TRANSIT OF VENUS: FIRST IN 122 YEARS • SCIENCE REVIVES FREUD

SCIENTIFIC AMERICAN

Using DNA to Program Synthetic Living Machines

MAY 2004
WWW.SCIAM.COM
$4.95

The TIME Before TIME

The Big Bang Might Not Have Been the Beginning

Do Fuel Cells Make Environmental Sense?

The Pervasive Future of Better GPS

may 2004

contents

features

SCIENTIFIC AMERICAN Volume 290 Number 5

COSMOLOGY

54 The Myth of the Beginning of Time

BY GABRIELE VENEZIANO

String theory suggests that the 13.7-billion-year-old universe we know is only part of an infinite expanse that predates the big bang.

ENERGY

66 Questions about a Hydrogen Economy

BY MATTHEW L. WALD

Fuel cells are generating excitement as clean alternatives for powering automobiles. But the environmental benefits of shifting to a hydrogen-based economy are cloudy.

BIOTECHNOLOGY

74 Synthetic Life

BY W. WAYT GIBBS

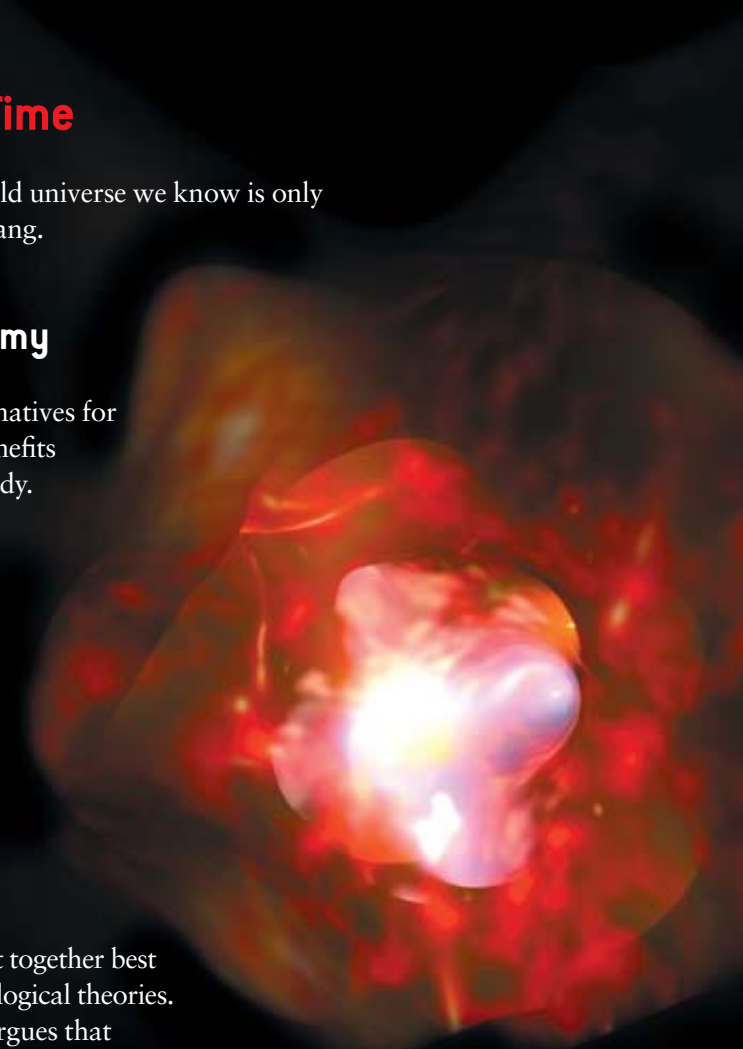
Biologists have tinkered with the genes inside organisms for decades. Now, with "circuits" of interacting DNA, they are beginning to create programmable living machines.

NEUROSCIENCE

82 Freud Returns

BY MARK SOLMS

Modern biological descriptions of the brain may fit together best when integrated with Freud's controversial psychological theories. *Also:* **Counterpoint** from J. Allan Hobson, who argues that Freud's thinking is still highly suspect.



54 Are time and space older than the big bang?

INFORMATION TECHNOLOGY

90 Retooling the Global Positioning System

BY PER ENGE

GPS units already serve more than 30 million users, from hikers to airline pilots. The next wave of improvements will make the technology even more accurate, reliable, useful and ubiquitous.

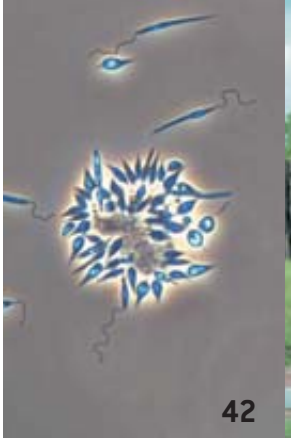
PLANETARY SCIENCE

98 The Transit of Venus

BY STEVEN J. DICK

When Venus crosses the face of the sun this June, scientists will celebrate one of the greatest stories in the history of astronomy.

departments



10 SA Perspectives
The White House bends science to its will.

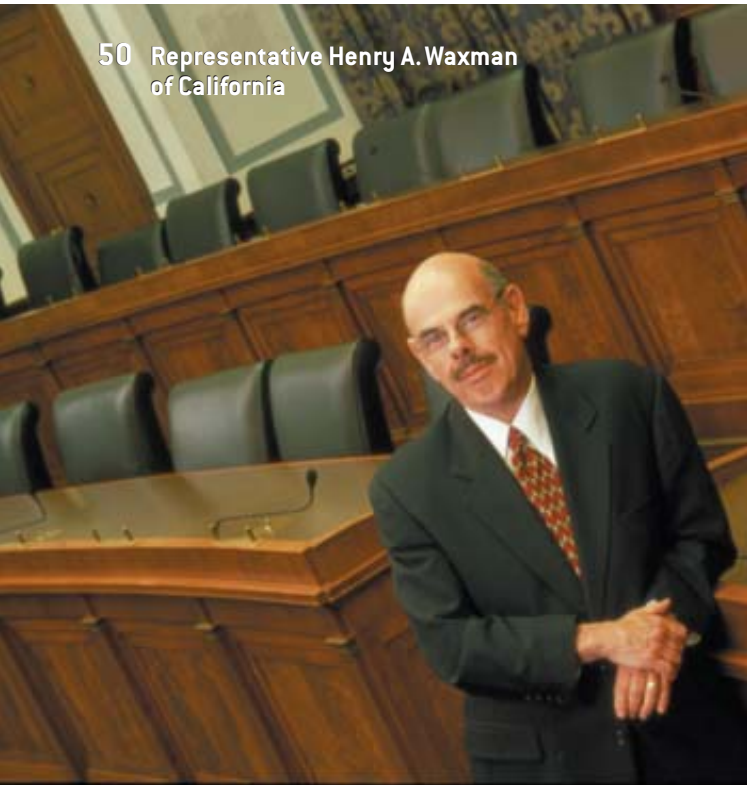
12 How to Contact Us

12 On the Web

14 Letters

18 50, 100 & 150 Years Ago

20 News Scan
- Replacement organs and problematic pigs.
- Planets stripped by their suns.
- A new therapeutic target for Alzheimer's?
- Weeding out superconductivity theories.
- Sea otters clean up after the *Exxon Valdez*.
- A Pacific flow could improve shipping.
- By the Numbers: Fall of the blue-collar class.
- Data Points: Biggest planetoids beyond Pluto.



42 Innovations
A married couple attacks the diseases of the developing world by making drugs, not profits.

48 Staking Claims
Bioprospectors stake their intellectual-property claims in the last frontier, the Great White South.

50 Insights
Representative Henry A. Waxman of California fights against political abuses of science.

106 Working Knowledge
Lasik and other laser eye surgeries.

108 Voyages
Petroglyphs and pristine marshes offer a glimpse of the deep past in Australia's Kakadu National Park.

112 Reviews
A 25th-anniversary special edition of *Machines Who Think* chronicles the fledgling science of artificial intelligence.

columns

46 Skeptic BY MICHAEL SHERMER
The facts never speak for themselves.

118 Puzzling Adventures BY DENNIS E. SHASHA
Jumping to a conclusion.

119 Anti Gravity BY STEVE MIRSKY
How TV can save U.S. health care.

120 Ask the Experts
How are temperatures close to absolute zero achieved? Why is air cooler at higher elevations?

Cover image by Tom Draper Design; MSX/IPAC/NASA (*background*)

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.

Bush-League Lysenkoism

Starting in the 1930s, the Soviets spurned genetics in favor of Lysenkoism, a fraudulent theory of heredity inspired by Communist ideology. Doing so crippled agriculture in the U.S.S.R. for decades. You would think that bad precedent would have taught President George W. Bush something. But perhaps he is no better at history than at science.

In February his White House received failing marks in a statement signed by 62 leading scientists, including 20 Nobel laureates, 19 recipients of the National Medal of Science, and advisers to the Eisenhower and Nixon administrations. It begins, "Successful application of science has played a large part in the policies that have made the United States of America the world's most powerful nation and its citizens increasingly prosperous and healthy. Although scientific input to the government is rarely the only factor in public policy decisions, this input should always be weighed from an objective and impartial perspective to avoid perilous consequences.... The administration of George W. Bush has, however, disregarded this principle."

STANDING UP for science—or stepping on it?

J. SCOTT APPLEWHITE *AP Photo*

Doubters of that judgment should read the report from the Union of Concerned Scientists (UCS) that accompanies the statement, "Restoring Scientific Integrity in Policy Making" (available at www.ucsusa.org). Among the affronts that it details: The administration misrepresented the findings of the National Academy of Sciences and other experts on climate change. It meddled with the discussion of climate change in an Environmental Protection Agency report until the EPA eliminated that section. It suppressed another EPA study that showed that the administration's proposed Clear Skies Act would do less than current law to reduce air pollution and mercury contamination of fish. It even dropped independent scientists from advisory committees on lead poisoning and drug abuse in favor of ones with ties to industry.

Let us offer more examples of our own. The Department of Health and Human Services deleted information from its Web sites that runs contrary to the president's preference for "abstinence only" sex education programs. The Office of Foreign Assets Control made it much more difficult for anyone from "hostile nations" to be published in the U.S., so some scientific journals will no longer consider submissions from them. The Office of Management and Budget has proposed overhauling peer review for funding of science that bears on environmental and health regulations—in effect, industry scientists would get to approve what research is conducted by the EPA.

None of those criticisms fazes the president, though. Less than two weeks after the UCS statement was released, Bush unceremoniously replaced two advocates of human embryonic stem cell research on his advisory Council on Bioethics with individuals more likely to give him a hallelujah chorus of opposition to it.

Blind loyalists to the president will dismiss the UCS report because that organization often tilts left—never mind that some of those signatories are conservatives. They may brush off this magazine's reproofs the same way, as well as the regular salvos launched by California Representative Henry A. Waxman of the House Government Reform Committee [see Insights, on page 52] and maybe even Arizona Senator John McCain's scrutiny for the Committee on Commerce, Science and Transportation. But it is increasingly impossible to ignore that this White House disdains research that inconveniences it.

THE EDITORS editors@sciam.com

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

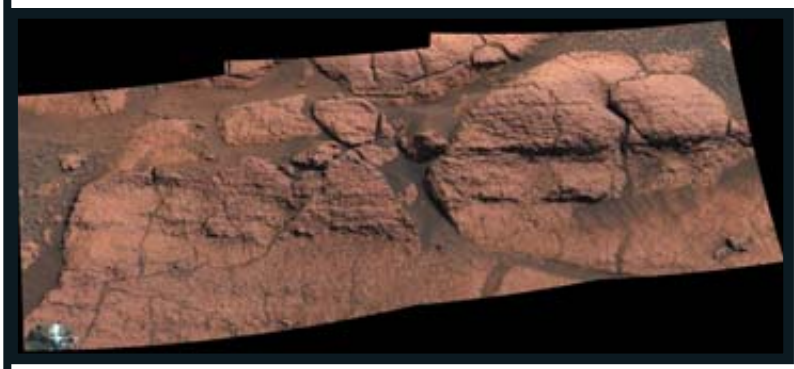
New York
Scientific American
415 Madison Ave.
New York, NY 10017-1111
212-451-8893
fax: 212-754-1138
Los Angeles
310-234-2699
fax: 310-234-2670
San Francisco
415-403-9030
fax: 415-403-9033
Detroit
Karen Teegarden & Associates
248-642-1773
fax: 248-642-6138
Midwest
Derr Media Group
847-615-1921
fax: 847-735-1457
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
Convergence Media
+91-22-2414-4808
fax: +91-22-2414-5594
Japan
Pacific Business, Inc.
+813-3661-6138
fax: +813-3661-6139
Korea
Biscom, Inc.
+822-739-7840
fax: +822-732-3662
Middle East
Peter Smith Media & Marketing
+44-140-484-1321
fax: +44-140-484-1320
Sweden
Publicitas Nordic AB
+46-8-442-7050
fax: +46-8-442-7059
U.K.
The Powers Turner Group
+44-207-592-8331
fax: +44-207-630-9922

On the Web

WWW.SCIENTIFICAMERICAN.COM

FEATURED THIS MONTH

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



COURTESY OF NASA

Mars Rover Reveals Red Planet's "Soaking Wet" Past

Recent data from the Mars rover Opportunity indicate that water once flowed on the Red Planet. "We've been able to read the telltale clues the water left behind, giving us confidence in that conclusion," says principal investigator Steve Squyres of Cornell University.

Fossil Human Teeth Fan Diversity Debate

The discovery in Ethiopia's Middle Awash region of a handful of nearly six-million-year-old teeth is adding fuel to a long-standing debate among scholars of human evolution. At issue is whether the base of our family tree is as streamlined as a saguaro or as shaggy as a shrub.

Ask the Experts

How do sunless tanners work?
Randall R. Wickett, professor of pharmaceutics and cosmetic science at the University of Cincinnati, explains.

SCIENTIFIC AMERICAN®
Established 1845

Letters

EDITORS@SCIAM.COM

ALL SORTS OF THINGS come in small packages, as readers learned from *Scientific American*'s January issue. In "Atoms of Space and Time," Lee Smolin discussed how the universe might be made up of discrete bits. In "Spring Forward," Daniel Grossman wrote about the ecosystem effects of incremental climate change. And in "RFID: A Key to Automating Everything," Roy Want described tiny tracking devices. Whether the theories, consequences and implications wrapped up in these small packages are good things or bad is a matter that inspired many responses. As the letters on the following pages show, a lot depends on your frame of reference.

QUANTUM CONTENTIONS

"Atoms of Space and Time," by Lee Smolin, discussed the theory of loop quantum gravity. One of the results the article expected was that high-energy waves, such as gamma rays from distant astronomical sources, would travel faster than less energetic radiation. But a December 16, 2003, NASA press release (see www.gsfc.nasa.gov/topstory/2003/1212einstein.html) reports that this has been found to be false.

Jacob Rosenberg
NASA

According to a news item in the September 2003 issue of *Astronomy* (see http://arXiv.org/abs/astro-ph/0301184), researchers at the University of Alabama contend that the sharpness of the optical images of distant galaxies indicates that time is not quantized at a value of approximately 10^{-43} second.

This finding is at odds not only with loop quantum gravity but with just about every theory of quantum gravity that I have encountered.

Kelly Mills
Bellaire, Mich.

SMOLIN REPLIES: *Several of us in the quantum-gravity community corresponded with Floyd Stecker, lead author of the research paper cited by Rosenberg. The bounds on the discreteness of spacetime found in the paper do not apply to loop quantum gravity. In particular, the analysis in the paper depends on the assumption that there is a preferred rest frame. That assumption contradicts the basic principles of both classical general relativity and loop quantum gravity. Consequently, Stecker's bounds do not apply to loop quantum gravity. The same considerations apply to bounds deduced by other researchers referenced in Stecker's paper. This case is an example of how new observations and experiments are playing a big role in the field of quantum gravity by ruling out some theories but not others. This is a very good thing—it is real science.*

To address Mills's comment, my understanding is that the research paper's claim is wrong, because the analysis does not model spacetime as a quantum system. Instead it models spacetime as a classical spacetime with ordinary, statistical noise. This would not be predicted by loop quantum gravity and other quantum theories of gravity that treat spacetime as a conventional quantum system.

REFUTING RFID FEARS

"RFID: A Key to Automating Everything," by Roy Want, described how radio-frequency identification chips work and revealed some of the dreams of the technology's advocates. But the sidebar "Dealing with the Darker Side" does not reflect reality, at least for retail.

Want writes that "one of the major worries for privacy advocates" is that retailers and marketers could learn what a customer buys, assuming he or she uses a credit or debit card (or loyalty card). This is not a new issue for consumers and is certainly not one brought on by the introduction of RFID. Retailers have been collecting and using similar information for many years.

The second fear Want discusses is that criminals will use RFID chips to gather information about customers. But the data contained on the chip will be encoded as a string of numbers and more than likely will be encrypted. Unless a criminal can decode the information, it will be useless. In fact, the chips might be comparable to today's bar codes. If you are not concerned about people reading bar codes on packaging in your trash to find out what you ate for lunch, you need not worry too much about RFID chips either. A more important privacy issue would be if criminals searched your trash or computer for bank statements, credit-card information and far more revealing receipts.

RFID could bring significant benefits to consumers, retailers and manufacturers alike. I look forward to the coming of this tiny revolution.

Christopher Allan
London

PATENT PATROL

"In Search of Better Patents," by Gary Stix [Staking Claims], advocates catching invalid patents by a post–grant review. This proposed solution misses the root of the problem, which is inadequate examinations of applications by the U.S. patent office. This failing is the result of incompetent or overworked examiners using inefficient workflow and information systems. Congress has reduced funding for the patent office for years, depriving it of the resources to hire enough fully capable examiners and to upgrade its workflow and information systems. There is no simple and inexpensive procedure for determining whether a device or method is new, useful and nonobvious. The task requires a thorough examination, supported by a complete prior-art search. In other words, it requires a well-funded and well-managed patent office.

John Stewart, patent agent
Orlando, Fla.

TROUBLING TEMPERATURE TRENDS

I am skeptical about highly charged packaging of the global-warming concept, so I appreciated Daniel Grossman's "Spring Forward." It avoids extending beyond clearly verifiable facts into the more bold claims and, ultimately, the moral and political arguments of radical environmentalists. I have a few questions regarding ecosystem shifting caused by global warming. First, haven't paleontology and evolution theory shown us the importance of fluctuation in the earth's environments to the evolution of life? Second, although we can detect extinctions of known species relatively easily, isn't it true that we have no easy way to detect the gradual and ongoing emergence of new species? And last, why be concerned about the extinction of species that evolution has pushed to fill very narrow, unstable niches? Shouldn't we expect that changes in local environments caused by global warming would open new niches to be filled by existing species waiting in the wings?

ADÉLIE PENGUIN **population around Palmer Station, Antarctica, is dropping.**

JOHN CONRAD *Corbis*

Jim Carnicelli
via e-mail

What, if anything, is being done to save the Adélie penguins? And what can the average person do to help protect plant and animal species that are endangered by global warming?

Doris Black
via e-mail

GROSSMAN REPLIES: *Regarding Carnicelli's points: It is true that if we adopt the perspective of a geologic timescale of, say, millions of years, then human-caused climate effects may not result in something that is "worse," just different. But on a human timescale of hundreds of years, we may suffer reduced biodiversity that could include losses of species that people appreciate, such as certain songbirds and New England's maple trees.*

In response to Black's query, there are no efforts to save the Adélies around Palmer Station. By way of explanation, penguin scientist Bill Fraser said to me, "How can you 'save' a species that is being negatively affected by what is actually a global-scale problem—climate warming?" Fortunately, Adélies exist in large and undiminished numbers in the Ross Sea and elsewhere. Readers who share Black's concern have a powerful tool at their disposal: energy conservation.

ERRATA: "In Search of Better Patents," by Gary Stix [Staking Claims], inaccurately reported that during a reexamination process before the U.S. patent office a patent holder can request to broaden the claims of a patent. It should have stated that the petition can be made to amend existing claims or add new ones, but the overall scope of the patent cannot be expanded.

Bill Fraser, a penguin scientist mentioned in "Spring Forward," by Daniel Grossman, is no longer affiliated with Montana State University. He now works through the Polar Oceans Research Group, a nonprofit organization.

Answers to This Month's Puzzle [*see page 118*]:

For the three-by-three grid, leaving any three corner squares empty at the start of the game will ensure that only one counter will remain at the end. For the four-by-four grid, you can start with one empty square anywhere in the grid and achieve the same goal. In the Jump Snatch game shown, the Jumper will win if he makes two jumps in the fourth move. For a full explanation of the May puzzle and for future puzzles and their solutions, visit www.sciam.com

50, 100 & 150 Years Ago

FROM *SCIENTIFIC AMERICAN*

Deathly Dust ▪ Living Clock ▪ Killer Whale

MAY 1954

RADIOACTIVE FOOD—"The second thermonuclear experiment at the explosion grounds in the Marshall Islands was said to be 600 times as forceful as the Hiroshima atomic bomb. The immediate brunt fell on a Japanese fishing vessel called *The Fortunate Dragon,* carrying a harvest of tuna and shark in its open hold. Caught 80 miles from the explosion, it was showered with a white ash of particles which blistered the 23 fishermen's skin and made the fish radioactive. When the ship made port, some of the fish were sold before the government could stop it. Overnight the Japanese people stopped eating fish; housewives shopped with Geiger counters; the price of tuna fell to one third with few takers. The Japanese newspapers looked upon the shower of 'death dust' as the third atomic bombing of Japan."

GOLLY . . . THEY *DID* LIKE IKE!—"But for the 1948 Democrats who left their party, General Eisenhower would not have gone to the White House. What were the motives behind this great swing of voters to the Republican candidate? A nationwide study was undertaken to provide as full an answer as possible to that intriguing question. A sizable number in each group appeared 'non-partisan' on the candidates' personal qualities, yet among strikingly large percentages of each group of voters, the General held high favor over Governor Stevenson. This strong leaning to Eisenhower as a person appears to have been the one factor which united all the groups that voted for him."

MAY 1904

FLOWER CLOCK—"The Louisiana Purchase Exposition opened at St. Louis, commemorating one of the most important centennials in American history. Its floral clock will be sixteen times larger than any timepiece in the world [*see illustration*]. It will keep accurate time, for beneath the vines and other plants, skilled artisans have constructed machinery similar to the works of a watch. The hands are long steel troughs, in which fertilized earth has been placed to supply nourishment to the vines that will cover the metal. The numerals of the hours will be dark tall foliage plants."

