers have received unauthorized notifications and subscription offers from companies such as Publishers Services Exchange, United Publishers Network, Global Publication Services, Publishers Access Service, Platinum Publishing Service and Lake Shore Publishers Service. These are not authorized representatives or agents of Scientific American, and they are charging far higher rates for subscriptions and renewals than the actual prices. Please forward any correspondence you may receive from these companies to L. Terlecki, Scientific American, 415 Madison Ave., New York, NY 10017.

BIOPHYSICS

Magnetic Moods

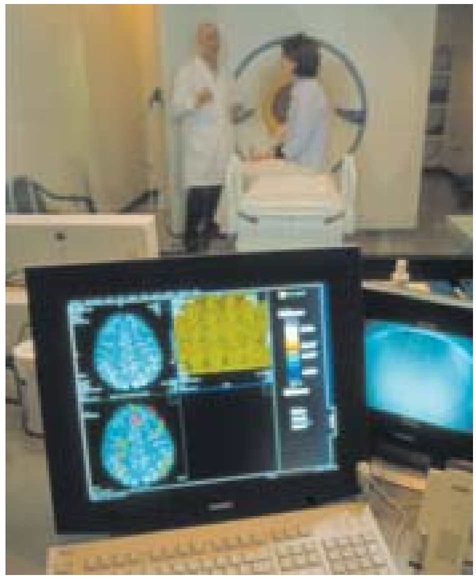
IN A SURPRISE, BRAIN SCAN CLEARS PATIENTS' GLOOM BY EMILY HARRISON

A 20-minute spell in an MRI tube is nobody's idea of a good time. So when several depressed patients exited a novel scanning session laughing, joking and exhibiting generally jovial behavior, researchers led by imaging physicist Michael Rohan and imaging center director Perry F. Renshaw at McLean psychiatric hospital in Belmont, Mass., quickly decided to investigate. What their preliminary study suggests is that the unique induced electrical fields associated with that particular type of magnetic resonance imaging session could improve the mood of patients with bipolar disorder.

The scan used in the study was an echo-planar magnetic resonance spectroscopic imaging (EP-MRSI) procedure, a fairly new method of MRI that McLean researchers were using to observe the effects of certain pharmaceuticals on bipolar subjects at the time of the serendipitous observation. Of the 30 individuals who received the EP-MRSI scans, 23 reported immediate mood improvement, the team says in the January issue of the *American Journal of Psychiatry*. The scans did not affect healthy individuals, eliminating the unsettling possibility that such electromagnetic therapy could be used to get a one-shot hit of happiness.

The EP-MRSI treatment had no apparent adverse effects, unlike chemical antidepressants and shock therapy. It also seems better than another effort at electromagnetic bipolar therapy: repetitive transcranial magnetic stimulation, in which an electromagnetic coil held near the head induces an intense electrical field (500 volts per meter) focused in the cortical region just below the instrument. Although the procedure alleviates depression fairly successfully, it is linked to seizure and severe scalp pain under the coil. In contrast, EP-MRSI fields are weak (less than one volt per meter) and uniformly distributed across the cortex.

The McLean team does not know why EP-MRSI exerts an antidepressant effect, but the researchers note that the electrical field's one-kilohertz pulse rate matches the natural firing rate of brain cells. They have recognized, too, that the axons of the corpus callosum, a bundle of nerve fibers in the central



UPLIFTING: Michael Rohan demonstrates a type of MRI, called EP-MRSI, that improves mood. While patients are in the tube, their brain activity, seen on a computer screen, is recorded.

JIM HARRISON

cortex that coordinates activity between the brain's right and left hemispheres, orient in the same right-to-left direction that the EP-MRSI electromagnetic pulses steadily travel. Some studies suggest that in episodes of bipolar disorder, these two hemispheres get out of balance; the electromagnetic pulses of EP-MRSI may affect that imbalance.

Since the completion of the preliminary study, McLean investigators have constructed a small tabletop device that delivers the same critical electromagnetic fields as a conventional MRI scanner and found it to be effective in animal trials. In fact, the electromagnetic therapy compared well with Prozac in reducing anxiety in rats. Rohan hopes to use the device in human trials within the year.

"It's only a first look, and we need to stay realistic," he warns. "But we're excited about its potential as a treatment for human depression." Rohan also sees interdisciplinary opportunities in future studies of the mechanism behind the effect, citing the need for input from electrophysiologists, physicists and neurologists, among others. "If we can interact with cells at that level with EP-MRSI, that makes it a pretty powerful tool for research."

HIGH ANXIETY ON THE RAT SWIM TEST

To test antidepressants on rats, scientists rely on a method called the forced-swim test. Rats in a closed container of water will try to escape for several minutes before giving up and merely treading water. Researchers call this quiescence a "depressed" state and have found that rats given chemical or electrical antidepressants swim, dive and attempt to climb the walls longer before treading water than untreated rats do. Different types of treatments affect different aspects of the escape behavior.

PHYSICS

Shattered Glass

SEEKING THE DENSEST MATTER: THE COLOR GLASS CONDENSATE **BY DAVID APPELL**

Physicists investigating heavy-particle collisions believe they are on the track of a universal form of matter, one common to very high energy particles ranging from protons to heavy nuclei such as uranium. Some think that this matter, called a color glass condensate, may explain new nuclear properties and the process of particle formation during collisions. Experimentalists have recently reported intriguing data that suggest a color glass condensate has actually formed in past work.

Particles such as protons and neutrons consist of smaller particles called quarks and gluons. Just as electrons have an electrical charge and transmit their force via photons, quarks have a "color" charge and transmit their force via gluons. But one major difference is that gluons, unlike photons, interact strongly with one another. As protons or heavy nuclei, such as gold, are accelerated to nearly the speed of light, the quarks and gluons inside flatten into a pancakelike structure, a relativistic effect called Lorentz contraction. The energy of acceleration also produces more gluons. The flattened multitude of gluons then begins to overlap, falling into the same quantum state, similar to the way atoms in a low-temperature Bose-Einstein condensate overlap and behave collectively as one gigantic atom.

Besides being similar to Bose condensates, the squashed matter "bears some resemblance to ordinary glasses," says Larry McLerran, a theorist at Brookhaven National Laboratory who first formulated the concept of a color glass condensate. For instance, the color fields produced by the gluons point in random directions, like the small, diffuse electrical fields generated by the orientation of atoms in glass. Just as regular glass is an amorphous solid for short periods (years) but flows over long intervals (centuries), these high-energy gluons are in a glassy planar state that changes very slowly compared with timescales typical of nuclear systems. This state is common to all extremely high energy particles and should enable physicists to describe the distributions and scattering probabilities of particles produced during collisions.

The color glass condensate can "shatter" in a collision. The shattering


PARTICLE SWARM resulting from a collision between deuterons and gold ions might indicate that a color glass condensate formed and then shattered.

BROOKHAVEN NATIONAL LABORATORY

can produce a quark-gluon plasma, a bulk form of quarks and gluons. Although a discovery has not yet been announced, many physicists believe that a quark-gluon plasma, which would provide clues about the early universe, has been created in heavy-ion collisions in the Relativistic Heavy Ion Collider (RHIC) at the Brookhaven lab.

As a precursor to the quark-gluon plasma, the color glass condensate should have been created if the plasma formed, as McLerran and some experimentalists believe it has. Electron-proton scattering in the HERA accelerator in Hamburg, Germany, provided indications of a color glass condensate. But perhaps the clearest signals have taken place in collisions in the RHIC: both in gold-gold and in deuteron-gold collisions. (Deuterons consist of one proton and one neutron.)

To detect a quark-gluon plasma, physicists examine the spray of particles emitted perpendicularly to the beam axis. But to tease out signs of a color glass condensate, detectors look at very small angles (about four degrees) relative to the beam axis. There the effects of a large number of very low momentum gluons dominate. Both deuteron-gold and gold-gold collisions produce fewer particles (relative to other proton-proton collisions) at these small forward angles, a sign that the gold nuclei was in a color glass condensate state. The effect was first seen by the multi-institutional group referred to as the BRAHMS collaboration (for Broad RAnge Hadron Magnetic Spectrometer); two other collabo-

CONDENSATES OF ANOTHER TYPE

Another kind of condensate made headlines in January, when physicists at JILA and the University of Colorado said they had made a "fermionic condensate." It consists of atoms that ordinarily remain single but are induced to pair up near absolute zero. The fermionic condensate provides a framework to understand superconductors and may lead to ones that work at room temperature. See "The Next Big Chill," by Graham P. Collins, News Scan; SCIENTIFIC AMERICAN, October 2003, and www.sciam.com/news_directory.cfm

rations—the PHOBOS and PHENIX—confirmed the BRAHMS data.

"I think this is a very interesting hint that something is happening here," comments Gunther Roland of PHOBOS. "But I think there's still a lot of work on the theoretical side that is needed to confirm the color glass condensate as the reason for the effect we're seeing experimentally."

Theorist Miklos Gyulassy of Columbia University believes, however, that the experimental evidence for a color glass condensate is too indirect: "What has been presented so far is not enough, for me." He says that the condensate should in fact appear for gluons moving with even lower momenta than have been measured. Direct evidence for the condensate might not happen until the more energetic proton collisions occur in the Large Hadron Collider at CERN near Geneva in about three years or until there is an upgrade at Brookhaven, probably a decade away.

David Appell is based in Lee, N.H.

ECOLOGY

Double Distress

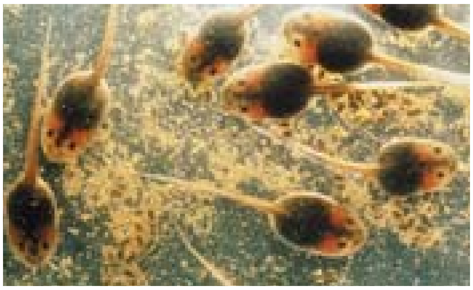
PESTICIDE KILLS FROGS ONLY IF PREDATORS ARE AROUND BY REBECCA RENNER

Amphibians are in decline, and the causes remain controversial. Among the earliest suspected culprits were pesticides, but the role of those toxic substances is not so obvious. Only a few reports have linked amphibian declines to pesticides. And even in those few studies, the pesticide concentrations appear to be too low to kill amphibians.

But University of Pittsburgh biologist Rick A. Relyea suggests that standard toxicology may greatly underestimate the power of pesticides on frogs in the wild. In the December 2003 *Ecological Applications,* he shows that carbaryl, a common pesticide sold as Sevin, is much more lethal to tadpoles—up to 46 times—when the pesticide is combined with another stressor: the presence of a predator.

Relyea kept tadpoles in water tanks that contained various amounts of carbaryl. Concentrations considered harmless on standard toxicity testing had little consequence, as expected. But many tadpoles died when the water contained tadpole-eating red-spotted newts, which were kept separate with netting.

Tadpoles are exquisitely sensitive to the smell of danger—for example, they react to just one dragonfly larva (another tadpole predator) in 1,000 liters of water, Relyea says. The carbaryl-newt data build on other findings by Relyea, who has now documented synergistic results in seven experiments with a total of six species of frog exposed to carbaryl. In an upcoming paper, he will also describe the same double-whammy effect of a predator and a common herbicide. His work "shows that the kinds of stressors prevalent in nature may be a key to understanding the real effects of pesticides on wildlife," says Yale University biologist David Skelly.



STRESSED: A toxic substance sickens tadpoles (seen here at about six weeks old) only when a predator lurks.

ADAM HART-DAVIS *Science Photo Library/Photo Researchers, Inc.*

The U.S. Environmental Protection Agency has reviewed Relyea's previous findings but believes that its regulations protect amphibians, because it bases its toxicity standards on Atlantic salmon, which are more sensitive to carbaryl than amphibians are. Still, the EPA remains interested in Relyea's synergistic effects, as are other herpetologists. "It's very difficult to prove that modern pesticides are a major cause" of amphibian declines, says biologist Donald W. Sparling of Southern Illinois University. Even DDT's role in wildlife problems took years to decipher, he notes: "We're going to have to rely on weight of evidence, and Relyea's study adds a very significant weight."

Rebecca Renner is based in Williamsport, Pa.

STILL MYSTERIOUS VANISHING ACT

The most important reason for amphibian decline is habitat loss, resulting mostly from human activities such as harvesting timber, draining wetlands and introducing nonnative species by, for example, stocking ponds and streams with trout. But amphibians are declining in "pristine" areas as well. Besides contamination from pesticides and other chemicals, these declines have been attributed to climate change, new diseases, parasites and higher levels of ultraviolet radiation. The emerging consensus is that no single overarching cause lies behind the global decline; instead several factors threaten amphibians to varying degrees.

news SCAN

BY THE NUMBERS

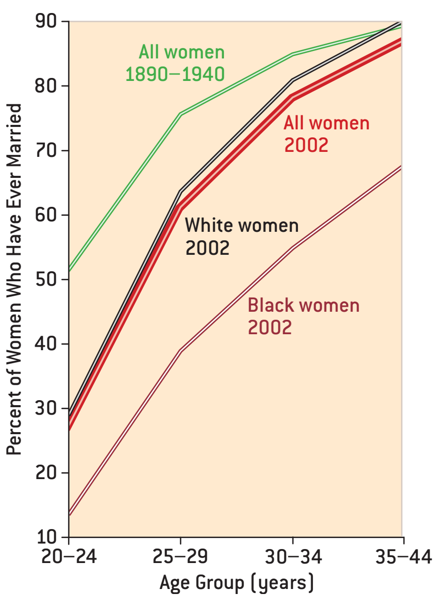
A Surplus of Women

THE SKEWED RATIO MEANS MARRIAGE IS LATER, NOT SOONER

BY RODGER DOYLE

In 19th-century America, men of marriageable age outnumbered women, in part because the immigrant stream was heavily male and because many young women died in childbirth. Changes in immigration and mortality now mean that the reverse is true. In 1890 there were 107 males for every 100 females in the 20- to 44-year-old group, but in 2002 the ratio had dropped to 98 per 100.

The present imbalance has led to exaggerated reports of female marriage prospects. For example, a widely publicized report in 1986 claimed that a white college-educated woman still single at 35 had a 5 percent chance of marrying; at 40, her chances declined to 1 percent. The conclusion seemed credible because it fed the stereotype that women who have a college degree have trouble finding a husband—a notion apparently originating in the late 19th century when marriage by female college graduates was low. A far more reliable forecast, based on more sophisticated analyses, comes from two Princeton University demographers, Joshua R. Goldstein and Catherine T. Kenney, who estimate that 97 percent of white female college graduates born between 1960 and 1964 will eventually marry.



Census data bolster their finding. Most women do not face a permanent single life but rather a delay in marriage, as illustrated by the chart. It shows that for all women aged 25 to 29 in 2002, 61 percent had ever married, compared with an average of 76 percent from 1890 to 1940. Among women who were 35 to 44 years old in 2002, however, 87 percent had married, only slightly less than the 89 percent recorded for the 1890 to 1940 period. Since the mid-1800s, more than 90 percent of women have eventually found husbands, and there is no reason to believe that the current generation of women will deviate much from this norm.

Besides the shortage of men, other factors have led to the postponement of marriage, including the increasing pursuit of higher education by women, the resurgence of feminism in the 1960s, and greater acceptance of premarital sex. According to one theory, the steep rise beginning in about 1970 in the number of women in professional schools resulted from the greater availability of the contraceptive pill. Controlling their own fertility enabled women to pursue education while not having to abstain from sex. Another deterrent to marriage is lack of information: when in school, women could easily meet and, with the help of friends, evaluate men, but such opportunities tend to diminish as women delay marriage.

Marriage is lowest among black women without a college degree: only 60 percent born between 1960 and 1964 will ever marry, according to Goldstein and Kenney's projection. Large numbers of black women have children out of wedlock, a circumstance that bears some relation to the scarcity of black men. Other elements have depleted the ranks of eligible men, including higher than average mortality rates, very high imprisonment rates and unavailability of good-paying jobs. Several studies suggest that welfare does not play a major role in lowering the black female marriage rate.

Rodger Doyle can be reached at rdoyle2@adelphia.net

RODGER DOYLE

STAYING ON THE MARRIAGE TRACK

Predicted percentage of women born between 1960 and 1964 ever marrying:

All races: 89
College graduates: 95
Without a college degree: 86

Whites: 93
College graduates: 97
Without a college degree: 92

Blacks: 64
College graduates *
Without a college degree: 60

*Not calculated but is greater than the rate of blacks who have not graduated from college

FURTHER READING

Too Many Women? The Sex Ratio Question. Marcia Guttentag and Paul F. Secord. Sage Publications, 1983.

Career and Marriage in the Age of the Pill. Claudia Goldin and Lawrence F. Katz in *American Economic Review,* Vol. 90, No. 2, pages 461–465; May 2000.

Marriage Delayed or Marriage Forgone? New Cohort Forecasts of First Marriage for U.S. Women. Joshua R. Goldstein and Catherine T. Kenney in *American Sociological Review,* Vol. 66, No. 4, pages 506–519; August 2001.

Why There Are No Good Men Left: The Romantic Plight of the New Single Woman. Barbara Dafoe Whitehead. Broadway Books, 2003.

SOURCES: Data in the chart are from U.S. decennial census, 1890–1940, and Census Bureau Current Population Survey, 2002. Data in table are from Goldstein and Kenney, 2001.

news
SCAN

DATA POINTS: HEAVY GOING

The periodic table gets two new members to fill in the bottom row: elements containing 113 and 115 protons, respectively. Scientists at Lawrence Livermore National Laboratory and the Joint Institute for Nuclear Research in Dubna, Russia, collided calcium 48 (20 protons) with americium 243 (95 protons) to make element 115 in two isotopic forms (172 and 173 neutrons). Element 115 released a helium nucleus to decay into element 113. These transuranic heavyweights support the idea that elements with "magic numbers" of both protons and neutrons would be extremely stable.

Lifetime of element 113: 1.2 second

Lifetime of element 115: 90 milliseconds

Heaviest natural element: 92 (uranium)

Heaviest known "doubly magic" isotope: lead 208
Protons: 82
Neutrons: 126

Next hypothesized doubly magic nucleus:
Protons: 114, 120 or 126
Neutrons: 184

Elements made:
1995 to 2004: 6 (111–116)
1985 to 1994: 1 (110)
1975 to 1984: 3 (107–109)

SOURCES: Lawrence Livermore National Laboratory; Physical Review C, February 1, 2004; Scientific American, September 1998.

BIOTECH

Rooting Out Bombs

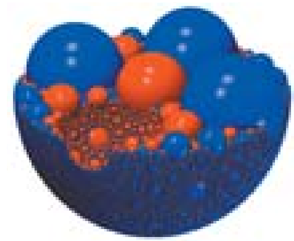
Land mines kill or injure some 26,000 people every year, and roughly 110 million remain unexploded in about 64 countries. Genetically engineered vegetation could help detect these hidden bombs. Biotechnology firm Aresa Biodetection in Copenhagen has modified the common garden weed thale cress (*Arabidopsis thaliana*). If their roots detect chemicals common to explosives, such as nitrogen dioxide, that leak out as mines corrode, the plants react as if it were autumn and change from green to red in three to six weeks. Aresa plans to test its plant, whose pollen has been rendered sterile, in small, restricted areas in Sri Lanka, Bosnia and other war-torn places. The hope is to clear mine-ridden land safely and cheaply so that farmers can resume cultivation. The company, which announced the plant's creation on January 24, is also working on plants to detect and remove heavy metals in polluted soil. *—Charles Choi*

RED WEED BAD: Explosives in soil cause plants to redden (*in each pot, quadrants to right of label "1"*).

PHYSICS

A Regal Bearing

When physicist Hans J. Herrmann of the University of Stuttgart in Germany heard a 1985 talk about tectonic plates sliding past each other with unexpectedly low friction, he began mulling over the nature of space-filling groups of ball bearings. He soon found theoretical arrays of two-dimensional disks that all turn in harmony, but a three-dimensional version proved elusive—no matter the arrangement, some balls would slip and rub, instead of turning against their neighbors. The physicist and his colleagues have now solved the problem theoretically. Imagine a sphere with six smaller spheres placed inside like the corners of a regular octahedron. The remaining space inside the big sphere can be completely filled with ever smaller spheres in a fractal pattern by a mathematical technique called inversion. Turn one sphere, and the rest turn without rubbing. A real bearing based on this model must consist of finitely many spheres, which Herrmann says would still be frictionless unless the balls were somehow forced out of place. Turn to the January 30 *Physical Review Letters* for the head-spinning result. *—JR Minkel*



ROLL ON: This theoretical model has ball bearings arranged so that none slips against another.

CHEMISTRY

New Path to Ammonia

For more than 70 years, synthesizing ammonia (NH_3) has meant cooking nitrogen and hydrogen gases at high temperature and pressure over a solid iron catalyst, a procedure called the Haber-Bosch process. Now chemist Paul J. Chirik and his colleagues at Cornell University have taken a major step in producing ammonia, which is crucial for fertilizers and other myriad products, under milder conditions—namely, in solution. They used a soluble complex made of two bulky hydrocarbon rings projecting from a zirconium atom. If the rings are just bulky enough, a molecule of nitrogen gas (N_2) will cuddle up and latch onto two zirconium complexes. A hydrogen atom binds to each nitrogen, and gentle heating breaks the nitrogen atoms apart, for still mysterious reasons. The metal complex ends up with extra hydrogen, which prevents it from fusing with additional nitrogen. The method won't replace the tried-and-true Haber-Bosch process, Chirik says, but it may open up faster synthesis of more complex nitrogen-containing molecules for dyes, rocket fuels and pharmaceuticals. The February 5 *Nature* has more. *—JR Minkel*

COURTESY OF ARESA BIODETECTION (*left*); HANS J. HERRMANN AND REZA MAHMOODI (*right*); ILLUSTRATION BY MATT COLLINS

MICROBIOLOGY

Leaving the Host Behind

No one thought the deadly germ that causes anthrax, *Bacillus anthracis,* could thrive outside a living host, but new data suggest that it grows just fine in ordinary dirt. University of Michigan Medical School microbiologist Philip C. Hanna and his colleagues experimented with local stream mud. They filtered out existing microbes and seeded the mud with dormant spores of a noninfectious anthrax strain. Hanna's team detected all stages of the germ's life. This result could explain why herds of cattle and big game experience anthrax outbreaks when hot, dry times follow rainy seasons: germs flourish after rainfall and concentrate in drinking holes after the water dries up. It remains uncertain how active the microbes are under natural conditions, given competitors that could restrain their spread. Still, the findings raise the question of whether anthrax trades genes with other soil microbes, "including ones for antibiotic resistance," Hanna said at the February meeting of the American Association for the Advancement of Science meeting in Seattle. *—Charles Choi*

ANTHRAX BACTERIA can grow outside a living host.

PSYCHOLOGY

Red with Prejudice

Anger may trigger spontaneous, automatic prejudices. Psychologists asked 87 volunteers to write in detail about events from their past about which they felt very angry, sad or emotionally neutral. Participants were subsequently assigned into two groups, color-coded either red or blue. Subjects had words from their written experiences linked to anger, sadness or neutrality flashed at them, followed by pictures of people from both color groups. Volunteers then were asked to quickly categorize fellow participants either positively or negatively. When angered, blue subjects evaluated red-coded individuals negatively, but not fellow blue members; the same was true for reds. Sadness and neutrality triggered no bias. The researchers, from Northeastern University and the University of Massachusetts at Amherst, say that the results could prove significant for professions that require snap decisions, such as law enforcement. The findings appear in the May issue of *Psychological Science.* *—Charles Choi*

THE MADDER I GET, the more biased I become.

PHILIP C. HANNA *University of Michigan Medical School* (*top*); LEFT LANE PRODUCTIONS/CORBIS (*bottom*)

BRIEF POINTS

- Researchers from Korea managed to create 30 cloned embryos of about 100 cells each, out of 242 donated eggs. Only one yielded viable stem cells, which turned into cartilage, muscle or bone when implanted into mice.

  Science online, February 12, 2004

- Greased quakes: Rocks abrading one another produce mineral powder that can combine with water to form a lubricating gel, a process that could reduce friction to zero and boost the release of earthquake energy.

  Nature, January 29, 2004

- A black hole in galaxy RXJ1242-11 has consumed about 1 percent of a star that ventured too close. The data, from x-ray flashes, provide the first observational evidence of the gravitational appetite of black holes.

  NASA announcement, February 18, 2004

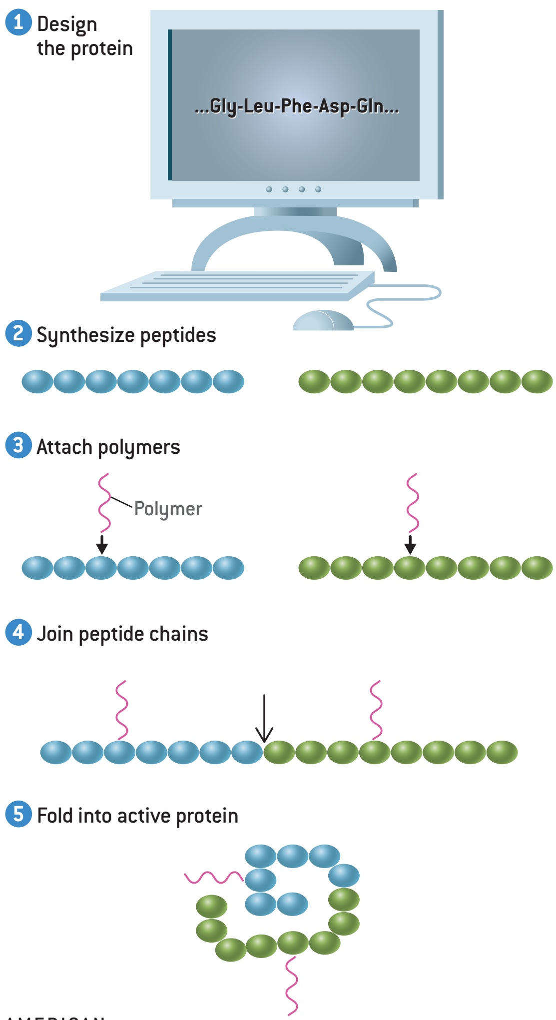
- Papaya king: The Y chromosome may have first emerged in the papaya plant. Its maleness gene is the most primitive yet found and probably resembles the human version before evolution pared the chromosome down.

  Nature, January 22, 2004

Making Proteins without DNA

A long odyssey produces a synthetic version of a biotech blockbuster By GARY STIX

Emil Fischer experimented with making polypeptides—chains of at least three amino acids—during the opening years of the 20th century. Fischer received the Nobel Prize in 1902 for his work on the synthesis of sugars and purines. But he never reached his goal of concocting a complete protein. Nearly 90 years later chemists were not doing that much better. The only practical methods of producing synthetic polypeptides had reached about the 50–amino acid mark, the size of the smallest proteins. But much of the attention had switched to recombinant-DNA methods that copied a gene and then inserted the clone into a cell that could pump out protein.



DESIGNER PROTEIN DRUGS can be built from synthesized peptides, chains of amino acids. Addition of polymers to the peptides improves their pharmacological properties, such as the duration of drug activity.

SAMUEL VELASCO

A few diehards, however, could still see the promise of synthetic chemistry. In 1989 biochemist Stephen B. H. Kent, along with colleagues from the California Institute of Technology used a synthetic process to make the HIV protease—the enzyme needed to make the virus fully functional. Then, along with collaborators from the National Cancer Institute, the team went on to determine the crystalline structure of the protein. "Some of us were too old-fashioned to stop making things by chemistry," Kent says. "We beat out people in the pharmaceutical industry who were trying to clone and express proteins."

But Kent was by no means completely satisfied with the methods he had used to make the HIV protease. Having moved to the Scripps Research Institute in La Jolla, Calif., he set about developing a means to supplant stepwise, solid-phase synthesis, a technique that would simply not work for larger proteins. Analogous to the stringing of beads, solid-phase synthesis links amino acids one by one into chains. But as a desired polypeptide grows, so, too, do unwanted by-products, until the synthesis becomes prohibitively inefficient. The workable limit for any given polypeptide maxes out at about 50 or 60 amino acids, corresponding only to the smallest proteins, such as insulin.

Kent's answer to this problem appeared in 1992 and 1994 *Science* papers, in which he and his collaborators spelled out how to link already assembled chains of 30 or so amino acids to produce larger proteins that folded up and behaved the way the natural proteins do. "That turned out to be a very powerful way of making proteins," he remarks. "Within a couple of years, we were making things with over 200 amino acids with a facility that had never been seen before."

In the middle of 1996 Kent left Scripps to devote all his time to a drug company he had founded, called Gryphon Sciences, that was trying to commercialize "synthetic micro proteins," which he predicted would rival recombinant proteins by the end of the decade. Kent's synthesis techniques worked well. But the company lacked direction. When its first chief executive left in 1998, the former academic, who was then Gryphon's president and chief scientific officer, also took over as chief executive. He set about trying to expand the company's push toward becoming what he called "Proteins R Us" to the pharmaceutical and biotechnology industries—a purveyor of synthetic proteins to other companies. The strategy did not quite work. "We were successful at helping the industry with their protein problems, but it was costing more than we were charging," Kent observes. "That's not a good business model, unless you're Enron."

Gryphon's version of erythropoeitin shows how chemists can tweak a protein drug's properties in a way that genetic engineers cannot.

Gryphon underwent a recapitalization in 2000 that ultimately brought $31 million in new money. Friedhelm Blobel, a consultant recruited to evaluate the business, agreed to take over the chief executive slot that had been occupied by Kent. The firm refocused on drug development, and the word "Sciences" in its name changed tellingly to "Therapeutics." Kent remained as chief scientific officer until 2001, when he accepted a professorship at the University of Chicago. He is now only an occasional consultant for Gryphon. "One works pretty hard at life, and you need to do things that are important to you according to your own set of values," Kent says. "I'm much more interested in science than I am in commercial success."

The revamped Gryphon had added an infrastructure for drug development, which includes pharmacologists, analytical chemists and regulatory officers. "It is much more focused and product-driven, which it wasn't before," says Gerd G. Kochendoerfer, the company's R&D director. Commercial recognition came in 2002, when Roche, the pharmaceutical giant, agreed to pay up to $155 million, along with royalties for exclusive rights to Gryphon's synthetic version of erythropoietin (EPO). In its recombinant form, this antianemia protein is one of the most successful biotechnology drugs of all time. Annual sales of recombinant EPO by Amgen and others totaled almost $7 billion throughout the world during 2002. Roche is now in an early phase of clinical trials of Gryphon's synthetic EPO. This compound complements a recombinant EPO that Roche sells in Europe and another EPO-based drug now entering the company's late-stage development pipeline.

Gryphon's EPO is more potent in rats and mice than EPO produced through recombinant-DNA methods and is said to have a more uniform composition. It also demonstrates activity for two to three times as long as first-generation recombinant molecules. The drug shows how chemists can tweak a compound in a way that genetic engineers cannot as of yet.

Borrowing a technique it patented based on research by two scientists at the University of Geneva, Gryphon inserts an unnatural amino acid (not one of the 20 encoded by DNA) into the EPO peptide backbone at two of the four sites where sugars are normally attached. The amino acid then serves as an anchor point for linking polymers to the peptide. The polymers act like sponges; their bloating prevents the protein from being rapidly excreted by the kidneys, allowing synthetic EPO to exert its red blood cell–generating action longer. The polymers also hinder the protein from being cut up by enzymes, and researchers are investigating whether they might diminish immune reactions experienced at times by patients who are injected with the recombinant EPO.

Not heard from yet are some suppliers of EPO drugs. Gryphon's synthetic erythropoiesis protein—and its method of manufacture—appears to differ sufficiently from recombinant ones to warrant its own patent. But Amgen has always been fiercely protective of its blockbuster, and a court battle would surprise no one.

Despite Kent's prediction, the era of synthetic protein drugs has not arrived. Drugmakers still prefer to analyze proteins to figure out how they can make a small molecule that mimics the larger entity's biological activity. But Gryphon's EPO, which could become the first completely synthetic protein drug, shows the potential of the process to manipulate medium-size proteins that measure up to 250 amino acids in size—a scale that includes growth factors and other pharmaceutically interesting compounds. Gryphon is working on "mirror image" proteins that can be used for screening drug candidates. Longer term, it wants to develop an HIV inhibitor and various cytokines (signaling proteins).

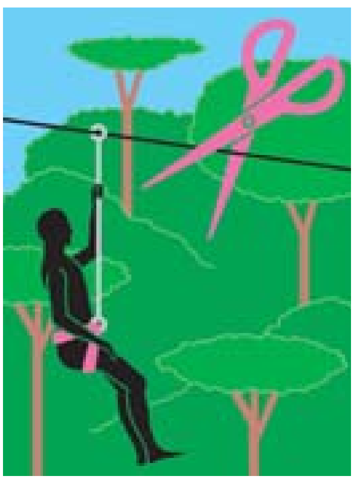
"We have total control over the structure of the protein," Kochendoerfer enthuses. "We can design what we want in every position of the protein and then put it there." So Gryphon has gone a long way toward proving that in the age of the genetic engineer, the synthetic chemist still has a substantial role to play. SA

Patent Enforcement

Vacationers to Costa Rica should check first with their tour operators' lawyers By GARY STIX

The U.S. intellectual-property system has distinguished itself in the past several years for such gems as patents on privatizing government, a method for using a playground swing, and a computerized system that handles reservations for going to the toilet. But patenting the obvious is by no means confined to the land of reality shows and SUVs.

In recent years, Costa Rica has given new meaning to the legal term "patent enforcement." It all has to do with the country's popular canopy tours, in which visitors strapped in a harness slide along a cable between treetop platforms. For Costa Rica, decade-old canopy tours are big business, generating a reported $120 million annually. It is estimated that a quarter of the more than a million tourists who come here every year patronize one of the 80-plus tour operations.



JENNIFER KANE

But the future of many of these businesses may now be in the balance because of a patent. In 1998 Darren Hreniuk, a transplanted Canadian entrepreneur, received a 20-year patent from Costa Rica's Industrial Property Registry for "an elevated forest transport system using harnesses and pulleys on a single horizontal line, using gravity for propulsion."

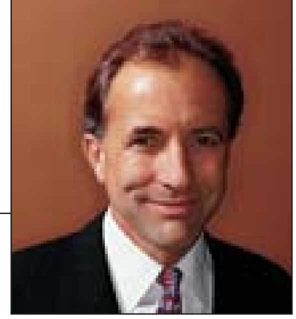
Last spring Hreniuk began to "enforce" his rights in the most literal sense of the word, according to reports in two Costa Rican newspapers, the *Tico Times* and *La Nacion*. With a cease-and-desist order from the country's Industrial Property Registry, issued on April 25, 2003, Hreniuk and police officers went to 14 canopy tour operators and tried to close them down unless each agreed to pay at least $75,000 for a franchise.