CLOCK made of flowers, St. Louis, 1904

HYDROELECTRICITY AND CO_2—"In San Francisco the cost of electric current for power and light is almost exactly one-seventh of what it was a few years ago, and it is possible to deliver at the factory on the coast, from the melting snows and glaciers of the Rockies, power at a smaller cost than that procured from steam. It has been estimated that the quantity of carbonic acid annually exhaled by the population of New York City is about 450,000 tons, and that this amount is less than three per cent of that produced by the fuel combustion of that city; so we may expect that, with the removal of this great source of contamination of the atmosphere, even the air of our greater cities will be practically as pure as that of the country."

MAY 1854

ORCA—"Lieut. Maury said that Captain Royes, a New England whaleman, wrote him a letter describing sixteen kinds of whales, one of them a strange fish, which the Lieutenant did not find named in any of the books. The Captain called it the 'Killer Whale,' and described him as thirty feet long, yielding about five barrels of oil, having sharp, strong teeth and on the middle of the back a fin, very stout, about four feet long. This 'Killer' is an exceedingly pugnacious fellow. He attacks the right whale, seizing him by the throat, biting till the blood spouts, or till another 'Killer' comes by and eats out the tongue of the tortured fish. This tongue of a right whale is an oily mass, weighing three or four tons. The 'Killer' scours the ocean from pole to pole, is in every sea, and all old whalemen have met him."

BIOTECH

Body Building

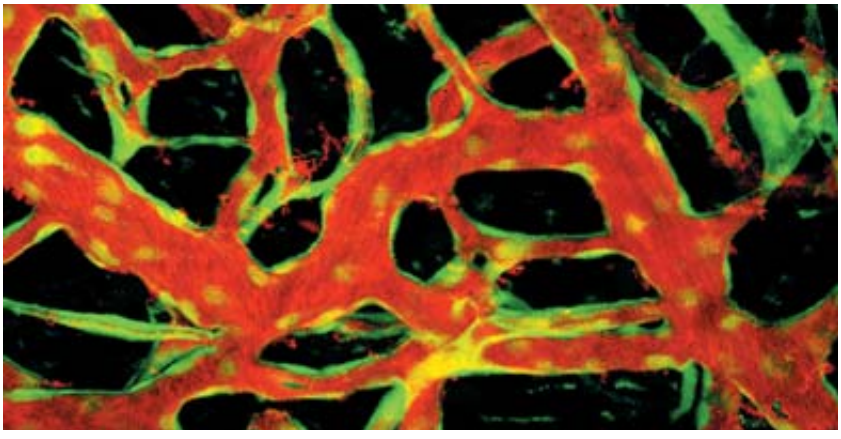
GROWING REPLACEMENT ORGANS IS STILL A LONG WAY OFF **BY CHRISTINE SOARES**

Six years ago Michael Sefton of the University of Toronto challenged his colleagues in the fledgling field of tissue engineering to build a functioning human heart within 10 years. With the isolation of human embryonic stem cells later that year, Sefton's challenge seemed all the more relevant: stem cells, after all, are nature's starting point for building working organs.

Now Sefton admits that the deadline on his Living Implants from Engineering ("LIFE") initiative was naive, and he thinks it will be at least another 10 to 20 years. "We need to be able to walk before we can run," he says, "and the worry today is, Can we make a vascularized piece of tissue or a tissue with two or three cell types in a controlled way?"

Thin sheets of skin and single blood vessels have been grown in the laboratory, and some versions have already been put through human clinical trials. Yet any whole organ would be a complex three-dimensional edifice comprising specialized cells, nerves and muscle, all interwoven with a dense web of veins and capillaries diffusing oxygen and nutrients. The main hurdles have been just getting multiple cell types to grow and work in harmony and spurring formation of the blood vessels required to nourish tissues more than a few hundredths of a millimeter thick.

By mimicking the natural 3-D shape in which an organ grows, tissue engineers are trying to get adjacent cells to "talk" to one another and complete the task of building the desired tissues. This approach has yielded "ink-jet"-dispensed dollops of cell aggregates "printed" in simple patterns that flow together, linking up into larger pieces of tissue. The next step will be to "print" designs using multiple cell types and eventually to print them layer on layer to create larger structures. A similar technique suspends living cells in a clear hydrogel matrix that can be layered or molded into 3-D shapes. Neither tactic has yielded the all-important vascular network needed to sustain thicker tissues.



BLOOD WORK: Rakesh K. Jain of Harvard Medical School grew this web of blood vessels inside a mouse on a scaffold seeded with human vascular endothelial cells (*green*) and muscle precursor cells. Infusing artificial organs with such complex vasculature has proved more difficult.

N. KOIKE ET AL. IN *NATURE*, VOL. 428, PAGES 138–139; 2004

WHEN HUMANS MEET PIGS

Custom-grown spare parts from stem cells are years away. That means animal organs may be the only realistic alternative for patients awaiting transplants. But xenotransplantation took a serious blow in January, when Jeffrey L. Platt of the Mayo Clinic and his colleagues confirmed that a virus present in most pigs, porcine endogenous retrovirus (PERV), could infect human cells in vivo. PERVs are harmless to pigs, but no one knows how they might react when transplanted into humans. The Mayo team injected human stem cells into fetal swine; after the pigs were born, the researchers found that PERV infected the host cells as well as the human cells. What is more, they detected chimeric cells containing fused pig and human DNA that were positive for PERV, too.

More progress has been made by seeding stem cells onto a variety of simple scaffolds impregnated with growth-promoting chemicals. Last fall, for example, researchers from the Massachusetts Institute of Technology and the Technion-Israel Institute of Technology reported generating tissues of neural, liver and cartilage cells, as well as formation of a "3D vessel-like network" on a biodegradable polymer scaffold seeded with human embryonic stem cells. When transplanted into a mouse, the constructs remained intact and appeared to connect with the animal's blood supply.

"RENAL UNIT"—a proto-kidney—produced urinelike liquid after 12 weeks of growth.

Still, scientists working with stem cells, embryonic or otherwise, admit that they are just beginning to learn tricks for controlling the kind of tissue the cells become and just starting to discern the cues cells give to one another as well as take from their natural environment during the course of organ development. "We don't have anything like [nature's] exquisite repertoire of tools," Sefton says.

And so most models for growing entire organs involve using some kind of living "bioreactor." In some cases, it could be the same patient in need of the organ. Anthony Atala of Wake Forest University, who once grew a simple bladder in a beaker and transplanted it into a dog, teamed up more recently with Robert P. Lanza, also now with Wake Forest, and others to grow a mini kidney inside a cow. Kidney progenitor cells were taken from a fetal clone of the cow in question, then implanted into the cow's body, where they developed into proto-organs with all the cell types of a normal kidney. These "renal units" even produced a urinelike liquid.

The idea of seeding an organ and letting the body do the rest of the construction might work for a kidney, because the patient could be treated with dialysis while the new organ was being generated, according to Jeffrey L. Platt, director of transplantation biology at the Mayo Clinic. For a patient suffering from lung or heart failure, however, growing a new organ would put too much strain on an already weak body. But every advance toward creating ever more complex tissues might yield a lifesaving patch for a moderately damaged heart or liver, Platt says, along with fresh insight into how nature builds bigger body parts.

COURTESY OF ROBERT P. LANZA AND ANTHONY J. ATALA

ASTRONOMY

Burning Down to Rock

GAS GIANTS MIGHT GET COOKED CLEAN TO THEIR SOLID CORES **BY CHARLES CHOI**

The first rocky worlds astronomers detect circling other stars could resemble Inferno more than Earth. The existence of such lava-coated planets, which may prove commonplace, will force a reconsideration of theories about planetary formation.

Since 1991 observers have discovered some 120 exoplanets—worlds outside our solar system. All but three appear, by their great size and low density, to be gas giants. Roughly a sixth are "hot Jupiters" surprisingly near their stars, all closer than Mercury is to our sun.

Some hot Jupiters live just too close to their stars for comfort. Last year the Hubble Space Telescope provided the first evidence of an evaporating atmosphere, from an exoplanet, HD 209458b, that circles its star at a distance of less than 1/20 the distance between the sun and Earth. The star roasts the exoplanet and rips at it with its gravity. The result: the exoplanet blows away at least 10,000 tons of gas a second, which streaks off in a vast plume 200,000 kilometers long. Astronomer Alfred Vidal-Madjar of the Institute

of Astrophysics in Paris and his team dubbed the world "Osiris," after the Egyptian god torn to pieces by his evil brother Set.

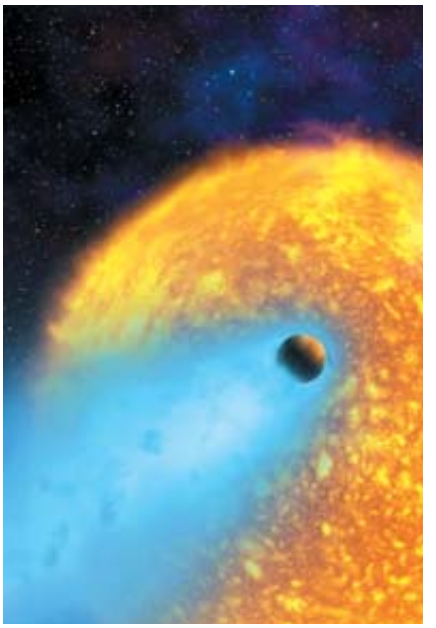
In contemplating the fate of Osiris, Vidal-Madjar and his team calculated how long it and other giants might live. At roughly 220 times Earth's mass, Osiris boasts a gravitational pull strong enough to hold its atmosphere until its star dies. But the researchers speculate the hellish rate of evaporation might completely scour all gas off smaller hot Jupiters or those closer to their stars than Osiris.

This could lead to a new class of planets—a dead giant's hard, bare heart. The astronomers named such worlds "chthonians," after primeval Greek deities of the underworld. In findings to appear in *Astronomy and Astrophysics*, astronomer Alain Lecavelier des Etangs of the Institute of Astrophysics and his co-workers figure that the four exoplanets discovered so far may one day become chthonians.

Though remnants of far larger worlds, chthonians would still weigh in at roughly 10 to 15 times Earth's mass and six to eight times Earth's diameter. With searing temperatures of roughly 1,000 degrees Celsius at their surfaces, they would look "like lava planets," Lecavelier des Etangs imagines. If chthonian exoplanets exist, "it is probable that they will be the first rocky planets to be detected around other stars," Vidal-Madjar remarks. (Three planets, two about three to four times Earth's mass and the third twice the mass of the moon, were discovered in the 1990s and most likely are solid, but they all orbit a pulsar.)

Spotting chthonians would help answer questions regarding planetary formation, explains astronomer Adam Burrows of the University of Arizona. Researchers think that worlds are born from disks of gas and dust encircling stars. The most popular idea proposes that solid cores amass from protoplanetary disks and behave like seeds, attracting gas to grow into giant planets.

The alternative theory suggests that giant planets may not possess hard cores. Instead they may have fluid centers, after having condensed directly from protoplanetary disks without forming solid



GAS GIANTS may lose their atmospheres to their stars, resulting in rocky worlds called chthonians.

ESA, ALFRED VIDAL-MADJAR *Institute of Astrophysics* AND NASA

hearts. Scientists have not conclusively identified whether the centers of giants in our own solar system are solid. Detecting chthonians could prove one scenario of planetary formation right.

The European Southern Observatory telescope in Chile has an outside chance of finding them next year: a new instrument there could detect planets as low as about 15 times Earth's mass by looking for the gravitational tugs each has on its star. The best chance to spot chthonians will come from the first space probes sensitive enough to see Earth-size planets: the French satellite COROT, scheduled for launch in 2006, and NASA's Kepler, around 2007. These missions might uncover several tens of chthonians, probably by spotting them when they pass in front of their stars, dimming them.

Burrows thinks that chthonian exoplanets may not turn out to be all rock. If a chthonian's star does not strip off its atmosphere, ices found in a giant's core might survive underneath. Lecavelier des Etangs says that chthonians might even support life, although it would almost certainly be "very different from what we know on Earth."

Charles Choi, a frequent contributor, is based in New York City.

NEUROSCIENCE

Downsized Target

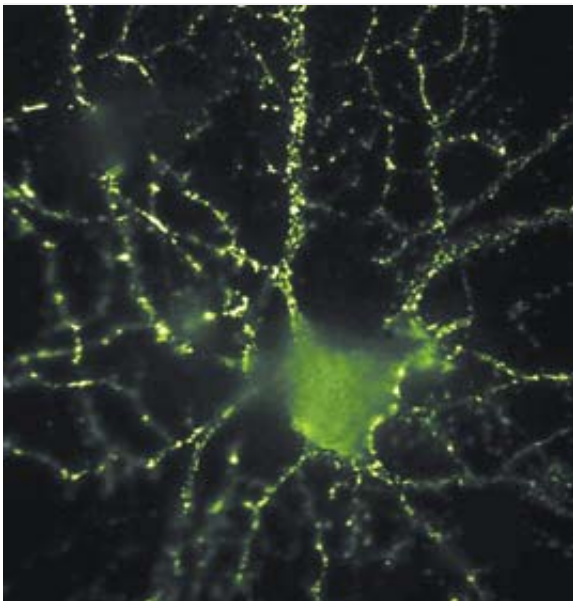
A TINY PROTEIN CALLED ADDL COULD BE THE KEY TO ALZHEIMER'S **BY TOM VALEO**

Scientists have long suspected that the protein clumps and tangles identified by Alois Alzheimer in 1907 somehow cause the disease that bears his name, probably by killing neurons. Now some researchers are blaming a much smaller form of protein, one that apparently produces memory deficits merely by binding to neurons and disrupting their ability to transmit signals. The search has begun for an antibody that would destroy these tiny proteins—or ADDLs—thereby preventing the onset of Alzheimer's disease and possibly even reversing the early symptoms.

The discovery of ADDLs explains glaring anomalies in the conventional thinking about Alzheimer's, which holds that fragments of amyloid precursor protein, produced by normal neurons, aggregate into sticky, insoluble plaques that damage neurons. The problem with this theory is that virtually every older person carries some amyloid plaque, but only a few develop Alzheimer's. Conversely, those with Alzheimer's often have relatively few plaques. Another proposed culprit is the presence of tangles of tau protein, which form inside neurons and coincide with the collapse of microtubules that support the cell body and transport nutrients. The tau tangles correlate much better with the disease but tend to appear later, suggesting that they are a consequence, not a cause.

In 1994 Caleb E. Finch, a neurogerontologist at the University of Southern California, attempted to create amyloid plaque by mixing a solution of amyloid precursor protein fragments with clusterin, a substance produced at higher levels in the brains of people with Alzheimer's. The clusterin did not trigger the formation of amyloid plaques, but the resulting solution profoundly disrupted the ability of the neurons to transmit signals.

Finch reported this finding to Grant A. Krafft and William L. Klein, two colleagues at Northwestern University, who set out to discover what was in the solution. Using an atomic-force microscope, they obtained extraordinary pictures of globules no one had ever seen. "They looked like little marbles," Krafft recalls. "It turned out these globules contained only a few of the amyloid peptide building blocks, whereas the long fibrils contained thousands, if not millions, of these subunits." The three scientists decided to call the substance ADDL, which stands for amyloid beta-derived diffusible ligand. (The molecule is derived from amyloid precursor protein; it diffuses throughout the brain instead of aggregating into fixed plaques; as a ligand, it attaches to receptors on neurons.)



ALZHEIMER'S ATTACK? Toxic proteins known as ADDLs (*yellow spots*) affix themselves to a human neuron. They may cause memory problems by disrupting the signals between neurons.

Klein developed an antibody that revealed how ADDLs attach to dendrites in the hippocampus, thereby disrupting signals needed to produce short-term memories. And last summer Klein, Krafft, Finch and their colleagues found huge quantities of ADDLs in postmortem brains from people with Alzheimer's, whereas brains from normal patients were virtually free of ADDLs. What is more, they discovered that neurons of mice functioned normally once the ADDLs were removed.

The obvious solution to treat Alzheimer's

GOOD FOR AMYLOID AND TAU

Protein globules called ADDLs shift the blame for Alzheimer's disease from amyloid plaques themselves to the tiny molecules that create them. Dennis J. Selkoe of Harvard University, who helped to develop experimental vaccines and other treatments against amyloid plaques, now believes that the globules are a more likely basis for the synaptic failure. But he thinks that the ADDL idea buttresses, not replaces, conventional thinking. "We amyloid aficionados consider it a refinement of the amyloid story," he remarks. "This is just an evolution of the theory that amyloid basically causes the disease."

Investigators of tau protein, another possible Alzheimer's culprit, also embrace ADDLs, because ADDLs provide a plausible explanation for the production of tau tangles. "I think tau is essential to Alzheimer's disease," maintains molecular biologist Lester I. Binder of Northwestern University, "but I think some form of amyloid is the trigger. That form could be ADDLs."

Y. GONG ET AL. IN *PNAS*, VOL. 100, NO. 18; SEPTEMBER 2, 2003

disease, in Krafft's opinion, is to remove the ADDLs or prevent them from forming. Attempts to eradicate amyloid plaques are misguided, he believes, and any attempt to intervene after neurons have started to die comes too late to do much good. "It's pretty clear to me that we're wasting about 90 percent of the Alzheimer's research budget on things that are worthless," he says.

While crafting their theory, Krafft, Klein and Finch acquired patent rights to ADDLs and formed their own corporation, Acumen Pharmaceuticals, which recently formed a partnership with Merck. "By partnering with Merck, Acumen can get the antibody and vaccine products to market much faster than if we tried to do it by ourselves," Krafft explains.

Merck has committed up to $48 million to Acumen for the right to develop an antibody against Alzheimer's and another $48 million if it succeeds in bringing to market a viable vaccine. That money, plus funding from other investors, will enable Acumen to devise three other ADDL-based strategies for preventing Alzheimer's, as well as diagnostic tests that would reveal early signs of the disease.

Tom Valeo, based in the Chicago area, writes a column on aging for the St. Petersburg Times *in Florida.*

PHYSICS

High-Temp Knockout

GONE: TWO POSSIBLE SUPERCONDUCTING "GLUES"

BY GRAHAM P. COLLINS

In the 18 years since they were discovered, high-temperature superconductors have remained an enigma. These copper oxide ceramics conduct electricity without loss at temperatures far higher than those needed for conventional superconductors, albeit still far below room temperature. Physicists know that in both types of material, the superconductivity is caused by electrons pairing up and gathering en masse in a single collective quantum state. But they do not know what "glue" causes the pairing in the high-temperature ("high-Tc") superconductors. Numerous ideas have been proposed, but none has been proved. A recent experimental study suggests that two important theoretical possibilities can be eliminated.

In low-temperature superconductors, the crucial interaction among the electrons is mediated by vibrations of the metal's lattice of positive ions. One electron distorts the lattice as it passes by, and microseconds later the distortion influences the electron's partner when it arrives on the scene. The lattice vibrations are called phonons—they behave just like particles, and their emission and absorption by the electrons generate a weak attractive interaction. Physicists refer to this conventional model as the BCS theory, after the scientists who worked out the mathematics in 1957.

After the discovery of high-Tc superconductors in 1986, physicists quickly determined that the unadorned BCS theory could not explain the behavior of the new materials. To begin with, thermal vibrations from high temperatures should overwhelm any attraction produced by phonons. (More recently, however, this limit on the critical temperature has been questioned.) Second, substituting different isotopes in a BCS superconductor changes the characteristics of the phonons (heavier atoms should vibrate more slowly) and consequently changes the critical temperature by a precise amount. The high-temperature superconductors change by different amounts. Other detailed features are also hard to explain with BCS.

Recently physicists have been studying a "kink," or bend, that appears in graphs that plot the energies of paired electrons as a clue to the force that causes the pairing. Many researchers have related the kink to a type of collective state among the electrons called a magnetic resonance. One experimental group has

made a case for phonons to be the cause of the kink—a result that would upset the conventional wisdom about unconventional superconductors.

Results of experimenters at McMaster University and Brookhaven National Laboratory seem to eliminate both the magnetic resonance and phonons as the glue. In this group's experiment, infrared light was shone on the superconductor, and the amount of light scattered at each wavelength provided a measure of the energies of the paired electrons. The physicists, led by Thomas Timusk of McMaster, found both a sharp peak in scattering at a particular frequency and a broad background of scattering across all frequencies. The sharp peak is clearly related to the kink seen in the other experiments, but it disappeared from view in so-called overdoped material, which has too many oxygen atoms for optimal superconductivity. (Overdoped materials superconduct, but at lower temperatures as the doping increases.) That rules out phonons as the cause of the peak and the kink; phonons should remain present in all materials, even the overdoped ones. Nor can phonons be responsible for the broad background: if they were, the background would cut off at high frequencies, which it does not.

The sharp peak's behavior—the conditions under which it is present—correlated well with what was expected for a magnetic resonance. But there's a gotcha: its disappearance in overdoped materials that nonetheless still superconduct. Consequently, it cannot be the cause of the superconductivity.

That leaves the broad background, which Timusk and his co-workers think is likely to be a signal of whatever process really is binding the electrons together in pairs. Michael Norman, a materials scientist at Argonne National Laboratory, argues that although this glue cannot be the much studied magnetic resonance, there are good reasons for believing it is magnetic in nature. And so the quest goes on. Two contenders are knocked out, but the puzzle remains.

ECOLOGY

The Oil and the Otter

SEA OTTERS CLEAN UP AFTER THE *EXXON VALDEZ* SPILL—AND GET SICK DOING SO **BY SONYA SENKOWSKY**

It has been 15 years since the *Exxon Valdez* oiled Alaska's Prince William Sound, and more than 12 since the last of the official restoration workers took off their orange slickers and headed home. But at least one cleanup crew never left the Sound: sea otters. The creatures, which were hit especially hard by the first effects of the spill, continue to feed on clams and other food in areas that still contain pockets of oil. Their diligent digging is helping release trapped petroleum—which appears to be sickening them. Ecologists are left

GREASY EATS: By digging for food, sea otters in Prince William Sound are cleaning up what remains of the mess left by the *Exxon Valdez*. The oil components are poisoning the otters.

JACK SMITH *AP Photo*

with a dilemma: remove the oil (and possibly cause more harm to the Sound) or let the animals continue to do the dirty work and pay the price.

Scientists had originally predicted that any remaining oil would have been carried by waves to shorelines by now. There exposure to air would transform the oil into a hardened asphalt residue lacking the more volatile and toxic components. "The assumption was that the oil wasn't subsurface, it wasn't low, it was up there in that 'bathtub ring,' and that's where the cleaning effort was focused," explains Stanley D. Rice, a laboratory program manager with the National Oceanic and Atmospheric Administration's Alaska Fisheries Science Center in Juneau.

But in 2001, with some animals continuing to show indications of oil exposure, NOAA researchers dug into those beaches and found far more *Exxon Valdez* oil than expected—much of it still liquid—in about 70 percent of the sites. The remaining residue "still has a pretty high complement of the toxic components of oil," remarks team leader Jeffrey W. Short.

Sea otters, which feed on clams, mussels and other invertebrates, reach their prey by diving and digging underwater pits. One otter can create thousands of pits in a year, moving five to seven cubic yards of sediment a day. These excavations release oil from surrounding sediment, helping it disperse, explains U.S. Geological Survey research wildlife biologist James L. Bodkin. He has been studying a group of about 70 sea otters from northern Knight Island, a region that lost 90 percent of its sea otter population after the spill. The otters are no longer becoming coated in oil and dying from hypothermia, but there is evidence that they are ingesting the contaminants. Researchers have recorded life spans reduced by between 10 and 40 percent compared with before the spill and noted swollen and discolored livers in some dead otters.

The sacrifices of today's sea otters, however, should have their benefits, Rice observes: "The [otters] that are new and coming along, they're going to

be entering a habitat that's cleaner." Decreasing levels of an enzyme called cytochrome P450-1A in the animals' blood, produced in response to toxic chemicals, indicate that an end to the prolonged oil exposure is near, according to USGS physiologist Brenda E. Ballachey and Purdue University pathologist Paul W. Snyder. "While they're still being exposed, there is less and less oil there every year," Rice notes.

With the possibility of seeking further restoration funds from Exxon on the horizon, scientists are debating whether a cleanup makes sense. "I think that if we had asked this question and had the data we have now several years ago, we probably would be out there cleaning up," Rice states. The effort generally involves mechanical tilling—essentially, plowing the affected area with heavy machinery. The method turns the ground and releases trapped oil, which is then broken down by microorganisms.

But the time may be fast approaching, Rice adds, when such intervention may not be wise. Although human cleanup efforts would more quickly make feeding safer for sea otters and other foragers, such as harlequin ducks, they would physically disrupt the environment and would not be beneficial to all organisms. "Maybe on some marginal beaches, you would do more harm than good," Rice surmises. "What might be a good idea for otters may not be a good idea for a clam or a mussel. There is no obvious choice."

Sonya Senkowsky, based in Anchorage, Alaska, may be reached at sonya@alaskawriter.com

OCEANS

Splash of Cold Water

NEWFOUND EDDY EXPLAINS MYSTERIOUS FLOWS

BY CHRISTINA REED

Neighbors often trim only the part of a tree that is growing over their own property lines. For decades, Japan and South Korea acted similarly, staying within their exclusive economic zones when studying the Sea of Japan, or the East Sea, as the Koreans refer to it. Then, in 1999, oceanographers from the two nations teamed up with the U.S. Navy to explore the Japan/East Sea in the first long-term underwater study of its circulation.

Now the team is showing abundant fruit from its labor. What the researchers uncovered changes the perspective of the ocean basin between the two Asian countries: a cold-water eddy swirling in and out where no one had noticed it before. Named after one of the islands in the Ulleung Basin, the Dok Cold Eddy explains previously misunderstood flows in the Sea of Japan that may help naval operations, commercial shipping and fishing.

"We found that this eddy has an extreme impact on the circulation of the entire Japan/East Sea," says Douglas A. Mitchell of the Naval Research Laboratory (NRL) at Stennis Space Center in Mississippi. Mitchell, who identified the Dok Cold Eddy earlier this year at the oceans meeting of the American Geophysical Union, notes that it had been overlooked even in the satellite data because of the political boundaries.

The investigators discovered the Dok Cold Eddy using instruments called inverted echo sounders stationed on the seafloor from June 1999 to July 2001. The devices measured the time it took for signals to bounce off the sea surface and return. The time interval depends on the density of water, which in turn depends on temperature. Mitchell converted the acoustic measurements into temperature and velocity profiles of the currents in the Sea of Japan. During the two-year period, an eddy 60 kilometers in diameter propagated in and out of the basin beginning in the north near Dok Island. A

DOK COLD EDDY emerges after cold water flows are pinched between Ulleung and Dok Islands. It moves southwest to the Korean coast, where it eventually disperses. By diverting warm flows, a persistent eddy can cool much of the basin. The eddy is about the size of those seen in the Gulf of California (*below*), made visible here by phytoplankton.

SEAWIFS PROJECT, NASA GODDARD SPACE FLIGHT CENTER AND ORBIMAGE (*satellite image*); EDWARD BELL (*map*)

closer look at past satellite data showed this local feature coming and going over a nine-year period.

Eddies abound throughout the ocean often as ephemeral features, but some, like the Dok Cold Eddy and those that spiral off the Gulf Stream, repeatedly appear. Understanding eddies is "a ripe area of research," comments ocean physicist Tommy Dickey of the University of California at Santa Barbara. "The biogeochemistry of eddies and even the physics behind them isn't all understood at this point," he notes.

The U.S. Navy contributed $2 million to the work for strategic purposes: Eddies create a density contrast with a wall of water behind which a submarine can hide. Instead of taking a straight path, the vessel's noises are refracted from the eddy, creating the impression to anyone listening that the sub is in a different location. "Submarines can hide behind these features and follow an eddy out," Mitchell says. Understanding eddy formation and movement will also improve the interpretation of acoustic noises.

The results could help commercial activity. By keeping tabs on the circulation patterns, shipping industries could better manage any hazardous spills as well as boost efficiency. As study co-author William Teague of the NRL puts it, "It is more expensive to drive against a current."

Local fisheries in the Sea of Japan are dependent on the body of water's physical properties for catching temperature-sensitive species. The sea contains a highly productive ecosystem that is difficult to regulate as a result of an uneven distribution of fish. Understanding the relation between the Dok Cold Eddy and temperature-sensitive animals may help improve fishery management in both Korea and Japan. In the second year of the monitoring experiment, the cold-water eddy diverted the northward warm-water current, preventing its return for five months. In previous years, fishermen blamed colder winters for the cold currents and resulting poor catches.

Just how the eddy helps or hinders the local ecology and fish stock in the sea will take further study. But with Japanese and Korean scientists continuing to look beyond their boundaries, together they will improve their knowledge of the important resource they have between them.

Christina Reed, who writes on ocean and planetary science, was the science coordinator for James Cameron's upcoming IMAX movie about extreme life at hydrothermal vents.

SCOPING THE FISH MARKET

Eddies are traditionally studied for their physical properties, but more recently their impact on ocean biology is under investigation. Ocean eddies have the capability of supplying new nutrients to a region. They may account for from 5 percent to "as much as 50 percent of the primary production in the open ocean," says oceanographer Claudia Benitez-Nelson of the University of South Carolina, referring to the rate of photosynthesis by phytoplankton and other organisms that form the base of the marine food chain. Picking a specific eddy to study, however, is difficult—like setting up a study of a future hurricane. Most often oceanographers look for areas where eddies regularly form and plan a cruise with satellite assistance to determine where and when they might occur.

BY THE NUMBERS

Blue-Collars in Eclipse

PRODUCTIVITY LED TO WORKING-CLASS DECLINE **BY RODGER DOYLE**

In 1840 manufacturing and other manual labor industries employed about 17 percent of the U.S. job force. These employees consisted of a heterogeneous group encompassing artisans, ditchdiggers, sailors and others who worked with their hands. The American blue-collar class began to take shape in the early 20th century, when management engineers wrested control of the manufacturing process from skilled laborers such as machinists to take advantage of the proliferating number of new tools. Through time-and-motion studies, they also prescribed the precise way people should do their jobs.

This "scientific management" in part created assembly-line production, which greatly increased productivity by eliminating the older rhythms of work. But the technique helped to generate millions of boring, closely supervised jobs. Some of the tasks required special clothing, including, in some cases, blue protective gear, which gave the class its name.

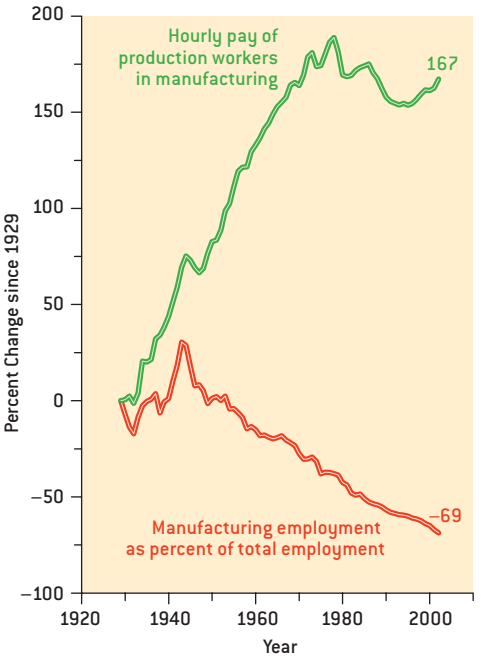
By the 1930s, with the coming of the New Deal and its pro-labor legislation, it might have seemed that workers would soon dominate the country, for they were organized, motivated and numerous. But American labor never became politically dominant, unlike labor in several European countries [see By the Numbers, May 1999]. In 1943, the peak year in terms of their numerical importance, blue-collar employees accounted for at least 40 percent of the job force.

The chart sums up the 20th-century history of American workers in manufacturing, by far the most important employer of blue-collars. Their relative importance has declined almost without pause since 1943. The drop traces primarily to vastly increased productivity: for example, the productivity of the average manufacturing worker in 2003 was 5.1 percent higher than in 2002. Competition from developing countries, often cited as the reason for the decline in manufacturing, has been a secondary factor [see By the Numbers, May 2002]. Real wages have generally stagnated in recent decades. But wages in manufacturing have trailed those in other blue-collar occupations such as construction and transportation, perhaps because manufacturing is less well protected against foreign competition.

Shifting perceptions among blue-collars themselves also drained their power. Sociologist David Halle of the University of California at Los Angeles showed that blue-collars tended to think of themselves as working folks united in opposition to plant management. But outside of the factory, they gravitated toward middle-class attitudes typical of white-collar employees, particularly if they were homeowners. Modern company practices, such as profit-sharing, also probably made it easier to lower working-class consciousness.

Manufacturing employment parallels trends in agriculture, which employed 63 percent of the workforce in 1840 compared with about 2 percent today. It would not be surprising if blue-collar jobs in manufacturing, now at about 8 percent, fell to the same 2 percent level before the 21st century ends.

Rodger Doyle can be reached at rdoyle2@adelphia.net



SOURCE: Calculated from data supplied by the Bureau of Labor Statistics. Numbers in the chart refer to 2002 statistics.

THE BULK OF BLUE-COLLARS

Production workers in 2002, in thousands:

Manufacturing: 11,217

Construction: 5,196

Transportation/Warehousing: 3,611

Utilities: 478

Mining: 436

SOURCE: Bureau of Labor Statistics

FURTHER READING

America's Working Man: Work, Home, and Politics among Blue-Collar Property Owners. David Halle. University of Chicago Press, 1984.

A Social History of the Laboring Classes: From Colonial Times to the Present. Jacqueline Jones. Blackwell Publishers, 1999.

American Workers, American Unions: The Twentieth Century. Third edition. Robert H. Zieger and Gilbert J. Gall. Johns Hopkins University Press, 2002.

RODGER DOYLE

DATA POINTS: BIG BEYOND PLUTO

This winter astronomers apparently discovered the two largest planetoids beyond Pluto. Called 2003 VB12 (tentatively named Sedna) and 2004 DW, they assume the top spot held by Quaoar, found in 2002. The new objects add to a growing list of large bodies found at the fringes of the solar system; Sedna's extreme location in particular provides evidence for a hypothesized distant collection of icy bodies called the Oort cloud. Astronomers expect to find five to 10 more in the next couple of years, some perhaps even bigger than Pluto.

Diameter, in kilometers, of:
Pluto: 2,300
Pluto's moon Charon: 1,300
Quaoar: 1,250
2004 DW: Up to 1,600
Sedna: Up to 1,700

Distance to the sun, in billions of kilometers:
Pluto: 4.4 to 7.4
2004 DW: 4.6 to 7
Sedna: 13 to 135

Time to orbit the sun:
2004 DW: 248 years
Sedna: 10,500 years

SOURCE: California Institute of Technology

OPTICS

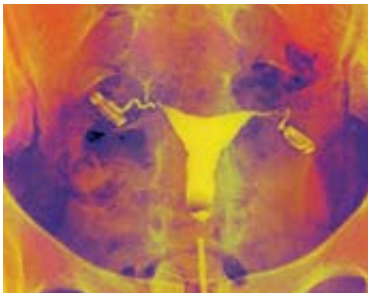
Attosecond Laser Pulses

An electron completes an orbit around a hydrogen atom in a mere 150 attoseconds—what the tick of a secondhand is to 200 million years. (One attosecond is 10^{-18} second.) Hoping to investigate such brief phenomena, physicists have made attosecond-scale laser flashes, typically by exciting electrons into ultimately releasing the flash. But precisely measuring the pulses has proved difficult; techniques have relied on indirect means or calculations based on how the pulse was made. A team led by Ferenc Krausz at the Vienna University of Technology has come up with a more accurate way. The group directed attosecond-scale x-ray flashes at neon atoms to strip the electrons off. Then a second light pulse sweeps the electrons sideways. Knocked clear, the electrons could have their energies measured. That enables researchers to determine the duration of the original pulse, which, in the data reported in the February 26 *Nature,* was 250 attoseconds long. —*Alexander Hellemans*

REPRODUCTION

More Eggs in One Basket

That women are born with all the eggs they will ever have may be a myth. Researchers have found that mice retain the ability to make egg-generating oocyte cells into adulthood. In juvenile female mice, follicles (oocytes encased in support cells) died rapidly enough that egg supplies should have been depleted in days or weeks. Still, mice can remain fertile past one year of age; moreover, follicle numbers overall remained virtually unchanged. This evidence suggests female mice have a previously undiscovered type of stem cell that continuously generates reproductive cells, just as males do. About 60 cells near each mouse ovary possessed chemicals typical of these stem cells. If these findings prove true in humans, theories about how a woman's reproductive system ages and how smoking, chemotherapy and radiation affect fertility will have to be reexamined. The report appears in the March 11 *Nature.* —*Charles Choi*



EGGED ON: The supply might not be finite.

ENVIRONMENT

Power Sludge

Roughly 33 billion gallons of wastewater are treated daily in the U.S. at an annual cost of more than $25 billion. A microbe-based device could offset the expense by generating electricity as it cleans sewage. The fuel cell, consisting in part of electrodes made of graphite and a carbon-plastic-platinum catalyst membrane, fills with wastewater. The germs in the sludge generate free electrons as their enzymes break down sugars, proteins and fats. In experiments, the invention produced 10 to 50 milliwatts of power per square meter of electrode surface (5 percent of the power needed to light one Christmas tree bulb). Meanwhile the fuel cell removed up to 78 percent of the water's organic muck. Environmental engineers at Pennsylvania State University say that their hand-size gadget could incorporate alternative materials to generate 10 to 20 times as much power. The findings appeared online in the February 21 *Environmental Science & Technology.* —*Charles Choi*

WASTEWATER could be a source of electricity.

SPL/PHOTO RESEARCHERS, INC. (*top*); LESTER LEFKOWITZ *Corbis* (*bottom*); ILLUSTRATION BY MATT COLLINS

TINY BLUEBERRY-SHAPED Mars rock (*inset*) found at "El Capitan" (*black-and-white area*) contained evidence of past water.

NASA/JPL/CORNELL (*background image*); NASA/JPL/U.S. GEOLOGICAL SURVEY (*inset*)

MARS

Foreshadowing Flashes in the Planum

Water may explain mysterious Martian flashes. For decades, astronomers have spied bursts of light on sites such as Meridiani Planum, where the rover Opportunity landed, even when the skies above the Red Planet were clear. Perhaps dunes or salt deposits left behind by ancient seas were reflecting sunlight. In 2002 NASA's Mars Odyssey orbiter found signs of ice lurking just below the planet's surface, including at Meridiani. This March, Opportunity sent back stunning evidence that water once drenched those very rocks. The rover's onboard spectrometers detected high concentrations of metal sulfate salts. Terrestrial rocks with that much sulfur either formed in water or soaked in it a long time. Rice-shaped indentations in the rock strongly resemble voids left by salt crystals grown in briny Earth water, and BB-size particles could have formed from minerals deposited in wet, porous rock. The flashes and the Odyssey results "support the Opportunity findings that there's something very interesting, and related to past Mars soaking, in this area," comments William Sheehan, an astronomer based in Willmar, Minn., who predicted and documented the most recent Martian flashes. *—JR Minkel*

BRIEF POINTS

- **Ten of 13 authors of a 1998 paper linking the childhood MMR vaccine to autism retracted their conclusions, in part because the selection of subjects may have been biased and because one author received undeclared funds from a group pursuing legal action on behalf of children allegedly damaged by the vaccine.**

  Lancet, March 6, 2004

- **Psychological stress appears to help trigger multiple sclerosis. Parents who lost a child were 50 percent more likely to develop the disease than those who did not; unexpected child deaths doubled the likelihood.**

  Neurology, March 9, 2004

- **Astronomers have detected the most distant galaxy yet, one with a redshift of 10, meaning that it is 13.2 billion light-years away. It may be among the first objects in the universe to generate their own light.**

  European Southern Observatory announcement, March 1, 2004

- **Populations of plants, insects and birds in the U.K. have dropped precipitously in the past 40 years—more evidence that the world is in the midst of its sixth mass extinction. The previous one wiped out the dinosaurs.**

  Science, March 19, 2004

PHYSICS

Nonstick Sliding

Friction arises when the atoms of a sliding surface "pluck" opposing atoms, producing vibrations that fritter energy away into heat. If the solids interact weakly enough, they should be able to rub without making vibrations—in other words, without friction. Ernst Meyer and his co-workers at the University of Basel have conclusively borne out this decades-old prediction by sliding a custom-made silicon tip over a crystal of salt. When the downward force on the tip is high, the atoms in the crystal get stretched like springs, and the tip repeatedly sticks and slips its way over the corrugated crystal surface, with each slip dissipating energy into heat. But when the force is low enough, the atomic bonds stay rigid, and the tip slides smoothly, producing essentially zero friction. The stick-and-slip results were scheduled to appear in an April issue of *Physical Review Letters*. *—JR Minkel*

NEAR-EARTH OBJECTS

Close Calls

For nine hours in January, a real-life *Deep Impact* looked possible. Thankfully, the first asteroid ever predicted to hit Earth within days (and with megaton force) turned out to be a false alarm. "I never said I was going to call the White House, as the 24/7 news media reported," says astronomer Clark R. Chapman of the Southwest Research Institute in Boulder, Colo. At a February conference on planetary defense, Chapman faced accusations that he overreacted to early data on Asteroid 2004 AS1, which passed Earth with distance to spare. Keeping watch for collisions today is the Spaceguard Survey, an international network of observatories, but it only looks for objects bigger than a kilometer across. Smaller threats, such as the 500-meter-wide 2004 AS1, can go undetected. To track them, Representative Dana Rohrabacher of California has proposed legislation to boost planetary defense funding from $3.5 million to $20 million annually. *—Ian Steer*

Making Drugs, Not Profits

A married couple attacks neglected diseases of the developing world By GARY STIX

When Victoria Hale left her job as a pharmacologist at Genentech in 1998, she made a list. It detailed areas that she felt the pharmaceutical industry had ignored: orphan drugs for metabolic disorders, treatments for substance abuse, modernization of contraceptives, and global infectious disease. Hale, an ebullient woman who also had more than five years of experience as a drug evaluator at the Food and Drug Administration, looked over what she had written and decided that, of the various choices, fighting infectious disease would have the most pronounced impact on public health.

To achieve her goal, however, would require setting up a venture that would differ radically from the traditional business models embraced by the pharmaceutical and biotechnology industries. To make drugs affordable in places where annual family incomes were often less than the cost of an MP3 player, the first thing that would have to be jettisoned was the profit motive.

Lacking business experience, Hale approached colleagues she had known at the FDA and Genentech, pestering each one to become chief executive to fulfill her notion of what a nonprofit drug company should be. They all told her that unless she took on leadership of the new entity herself, it would never come to fruition. Eventually that inevitability sunk in—and she began to assume the mantle of chief executive for a still emerging concept. In her quest, Hale made her way to the World Health Organization in Geneva. Never having been a member of the global health community, Hale says that she encountered a somewhat perplexed reaction at first. Was this just a naive visionary from California going through a midcareer crisis without any clear idea of what she was getting into?

DYNAMIC DUO: Ahvie Herskowitz and Victoria Hale, a husband-and-wife team, took the novel step of starting a nonprofit drug company.

Even on her first trip to Switzerland in 1999, however, she gained answers to some basic questions. The most important one had to do with identifying the mission for a nonprofit drug company with virtually no resources except human capital. Philippe M. P. Desjeux of WHO suggested that an opportunity existed for an off-patent antibiotic that needed one last clinical trial to prove its worth as a drug against a deadly parasite.

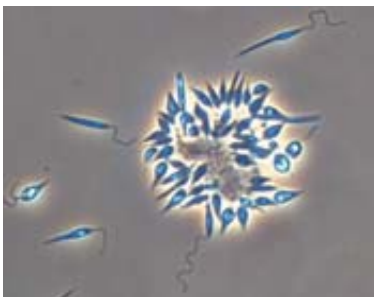
Leishmaniasis—also called kala-azar ("black fever" in Hindi)—is a parasitic disease that is transmitted by sand flies. Left untreated, visceral leishmaniasis, the internal form of the disease, results in almost certain death. (There is also a disfiguring cutaneous form.) Every year 500,000 new visceral cases emerge around the world, and 200,000 or more deaths are not unusual. Most visceral leishmaniasis cases are concentrated in poor populations in just a few countries: India, Bangladesh, Nepal, Sudan and Brazil. Despite the disease's grim epidemiology, its incidence is markedly less than a global monster like malaria and is more manageable for distributing a newly approved pharmaceutical.

All that was necessary for the antibiotic's approval in India was a late-stage (Phase III) clinical trial. Desjeux gave Hale a list of scientists and clinicians in India and elsewhere who were among the world's leading ex-

ONEWORLD HEALTH

perts on the disease. Two weeks later Hale called Desjeux back. She had already visited India, talked with many of the experts on Desjeux's list and was raring to move ahead, ready to join the ranks of the closely knit community of health workers called leishmaniacs.

India was the perfect place to begin a trial. The parasite, a protozoan called *Leishmania donovani,* had become resistant to the drugs of choice, compounds based on the element antimony. One other treatment, amphotericin, at about $100 for a course of therapy, matches the annual income of many of the households where kala-azar claims its victims. A family might have to sell a cow or another prized possession to come up with the money. Amphotericin is also toxic and requires a hospital stay. If WHO's antibiotic, paromomycin, could be deployed, it might be used on an outpatient basis at a cost of $1 a day, eliminating the parasite in three weeks. "It's incredible what impact this could have," Hale declares. "It could change the world, the whole fate of a community."



LEISHMANIA DONOVANI: These single-celled organisms cause visceral leishmaniasis.