An attorney representing the besieged tour outfits entered a series of legal motions, and the registry's order was suspended for a number of months. But a Costa Rican Supreme Court ruling last November unfreezing the order sent Hreniuk back on the warpath. After the court's decision, Hreniuk and officials from the police and the Industrial Property Registry then tried to shut down several tour operators—and they did so reportedly by cutting cables and destroying platforms. By December the Costa Rican security minister had suspended the registry's order again—and then lawsuits brought against registry director Liliana Alfaro led to her suspension for two months.

One of the disputes surrounding the case centers on "prior art": previous technology that would undermine the claim in Hreniuk's patent application that his treetop apparatus is new and inventive. And putting this controversy to rest may be as simple as going to the Juan Santamaría Museum near the capital, San José, to inspect a piece of prior art that is actual art. There a painting shows soldiers crossing above the Barranca River in 1860 using ropes and pulleys during a pitched battle.

The wrangling over the canopy tours has spawned a bemused audience in the world capital of intellectual property. Patent gadfly Gregory Aharonian follows the case in his *Internet Patent News Service*, commenting on how U.S. Patent and Trademark Office examiners are not the only ones who ignore the wealth of prior art that is not contained in patent databases. The Juan Santamaría Museum painting is but one example of so-called nonpatent prior art. (Aharonian makes his living doing prior-art searches.)

The U.S. leads by example in novel forms of patenting. Still, a grudging respect may linger for Costa Rica, a country that once issued a trademark for the word "ecotourism." If pressed, the legions of American litigators, mired in the point-counterpoint of legal briefs, might reluctantly acknowledge a secret admiration for the slash-and-upturn methods employed by Hreniuk and the Costa Rican authorities in enforcing intellectual-property rights. SA



Magic Water and Mencken's Maxim

Social critic H. L. Mencken offers a lesson on how to respond to outrageous pseudoscientific claims By MICHAEL SHERMER

Henry Louis Mencken was a stogie-chomping, QWERTY-pounding social commentator in the first half of the 20th century who never met a man or a claim he didn't like ... to disparage, critique or parody with wit that would shame Dennis Miller back to *Monday Night Football*. Stupidity and quackery were favorite targets for Mencken's barbs. "Nature abhors a moron," he once quipped. "No one in this world, so far as I know ... has ever lost money by underestimating the intelligence of the great masses of the plain people," he famously noted. Some claims are so preposterous, in fact, that there is only one rejoinder: "One horselaugh is worth ten thousand syllogisms." I call this "Mencken's maxim," and I find that it is an appropriate response to preposterous claims made about magic water sold on the Web. I offer as a holotype of Mencken's maxim the following: Golden 'C' Lithium Structured Water (www.luminanti.com/goldenc.html).

"One horselaugh is worth ten thousand syllogisms."

This "is pure water infused with the energies of the Golden 'C' crystal, a very special and extremely rare stone mined near San Diego at the turn of the 20th century." The stone "contains more lithium than any other stone on the planet" and "emits a signature one-of-a-kind healing energy." How does the Golden 'C' water get these magical qualities? Crystal and water are placed in a ceramic container in a "dark and quiet space" for 24 hours, then the water is poured into "violet glass bottles" that "energize it." Finally, "each violet bottle is placed precisely within a special copper pyramid, specially designed to have the exact Sacred Geometry to create a Pillar of Light Jacob's Ladder vortex."

At only $15 per half-ounce, Golden 'C' water is a bargain because it "aligns and balances chakras and meridians; acts as a negative ion generator; clears stressful emotions and negative thought forms; clears all negative energy from crystals, food, rooms, people and pets; eases stress; disperses anger; improves immune system; clears bed of nightmare energy and previous energy of dreams; improves mental concentration; facilitates deeper meditations; hydrates and soothes skin; creates environment for visionary dreams." And, most important, it "clears and protects from electromagnetic pollution such as kitchen appliances, TV, microwave emissions from ovens and the environment, electrical clocks, stereos, high electrical wire lines, etc." As evidence we are offered this factoid: "Using an instrument to measure wavelengths of light, Holy Water from Lourdes, France, registered 156,000 angstroms of light. Golden 'C' water registered 250,000 angstroms of light!"

Wait! That's not the best Mencken moment to come. Just below the order button a warning label reads: "Note: no actual lithium is in the water. Only the energetics of lithium and the other minerals is contained in the water." Maybe that explains another disclaimer, perhaps written with attorneys in mind: "No therapeutic, drug or healing claims related to the physical body are made in the use of Golden 'C' Lithium Structured Water." One is advised, however, to keep it refrigerated.

In case any credulity remains, according to Ray Beiersdorfer, professor of geochemistry at Youngstown State University, "exposing ordinary water to lithium crystals, or any other crystals for that matter, cannot fundamentally alter the molecular structure of the water. The chemical structure within the water molecule, as defined by bond length and orientation, doesn't change. The claim that the chemical structure of liquid water changes because of exposure to a relatively insoluble crystal is nonsense."

For another Mencken moment, check out tachyonized superconductor water at www.tachyon-energy-products.com. Its promoter, Gene Latimer, explains its benefits: "I am now living in a radically different electromagnetic field environment that appears to be harmonizing the chaotic impact of electrical Alternating Current on the life forms in our house." *All* the lifeforms? Wow! And guess what? Tachyon is not limited to water. You can order tachyonized gel, algae, spirulina, herbs, mattress pads, massage oil and even "star dust." Sprinkle lightly.

We would all do well to follow another Mencken observation: "I believe it is better to tell the truth than to lie.... And I believe that it is better to know than to be ignorant." Amen to that, brother. SA

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

BRAD HINES

Draining the Language out of Color

Words mold many aspects of thought, says linguist Paul Kay, but not all aspects. The proof lies in the names the world's languages give to colors By PHILIP E. ROSS

Would a rose by any other name really smell as sweet? Do our words shape our thoughts, so that "we dissect nature along lines laid down by our native languages," as the linguist Benjamin L. Whorf asserted half a century ago? Is language a straitjacket?

Perhaps to some extent, allows Paul Kay, 69, emeritus professor of linguistics at the University of California at Berkeley. Those are hardly fighting words, and Kay, dressed in fuzzy shoes and a fuzzy sweater, his feet up on his desk, doesn't seem a pugnacious fellow. Yet he and his former colleague, Brent Berlin (now at the University of Georgia), have been at the center of a 35-year running debate concerning Whorf's hypothesis, called linguistic relativity. "Our work has been interpreted by some people as undermining linguistic relativity, but it applies only to a very restricted domain: color," Kay remarks.

AMANDA MARSALIS

PAUL KAY: IS LANGUAGE A STRAITJACKET?

- **Kay discovered that languages build their basic color vocabularies in a constrained process, suggesting that people everywhere perceive color in quite similar ways.**
- **English has 11 basic color terms: black, white, red, green, yellow, blue, brown, pink, orange, purple and gray; Russian adds *goluboy* (light blue).**
- **Modern tints: According to Kay, the number of words that a culture uses to describe colors corresponds with its degree of industrialization.**

Relativity can be demonstrated, somewhat trivially, to any landlubber who sees a mere boat when his mariner friend cannot help seeing a ship, skiff or scow. The stakes are raised considerably, however, when one extends the argument from man-made concepts to natural phenomena, as Whorf did in an essay, published posthumously in 1956. He argued that "we cut nature up, organize it into concepts, and ascribe significances as we do largely because we are parties to an agreement to organize it in this way, an agreement that holds throughout our speech community and is codified in the patterns of our language."

In sifting through the color terms of the world's far-flung languages, Berlin and Kay were reacting to ambitious statements, amplifying Whorf's, in many standard linguistics textbooks, that chose color as "the hard case, the locus classicus, of relativity," Kay says. "They must have thought that if arbitrary linguistic categories determine perception of color, then this must determine the perception of everything."

Kay and Berlin hit on color terminology in the early 1960s while comparing notes on their field research. Kay, a New York City–born, New Orleans–bred cultural anthropologist, had just returned from 15 months in Tahiti. Berlin, a linguistic anthropologist reared in Oklahoma, had been researching a Mayan language of southern Mexico. "We found that in both our languages, all the major color terms but one were exactly like those in English, and in the one area of difference, they differed in exactly the same way." (They grouped green and blue to form what Kay and Berlin called "grue.") That two such profoundly unrelated languages should name colors alike seemed to point to some universal linguistic pattern.

In the mid-1960s Berlin and Kay ended up at Berkeley. They had their graduate students scour the Bay Area for native speakers of foreign languages, quizzing them with standard color chips, not unlike those used as samples for paint. Their object was to establish the meanings of basic color terms—that is, those that could not be analyzed into simpler terms (such as "blue-green") and were not defined as characteristic of a given object (such as "salmon"). Later Berlin and Kay collaborated with other researchers to expand their sample to 110 languages.

Color lexicons vary, first of all, in sheer size: English has 11 basic terms, Russian and Hungarian have 12, yet the New Guinean language Dani has just two. One of the two encompasses black, green, blue and other "cool" colors; the other encompasses white, red, yellow and other "warm" colors. Those languages with only three terms almost always have "black-cool," "white-light" and "red-yellow-warm." Those having a fourth usually carve out "grue" from the "black-cool" term.

The tree of possibilities turned out to have branching points, some of them rather rare. Still, the manner in which languages can build up their color words is tightly constrained, suggesting the existence of universal constraints on semantic variation.

Berlin and Kay published *Basic Color Terms* in 1969, starting years of speculation that the pattern they discovered reflected a universal trait of neurophysiology. "I have been associated with the argument that it had to do with higher-order neural processing, but it now turns out that there is no physiological evidence for or against that view," Kay says. He is therefore an agnostic on why languages build color vocabulary the way they do.

Back in the 1960s a model of color vision propounded by the late Russell L. De Valois, a Berkeley psychologist, had been interpreted as establishing that the categories red, yellow, green and blue were hardwired into the brain. That interpretation, however, fell apart after the model failed to predict the mix of frequencies that the eye perceives as "pure" colors (for instance, the model did not explain why the reddest-looking red contains a touch of blue). That left no physiological rationale for color categories. A more recent theory attributes universals in color vocabularies to the way the world is colored—that is, to the natural distribution of wavelengths.

In any case, Kay notes, the degree to which the perceived world is man-made seems to explain the variation in the *number* of color words. Hunter-gatherers need fewer color words because color data rarely provide much crucially distinguishing information about a natural object or scene. Industrial societies get a bigger informational payoff from color words.

Kay plans to probe subtler aspects of color-naming. With Terry Regier of the University of Chicago, he wants to study variation across color-term systems and compare the results with predictions made by a number of psychological models. If, for instance, one or more basic colors belong to a color category, can that information be used to project the boundaries of the entire category? Do those languages that have the "grue" color, say, make the category larger than "green" and "blue" combined, smaller, or the same size? The goal is to explain a specific cognitive universal—color—in terms of general psychological processes.

Although linguistic relativity does not apply to the naming of colors, Kay explains, there is no reason to rule it out in the naming of other domains—size, sharpness, degree of consanguinity or whatever. He also sees no reason why language may not shape the way we think about some aspects of color (other than its names): "There is a wealth of evidence showing that what people treat as the same or as different depends on what languages they speak.

"Two key questions must always be kept separate," Kay adds. "One is, do different languages give rise to different ways of thought? The other is, how different are languages?" It is possible, he says, that the respective answers are "yes" and "not very."

It is unlikely that the world's languages are so different from one another that their speakers think in ways that are incommensurable.

One of the most interesting inquiries into these questions is being conducted at the Max Planck Institute for Psycholinguistics in Nijmegen, the Netherlands, where Stephen C. Levinson and his associates are studying the psychological consequences of the differing ways in which languages describe space. Several languages lack subjective terms analogous to "left" and "right," using instead absolute directions, akin to "north" and "south." In such a language, one might say, "There's a fly to the north of your nose."

Presented with an arrow pointing to their left, speakers of Guugu Yimithirr, a language of Australia, will later draw it pointing to the left only if they are still facing in the direction in which they saw the arrow in the first place. If, however, they turn around, they will draw it pointing to the right—that is, in the same absolute direction as the original arrow.

Here, then, is an example of language categories molding thought and behavior in a striking way. Kay concludes that linguistic relativists may be correct that the languages people speak mold their thoughts. "But it is unlikely that the various languages of the world are so different from one another, in underlying conceptual structure, that the ways their speakers think are incommensurable." Or as Terence Africanus, the Roman essayist, put it: "I am a man; nothing human is alien to me." SA

The Other Half

of the Brain

MOUNTING EVIDENCE SUGGESTS THAT GLIAL CELLS, OVERLOOKED FOR HALF A CENTURY, MAY BE NEARLY AS CRITICAL TO THINKING AND LEARNING AS NEURONS ARE

By R. Douglas Fields

JEFF JOHNSON *Hybrid Medical Animation*

GLIAL CELLS (*red*) outnumber neurons nine to one in the brain and the rest of the nervous system.

The recent book *Driving Mr. Albert* tells the true story of pathologist Thomas Harvey, who performed the autopsy of Albert Einstein in 1955. After finishing his task, Harvey irreverently took Einstein's brain home, where he kept it floating in a plastic container for the next 40 years. From time to time Harvey doled out small brain slices to scientists and pseudoscientists around the world who probed the tissue for clues to Einstein's genius. But when Harvey reached his 80s, he placed what was left of the brain in the trunk of his Buick Skylark and embarked on a road trip across the country to return it to Einstein's granddaughter.

One of the respected scientists who examined sections of the prized brain was Marian C. Diamond of the University of California at Berkeley. She found nothing unusual about the number or size of its neurons (nerve cells). But in the association cortex, responsible for high-level cognition, she did discover a surprisingly large number of nonneuronal cells known as glia—a much greater concentration than that found in the average Albert's head.

An odd curiosity? Perhaps not. A growing body of evidence suggests that glial cells play a far more important role than historically presumed. For decades, physiologists focused on neurons as the brain's prime communicators. Glia, even though they outnumber nerve cells nine to one, were thought to have only a maintenance role: bringing nutrients from blood vessels to neurons, maintaining a healthy balance of ions in the brain, and warding off pathogens that evaded the immune system. Propped up by glia, neurons were free to communicate across tiny contact points called synapses and to establish a web of connections that allow us to think, remember and jump for joy.

That long-held model of brain function could change dramatically if new findings about glia pan out. In the past several years, sensitive imaging tests have shown that neurons and glia engage in a two-way dialogue from embryonic development through old age. Glia influence the formation of synapses and help to determine which neural connections get stronger or weaker over time; such changes are essential to learning and to storing long-term memories. And the most recent work shows that glia also communicate among themselves, in a separate but parallel network to the neural network, influencing how well the brain performs. Neuroscientists are cautious about assigning new prominence to glia too quickly, yet they are excited by the prospect that more than half the brain has gone largely unexplored and may contain a trove of information about how the mind works.

See Me, Hear Me

THE MENTAL PICTURE most people have of our nervous system resembles a tangle of wires that connect neurons. Each neuron has a long, outstretched branch—an axon—that carries electrical signals to buds at its end. Each bud emits neurotransmitters—chemical messenger molecules—across a short synaptic gap to a twiglike receptor, or dendrite, on an adjacent neuron. But packed around the neurons and axons is a diverse population of glial cells. By the time of Einstein's death, neuroscientists suspected that glial cells might contribute to information processing, but convincing evidence eluded them. They eventually demoted glia, and research on these cells slid into the backwater of science for a long time.

Astrocyte glia activate distant neurons to help form memories.

Neuroscientists failed to detect signaling among glia, partly because they had insufficient analytical technology but primarily because they were looking in the wrong place. They incorrectly assumed that if glia could chatter they would use the same electrical mode of communication seen in neurons. That is, they would generate electrical impulses called action potentials that would ultimately cause the cells to release neurotransmitters across synapses, igniting more impulses in other neurons. Investigators did discover that glia had many of the same voltage-sensitive ion channels that generate electrical signals in axons, but they surmised that these channels merely allowed glia to sense indirectly the level of activity of adjacent neurons. They found that glial cells lacked the membrane properties required to actually propagate their own action potentials. What they missed, and what advanced imaging techniques have now revealed, is that glia rely on chemical signals instead of electrical ones to convey messages.

Valuable insights into how glia detect neuronal activity emerged by the mid-1990s, after neuroscientists established that glia had a variety of receptors on their membranes that could respond to a range of chemicals, including, in some cases, neurotransmitters. This discovery suggested that glia might communicate using chemical signals that neurons did not recognize and at times might react directly to neurotransmitters emitted by neurons.

Overview/*Glia*

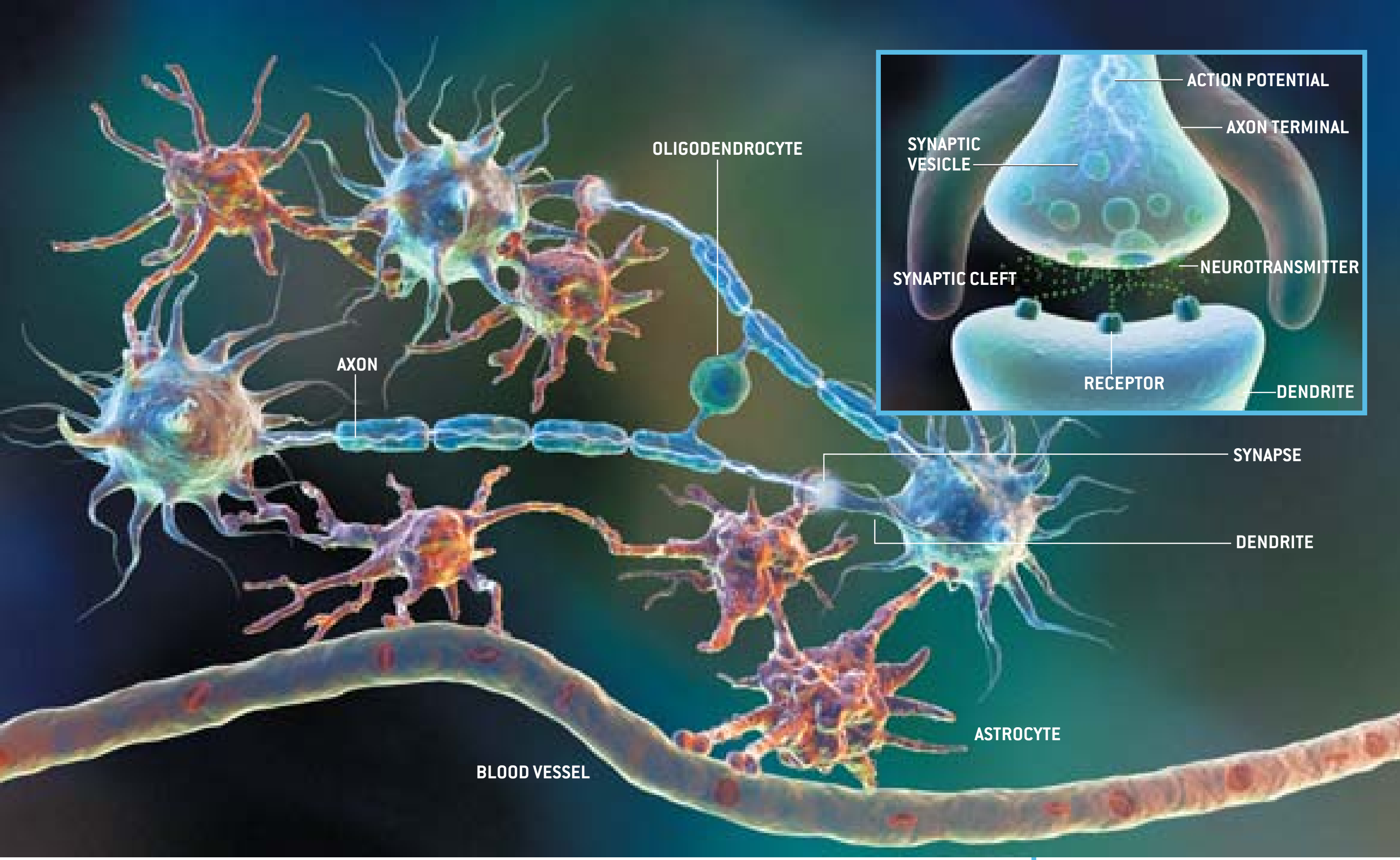
- For decades, neuroscientists thought neurons did all the communicating in the brain and nervous system, and glial cells merely nurtured them, even though glia outnumber neurons nine to one.
- Improved imaging and listening instruments now show that glia communicate with neurons and with one another about messages traveling among neurons. Glia have the power to alter those signals at the synaptic gaps between neurons and can even influence where synapses are formed.
- Given such prowess, glia may be critical to learning and to forming memories, as well as repairing nerve damage. Experiments are getting under way to find out.

To prove such assertions, scientists first had to show that glia actually do "listen in" on neuronal communication and take action based on what they "hear." Earlier work indicated that an influx of calcium into glial cells could be a sign that they had been stimulated. Based on that notion, investigators devised a laboratory method called calcium imaging to see whether glial cells known as terminal Schwann cells—which surround synapses where nerves meet muscle cells—were sensitive to neuronal signals emitted at these junctions. The method confirmed that Schwann cells, at least, did respond to synaptic firing and that the reaction involved an influx of calcium ions into the cells.

But were glia limited only to eavesdropping on neuronal activity, by scavenging traces of neurotransmitter leaking from a synapse? More general-function Schwann cells also surround axons all along nerves in the body, not just at synapses, and oligodendrocyte glia cells wrap around axons in the central nervous system (brain and spinal cord). At my National Institutes of Health lab, we wanted to know if glia could monitor neural activity anywhere as it flowed through axons in neural circuits. If so, how was that communication mediated? More important, how exactly would glia be affected by what they heard?

To find answers, we cultured sensory neurons (dorsal root ganglion, or DRG, cells) from mice in special lab dishes equipped with electrodes that would enable us to trigger action potentials in the axons. We added Schwann cells to some cultures and oligodendrocytes to others.

We needed to tap independently into the activity of the axons and the glia to determine if the latter were detecting the axon messages. We used a calcium-imaging technique to record visually what the cells were doing, introducing dye that fluoresces if it binds to calcium ions. When an axon fires, voltage-sensitive ion channels in the neuron's membrane open, allowing calcium ions to enter. We would therefore expect to see the firing as a flash of green fluorescence lighting up the entire neuron from the inside. As the concentration of calcium rose in a cell, the fluorescence would get brighter. The intensity could be measured by a photomultiplier tube, and images of the glowing cells could be digitized and displayed in pseudocolor on a monitor in real time—looking something like the radar images of rainstorms shown on weather reports. If glial cells heard the neu-



GLIA AND NEURONS work together in the brain and spinal cord. A neuron sends a message down a long axon and across a synaptic gap to a dendrite on another neuron. Astrocyte glia bring nutrients to neurons as well as surround and regulate synapses. Oligodendrocyte glia produce myelin that insulates axons. When a neuron's electrical message (action potential) reaches the axon terminal (*inset*), the message induces vesicles to move to the membrane and open, releasing neurotransmitters (signaling molecules) that diffuse across a narrow synaptic cleft to the dendrite's receptors. Similar principles apply in the body's peripheral nervous system, where Schwann cells perform myelination duties.

JEFF JOHNSON *Hybrid Medical Animation*

ronal signals and did so in part by taking up calcium from their surroundings, they would light up as well, only later.

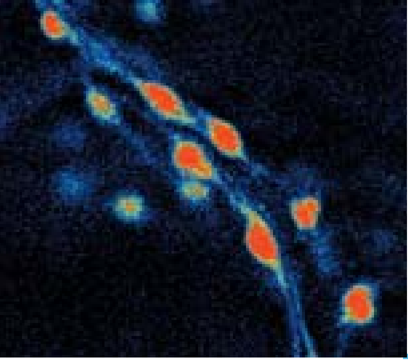
Staring at a computer monitor in a darkened room, my NIH colleague, biologist Beth Stevens, and I knew that after months of preparation our hypothesis was about to be tested with the flick of a switch. When we turned on the stimulator, the DRG neurons responded instantly, changing from blue to green to red and then white on a pseudocolor scale of calcium concentration, as calcium flooded into the axons. Initially, there were no changes in the Schwann cells or oligodendrocytes, but about 15 long seconds later the glia suddenly began to light up like bulbs on a string of Christmas lights [*see illustration on page 59*]. Somehow the cells had detected the impulse activity in the axons and responded by raising the concentration of calcium in their own cytoplasm.

Glia Communicating with Glia

THUS FAR WE HAD confirmed that glia sense axon activity by taking in calcium. In neurons, calcium activates enzymes that produce neurotransmitters. Presumably, the influx in glial cells would also activate enzymes that would marshal a response. But what response was the cell attempting? More fundamentally, what exactly had triggered the calcium influx?

Clues came from previous work on other glial cells in the brain known as astrocytes. One of their functions is to carry nutrients from capillaries to nerve cells; another is to maintain the optimal ionic conditions around neurons necessary for firing impulses. Part of the latter job is to remove excess neurotransmitters and ions that neurons release when they fire. In a classic 1990 study, a group led by Stephen J. Smith of Yale University (now at Stanford University) used calcium imaging to show that the calcium concentration in an astrocyte would rise suddenly when the neurotransmitter glutamate was added to a cell culture. Calcium waves soon spread throughout all the astrocytes in the culture. The astrocytes were responding as if the neurotransmitter had just been released by a neuron, and they were essentially discussing the news of presumed neuronal firing among themselves.

Some neuroscientists wondered whether the communication occurred because calcium ions or related signaling molecules simply passed through open doorways connecting abutting astrocytes. In 1996 S. Ben Kater and his colleagues at the University of Utah defused that suspicion. Using a sharp microelectrode,



Schwann glia could be key to treating nerve diseases such as MS.

they cut a straight line through a layer of astrocytes in culture, forming a cell-free void that would act like a highway separating burning forests on either side. But when they stimulated calcium waves on one side of the break, the waves spread to astrocytes across the void with no difficulty. The astrocytes had to be sending signals through the extracellular medium, rather than through physical contact.

Intensive research in many laboratories over the next few years showed similar results. Calcium responses could be induced in astrocytes by adding neurotransmitters or by using electrodes to stimulate the release of neurotransmitters from synapses. Meanwhile physiologists and biochemists were finding that glia had receptors for many of the same neurotransmitters neurons use for synaptic communication, as well as most of the ion channels that enable neurons to fire action potentials.

ATP Is the Messenger

THESE AND OTHER RESULTS led to confusion. Glial communication is controlled by calcium influxes, just as neuronal communication is. But electrical impulses trigger calcium changes in neurons, and no such impulse exists in or reaches glia. Was glial calcium influx initiated by a different electrical phenomenon or some other mechanism?

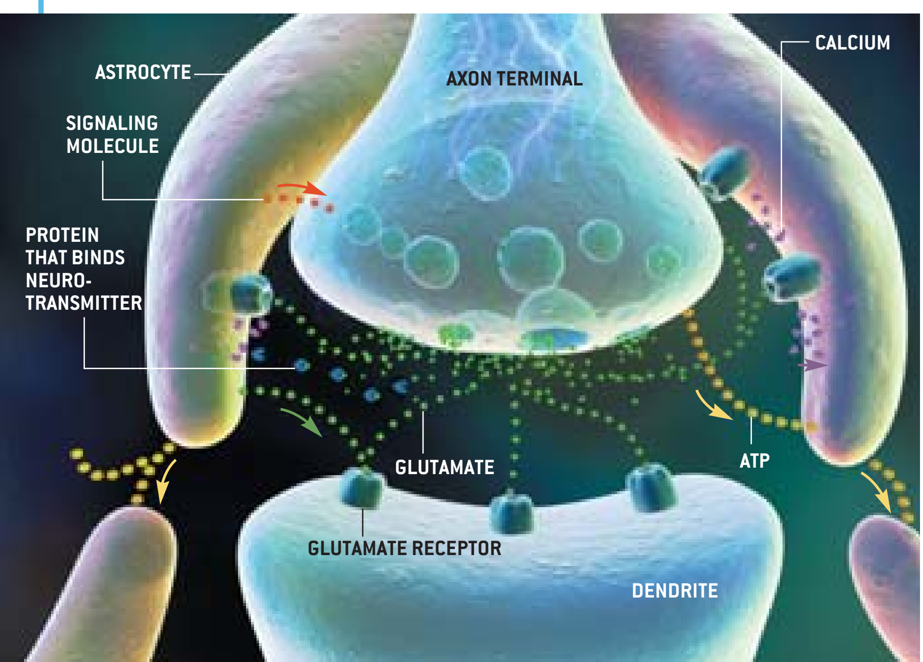
In their glial experiments, researchers were noticing that a familiar molecule kept cropping up—ATP (adenosine triphosphate), known to every biology student as the energy source for cellular activities. Although it makes a great power pack, ATP also has many features that make it an excellent messenger molecule between cells. It is highly abundant inside cells but rare outside of them. It is small and therefore diffuses rapidly, and it breaks down quickly. All these traits ensure that new messages conveyed by ATP molecules are not confused with old messages. Moreover, ATP is neatly packaged inside the tips of axons, where neurotransmitter molecules are stored; it is released together with neurotransmitters at synapses and can travel outside synapses, too.

In 1999 Peter B. Guthrie and his colleagues at the University of Utah showed conclusively that when excited, astrocytes release ATP into their surroundings. The ATP binds to receptors on nearby astrocytes, prompting ion channels to open and allow an influx of calcium. The rise triggers ATP release from those cells, setting off a chain reaction of ATP-mediated calcium responses across the population of astrocytes.

A model of how glia around an axon sense neuronal activity and then communicate to other glia residing at the axon's synapse was coming together. The firing of neurons somehow induces glial cells around an axon to emit ATP, which causes calcium intake in neighboring glia, prompting more ATP release, thereby activating communication along a string of glia that can reach far from the initiating neuron. But how could the glia in our experiment be detecting the neuronal firing, given that the axons made no synaptic connections with the glia and the axonal glia were nowhere near the synapse? Neurotransmitters were not the answer; they do not diffuse out of axons (if they did, they could act in unintended places in the brain, wreaking havoc). Perhaps ATP, which is released along with neuro-

THE AUTHOR

R. DOUGLAS FIELDS is chief of the Nervous System Development and Plasticity Section at the National Institute of Child Health and Human Development and adjunct professor in the Neurosciences and Cognitive Science Program at the University of Maryland. He was a postdoctoral fellow at Yale and Stanford universities. Fields enjoys rock climbing, scuba diving, and building acoustic guitars and Volkswagen engines.



ASTROCYTES REGULATE SIGNALING across synapses in various ways. An axon transmits a signal to a dendrite by releasing a neurotransmitter (*green*)—here, glutamate. It also releases the chemical ATP (*gold*). These compounds then trigger an influx of calcium (*purple*) into astrocytes, which prompts the astrocytes to communicate among themselves by releasing their own ATP. Astrocytes may strengthen the signaling by secreting the same neurotransmitter, or they may weaken the signal by absorbing the neurotransmitter or secreting proteins that bind to it (*blue*), thereby preventing it from reaching its target. Astrocytes can also release signaling molecules (*red*) that cause the axon to increase or decrease the amount of neurotransmitter it releases when it fires again. Modifying the connections among neurons is one way the brain revises its responses to stimuli as it accumulates experience—how it learns. In the peripheral nervous system, Schwann cells surround synapses.

JEFF JOHNSON *Hybrid Medical Animation*

transmitters when axons fire, was somehow escaping along the axon.

To test this notion, we electrically stimulated pure cultures of DRG axons and then analyzed the medium. By exploiting the enzyme that allows fireflies to glow—a reaction that requires ATP—we were able to detect the release of ATP from axons by seeing the medium glow when axons fired. We then added Schwann cells to the culture and measured the calcium responses. They also lit up after axons fired an action potential. Yet when we added the enzyme apyrase, which rapidly destroys ATP—thereby intercepting the ATP before it could reach any Schwann cells—the glia remained dark when axons fired. The calcium response in the Schwann cells had been blocked, because the cells never received the ATP message.

ATP released from an axon was indeed triggering calcium influx into Schwann cells. Using biochemical analysis and digital microscopy, we also showed that the influx caused signals to travel from the cells' membrane to the nucleus, where the genes are stored, causing various genes to switch on. Amazingly, by firing to communicate with other neurons, an axon could instruct the readout of genes in a glial cell and thus influence its behavior.

Axons Control Glia's Fate

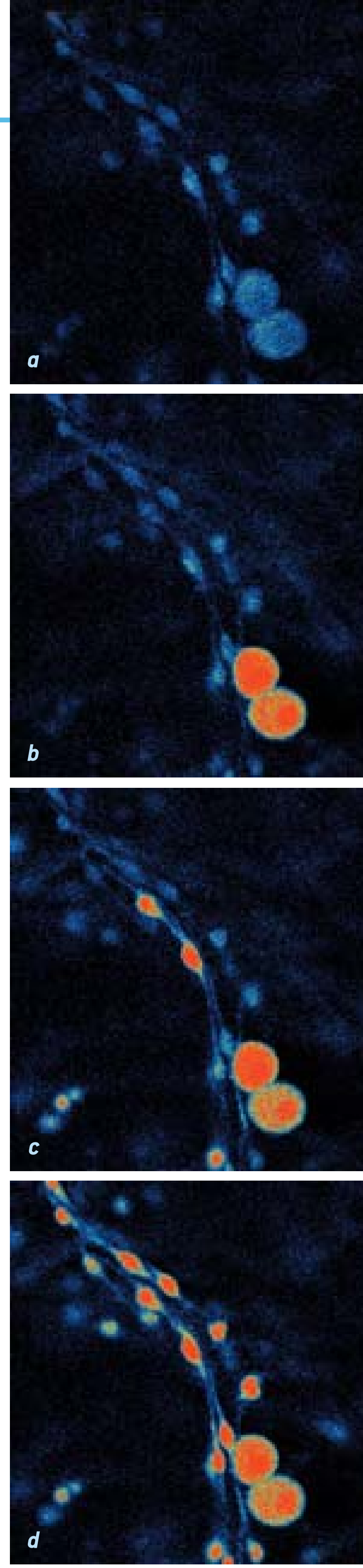
TO THIS POINT, work by us and others had led to the conclusion that a glial cell senses neuronal action potentials by detecting ATP that is either released by a firing axon or leaked from the synapse. The glial cell relays the message inside itself via calcium ions. The ions activate enzymes that release ATP to other glial cells or activate enzymes that control the readout of genes.