Hale's husband, Ahvie Herskowitz, a physician, had also decided to leave his job. He had been running large clinical trials for the Ischemia Research and Education Foundation. Both he and Hale started their own drug-development consultancy. The work gave Hale enough time to travel and explore her idea. Herskowitz often became locked in discussions with his wife about whether a profitless drug company would really be practical. He, too, began to devote more time to the venture. As a child of Holocaust survivors, Herskowitz was driven by some of the same impulses as Hale: "I am lucky to be alive—I was successful professionally and felt I needed to help those less fortunate."

In 2000 the pair launched the Institute for OneWorld Health, with Hale as chief executive and Herskowitz as chief medical officer. The first obstacle was bureaucratic: getting Internal Revenue Service approval for a nonprofit pharmaceutical company, a designation that, at first glance, seems like an oxymoron—and one that the agency had difficulty grasping. But the timing of Hale's vision for a new type of drug company was impeccable. Just about when the two were getting started, the Bill & Melinda Gates Foundation was coming into its own. When Hale approached the foundation, the officers told her that they were already supplying money for leishmaniasis through support for vaccine research for the disease. Hale emphasized that vaccines for parasites had a pitiful track record—malaria being a notable case in point. In 2002 the Gates Foundation agreed to provide $4.7 million, most of it for a Phase III leishmaniasis trial. "They're doing great stuff," Bill Gates says. "Just take that one thing, kala-azar. Hey, that's going to be a medicine [paromomycin] that is going to save a lot of lives." Late last year the foundation decided to supply another $5.3 million.

Last May, OneWorld and WHO started a clinical trial of paromomycin that has since enrolled 670 patients in Bihar state—the largest for an antiparasite drug ever conducted in India, according to Herskowitz. OneWorld will also use the study to seek approval in the U.S. or a European country, thereby meeting a set of international guidelines that will enable rapid approval wherever the disease is endemic.

If Indian regulators give the nod next year, OneWorld will pay up-front costs for manufacturing the first batches in India—and then future revenues will go to the drugmakers there. The biggest challenge will be to build a distribution system to ensure that the drug gets supplied to those who need it. "Pharmaceuticals haven't penetrated into the depths of these communities as much as Coca-Cola," Herskowitz remarks. In the past, India has had in place an emergency system that was mobilized when the disease reached epidemic proportions—and a collaboration of OneWorld, WHO and the Indian government will try to construct its supply network on this model.

Word of OneWorld's work has spread, and the company receives frequent calls from scientists and executives at other pharmaceutical firms who wonder how they can play a part in the nonprofit's mission. Celera Genomics licensed to OneWorld royalty-free a drug for Chagas disease that it inherited when the company acquired a smaller biotech firm. And Yale University and the University of Washington licensed on the same terms another compound for the parasitic disease, which afflicts 16 million to 18 million people in Mexico and Central and South America and causes 50,000 fatalities every year. The Chagas treatments, with some of the development work funded by the Gates money, will test the company's ability to take a drug all the way through the clinical trial process. And OneWorld has the makings of a pipeline—it has early-stage development programs for drugs to treat malaria and diarrhea.

At a juncture when the global pharmaceutical industry is under siege for the prices it charges, Hale and the 25 employees of OneWorld have demonstrated that the spirit of the entrepreneur can be directed toward supplying something besides simple knockoffs of cholesterol and depression medication. SA

SINCLAIR STAMMERS *Photo Researchers, Inc.*

The Enchanted Glass

Francis Bacon and experimental psychologists show why the facts in science never just speak for themselves By MICHAEL SHERMER

In the first trimester of the gestation of science, one of science's midwives, Francis Bacon, penned an immodest work entitled *Novum Organum* ("new tool," after Aristotle's *Organon*) that would open the gates to the "Great Instauration" he hoped to inaugurate through the scientific method. Rejecting both the unempirical tradition of scholasticism and the Renaissance quest to recover and preserve ancient wisdom, Bacon sought a blend of sensory data and reasoned theory.

Cognitive barriers that color clear judgment presented a major impediment to Bacon's goal. He identified four: idols of the cave (individual peculiarities), idols of the marketplace (limits of language), idols of the theater (preexisting beliefs) and idols of the tribe (inherited foibles of human thought).

We have a cognitive bias to see ourselves in a more positive light than others see us.

Experimental psychologists have recently corroborated Bacon's idols, particularly those of the tribe, in the form of numerous cognitive biases. The self-serving bias, for example, dictates that we tend to see ourselves in a more positive light than others see us: national surveys show that most businesspeople believe that they are more moral than other businesspeople, and psychologists who study moral intuition think they are more moral than other such psychologists. In one College Entrance Examination Board survey of 829,000 high school seniors, less than 1 percent rated themselves below average in "ability to get along with others," and 60 percent put themselves in the top 10 percent. And according to a 1997 *U.S. News and World Report* study on who Americans believe are most likely to go to heaven, 52 percent said Bill Clinton, 60 percent thought Princess Diana, 65 percent chose Michael Jordan and 79 percent selected Mother Teresa. Fully 87 percent decided that the person most likely to see paradise was the survey taker!

Princeton University psychology professor Emily Pronin and her colleagues tested an idol called bias blind spot, in which subjects recognized the existence and influence of eight different cognitive biases in other people but failed to see those same biases in themselves. In one study on Stanford University students, when asked to compare themselves with their peers on such personal qualities as friendliness and selfishness, they predictably rated themselves higher. Even when the subjects were warned about the "better than average" bias and asked to reconsider their original assessments, 63 percent claimed that their initial evaluations were objective, and 13 percent even claimed to be too modest.

In a second study, Pronin randomly assigned subjects high or low scores on a "social intelligence" test. Unsurprisingly, those who were given high marks rated the test as being fairer and more useful than those receiving low marks. When the subjects were then asked if it was possible that they had been influenced by the score on the test, they responded that other participants had been far more biased than they were. In a third study, in which Pronin queried subjects about what method they used to assess their own biases and those of others, she found that people tend to use general theories of behavior when evaluating others but use introspection when appraising themselves. In what is called the introspection illusion, people do not believe that others can be trusted to do the same: okay for me but not for thee.

Psychologist Frank J. Sulloway of the University of California at Berkeley and I made a similar discovery of an attribution bias in a study we conducted on why people say they believe in God and why they think other people do so. In general, most individuals attribute their own faith to such intellectual reasons as the good design and complexity of the world, whereas they attribute others' belief in God to such emotional reasons as that it is comforting, that it gives meaning and that it is how they were raised.

None of these findings would surprise Francis Bacon, who, four centuries ago, noted: "For the mind of man is far from the nature of a clear and equal glass, wherein the beams of things should reflect according to their true incidence; nay, it is rather like an enchanted glass, full of superstition and imposture, if it be not delivered and reduced." SA

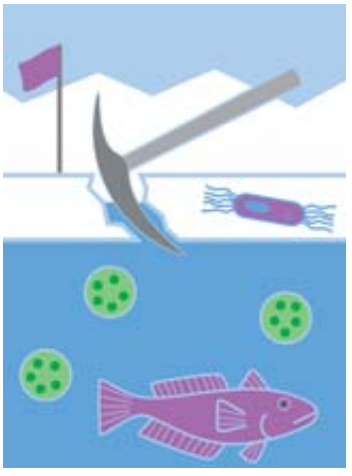
Michael Shermer is publisher of Skeptic *(www.skeptic.com) and author of* The Science of Good and Evil.

BRAD HINES

Patents on Ice

Antarctica as a last frontier for bioprospectors—and their intellectual property By GARY STIX

A patent and trademark office has yet to open its doors on McMurdo Sound or at Prydz Bay. But the microbes and fish that live in Antarctica and its environs have already become the subject of patent claims. The Spanish patent office granted a patent in 2002 for wound healing and other treatments with a glycoprotein drawn from the bacterium *Pseudoalteromonas antarctica*. Also that year Germany handed out a patent for a skin treatment using an extract from the green alga *Prasiola crispa ssp. antarctica*. And an application now before the U.S. Patent and Trademark Office covers a process for producing antifreeze peptides discovered in Antarctic bacteria.



In all, it is estimated that more than 40 patents have been granted worldwide that rely on Antarctic flora and fauna, and the U.S. patent office has received in excess of 90 filings. These numbers are not large—and no commercial enterprise is engaged in industrial harvesting of the continent's biota. But drug companies bring in tens of billions of dollars every year from natural compounds or synthetic knockoffs inspired by them. Interest in developing pharmaceuticals from Antarctica's novel life-forms, extremophiles—which withstand cold, aridity and salinity—will continue to grow. A case in point is AMRAD Natural Products, an Australian pharmaceutical company that struck a deal with the Antarctic Cooperative Research Center at the University of Tasmania in 1995 to screen about 1,000 microbial samples a year for antibiotics and a range of other pharmaceuticals.

One or two blockbuster drugs derived from Antarctic bacteria could spur a veritable stampede. A United Nations study released in February cautioned that the push to exploit extremophiles requires new rules to protect the continent's fragile ecosystem. Regulation of these activities presents special challenges. The Antarctic Treaty System pledges to protect the continent's environment but does not address bioprospecting directly, which could encourage more of these endeavors. Moreover, existing international policies on bioprospecting are of limited use. For instance, although the Convention on Biological Diversity has established a framework for allowing access to biological resources, it assumes that individual states in fact have sovereignty over these resources, a presumption that does not hold for Antarctica.

Moreover, it is already a problem to figure out who is doing the collecting and for what purpose. Bioprospecting often involves consortia composed of public and private entities. Delineating where scientific research ends and commercial activity begins becomes a difficult task, notes a report from the U.N. University Institute of Advanced Studies entitled "The International Regime for Bioprospecting: Existing Policies and Emerging Issues for Antarctica." The document was drafted in preparation for a biodiversity meeting, the Seventh Conference of Parties to the Convention on Biological Diversity, held in Kuala Lumpur, Malaysia, this past February.

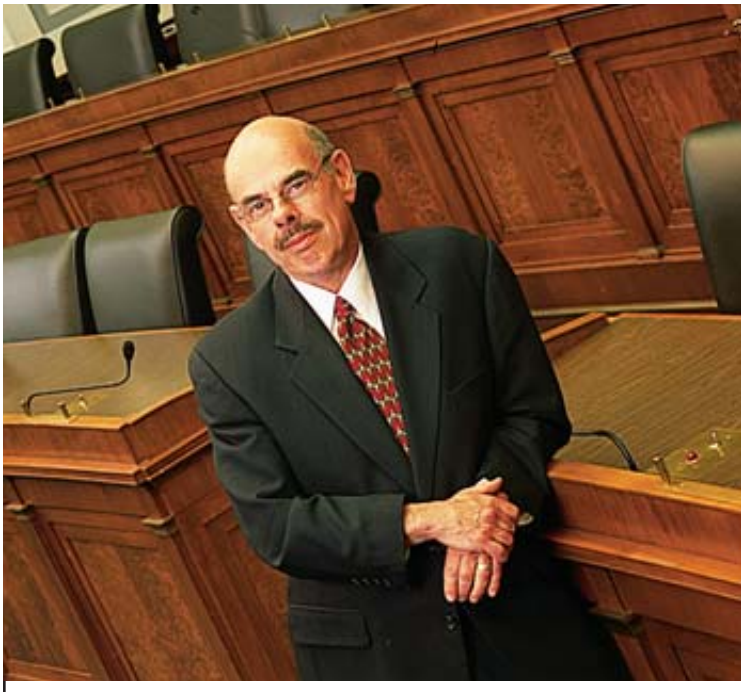
The report calls for the development of regulations to govern bioprospecting that would address a series of basic questions: Who owns the continent's genetic resources? How can scientists legitimately acquire biomaterials? What measures should researchers take to protect extremophiles? Who owns the products that eventually get marketed commercially from these discoveries? And would bioprospecting violate a provision of the Antarctic Treaty System requiring that scientific results be shared freely? Determining the answers now might help waylay the legal entanglements that will inevitably occur if bioprospecting thrives and a swarm of extremophile collectors descend on Prydz Bay and other entry points to the frozen continent. SA

JENNIFER KANE

Science's Political Bulldog

Representative Henry A. Waxman blasts away at the White House for alleged abuse of science. Sure, it's politics—but it could restore confidence in the scientific process By JULIE WAKEFIELD

To hear Henry A. Waxman bemoan how predetermined beliefs are jeopardizing scientific freedom, you might think you are in another age or in some struggling new country. But there, outside his corner office, is the gleaming dome of the Capitol, its perimeter tightened with bollards and the latest surveillance. "Science is very much under attack with the Bush administration," Waxman declares from his suite in the Rayburn Office Building. "If the science doesn't fit what the White House wants it to be, it distorts the science to fit into what its preconceived notions are about what it wants to do."



KAY CHERNUSH

HENRY A. WAXMAN: KEEPING HOUSE

- Entered Congress in 1974 with other reform-minded Democrats who swept into office in the midterm elections after Watergate.
- Holds degrees in political science and in law from the University of California at Los Angeles.
- On his career: "My parents would have preferred that I be a doctor rather than a lawyer and then later a congressman. But that wasn't my strength."

As the ranking minority member on the House Government Reform Committee, the 64-year-old California Democrat has become a leading voice railing against the White House's science policy—or lack thereof. The charges are not new—word of such politicization began percolating almost as soon as George W. Bush took office, and until recently, many scientists who complained in private held their tongues in public. Waxman has given scientists' fears a voice, and a growing crowd of scientific organizations, advocacy groups and former officials are adding to the chorus.

Waxman launched his first formal salvo last August. Pulling together reports and editorials from various sources (including *Scientific American*), his office issued a report detailing political interference in more than 20 areas affecting health, environmental and other research agencies. Examples include deleting information from Web sites, stacking advisory committees with candidates with uncertain qualifications and questionable industry ties, and suppressing information and projects inconvenient to White House policy goals, such as those having to do with global warming. And he charges that the beneficiaries of these distortions are for the most part Bush's political supporters, including the Traditional Values Coalition, a church-based policy group in Washington, D.C., and oil lobbyists.

To Waxman, who became interested in health issues in 1969 when he was appointed to the California State Assembly Health Committee, the assaults on the National Institutes of Health are especially offensive. For example, after prompting by Republican members of Congress, NIH officials started contacting a "hit list" of 150 investigators compiled by the Traditional Values Coalition. The organization charged that the NIH was funding smarmy sex studies and denounced the projects that look at such behaviors as truck-stop prostitution and the sexual habits of seniors.

Although no grants were rescinded, many viewed the calls as an attempt to stifle the scientific process, considering that all 200 of the grants in question had already undergone peer review. At the University of California at San Francisco, where about 17 investigators were contacted, the message was clear: "Look out: Big Brother is watching," recounts Keith R. Yamamoto, executive vice dean at the medical school.

"I just think we need to make sure the jewel of U.S. government policy—the NIH, which I think is a national treasure—not be hurt in any way by those who would try to inject politics into scientific research," Waxman states. NIH officials declined to comment for this story. But in a previous interview, NIH director Elias A. Zerhouni stated that he has not seen many solid cases of political interference and invited researchers who encountered such pressure to come forward [see "A Biomedical Politician," by Carol Ezzell, Insights, September 2003].

Beyond grants, scientific publishing also seems to be under fire. The Office of Foreign Assets Control, part of the U.S. Treasury, has pressured professional organizations—such as the American Society for Microbiology and the Institute of Electrical and Electronics Engineers—to virtually ban papers originating in Iran, Cuba, Sudan and Libya. The rationale: the ban is part of the U.S. trade embargo policy with these countries. Publishing their papers requires special licenses.

"If the science doesn't fit what the White House wants it to be, it distorts the science." —*Henry Waxman*

Perhaps more contentious is the Office of Management and Budget's proposal to centrally peer-review the science behind new federal regulations. The plan, which could be implemented by the summer, is a way to "enhance the competence and credibility of science used by regulators," according to John D. Graham, an OMB administrator. For example, "the lack of adequate peer review contributed to childhood deaths due to passenger air bag deployment," Graham says—specifically, federal agencies failed to consider risk assessments performed by automakers indicating that kids seated in cars with passenger air bags need to be restrained properly in the back seat.

Critics such as Waxman see it differently. They call the proposal an insidious way to use scientific uncertainty to stall regulations that are likely to be costly to industry by adding layers of review—and by including potentially biased ones. "It's very heavy-handed of the OMB to come in and regulate peer review," Waxman charges. Moreover, he adds, the OMB's notion of the process has fallen short in the recent past. In the debate over the environment, the Bush administration has quashed findings that run counter to policy decisions. And its actions extend beyond its rejection of the Kyoto protocol. For example, the White House suppressed for several months a 2003 Environmental Protection Agency report detailing that a Senate Clean Air bill would prevent substantially more deaths from mercury contamination than the administration's proposed Clear Skies Act.

The Union of Concerned Scientists outlined these and other allegations in a report issued this February. Along with the report, 62 prominent scientists—including Nobel laureates and National Medal of Science winners—signed a statement calling for the restoration of scientific integrity to federal policymaking.

"The peer-review situation at the OMB is frightening on many levels," says Neal Lane, a signatory of the statement who headed the National Science Foundation and served as presidential science adviser under Bill Clinton. "The integrity of information is going to be seriously undermined in a process that requires political approval." He points out that whereas the heads of the NIH and other far-flung agencies are all political appointees, the OMB is part of the White House.

Although science has historically been political to some degree, "it's unprecedented what we're now seeing," Waxman contends. "We've had people from the Nixon administration, Republicans who served in the EPA"—Russell E. Train and William D. Ruckelshaus—"decry what's being done."

Some scholars remain skeptical about whether science has become more political. "When people are seeking political advantage, there isn't much that is sacred," observes economist Lester Lave of Carnegie Mellon University. "Since scientists enjoy a positive reputation with the public, members of Congress and other decision makers, there is some attempt to line up Nobel Prize winners, professional society presidents or large numbers of university people to support or oppose a position. There is nothing new here." And even Lane notes a considerable amount of "polemic" mixed with the concrete cases of interference outlined in Waxman's August report.

Bush administration officials have countered that Waxman himself is using scientists' concerns for his own political gain. "He's just playing politics by continuing to attack the president's policies. He's not offering constructive ways to enhance science policy," says Mary Ellen Grant, a spokesperson for the Republican National Committee.

Waxman is undeterred. As he did in many of his past reform campaigns, he established a "tipline" for scientists to register additional examples of politicization. But he has not been able to round up support for congressional hearings as he did against the tobacco industry in 1994. The Republican congressional majority's lack of interest in the issue has frustrated him. Still, he hopes to effect change: "It should be enough to bring it under public scrutiny, because [the administration] can't defend those kinds of actions." SA

Julie Wakefield, based in Washington, D.C., is writing a book on the adventures of Edmond Halley.

the myth of TIME

The Beginning of

By Gabriele Veneziano

String theory suggests that the
BIG BANG was not the origin of the universe
but simply the outcome of a preexisting state

Was the big bang really the beginning of time?

Or did the universe exist before then? Such a question seemed almost blasphemous only a decade ago. Most cosmologists insisted that it simply made no sense—that to contemplate a time before the big bang was like asking for directions to a place north of the North Pole. But developments in theoretical physics, especially the rise of string theory, have changed their perspective. The pre-bang universe has become the latest frontier of cosmology.

The new willingness to consider what might have happened before the bang is the latest swing of an intellectual pendulum that has rocked back and forth for millennia. In one form or another, the issue of the ultimate beginning has engaged philosophers and theologians in nearly every culture. It is entwined with a grand set of concerns, one famously encapsulated in an 1897 painting by Paul Gauguin: *D'ou venons-nous? Que sommes-nous? Ou allons-nous?* "Where do we come from? What are we? Where are we going?" The piece depicts the cycle of birth, life and death—origin, identity and destiny for each individual—and these personal concerns connect directly to cosmic ones. We can trace our lineage back through the generations, back through our animal ancestors, to early forms of life and protolife, to the elements synthesized in the primordial universe, to the amorphous energy deposited in space before that. Does our family tree extend forever backward? Or do its roots terminate? Is the cosmos as impermanent as we are?

The ancient Greeks debated the origin of time fiercely. Aristotle, taking the no-beginning side, invoked the principle that out of nothing, nothing comes. If the universe could never have gone from nothingness to somethingness, it must always have existed. For this and other reasons, time must stretch eternally into the past and future. Christian theologians tended to take the opposite point of view. Augustine contended that God exists outside of space and time, able to bring these constructs into existence as surely as he could forge other aspects of our world. When asked, "What was God doing *before* he created the world?" Augustine answered, "Time itself being part of God's creation, there was simply no *before!*"

Einstein's general theory of relativity led modern cosmologists to much the same conclusion. The theory holds that space and time are soft, malleable entities. On the largest scales, space is naturally dynamic, expanding or contracting over time, carrying matter like driftwood on the tide. Astronomers confirmed in the 1920s that our universe is currently expanding: distant galaxies move apart from one another. One consequence, as physicists Stephen Hawking and Roger Penrose proved in the 1960s, is that time cannot extend back indefinitely. As you play cosmic history backward in time, the galaxies all come together to a single infinitesimal point, known as a singularity—almost as if they were descending into a black hole. Each galaxy or its precursor is squeezed down to zero size. Quantities such as density, temperature and spacetime curvature become infinite. The singularity is the ultimate cataclysm, beyond which our cosmic ancestry cannot extend.

Strange Coincidence

THE UNAVOIDABLE singularity poses serious problems for cosmologists. In particular, it sits uneasily with the high degree of homogeneity and isotropy that the universe exhibits on large scales. For the cosmos to look broadly the same everywhere, some kind of communication had to pass among distant regions of space, coordinating their properties. But the idea of such communication contradicts the old cosmological paradigm.

To be specific, consider what has happened over the 13.7 billion years since the release of the cosmic microwave background radiation. The distance between galaxies has grown by a factor of about 1,000 (because of the expansion), while the radius of the observable universe has grown by the much larger factor of about 100,000 (because light outpaces the expansion). We see parts of the universe today that we could not have seen 13.7 billion years ago. Indeed, this is the first time in cosmic history that light from the most distant galaxies has reached the Milky Way.

Nevertheless, the properties of the Milky Way are basically the same as those of distant galaxies. It is as though you showed up at a party only to find you were wearing exactly the same clothes as a dozen of your closest friends. If just two of you were dressed the same, it might be explained away as coincidence, but a dozen suggests that the partygoers had coordinated their attire in advance. In cosmology, the number is not a dozen but tens of thousands—the number of independent yet statistically identical patches of sky in the microwave background.

One possibility is that all those regions of space were endowed at birth with identical properties—in other words, that the homogeneity is mere coincidence. Physicists, however, have thought about two more natural ways out of the impasse: the early universe was much smaller or much older than in standard cosmology. Either (or both, acting together) would have made intercommunication possible.