This insight made us wonder what functions the genes might be controlling. Were they telling the glia to act in ways that would influence the neurons around them? Stevens set out to answer this question by focusing on the process that prompts production of the myelin insulation around axons, which clearly would affect a neuron. This insulation is key to the conduction of nerve impulses at high speed over long distances. Its growth enables a baby to gradually hold up its head, and its destruction by diseases such as multiple sclerosis causes severe impairment.

We turned to myelin because we were curious about how an immature Schwann cell on an axon in the peripheral nervous system of a fetus or infant knows which axons will need myelin and when to start sheathing those axons and, alternatively, how it knows if it should transform itself into a cell that will not make insulation. Generally, only large-diameter axons need myelin. Could axon impulses or ATP release affect these decisions? We found that Schwann cells in culture proliferated more slowly when gathered around axons that were firing than around axons that were quiet. Moreover, the Schwann cells' development was arrested and myelin formation was blocked. Adding ATP produced the same effects.

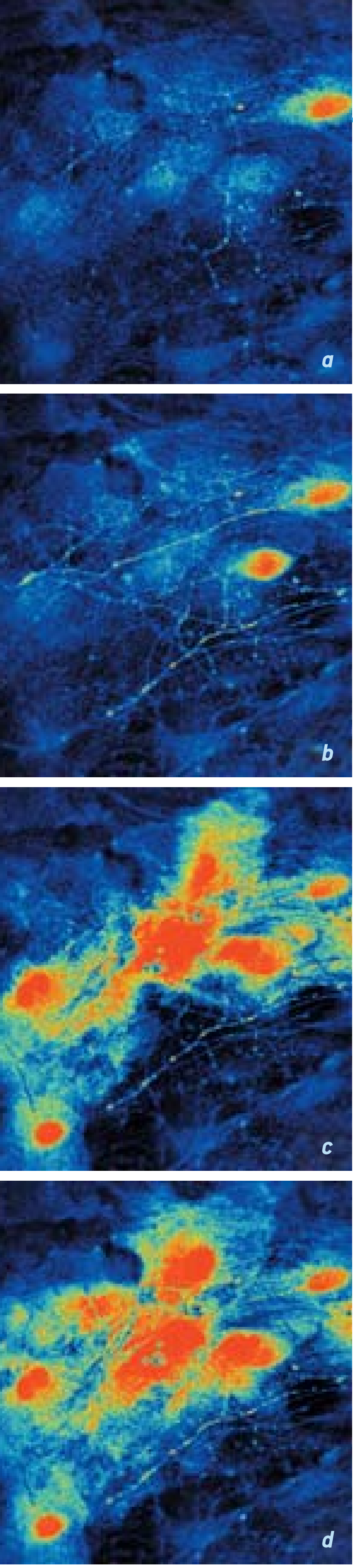
Working with Vittorio Gallo and his colleagues in the adjacent NIH lab, however, we found a contrasting situation with the oligodendrocyte glia that form myelin in the brain. ATP did not inhibit their proliferation, but adenosine, the substance left when phosphate molecules in ATP are removed, stimulated the cells to mature and form myelin. The two findings indicate that different receptors on glia provide a clever way for a neuron to send separate messages to glial cells in the central or peripheral nervous system without having to make separate messenger molecules or specify message destinations.

Better understanding of myelination is important. Every year thousands of people die and countless more are paralyzed or blinded because of demyelinating disease. Multiple sclerosis, for example, strikes one in 700 people. No one knows what exactly initiates myelination, but adenosine is the first substance derived from an axon that has been found to stimulate the process. The fact that adenosine is released from axons in response to axon firing means activity in the brain actually influences myeli-



MOVIE MADE using scanning-laser confocal microscopy (*later colorized*) shows that glial cells respond to chattering neurons. Sensory neurons (*two large bodies,* 20 microns in diameter) (*a*) and Schwann glial cells (*smaller bodies*) were mixed in a lab culture containing calcium ions (*invisible*). Dye that would fluoresce if calcium ions bound to it was introduced into the cells. A slight voltage applied to the neurons prompted them to fire action potentials down axons (*long lines*), and the neurons immediately lit up (*b*), indicating they had opened channels on their membranes to allow calcium to flow inside. Twelve seconds later (*c*), as the neurons continued to fire, Schwann cells began to light up, indicating they had begun taking in calcium in response to the signals traveling down axons. Eighteen seconds after that (*d*), more glia had lit up, because they had sensed the signals. The series shows that glia tap into neuronal messages all along the lines of communication, not just at synapses where neurotransmitters are present.

COURTESY OF R. DOUGLAS FIELDS; SOURCE: DERIVED FROM SUPPLEMENTARY MATERIAL FOR B. STEVENS AND R. D. FIELDS IN *SCIENCE*, VOL. 287, PAGES 2267–2271; MARCH 24, 2000. USED WITH PERMISSION



HOW DO GLIA communicate? Glia called astrocytes (*a*) and sensory neurons (*not shown*) were mixed in a lab culture containing calcium ions. After a neuron was stimulated to fire action potentials down long axons (*lightning bolts*) (*b*), glia began to light up, indicating they sensed the message by beginning to absorb calcium. After 10 and 12.5 seconds (*c* and *d*), huge waves of calcium flux were sweeping across the region, carrying signals among many astrocytes. Green to yellow to red depicts higher calcium concentration.

COURTESY OF R. DOUGLAS FIELDS; SOURCE: DERIVED FROM SUPPORTING ONLINE MATERIAL FOR R. D. FIELDS AND B. STEVENS-GRAHAM IN *SCIENCE*, VOL. 298, PAGES 556–562; OCTOBER 18, 2002. USED WITH PERMISSION

nation. Such findings could mark paths to treatment. Drugs resembling adenosine might help. Adding adenosine to stem cells could perhaps turn them into myelinating glia that are transplanted into damaged nerves.

Outside the Neuronal Box

EXPERIMENTS IN OUR LAB and others strongly suggest that ATP and adenosine mediate the messages coursing through networks of Schwann and oligodendrocyte glia cells and that calcium messages are induced in astrocyte glia cells by ATP alone. But do glia have the power to regulate the functioning of neurons, other than by producing myelin?

The answer appears to be "yes." Richard Robitaille of the University of Montreal saw the voltage produced by synapses on frog muscle become stronger or weaker depending on what chemicals he injected into Schwann cells at the synapse. When Eric A. Newman of the University of Minnesota touched the retina of a rat, waves of calcium sent by glia changed the visual neurons' rate of firing. Studying slices of rat brain taken from the hippocampus—a region involved in memory—Maiken Nedergaard of New York Medical College observed synapses increase their electrical activity when adjacent astrocytes stimulated calcium waves. Such changes in synaptic strength are thought to be the fundamental means by which the nervous system changes its response through experience—a concept termed plasticity, suggesting that glia might play a role in the cellular basis of learning.

One problem arises from these observations. Like a wave of cheering fans sweeping across a stadium, the calcium waves spread throughout the entire population of astrocytes. This large-scale response is effective for managing the entire group, but it cannot convey a very complex message. The equivalent of "Go team!" might be useful in coordinating general activity in the brain during the sleep-wake cycle or during a seizure, but local conversations are necessary if glial cells are to be involved in the intricacies of information processing.

In a footnote to their 1990 paper, Smith and his colleagues stated that they believed neurons and glia carried on more discrete conversations. Still, the researchers lacked experimental methods precise enough to deliver a neurotransmitter in a way that resembled what an astrocyte would realistically experience at a synapse. In 2003 Philip G. Haydon of the University of Pennsylvania achieved this objective. He used improved laser technology to release such a small quantity of glutamate in a hippocampal brain slice that it would be detected by only a single astrocyte. Under this condition, Haydon observed that an astrocyte sent specific calcium signals to just a small number of nearby astrocytes. As Haydon put it, in addition to calcium waves that affect astrocytes globally, "there is short-range connectivity between astrocytes."

In other words, discrete astrocyte circuits in the brain coordinate activity with neuronal circuits. (The physical or biochemical factors that define these discrete astrocytic circuits are unknown at present.) Investigation by others has also indicated that astrocytes may strengthen signaling at synapses by secreting the same neurotransmitter the axon is releasing—in effect, amplifying the signal.

The working hypothesis that Haydon and I, along with our colleagues, are reaching from these discoveries is that communication among astrocytes helps to activate neurons whose axons terminate relatively far away and that this activity, in turn, contributes to the release of neurotransmitters at distant synapses. This action would regulate how susceptible remote synapses are to undergoing a change in strength, which is the cellular mechanism underlying learning and memory.

Results announced at the Society for Neuroscience's annual meeting in November 2003 support this notion and possibly expand the role of glia to include participation in the formation of new synapses [*see box on opposite page*]. Some of the findings build on research done two years earlier by Ben A. Barres, Frank W. Pfrieger and their colleagues at Stanford, who reported that rat neurons grown in culture made more synapses when in the presence of astrocytes.

Working in Barres's lab, postdoctoral students Karen S. Christopherson and Erik M. Ullian have subsequently found that a protein called thrombospondin, presumably from the astrocyte, was the chemical messenger that spurred synapse building. Thrombospondin plays various biological roles but was not thought to be a major factor in the nervous system. The more thrombospondin they added to the astrocyte culture, though, the more synapses appeared. Thrombospondin may be responsible for bringing together proteins and other compounds needed to create a synapse when young nerve networks grow and therefore might contribute to the modification of synapses as the networks age.

Future experiments could advance our emerging understanding of how glia affect our brains. One challenge would be to show that memory—or a cellular analogue of memory, such as long-term potentiation—is affected by synaptic astrocytes. Another challenge would be to determine precisely how remote synapses might be influenced by signals sent through astrocyte circuits.

Perhaps it should not be surprising that astrocytes can affect synapse formation at a distance. To form associations between stimuli that are processed by different circuits of neurons—the smell of a certain perfume, say, and the emotions it stirs about the person who wears it—the brain must have ways to establish fast communication between neuronal circuits that are not wired together directly. If neurons are like telephones communicating electrically through hardwired synaptic connections, astrocytes may be like cell phones, communicating with chemical signals that are broadcast widely but can be detected only by other astrocytes that have the appropriate receptors tuned to receive the message. If signals can travel extensively through astrocyte circuits, then glia at one site could activate distant glia to coordinate the firing of neural networks across regions of the brain.

Comparisons of brains reveal that the proportion of glia to neurons increases greatly as animals move up the evolutionary ladder. Haydon wonders whether extensive connectivity among astrocytes might contribute to a greater capacity for learning. He and others are investigating this hypothesis in new experiments. Perhaps a higher concentration of glia, or a more potent type of glia, is what elevates certain humans to genius. Einstein taught us the value of daring to think outside the box. Neuroscientists looking beyond neurons to see how glia may be involved in information processing are following that lead. SA

GLIA CONTROL SYNAPSES

FOR YEARS, scientists assumed that only neurons specify the connections they make to other neurons. But evidence shows that glia can strongly influence how many synapses a neuron forms and where it forms them.

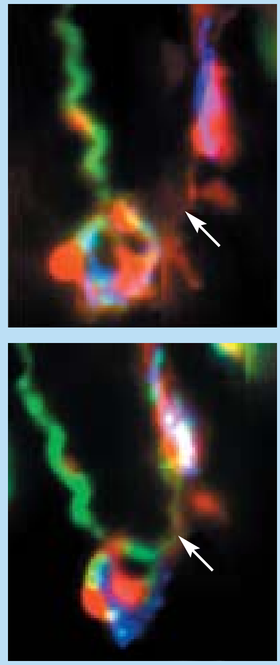
Ben A. Barres and his colleagues at Stanford University found that when they grew neurons from a rat's retina in a lab culture devoid of glial cells known as astrocytes, the neurons created very few synapses. When the researchers added astrocytes or culture medium that had been in contact with astrocytes, synapses formed abundantly. Barres could see the synapses and count them through a microscope as well as detect them by recording electrical activity (a sign that signals were flowing through synapses) with a microelectrode. He then detected in the medium two chemicals that are released by astrocytes to stimulate synapse formation—a fatty complex called apoE/cholesterol and the protein thrombospondin.

Meanwhile Jeff W. Lichtman's group at Washington University recorded muscle synapses in mice over several days or weeks as they formed and as they were removed during development (the time when unneeded synapses get pruned) or after injury. When the images were spliced into a time-lapse movie, it appeared that both synapse formation and elimination were influenced by nonneuronal cells, seen as ghostlike forces acting on the axon terminal.

Most recently, Le Tian, Wesley Thompson and their associates at the University of Texas at Austin experimented with a mouse that had been engineered so that its Schwann glia cells fluoresced. This trait allowed Thompson's team to collaborate with Lichtman's group and watch glial cells operate at the junction where neurons meet muscle—a feat previously not possible. After a muscle axon is injured or cut, it withdraws, but a cluster of neurotransmitter receptors remains on the recipient side of a synapse. Investigators knew that an axon can regenerate and find its way back to the abandoned receptors by following the Schwann cells that remain.

But what happens if the axon cannot find its way? Tracking the fluorescence, Thompson's group saw that Schwann cells at intact synapses somehow sensed that a neighboring synapse was in trouble. Mysteriously, the Schwann cells sprouted branches that extended to the damaged synapse, forming a bridge along which the axon could grow a new projection to the receptors (*photographs*).

This work clearly shows that glia help to determine where synaptic connections form. Researchers are now trying to exploit this power to treat spinal cord injuries by transplanting Schwann cells into damaged spinal regions in lab animals. *—R.D.F.*



GLIA CAN GUIDE the formation of synapses. Neurobiologist Le Tian severed a muscle nerve synapse in a mouse whose cells had been engineered to fluoresce. Two days later (*top*) Schwann glia cells (*dark red*) had formed a bridge across the divide (*arrow*). In another two days (*bottom*), an axon (*green*) had regrown along the bridge to create a synapse.

LE TIAN AND WESLEY THOMPSON

MORE TO EXPLORE

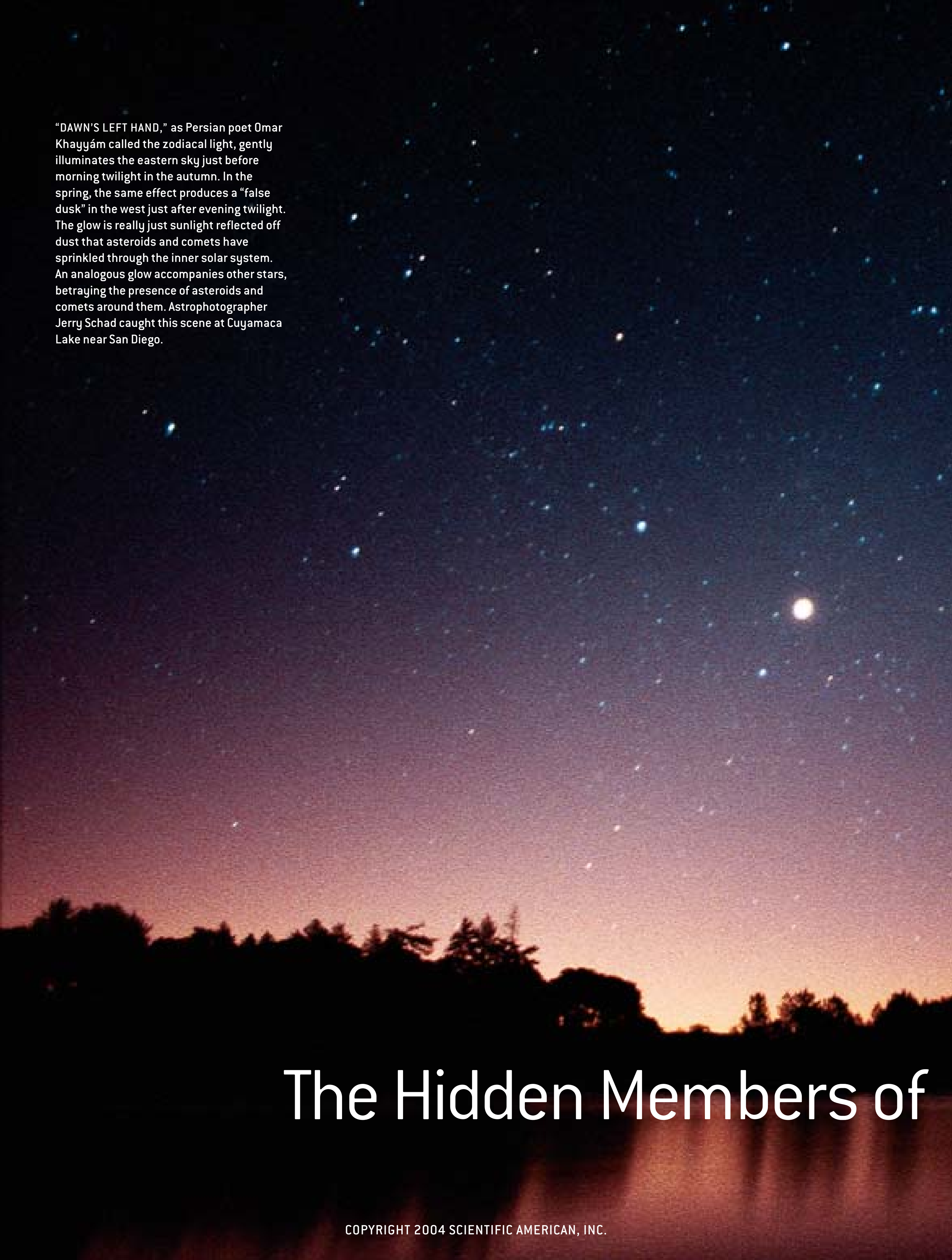
Driving Mr. Albert: A Trip across America with Einstein's Brain. Michael Paterniti. Delta, 2001.

New Insights into Neuron-Glia Communication. R. D. Fields and B. Stevens-Graham in *Science*, Vol. 298, pages 556–562; October 18, 2002.

Adenosine: A Neuron-Glial Transmitter Promoting Myelination in the CNS in Response to Action Potentials. B. Stevens, S. Porta, L. L. Haak, V. Gallo, and R. D. Fields in *Neuron,* Vol. 36, No. 5, pages 855–868; December 5, 2002.

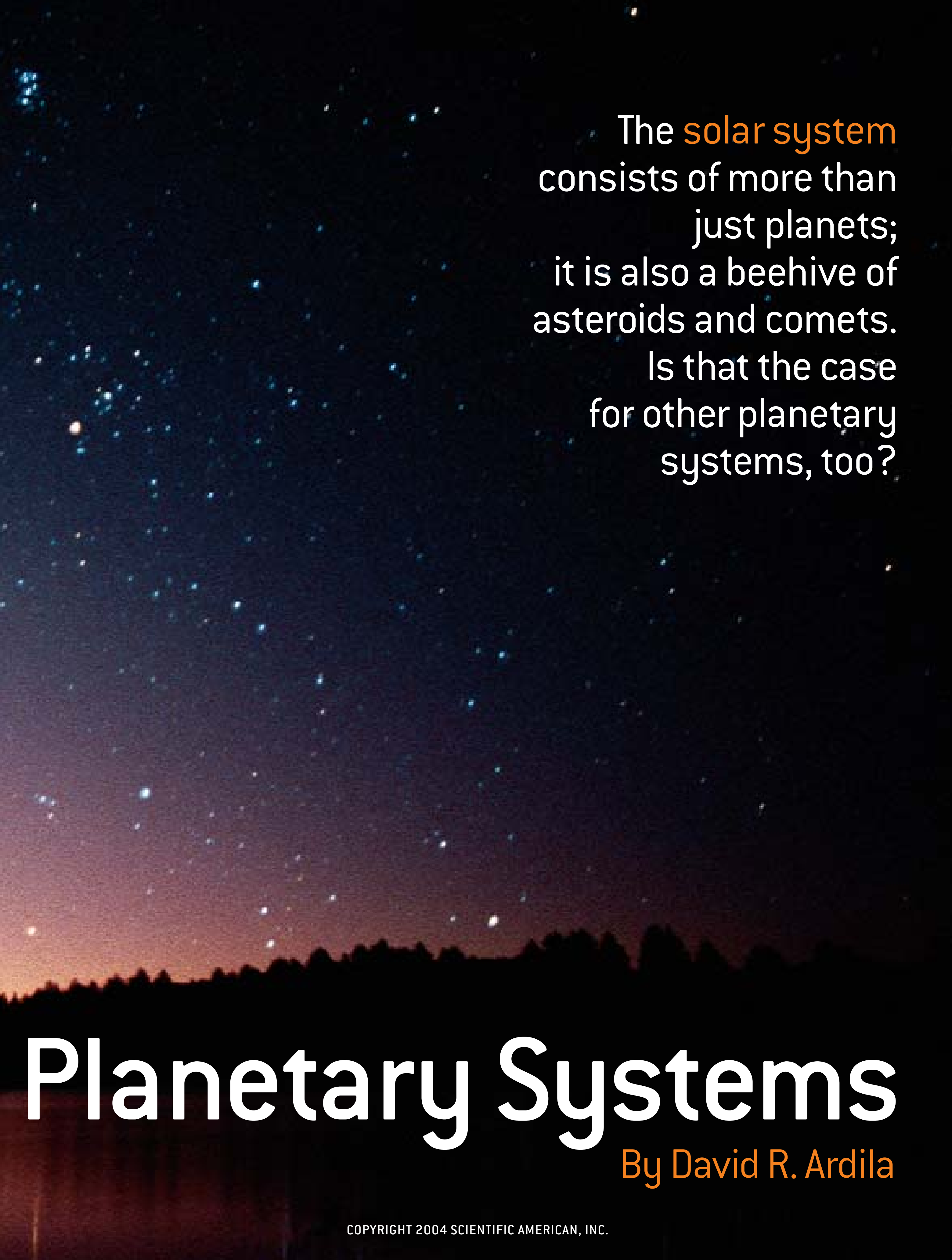
Astrocytic Connectivity in the Hippocampus. Jai-Yoon Sul, George Orosz, Richard S. Givens, and Philip G. Haydon in *Neuron Glia Biology,* Vol. 1, pages 3–11; 2004.

Also see the journal *Neuron Glia Biology:* **www.journals.cambridge.org/jid_NGB**



“DAWN’S LEFT HAND,” as Persian poet Omar Khayyám called the zodiacal light, gently illuminates the eastern sky just before morning twilight in the autumn. In the spring, the same effect produces a “false dusk” in the west just after evening twilight. The glow is really just sunlight reflected off dust that asteroids and comets have sprinkled through the inner solar system. An analogous glow accompanies other stars, betraying the presence of asteroids and comets around them. Astrophotographer Jerry Schad caught this scene at Cuyamaca Lake near San Diego.

The Hidden Members of



The solar system consists of more than just planets; it is also a beehive of asteroids and comets. Is that the case for other planetary systems, too?

Planetary Systems

By David R. Ardila

Does our solar system represent the rule or the exception?

Do similar collections of worlds surround other stars in the galaxy, or is the sun peculiar? Although this is one of the fundamental questions driving modern astronomy, the answer remains elusive. Over the past nine years, astronomers have discovered at least 111 planets around sunlike stars by looking for the slight back-and-forth motion that these bodies impart to their parent suns. Yet this technique detects only the most massive and tightly orbiting objects. If extraterrestrial astronomers applied the same method to our solar system, they might manage to identify Jupiter, and maybe Saturn, but they would completely miss the smaller bodies that make the sun's family so rich and varied: asteroids, comets and the terrestrial planets.

How can astronomers detect those smaller bodies and paint a more complete picture of the diversity of planetary systems? A clue appears in the western sky in the spring, right after sunset. If you watch closely, you might see the zodiacal light, a faint triangle of light extending up from the horizon. The zodiacal light is produced by sunlight bouncing off interplanetary dust particles in our solar system. The triangle of light stretches along the sun's path in the sky, indicating that the dust forms a disk in the plane of Earth's orbit. What makes the dust interesting is that it should not be there. The individual dust particles are so small—about 20 to 200 microns across, judging from the color of the zodiacal light—that sunlight quickly causes them to spiral into the sun and burn up. Dust particles that are even smaller are quickly blown away from the solar system by radiation pressure. Therefore, for dust to be present it must be replenished continuously.

Astronomers believe that these dust particles result from collisions among asteroids and the evaporation of comets in their travels close to the sun. In the main asteroid belt, located between the orbits of Mars and Jupiter, collisions are common. An impact throws off dust, and it may crack the two bodies into shards that continue to grind against one another for millions of years afterward, littering space with still more dust. In the case of comets, sunlight boils off the dirty ice at their surface, creating spectacular tails that disappear in a trail of dust—dust that we then see as meteor showers.

Once released, the debris dust spreads through much of the inner solar system, out to the orbit of Jupiter. Its total mass at any given time amounts to just a thousandth of the mass of the moon. But because of its huge surface area, the dust outshines the planets by a factor of 100. Extraterrestrial observers looking in our direction would see it before they saw Jupiter or Earth.

The same process occurs around other stars. Twenty years ago the Infrared Astronomical Satellite (IRAS), expecting to obtain routine calibration observations of the star Vega, discovered evidence of a debris disk around it. By the early 1990s further analysis of IRAS data had suggested the presence of debris disks around 100 or so stars. Most of the disks, however, could not be seen directly; their presence had to be deduced indirectly. Only since the late 1990s have ground-based and orbiting observatories provided astronomers with detailed images of a handful of disks. The latest contributions have come from the Advanced Camera for Surveys (ACS), an instrument installed on the Hubble Space Telescope in 2002, and the Spitzer Space Telescope, the infrared counterpart to Hubble, launched in August 2003.

What the recent images show is wonderfully unexpected. Far from appearing featureless, some disks look like gigantic versions of the rings of Saturn, and some contain large blobs, holes and spirals. Some of these features may be caused by unseen giant planets. After all, the presence of a debris disk implies the existence of asteroids or comets, which are by-products of the planet formation process—either fragments of larger objects that were destroyed by collisions (in the case of most asteroids) or "planetesimals," the building blocks of planets, some of which never coalesced to make larger bodies (in the case of comets). In our solar system, asteroids and comets coexist with rocky and giant planets, and perhaps the same is true for other planetary systems as well.

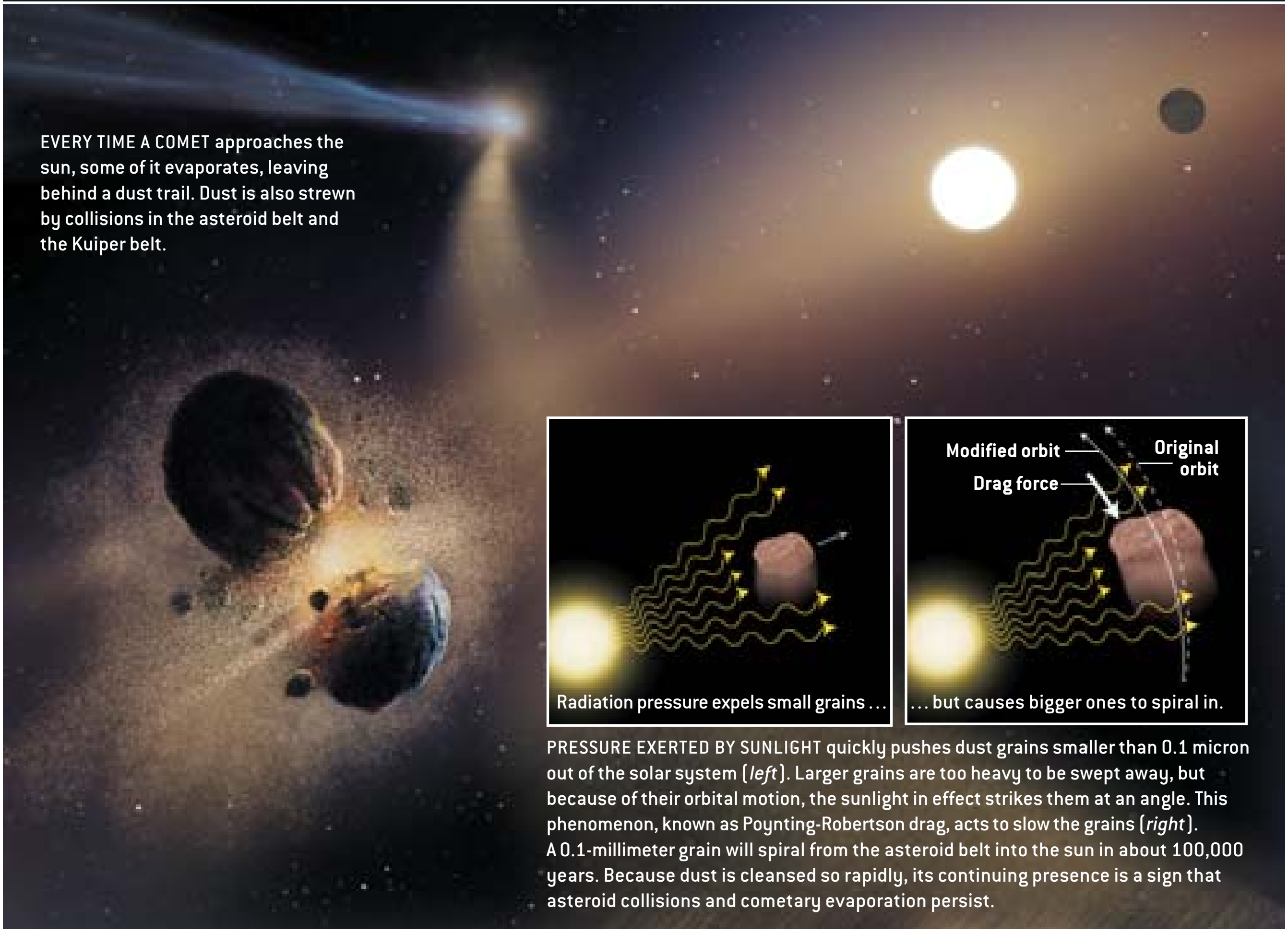
Traditionally, the study of planet formation has proved frustrating: with only the solar system as a close example, astronomers are not sure whether their theories apply to other planetary systems. The observations of debris disks around stars of different masses and ages are helping to place our solar system in context.

Overview/*Circumstellar Debris Disks*

- One of the great advances of astronomy over the past decade has been the discovery of planets outside our solar system—the first tangible clue that we may not be alone in the universe.
- But this discovery is just a small piece of the puzzle. Planet finders generally see the very biggest worlds; they miss the panoply of bodies, such as asteroids and comets, that makes our solar system a rich and diverse place. Not only are these objects the most numerous of the sun's progeny, they are important as leftover building blocks of larger worlds.
- To get at those smaller bodies, astronomers look for the dust scattered when objects collide. Observers have found more than 100 stars with dusty debris disks and obtained images of about a dozen of them. The study of the disks suggests that planet formation occurs elsewhere in much the same way that it took place in our solar system.

JERRY SCHAD (*preceding pages*)

WHAT THE DUST REVEALS



EVERY TIME A COMET approaches the sun, some of it evaporates, leaving behind a dust trail. Dust is also strewn by collisions in the asteroid belt and the Kuiper belt.

PRESSURE EXERTED BY SUNLIGHT quickly pushes dust grains smaller than 0.1 micron out of the solar system (*left*). Larger grains are too heavy to be swept away, but because of their orbital motion, the sunlight in effect strikes them at an angle. This phenomenon, known as Poynting-Robertson drag, acts to slow the grains (*right*). A 0.1-millimeter grain will spiral from the asteroid belt into the sun in about 100,000 years. Because dust is cleansed so rapidly, its continuing presence is a sign that asteroid collisions and cometary evaporation persist.

Debris Everywhere

IRAS WAS ONE OF the most productive satellites in the history of astronomy. Though functional for only 10 months in 1983, the observations it performed continue to be a major source of information for astronomers. The satellite undertook a complete survey of the sky in mid- and far-infrared light, with wavelengths between 12 and 100 microns. This part of the spectrum is difficult or impossible to detect from the ground because Earth's atmosphere blocks most of it. For matter to emit light mainly in the far infrared, it has to be relatively cold, from 50 to 100 kelvins. Astronomers expected that normal stars, having surface temperatures of thousands of kelvins, would be almost invisible to IRAS.

Yet the satellite discovered that some stars shine brightly at these wavelengths and may emit tens or even hundreds of times as much infrared light as normal stars. The infrared excess suggests the presence of dust around the star. The idea is that starlight heats the material, which then emits infrared light, producing a bump in the stellar spectrum [*see illustration on page 67*]. These stars are old enough that the dust cannot be a leftover of their formation; it has to be transient, as in our solar system, and so it must be coming from the collision or evaporation of unseen bodies.

IRAS did not have sufficiently high resolution to see most of the disks directly. In the images, all but four of them appear as featureless points. But the disks' brightness provides a crude way to estimate their size: larger disks produce greater infrared excesses. The disks thus appear to range in radius from 100 to 1,000 astronomical units, or between 20 and 200 times the distance from the sun to Jupiter. Analysis of the spectra indicates that the composition of the dust resembles that of solar system comets.