The most popular choice follows the first alternative. It postulates that the universe went through a period of accelerating expansion, known as inflation, early in its history. Before this phase, galaxies or their precursors were so closely packed that they could easily coordinate their properties. During inflation, they fell out of contact because light was unable to keep pace with the frenetic expansion. After inflation ended, the expansion began to decelerate, so galaxies gradually came back into one another's view.

Physicists ascribe the inflationary spurt to the potential energy stored in a new quantum field, the inflaton, about 10^{-35} second after the big bang. Potential energy, as opposed to rest mass or kinetic energy, leads to gravitational repulsion. Rather than slowing down the expansion, as the gravitation of ordinary matter would, the inflaton accelerated it. Proposed in 1981, inflation has explained a wide variety of observations with precision [see "The Inflationary Universe," by Alan H. Guth and Paul J. Steinhardt; SCIENTIFIC AMERICAN, May 1984; and "Four Keys to Cosmology," Special report; SCIENTIFIC AMERICAN, February]. A number of possible theoretical problems remain, though, beginning with the questions of what exactly the inflaton was and what gave it such a huge initial potential energy.

A second, less widely known way to solve the puzzle follows the second alternative by getting rid of the singularity. If time did not begin at the bang, if a long era preceded the onset of the present cosmic expansion, matter could have had plenty of time to arrange itself smoothly. Therefore, researchers have reexamined the reasoning that led them to infer a singularity.

One of the assumptions—that relativity theory is always valid—is question-

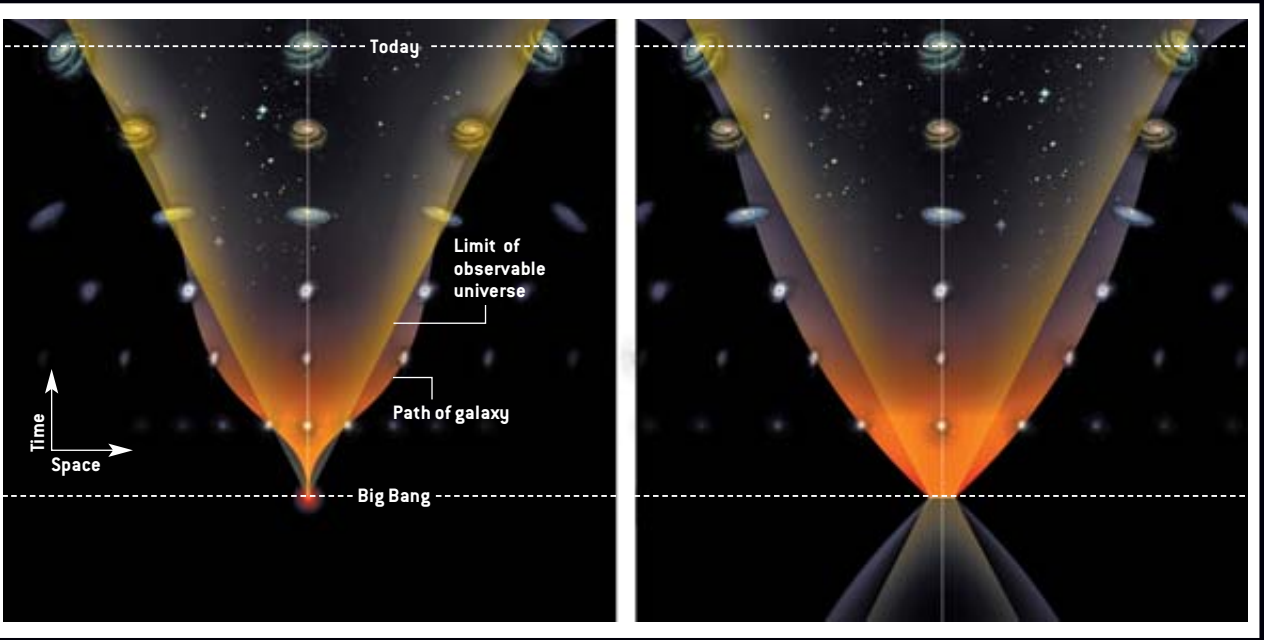
Overview/*String Cosmology*

- Philosophers, theologians and scientists have long debated whether time is eternal or finite—that is, whether the universe has always existed or whether it had a definite genesis. Einstein's general theory of relativity implies finiteness. An expanding universe must have begun at the big bang.
- Yet general relativity ceases to be valid in the vicinity of the bang because quantum mechanics comes into play. Today's leading candidate for a full quantum theory of gravity—string theory—introduces a minimal quantum of length as a new fundamental constant of nature, making the very concept of a bangian genesis untenable.
- The bang still took place, but it did not involve a moment of infinite density, and the universe may have predated it. The symmetries of string theory suggest that time did not have a beginning and will not have an end. The universe could have begun almost empty and built up to the bang, or it might even have gone through a cycle of death and rebirth. In either case, the pre-bang epoch would have shaped the present-day cosmos.

ALFRED T. KAMAJIAN *(preceding pages)*

Two Views of the Beginning

In our expanding universe, galaxies rush away from one another like a dispersing mob. Any two galaxies recede at a speed proportional to the distance between them: a pair 500 million light-years apart separates twice as fast as one 250 million light-years apart. Therefore, all the galaxies we see must have started from the same place at the same time—the big bang. The conclusion holds even though cosmic expansion has gone through periods of acceleration and deceleration; in spacetime diagrams (*below*), galaxies follow sinuous paths that take them in and out of the observable region of space (*yellow wedge*). The situation became uncertain, however, at the precise moment when the galaxies (or their ancestors) began their outward motion.



In standard big bang cosmology, which is based on Einstein's general theory of relativity, the distance between any two galaxies was zero a finite time ago. Before that moment, time loses meaning.

In more sophisticated models, which include quantum effects, any pair of galaxies must have started off a certain minimum distance apart. These models open up the possibility of a pre-bang universe.

able. Close to the putative singularity, quantum effects must have been important, even dominant. Standard relativity takes no account of such effects, so accepting the inevitability of the singularity amounts to trusting the theory beyond reason. To know what really happened, physicists need to subsume relativity in a quantum theory of gravity. The task has occupied theorists from Einstein onward, but progress was almost zero until the mid-1980s.

Evolution of a Revolution

TODAY TWO APPROACHES stand out. One, going by the name of loop quantum gravity, retains Einstein's theory essentially intact but changes the procedure for implementing it in quantum mechanics [see "Atoms of Space and Time," by Lee Smolin; SCIENTIFIC AMERICAN, January]. Practitioners of loop quantum gravity have taken great strides and achieved deep insights over the past several years. Still, their approach may not be revolutionary enough to resolve the fundamental problems of quantizing gravity. A similar problem faced particle theorists after Enrico Fermi introduced his effective theory of the weak nuclear force in 1934. All efforts to construct a quantum version of Fermi's theory failed miserably. What was needed was not a new technique but the deep modifications brought by the electroweak theory of Sheldon L. Glashow, Steven Weinberg and Abdus Salam in the late 1960s.

The second approach, which I consider more promising, is string theory—a truly revolutionary modification of Einstein's theory. This article will focus on it, although proponents of loop quantum gravity claim to reach many of the same conclusions.

String theory grew out of a model that I wrote down in 1968 to describe the world of nuclear particles (such as protons and neutrons) and their interactions. Despite much initial excitement, the model failed. It was abandoned several

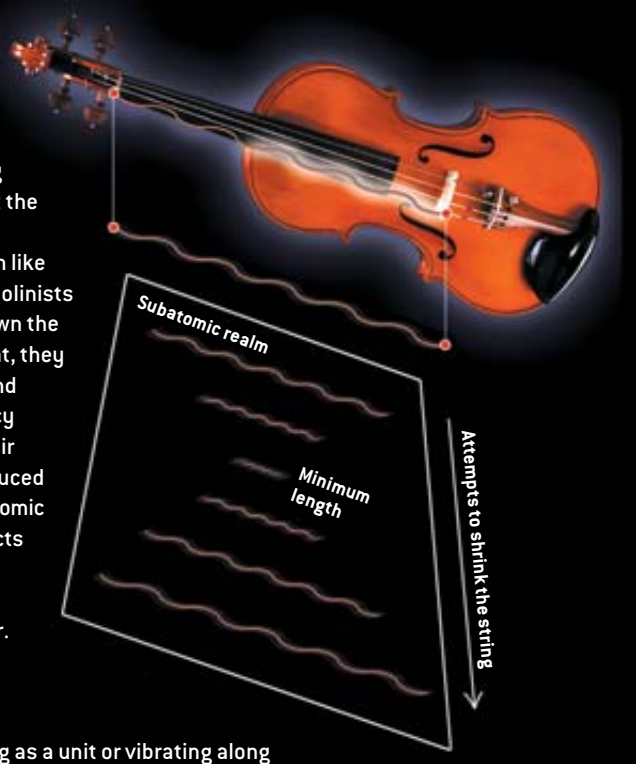
THE AUTHOR

GABRIELE VENEZIANO, a theoretical physicist at CERN, was the father of string theory in the late 1960s—an accomplishment for which he received this year's Heineman Prize of the American Physical Society and the American Institute of Physics. At the time, the theory was regarded as a failure; it did not achieve its goal of explaining the atomic nucleus, and Veneziano soon shifted his attention to quantum chromodynamics, to which he made major contributions. After string theory made its comeback as a theory of gravity in the 1980s, Veneziano became one of the first physicists to apply it to black holes and cosmology.

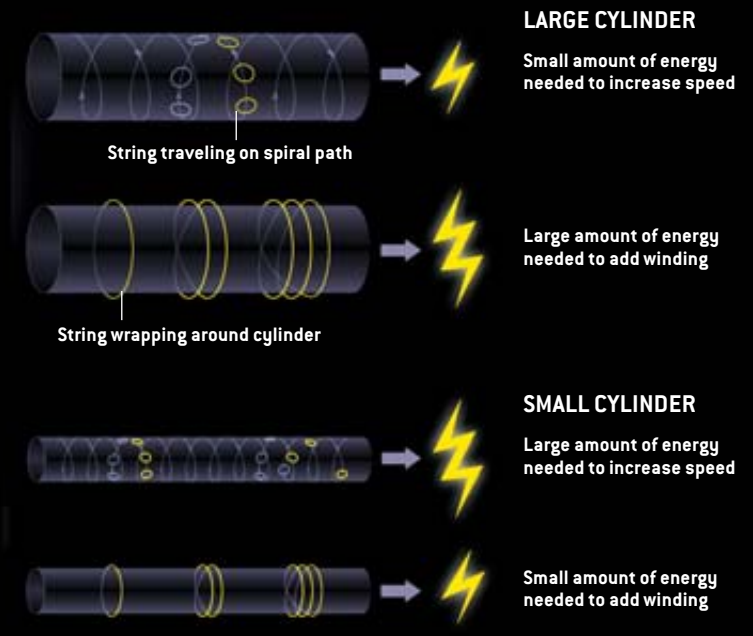
SAMUEL VELASCO

String Theory 101

String theory is the leading (though not only) theory that tries to describe what happened at the moment of the big bang. The strings that the theory describes are material objects much like those on a violin. As violinists move their fingers down the neck of the instrument, they shorten the strings and increase the frequency (hence energy) of their vibrations. If they reduced a string to a sub-subatomic length, quantum effects would take over and prevent it from being shortened any further.



In addition to traveling as a unit or vibrating along its length, a subatomic string can wind up like a spring. Suppose that space has a cylindrical shape. If the circumference is larger than the minimum allowed string length, each increase in the travel speed requires a small increment of energy, whereas each extra winding requires a large one. But if the circumference is smaller than the minimum length, an extra winding is less costly than an extra bit of velocity. The net energy—which is all that really matters—is the same for both small and large circumferences. In effect, the string does not shrink. This property prevents matter from reaching an infinite density.



years later in favor of quantum chromodynamics, which describes nuclear particles in terms of more elementary constituents, quarks. Quarks are confined inside a proton or a neutron, as if they were tied together by elastic strings. In retrospect, the original string theory had captured those stringy aspects of the nuclear world. Only later was it revived as a candidate for combining general relativity and quantum theory.

The basic idea is that elementary particles are not pointlike but rather infinitely thin one-dimensional objects, the strings. The large zoo of elementary particles, each with its own characteristic properties, reflects the many possible vibration patterns of a string. How can such a simple-minded theory describe the complicated world of particles and their interactions? The answer can be found in what we may call quantum string magic. Once the rules of quantum mechanics are applied to a vibrating string—just like a miniature violin string, except that the vibrations propagate along it at the speed of light—new properties appear. All have profound implications for particle physics and cosmology.

First, quantum strings have a finite size. Were it not for quantum effects, a violin string could be cut in half, cut in half again and so on all the way down, finally becoming a massless pointlike particle. But the Heisenberg uncertainty principle eventually intrudes and prevents the lightest strings from being sliced smaller than about 10^{-34} meter. This irreducible quantum of length, denoted l_s, is a new constant of nature introduced by string theory side by side with the speed of light, c, and Planck's constant, h. It plays a crucial role in almost every aspect of string theory, putting a finite limit on quantities that otherwise could become either zero or infinite.

Second, quantum strings may have angular momentum even if they lack mass. In classical physics, angular momentum is a property of an object that rotates with respect to an axis. The formula for angular momentum multiplies together velocity, mass and distance from the axis; hence, a massless object can have no angular momentum. But quan-

SAMUEL VELASCO

tum fluctuations change the situation. A tiny string can acquire up to two units of h of angular momentum without gaining any mass. This feature is very welcome because it precisely matches the properties of the carriers of all known fundamental forces, such as the photon (for electromagnetism) and the graviton (for gravity). Historically, angular momentum is what clued in physicists to the quantum-gravitational implications of string theory.

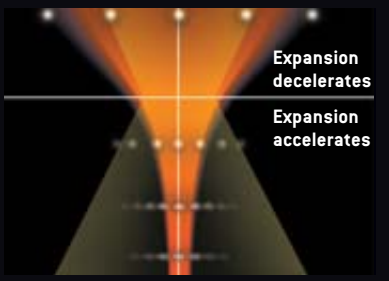
Third, quantum strings demand the existence of extra dimensions of space, in addition to the usual three. Whereas a classical violin string will vibrate no matter what the properties of space and time are, a quantum string is more finicky. The equations describing the vibration become inconsistent unless spacetime either is highly curved (in contradiction with observations) or contains six extra spatial dimensions.

Fourth, physical constants—such as Newton's and Coulomb's constants, which appear in the equations of physics and determine the properties of nature—no longer have arbitrary, fixed values. They occur in string theory as fields, rather like the electromagnetic field, that can adjust their values dynamically. These fields may have taken different values in different cosmological epochs or in remote regions of space, and even today the physical "constants" may vary by a small amount. Observing any variation would provide an enormous boost to string theory. [*Editors' note: An upcoming article will discuss searches for these variations.*]

One such field, called the dilaton, is the master key to string theory; it determines the overall strength of all interactions. The dilaton fascinates string theorists because its value can be reinterpreted as the size of an extra dimension of space, giving a grand total of 11 spacetime dimensions.

PRE–BIG BANG SCENARIO

A pioneering effort to apply string theory to cosmology was the so-called pre–big bang scenario, according to which the bang is not the ultimate origin of the universe but a transition. Beforehand, expansion accelerated; afterward, it decelerated (at least initially). The path of a galaxy through spacetime (*right*) is shaped like a wineglass.

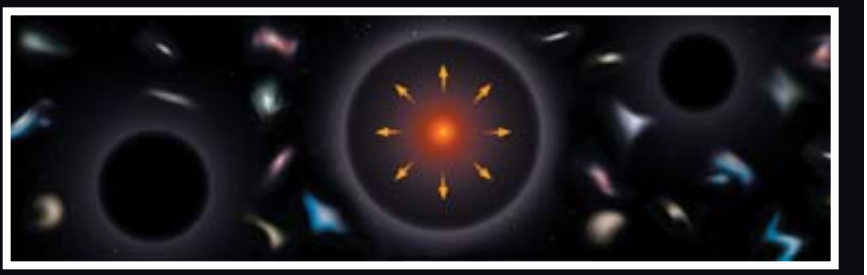


The universe has existed forever. In the distant past, it was nearly empty. Forces such as gravitation were inherently weak.

The forces gradually strengthened, so matter began to clump. In some regions, it grew so dense that a black hole formed.

Space inside the hole expanded at an accelerating rate. Matter inside was cut off from matter outside.

Inside the hole, matter fell toward the middle and increased in density until reaching the limit imposed by string theory.



When matter reached the maximum allowed density, quantum effects caused it to rebound in a big bang. Outside, other holes began to form—each, in effect, a distinct universe.

Tying Down the Loose Ends

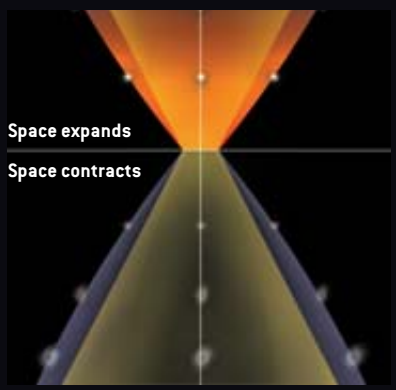
FINALLY, QUANTUM strings have introduced physicists to some striking new symmetries of nature known as dualities, which alter our intuition for what happens when objects get extremely small. I have already alluded to a form of duality: normally, a short string is lighter than a long one, but if we attempt to squeeze down its size below the fundamental length l_s, the string gets heavier again.

Another form of the symmetry, T-duality, holds that small and large extra dimensions are equivalent. This symmetry arises because strings can move in more complicated ways than pointlike particles can. Consider a closed string (a loop) located on a cylindrically shaped space, whose circular cross section represents one finite extra dimension. Besides vibrating, the string can either turn

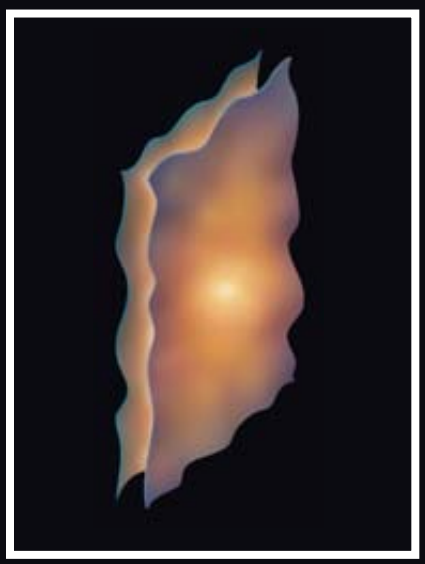
SAMUEL VELASCO

EKPYROTIC SCENARIO

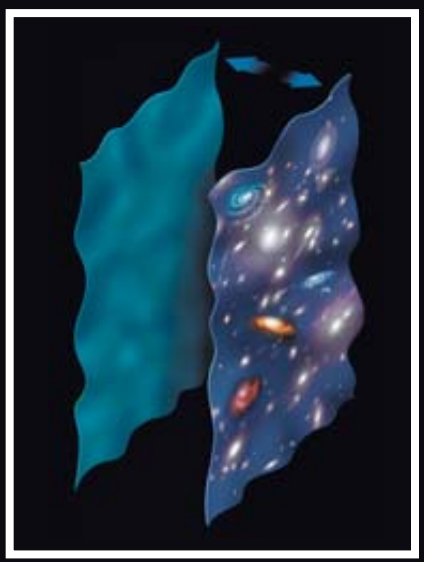
If our universe is a multidimensional membrane, or simply a "brane," cruising through a higher-dimensional space, the big bang may have been the collision of our brane with a parallel one. The collisions might recur cyclically. Each galaxy follows an hourglass-shaped path through spacetime (*below*).



Two nearly empty branes pull each other together. Each is contracting in a direction perpendicular to its motion.



The branes collide, converting their kinetic energy into matter and radiation. This collision is the big bang.



The branes rebound. They start expanding at a decelerating rate. Matter clumps into structures such as galaxy clusters.

In the cyclic model, as the branes move apart, the attractive force between them slows them down. Matter thins out.

The branes stop moving apart and start approaching each other. During the reversal, each brane expands at an accelerated rate.

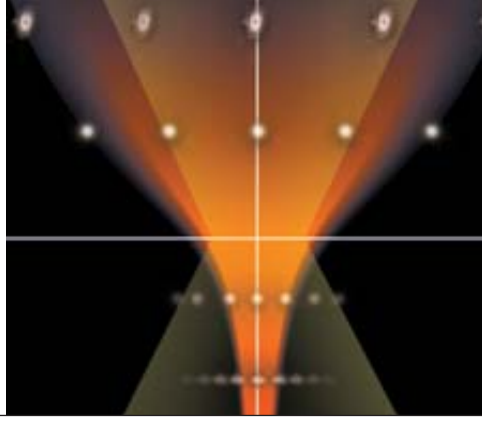
as a whole around the cylinder or wind around it, one or several times, like a rubber band wrapped around a rolled-up poster [*see illustration on page 58*].

The energetic cost of these two states of the string depends on the size of the cylinder. The energy of winding is directly proportional to the cylinder radius: larger cylinders require the string to stretch more as it wraps around, so the windings contain more energy than they would on a smaller cylinder. The energy associated with moving around the circle, on the other hand, is inversely proportional to the radius: larger cylinders allow for longer wavelengths (smaller frequencies), which represent less energy than shorter wavelengths do. If a large cylinder is substituted for a small one, the two states of motion can swap roles. Energies that had been produced by circular motion are instead produced by winding, and vice versa. An outside observer notices only the energy levels, not the origin of those levels. To that observer, the large and small radii are physically equivalent.

Although T-duality is usually described in terms of cylindrical spaces, in which one dimension (the circumference) is finite, a variant of it applies to our or-

SAMUEL VELASCO

Strings abhor infinity. They cannot collapse to an infinitesimal point, so they avoid the paradoxes that collapse would entail.



dinary three dimensions, which appear to stretch on indefinitely. One must be careful when talking about the expansion of an infinite space. Its overall size cannot change; it remains infinite. But it can still expand in the sense that bodies embedded within it, such as galaxies, move apart from one another. The crucial variable is not the size of the space as a whole but its scale factor—the factor by which the distance between galaxies changes, manifesting itself as the galactic redshift that astronomers observe. According to T-duality, universes with small scale factors are equivalent to ones with large scale factors. No such symmetry is present in Einstein's equations; it emerges from the unification that string theory embodies, with the dilaton playing a central role.

For years, string theorists thought that T-duality applied only to closed strings, as opposed to open strings, which have loose ends and thus cannot wind. In 1995 Joseph Polchinski of the University of California at Santa Barbara realized that T-duality did apply to open strings, provided that the switch between large and small radii was accompanied by a change in the conditions at the end points of the string. Until then, physicists had postulated boundary conditions in which no force acted on the ends of the strings, leaving them free to flap around. Under T-duality, these conditions become so-called Dirichlet boundary conditions, whereby the ends stay put.