The spectrum also reveals the geometry of the dust. The disks observed by IRAS span a range of temperatures: the inner parts, being close to the star, are warmer than the zone farther out. Inter-

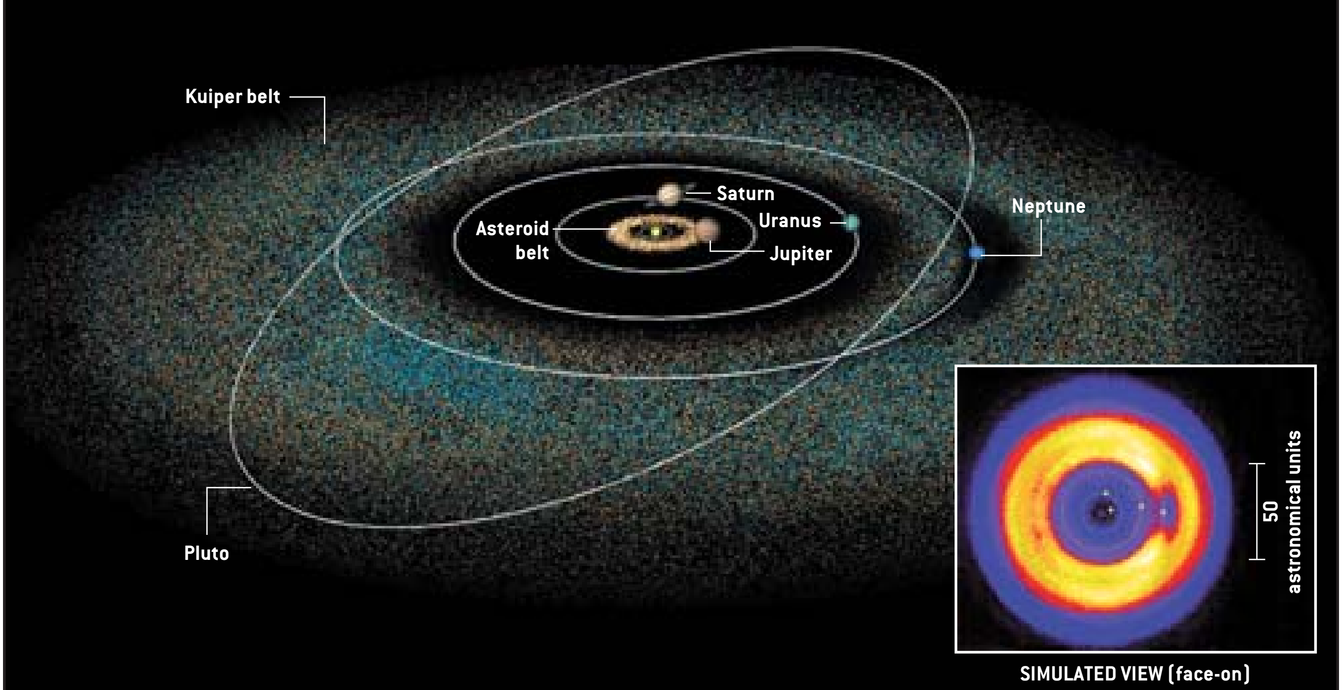
THE AUTHOR

DAVID R. ARDILA grew up in Colombia, where he pursued physics at the University of Los Andes in Bogotá. He earned his Ph.D. from the University of California, Berkeley, and in 2002 joined the Advanced Camera for Surveys Science Team at Johns Hopkins University. Besides debris disks, he studies protostellar disks, planet formation and brown dwarfs. He resides in Baltimore with his wife, Debi, and their four-year-old son, Alejandro. Having never lived before in a place with four seasons, the whole family is still shocked by them.

DON DIXON

SCULPTED BY PLANETS

Not only does dust indicate the existence of asteroids and comets, it can reveal planets. In our solar system the gravity of the giant planets is thought to create patterns in the Kuiper belt dust. An outside viewer would see gaps and clumps in an infrared image (*inset*) and deduce the presence of the giant planet. This simulated view is similar to what astronomers have seen around other stars.



SIMULATED VIEW (face-on)

estingly, most of the disks do not seem to have dust much hotter than 200 kelvins, which is cooler than would be expected if the disks extended all the way down to their stars. Thus, the disks appear to have inner holes. The intimation of these holes offered astronomers the first sign that disks had a structure—one that might be caused by hidden planets.

In 1984, following up on the IRAS findings, Bradford A. Smith, then at the University of Arizona, and Richard J. Terrile of the Jet Propulsion Laboratory in Pasadena, Calif., observed one of the stars, Beta Pictoris, using the 2.5-meter telescope at Las Campanas Observatory in Chile. To detect the faint disk material against the bright stellar glare, they used a coronagraph, a small mask put on the telescope to block the direct starlight. The visible-light image revealed a magnificent edge-on disk extending more than 400 astronomical units away from the star. More recent observations have put the disk's radius at more than 1,400 astronomical units.

Beta Pictoris is a special case. It is relatively near to us (only 60 light-years away), and its disk is very large, very bright, and edge-on (which increases its apparent brightness). This combination of factors made it relatively easy to see. Unfortunately, astronomers were unable to use coronagraphs to observe other disks. As seen from the ground in visible light, stars have a certain apparent size in the sky, determined by the blurring that Earth's atmosphere produces in the starlight. A coronagraphic mask large enough to block the star usually ends up hiding the disk as well. At longer wavelengths, such as the far infrared, stars are faint, so the disks should be easier to see. But the atmosphere absorbs light at these wavelengths. Light with even longer wavelengths, close to a millimeter, can be detected from the ground, but until the late 1990s the instruments capable of detecting this "submillimeter" light had low resolution and low sensitivity. Thirteen years passed before instrument technology advanced enough to image other disks.

The breakthrough came with the development of the Submillimeter Common-User Bolometer Array (SCUBA), a very sensitive camera capable of detecting light with wavelengths around one millimeter. In 1997 a group led by Wayne S. Holland and Jane S. Greaves, then at the Joint Astronomy Center in Hawaii, used SCUBA, mounted on the James Clerk Maxwell Telescope on Mauna Kea, to take images of several IRAS stars. These and other images confirmed the existence of disks around stars besides Beta Pictoris. Since then, about a dozen disks have been resolved by SCUBA, ground-based mid-infrared detectors and the Hubble Space Telescope. In some cases, however, the far-infrared excess resulted from background objects or nearby interstellar clouds unrelated to the target star.

Comet Swarms

MOST OF THE DEBRIS disks around other stars are much cooler and larger than the zodiacal disk. A partial explanation of this difference emerged a decade ago from a parallel line of work. In the early 1990s astronomers confirmed the existence of the Kuiper belt, a long-hypothesized band of icy bodies that extends from the orbit of Neptune past that of Pluto [see "The Kuiper Belt," by Jane X. Luu and David C. Jewitt; Scientific American, May 1996]. Collisions in this belt should create a second debris disk—one that is comparatively cool. The disk is difficult to see from Earth, immersed as we are in the warm, bright glow of the zodiacal light.

The Kuiper debris disk, rather than the zodiacal debris disk, appears to be the analogue of the debris disks around most of the stars. In some cases, astronomers have detected a smaller, warmer disk—analogous to the zodiacal debris disk—in addition to the large cool disk.

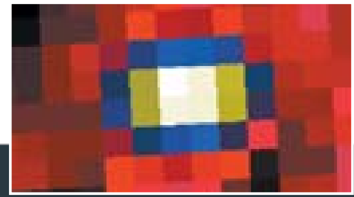
Even though our Kuiper debris disk extends farther from the sun than the zodiacal disk and contains perhaps 10 times more dust, it is still much smaller than disks elsewhere. The Beta Pictoris disk has at least 10,000 times as much dust as our solar system does. The rate at which the planetesimals collide and strew dust depends on the square of the number of objects, and so, other things being equal, Beta Pictoris must have 100 times as many planetesimals as the sun.

Astronomers believe that the abundance of dust is related to the system's youth: the sun is 4.5 billion years old, but Beta Pictoris is only 15 million years old. In fact, observations suggest that the amount of dust decreases with time [*see illustration on page 69*], probably because the parent population of planetesimals erodes. The very impacts that produce the dust also destroy the bodies. Furthermore, gravitational interactions with planets can either eject asteroids and comets or cast them into the central star. The spectrum of Beta Pictoris shows absorption lines that appear and disappear, and astronomers have concluded that they are the result of comets falling into the star and burning up. As many as 200 comets a year could meet this fate.

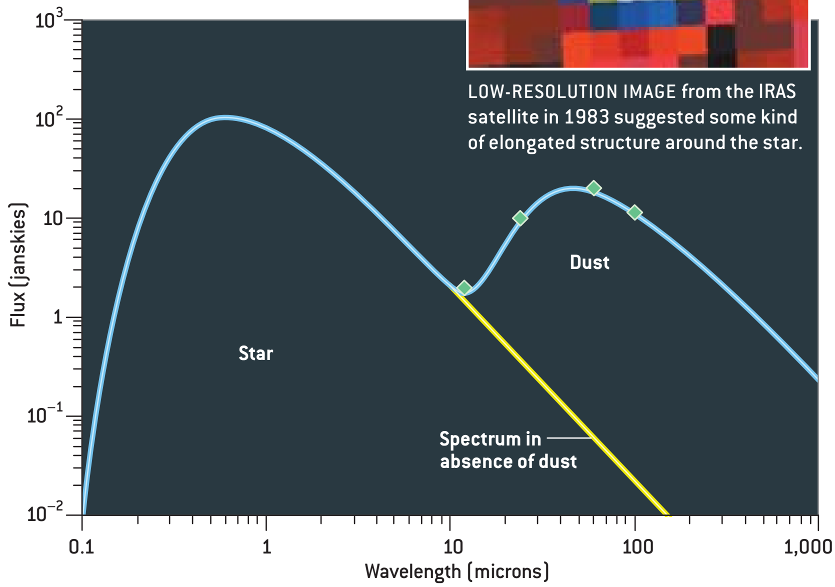
When the solar system was young, it, too, must have been thick with asteroids and comets. Just as it gradually thinned out, the disks around other stars could be

BETA PICTORIS

The dust disk around the star Beta Pictoris, which is located about 63 light-years from Earth, is the best-studied one outside our solar system.

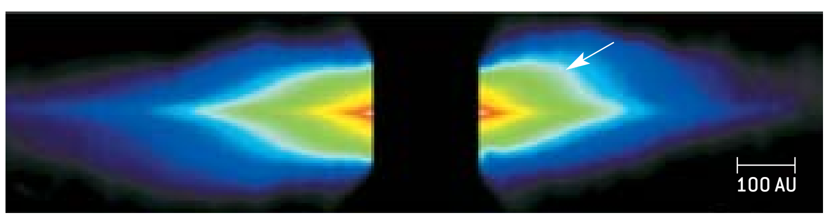


LOW-RESOLUTION IMAGE from the IRAS satellite in 1983 suggested some kind of elongated structure around the star.

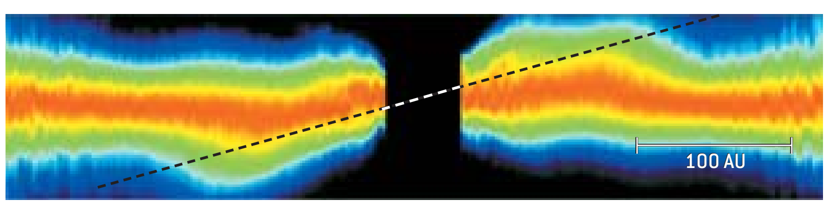


SPECTRAL ENERGY DISTRIBUTION of Beta Pictoris deviates from that of a star. The excess energy at infrared wavelengths is a telltale sign of dust. The diamonds are IRAS measurements.

MORE DETAILED IMAGE made in visible light in 1984 showed a disk edge-on. The dark straight lines and circular rings are artifacts of the instrument used to hide the star.



HUBBLE SPACE TELESCOPE visible-light image in 1995 revealed a flaring of the disk (*arrow*), perhaps caused by a brown dwarf or passing star. Color represents brightness.



HUBBLE IMAGE in 1997 found a warp (*dotted line*) in the inner region of the disk. This pattern, along with other features, suggests the presence of a giant planet on an inclined orbit.

SKYVIEW IMAGE ARCHIVE (skyview.gsfc.nasa.gov) (*top*); LUCY READING; SOURCE: DAVID R. ARDILA (*graph*); RICHARD J. TERRILE *Jet Propulsion Laboratory* AND BRADFORD A. SMITH (*top middle*); AL SCHULTZ *Computer Sciences Corporation/Space Telescope Science Institute* AND NASA (*bottom middle*); SALLY HEAP *NASA Goddard Space Flight Center* (*bottom*)

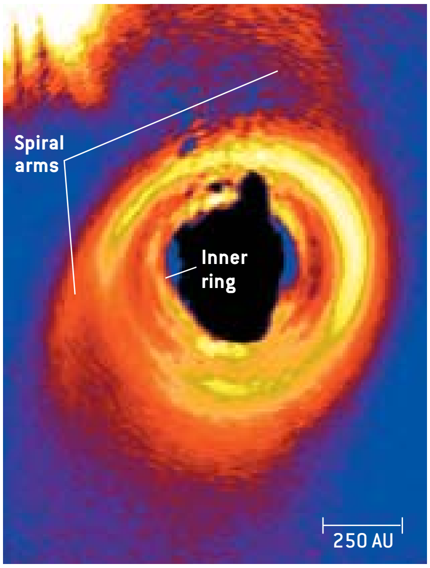
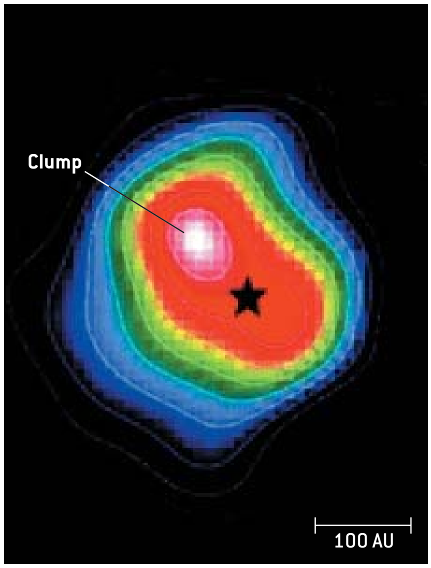
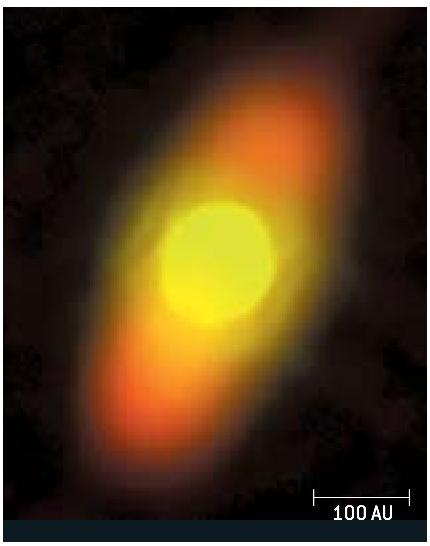


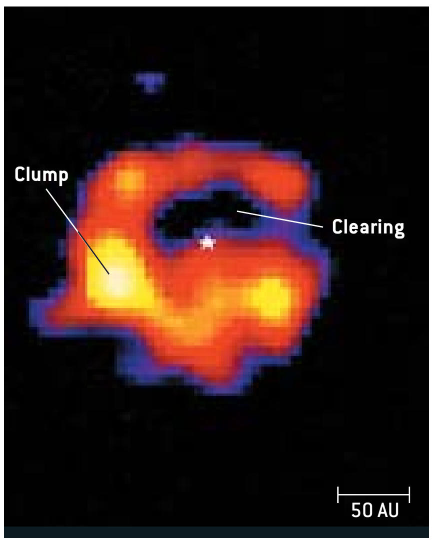
IMAGE OF DUST around the star HD 141569, 330 light-years away, shows two large spiral arms. They could be caused by the double stellar companion at the upper left corner. An inner ring may imply an unseen planet. Color represents the density of material seen in visible light.



VEGA SYSTEM, 25 light-years away, has a bright clump at the top left, which simulations attribute to a planet with twice the mass of Jupiter. The black star marks the position of the central star. This image was taken in submillimeter light. Like HD 141569, Vega has twice the mass of the sun.



FOMALHAUT SYSTEM, 25 light-years away, is larger on the lower left side than on the upper right side, perhaps because of a recent asteroid collision. The center of the ring is filled with warm dust similar to the zodiacal dust in our solar system. This image was taken in far-infrared light.



EPSILON ERIDANI SYSTEM, 10 light-years away, has clumps and clearings, the possible handiwork of a Saturn-mass planet on an elongated orbit. Independent but controversial spectral data suggest another planet closer in. This image was taken in submillimeter light.

evolving toward systems that look rather like our own.

Although age could account for some of the differences between these systems and our solar system, astronomers must be cautious in interpreting their observations. Most of the stars with an infrared excess are more massive than the sun. That may be a selection effect: massive stars are hotter, and so they tend to heat the dust more, making it easier to see. More massive stars may also start with a larger, dustier disk. It remains unclear to what extent the conclusions drawn from these systems can be applied more generally. For example, hotter stars might dissipate their original planet-forming disks faster, thereby affecting the rate at which planets form or planetesimals evolve.

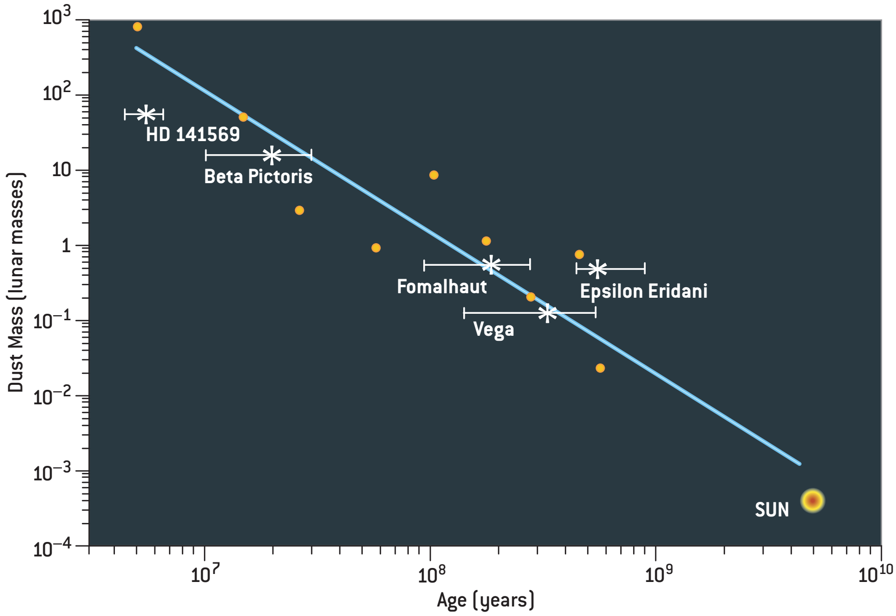
Unseen Planets

DO THESE SYSTEMS have planets as well as planetesimals? Even the youngest ones are old enough that giant planets would have had time to form. The disks contain very little gas, suggesting that they no longer have the ability to form giant planets (which are mostly gaseous). Either the systems have already created such planets or they are never going to. Figuring out which is the case will not be easy. The samples of known planets and known disks do not overlap: to this day, astronomers have been unable to confirm the presence of a debris disk around any star (other than the sun) known to have planets. Planets in debris disks, and debris disks in other planetary systems, have so far escaped detection.

The disks themselves, however, might betray the presence of planets. The rain of comets onto Beta Pictoris is hard to understand unless a planet is present to exert a gravitational influence. Furthermore, on the handful of resolved disk images, astronomers see large-scale features: rings, warps, blobs and, in one case, a large spiral [*see illustrations at left*]. A planet in an inclined orbit can pull dust into inclined orbits, giving the disk a warp. Planets can also sweep away dust, creating cavities and rings, or they can produce a wake of dust that looks like a blob; our own Earth leaves such a wake in the zodiacal dust.

On the other hand, the case for planets hiding in these disks is not as firmly established as astronomers would like. The planets required to produce the observed features would have to be far from the central star—farther than Neptune is from our sun and perhaps too far for planets ever to have arisen (the farther the planet, the longer it takes to form). It may be possible that some planets formed closer to their central stars and then moved outward, as has been suggested for Neptune itself. To conserve angular momentum, such outward motion requires another large planet, like Jupiter, to move inward. In general, however, evidence for this second planet has not emerged. The

NASA, MARK CLAMPIN *Space Telescope Science Institute* AND COLLEAGUES, ACS SCIENCE TEAM AND EUROPEAN SPACE AGENCY (*top left*); WAYNE S. HOLLAND *Joint Astronomy Center* AND COLLEAGUES (*top right*); KARL R. STAPELFELDT *Jet Propulsion Laboratory* AND NASA/JPL-CALTECH (*bottom left*); JANE GREAVES *Joint Astronomy Center* AND COLLEAGUES (*bottom right*)



AS STARS GET OLDER, their dust goes away—presumably because the asteroids and comets, from which the dust comes, are gradually destroyed. The age relationship is not a strict one, but it does suggest that all systems evolve in basically the same way. The data were taken by the Infrared Space Observatory. The orange dots represent star clusters. The point labeled "Sun" represents only the zodiacal dust; the amount of dust in the Kuiper belt is unknown but may be 10 times as great.

data are ambiguous: different groups of researchers argue for planets of different sizes in different places.

Various processes besides the gravitational pull of planets could be responsible for some of these features. A number of astronomers argue that all young planetary systems have rings; as planetesimals grow and coalesce into planets, they disrupt the disks, increasing the occurrence of collisions and the rate of dust production. Others argue that rings of dust can spontaneously form at the edges of gaseous disks; the rapid change in gas pressure at the edge slows down dust particles that would otherwise be expelled from the planetary system.

Understanding these features has become a major goal for astronomers. Last year my colleagues and I observed the star HD 141569 using the ACS coronagraph on board the Hubble Space Telescope. Earlier images had shown two rings around this star. Our images revealed long spiral arms of dust, not unlike those seen in spiral galaxies, and suggested that the rings observed earlier are actually fragments of spirals. HD 141569 has two companion stars, and we believe that less than 100,000 years ago the companions passed close to the disk, disrupting and stretching it. This interaction could have created the spirals. Other researchers contend that repeated encounters with the companions have molded the disk. In this way, bodies other than planets may be responsible for some of the debris disks' characteristics.

The uncertainties in the interpretation of the images stem mainly from the fact that the sample of resolved disks remains very small. Each is a special case. But the Spitzer Space Telescope is changing that situation dramatically. One of its main goals is to collect an extensive sample of debris disks, some of which can be seen directly. The telescope has detectors that work at far-infrared wavelengths, as IRAS's did, but are 1,000 times as sensitive. They can detect minute amounts of dust and so obtain a more complete sample of disks. Last December, Karl R. Stapelfeldt of JPL and his colleagues released the first Spitzer images of a debris disk, around the star Fomalhaut. At a wavelength of 70 microns, the shape of the disk becomes evident—an edge-on ring almost 200 astronomical units in radius. One side appears brighter than the other, perhaps the result of a recent asteroid collision or the gravitational influence of an unseen planet. At a wavelength of 24 microns, a concentration of warm material appears close to the star—a clear analogue of our own zodiacal cloud, suggesting that Fomalhaut may have something akin to an asteroid belt.

The debris disks have shown astronomers that stars other than our sun have asteroids and comets—one of the consequences of the planet formation process. This result implies that the solar system fundamentally resembles other systems. On the other hand, even the smallest debris disk detected contains 50 times as much dust as the solar system does. Is that because planets in our solar system have already expelled most planetesimals? Is it because the sun started with an unusually small disk? Or is it because observers have not yet used instruments sensitive enough to uncover true analogues to our solar system?

Astronomers still need to piece together a detailed and consistent picture of how planets form around stars of different masses and how these planets evolve over time. Further observations by Hubble, Spitzer and ground-based observatories will prove crucial to answering these questions. We may then know how exceptional our solar system is among all possible planetary systems. SA

MORE TO EXPLORE

Planetary Material around Main-Sequence Stars. A. M. Lagrange, D. E. Backman and P. Artymowicz in *Protostars and Planets IV.* Edited by Vincent Mannings, Alan Boss and Sara Russell. University of Arizona, 2000.

Star Factories: The Birth of Stars and Planets. Ray Jayawardhana. Raintree/Steck-Vaughn, 2000.

Dusty Circumstellar Disks. Benjamin Zuckerman in *Annual Review of Astronomy and Astrophysics,* Vol. 39, pages 549–580; September 2001.

Distant Wanderers: The Search for Planets beyond the Solar System. Bruce Dorminey. Springer-Verlag, 2001.

HST/ACS Coronagraphic Imaging of the Circumstellar Disk around HD1415659A. Mark Clampin et al. in *Astronomical Journal,* Vol. 126, No. 1, pages 385–392; July 2003. Available online at **arXiv.org/abs/astro-ph/0303605**

LUCY READING AND DAVID R. ARDILA; SOURCE: CELESTE SPANGLER *California Institute of Technology*


0,000,000'S OF STYLES!
EVEN MORE THAN EVER
WORLD O' FLAVORS
TRY THIS
MEGA·CHOICE MALL
SOCKS! 1,000 colors!
12,000,000 TITLES
ZILLIONS OF CHOICES!
TRY 'EM ALL
SHOP HERE ALL DAY!! TRY EVERY CAR!
CHOICES 'R' U

The Tyranny of Choice

Logic suggests that having options allows people to select precisely what makes them happiest. But, as studies show, abundant choice often makes for misery

By Barry Schwartz

Americans today choose among more options in more parts of life than has ever been possible before. To an extent, the opportunity to choose enhances our lives. It is only logical to think that if some choice is good, more is better; people who care about having infinite options will benefit from them, and those who do not can always just ignore the 273 versions of cereal they have never tried. Yet recent research strongly suggests that, *psycho*logically, this assumption is wrong. Although some choice is undoubtedly better than none, more is not always better than less.

This evidence is consistent with large-scale social trends. Assessments of well-being by various social scientists—among them, David G. Myers of Hope College and Robert E. Lane of Yale University—reveal that increased choice and increased affluence have, in fact, been accompanied by *decreased* well-being in the U.S. and most other affluent societies. As the gross domestic product more than doubled in the past 30 years, the proportion of the population describing itself as "very happy" declined by about 5 percent, or by some 14 million people. In addition, more of us than ever are clinically depressed. Of course, no one believes that a single factor explains decreased well-being, but a number of findings indicate that the explosion of choice plays an important role.

Thus, it seems that as society grows wealthier and people become freer to do whatever they want, they get less happy. In an era of ever greater personal autonomy, choice and control, what could account for this degree of misery?

Along with several colleagues, I have recently conducted research that offers insight into why many people end up unhappy rather than pleased when their options expand. We began by making a distinction between "maximizers" (those who always aim to make the best possible choice) and "satisficers" (those who aim for "good enough," whether or not better selections might be out there). We borrowed the term "satisficers" from the late Nobel Prize–winning psychologist and economist Herbert A. Simon of Carnegie Mellon University.

In particular, we composed a set of statements—the Maximization Scale—to diagnose people's propensity to maximize. Then we had several thousand people rate themselves from 1 to 7 (from "completely disagree"

THE MAXIMIZATION SCALE

THE STATEMENTS BELOW distinguish maximizers from satisficers. Subjects rate themselves from 1 to 7, from "completely disagree" to "completely agree," on each statement. We generally consider people whose average rating is higher than 4 to be maximizers. When we looked at averages from thousands of subjects, we found that about a third scored higher than 4.75 and a third lower than 3.25. Roughly 10 percent of subjects were extreme maximizers (averaging greater than 5.5), and 10 percent were extreme satisficers (averaging lower than 2.5.) *—B.S.*

1. Whenever I'm faced with a choice, I try to imagine what all the other possibilities are, even ones that aren't present at the moment.
2. No matter how satisfied I am with my job, it's only right for me to be on the lookout for better opportunities.
3. When I am in the car listening to the radio, I often check other stations to see if something better is playing, even if I am relatively satisfied with what I'm listening to.
4. When I watch TV, I channel surf, often scanning through the available options even while attempting to watch one program.
5. I treat relationships like clothing: I expect to try a lot on before finding the perfect fit.
6. I often find it difficult to shop for a gift for a friend.
7. Renting videos is really difficult. I'm always struggling to pick the best one.
8. When shopping, I have a hard time finding clothing that I really love.
9. I'm a big fan of lists that attempt to rank things (the best movies, the best singers, the best athletes, the best novels, etc.).
10. I find that writing is very difficult, even if it's just writing a letter to a friend, because it's so hard to word things just right. I often do several drafts of even simple things.
11. No matter what I do, I have the highest standards for myself.
12. I never settle for second best.
13. I often fantasize about living in ways that are quite different from my actual life.

to "completely agree") on such statements as "I never settle for second best." We also evaluated their sense of satisfaction with their decisions.

We did not define a sharp cutoff to separate maximizers from satisficers, but in general, we think of individuals whose average scores are higher than 4 (the scale's midpoint) as maximizers and those whose scores are lower than the midpoint as satisficers. People who score highest on the test—the greatest maximizers—engage in more product comparisons than the lowest scorers, both before and after they make purchasing decisions, and they take longer to decide what to buy. When satisficers find an item that meets their standards, they stop looking. But maximizers exert enormous effort reading labels, checking out consumer magazines and trying new products. They also spend more time comparing their purchasing decisions with those of others.

Naturally, no one can check out every option, but maximizers strive toward that goal, and so making a decision becomes increasingly daunting as the number of choices rises. Worse, after making a selection, they are nagged by the alternatives they have not had time to investigate. In the end, they are more likely to make better objective choices than satisficers but get less satisfaction from them. When reality requires maximizers to compromise—to end a search and decide on something—apprehension about what might have been takes over.

We found as well that the greatest maximizers are the least happy with the fruits of their efforts. When they compare themselves with others, they get little pleasure from finding out that they did better and substantial dissatisfaction from finding out that they did worse. They are more prone to experiencing regret after a purchase, and if their acquisition disappoints them, their sense of well-being takes longer to recover. They also tend to brood or ruminate more than satisficers do.

Does it follow that maximizers are less happy in general than satisficers? We tested this by having people fill out a variety of questionnaires known to be reliable indicators of well-being. As might be expected, individuals with high maximization scores experienced less satisfaction with life and were less happy, less optimistic and more depressed than people with low maximization scores. Indeed, those with extreme maximization ratings had depression scores that placed them in the borderline clinical range.

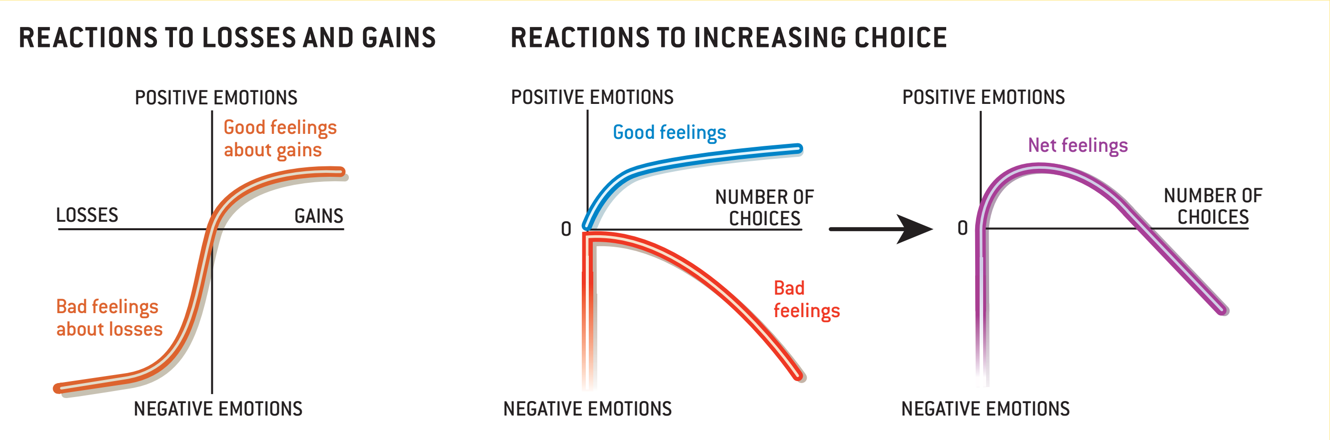
Recipe for Unhappiness

SEVERAL FACTORS EXPLAIN why more choice is not always better than less, especially for maximizers. High among these are "opportunity costs." The quality of any given option cannot be assessed in isolation from its alternatives. One of the "costs" of making a selection is losing the opportunities that a different option would have afforded. Thus, an opportunity cost of vacationing on the beach in Cape Cod might be missing the fabulous restaurants in the Napa Valley. If we assume that opportunity costs reduce the overall desirability of the most preferred choice, then the more alternatives there are, the deeper our sense of loss will be and the less satisfaction we will derive from our ultimate decision.

Lyle Brenner of the University of Florida and his collabora-

FEELINGS EVOKED BY EVER MORE CHOICES

EARLY DECISION-MAKING RESEARCH by Daniel Kahneman and Amos Tversky showed that people respond much more strongly to losses than gains (*depicted schematically in left graph*). Similarly, my co-workers and I believe that feelings of well-being initially rise as choice increases (*blue line in schematic plot, center graph*) but then level off quickly (good feelings satiate). Meanwhile, although zero choice (*at the y axis*) evokes virtually infinite unhappiness, bad feelings escalate (*red line*) as we go from having few choices to many. The net result (*purple line in right graph*) is that, at some point, added choice only decreases happiness. *—B.S.*



BARRY SCHWARTZ

tors demonstrated the effects of opportunity costs when they had subjects put a dollar value on subscriptions to magazines or flights from San Francisco to attractive locations. Some attached prices to a single magazine subscription or a single destination. Others attached prices to the same magazine or destination when it was part of a group containing three others. Prices were consistently lower when a given alternative was evaluated as part of a group than when it was evaluated in isolation.

Why might this be so? When you assign a value to, say, *Newsweek,* as part of a group that also contains *People,* the *New Republic* and *Us,* your tendency will be to compare the various magazines. Perhaps you judge *Newsweek* to be more informative than *People* but less entertaining. Each comparison that *Newsweek* wins will be a gain, but each comparison that it loses will be a loss, an opportunity cost. Any particular magazine will both benefit and suffer from comparison with others. But we know from the research of Nobelist psychologist Daniel Kahneman of Princeton University and his late colleague Amos Tversky of Stanford that losses (in this case, opportunity costs) have a much greater psychological impact than gains. Losses make us hurt more than gains make us feel good.

Sometimes opportunity costs may create enough conflict to produce paralysis. For example, participants in one study were offered $1.50 for filling out some questionnaires. After they finished, they were offered a fancy metal pen instead of the $1.50 and told that the pen usually costs about $2. Seventy-five percent chose the pen. In a second condition, subjects were offered the $1.50 or a choice between that same metal pen and a pair of less expensive felt-tipped pens (also worth about $2 together). Now fewer than 50 percent chose any pen.

The problem of opportunity costs will be worse for a maximizer than for a satisficer. The latter's "good enough" philosophy can survive thoughts about opportunity costs. In addition, the "good enough" standard leads to much less searching and inspection of alternatives than the maximizer's "best" standard. With fewer choices under consideration, a person will have fewer opportunity costs to subtract.

Regret Adds to Costs

JUST AS PEOPLE FEEL sorrow about the opportunities they have forgone, they may also suffer regret about the option they settle on. My colleagues and I devised a scale to measure proneness to feeling regret, and we found that people with high sensitivity to regret are less happy, less satisfied with life, less optimistic and more depressed than those with low sensitivity. Not surprisingly, we also found that people with high regret sensitivity tend to be maximizers. Indeed, we think that worry over future regret is a major reason that individuals become maximizers. The only way to be sure you will not regret a decision is by making the best possible one. Unfortunately, the more options you have and the more opportunity costs you incur, the more likely you are to experience regret.

Regret may be one reason for our aversion to losses. Have you ever bought an expensive pair of shoes only to discover that they are so uncomfortable that you cannot wear them for more than 10 minutes without hobbling? Did you toss them out, or are they still sitting in the back of your closet? Chances are you had a hard time throwing them away. Having bought the shoes, you incurred an actual, or "sunk," cost, and you are going to keep them around in the hope that eventually you will get your money's worth out of them. To give the shoes away or throw them out would force you to acknowledge a mistake—a loss.

In a classic demonstration of the power of sunk costs, people were offered season subscriptions to a local theater company. Some were offered the tickets at full price and others at a discount. Then the researchers simply kept track of how often the ticket purchasers actually attended the plays over the course of the season. Full-price payers were more likely to show up at performances than discount payers. The reason for this, the investigators argued, was that the full-price payers would experience more regret if they did not use the tickets because not using the more costly tickets would constitute a bigger loss.

Several studies have shown that two of the factors affecting regret are how much one feels personal responsibility for the result and how easy it is to imagine a better alternative. The availability of choice obviously exacerbates both these factors. When you have no options, what can you do? You will feel disappointment, maybe; regret, no. With no options, you just do the best you can. But with many options, the chances increase that a really good one is out there, and you may well feel that you ought to have been able to find it.

LESSONS

Choose when to choose.
We can decide to restrict our options when the decision is not crucial. For example, make a rule to visit no more than two stores when shopping for clothing.

Learn to accept "good enough."
Settle for a choice that meets your core requirements rather than searching for the elusive "best." Then stop thinking about it.

Don't worry about what you're missing.
Consciously limit how much you ponder the seemingly attractive features of options you reject. Teach yourself to focus on the positive parts of the selection you make.

Control expectations.
"Don't expect too much, and you won't be disappointed" is a cliché. But that advice is sensible if you want to be more satisfied with life. *—B.S.*

Adaptation Dulls Joy

A PHENOMENON called adaptation also contributes to the fallout we face from too many choices. Simply put, we get used to things, and as a result, very little in life turns out quite as good as we expect it to be. After much anguish, you might decide to buy a Lexus and then try to put all the attractions of other models out of your mind. But once you are driving your new car, adaptation begins, and the experience falls just a little bit flat. You are hit with a double whammy—regret about what you did not choose and disappointment with what you did, even if your final decision was not bad.

Because of adaptation, enthusiasm about positive experiences does not sustain itself. Daniel T. Gilbert of Harvard University and Timothy D. Wilson of the University of Virginia and their collaborators have shown that people consistently mispredict how long good experiences will make them feel good and how long bad experiences will make them feel bad. The waning of pleasure or enjoyment over time always seems to come as an unpleasant surprise.

And it may cause more disappointment in a world of many options than in a world of few. The opportunity costs associated with a decision and the time and effort that go into making it are "fixed costs" that we "pay" up front, and those costs then get "amortized" over the life of the decision. The more we invest in a decision, the more satisfaction we expect to realize from our investment. If the decision provides substantial satisfaction for a long time after it is made, the costs of making it recede into insignificance. But if the decision provides pleasure for only a short time, those costs loom large. Spending four months deciding what stereo to buy is not so bad if you really enjoy that stereo for 15 years. But if you end up being excited by it for six months and then adapting, you may feel like a fool for having put in all that effort.

The Curse of High Expectations

A SURFEIT OF alternatives can cause distress in yet another way: by raising expectations. In the fall of 1999 the *New York Times* and CBS News asked teenagers to compare their experiences with those their parents had growing up. Fifty percent of children from affluent households said their lives were harder. When questioned further, these adolescents talked about high expectations, both their own and their parents'. They talked about "too muchness": too many activities, too many consumer choices, too much to learn. As one commentator put it, "Children feel the pressure ... to be sure they don't slide back. Everything's about going forward.... Falling back is the American nightmare." So if your perch is high, you have much further to fall than if your perch is low.

The amount of choice we now have in most aspects of our lives contributes to high expectations. When I was on vacation a few years ago in a tiny seaside town on the Oregon coast, I went into the local grocery store to buy ingredients for dinner.

THE AUTHOR

BARRY SCHWARTZ is Dorwin Cartwright Professor of Social Theory and Social Action in the department of psychology at Swarthmore College, where he has taught since 1971. He recently published a book on the consequences of excessive choice (see "More to Explore," on opposite page) and has written several other books, including *The Battle for Human Nature* and *The Costs of Living*.

MATT COLLINS

The store offered about a dozen options for wine. What I got was so-so, but I did not expect to be able to get something very good and, hence, was satisfied with what I had. If instead I had been shopping in a store that offered an abundance of choices, my expectations would have been a good deal higher and that same so-so wine might have left me sorely disappointed.

Alex C. Michalos of the University of Northern British Columbia has pointed out that all our evaluations of the things we do and buy depend on comparison—to past experiences, to what we were hoping for, and to what we expected. When we say that some experience was good, what we mean, in part, is that it was better than we expected it to be. So high expectations almost guarantee that experiences will fall short, especially for maximizers and especially when regret, opportunity costs, and adaptation do not factor into our expectations.

A Link to Depression?

THE CONSEQUENCES OF unlimited choice may go far beyond mild disappointment, to suffering. As I indicated earlier, Americans are showing a decrease in happiness and an increase in clinical depression. One important contributing factor, I think, is that when we make decisions, experience the consequences and find that they do not live up to expectations, we blame ourselves. Disappointing outcomes constitute a personal failure that could and should have been avoided if only we had made a better choice.

The research that my colleagues and I have done suggests that maximizers are prime candidates for depression. With group after group of people, varying in age (including young adolescents), gender, educational level, geographic location, race and socioeconomic status, we have found a strong correlation between maximizing and measures of depression. If the experience of disappointment is relentless, if virtually every choice you make fails to live up to expectations and aspirations, and if you consistently take personal responsibility for the disappointments, then the trivial looms larger and larger, and the conclusion that you cannot do anything right becomes devastating. Although depression has many sources, and the relation among choice, maximizing and depression requires more study, there is good reason to believe that overwhelming choice at least contributes to the epidemic of unhappiness spreading through modern society.

What Can Be Done

THE NEWS I HAVE reported is not good. We get what we say we want, only to discover that what we want does not satisfy us to the degree that we expect. Does all this mean that we would all be better off if our choices were severely restricted, even eliminated? I do not think so. The relation between choice and well-being is complicated. A life without significant choice would be unlivable. Being able to choose has enormous important positive effects on us. But only up to a point. As the number of choices we face increases, the psychological benefits we derive start to level off. At the same time, some of the negative effects of choice that I have discussed begin to appear, and rather than leveling off, they accelerate. A quarter of a century ago the late Clyde H. Coombs of the University of Michigan at Ann Arbor and George S. Avrunin of the University of Massachusetts at Amherst noted that good things "satiate" and bad things "escalate." Much the same can be said of feelings. Indeed, a point is reached at which increased choice brings increased misery rather than increased opportunity. It appears that American society has long since passed that point.

Few Americans would favor passing laws to limit choices. But individuals can certainly take steps to mitigate choice-related distress. Such actions require practice, discipline and perhaps a new way of thinking, but each should bring its own rewards [*see box on opposite page*].

Beyond those individual strategies, I think our society would be well served to rethink its worship of choice. As I write this, public debate continues about privatization of Social Security (so people could select their retirement investments), privatization of Medicare and prescription drug benefits (so people could choose their own health plans), and choice in public education. And in the private sphere, medical ethicists treat the idea of "patient autonomy" as sacrosanct, as if it goes without saying that having patients choose their treatments will make them better off. Software developers design their products so that users can customize them to their own specific needs and tastes, as if the resulting complexity and confusion are always a price worth paying to maximize user flexibility. And manufacturers keep offering new products or new versions of old products, as if we needed more variety. The lesson of my research is that developments in each of these spheres may well rest on assumptions that are deeply mistaken. SA

MATT COLLINS

MORE TO EXPLORE

Choices, Values, and Frames. Daniel Kahneman and Amos Tversky in *American Psychologist,* Vol. 39, pages 341–350; April 1984.

The Loss of Happiness in Market Democracies. Robert E. Lane. Yale University Press, 2001.

Maximizing versus Satisficing: Happiness Is a Matter of Choice. Barry Schwartz, Andrew Ward, John Monterosso, Sonya Lyubomirsky, Katherine White and Darrin Lehman in *Journal of Personality and Social Psychology,* Vol. 83, No. 5, pages 1178–1197; 2002.

The Paradox of Choice: Why More Is Less. Barry Schwartz. Ecco/HarperCollins Publishers, 2004.

the first Nanochips

As scientists and engineers continue to push back the limits of chipmaking technology, they have quietly entered into the nanometer realm

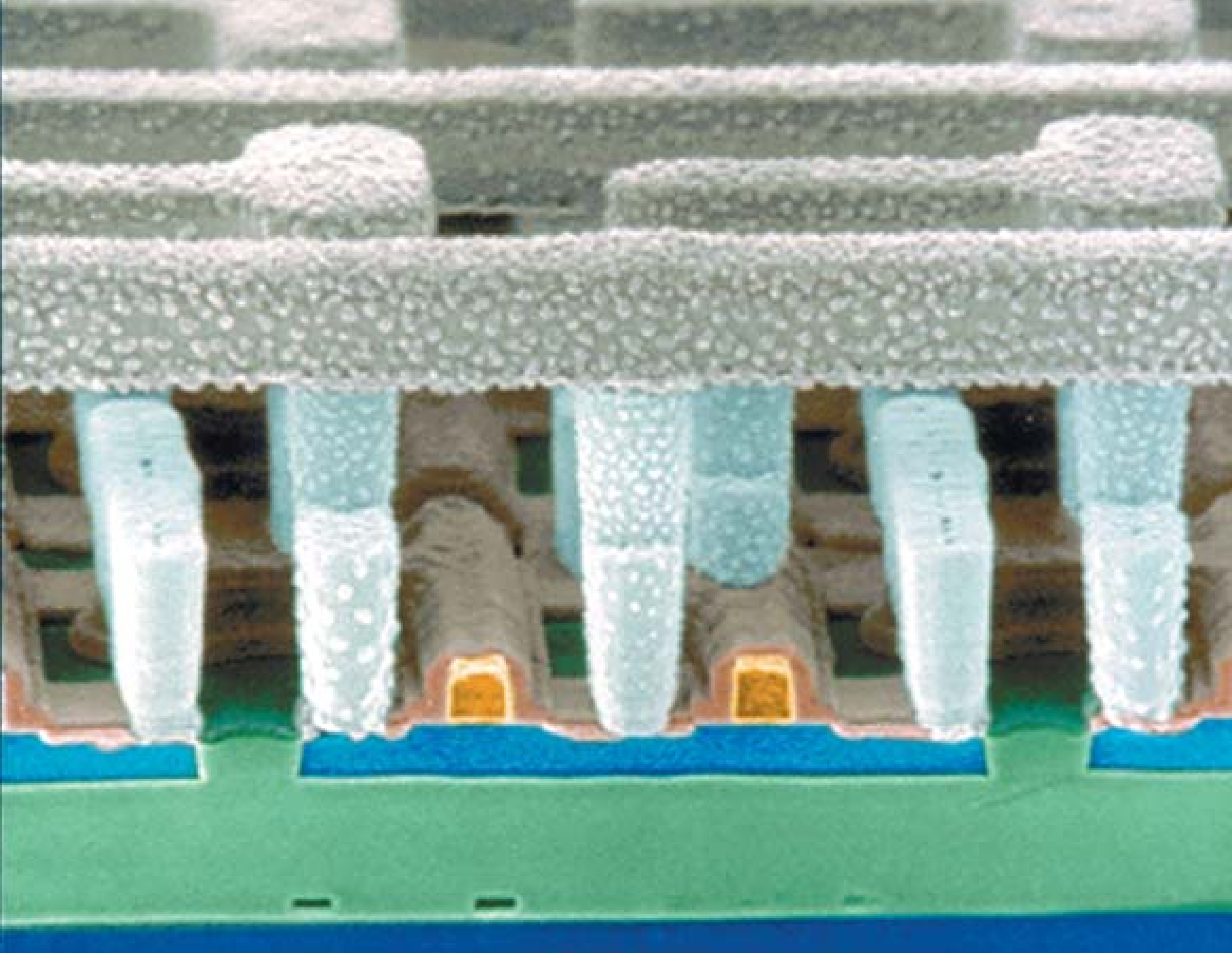
By G. Dan Hutcheson

For most people,

the notion of harnessing nanotechnology for electronic circuitry suggests something wildly futuristic. In fact, if you have used a personal computer made in the past few years, your work was most likely processed by semiconductors built with nanometer-scale features. These immensely sophisticated microchips—or rather, nanochips—are now manufactured by the millions, yet the scientists and engineers responsible for their development receive little recognition. You might say that these people are the Rodney Dangerfields of nanotechnology. So here I would like to trumpet their accomplishments and explain how their efforts have maintained the steady advance in circuit performance to which consumers have grown accustomed.

The recent strides are certainly impressive, but, you might ask, is semiconductor manufacture really nanotechnology? Indeed it is. After all, the most widely accepted definition of that word applies to something with dimensions smaller than 100 nanometers, and the first transistor gates under this mark went into production in 2000. Integrated circuits coming to market now have gates that are a scant 50 nanometers wide. That's 50 billionths of a meter, about a thousandth the width of a human hair.

Having such minuscule components conveniently allows one to stuff a lot into a compact package, but saving space per se is not the impetus behind the push for extreme miniaturization. The reason to make things small is that it lowers the



INNNOVATION BOOSTS the performance of microprocessors, providing techniques that go beyond shrinking the size of transistors. The chip shown here, enlarged some 50,000 times, improves speed and saves power by placing silicon for the transistors (*light blue*) above a layer of oxide (*green*).

unit cost for each transistor. As a bonus, this overall miniaturization shrinks the size of the gates, which are the parts of the transistors that switch between blocking electric current and allowing it to pass. The more narrow the gates, the faster the transistors can turn on and off, thereby raising the speed limits for the circuits using them. So as microprocessors gain more transistors, they also gain more speed.

The desire for boosting the number of transistors on a chip and for running it faster explains why the semiconductor industry, just as it crossed into the new millennium, shifted from manufacturing microchips to making nanochips. How it quietly passed this milestone, and how it continues to advance, is an amazing story of people overcoming some of the greatest engineering challenges of our time—challenges every bit as formidable as those encountered in building the first atomic bomb or sending a person to the moon.

Straining to Accelerate

THE BEST WAY to get a flavor for the technical innovations that helped to usher in the current era of nanochips is to survey improvements that have been made in each of the stages required to manufacture a modern semiconductor—say, the microprocessor that powers the computer on which I typed this text. That chip, a Pentium 4, contains some 42 million transistors intricately wired together. How in the world was this marvel of engineering constructed? Let us survey the steps.

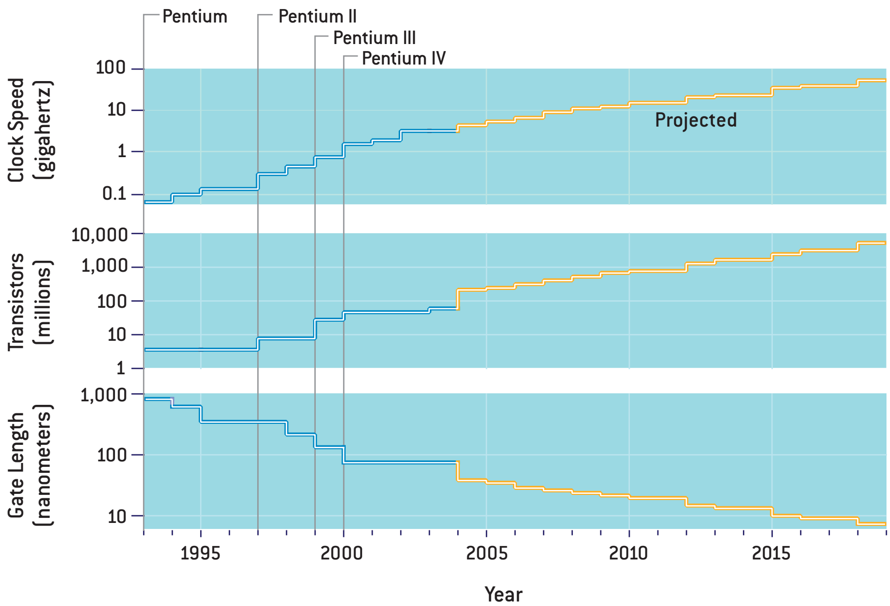
Before the chipmaking process even begins, one needs to obtain a large crystal of pure silicon. The traditional method for doing so is to grow it from a small seed crystal that is immersed in a batch of molten silicon. This process yields a cylindrical ingot—a massive gem-quality crystal—from which many thin wafers are then cut.

It turns out that such single-crystal ingots are no longer good enough for the job: they have too many "defects," dislocations in the atomic lattice that hamper the silicon's ability to conduct and otherwise cause trouble during chip manufacture. So chipmakers now routinely deposit a thin, defect-free layer of single-crystal silicon on top of each wafer by exposing it to a gas containing silicon. This technique improves the speed of the transistors, but engineers have been pushing hard to do even better using something called silicon-on-insulator technology, which involves putting a thin layer of insulating oxide slightly below the surface of the wafer. Doing so lowers the capacitance (the ability to store electrical charge) between parts of the transistors and the underlying silicon substrate, capacitance that would otherwise sap speed and waste power. Adopting a silicon-on-insulator geometry can boost the rate at which the transistors can be made to switch on and off (or, alternatively, reduce the power needed) by up to 30 percent. The gain is equivalent to what one gets in moving one generation ahead in feature size.

IBM pioneered this technology and has been selling integrated circuits made with it for the past five years. The process IBM developed, dubbed SIMOX, short for *s*eparation by *im*plantation of *ox*ygen, was to bombard the silicon with oxygen atoms (or rather, oxygen ions, which have electrical charge and can thus be readily accelerated to high speeds). These ions implant themselves deep down, relatively speaking, where they combine with atoms in the wafer and form a layer of silicon dioxide. One difficulty with this approach is that the passage of oxygen ions through the silicon creates many defects, so the surface has to be carefully heated afterward to mend disruptions to the crystal lattice. The greater problem is that oxygen implantation is inherently slow, which makes it costly. Hence, IBM reserved its silicon-on-insulator technology for its most expensive chips.

A new, faster method for accomplishing the same thing is, however, gaining ground. The idea is to first form an insulating oxide layer directly on top of a silicon wafer. One then flips the oxidized surface over and attaches it onto another, untreated wafer. After cleverly pruning off most of the silicon above the oxide layer, one ends up with the desired arrangement: a thin stratum of silicon on top of the insulating oxide layer on top of a bulk piece of silicon, which just provides physical support.

The key was in developing a precision slicing method. The French company that did so, Soitec, aptly trademarked the name Smart Cut for this technique, which requires shooting hydrogen ions through the oxidized surface of the first wafer so that they implant themselves at a prescribed depth within the underlying silicon. (Implanting hydrogen can be done more rapidly than implanting oxygen, making this process relatively inexpensive.) Because the hydrogen ions do most of their damage right where they stop, they produce a level within the silicon that is quite fragile. So after flipping this



MICROPROCESSOR components have entered the nano realm during the past decade, as illustrated by the evolution of Intel's Pentium series (*blue*), which shows remarkable gains in the speed and quantity of transistors, both of which rise as the gate length of the transistors diminishes. If the semiconductor industry even comes close to matching its forecasts (*yellow*), these trends should continue.

LUCY READING; SOURCES: INTEL AND INTERNATIONAL TECHNOLOGY ROADMAP FOR SEMICONDUCTORS

Overview/*Smaller, Faster, Better Chips*

- In 2000 the semiconductor industry quietly began producing "nanochips"—chips with features measuring less than 100 nanometers (roughly one thousandth the thickness of a human hair). These devices are found in the average desktop computer today.
- Reducing the size of features boosts speed and improves the economics of manufacture by allowing more transistors (often more than 50 million) to be put on a single chip. In just a few years, a typical microprocessor will contain about 10 times that number.
- Significant gains have accrued from several new forms of technology, including improved materials and methods to correct for distortions that occur from optical diffraction when patterning the chips.

treated wafer over and attaching it to a wafer of bulk silicon, one can readily cleave the top off at the weakened plane. Any residual roughness in the surface can be easily polished smooth. Even IBM now employs Smart Cut for making some of its high-performance chips, and AMD (Advanced Micro Devices in Sunnyvale, Calif.) will use it in its upcoming generation of microprocessors.

The never-ending push to boost the switching speed of transistors has also brought another very basic change to the foundations of chip manufacture, something called strained silicon. It turns out that forcing the crystal lattice of silicon to stretch slightly (by about 1 percent) increases the mobility of electrons passing through it considerably, which in turn allows the transistors built on it to operate faster. Chipmakers induce strain in silicon by bonding it to another crystalline material—in this case, a silicon-germanium blend—for which the lattice spacing is greater. Although the technical details of how this strategy is being employed remain closely held, it is well known that many manufacturers are adopting this approach. Intel, for example, is using strained silicon in an advanced version of its Pentium 4 processor called Prescott, which began selling late last year.

Honey, I Shrunk the Features

ADVANCES IN the engineering of the silicon substrate are only part of the story: the design of the transistors constructed atop the silicon has also improved tremendously in recent years. One of the first steps in the fabrication of transistors on a digital chip is growing a thin layer of silicon dioxide on the surface of a wafer, which is done by exposing it to oxygen and water vapor, allowing the silicon, in a sense, to rust (oxidize). But unlike what happens to the steel body of an old car, the oxide does not crumble away from the surface. Instead it clings firmly, and oxygen atoms required for further oxidization must diffuse through the oxide coating to reach fresh silicon underneath. The regularity of this diffusion provides chipmakers with a way to control the thickness of the oxide layers they create.

For example, the thin oxide layers required to insulate the gates of today's tiny transistors can be made by allowing oxygen to diffuse for only a short time. The problem is that the gate oxide, which in modern chips is just several atoms thick, is becoming too slim to lay down reliably. One fix, of course, is to make this layer thicker. The rub here is that as the thickness of the oxide increases, the capacitance of the gate decreases. You might ask: Isn't that a good thing? Isn't capacitance bad? Often capacitance is indeed something to be avoided, but the gate of a transistor operates by inducing electrical charge in the silicon below it, which provides a channel for current to flow. If the capacitance of the gate is too low, not enough charge will be present in this channel for it to conduct.

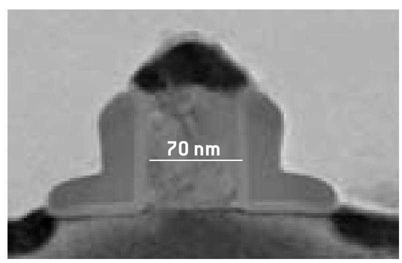
The solution is to use something other than the usual silicon dioxide to insulate the gate. In particular, semiconductor manufacturers have been looking hard at what are known as high-K (high-dielectric-constant) materials, such as hafnium oxide and strontium titanate, ones that allow the oxide layer to be made thicker, and thus more robust, without compromising the ability of the gate to act as a tiny electrical switch.

Placing a high-K insulator on top of silicon is, however, not nearly as straightforward as just allowing it to oxidize. The task is best accomplished with a technique called atomic-layer deposition, which employs a gas made of small molecules that naturally stick to the surface but do not bond to one another. A single-molecule-thick film can be laid down simply by exposing the wafer to this gas long enough so that every spot becomes covered. Treatment with a second gas, one that reacts with the first to form the material in the coating, creates the molecule-thin veneer. Repeated applications of these two gases, one after the next, deposit layer over layer of this substance until the desired thickness is built up.

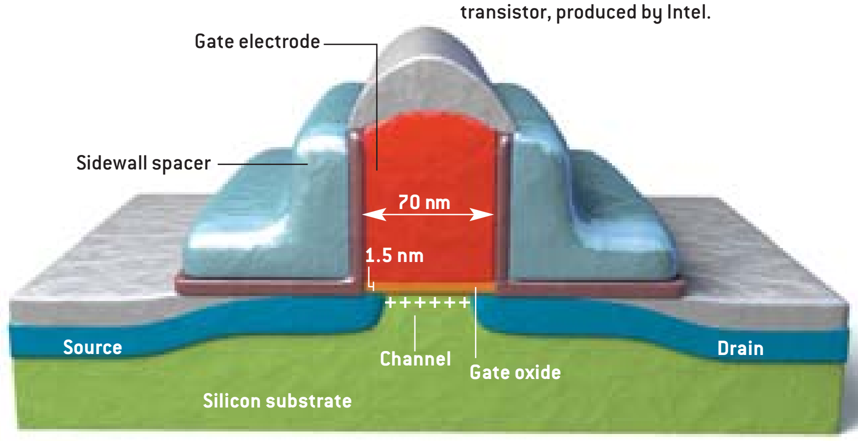
After the gate insulator is put in place, parts of it must be selectively removed to achieve the appropriate pattern on the wafer. The procedure for doing so (lithography) constitutes a key part of the

FIELD-EFFECT TRANSISTOR

THE FUNDAMENTAL BUILDING BLOCK of a microprocessor is the field-effect transistor, which acts as a simple switch. The proper voltage applied to the gate electrode induces charge along the channel, which then carries current between the source and the drain, turning the switch on. With sufficiently small gates, these transistors can switch on and off billions of times each second.



FIRST-GENERATION nanometer-scale transistor, produced by Intel.

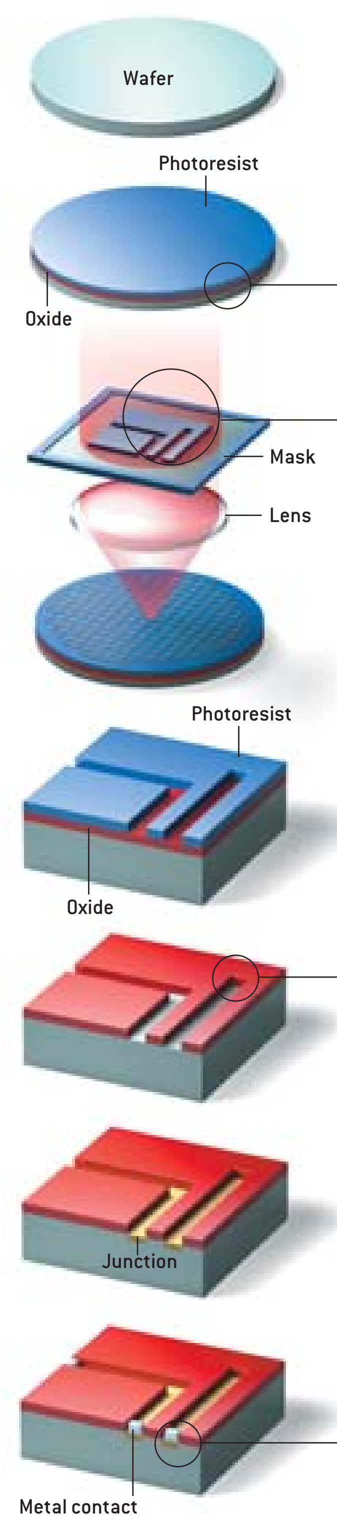


INTEL (*micrograph*); BRYAN CHRISTIE DESIGN (*illustration*)

BASIC CHIPMAKING PROCESS

A CIRCULAR WAFER of silicon about the size of a dinner plate provides the starting point for the step-by-step chipmaking process, which sculpts transistors and their interconnections. Some of the manipulations shown below are repeated many times in the course of production, to build complex structures one layer at a time.

BASIC CHIPMAKING PROCESS



1 Steam oxidizes surface (*red layer*)

2 Photoresist (*dark blue layer*) coats oxidized wafer

3 Lithography transfers desired pattern from mask to wafer

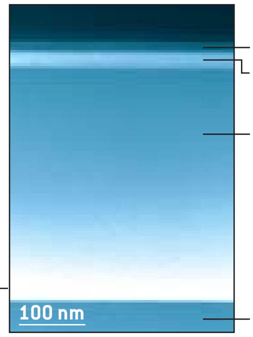
4 Chemicals and baking harden unexposed photoresist. Other parts of photoresist are removed

5 Chemical etching selectively strips off the oxide where no photoresist protects it. The rest of the photoresist is removed.

6 Ions shower etched areas, forming source and drain junctions

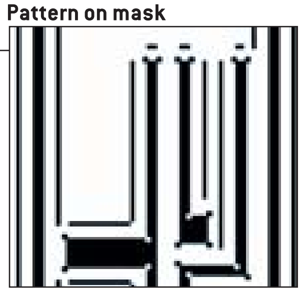
7 Metal contacts are added using lithography during later stages of fabrication

REFINEMENTS IN CHIPMAKING

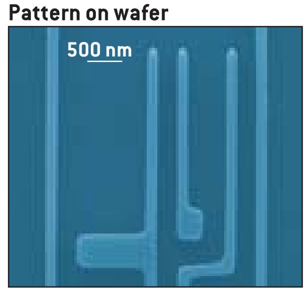


PERFORMANCE HAS IMPROVED with the growing use of wafers having a buried oxide layer or those fashioned to have a thin layer of strained silicon at the top—or by employing both techniques at once, as shown here.

Pattern on mask



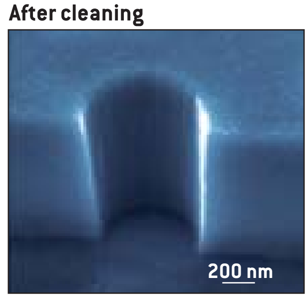
Pattern on wafer



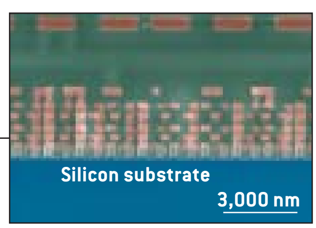
IF MODERN TECHNIQUES such as "optical proximity correction" are applied to compensate for the blurring effects of diffraction, photolithography can create features smaller than the wavelength of light used in projecting the pattern. In this example of optical proximity correction, a complicated pattern used for the mask (*left*) results in crisp features on the chip (*right*).

Before cleaning

After cleaning



AS FEATURE SIZES SHRINK, removing photoresist and residues that remain after etching (*left*) becomes difficult. But supercritical carbon dioxide can penetrate tiny openings and dislodge particles without leaving traces of cleaning fluid behind (*right*).



AS MANY AS EIGHT levels of wiring now connect the millions of transistors found on a typical microprocessor. Aluminum, the metal long used for this purpose, has given way to copper, which is more difficult to emplace but improves the speed and integrity of the signals carried on the wires.

technology needed to create transistors and their interconnections. Semiconductor lithography employs a photographic mask to generate a pattern of light and shadows, which is projected on a wafer after it is coated with a light-sensitive substance called photoresist. Chemical processing and baking harden the unexposed photoresist, which protects those places in shadow from later stages of chemical etching.

Practitioners once believed it impossible to use lithography to define features smaller than the wavelength of light employed, but for a few years now, 70-nanometer features have been routinely made using ultraviolet light with a wavelength of 248 nanometers. To accomplish this magic, lithography had to undergo some dramatic changes. The tools brought to bear have complicated names—optical proximity correction, phase-shifting masks, excimer lasers—but the idea behind them is simple, at least in principle. When the size of the features is smaller than the wavelength of the light, the distortions, which arise through optical diffraction, can be readily calculated and corrected for. That is, one can figure out an arrangement for that mask that, after diffraction takes place, yields the desired pattern on the silicon. For example, suppose a rectangle is needed. If the mask held a plain rectangular shape, diffraction would severely round the four corners projected on the silicon. If, however, the pattern on the mask were designed to look more like a dog bone, the result would better approximate a rectangle with sharp corners.

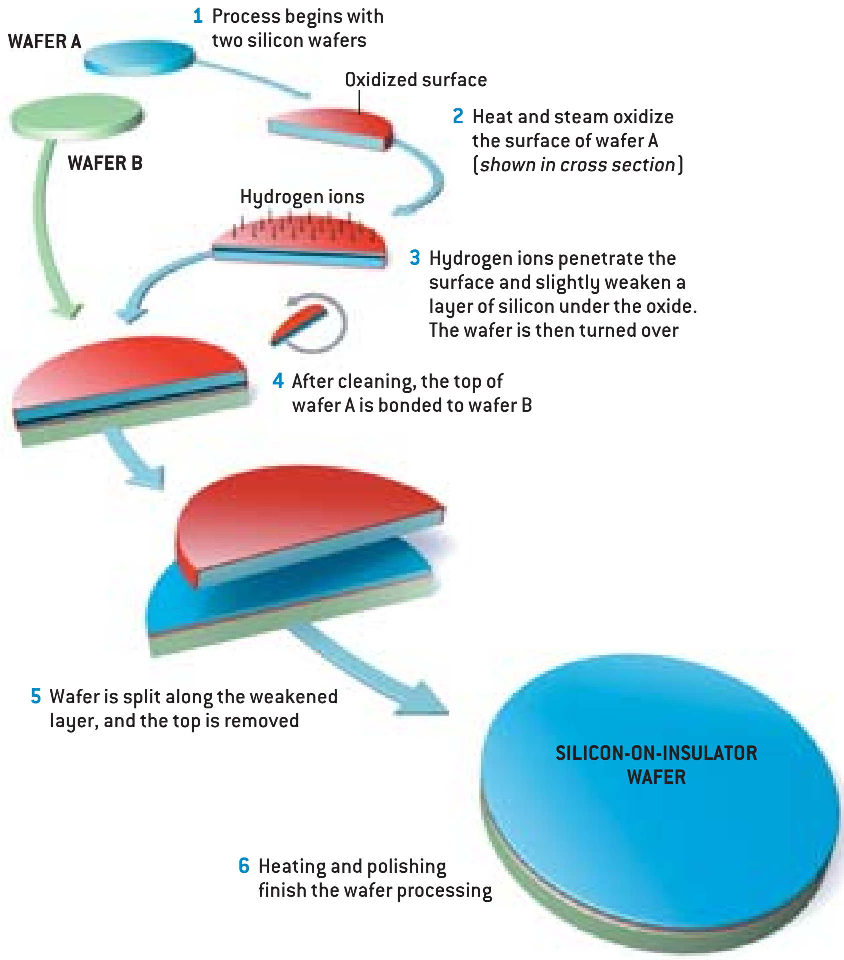
This general strategy now allows transistors with 50-nanometer features to be produced using light with a wavelength of 193 nanometers. But one can push these diffraction-correction techniques only so far, which is why investigators are trying to develop the means for higher-resolution patterning. The most promising approach employs lithography, but with light of much shorter wavelength—what astronomers would call "soft" x-rays or, to keep with the preferred term in the semiconductor industry, extreme ultraviolet.