Any given string can mix both types of boundary conditions. For instance, electrons may be strings whose ends can move around freely in three of the 10 spatial dimensions but are stuck within the other seven. Those three dimensions form a subspace known as a Dirichlet membrane, or D-brane. In 1996 Petr Horava of the University of California at Berkeley and Edward Witten of the Institute for Advanced Study in Princeton, N.J., proposed that our universe resides on such a brane. The partial mobility of electrons and other particles explains why we are unable to perceive the full 10-dimensional glory of space.

Taming the Infinite

ALL THE MAGIC properties of quantum strings point in one direction: strings abhor infinity. They cannot collapse to an infinitesimal point, so they avoid the paradoxes that collapse entails. Their nonzero size and novel symmetries set upper bounds to physical quantities that increase without limit in conventional theories, and they set lower bounds to quantities that decrease. String theorists expect that when one plays the history of the universe backward in time, the curvature of spacetime starts to increase. But instead of going all the way to infinity (at the traditional big bang singularity), it eventually hits a maximum and shrinks once more. Before string theory, physicists were hard-pressed to imagine any mechanism that could so cleanly eliminate the singularity.

Conditions near the zero time of the big bang were so extreme that no one yet knows how to solve the equations. Nevertheless, string theorists have hazarded guesses about the pre-bang universe. Two popular models are floating around.

The first, known as the pre–big bang scenario, which my colleagues and I began to develop in 1991, combines T-duality with the better-known symmetry of time reversal, whereby the equations of physics work equally well when applied backward and forward in time. The combination gives rise to new possible cosmologies in which the universe, say, five seconds before the big bang expanded at the same pace as it did five seconds after the bang. But the rate of change of the expansion was opposite at the two instants: if it was decelerating after the bang, it was accelerating before. In short, the big bang may not have been the origin of the universe but simply a violent transition from acceleration to deceleration.

The beauty of this picture is that it automatically incorporates the great insight of standard inflationary theory—namely, that the universe had to undergo a period of acceleration to become so homogeneous and isotropic. In the standard theory, acceleration occurs after the big bang because of an ad hoc inflaton field. In the pre–big bang scenario, it occurs before the bang as a natural outcome of the novel symmetries of string theory.

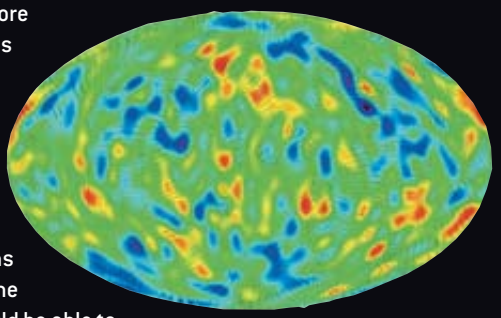
According to the scenario, the pre-bang universe was almost a perfect mirror image of the post-bang one [*see illustration on page 59*]. If the universe is eternal into the future, its contents thinning to a meager gruel, it is also eternal into the past. Infinitely long ago it was nearly empty, filled only with a tenuous, widely dispersed, chaotic gas of radiation and matter. The forces of nature, controlled by the dilaton field, were so feeble that particles in this gas barely interacted.

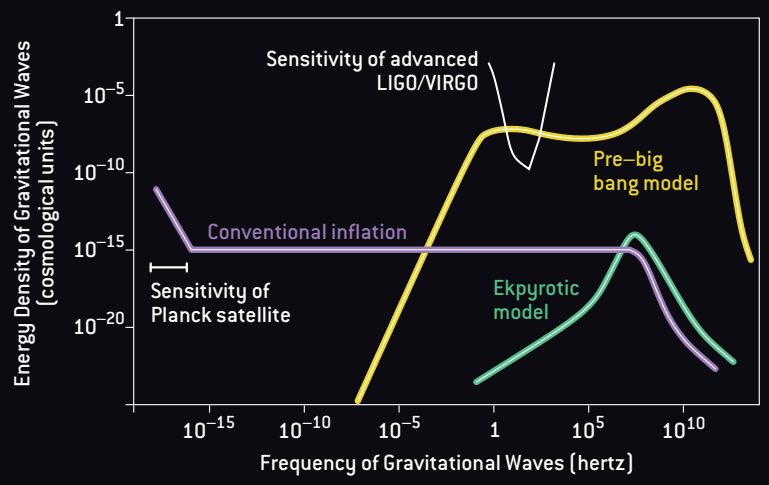
As time went on, the forces gained in strength and pulled matter together. Randomly, some regions accumulated matter at the expense of their surroundings. Eventually the density in these regions became so high that black holes started to form. Matter inside those regions was then cut off from the outside, breaking up the universe into disconnected pieces.

Inside a black hole, space and time swap roles. The center of the black hole is not a point in space but an instant in time. As the infalling matter approached the center, it reached higher and higher densities. But when the density, temperature

OBSERVATIONS

Observing the pre-bang universe may sound like a hopeless task, but one form of radiation could survive from that epoch: gravitational radiation. These periodic variations in the gravitational field might be detected indirectly, by their effect on the polarization of the cosmic microwave background (*simulated view, below*), or directly, at ground-based observatories. The pre–big bang and ekpyrotic scenarios predict more high-frequency gravitational waves and fewer low-frequency ones than do conventional models of inflation (*bottom*). Existing measurements of various astronomical phenomena cannot distinguish among these models, but upcoming observations by the Planck satellite as well as the LIGO and VIRGO observatories should be able to.





and curvature reached the maximum values allowed by string theory, these quantities bounced and started decreasing. The moment of that reversal is what we call a big bang. The interior of one of those black holes became our universe.

Not surprisingly, such an unconventional scenario has provoked controversy. Andrei Linde of Stanford University has argued that for this scenario to match observations, the black hole that gave rise to our universe would have to have formed with an unusually large size—much larger than the length scale of string theory. An answer to this objection is that the equations predict black holes of all possible sizes. Our universe just happened to form inside a sufficiently large one.

A more serious objection, raised by Thibault Damour of the Institut des Hautes Études Scientifiques in Bures-sur-Yvette, France, and Marc Henneaux of the Free University of Brussels, is that matter and spacetime would have behaved chaotically near the moment of the bang, in possible contradiction with the observed regularity of the early universe. I have recently proposed that a chaotic state would produce a dense gas of miniature "string holes"—strings that were so small and massive that they were on the verge of becoming black holes. The behavior of these holes could solve the problem identified by Damour and Henneaux. A similar proposal has been put forward by Thomas Banks of Rutgers University and Willy Fischler of the University of Texas at Austin. Other critiques also exist, and whether they have uncovered a fatal flaw in the scenario remains to be determined.

Bashing Branes

THE OTHER LEADING model for the universe before the bang is the ekpyrotic ("conflagration") scenario. Developed three years ago by a team of cosmologists and string theorists—Justin Khoury of Columbia University, Paul J. Steinhardt of Princeton University, Burt A. Ovrut of the University of Pennsylvania, Nathan Seiberg of the Institute for Advanced Study and Neil Turok of the University of Cambridge—the ekpyrotic scenario relies on the idea that our universe is one of many D-branes floating within a higher-dimensional space. The branes exert attractive forces on one another and occasionally collide. The big bang could be the impact of another brane into ours [*see illustration on page 62*].

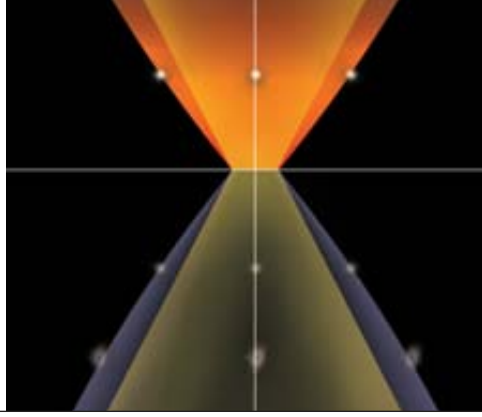
In a variant of this scenario, the collisions occur cyclically. Two branes might hit, bounce off each other, move apart, pull each other together, hit again, and so on. In between collisions, the branes behave like Silly Putty, expanding as they recede and contracting somewhat as they come back together. During the turnaround, the expansion rate accelerates; indeed, the present accelerating expansion of the universe may augur another collision.

The pre–big bang and ekpyrotic scenarios share some common features. Both begin with a large, cold, nearly empty universe, and both share the difficult (and unresolved) problem of making the transition between the pre- and the post-bang phase. Mathematically, the main difference between the scenarios is the behavior of the dilaton field. In the pre–big bang, the dilaton begins with a low value—so that the forces of nature are weak—and steadily gains strength. The opposite is true for the ekpyrotic scenario, in which the collision occurs when forces are at their weakest.

The developers of the ekpyrotic theory initially hoped that the weakness of the forces would allow the bounce to be analyzed more easily, but they were still confronted with a difficult high-curvature situation, so the jury is out on whether the scenario truly avoids a singularity.

NASA (*top*); GABRIELE VENEZIANO *CERN* (*bottom*)

Vestiges of the pre-bangian epoch might show up in galactic and intergalactic magnetic fields.



Also, the ekpyrotic scenario must entail very special conditions to solve the usual cosmological puzzles. For instance, the about-to-collide branes must have been almost exactly parallel to one another, or else the collision could not have given rise to a sufficiently homogeneous bang. The cyclic version may be able to take care of this problem, because successive collisions would allow the branes to straighten themselves.

Leaving aside the difficult task of fully justifying these two scenarios mathematically, physicists must ask whether they have any observable physical consequences. At first sight, both scenarios might seem like an exercise not in physics but in metaphysics—interesting ideas that observers could never prove right or wrong. That attitude is too pessimistic. Like the details of the inflationary phase, those of a possible pre-bangian epoch could have observable consequences, especially for the small variations observed in the cosmic microwave background temperature.

First, observations show that the temperature fluctuations were shaped by acoustic waves for several hundred thousand years. The regularity of the fluctuations indicates that the waves were synchronized. Cosmologists have discarded many cosmological models over the years because they failed to account for this synchrony. The inflationary, pre–big bang and ekpyrotic scenarios all pass this first test. In these three models, the waves were triggered by quantum processes amplified during the period of accelerating cosmic expansion. The phases of the waves were aligned.

Second, each model predicts a different distribution of the temperature fluctuations with respect to angular size. Observers have found that fluctuations of all sizes have approximately the same amplitude. (Discernible deviations occur only on very small scales, for which the primordial fluctuations have been altered by subsequent processes.) Inflationary models neatly reproduce this distribution. During inflation, the curvature of space changed relatively slowly, so fluctuations of different sizes were generated under much the same conditions. In both the stringy models, the curvature evolved quickly, increasing the amplitude of small-scale fluctuations, but other processes boosted the large-scale ones, leaving all fluctuations with the same strength. For the ekpyrotic scenario, those other processes involved the extra dimension of space, the one that separated the colliding branes. For the pre–big bang scenario, they involved a quantum field, the axion, related to the dilaton. In short, all three models match the data.

Third, temperature variations can arise from two distinct processes in the early universe: fluctuations in the density of matter and rippling caused by gravitational waves. Inflation involves both processes, whereas the pre–big bang and ekpyrotic scenarios predominantly involve density variations. Gravitational waves of certain sizes would leave a distinctive signature in the polarization of the microwave background [see "Echoes from the Big Bang," by Robert R. Caldwell and Marc Kamionkowski; SCIENTIFIC AMERICAN, January 2001]. Future observatories, such as European Space Agency's Planck satellite, should be able to see that signature, if it exists—providing a nearly definitive test.

A fourth test pertains to the statistics of the fluctuations. In inflation the fluctuations follow a bell-shaped curve, known to physicists as a Gaussian. The same may be true in the ekpyrotic case, whereas the pre–big bang scenario allows for sizable deviation from Gaussianity.

Analysis of the microwave background is not the only way to verify these theories. The pre–big bang scenario should also produce a random background of gravitational waves in a range of frequencies that, though irrelevant for the microwave background, should be detectable by future gravitational-wave observatories. Moreover, because the pre–big bang and ekpyrotic scenarios involve changes in the dilaton field, which is coupled to the electromagnetic field, they would both lead to large-scale magnetic field fluctuations. Vestiges of these fluctuations might show up in galactic and intergalactic magnetic fields.

So, when did time begin? Science does not have a conclusive answer yet, but at least two potentially testable theories plausibly hold that the universe—and therefore time—existed well before the big bang. If either scenario is right, the cosmos has always been in existence and, even if it recollapses one day, will never end. SA

MORE TO EXPLORE

The Elegant Universe. Brian Greene. W. W. Norton, 1999.

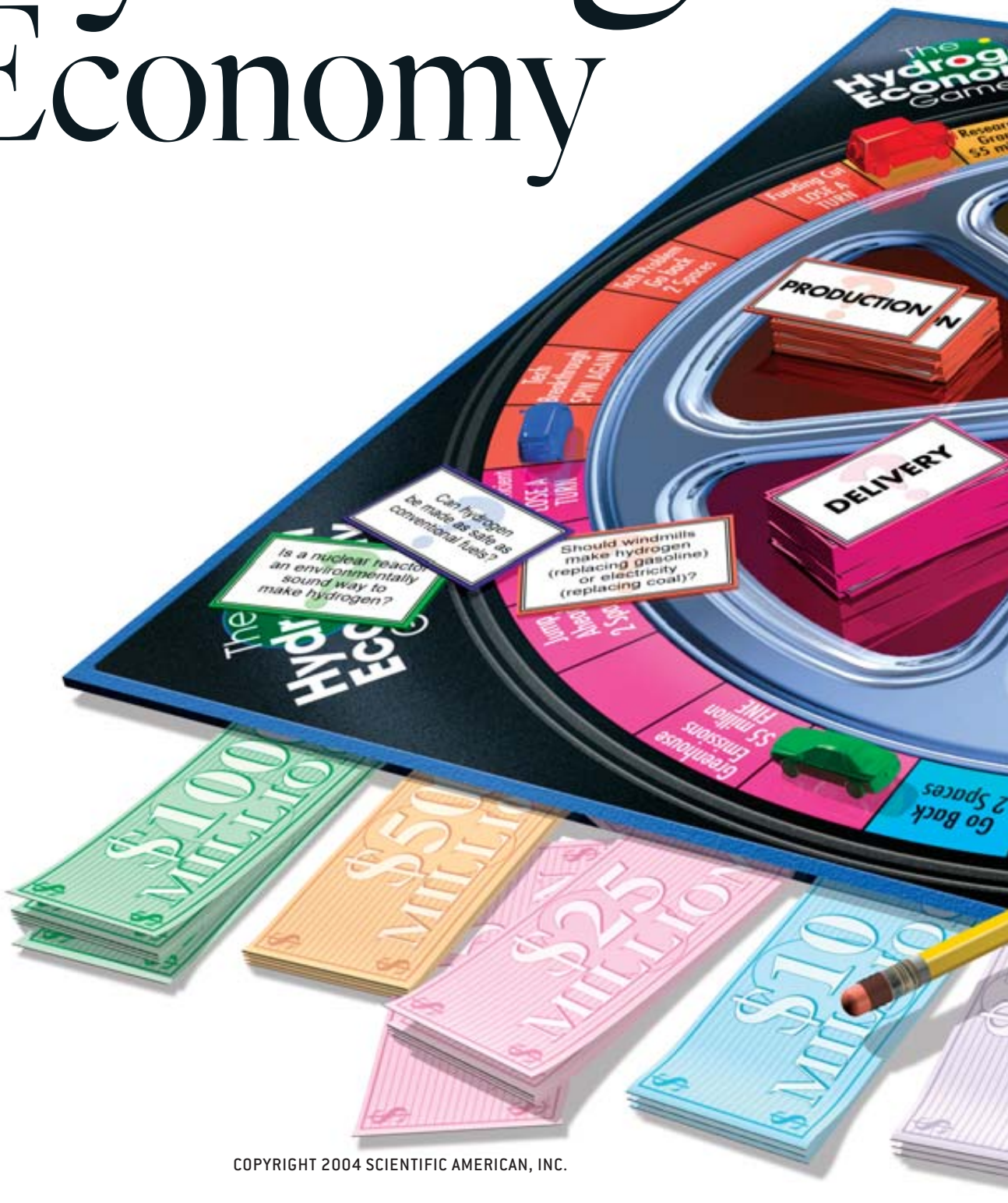
Superstring Cosmology. James E. Lidsey, David Wands and Edmund J. Copeland in *Physics Reports,* Vol. 337, Nos. 4–5, pages 343–492; October 2000. **hep-th/9909061**

From Big Crunch to Big Bang. Justin Khoury, Burt A. Ovrut, Nathan Seiberg, Paul J. Steinhardt and Neil Turok in *Physical Review D,* Vol. 65, No. 8, Paper no. 086007; April 15, 2002. **hep-th/0108187**

A Cyclic Model of the Universe. Paul J. Steinhardt and Neil Turok in *Science,* Vol. 296, No. 5572, pages 1436–1439; May 24, 2002. **hep-th/0111030**

The Pre–Big Bang Scenario in String Cosmology. Maurizio Gasperini and Gabriele Veneziano in *Physics Reports,* Vol. 373, Nos. 1–2, pages 1–212; January 2003. **hep-th/0207130**

Questions about a Hydrogen Economy



Much excitement surrounds the progress in fuel cells, but the quest for a hydrogen economy is no trivial pursuit

By Matthew L. Wald

In the fall of 2003, a few months after President George W. Bush announced a $1.7-billion research program to develop a vehicle that would make the air cleaner and the country less dependent on imported oil, Toyota came to Washington, D.C., with two of them. One, a commercially available hybrid sedan, had a conventional, gasoline-fueled internal-combustion engine supplemented by a battery-powered electric motor. It got about 50 miles to the gallon, and its carbon dioxide emissions were just over half those of an average car. The other auto, an experimental SUV, drove its electric motor with hydrogen fuel cells and emitted as waste only water purer than Perrier and some heat. Which was cleaner?

Answering that question correctly could have a big impact on research spending, on what vehicles the government decides to subsidize as it tries to incubate a technology that will wean us away from gasoline and, ultimately, on the environment. But the answer is not what many people would expect, at least according to Robert Wimmer, research manager for technical and regulatory affairs at Toyota. He said that the two vehicles were about the same.

Overview/*Hydrogen Economy*

- Per a given equivalent unit of fuel, hydrogen fuel cells in vehicles are about twice as efficient as internal-combustion engines. Unlike conventional engines, fuel cells emit only water vapor and heat.
- Hydrogen doesn't exist freely in nature, however, so producing it depends on current energy sources. Sources of hydrogen are either expensive and not widely available (including electrolysis using renewables such as solar, wind or hydropower), or else they produce undesirable greenhouse gases (coal or other fossil fuels).
- Ultimately hydrogen may not be the universal cure-all, although it may be appropriate for certain applications. Transportation may not be one of them.

Wimmer and an increasing number of other experts are looking beyond simple vehicle emissions, to the total effect on the environment caused by the production of the vehicle's fuel and its operation combined. Seen in a broader context, even the supposed great advantages of hydrogen, such as the efficiency and cleanliness of fuel cells, are not as overwhelming as might be thought. From this perspective, coming in neck and neck with a hybrid is something of an achievement; in some cases, the fuel-cell car can be responsible for substantially more carbon dioxide emissions, as well as a variety of other pollutants, the Department of Energy states. And in one way the hybrid is, arguably, superior: it already exists as a commercial product and thus is available to cut pollution now. Fuel-cell cars, in contrast, are expected on about the same schedule as NASA's manned trip to Mars and have about the same level of likelihood.

If that sounds surprising, it is also revealing about the uncertainties and challenges that trail the quest for a hydrogen economy—wherein most energy is devoted to the creation of hydrogen, which is then run through a fuel cell to make electricity. Much hope surrounds the advances in fuel cells and the possibility of a cleaner hydrogen economy, which could include not only transportation but also power for houses and other buildings. Last November U.S. Energy Secretary Spencer Abraham told a Washington gathering of energy ministers from 14 countries and the European Union that hydrogen could "revolutionize the world in which we live." Noting that the nation's more than 200 million motor vehicles consume about two thirds of the 20 million barrels of oil the U.S. uses every day, President Bush has called hydrogen the "freedom fuel."

But hydrogen is not free, in either dollars or environmental damage. The hydrogen fuel cell costs nearly 100 times as much per unit of power produced as an internal-combustion engine. To be price competitive, "you've got to be at a nickel a watt, and we're at $4 a watt," says Tim R. Dawsey, a research associate at Eastman Chemical Company, which makes polymers for fuel cells. Hydrogen is also about five times as expensive, per unit of usable energy, as gasoline. Simple dollars are only one speed bump on the road to the hydrogen economy. Another is that

SLIM FILMS (*preceding pages*)

FACE OFF: If total life-cycle environmental impact of a given fuel is included, the Toyota Prius (*right*), a hybrid that has a gasoline internal-combustion engine supplemented by an electric motor, compares favorably with the company's experimental hydrogen fuel-cell SUV (*left*).

supplying the energy required to make pure hydrogen may itself cause pollution. Even if that energy is from a renewable source, like the sun or the wind, it may have more environmentally sound uses than the production of hydrogen. Distribution and storage of hydrogen—the least dense gas in the universe—are other technological and infrastructure difficulties. So is the safe handling of the gas. Any practical proposal for a hydrogen economy will have to address all these issues.

Which Sources Make Sense?

HYDROGEN FUEL CELLS have two obvious attractions. First, they produce no pollution at point of use [see "Vehicle of Change," by Lawrence D. Burns, J. Byron McCormick and Christopher E. Borroni-Bird; SCIENTIFIC AMERICAN, October 2002]. Second, hydrogen can come from myriad sources. In fact, the gas is not a fuel in the conventional sense. A fuel is something found in nature, like coal, or refined from a natural product, like diesel fuel from oil, and then burned to do work. Pure hydrogen does not exist naturally on earth and is so highly processed that it is really more of a carrier or medium for storing and transporting energy from some original source to a machine that makes electricity. "The beauty of hydrogen is the fuel diversity that's possible," said David K. Garman, U.S. assistant secretary for energy efficiency and renewable energy. Each source, however, has an ugly side.

For instance, a process called electrolysis makes hydrogen by splitting a water molecule with electricity [*see illustration on page 72*]. The electricity could come from solar cells, windmills, hydropower or safer, next-generation nuclear reactors [see "Next-Generation Nuclear Power," by James A. Lake, Ralph G. Bennett and John F. Kotek; SCIENTIFIC AMERICAN, January 2002]. Researchers are also trying to use microbes to transform biomass, including parts of crops that now have no economic value, into hydrogen. In February researchers at the University of Minnesota and the University of Patras in Greece announced a chemical reactor that generates hydrogen from ethanol mixed with water. Though appealing, all these technologies are either unaffordable or unavailable on a commercial scale and are likely to remain so for many years to come, according to experts.