Semiconductor manufacturers face daunting challenges as they move to extreme ultraviolet lithography, which reduces the wavelengths (and thus the size of the features that can be printed) by an order of magnitude. The prototype systems built so far are configured for a 13-nanometer wavelength. They are truly marvels of engineering—on both macroscales and nanoscales.

Take, for instance, the equipment needed to project images onto wafers. Because all materials absorb strongly at extreme ultraviolet wavelengths, these cameras cannot employ lenses, which would be essentially opaque. Instead the projectors must use rather sophisticated mirrors. For the same reason, the masks must be quite different from the glass screens used in conventional lithography. Extreme ultraviolet work demands masks

SLICING A NANOCHIP

SILICON-ON-INSULATOR technology, which has helped improve chip performance considerably, has become cheaper and easier to adopt, thanks to a technique called Smart Cut, developed by Soitec, a French company.



BRYAN CHRISTIE DESIGN

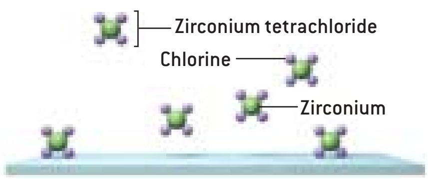
THE AUTHOR

G. DAN HUTCHESON is chief executive officer and president of VLSI Research Inc., a market research and economic analysis firm serving the semiconductor industry. Hutcheson, who holds a master's degree in economics from San Jose State University, has constructed various quantitative models that chipmakers can use to forecast costs and to guide them in procuring equipment. As an industry analyst, he follows the emerging technologies of semiconductor manufacture and provides summaries of the latest research advances and manufacturing trends to interested companies.

that absorb and reflect light. To construct them, dozens of layers of molybdenum and silicon are laid down, each just a few nanometers thick. Doing so produces a highly reflective surface onto which a patterned layer of chromium is applied to absorb light in just the appropriate places.

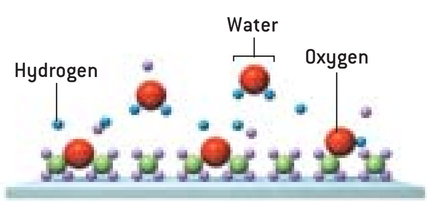
As with other aspects of chipmaking, these masks must be free from imperfections. But because the wavelengths are so small, probing for defects proves a considerable challenge. Scientists and engineers from industry, academe and government laboratories from across the U.S. and Europe are collaboratively seeking solutions to this and other technical hurdles that must be overcome before extreme ultraviolet lithography becomes practical. But for the time being, chipmakers must accept the limits of conventional lithography and maintain feature sizes of at least 50 nanometers or so.

ATOMIC-LAYER DEPOSITION allows chipmakers to lay down coatings that are extremely thin. Cycling through these steps repeatedly builds up the coating—one molecule of thickness at a time.

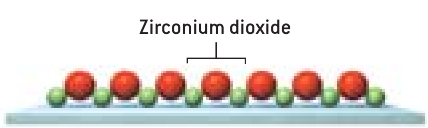


1 The surface is exposed to the first of two gases, here zirconium tetrachloride ($ZrCl_4$).

2 Molecules of $ZrCl_4$ adhere to the surface but not to one another.



3 The coated surface is exposed to a second gas, in this case steam (H_2O).



4 The $ZrCl_4$ on the surface reacts with the water (H_2O) to form a single-molecule-thick veneer of the desired material, zirconium dioxide (ZrO_2).

BRYAN CHRISTIE DESIGN

Using lithography to imprint such features on a film of photoresist is only the first in a series of manipulations used to sculpt the wafer below. Process engineers must also figure out how to remove the exposed parts of the photoresist and to etch the material that is uncovered in ways that do not eat into adjacent areas. And one must be able to wash off the photoresist and the residues left over after etching—a mundane task that becomes rather complicated as the size of the features shrinks.

The problem is that, seen at the nanometer level, the tiny features put on the chip resemble tall, thin skyscrapers, separated by narrow chasms. At this scale, traditional cleaning fluids act as viscous tidal waves and could easily cause things to topple. Even if that catastrophe can be avoided, these liquids have a troubling tendency to get stuck in the nanotechnology canyons.

An ingenious solution to this problem emerged during the 1990s from work done at Los Alamos National Laboratory: supercritical fluids. The basic idea is to use carbon dioxide at elevated pressure and temperature, enough to put it above its so-called critical point. Under these conditions, CO_2 looks something like a liquid but retains an important property of a gas—the lack of viscosity. Supercritical carbon dioxide thus flows easily under particles and can mechanically dislodge them more effectively than can any wet chemical. (It is no coincidence that supercritical carbon dioxide has recently become a popular means to dry-clean clothes.) And mixed with the proper co-solvents, supercritical carbon dioxide can be quite effective in dissolving photoresist. What is more, once the cleaning is done, supercritical fluids are easy to remove: lowering the pressure—say, to atmospheric levels—causes them to evaporate away as a normal gas.

With the wafer cleaned and dried in this way, it is ready for the next step: adding the junctions of the transistors—tubs on either side of the gate that serve as the current "source" and "drain." Such junctions are made by infusing the silicon with trace elements that transform it from a semiconductor to a conductor. The usual tactic is to fire arsenic or boron ions into the surface of the silicon using a device called an ion implanter. Once emplaced, these ions must be "activated," that is, given the energy they need to incorporate themselves into the crystal lattice. Activation requires heating the silicon, which often has the unfortunate consequence of causing the arsenic and boron to diffuse downward.

To limit this unwanted side effect, the temperature must be raised quickly enough that only a thin layer on top heats up. Restricting the heating in this way ensures that the surface will cool rapidly on its own. Today's systems ramp up and down by thousands of degrees a second. Still, the arsenic and boron atoms diffuse too much for comfort, making the junctions thicker than desired for optimum speed. A remedy is, however, on the drawing board—laser thermal processing, which can vary the temperature at a rate of up to five *billion* degrees a second. This technology, which should soon break out of the lab and onto the factory floor, holds the promise of preventing virtually all diffusion and yielding extremely shallow junctions.

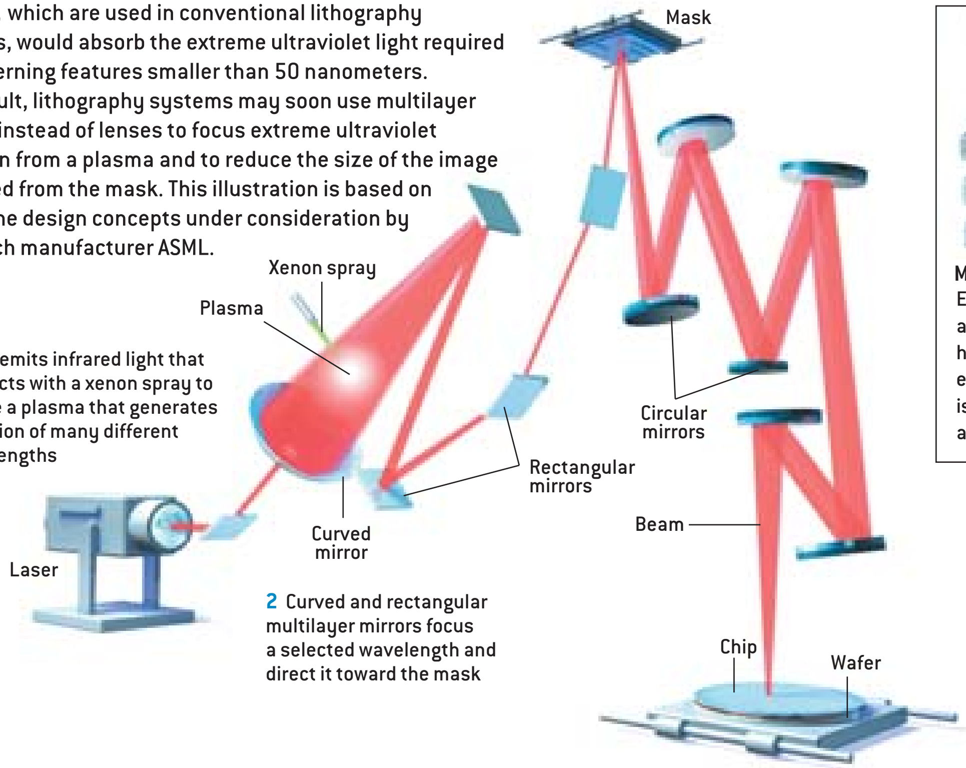
Once the transistors are completed, millions of capacitors are often added to make dynamic random-access memory, or DRAM. The capacitors used for DRAM have lately become so small that manufacturing engineers are experiencing the same kinds of problems they encounter in fashioning transistor gates. Indeed, here the problems are even more urgent, and the answer, again, appears to be atomic-layer deposition, which was adopted for the production of the latest generation of DRAM chips.

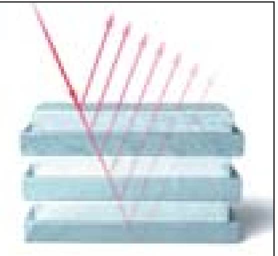
New Meets Old

ATOMIC-LAYER DEPOSITION can also help in the next phase of chip manufacture, hooking everything together. The procedure is to first lay down an insulating layer of glass on which a pattern of lines is printed and etched. The grooves are then filled with metal to form the

EXTREME ULTRAVIOLET LITHOGRAPHY

LENSES, which are used in conventional lithography systems, would absorb the extreme ultraviolet light required for patterning features smaller than 50 nanometers. As a result, lithography systems may soon use multilayer mirrors instead of lenses to focus extreme ultraviolet radiation from a plasma and to reduce the size of the image projected from the mask. This illustration is based on one of the design concepts under consideration by the Dutch manufacturer ASML.





MULTILAYER MIRROR
Each layer reflects only a small amount of the light hitting it. Yet the cumulative effect of the many layers is sufficient to create an effective reflector.

wires. These steps are repeated to create six to eight layers of crisscrossing interconnections. Although the semiconductor industry has traditionally used aluminum for this bevy of wires, in recent years it has shifted to copper, which allows the chips to operate faster and helps to maintain signal integrity. The problem is that copper contaminates the junctions, so a thin conductive barrier (one that does not slow the chip down) needs to be placed below it. The solution was atomic-layer deposition.

The switch to copper also proved challenging for another reason: laying down copper is inherently tricky. Many high-tech approaches were attempted, but none worked well. Then, out of frustration, engineers at IBM tried an old-fashioned method: electroplating, which leaves an uneven surface and has to be followed with mechanical polishing. At the time, the thought of polishing a wafer—that is, introducing an abrasive grit—was anathema to managers in this industry, which is downright obsessed with cleanliness. Hence, the engineers who originally experimented with this approach at IBM did so without seeking permission from their supervisor. They were delighted to discover that the polishing made the wafer more amenable to lithographic patterning (because the projection equipment has a limited depth of focus), that it removed troublesome defects from the surface and that it made it easier to deposit films for subsequent processing steps.

The lesson to be learned here is that seemingly antiquated methods can be just as valuable as cutting-edge techniques. Indeed, the semiconductor industry has benefited a great deal in recent years from combinations of old and new. That it has advanced as far as it has is a testament to the ingenious ability of countless scientists and engineers to continually refine the basic method of chip manufacture, which is now more than four decades old.

Will the procedures used for fabricating electronic devices four decades down the road look anything like those currently employed? Although some futurists would argue that exotic forms of nanotechnology will revolutionize electronics by midcentury, I'm betting that the semiconductor industry remains pretty much intact, having by then carried out another dazzling series of incremental technical advances, ones that are today beyond anyone's imagination. SA

MORE TO EXPLORE

Toward Point One. Gary Stix in *Scientific American,* Vol. 272, No. 2, pages 90–95; February 1995.

Technology and Economics in the Semiconductor Industry. G. Dan Hutcheson and Jerry D. Hutcheson in *Scientific American,* Vol. 274, No. 1, pages 54–62; January 1996.

Handbook of Semiconductor Manufacturing Technology. Edited by Yoshio Nishi and Robert Doering. Marcel Dekker, 2000.

2003 International Technology Roadmap for Semiconductors. Available online at **http://public.itrs.net/Files/2003ITRS/Home2003.htm**

News from International SEMATECH, a global consortium of leading semiconductor manufacturers, is available online at **www.sematech.org**

BRYAN CHRISTIE DESIGN

THREE-“LETTER” SEQUENCES, or codons, of DNA and RNA (*shown*) encode the individual amino acids that build and maintain all life on earth.

Evolution Encoded

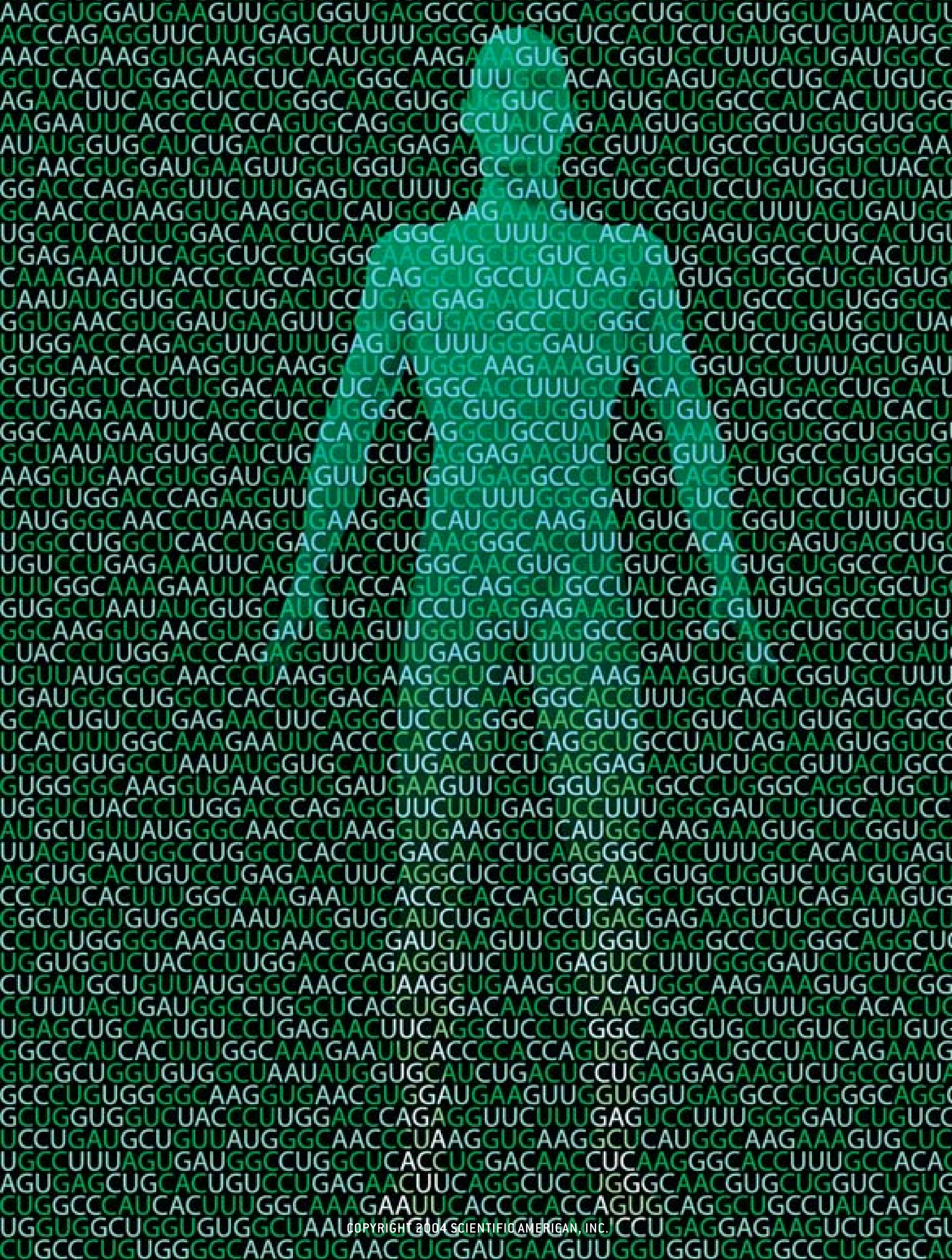
New discoveries about the rules governing how genes encode proteins have revealed nature’s sophisticated “programming” for protecting life from catastrophic errors while accelerating evolution

On April 14, 2003, scientists announced to the world that they had finished sequencing the human genome—logging the three billion pairs of DNA nucleotides that describe how to make a human being. But finding all the working genes amid the junk in the sequence remains a further challenge, as does gaining a better understanding of how and when genes are activated and how their instructions affect the behavior of the protein molecules they describe. So it is no wonder that Human Genome Project leader Francis S. Collins has called the group’s accomplishment only “the end of the beginning.”

Collins was also alluding to an event commemorated that same week: the beginning of the beginning, 50 years earlier, when James D. Watson and Francis H. Crick revealed the structure of the DNA molecule itself. That, too, was an exciting time. Scientists knew that the molecule they were finally able to visualize contained nothing less than the secret of life, which permitted organisms to store themselves as a set of blueprints and convert this stored information back into live metabolism. In subsequent years, attempts to figure out how this conversion took place captivated the scientific world. DNA’s alphabet was known to consist of only four types of nucleotide. So the information encoded in the double helix had to be *decoded* according

SLIM FILMS

By Stephen J. Freeland and Laurence D. Hurst



to some rules to tell cells which of 20 amino acids to string together to constitute the thousands of proteins that make up billions of life-forms. Indeed, the entire living world had to be perpetually engaged in frenetic decryption, as eggs hatched, seeds germinated, fungus spread and bacteria divided.

But so little was understood at the time about the cellular machinery translating DNA's message that attempts to crack this genetic code focused on the mathematics of the problem. Many early proposals proved wrong, a few spectacularly, although their sheer ingenuity and creativity still provide fascinating reading. In fact, when the actual code was finally deciphered during the 1960s, it nearly disappointed. Nature's version looked less elegant than several of the theorists' hypotheses.

Only in recent years have new discoveries about the code revealed just how sophisticated a piece of programming it really is. Why nature chose these basic rules and why they have survived three billion or so years of natural selection have started to become clear. We can now show that the code's rules may actually speed evolution while protecting life from making disastrous errors in protein synthesis. Studying the code is also providing clues to solving some of those remaining challenges facing laboratories in the post-genome era. In going back to the *very* beginning to understand the rules of life's underlying code, we are discovering tools for future research.

When we speak of the "code" and "decoding," we are being quite literal. Genetic instructions are stored in DNA and RNA, both made of one type of biochemical molecule, nucleic acid. But organisms are mostly built from (and by) a very different type of molecule, protein. So although a gene is traditionally defined as the sequence of nucleotides that describes a single protein, the genetic sentence containing that description must first be translated from one system of symbols into an entirely different kind of system, rather like converting from Morse code to English.

CCG

Certain codons were just redundant. MANY CAME TO VIEW NATURE'S REAL CODE as little more than a random accident.

Cracking the Code

WHEN WATSON and Crick described DNA's structure in 1953, they and their contemporaries could see that genes are written in an alphabet of just four "letters"—the bases adenine, cytosine, guanine and thymine (A, C, G and T) that distinguish each nucleotide and form the rungs in DNA's now familiar twisted-ladder shape. The protein alphabet, in contrast, contained 20 different amino acids, so the need for a multinucleotide genetic "word" to specify any given amino acid was obvious. Two-letter combinations of the four bases would yield only 16 possible words, or "codons." But triplet combinations produce 64 possible codons, which would be plenty.

Little else was obvious at the time about how genes might be translated into proteins. Today we understand that gene sequences do use three-letter codons to specify individual amino acids and that several steps are needed for the gene's sequence of bases to be converted into a sequence of amino acids. The DNA gene is first copied and edited into a transcript made of RNA, employing similar nucleic acid bases, except that DNA's thymine is replaced by uracil. This messenger RNA (mRNA) version of the gene is then read by cellular machinery, three letters at a time, while tiny cellular butlers known as transfer RNAs (tRNA) fetch the specified amino acids to be strung together.

But in the early 1950s this process was a black box, leaving only an intriguing mathematical puzzle. And the first proposed solution came not from a biologist but from physicist George Gamow, better known as an originator of the big bang theory. His "diamond code," published in 1954, elegantly combined the arithmetic of getting 20 amino acid meanings from a four-nucleotide alphabet with the physical structure of DNA itself.

Gamow theorized that at every turn in the double helix there was a diamond-shaped space bounded at its four corners by nucleotides. These gaps would allow DNA to act as a template against which amino acids would line up, determined by the nucleotide combinations present at each twist. His model eliminated one corner of each diamond, then sorted the 64 possible three-nucleotide codons into chemically related groups. It also allowed meaningful codons to overlap, depending on the "reading frame," or where one began reading the sequence of letters along the length of the DNA molecule. This kind of data compression was an efficien-

Overview/*Life's Code*

- Genetic instructions for the manufacture of proteins are written in three-letter "words" called codons, each specifying one of 20 amino acids or a "stop translating" sign. The arrangement of these codons and their amino acid meanings was once considered random, but recent discoveries indicate that natural selection has chosen and maintained this order.
- Computer simulations reveal why: compared with hypothetical alternatives, the standard code is exceptionally good at minimizing the harm caused by errors in genes themselves or in the process of translating genes into proteins.

NATURE'S CODE

IF A GENE SEQUENCE is a "sentence" describing a protein, then its basic units are three-letter "words," or "codons," each of which translates into one of 20 amino acids or a "stop translating" signal. Cellular machinery transcribes DNA genes into RNA versions—whose nucleotide building blocks are represented by the letters A, C, G and U—and then translates the RNA genes, codon by codon, into a corresponding amino acid sequence. Nature's exact amino acid definitions (*below*) were worked out during the early1960s. But the significance of patterns in the code would not be fully appreciated for several decades.

SYNONYMS AND SIMILARITIES

Many of the 64 possible three-letter codons specify the same amino acid, providing alternative ways for genes to spell out most proteins. These synonymous codons tend to differ by just a single letter, usually the last, forming a pattern of blocks. Codons for amino acids with similar affinities for water also tend to differ by their last letter, and codons sharing the same first letter often code for amino acids that are products or precursors of one another. These features, as it turns out, are crucial to the survival of all organisms and may even help speed their evolution.

First Nucleotide Position	Second Nucleotide Position: U	C	A	G
U	UUU Phenylalanine	UCU Serine	UAU Tyrosine	UGU Cysteine
U	UUC Phenylalanine	UCC Serine	UAC Tyrosine	UGC Cysteine
U	UUA Leucine	UCA Serine	UAA STOP	UGA STOP
U	UUG Leucine	UCG Serine	UAG STOP	UGG Tryptophan
C	CUU Leucine	CCU Proline	CAU Histidine	CAU Arginine
C	CUC Leucine	CCC Proline	CAC Histidine	CAC Arginine
C	CUA Leucine	CCA Proline	CAA Glutamine	CAA Arginine
C	CUG Leucine	CCG Proline	CAG Glutamine	CAG Arginine
A	AUU Isoleucine	ACU Threonine	AAU Asparagine	AAU Serine
A	AUC Isoleucine	ACC Threonine	AAC Asparagine	AAC Serine
A	AUA Isoleucine	ACA Threonine	AAA Lysine	AAA Arginine
A	AUG Methionine	ACG Threonine	AAG Lysine	AAG Arginine
G	GUU Valine	GCU Alanine	GAU Aspartate	GAU Glycine
G	GUC Valine	GCC Alanine	GAC Aspartate	GAC Glycine
G	GUA Valine	GCA Alanine	GAA Glutamate	GAA Glycine
G	GUG Valine	GCG Alanine	GAG Glutamate	GAG Glycine

cy prized by coding theorists of the day. Unfortunately, amino acid chains were soon discovered that could not be accounted for by Gamow's or any other overlapping codes.

At the same time, evidence was suggesting that DNA and amino acids were not interacting with one another directly. Crick developed a hypothesis that so-called adaptor molecules could be serving as intermediaries, and in 1957 he put forth a set of rules by which they might operate. Simply put, Crick's adaptors recognized only 20 meaningful codons designating each of the 20 amino acids, making the remainder of the 64 possible triplets "nonsense." Rather than overlapping, Crick's code was "commaless": meaningless codons were effectively invisible to the adaptors, so nature needed no figurative punctuation to designate the start of a reading frame. The commaless concept was so streamlined that it immediately won near universal acceptance—that is, until the data again proved an elegant theory wrong.

In the early 1960s experiments showed that even supposed nonsense codons could provoke protein synthesis in a beaker, and by 1965 the actual amino acid meanings of all 64 possible triplet codons had been worked out in the lab. No tidy numerology was apparent: certain codons were just redundant, with some individual amino acids specified by two, four, even six different codons. After all the enthusiastic speculation, many came to view nature's real code as little more than a random accident of history.

Frozen Accident?

INDEED, ONCE THE CODE was deciphered, scientists found that organisms as different as humans and bacteria employed the exact same coding rules. Seemingly no variations had occurred in the billions of years since the three basic domains of life—archaea, bacteria and eukaryotes—diverged from a single ancient common ancestor. Consequently, the simple and persuasive "frozen accident" argument, put forth by Crick himself in 1968, came to dominate scientific thinking until recently. "The allocation of codons to amino acids at this point was entirely a matter of chance," he wrote, but once the code had appeared in any form, it was so fundamental to life that any further changes would have been catastrophic.

Darwinian natural selection rests on the premise that sometimes a small change in a single gene can prove beneficial if it allows organisms to fare better in their environment. But altering an organism's decoding rules would be tantamount to simultaneously introducing changes at countless sites throughout its genetic material, producing an utterly dysfunctional metabolism. It would be the difference between introducing a single typo and rewiring the entire typewriter keyboard.

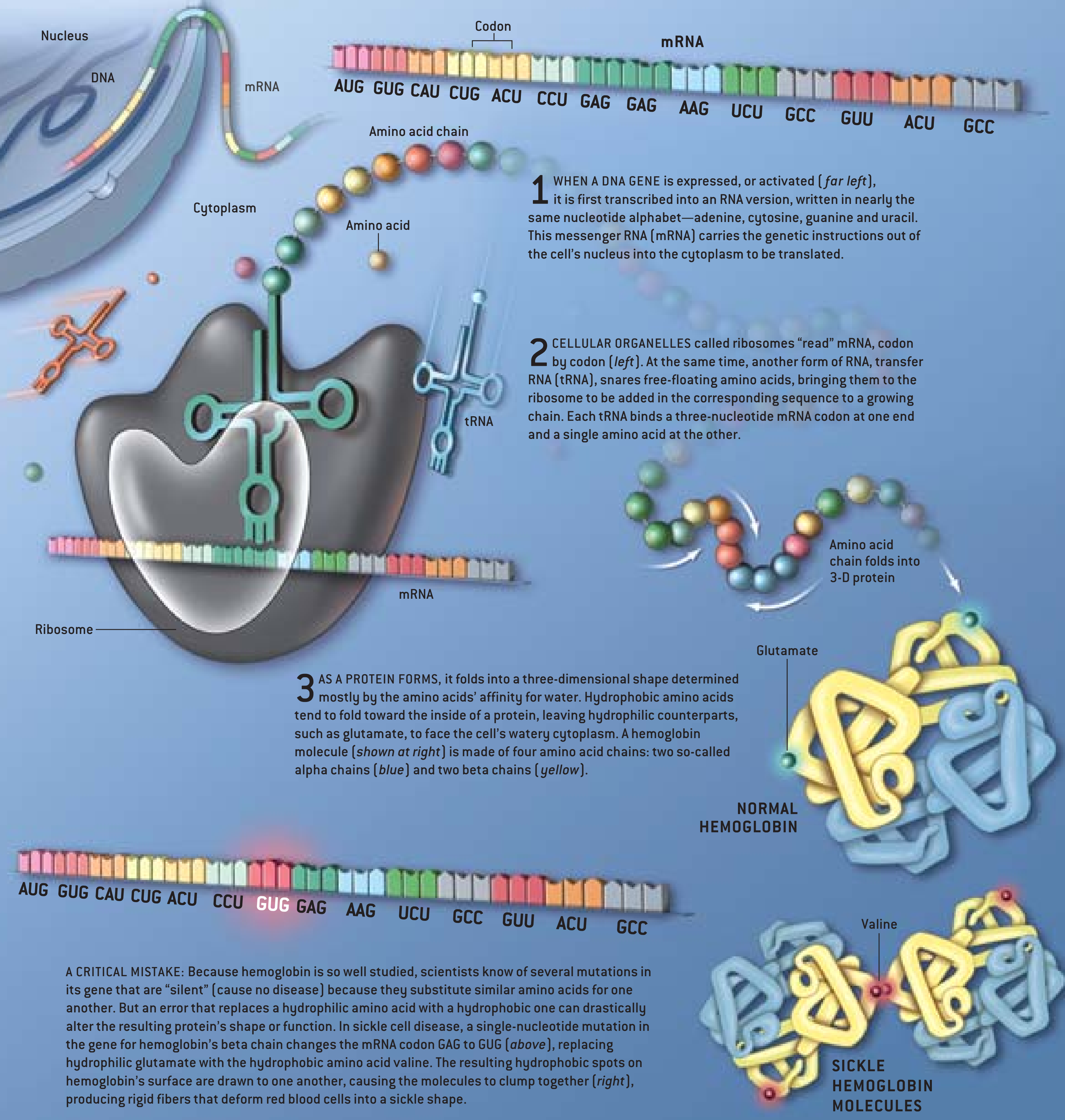
This attractively straightforward reasoning, however, has since proved simplistic. Although most living systems do employ the standard genetic code, scientists now know of at least 16 variants, distributed across a diverse array of evolutionary lineages, that assign different meanings to certain codons. The underlying system remains the same: triple-nucleotide codons are translated into amino acids. But where most organisms would read the RNA codon "CUG" to mean the amino acid leucine, many species of the fungus *Candida* translate CUG as serine. Mitochondria, the tiny power generators within all kinds of cells, have their own

LUCY READING

Protecting Proteins

Nature's code minimizes the effects of genetic errors, whether resulting from mutations in genes themselves or mistakes in the translation process. A gene sequence is translated into a corresponding amino acid sequence, which dictates the final three-dimensional structure of the encoded protein (*1, 2* and *3*).

Even when an error inserts the wrong amino acid, the code's arrangement ensures that the substitute is often chemically similar to the intended amino acid, and the final protein is relatively unaltered. An exception that illustrates the kind of harm a single-nucleotide error can cause is sickle cell disease (*at bottom*).



1 WHEN A DNA GENE is expressed, or activated (*far left*), it is first transcribed into an RNA version, written in nearly the same nucleotide alphabet—adenine, cytosine, guanine and uracil. This messenger RNA (mRNA) carries the genetic instructions out of the cell's nucleus into the cytoplasm to be translated.

2 CELLULAR ORGANELLES called ribosomes "read" mRNA, codon by codon (*left*). At the same time, another form of RNA, transfer RNA (tRNA), snares free-floating amino acids, bringing them to the ribosome to be added in the corresponding sequence to a growing chain. Each tRNA binds a three-nucleotide mRNA codon at one end and a single amino acid at the other.

3 AS A PROTEIN FORMS, it folds into a three-dimensional shape determined mostly by the amino acids' affinity for water. Hydrophobic amino acids tend to fold toward the inside of a protein, leaving hydrophilic counterparts, such as glutamate, to face the cell's watery cytoplasm. A hemoglobin molecule (*shown at right*) is made of four amino acid chains: two so-called alpha chains (*blue*) and two beta chains (*yellow*).

A CRITICAL MISTAKE: Because hemoglobin is so well studied, scientists know of several mutations in its gene that are "silent" (cause no disease) because they substitute similar amino acids for one another. But an error that replaces a hydrophilic amino acid with a hydrophobic one can drastically alter the resulting protein's shape or function. In sickle cell disease, a single-nucleotide mutation in the gene for hemoglobin's beta chain changes the mRNA codon GAG to GUG (*above*), replacing hydrophilic glutamate with the hydrophobic amino acid valine. The resulting hydrophobic spots on hemoglobin's surface are drawn to one another, causing the molecules to clump together (*right*), producing rigid fibers that deform red blood cells into a sickle shape.

genomes, and many have also developed their own codon assignments. For instance, in the mitochondrial genome of baker's yeast (*Saccharomyces cerevisiae*), four of the six codons that normally encode leucine instead encode threonine.

As discoveries of these variations proliferated during the 1990s, it became clear that the code is not frozen at all. It can evolve, which means that it probably *did* evolve. So nature's standard codon–amino acid assignments, refined and preserved by billions of years of natural selection, are no accident. In fact, their arrangement does an excellent job of minimizing the impact of accidents.

Damage Control

EVERY CODING SYSTEM has to contend with the possibility of mistakes, but not all errors are equally damaging. In English, vowels and consonants are very different, so that replacement of "s" by "a" makea thia meaaage aignificantly leaa eaay to underatand. In contrast, the letters "s" and "z" have a similar sound, zuch that thiz phraze remainz eazily underztandable. For an error-prone system, a good coding strategy would be one that reduces the effect of the inevitable occasional mistake.

In a living organism, errors come in many forms. Sometimes the original DNA version of a gene changes (a mutation). Sometimes the wrong adaptor (tRNA) binds to the mRNA transcript of a gene, misincorporating an amino acid into a nascent protein [*see box on opposite page*]. But even when scientists considered the code a product of chance, they noticed that it did seem to be arranged well in terms of ensuring that individual errors are of little consequence. As early as 1965 Carl R. Woese, then at the University of Illinois, observed that similar codons (those sharing two of three letters) usually specify similar amino acids, so a mistake here or there does not greatly affect the resulting protein.

The code can evolve, which means that it probably did evolve. NATURE'S AMINO ACID ASSIGNMENTS are no accident.

CCG

Defining "similar" with regard to amino acids can be complex: the 20 amino acids differ from one another in all sorts of properties, from size to shape to electric charge. What Woese and others noted is that codons sharing two out of three bases tend to code for amino acids that are much alike in the extent to which they are repelled by or attracted to water. This property is crucial to the ultimate functioning of the protein. A newly made amino acid chain folds into a distinctive shape depending on the positioning of hydrophobic amino acids, which like to cluster together away from the cell's watery cytoplasm, leaving hydrophiles to form the protein's surface.

The remarkable feature of the genetic code is that when a single-nucleotide error occurs, the actual and intended amino acids are often similar in hydrophobicity, making the alteration in the final protein relatively harmless. But just *how* efficient is the code in this regard? This is where, in 1998, we stepped in to develop the observations of earlier scientists.

Testing the Code

FIRST WE TOOK a quantitative measure of the 20 amino acids' hydrophobicity. Next we used those values to calculate the genetic code's error value, which we define as the average change in the resulting amino acids' hydrophobicity caused by all possible single-letter changes to all 64 codons of the code. This value represents the genetic code's susceptibility to errors but is of little meaning on its own. We needed to know how nature's coding system stacks up against possible alternatives.

To generate these hypothetical alternative codes, we had to begin with certain assumptions about realistic restrictions under which a code would operate in a world made of DNA, RNAs and amino acids. One observation is that mistakes in translation of mRNA into a corresponding amino acid occur most frequently at the codon's third position. This spot is simply where the binding affinity between the mRNA and tRNA is weakest, which is why Crick dubbed the phenomenon "wobble." But synonymous codons—those coding for the same amino acid—usually differ by only their last letters, so such mistranslations often yield the same amino acid meaning.

Although this grouping of synonymous codons in itself reduces the error value of the code, the mechanics of wobble make the arrangement more likely to be a biochemical limitation rather than an evolutionary adaptation. Thus, to err on the side of caution when deriving our measure, we should consider only alternative codes that share this feature. Moreover, it is impossible to put a hydropho-

THE AUTHORS

STEPHEN J. FREELAND and *LAURENCE D. HURST* use bioinformatics to study evolutionary biology. Freeland is assistant professor of bioinformatics at the University of Maryland, Baltimore County, where he is working to convert insights about genetic code evolution into practical approaches to exploring genome data. His students are currently testing his theories by reengineering a human cancer gene to express in *Escherichia coli* according to that organism's codon preferences. Freeland earned his Ph.D. in evolutionary theory at the University of Cambridge, studying under Royal Society Research Fellow Laurence Hurst, who is now professor of evolutionary genetics at the University of Bath in England. Hurst's research concentrates on understanding the structure and evolution of genetic systems, particularly the evolutionary origin of phenomena such as sexes, the order of genes on chromosomes, genomic imprinting and the genetic code itself.

TAMI J. TOLPA (*opposite page*)

bicity value on the codons assigned to the "stop" signal, so we kept their number and codon assignments the same in all alternative codes.

Using these technical assumptions, we generated alternatives by randomizing the 20 meanings among the 20 codon blocks. This still defined some 2.5×10^{18} possible configurations (approximately equal to the number of seconds that have elapsed since the earth formed). So we took large random samples of these possibilities and found that from a sample of one million alternative codes only about 100 had a lower error value than the natural code [*see box below*].

Still more striking was our finding when we incorporated additional restrictions to reflect observed patterns in the way DNA tends to mutate and the ways in which genes tend to be mistranscribed into RNA. Under these "real world" conditions, the natural code's error value appeared orders of magnitude better still, outperforming all but one in a million of the alternatives.

A straightforward explanation for the genetic code's remarkable resilience is that it results from natural selection. Perhaps there were once many codes, all with different degrees of error susceptibility. Organisms whose codes coped best with error were more likely to survive, and the standard genetic code simply won in the struggle for existence. We know that variant codes are possible, so this assumption is reasonable.

Evidence for error minimization as the driving evolutionary force behind the arrangement of the code has its critics, however. Sophisticated computer searches can certainly improve on nature's choice, even when they accept the premise that a "good" code is one that minimizes the change in amino acid hydrophobicity caused by genetic errors. But computer predictions for an optimal code are limited to the criteria provided by the programmer, and most of the "better" codes that have been described so far are based on oversimplified assumptions about the types of errors that a code encounters in the real world. For example, they ignore the wobble phenomenon, which prevents their algorithms from perceiving the advantage of having synonymous codons differ only in their third letter.

This shortcoming emphasizes a second problem with designer-optimized codes. Natural selection is a "blind designer," in that it can only grope toward an ideal by choosing the best alternative within a population of variants at each generation. When we simulate natural selection in this manner, we find that the degree of error minimization achieved by the standard genetic code is still rather impressive: typically less than 3 percent of random theoretical codes can evolve under selection to match its resilience.

In other words, the diamond and commaless codes once looked superior to nature's own, and computers may generate yet more mathematically idealized codes. But merely demonstrating the possibility of better codes without taking into account the evolutionary process is of dubious relevance to understanding the strength of natural selection's choice.

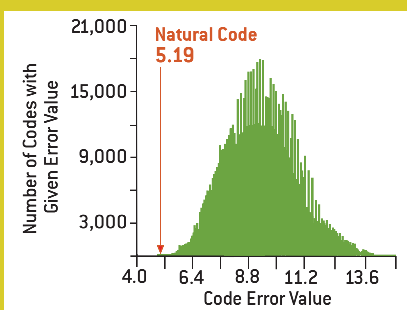
Indeed, the standard code is not only a product of natural selection; it may act as a search algorithm to speed evolution. The impact-minimizing properties of the code, with its blocks of both synonymous codons and those specifying biochemically similar amino acids, achieve more than damage control. "Smaller" mutations, in contrast with extreme alterations, are statistically more likely to be beneficial, so by minimizing the effects of any mutation, the code maximizes the likelihood that a gene mutation will lead to an improvement in the resulting protein.

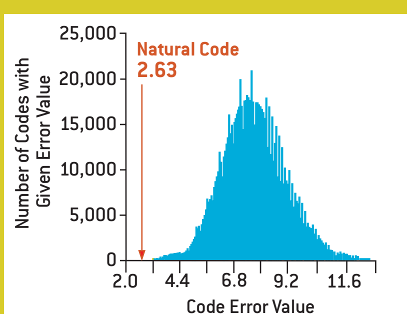
CCG

By minimizing the effects of any mutation, THE CODE MAXIMIZES THE LIKELIHOOD that a gene mutation will improve the resulting protein.

Testing the Code

DAMAGE TO PROTEINS resulting from gene mutations or mistranslations is minimized when the errors substitute amino acids with similar affinities for water (hydrophobicity). If we define a code's error value as the average change in amino acid hydrophobicity caused by all possible single-base changes within all codons of a code, then a high error value indicates that a code is very vulnerable to errors, and a low value means that a code minimizes their harm. We generated a large random sample of possible codes and found that only 100 of one million alternatives had a lower error value than nature's code (*top*). When we factor in real-world patterns in the way genes mutate and are mistranslated, nature's code outperforms all but one in a million of the alternatives (*bottom*).





LUCY READING

Using the Code

UNDERSTANDING THE FORCES that shaped the code and how it in turn shapes evolution does more than provide an opportunity to admire nature's skill as a primordial software designer. These insights can also help solve some of the toughest problems facing laboratories in 2004.

Sifting through reams of raw genome sequence data to find the actual genes is a priority in molecular biology, but current searches are limited to matching the characteristics of genes that we already know about. Taking into account the way that the genetic code filters gene mutations can enhance these searches by allowing scientists to recognize highly diversified genes and perhaps infer the function of the proteins they encode. Researchers can even derive clues about the folded protein shape that an amino acid sequence dictates by looking at the error-minimizing properties of its codons and how substitutions might affect amino acid size, charge or hydrophobicity.

Biologists can also apply our awareness of organisms that deviate from the standard code to "disguise" genes for research. Because a single code is nearly universal to all life, it has become common practice to take a gene of interest, such as a human cancer gene, and insert it into an organism, such as *Escherichia coli,* that will churn out the protein the gene encodes. But occasionally the organism fails to express the gene at all, or it produces less of the protein than expected or a slightly different version of the protein found in humans.

This problem can play havoc with biology research, but we now realize that sometimes the failure arises because the organisms exhibit different preferences among synonymous codons. For example, the standard code contains six codons for the amino acid arginine, and human genes tend to favor using the codons AGA and AGG. *E. coli,* however, very rarely uses AGA and often mistranslates it. Knowing these variations and preferences enables us to design versions of the human gene that will work reliably when moved between different organisms.

One of our labs (Freeland's) is developing software applications to help molecular biologists turn such theoretical observations about the code into practical tools for genetic engineering, gene finding, and predicting protein shapes. And both of us, along with other researchers, are investigating how the code itself came to be—how RNA first started interacting with amino acids, how their association developed into a system of formal coding and how the amino acid alphabet expanded during early evolution.

This approach may allow inroads into many additional unresolved questions: Why 20 and only 20 standard amino acids? Why are some amino acids assigned six codons, whereas others have just one or two? Could this pattern have anything to do with minimizing error? Cracking the code has proved merely the start to understanding its meaning. SA

The Evolving Code

At least 16 organisms from a diverse array of evolutionary lineages deviate from nature's standard code in the amino acid "meaning" they assign to specific codons. Many species of the green alga *Acetabularia,* for example, translate the standard "stop" codons UAG and UAA as the amino acid glycine. To *Candida* fungi, the RNA codon CUG, which normally means leucine, instead specifies serine. The existence of such variations demonstrates that the code can evolve and may provide clues about how it did. In all three domains of life, a nonstandard 21st amino acid, selenocysteine, is sometimes fabricated in response to the standard stop codon UGA. Selenocysteine is created by chemical tweaking of serine while that amino acid is still attached to its tRNA in the ribosome. In two domains (archaea and bacteria), a 22nd amino acid, pyrrolysine, is produced in the same manner, in response to the standard stop codon UAG.

MARINE ALGA *Acetabularia* can grow up to five centimeters tall, but each stalk is a single cell—the largest known to science.

The code used by early life probably did not specify as many as 20 amino acids. Indeed, the more complex ones are produced solely as biologically modified derivatives of simpler ones. In several bacterial species, for example, the amino acid glutamine is produced from its biochemical cousin glutamate while the latter is still attached to its tRNA. This phenomenon suggests that novel amino acids may have arisen as modifications to a smaller primordial set and that the newcomers "captured" a subset of the tRNAs and codons assigned to their chemically simpler relatives, just as certain codons appear to have been captured by standard amino acids in the organisms known to employ variant codes. These discoveries raise the question of how many more variant codes may be out there and whether the standard code will eventually expand to contain far more amino acids. *—S.J.F. and L.D.H.*

MORE TO EXPLORE

A Quantitative Measure of Error Minimization in the Genetic Code. David Haig and Laurence D. Hurst in *Journal of Molecular Evolution,* Vol. 33, No. 5, pages 412–417; November 1991.

The Invention of the Genetic Code. Brian Hayes in *American Scientist,* Vol. 86, No. 1, pages 8–14; January-February 1998.

The Genetic Code Is One in a Million. Stephen J. Freeland and Laurence D. Hurst in *Journal of Molecular Evolution,* Vol. 47, No. 3, pages 238–248; September 1998.

The Case for an Error-Minimizing Genetic Code. Stephen J. Freeland, Nick Keulmann and Tao Wu in *Origins of Life and Evolution of the Biosphere,* Vol. 33, Nos. 4–5, pages 457–477; October 2003.

More information about the evolution of the genetic code can be found on Stephen J. Freeland's laboratory Web site at **www.evolvingcode.net**

BLASTOFFS ON A BUDGET

Private ventures seeking to make routine access to space affordable see big potential in going small

By Joan C. Horvath

HYBRID ROCKET ENGINE on Scaled Composites's SpaceShipOne burned for 15 seconds during its first powered aerial test in mid-December (*above* and *opposite page*). The suborbital spacecraft thrust its way to almost Mach 1.2 and then coasted to a 68,000-foot altitude.

SpaceShipOne test flight

Well into the 1860s, the American West remained divided from the East by the harsh terrain of the country's broad, untamed interior, particularly the steep peaks of the Sierra Nevada. Then four Sacramento merchants began raising money to fund a seemingly impossible project: to build a railroad across the high Sierras and thus unite the continent. Derided by the press, the moneymen, top engineers and politicians, the ambitious enterprise nonetheless overcame daunting technical obstacles and eventually succeeded. The so-called Big Four investors, Collis P. Huntington, Mark Hopkins, Charles Crocker and Leland Stanford, became enormously rich as settlers arrived in the newly opened lands. Railroads prospered from the short-haul traffic for a burgeoning population they helped to create.

Today a group of entrepreneurs has a comparable but loftier aim: to provide cheap, reliable transportation to low Earth orbit. Their high-flying goal comes with similarly steep challenges. Like the rail pioneers, private rocket builders are trying to create a market where none currently exists while keeping costs affordable. Further, they must develop a regular taxi service to space that is sufficiently safe to attract customers. (It will be some time before any flights to space will be as safe as passenger airline flights, though.) Finally, the entrepreneurs must surmount evolving governmental regulatory hurdles.

SCALED COMPOSITES

Unlike the established Big Aerospace companies, which have traditionally focused on meeting objectives set by the military and government agencies that emphasize top performance over monetary concerns, the new rocket firms see large potential in thinking small: they seek more modest-size payloads that weigh as little as a few kilograms and launch costs measured in hundreds of thousands of dollars rather than tens of millions.

Of course, the unforgiving physics of orbital launch makes getting into space extremely difficult, no matter whether one goes big or small. So much energy must be expended to break the bonds of Earth's gravity that current launchers can barely carry enough fuel to get to orbit. This requirement means that much of the vehicle structure is thrown away after a single use. So far this reality has limited spaceflight to multistage expendable rockets or the space shuttle, which is assisted by strap-on fuel tanks and solid-fuel boosters that are jettisoned during ascent. Reusable spacecraft are seen as the way to keep launch costs under control and to make regular scheduled service possible.

The "small" space industry gained new vigor in 1995, when the $10-million X Prize competition was announced to spur the development of a reusable suborbital launch vehicle. Modeled on the Orteig Prize that Charles Lindbergh won by flying nonstop across the Atlantic, the award is intended to push private space entrepreneurs toward a goal that is easier to achieve than the orbital launch for which they had been aiming. "We redefined what was valuable" to lower the barrier to entry, says X Prize founder and president Peter H. Diamandes. The award will go to the team that first takes three people to a 100-kilometer suborbital altitude, returns to Earth, and repeats the feat within 14 days after replacing less than 10 percent of the dry weight of the vehicle. The race, Diamandes says, is forcing contestants to work through the many regulatory and liability issues related to private spaceflight and to determine real-world costs. More than 20 companies have entered the race officially, and about a dozen are building hardware. Anticipation is high that a winner may soon emerge (perhaps even by the time this article appears).

In the near term, suborbital tourism for the affluent is seen by many as the best market opportunity. For several years, Space Adventures in Arlington, Va., has been accepting deposits for seats on future private spaceflights it hopes might take place by 2005. The company has already made it possible for some 2,000 enthusiasts to take wild rides in Russian MiG fighters, experience zero-g conditions in parabolic flights onboard an Ilyushin 76 and, in two cases, stay in the International Space Station. In 2000 Space Adventures commissioned a study by Harris Interactive that estimates that more than 100,000 people in the U.S. and Canada were prepared to pay $100,000 apiece to fly to suborbital altitudes. Another study, conducted by Maryland-based consulting firm Futron, projects 15,000 suborbital tourists annually by 2021, representing a $700-million market.

Overview/*Access to Space*

- A private-sector space-launch industry is slowly rising. Prototype vehicles intended to provide routine, low-cost access to suborbital altitudes are being flight-tested.
- Highly capable small satellites and payloads, now under development, could provide an early market for these smaller rocket launchers.
- Stiff challenges from regulatory, financial and legal liability issues may prove more difficult to overcome than technological ones.

WHITE KNIGHT, the mother ship for SpaceShipOne (*seen attached below the large airplane's fuselage*), will release the three-seat spacecraft at about 50,000 feet. Scaled Composites's rocket will then climb to suborbital altitude.

SCALED COMPOSITES

Thinking Outside the Gantry

PRIVATE LAUNCH CAPABILITIES have been a long time coming. During the past two decades, firms have created a veritable armada of space vehicle designs, and some reached the prototype stage. The design concepts have been diverse: aircraft that drop rockets in flight or tow a rocket plane; rockets carried to altitude by balloons; and piloted, first- and second-stage winged boosters. The planned propellants have included kerosene, alcohol, peroxide and rubbery solids. And the range of reentry schemes has been similarly wide: parachutes, helicopterlike rotating blades and even crushable landing structures. Why, then, isn't the sky dark with launch vehicles?

Mostly, it's about the money. At the peak of the economic boom of the 1990s, some private space ventures secured significant funding resources. Often these were the companies with the most elaborate technology, because that factor fit the funding criteria investors used at the time. Other companies, born of the traditional aerospace industry culture in which no metal is cut until the design plan is complete, spent their early venture capital on design and could not attract money to actually build anything. "The little guys have a lot of fun and early rapid progress, but they run out of capital and/or energy before they achieve significant programmatic success," says John Jurist, an investor in the private space-access business.

Government-funded projects to develop a reusable single-stage-to-orbit space-transport system stalled in the 1990s, even though McDonnell Douglas demonstrated its feasibility with the Delta Clipper Experimental (DC-X) vehicle. The DC-X showed promise in tests when it took off vertically, hovered, moved sideways and then landed upright. Unfortunately, funding responsibility switched from the Strategic Defense Initiative Organization to NASA as part of a change in space policy. The DC-X then became a test bed for an even higher-risk project to build a low-cost, reusable single-stage launch vehicle. Despite the efforts of the contractor, Lockheed Martin, NASA's X-33 rocket plane program was never able to overcome problems with its advanced fuel tank and engine. The X-33 was canceled in 2001. Around the same time, the government contract for the X-34, a reusable vehicle designed to fly to orbit after being dropped from an L-1011 airliner, was allowed to expire. More than $1 billion went toward these defunct projects.

Why isn't the sky dark with launch vehicles? MOSTLY, IT'S ABOUT THE MONEY.

LONG-EZ SPORT PLANE was fitted with an experimental liquid-fueled rocket motor for aerial tests by engineers at XCOR Aerospace.

COURTESY OF MIKE MASSEE *XCOR Aerospace*

Going Small

MEANWHILE PRIVATE SPACE entrepreneurs spied the possibility of making profits by providing routine, low-cost launch services for small spacecraft—an underserved market. To save money, today's experimental payloads have to hitch rides to space on rockets built to boost multimillion-dollar payloads. The lack of affordable rides on a routine schedule has severely hampered university space researchers, such as those investigating microgravity. "While I'd love to see lower launch costs, what we could really use is a system where the launch costs scale to low-mass spacecraft" in the range of tens of kilograms to less than a kilogram, says Chris Kitts, who builds small satellites at Santa Clara University. "I think the university market could be opened dramatically by focusing on this angle." Kitts estimates that if rides were available, as many as 10 university spacecraft would be launched a year at current prices.

That number could grow to 20 or 30 small payloads annually should a dedicated, low-cost launcher come to market, says longtime small-satellite designer and advocate Rex Ridenoure. Lightweight spacecraft with a single objective that need to remain in outer space for a week or less form a solid niche market, he notes. After all, NASA launched an average of 25 sounding rockets (conventional suborbital boosters) annually between 1990 and 2002, according to a market study conducted by the Aerospace Corporation, a private think tank in El Segundo, Calif., that works for the U.S. Air Force.

Although some companies do offer small-scale launch services, the price is still too high for many potential customers. Surrey Satellite Technology in the U.K., for one, has made a business of lofting to space mostly university research payloads weighing as much as half a ton tucked in as secondary payloads on big Russian boosters (for from $35,000 to $70,000), or on conventional Russian launchers, but each separate flight costs $10 million to $13 million. Orbital Science's Pegasus, an expendable rocket that is air-dropped from an L-1011 airliner, also provides orbital transport for smaller payloads from about $14 million to $30 million per launch.

Currently dozens of small companies are working to design the next generation of low-cost orbital launchers. Several call the dusty, tumbleweed-strewn Mojave Airport, two hours' drive north of Los Angeles, home. It might seem an unlikely center for private space activity, but the airport in what some have dubbed "Kerosene Valley" harbors many engineers eager to start a technology revolution. The airport's classic Auger Inn restaurant—a wry reference to the corkscrew nosedives sometimes experienced by aviation test pilots—is newly adorned with a blue scalloped awning advertising its new name: the Voyager Café. From its runway-side booths, patrons can watch flying vehicles of varying apparent engineering feasibility spring from sketches on napkins and, sometimes, later trundle about the desert site.

The kind of seat-of-the-pants engineering that characterizes this place sits in a corner of an old hangar down the road. Back when the café was still the inn, engineers at Mojave's XCOR Aerospace took a Long-EZ—a two-seat homebuilt sport plane—and fitted it with a liquid-fueled rocket engine in place of a motor and propeller [*see illustration on this page*]. XCOR engineers

THE AUTHOR

JOAN C. HORVATH is chief executive officer of Takeoff Technologies LLC, a technology-strategy consulting firm in southern California. She works to encourage emerging sectors of the private launch industry to collaborate with universities, localities interested in becoming spaceports and mainstream aerospace companies. Horvath is also co-founder and executive director of Global Space League, a nonprofit organization based in Frederick, Okla., that finds transport for student science experiments designed to operate in exotic places—eventually, she hopes, in actual orbit. Horvath spent 16 years at the Jet Propulsion Laboratory and holds engineering degrees from the Massachusetts Institute of Technology and the University of California, Los Angeles. She teaches graduate courses at the University of Phoenix and serves as president of the Pasadena Figure Skating Club.

constructed this somewhat chimeric EZ-Rocket to perform aerial tests on their new engine, which they hope will power a rocket plane to the edges of space.

Safe but Cheap

FOR THESE SUBORBITAL ENGINEERS, operational safety and low cost are driving their designs rather than attempting to eke out the last bit of performance from the engines, as Big Aerospace personnel might do to fulfill government requirements. Burt Rutan, a Kerosene Valley legend, sums up the local engineering philosophy: "Safety, of course, is paramount, but minimum cost is critical." Rutan's Mojave-based company, Scaled Composites, famous for creating novel composite aircraft designs such as Voyager (the first aircraft to fly around the world without refueling), is mounting a strong effort to establish low-cost suborbital spaceflight.

Building on design work that began in 1996, Scaled Composites began developing a launch aircraft called the White Knight, a three-seat rocket plane known as SpaceShipOne and associated ground operations [*see illustrations on pages 92 to 94*]. Says Rutan: "I realized that I could build a suborbital rocket and I could fly myself into space—that this was actually within my grasp."

Rutan's design goal was to simplify or eliminate as many operational systems as possible. SpaceShipOne, for example, features an unusual "feather" tailboom-raising mechanism that places the craft in a high-drag configuration, allowing it to descend from altitude safely on a nearly straight-down reentry trajectory with minimal steering and control inputs from the pilot. He characterizes more standard reentry vehicle designs, which typically minimize the g-forces on the occupants using more complex systems, as being more dangerous and expensive. Another unconventional design element, the hybrid rocket engine that powers the suborbital vehicle, provides unique operational options. Unlike a solid-fueled rocket, a hybrid power plant can be throttled by altering the flow of oxidizer to the rubbery solid fuel, but it is simpler and cheaper than liquid-fueled systems [see "Hybrids Take Off," by Steven Ashley; News Scan, SCIENTIFIC AMERICAN, June 2003]. During a flight test in mid-December of last year, SpaceShipOne reached Mach 1.2.

All of this development work was accomplished at a cost in the tens of millions of dollars, which was provided by Paul G. Allen, the billionaire investor. Allen intends to win the X Prize and thus prove out the technology for commercial use. He cites regulatory issues and the unknown time required to make the vehicles sufficiently safe for passengers as the major impediments to business success. Rutan meanwhile conceives of his current system as a pure flight demonstrator. He believes the cost to the company per test flight will initially be in the range of $90,000 to $100,000, a figure that could possibly be halved as the operators gain greater experience. A commercially viable launch vehicle, he thinks, will need to offer six or seven seats with sufficient room for passengers to move around to enjoy microgravity conditions.

A somewhat more risky concept comes from Mesquite, Tex.–based Armadillo Aerospace, founded by video game designer John Carmack. Company engineers are developing a spacecraft that would take off vertically on the thrust of a rocket engine fueled with purified hydrogen peroxide. The Black Armadillo is to return by parachute, with final deceleration occurring when its crushable nose strikes the ground [*see illustration above*]. Carmack, known for writing the software engines for popular video games such as *Doom* and *Quake,* is producing the code that will control the rocket engines. He notes that writing engine-control code gives him fewer headaches than producing driver software for the fickle video game market.

CRUSHABLE NOSE CONE will cushion the landing when Armadillo Aerospace's Black Armadillo spacecraft returns from its planned suborbital flight. A subscale model is shown parachuting down in a drop test (*inset*). After the experiment, engineers inspected how the model's nose structure collapsed to absorb the forces of impact.

ARMADILLO AEROSPACE

Private-public partnerships to finance space access are also beginning to emerge. In early January the state of Oklahoma announced that investors in another private spacecraft builder, Rocketplane, would qualify for a special tax credit that could be worth up to $17 million over five years, according to the plan's architect, State Senator Gilmer N. Capps. Starting in the fall of 2006, Rocketplane's management intends to offer suborbital flights to an altitude of 60 miles in its rocket-powered, winged vehicle for around $100,000.

Taking another tack, PayPal co-founder Elon Musk's SpaceX venture has focused on cutting by a factor of three the operational costs of flying a relatively conventional unmanned rocket. The concept is to build a simple "truck" design that burns liquid oxygen and economical, easy-to-handle, high-grade kerosene, as many Russian launchers do. After reentry, the truck would be plucked from the sea and 80 percent of its structure reused.

The Paperwork Barrier

ASSUMING THAT privately funded companies can establish a market, find enough funding and surmount the technological challenges, they will then have to face a seemingly intractable obstacle: government regulations. For launch vehicles originating from the U.S., the Federal Aviation Administration has jurisdiction. Current rules governing rockets assume that anyone onboard is a crew member who is knowingly taking personal risks. This licensing procedure contrasts greatly with the established certification process for passenger aircraft: an expensive and lengthy agenda that attempts to ensure both passenger and public safety at nearly zero risk levels. The licensing for winged, passenger rockets falls into a gray zone. Pending House legislation, the Commercial Space Act of 2003 (HR 3245) tries to address ambiguities regarding whether winged, rocket-powered launch vehicles should be defined as "rockets" or "airplanes." This distinction is important because regulations for rockets are intended to keep people on the ground safe, whereas airplane passenger regulations are aimed primarily at maintaining passenger safety—hence the difference between "licensing" and "certification." This bill would also set out the informed-consent requirements for crews and passengers. Another abiding issue that needs governmental attention concerns export-control laws designed to hinder international arms traffic; such laws place barriers on U.S. space-access firms hiring noncitizen employees or selling to overseas users.

Patricia Grace Smith, the FAA's associate administrator for commercial space transportation, has a perspective rooted in the safety successes of the civil aviation community. She holds that "the crew is part of the flight-safety system." In other words, the FAA tends to trust the pilot's judgment above that of an automated control system.

Smith expects that the nascent industry for piloted, reusable launch vehicles will undergo three phases. Today trailblazers—"very aggressive, visionary entrepreneurs"—fight a tough procedural struggle "versus very traditional institutions." As of press time, the FAA was processing three entrepreneurial launch license applications. As part of the U.S. licensing process, launch vehicle companies have to show that they can cover damages to third parties (hit by flying debris, for example). This indemnity is complicated by international treaties governing injury caused by space vehicle accidents. The launching entity must provide coverage up to the maximum probable loss (calculated by Smith's FAA office) that could reasonably be caused by the vehicle if it crashed. Losses beyond the insured amount would come under the purview of the secretary of transportation, who could approve federal coverage up to $1.5 billion. This government provision expires in 2004, but the pending Senate bill S 1260 (the Senate version of HR 3245) would extend it to 2009.

The FAA is also in charge of licensing spaceports, locations from which it is legal to fly to space. Mojave Airport is reportedly well on its way to being officially sanctioned as the first nongovernmental space-launch facility.

Once the initial flights establish basic launch capabilities, Smith believes they would be followed by a transitional stage, in which businesses would fly a limited number of passengers who understood and accepted the risks. There is significant controversy over precisely how much risk the federal government should allow passengers to take. This issue will be debated in the House and Senate over the coming months as HR 3245 and S 1260 move out of committee. Eventually routine operations of an affordable space transportation system in which passengers have a reasonable expectation of safety—nearly that of the commercial aviation system—will come to pass, Smith thinks.

Not everyone is happy with the current regulatory processes. Rutan perceives the current launch-licensing procedure as demanding the same levels of safety for early research flights as it does for flying paying passengers. He points out that the process requires environmental-impact analyses and estimates of failure probabilities based on the severely limited data that are available. Rutan favors an approach more like aircraft certification than rocket-launch licensing. Nevertheless, he has applied for a launch license for his upcoming X Prize flights.

When pondering the challenges facing today's private space pioneers, I often recall the Apollo moon landings. Once, as a youngster on a Florida vacation, I aimed my Instamatic camera at the huge Vehicle Assembly Building at Kennedy Space Center where technicians assembled the mighty Saturn 5. I was surprised, although I shouldn't have been, when the resulting photograph captured just part of the gigantic hangar. Back then everything associated with spaceflight was enormous. Now, nursing a cup of coffee at the Voyager Café while working over annual expenditures, rates of return, breakeven points and price per launch for one of my consulting projects, I realize that one thing is apparent: real-world access to space has finally come within the grasp of entrepreneurial dreamers and venture capitalists rather than being only the domain of national governments. SA

Burt Rutan believes test flights of his suborbital vehicle WILL COST FROM $90,000 TO $100,000.

MORE TO EXPLORE

Suborbital Reusable Launch Vehicles and Applicable Markets. Jared Martin and G. W. Law. U.S. Department of Commerce, October 2002. Available at **www.technology.gov/space/library/reports/2002-10-suborbital-LowRes.pdf**

Affordable Spaceship: Burt Rutan's Quest for Space. Michael A. Dornheim in *Aviation Week and Space Technology,* Vol. 158, No. 16, pages 64–73; April 21, 2003.

Encyclopedia Astronautica provides background information on rockets: **www.astronautix.com**

The Federal Aviation Administration's Office of the Associate Administrator for Commercial Space Transportation gives the current status of regulatory issues and relevant legislation: **http://ast.faa.gov**

Information about space start-up companies and hobbyist activities: **www.hobbyspace.com**

Information about X Prize entrants: **www.xprize.org**

FUEL INJECTION

Complete Burn

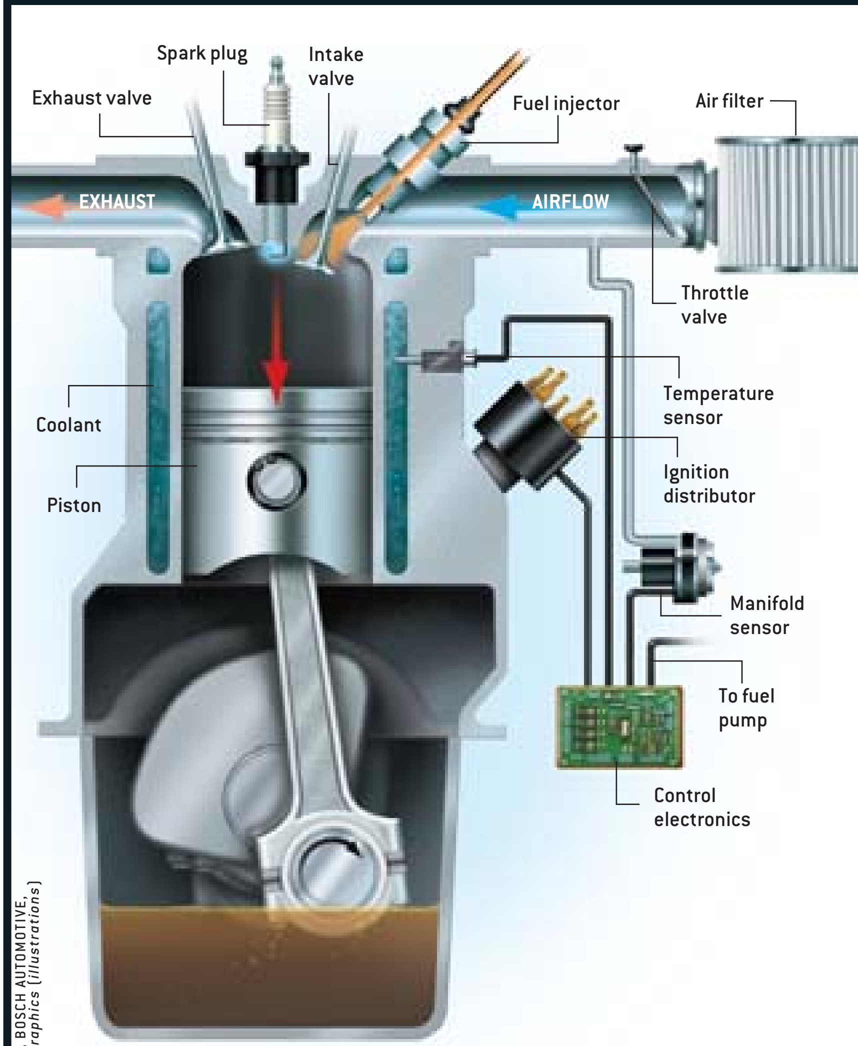
"Hey, this sporty model is hot," the auto salesman raves. "It's got sequential multipoint fuel injection!" Yeah, well, so does virtually every other passenger vehicle now in production.

For decades, the good old carburetor acted like a funnel that allowed gasoline and air to be sucked into a car engine's cylinders. Spark plugs ignited the mixture in mini explosions that drove the pistons. The carburetor worked well enough but struggled to finely control the fuel-to-air ratio or even to deliver gasoline equally to each cylinder, limiting fuel economy and creating pollutants and rough engine operation.

By the late 1970s electronics came along that could master fuel delivery; for a few years, automakers used throttle-body injectors that were basically computer-run carburetors. But the real improvement came in the mid-1980s as new cars were fitted with multipoint fuel injectors—nozzles that precisely delivered fuel to each cylinder. The gasoline burned much more uniformly: fuel efficiency and power output rose, and emissions fell. In the 1990s sequential multipoint fuel injection took over [*see left illustration*]; the engine computer controlled each injector individually, instead of en masse, for still greater benefits. Each step has required more advanced injectors, feedback sensors and computer processors.

Superior fuel control has been the driving force behind the long progression from carburetors, notes John German, manager of environmental and energy analysis at American Honda Motor Company in Ann Arbor, Mich. The evolution is now headed toward so-called direct injection, in which gasoline is sent straight into the cylinder rather than being premixed with air [*see right illustration*]. Engineers at car companies and facilities such as Sandia National Laboratories are experimenting with ways to improve the shape of the fuel spray and with how gasoline swirls or stratifies inside the cylinder for a highly thorough and clean burn.