Hydrogen could be derived from coal-fired electricity, which is the cheapest source of energy in most parts of the country. Critics argue, though, that if coal is the first ingredient for the hydrogen economy, global warming could be exacerbated through greater release of carbon dioxide.

Or hydrogen could come from the methane in natural gas, methanol or other hydrocarbon fuel [*see illustration on page 72*]. Natural gas can be reacted with steam to make hydrogen and carbon dioxide. Filling fuel cells, however, would preclude the use of natural gas for its best industrial purpose today: burning in high-efficiency combined-cycle turbines to generate electricity. That, in turn, might again lead to more coal use. Combined-cycle plants can turn 60 percent of the heat of burning natural gas into electricity; a coal plant converts only about 33 percent. Also, when burned, natural gas produces just over half as much carbon dioxide per unit of heat as coal does, 117 pounds per million Btu versus 212. As a result, a kilowatt-hour of electricity made from a new natural gas plant has slightly over one fourth as much carbon dioxide as a kilowatt-hour from coal. (Gasoline comes between coal and natural gas, at 157 pounds of carbon dioxide per million Btu.) In sum, it seems better for the environment to use natural gas to make electricity for the grid and save coal, rather than turning it into hydrogen to save gasoline.

Two other fuels could be steam-reformed to give off hydrogen: the oil shipped from Venezuela or the Persian Gulf and, again, the coal from Appalachian mines. To make hydrogen from fossil fuels in a way that does not add to the release of climate-changing carbon dioxide, the carbon must be captured so that it does not enter the atmosphere. Presumably this process would be easier than sequestering carbon from millions of

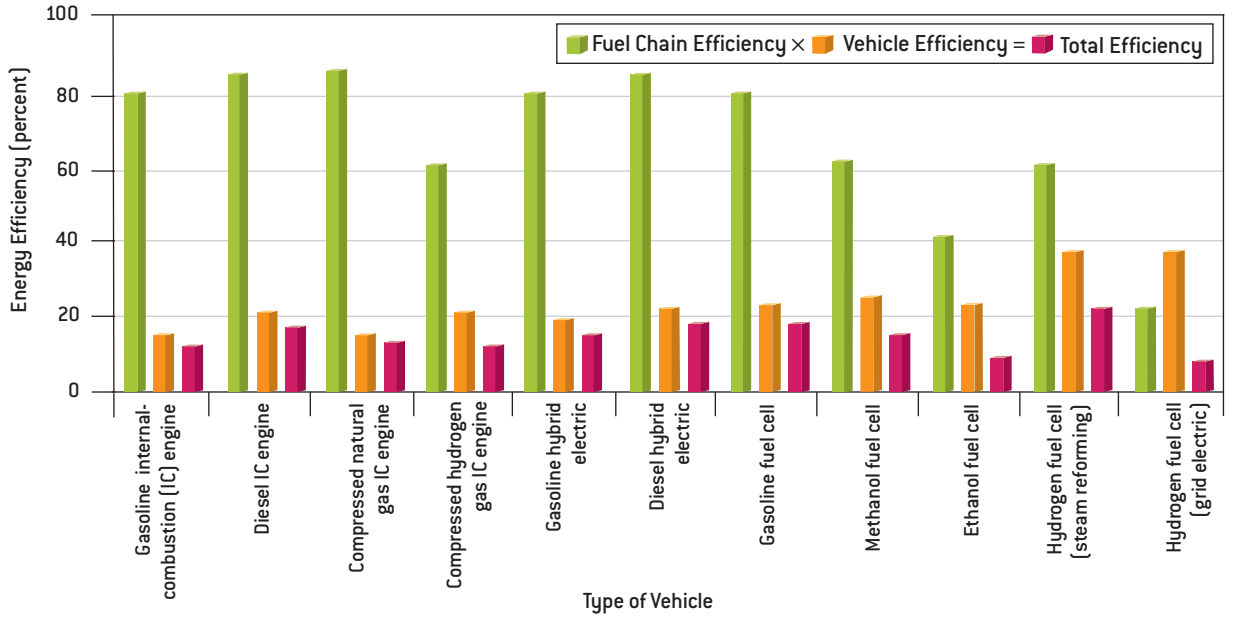
THE AUTHOR

MATTHEW L. WALD is a reporter at the *New York Times,* where he has been covering energy since 1979. He has written about oil refining; coal mining; electricity production from coal, natural gas, uranium, wind and solar energy; electric and hybrid automobiles; and air pollution from energy use. His current assignment is in Washington, D.C., where he also writes about transportation safety and other technical topics.

TOYOTA

WELL-TO-WHEELS ENERGY EFFICIENCY

Total energy efficiency includes not only vehicle operation but also the energy required to produce fuel. Extracting oil, refining gasoline and trucking that fuel to filling stations for internal-combustion engines is more efficient than creating hydrogen for fuel cells.



tailpipes. Otherwise, the fuels might as well be burned directly.

"If you look at it from the whole system, not the individual sector, you may do better to get rid of your coal-fired power plants, because coal is such a carbon-intensive fuel," says Michael Wang, an energy researcher at Argonne National Laboratory. Coal accounts for a little more than half the kilowatt-hours produced in the U.S.; about 20 percent is from natural gas. The rest comes from mostly carbon-free sources, primarily nuclear reactors and hydroelectricity. Thus, an effort to replace the coal-fired electric plants would most likely take decades.

In any case, if hydrogen were to increase suddenly in supply, fuel cells might not even be the best use for the gas. In a recent paper, Reuel Shinnar, professor of chemical engineering at the City College of New York, reviewed the alternatives for power and fuel production. Rather than the use of hydrogen as fuel, he suggested something far simpler: increased use of hydrocracking and hydrotreating. The U.S. could save three million barrels of oil a day that way, Shinnar calculated. Hydrocracking and hydrotreating both start with molecules in crude oil that are unsuitable for gasoline because they are too big and have a carbon-to-hydrogen ratio that is too heavy with carbon. The processes are expensive but still profitable, because they allow the refineries to take ingredients that are good for only low-value products, such as asphalt and boiler fuel, and turn them into gasoline. It is like turning chuck steak into sirloin.

What about Conversion Costs?

IF HYDROGEN PRODUCTION is dirty and expensive, could its impressive energy efficiency at point of use make up for those downsides? Again, the answer is complicated.

A kilo of hydrogen contains about the same energy as a gallon of unleaded regular gas—that is, if burned, each would give off about the same amount of heat. But the internal-combustion engine and the fuel cell differ in their ability to extract usable work from that fuel energy. In the engine, most of the energy flows out of the tailpipe as heat, and additional energy is lost to friction inside the engine. In round numbers, advocates and detractors agree, a fuel cell gets twice as much work out of a kilo of hydrogen as an engine gets out of a gallon of gas. (In a stationary application—such as a basement appliance that takes the hydrogen from natural gas and turns it into electricity to run the household—efficiency could be higher, because the heat given off by the fuel-cell process could also be used—for example, to heat tap water.)

There is, in fact, a systematic way to evaluate where best to use each fuel. A new genre of energy analysis, "well to wheels," compares the energy efficiency of every known method to turn a vehicle's wheels [*see illustration above*]. The building block of the well-to-wheels performance is "conversion efficiency." At every step of the energy chain, from pumping oil out of the ground to refining it to burning it in an engine, some of the original energy potential of the fuel is lost.

The first part of the well-to-wheels determination is what engineers call "well to tank": what it takes to make and deliver a fuel. When natural gas is cracked for hydrogen, about 40 percent of the original energy potential is lost in the transfer, according to the DOE Office of Energy Efficiency and Renewable Energy. Using electricity from the grid to make hydrogen by electrolysis of water causes a loss of 78 percent. (Despite the lower efficiency of electrolysis, it is likely to predominate in the early stages of a hydrogen economy because it is convenient—producing the hydrogen where it is needed and thus avoiding ship-

LUCY READING; SOURCE: DEPARTMENT OF ENERGY

ping problems.) In contrast, pumping a gallon of oil out of the ground, taking it to a refinery, turning it into gasoline and getting that petrol to a filling station loses about 21 percent of the energy potential. Producing natural gas and compressing it in a tank loses only about 15 percent.

The second part of the total energy analysis is "tank to wheels," or the fraction of the energy value in the vehicle's tank that actually ends up driving the wheels. For the conventional gasoline internal-combustion engine, 85 percent of the energy in the gasoline tank is lost; thus, the whole system, well to tank combined with tank to wheels, accounts for a total loss of 88 percent.

The fuel cell converts about 37 percent of the hydrogen's energy value to power for the wheels. The total loss, well to wheels, is about 78 percent if the hydrogen comes from steam-reformed natural gas. If the source of the hydrogen is electrolysis from coal, the loss from the well (a mine, actually) to tank is 78 percent; after that hydrogen runs through a fuel cell, it loses another 43 percent, with the total loss reaching 92 percent.

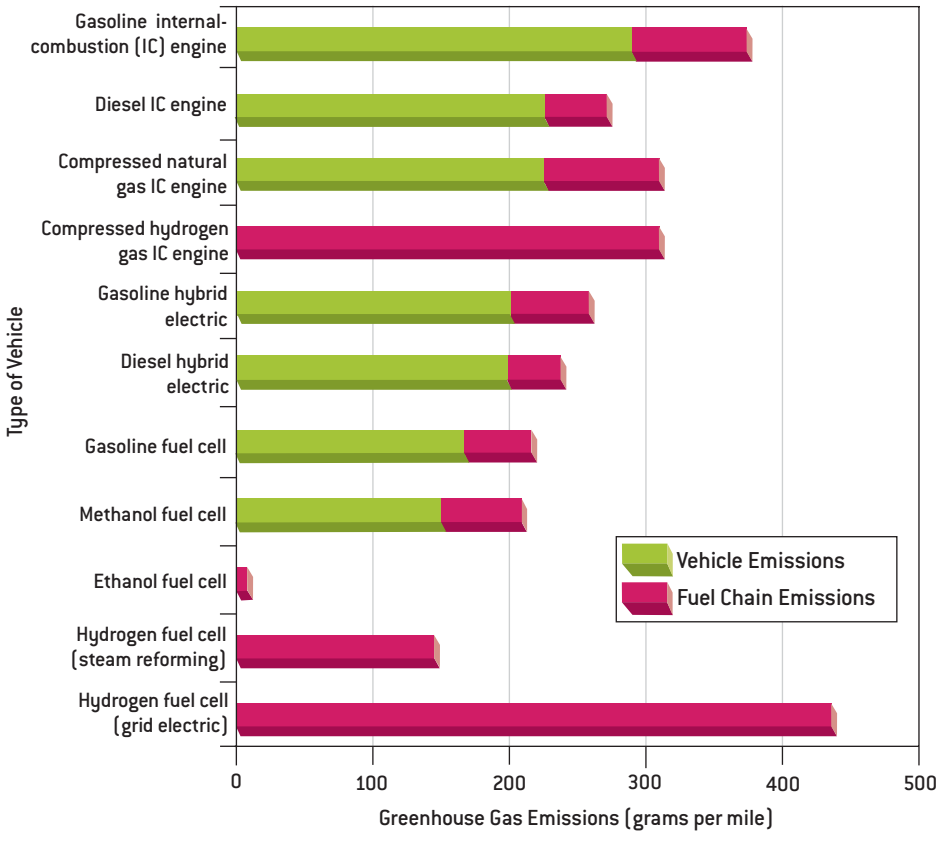
Wally Rippel, a research engineer at AeroVironment in Monrovia, Calif., who helped to develop the General Motors EV-1 electric car and the NASA Helios Solar Electric airplane, offers another way to look at the situation. He calculates that in a car that employs an electric motor to turn the wheels, a kilowatt-hour used to recharge batteries will propel the auto three times as far as if that same kilowatt-hour were instead used to make hydrogen for a fuel cell.

All these facts add up to an argument *not* to use electricity to make hydrogen and then go back to electricity again with an under-the-hood fuel cell. But there is one strong reason to go through inefficient multiple conversions. They may still make economic sense, and money is what has shaped the energy markets so far. That is, even if the hydrogen system is very wasteful of energy, there are such huge differences in the cost of energy from various sources that it might make sense to switch to a system that lets us go where the cheapest energy is.

Walter "Chip" Schroeder, president and chief executive of Proton Energy Systems, a Connecticut company that builds electrolysis machines, explains the economic logic. Coal at current prices (which is to say, coal at prices that are likely to prevail for years to come) costs a little more than 80 cents per million Btu. Gasoline at $1.75 a gallon (which seems pricey at the moment but in a few months or years could look cheap) is about $15.40. The mechanism for turning a Btu from coal into a Btu that will run a car is cumbersome, but in the transition, "you end up with wine, not water," he says. Likewise, he describes his device to turn water into hydrogen as an "arbitrage machine." "Arbitrage" is the term used by investment bankers or stock or commodities traders to describe buying low and selling high, but it usually refers to small differences in the price

TOTAL EMISSIONS OF VEHICLES

Emissions of greenhouse gases (carbon dioxide or equivalent) vary depending on the combined effects of the vehicle's operation and the source of the fuel. Fuel-cell vehicles emit no greenhouse gases themselves, but the creation of the hydrogen fuel can be responsible for more emissions overall than conventional gasoline internal-combustion engines are. (The Energy Department calculates that ethanol derived from corn has almost no greenhouse gas emissions, because carbon emitted by ethanol use is reabsorbed by new corn.)



LUCY READING; SOURCE: DEPARTMENT OF ENERGY

of a stock or the value of a currency between one market or another. "You can't make reasonable policy without understanding just how extreme the value differentials in our energy marketplace are," Schroeder says.

How to Deliver the Hydrogen?

DIFFERENT SOURCES of energy may not be as fungible as money is in arbitrage, however. There is a problem making hydrogen conveniently available at a good cost, at least if the hydrogen is going to come from renewable sources such as solar, hydropower or wind that are practical in only certain areas of the country.

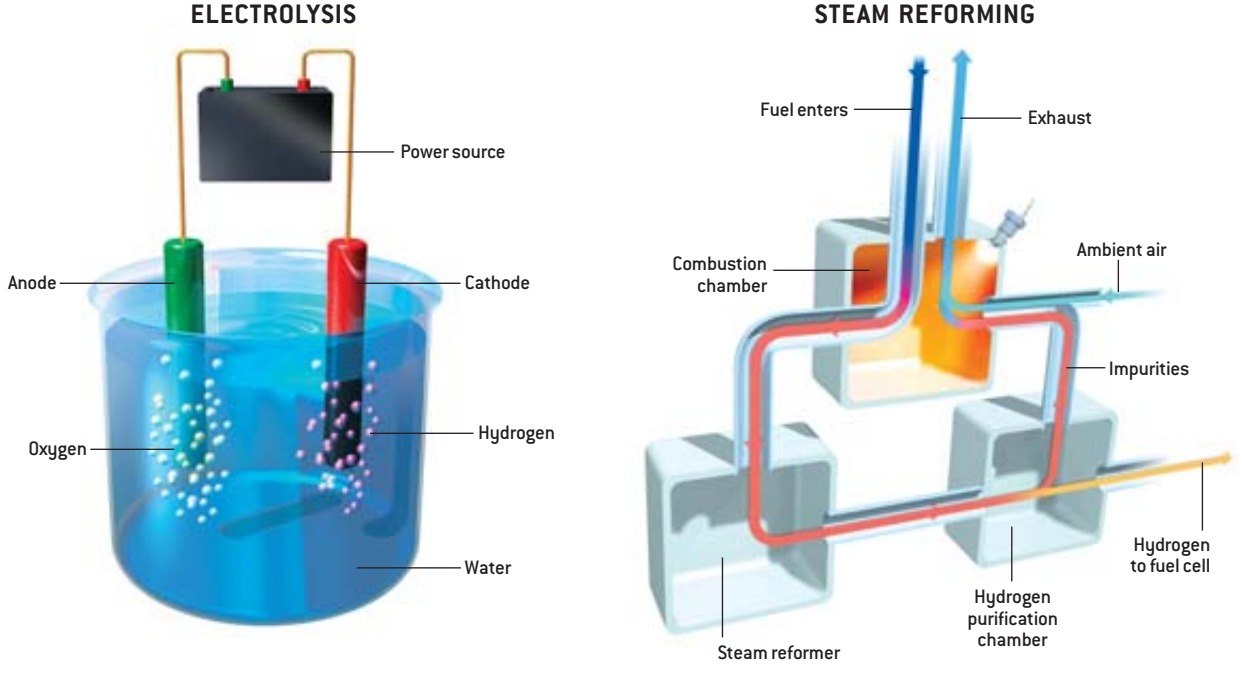
Hydrogen from wind, for example, is competitive with gasoline when wind power costs three cents a kilowatt-hour, says Garman of the DOE. That occurs where winds blow steadily. "Where I might get three-cent wind tends to be in places where people don't live," he notes. In the U.S., such winds exist in a belt running from Montana and the Dakotas to Texas. The electric power they produce would have a long way to go to reach the end users—with energy losses throughout the grid along the way. "You can't get the electrons out of the Dakotas because of transmission constraints," Garman points out. "Maybe a hydrogen pipeline could get the tremendous wind resource carried to Chicago," the nearest motor-fuel market.

That is, if such a pipeline were even practical to build. Given hydrogen's low density, it is far harder to deliver than, for instance, natural gas. To move large volumes of any gas requires compressing it, or else the pipeline has to have a diameter similar to that of an airplane fuselage. Compression takes work, and that drains still more energy from the total production process. Even in this instance, managing hydrogen is trickier than dealing with other fuel gases. Hydrogen compressed to about 790 atmospheres has less than a third of the energy of the methane in natural gas at the same pressure, points out a recent study by three European researchers, Ulf Bossel, Baldur Eliasson and Gordon Taylor.

A related problem is that a truck that could deliver 2,400 kilos of natural gas to a user would yield only 288 kilos of hydrogen pressurized to the same level, Bossel and his colleagues find. Put another way, it would take about 15 trucks to deliver the hydrogen needed to power the same number of cars that could be served by a single gasoline tanker. Switch to liquid hydrogen, and it would take only about three trucks to equal the one gasoline tanker, but hydrogen requires substantially more effort to liquefy. Shipping the hydrogen as methanol that could be reformed onboard the vehicle [*see illustration below*] would ease transport, but again, the added transition has an energy penalty. These facts argue for using the hydrogen where it is pro-

CREATING HYDROGEN

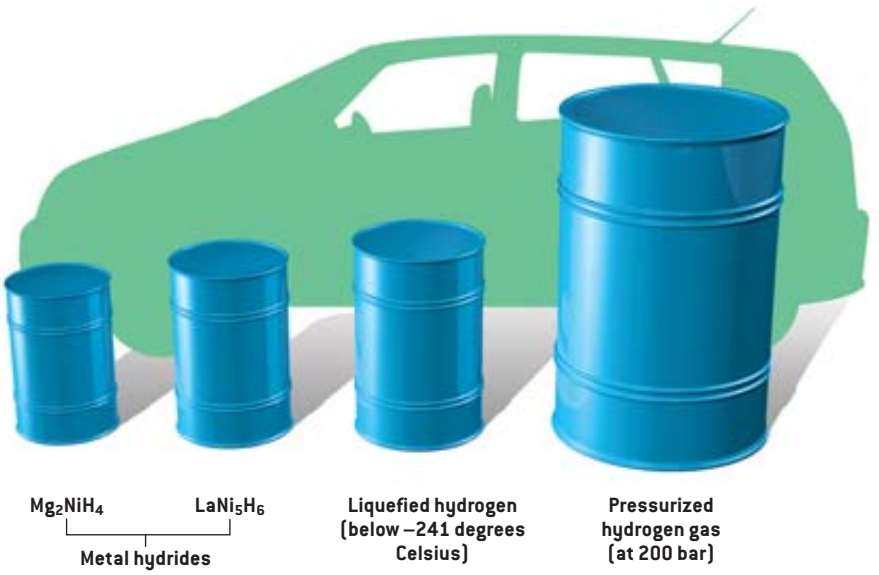
Two main methods are known for extracting hydrogen, which does not occur in pure form naturally on the earth. Electrolysis (*left*) uses electric current to split molecules of water (H_2O). A cathode (negative terminal) attracts hydrogen atoms, and an anode (positive) attracts oxygen; the two gases bubble up into air and can be captured. In steam reforming (*right*), a hydrocarbon such as methanol (CH_3OH) first vaporizes in a heated combustion chamber. A catalyst in the steam reformer breaks apart fuel and water vapor to produce components including hydrogen, which is then separated and routed to a fuel cell.



SLIM FILMS; SOURCE: IDATECH

SAME HYDROGEN, DIFFERENT VOLUMES

Containing the lightest gas in the universe onboard a car presents a challenge, as is clear from the differences in volume of some options for storing four kilograms of hydrogen—enough for a 250-mile driving range. (Four kilograms of hydrogen holds about the same energy as four gallons of gasoline. Because fuel cells are about twice as efficient as internal-combustion engines, that four kilograms takes the car as far as eight gallons of gasoline.) Current alternatives, including tanks that hold pressurized gas or liquefied hydrogen, are too big. Experimental metal hydrides or other solid-state technologies might be able to release hydrogen on demand and be recharged later, but they also carry a weight penalty or an energy penalty for the chemical transformations.



duced, which may be distant from the major motor-fuel markets.

No matter how hydrogen reaches its destination, the difficulties of handling the elusive gas will not be over. Among hydrogen's disadvantages is that it burns readily. All gaseous fuels have a minimum and maximum concentration at which they will burn. Hydrogen's range is unusually broad, from 2 to 75 percent. Natural gas, in contrast, burns between 5 and 15 percent. Thus, as dangerous as a leak of natural gas is, a hydrogen leak is worse, because hydrogen will ignite at a wider range of concentrations. The minimum energy necessary to ignite hydrogen is also far smaller than that for natural gas.

And when hydrogen burns, it does so invisibly. NASA published a safety manual that recommends checking for hydrogen fires by holding a broom at arm's length and seeing if the straw ignites. "It's scary—you cannot see the flame," says Michael D. Amiridis, chair of the department of chemical engineering at the University of South Carolina, which performs fuel-cell research under contract for a variety of companies. A successful fuel-cell car, he says, would have "safety standards at least equivalent to the one I have now." A major part of the early work on developing a hydrogen fueling supply chain has been building warning instruments that can reliably detect hydrogen gas.