Direct-injection engines operate at higher pressures, which some makers say is impractical, and they produce troublesome nitrous oxides. Nevertheless, early models are for sale in Europe and Japan. Direct injection will provide even better fuel control if cost and emissions problems can be licked. —*Mark Fischetti*



PORT FUEL-INJECTION ENGINES spray gasoline into an incoming airstream. The mixture enters the combustion chamber at about atmospheric pressure through an intake valve (port). Depressing the accelerator opens the throttle, letting in more air, and the engine computer instructs the injector to spray a bit longer. To respond to loads such as hills and hard acceleration, the computer also adjusts the spark timing and air intake, as well as the transmission gearing.

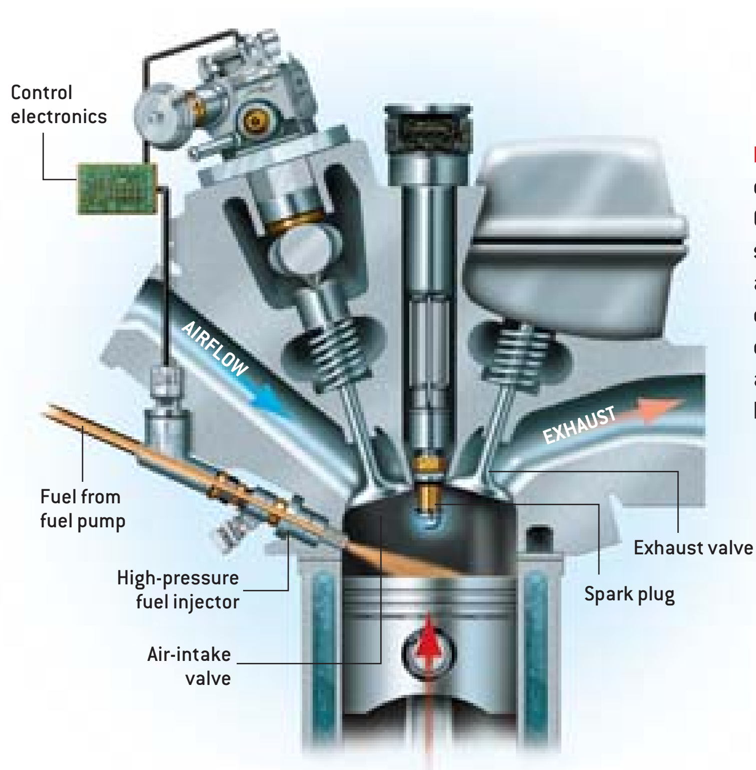
SOURCES: HONDA, SANDIA NATIONAL LABORATORIES, BOSCH AUTOMOTIVE, DENSO ELECTRONICS; KENT SNODGRASS *Precision Graphics* (*illustrations*)

DID YOU KNOW . . .

➤ SPARKLE: Engineers are prototyping engines fired by a process called homogeneous-charge compression ignition that could provide greater fuel economy and fewer emissions than contemporary designs can. The power plant runs on premixed fuel and air, like a fuel-injection engine, but ignites the mixture with high pressure instead of a spark plug, like a diesel engine [see “A Low-Pollution Engine Solution,” by Steven Ashley; SCIENTIFIC AMERICAN, June 2001].

➤ DIESEL POWER: Diesel engines get much greater mileage than gasoline engines because diesel fuel harbors about 12 percent more potential energy than gasoline. Also, diesels operate at higher compression ratios (14:1 to 25:1, versus 8:1 to 12:1 for gasoline) that extract more energy from combustion. Diesels use direct injection for precise fuel control and do not have a throttle [*see illustration on opposite page*], avoiding losses incurred in a gasoline engine by having to force air past a partially closed throttle valve; power output is controlled by altering the amount of fuel injected.

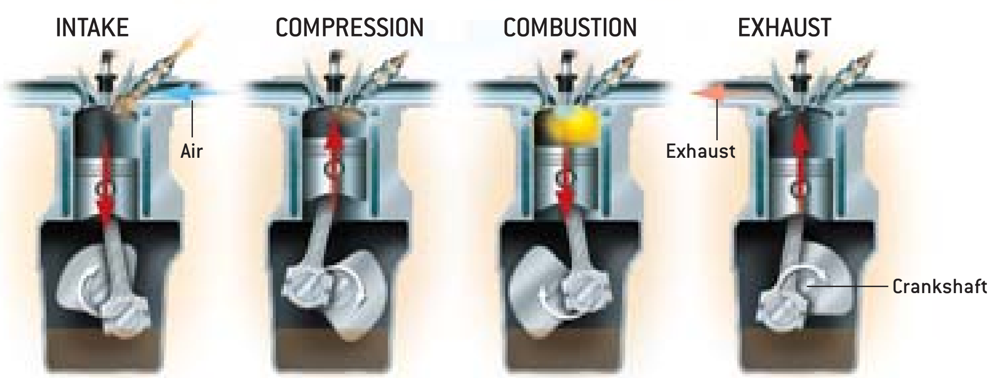
➤ DIRTY BUSINESS: Diesel engines produce more soot than gasoline engines and require complex emissions equipment to meet U.S. standards. Manufacturers are reluctant to install the hardware because it adds cost; however, diesels are prevalent in Europe and Japan, in part because pollution limits are lower and because fuel prices are higher, which make diesels’ greater mileage worth the cost.



DIRECT-INJECTION ENGINES send gasoline directly into the combustion chamber after the air-intake valve closes. Gasoline vaporizes better than it does in port fuel-injection systems, raising power output, and the injection timing allows the engine to run on a very lean air-fuel mixture, enhancing fuel economy. But advanced control and emissions technologies are needed. Diesel engines are also “direct,” but the piston compresses the air at pressures high enough to ignite the fuel without a spark plug.

PETER WITZE *Sandia National Laboratories*

ATOMIZED GASOLINE (*colorized*) emitted by a fuel injector hits an air-intake valve as it begins to open, in an experiment at Sandia National Laboratories.



FOUR-STROKE CYCLE is used by virtually all automobile engines to turn a crankshaft that powers the wheels. The intake valve opens so that air and gasoline can enter as the piston drops. The rising piston compresses the air-fuel mixture, and a spark ignites it. The force of combustion drives the piston down. The exhaust valve opens so that the rising piston can expel by-products.

Topic suggested by reader Bob Merritt.
Send ideas to workingknowledge@sciam.com

Plug-and-Play Robots

PERSONAL ROBOTS MAY SOON BE AS CHEAP AND CUSTOMIZABLE AS PERSONAL COMPUTERS **BY W. WAYT GIBBS**

"Could this be the place?" I wonder as I stand before a nondescript storefront, formerly a tattoo parlor, in the tiny borough of Youngwood, Pa. The windows are covered by blinds; the door bears forbidding bars. The building lacks a sign or even a house number. It seems an odd location from which to launch an ambitious new species of robot.

But when Thomas J. Burick opens the door and I see three prototype "PC-Bots" sitting on his small workbench, I realize that this 34-year-old entrepreneur is no ordinary inventor. The half-meter-high robots look like R2-D2 droids that have been redesigned by Cadillac. Burick says that he spent a year honing their appearance, something almost unheard of in serious robotics, where function usually trumps form.

To Burick, form is function, and it is very important that he get the design right the first time. This is his life's dream, and it has consumed his life savings. "As a kid, I watched *Lost in Space* on TV and thought it was so fantastic that Will Robinson had this machine who protected him and was the best friend anyone could have," he says, with boyish earnestness. In seventh grade Burick built a voice-controlled mobile robot, and in high school he constructed an autonomous fire-fighting rover.

As he got older, Burick dabbled with the low-level microcontrollers, servos and sensors used these days by amateur robot builders, but he finally gave up in frustration. His jobs had been in retail, selling consumer electronics rather than making them. With no formal training in programming or electrical engineering, "the learning curve was too steep," he says. "I thought there has got to be a better way."

PC-BOT PLATFORM can be customized with PC accessories, allowing the basic 912 model (*left*) to be transformed into a DVD-playing 912 MP3 (*center*) or a 912 HMV security rover (*right*).

THOMAS J. BURICK *White Box Robotics*

The better way occurred to him 18 months ago as he was repairing one of the generic "white box" PCs he assembles from components and sells at his small computer store half a block up the street. What if robots could be built up from interchangeable, commodity parts just the way that desktop computers are? Better yet, what if there were a robotic platform that could accept the thousands of plug-and-play PC peripherals and accessories already on the market?

Soon there will be. White Box Robotics, Burick's nascent company, is preparing three varieties of its mobile robotic chassis for mass production this summer. Sitting on the workbench, each of the final prototypes for the three robots looks quite distinct.

"This one is customized for security," Burick says, putting his hand on the 912 HMV. "It's painted with the same paint used on the Hummer H2," he notes. Hella driving lamps are mounted on its front, and webcams peer out from its head and belly. "It could patrol your house while you are away and e-mail or page you if it detects a loud noise or an unfamiliar person."

"And we call this one the 912 MP3," Burick says, gesturing toward a white robot of the same size and shape but sporting a color LCD screen in its back, a stereo control panel in its midriff and a striking blue lamp in its head instead of a camera. "We designed this to appeal to young people, who could use it as their bedroom computer. It can download music and play DVDs in response to voice commands."

But at the moment it is the third robot, the basic model, that illustrates Burick's idea best. Its exterior shell has been removed to reveal an inner skeleton, common to all three siblings, that allows them to work like PCs on wheels.

A simple metal frame holds up to six shelves at various positions. "All of the electronics mount on these trays, which slide out," he explains. One tray holds the motherboard, which is a diminutive Mini-ITX made by VIA that crams an Intel-compatible processor with 512 megabytes of memory, video, audio and networking chips all onto a circuit board the size of a bread plate. Another tray contains two hefty but inexpensive 12-volt batteries that allow the robot to run for three hours between charges. Standard laptop hard drives, CD burners, DVD drives—virtually any PC gadget you can find at CompUSA—can be mounted securely to these trays inside the robot's body.

"These robots run Windows XP, so they can do anything your PC will do," Burick points out. But they can also do things that no PC can—move, for example. Below the metal frame is a drive assembly: two motors connected to two four-inch wheels and a spring-mounted hard plastic ball. "This Delrin ball keeps the robot stable even on uneven surfaces and allows it to track its position precisely even as it turns and spins," he notes. "And this whole drive assembly comes off with four bolts, so if you want to replace it—with tank-style treads, for example—it's a five-minute job." Eventually he hopes to offer an optional drive mechanism capable of climbing stairs.

With an oomph, Burick lifts the 40-pound machine down to the floor and turns it on. To my surprise, the 18-inch-tall rover is quieter than my laptop computer. Just like any PC, it has ports on its backside for a monitor, keyboard, mouse, Ethernet cable and so on. But untethering the machine is only a matter of plugging an inexpensive WiFi receiver into the motherboard. Burick uses his laptop to log in over the wireless link. "Okay, this software is now running on the robot," he says as he launches the Robot Control Center, a program he has licensed from Evolution Robotics.

Clicking on buttons in the program, he drives the robot toward a 20-year-old Heath Hero Junior robot gathering dust in the corner. When it gets within three feet or so, the 912 stops and says in its synthesized voice, "Hello, Hero Junior."

It's a nice trick, and as I try the software myself, I find that creating such pseudointelligent behaviors is quite sim-

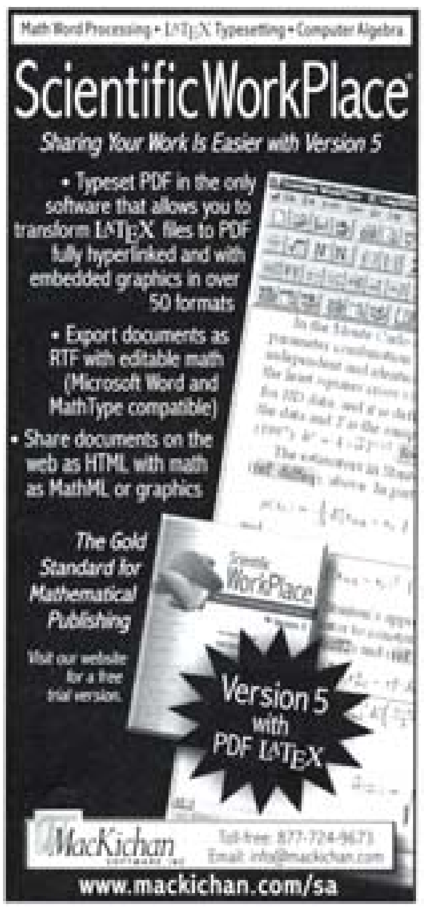
ROBOT CONTROL software from Evolution Robotics makes programming a PC-Bot simple. In this example, a 912 HMV patrolling a floor could recognize a person, speak a phrase, then snap a photograph and e-mail it to a supervisor.

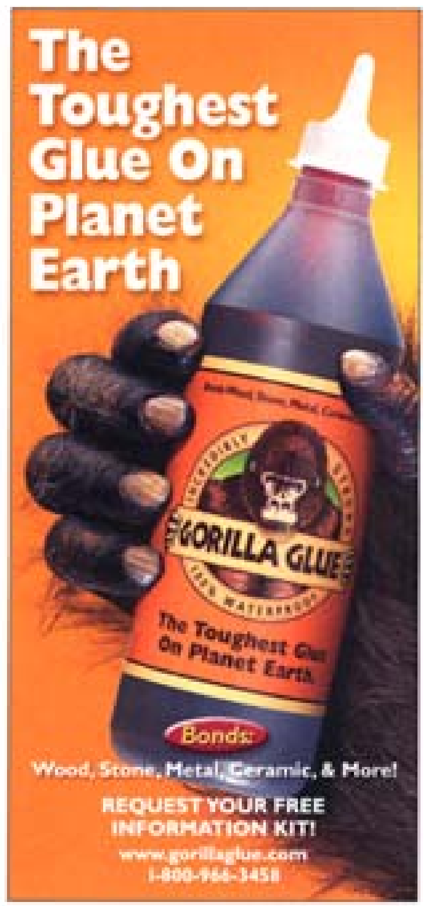
COURTESY OF WHITE BOX ROBOTICS, © EVOLUTION ROBOTICS

ple. A small window shows onscreen what the robot sees through its "eye" camera. Burick sets a model of the B9 robot from *Lost in Space* on the floor, and I turn the 912 to look at it. I click the "capture" button, and the machine adds it to a memorized list of objects that it can recognize. I back the 912 up, turn it around, and create a new behavior by checking a box here and making a menu selection there. Six clicks later I have taught it to speak a phrase whenever it sees the B9 toy. And it works: even from four feet away and at an odd angle, the 912 recognizes the toy and says, "Hey, get out of my way!"

It is a trivial example of a powerful combination: easy-to-use software and easily customized hardware. Plug in a microphone, and the robot can respond to voice commands. Attach an infrared sensor and tuck a few speakers in its case, and it becomes a CD player that follows you around the house. Bolt on a couple of gripper arms, and you can program it to empty the kitty litter box every other day.

Because the chassis can be stamped out by the same factories that make computer cases, the PC-Bots will cost in the neighborhood of $1,000—about as much as a decent desktop system, Burick declares: "My goal is to make it affordable enough that a 14-year-old can buy one with the money earned on a paper route. I want people to use this platform like a blank canvas, to let their imaginations run wild." SA

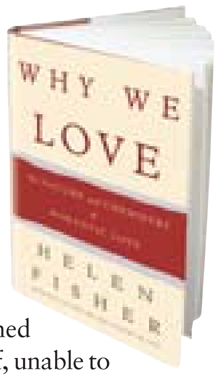




The Brain in Love

USING NEUROCHEMISTRY TO TRY TO UNRAVEL THE EXPERIENCE OF ROMANTIC PASSION **BY BARBARA SMUTS**

WHY WE LOVE: THE NATURE AND CHEMISTRY OF ROMANTIC LOVE
by Helen Fisher
Henry Holt, New York, 2004 ($25)



A male baboon named Sherlock sat on a cliff, unable to take his eyes off his favorite female, Cybelle, as she foraged far below. Each time Cybelle approached another adult male, Sherlock froze with tension, only to relax again when she ignored a potential rival. Finally, Cybelle glanced up and met his gaze. Instantly Sherlock flattened his ears and narrowed his eyes in what baboon researchers call the come-hither face. It worked; seconds later Cybelle sat by her guy, grooming him with gusto.

After observing many similar scenarios, I realized that baboons, like humans, develop intense attractions to particular members of the opposite sex. Baboon heterosexual partnerships bear an intriguing resemblance to ours, but they also differ in important ways. For instance, baboons can simultaneously be "in love" with more than one individual, a capacity that, according to anthropologist Helen Fisher, most humans lack.

Fisher is well known for her three previous books (*The Sex Contract, Anatomy of Love* and *The First Sex*), which bring an evolutionary perspective to myriad aspects of sex, love, and sex differences. This book is the best, in my view, because it goes beyond observable behaviors to consider their underlying brain mechanisms. Most people think of romantic love as a feeling. Fisher, however, views it as a drive so powerful that it can override other drives, such as hunger and thirst, render the most dignified person a fool, or bring rapture to an unassuming wallflower.

This original hypothesis is consistent with the neurochemistry of love. While emphasizing the complex and subtle interplay among multiple brain chemicals, Fisher argues convincingly that dopamine deserves center stage. This neurotransmitter drives animals to seek rewards, such as food and sex, and is also essential to the pleasure experienced when such drives are satisfied. Fisher thinks that dopamine's action can explain both the highs of romantic passion (dopamine rising) and the lows of rejection (dopamine falling). Citing evidence from studies of humans and other animals, she also demonstrates marked parallels between the behaviors, feelings and chemicals that underlie romantic love and those associated with substance addiction. Like the alcoholic who feels compelled to drink, the impassioned lover cries that he will die without his beloved.

Dying of a broken heart is, of course, not adaptive, and neither is forsaking family and fortune to pursue a sweetheart to the ends of the earth. Why then, Fisher asks, has evolution burdened humans with such seemingly irrational passions? Drawing on evidence from living primates, paleontology and diverse cultures, she argues that the evolution of large-brained, helpless hominid infants created a new imperative for mother and father to cooperate in child-rearing. Romantic love, she contests, drove ancestral women and men to come together long enough to conceive, whereas attachment, another complex of feelings with a different chemical basis, kept them together long enough to support a child until weaning (about four years). Evidence indicates that as attachment grows, passion recedes. Thus, the same feelings that bring parents together often force them apart, as one or both fall in love with someone new. In this scenario, broken hearts and self-defeating crimes of passion become the unfortunate by-products of a biological system that usually facilitates reproduction.

Fisher's theory of how human pair-

OLIVE BABOONS, an adult female (*left*) and male, snuggle during an afternoon rest period in Kenya. Among baboons, only pairs who have formed long-term friendships have been observed in such intimate contact.

BARBARA SMUTS

bonding evolved is just one of several hypotheses under debate today, and she does not discuss these alternatives. Similarly, some of her ideas about love's chemistry are quite speculative (which she fully acknowledges). No one familiar with the evidence, however, can disagree that romantic love is a human universal that requires an evolutionary explanation, and Fisher, more than any other scientist, has brought this important point to public awareness.

Like the words of a talented lover, Fisher's prose is charming and engaging. Love poems, both modern and classic, enliven her narrative, along with poignant examples of romantic passion from other times and cultures. One chapter is a litany to passion in other animals, a vivid reminder that we are not the only species that feels deeply. Another provides new insight into the obsessive attempts of abandoned lovers to rekindle romance. Toward the end of the book, Fisher helps to redeem the self-help genre, rooting her advice in hard science. She shows how you might "trick the brain" to maintain enduring passion or recover more quickly from the pain of rejection: "Someone is camping in your brain," she reminds us, and "you must throw the scoundrel out." Engaging in activities known to increase dopamine might help; after all, love is not our only source of intense pleasure.

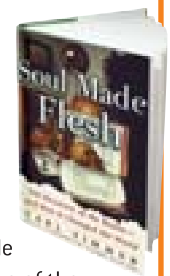
In hands as skilled and sensitive as Fisher's, scientific analysis of love only adds to its magic. If you forgot to give your beloved a gift on Valentine's Day, it's not too late to woo him or her anew with this book, which is likely to fascinate and delight anyone who has ever been in love. SA

Barbara Smuts is a professor in the psychology department at the University of Michigan at Ann Arbor. She is author of Sex and Friendship in Baboons *(reprinted with a new preface, Harvard University Press, 1999).*

THE EDITORS **RECOMMEND**

SOUL MADE FLESH: THE DISCOVERY OF THE BRAIN—AND HOW IT CHANGED THE WORLD

by Carl Zimmer. Free Press, New York, 2004 ($26)

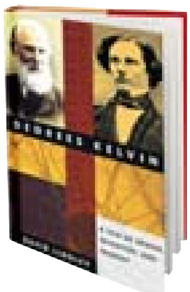


Thomas Willis, an eminent physician in 17th-century England, published in 1664 a book that became a medical classic: *Cerebri anatome,* or *The Anatomy of the Brain,* based on his pioneering and painstaking dissections of human brains. Science writer Zimmer, describing the work and its consequences, says that Willis and his team had produced more than a map of the brain. "They had, for the first time, created a unified treatment of the brain and the nerves." And to far-reaching effect. Willis "did for the brain and nerves what William Harvey had done for the heart and blood: made them a subject of modern scientific study." Zimmer draws a vivid picture of the background against which Willis and other scientists of the time worked.

DEGREES KELVIN: A TALE OF GENIUS, INVENTION, AND TRAGEDY

by David Lindley. Joseph Henry Press, Washington, D.C., 2003 ($27.95)

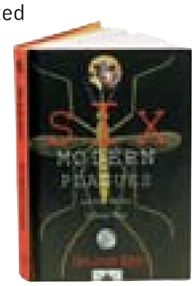
William Thomson, Lord Kelvin (1824–1907), made major contributions to 19th-century physics and technology but is mainly known today through the attachment of his name to a scale of temperature. Lindley, an astrophysicist who now focuses on writing about science, brings Kelvin to life in this excellent biography. The young Kelvin, Lindley writes, "made astonishing progress in the quest to understand the nature of heat, work, and energy, and in the parallel effort to elucidate the nature of electricity and magnetism." Kelvin's theory of undersea signal transmission was fundamental for the installation of transatlantic cables, and he was involved in work on power generation and navigation instruments. The "tragedy" of the book's title is that the old Kelvin became something of a crank, sticking "with blind stubbornness" to ideas about radioactivity, electromagnetism and the age of the earth in the face of contrary evidence accumulating at the turn of the century. But if Kelvin could come back today, Lindley says, he "would after being taken aback by the dizzying scope of modern theoretical physics decide that, after all, it was exactly what he had been trying to say."



SIX MODERN PLAGUES: AND HOW WE ARE CAUSING THEM

by Mark Jerome Walters. Island Press, Washington, D.C., 2003 ($22)

In the late 1960s the U.S. surgeon general declared that Americans could "close the book on infectious diseases." In 1999 the World Health Organization reported that "diseases that seemed to be subdued … are fighting back with renewed ferocity." And then there are the new or transformed diseases that have made headlines in recent years. Walters, a veterinarian and journalist, focuses on six of them: mad cow disease, AIDS, salmonella DT 104, Lyme disease, hantavirus and West Nile virus. In an epilogue, he briefly describes the latest headline maker, severe acute respiratory syndrome (SARS). "The larger story," he writes, "is not simply that humans and other animals are falling victim to new diseases; it is that we are causing or exacerbating many of them, not least of all through the radical changes we have made to the natural environment."



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

Bluffhead

BY DENNIS E. SHASHA

In my gambling years, between the ages of eight and 12, I played roulette, poker, blackjack and any other game that had a three-cent ante. One of my favorite games, though, was one we called "Bluffhead." Each person takes a card from a shuffled deck and holds it, face out, to his or her forehead. In other words, players see everyone else's cards but not their own. The best card wins. Ace is high and suits don't matter, so ties are possible.

This puzzle has to do with inferring information about the cards people hold by hearing what the players say. To be concrete, suppose that we have three players. Caroline always speaks first, then David, then Jordan, and back to Caroline and so on. Each player makes one of the statements listed at the bottom right of this page. Assume the players are perfect logicians and reveal information only through these phrases; also, they say the strongest thing they can—that is, they choose the statement that is true and highest on the list.

To warm up, suppose Caroline says, "I don't know." Then David says, "I lose," and Jordan says, "I lose." Hearing all this, we (and the players) know Caroline must have an ace and the others do not. To understand why, see the illustration below.

Your Challenge: What can you infer about the cards from the following scenarios?

Game 1: Caroline says, "I don't know." David then says, "I don't win," and Jordan follows with, "I win."

Game 2: Caroline says, "I don't know." David then says, "I don't win," and Jordan follows with, "I tie as a winner."

Game 3: Caroline says, "I don't know." David then says, "I don't win," and Jordan follows with, "I don't win."

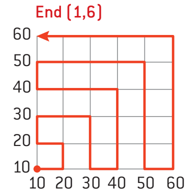
Game 4: Caroline says, "I don't know." David then says, "I don't know," and Jordan follows with, "I don't know." At the start of the next round, Caroline responds, "I lose." Can you predict what David and Jordan say next?

Game 5: Caroline, David and Jordan each say, "I don't know" in order, then say, "I don't know" in a second round. In the third round, Caroline and David say, "I don't know," but Jordan says, "I win." Which card does Jordan have, and what might he see? SA

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

Answer to Last Month's Puzzle

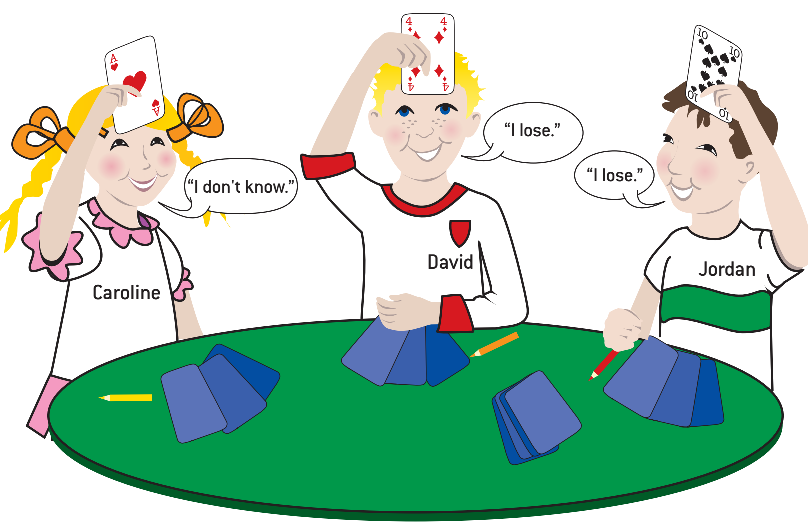
One good solution to the grid puzzle is shown in the illustration below. The total duration of the journey is 12 hours and six minutes.



Web Solution

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

SOLUTION TO WARM-UP PROBLEM: Even without seeing the players' cards depicted here, we would know from the spoken words that Caroline has an ace and the others do not. If Caroline saw an ace, she would know she was not the sole winner and would say, "I don't win." But not seeing an ace and having no information about her own card, she has to say, "I don't know." Her uncertainty tells the boys that neither of them has an ace. By seeing that Caroline has an ace, they both realize that they lose.



THE CHOICES

"I win."
(I have a higher card than everyone else.)

"I lose."
(Someone has a higher card than I do.)

"I tie as a winner."
(I and at least one other player tie with the highest card.)

"I don't win."
(I either tie as a winner or lose.)

"I don't lose."
(I either win or tie as a winner.)

"I don't know."

CORNELIA BLIK

Visiting Royalty

MONARCHS FROM MAINE AND MICHIGAN WAIT OUT THE WINTER ON A MEXICAN MOUNTAINSIDE **BY STEVE MIRSKY**

There are many compelling vistas for the traveler heading west of Mexico City on Route 15. The extinct Nevado de Toluca volcano, for example, dominates the scene to the left. Within its now quiet caldera lie two lakes, which, at an altitude of more than 15,000 feet, are among the highest places in the world where people scuba dive. The volcano's lakes thus offer the opportunity to get altitude sickness and the bends at the same time.

Then there are the views to the right as Route 15 turns from a busy highway into a twisting, two-lane mountain road. Pondering the sheer drops from the switchbacks my bus is negotiating can be the cause of—or cure for, depending on one's particular physiology—the gastrointestinal problems this part of the world is famous for inadvertently inflicting on gringos like me.

The bus, carrying a mostly gringo group of junketing journalists on this late January day, is wending its way to the small mountain town of Angangueo, where we'll spend the night. The next morning we will continue on to Sierra Chincua. This reserve is one of perhaps 20 pockets of fir forest, totaling only about 65 acres, that have just the right balance of temperature—usually—and humidity to be attractive to the quarter of a billion monarch butterflies (*Danaus plexippus*) that fly to them from eastern Canada and the U.S. to spend the winter. The phenomenon is not unlike that other great natural migration, spring break. The butterflies, however, will leave as mature adults.

After five hours, we at last pull up to our inn, along the steep, narrow main drag of Angangueo at dusk. How steep is the drag? The sidewalks have stairs built into them. After dinner, we see a map of the region in which landowners get paid to not cut down the fir trees that protect the butterflies through the winter. (The situation is streamlined in the U.S., where we pay landowners not to grow crops in the first place.)

In the morning we get back on the bus and climb further into the mountains until we reach the entrance to Sierra Chincua, where it's not just unusually cold, it's white—the night brought snow, and more soon begins to fall. Butterflies actually flying are probably off the menu for today. We're now at an elevation of more than 10,000 feet, with about two miles of hilly terrain to traverse to get to the fir trees we expect to be shielding the butterflies from the weather. Most of our party hops onto horses to make the trip. Claudio Angelo, a Brazilian science journalist, and I decide to hoof it ourselves, figuring that the walk will warm us up. "Besides," he says, "I don't feel like sitting on a wet, sweaty, smelly horse." A few minutes of hiking up hilly, muddy terrain at altitude, however, makes me paraphrase the famous words of Nietzsche: That which doesn't kill you makes you smell strong. In fact, the horses are giving us dirty looks.

Finally, we reach the butterflies, which have been grounded by the weather. Two days before, I had strolled down the Calle de los Muertos (Avenue of the Dead) amid the pyramids at Teotihuacán. The trail at Sierra Cinchua looks like a road of death as well, littered with thousands of butterflies. Turns out it's merely the calle de los frios. Walking carefully among the cold monarchs, we see that—as was said of Monty Python's parrot—they're not dead, they're resting, protected by a natural antifreeze compound in their bodies.

I pick one up and warm it with my breath. Suddenly the wings open. Startled, I nearly stagger backward, but I know that any random footsteps could probably take out about 20 other monarchs. So I stand my ground and place the almost fluttering butterfly on a tree trunk. This one may be lucky enough to begin the trip back to the U.S. in the spring. Here's hoping it and its northbound progeny avoid the windshields of southbound college students. SA

ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

How do dimples on golf balls affect their flight?

—T. Grinham, Manchester, England

Tom Veilleux, a senior scientist, and Vince Simonds, director of aerodynamic research at the Top-Flite Golf Company, explain:

Dimples reduce drag and improve lift, so golf balls fly farther. A smooth golf ball hit by a professional would travel only about half as far as one with dimples.

Engineers and scientists in the golf industry study the impact between a golf club and a ball to determine the so-called launch conditions. The impact, which typically lasts just 1/2,000 of a second, establishes the ball's velocity, launch angle and spin rate. Gravity and aerodynamics then take over the ball's trajectory (no matter how much the golfer hopes or curses). As a result, aerodynamic optimization—achieved through dimple-pattern design—is critical.

Air exerts a force on any object, such as a golf ball, moving through it. Holding your arm out of the window of a moving car easily illustrates this phenomenon. Aerodynamicists break down the force into two components: lift and drag. Drag acts to oppose motion, whereas lift acts in a direction perpendicular and upward to motion.

A speeding golf ball has a high-pressure area at its front. Air flows over the front contours and eventually separates from the ball toward the back. The ball also leaves behind a turbulent wake region, where the airflow is fluctuating or agitated, resulting in lower pressure at the rear. The size of the wake affects the amount of drag. Dimples create a thin, turbulent boundary layer of air that clings to the ball's surface. This design allows the smoothly flowing air to follow the ball's shape a little farther around the backside, decreasing the size of the wake. A dimpled ball thus has about half the drag of a smooth ball.

Most golf balls have between 300 and 500 dimples, which each have an average depth of about 0.010 inch. The lift and drag forces are very sensitive to dimple depth: a difference of as little as 0.001 inch can produce radical changes to trajectory and travel distance. Dimples have traditionally been spherical, but other shapes can also optimize aerodynamic performance; the HX golf ball by Callaway, for example, uses hexagons.

SARA CHEN

Dimples also affect lift. A smooth ball with backspin creates lift by warping the airflow such that the ball acts like an airplane's wing. The spinning action makes the air pressure on the bottom of the ball higher than the air pressure on the top; this imbalance creates an upward force on the ball. Ball spin contributes about one half of a golf ball's lift. The other half is provided by the dimples, which allow for optimization of the lift force. [For more on golf-ball lift, see "Flight Control," by Mark Fischetti, Working Knowledge; Scientific American, June 2001.]

How does club soda remove red wine stains?

—K. Hug, New Bern, N.C.

Biochemist Pete Wishnok of the biological engineering division at the Massachusetts Institute of Technology provides an answer:

There is no particularly good chemical reason why club soda should remove stains: it is essentially water with carbon dioxide and salts dissolved in it. It is weakly acidic, so it might decolorize stains that can act as acid-base indicators.

Club soda does seem effective on occasion. It has worked dozens of times for wine spills on our living room carpet (which probably says more about our lifestyle than it does about club soda). But it did not work at all on the tablecloth during Christmas dinner (the laundromat did the trick the next day).

A common theory suggests that the secret ingredient is the bubbles, and there may be something to that. Probably more important, however, is the fabric involved and how fast you can run. If your carpet is a synthetic that absorbs stains slowly—and if you arrive quickly with lots of paper towels—the club soda may simply act as a carrier in blotting everything up. It may also dilute the wine and help keep the color from setting so that a detergent can finish the job later. A particular stain remover might work better with certain fabrics, so it is a good idea to keep a few cleaning agents, other than club soda, handy. My conclusion: if club soda works, plain water probably works as well. SA

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


ERROR MESSAGES
The attempt to load "Ed" failed.
"Listening To Mrs. Cumberbund" has unexpectedly quit.
Not enough memory to keep program open. Try closing some windows, or alloting more memory to program.
An error has occurred (-413507B)
BUS STOP
SALE 75%
R. Chast