A Role for Hydrogen

DESPITE THE TECHNOLOGICAL and infrastructure obstacles, a hydrogen economy may be coming. If it is, it will most likely resemble the perfume economy, a market where quantities are so small that unit prices do not matter. Chances are good that it will start in cellular phones and laptop computers, where consumers might not mind paying $10 a kilowatt-hour for electricity from fuel cells; a recent study by the fuel-cell industry predicts that the devices could be sold in laptop computers this year. It might eventually move to houses, which will run nicely on five kilowatts or so and where an improvement in carbon efficiency is highly desirable because significant electricity demand exists almost every hour of the day. But hydrogen cells may not appear in great numbers in driveways, where cars have a total energy requirement of about 50 kilowatts apiece but may run only an average of two hours a day—a situation that is exactly backward from where a good engineer would put a device like a fuel cell, which has a low operating cost but a high cost per unit of capacity. Although most people may have heard of fuel cells as alternative power sources for cars, cars may be the last place they'll end up on a commercial scale.

If we need to find substitutes for oil for transportation, we may look to several places before hydrogen. One is natural gas, with very few technical details to work out and significant supplies available. Another is electricity for electric cars. Battery technology has hit some very significant hurdles, but they might be easier to solve than those of fuel cells. If we have to, we can run vehicles on methanol from coal; the Germans did it in the 1940s, and surely we could figure it out today.

Last, if we as a society truly support the development of renewable sources such as windmills and solar cells, they could replace much of the fossil fuels used today in the electric grid system. With that development, plus judicious conservation, we would have a lot of energy left over for the transportation sector, the part of the economy that is using up the oil and making us worry about hydrogen in the first place. SA

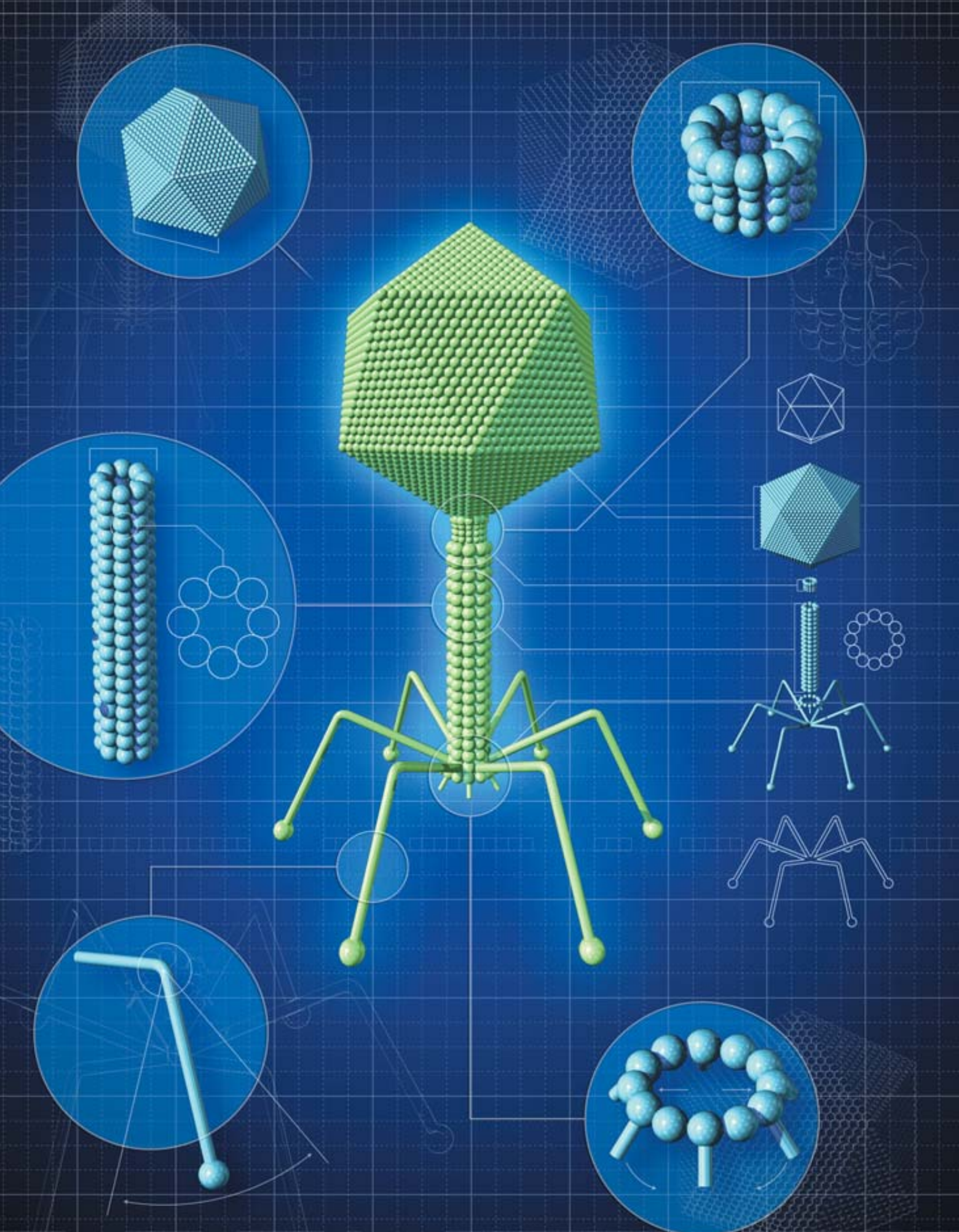
MORE TO EXPLORE

The Hydrogen Economy: Opportunities, Costs, Barriers, and R&D Needs. National Academies Press, 2004.

The Hype about Hydrogen. Joseph J. Romm. Island Press, 2004.

U.S. Department of Energy, Energy Efficiency and Renewable Energy Web pages on hydrogen: **www.eere.energy.gov/hydrogenandfuelcells/**

SLIM FILMS; SOURCE: *NATURE*



Biologists are crafting libraries of interchangeable DNA parts and assembling them inside microbes to create programmable, living machines

SYNTHETIC LIFE

By W. Wayt Gibbs

Evolution is a wellspring of creativity; 3.6 billion years of mutation and competition have endowed living things with an impressive range of useful skills. But there is still plenty of room for improvement. Certain microbes can digest the explosive and carcinogenic chemical TNT, for example—but wouldn't it be handy if they glowed as they did so, highlighting the location of buried land mines or contaminated soil? Wormwood shrubs generate a potent medicine against malaria but only in trace quantities that are expensive to extract. How many millions of lives could be saved if the compound, artemisinin, could instead be synthesized cheaply by vats of bacteria? And although many cancer researchers would trade their eyeteeth for a cell with a built-in, easy-to-read counter that ticks over reliably each time it divides, nature apparently has not deemed such a thing fit enough to survive in the wild.

It may seem a simple matter of genetic engineering to rewire cells to glow in the presence of a particular toxin, to manufacture an intricate drug, or to keep track of the cells' age. But creating such biological devices is far from easy. Biologists have been transplanting genes from one species to another for 30 years, yet genetic engineering is still more of a craft than a mature engineering discipline.

"Say I want to modify a plant so that it changes color in the presence of TNT," posits Drew Endy, a biologist at the Massachusetts Institute of Technology. "I can start tweaking genetic pathways in the plant to do that, and if I am lucky, then after a year or two I may get a 'device'—one system. But doing that once doesn't help me build a cell that swims around and eats plaque from artery walls. It doesn't help me grow a little microlens. Basically the current practice produces pieces of art."

Endy is one of a small but rapidly growing number of scientists who have set out in recent years to buttress the foundation of genetic engineering with what they call synthetic biology. They are designing and building living systems that behave in predictable ways, that use interchangeable parts, and in some cases that operate with an expanded genetic code, which allows them to do things that no natural organism can.

This nascent field has three major goals: One, learn about life by building it, rather than by tearing it apart. Two, make genetic engineering worthy of its

REDESIGNED VIRUSES will help biologists learn how to build reliable genetic machines. A group at the Massachusetts Institute of Technology has reorganized the genome of the T7 bacteriophage drawn here.

BRYAN CHRISTIE DESIGN

DREW ENDY (*pictured*) and others at M.I.T. have designed and built more than 140 "BioBricks" (*in vials*). Each is a piece of DNA that performs a well-characterized function and interacts well with other genetic parts.

name—a discipline that continuously improves by standardizing its previous creations and recombining them to make new and more sophisticated systems. And three, stretch the boundaries of life and of machines until the two overlap to yield truly programmable organisms. Already TNT-detecting and artemisinin-producing microbes seem within reach. The current prototypes are relatively primitive, but the vision is undeniably grand: think of it as Life, version 2.0.

A Light Blinks On

THE ROOTS OF SYNTHETIC BIOLOGY extend back 15 years to pioneering work by Steven A. Benner and Peter G. Schultz. In 1989 Benner led a team at ETH Zurich that created DNA containing two artificial genetic "letters" in addition to the four that appear in life as we know it. He and others have since invented several varieties of artificially enhanced DNA. So far no one has made genes from altered DNA that are functional—transcribed to RNA and then translated to protein form—within living cells. Just within the past year, however, Schultz's group at the Scripps Research Institute developed cells (containing normal DNA) that generate unnatural amino acids and string them together to make novel proteins [*see box on page 80*].

Overview/*Synthetic Biology*

- Molecular biology has been largely a reductive science that deduces the operation of living systems by breaking them apart.
- A growing number of synthetic biologists are taking a different approach: building machines from interchangeable DNA parts. The devices work inside living cells, from which they derive energy, raw materials, and the ability to move and reproduce.
- Synthetic biology has already produced microbes with a variety of unnatural talents. Some produce complex chemical ingredients for drugs; others make artificial amino acids, remove heavy metals from wastewater or perform simple binary logic.

Benner and other "old school" synthetic biologists see artificial genetics as a way to explore basic questions, such as how life got started on earth and what forms it may take elsewhere in the universe. Interesting as that is, the recent buzz growing around synthetic biology arises from its technological promise as a way to design and build machines that work inside cells. Two such devices, reported simultaneously in 2000, inspired much of the work that has happened since.

Both devices were constructed by inserting selected DNA sequences into *Escherichia coli,* a normally innocuous bacterium in the human gut. The two performed very different functions, however. Michael Elowitz and Stanislaus Leibler, then at Princeton University, assembled three interacting genes in a way that made the *E. coli* blink predictably, like microscopic Christmas tree lights [*see box on opposite page*]. Meanwhile James J. Collins, Charles R. Cantor and Timothy S. Gardner of Boston University made a genetic toggle switch. A negative feedback loop—two genes that interfere with each other—allows the toggle circuit to flip between two stable states. It effectively endows each modified bacterium with a rudimentary digital memory.

To engineering-minded biologists, these experiments were energizing but also frustrating. It had taken nearly a year to create the toggle switch and about twice that time to build the flashing microbes. And no one could see a way to connect the two devices to make, for example, blinking bacteria that could be switched on and off.

"We would like to be able to routinely assemble systems from pieces that are well described and well behaved," Endy remarks. "That way, if in the future someone asks me to make an organism that, say, counts to 3,000 and then turns left, I can grab the parts I need off the shelf, hook them together and predict how they will perform." Four years ago parts such as these were just a dream. Today they fill a box on Endy's desk.

Building with BioBricks

"THESE ARE GENETIC PARTS," Endy says as he holds out a container filled with more than 50 vials of clear, syrupy fluid. "Each of these vials contains copies of a distinct section of DNA that either performs some function on its own or can be

SAM OGDEN

HOW A GENETIC PART WORKS

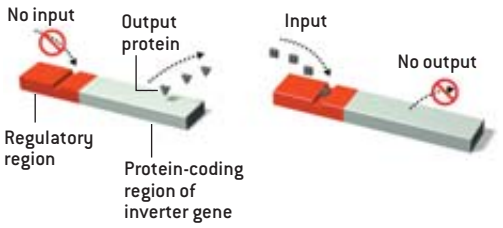
Assemblies of genes and regulatory DNA can act as the biochemical equivalent of electronic components, performing Boolean logic.

A COMPONENT

A biochemical inverter performs the Boolean NOT operation in response to an input signal, in the form of a protein encoded by another gene.

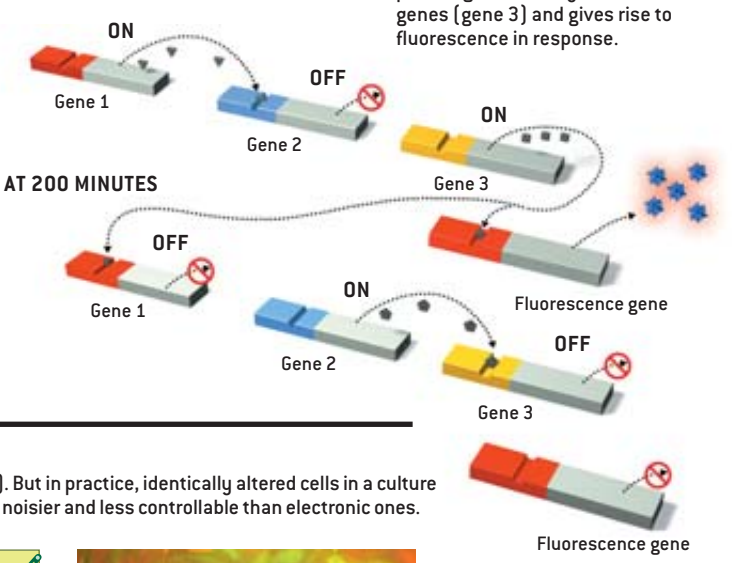
ON
When no input protein is present (input = 0), the inverter gene is "on"—it gives rise to its encoded protein (output = 1).

OFF
When input protein is abundant (input = 1), the inverter gene turns off (output = 0).



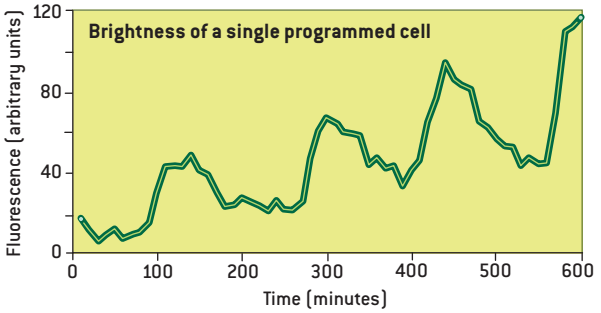
A CIRCUIT

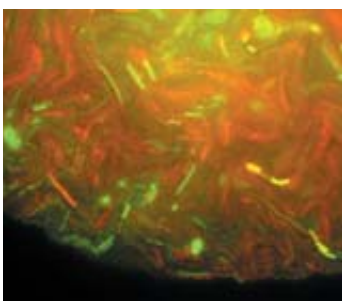
One simple genetic circuit connects three inverters, each of which contains a different gene (gene 1, 2 or 3). The genes oscillate between on and off states as the signal propagates through the circuit. The behavior is monitored through a gene (*far right*) that intercepts some of the output protein generated by one of the inverter genes (gene 3) and gives rise to fluorescence in response.



A CIRCUIT IN ACTION

Cells containing such a circuit blink on and off repeatedly (*graph*). But in practice, identically altered cells in a culture (*photograph*) blink at varying rates, because genetic circuits are noisier and less controllable than electronic ones.





used by a cell to make a protein that does something useful. What is important here is that each genetic part has been carefully designed to interact well with other parts, on two levels." At a mechanical level, individual BioBricks (as the M.I.T. group calls the parts) can be fabricated and stored separately, then later stitched together to form larger bits of DNA. And on a functional level, each part sends and receives standard biochemical signals. So a scientist can change the behavior of an assembly just by substituting a different part at a given spot.

"Interchangeable components are something we take for granted in other kinds of engineering," Endy notes, but genetic engineering is only beginning to draw on the power of the concept. One advantage it offers is abstraction. Just as electrical engineers need not know what is inside a capacitor before they use it in a circuit, biological engineers would like to be able to use a genetic toggle switch while remaining blissfully ignorant of the binding coefficients and biochemical makeup of the promoters, repressors, activators, inducers and other genetic elements that make the switch work. One of the vials in Endy's box, for example, contains an inverter BioBrick (also called a NOT operator). When its input signal is high, its output signal is low, and vice versa. Another BioBrick performs a Boolean AND function, emitting an output signal only when it receives high levels of both its inputs. Because the two parts work with compatible signals, connecting them creates a NAND (NOT AND) operator. Virtually any binary computation can be performed with enough NAND operators.

Beyond abstraction, standardized parts offer another powerful advantage: the ability to design a functional genetic system without knowing exactly how to make it. Early last year a class of 16 students was able in one month to specify four genetic programs to make groups of *E. coli* cells flash in unison, as fireflies sometimes do. The students did not know how to create DNA sequences, but they had no need to. Endy hired a DNA-synthesis company to manufacture the 58 parts called for in their designs. These new BioBricks were then added to

BRYAN CHRISTIE DESIGN (*top*); LUCY READING (*graph*); MICHAEL ELOWITZ (*bottom right*)

Living machines reproduce, but as they do, they mutate.

M.I.T.'s Registry of Standard Biological Parts. That online database today lists more than 140 parts, with the number growing by the month.

Hijacking Cells

AS USEFUL AS IT HAS BEEN to apply the lessons of other fields of engineering to genetics, beyond a certain point the analogy breaks down. Electrical and mechanical machines are generally self-contained. That is true for a select few genetic devices: earlier this year, for example, Milan Stojanovic of Columbia University contrived test tubes of DNA-like biomolecules that play a chemical version of tic-tac-toe. But synthetic biologists are mainly interested in building genetic devices within living cells, so that the systems can move, reproduce and interact with the real world. From a cell's point of view, the synthetic device inside it is a parasite. The cell provides it with energy, raw materials and the biochemical infrastructure that decodes DNA to messenger RNA and then to protein.

The host cell, however, also adds a great deal of complexity. Biologists have invested years of work in computer models of *E. coli* and other single-celled organisms [see "Cybernetic Cells," SCIENTIFIC AMERICAN, August 2001]. And yet, acknowledges Ron Weiss of Princeton, "if you give me the DNA sequence of your genetic system, I can't tell you what the bacteria will do with it." Indeed, Endy recalls, "about half of the 60 parts we designed in 2003 initially couldn't be synthesized because they killed the cells that were copying them. We had to figure out a way to lower the burden that carrying and replicating the engineered DNA imposed on the cells." (Eventually 58 of the 60 parts were produced successfully.)

One way to deal with the complexity added by the cells' native genome is to dodge it: the genetic device can be sequestered on its own loop of DNA, separate from the chromosome of the organism. Physical separation is only half the solution, however, because there are no wires in cells. Life runs on "wetware," with many protein signals simply floating randomly from one part to another. "So if I have one inverter over here made out of proteins and DNA," Endy explains, "a protein signal meant for that part will also act on any other instance of that inverter anywhere else in the cell," whether it lies on the artificial loop or on the natural chromosome.

One way to prevent crossed signals is to avoid using the same part twice. Weiss has taken this approach in constructing a "Goldilocks" genetic circuit, one that lights up when a target chemical is present but only when the concentration is not too high and not too low [*see illustration on opposite page*]. Tucked inside its various parts are four inverters, each of which responds to a different protein signal. But this strategy makes it much more difficult to design parts that are truly interchangeable and can be rearranged.

Endy is testing a solution that may be better for some systems. "Our inverter uses the same components [as one of Weiss's], just arranged differently," Endy says. "Now the input is not a protein but rather a rate, specifically the rate at which a gene is transcribed. The inverter responds to how many messenger RNAs are produced per second. It makes a protein, and that protein determines the rate of transcription going out [by switching on a second gene]. So I send in TIPS—transcription events per second—and as output, I get TIPS. That is the common currency, like a current in an electrical circuit." In principle, the inverter could be removed and replaced with any other BioBrick that processes TIPS. And TIPS signals are location-specific, so the same part can be used at several places in a circuit without interference.

The TIPS technique will be tested by a new set of genetic systems designed by students who took a winter course at M.I.T. this past January. The aim this year was to reprogram cells to work cooperatively to form patterns, such as polka dots, in a petri dish. To do this the cells must communicate with one another by secreting and sensing chemical nutrients.

"This year's systems were about twice the size of the 2003 projects," Endy says. It took 13 months to get the blinking *E. coli* designs built and into cells. But in the intervening year the inventory of BioBricks has grown, the speed of DNA synthesis has shot up, and the engineers have gained experience assembling genetic circuits. So Endy expects to have the 2004 designs ready for testing in just five months, in time to show off at the first synthetic biology conference, scheduled for this June.

Rewriting the Book of Life

THE SCIENTISTS WHO ATTEND that conference will no doubt commiserate about the inherent difficulty of engineering a relatively puny stretch of DNA to work reliably within a cell that is constantly changing. Living machines reproduce, but as they do they mutate.

"Replication is far from perfect. We've built circuits and seen them mutate in half the cells within five hours," Weiss reports. "The larger the circuit is, the faster it tends to mutate." Weiss and Frances H. Arnold of the California Institute of Technology have evolved circuits with improved performance

SAM OGDEN

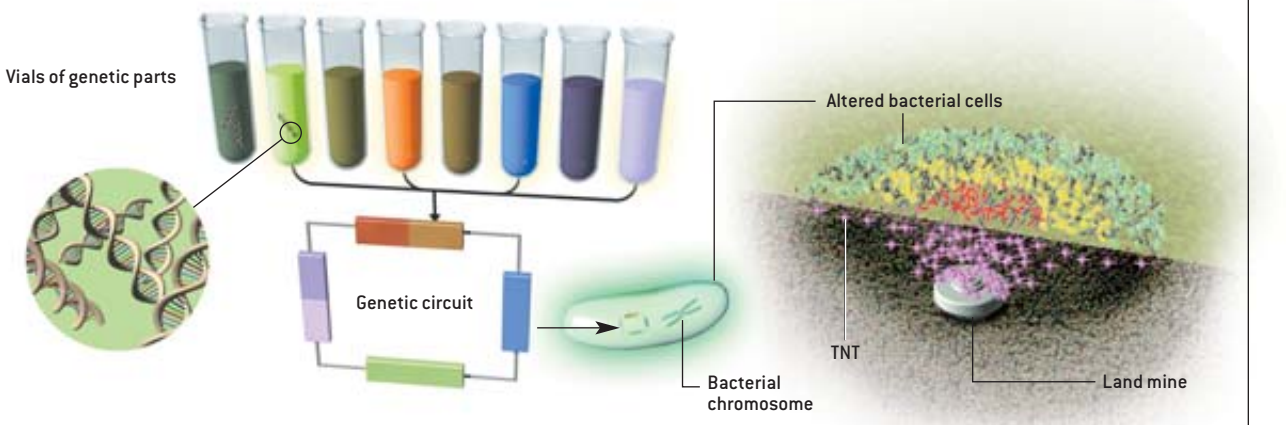
BUILDING A GENETIC MACHINE

A living TNT detector that reveals buried land mines could be made using genetic "Goldilocks" circuits that fluoresce only when the concentration of TNT is just right.

CONSTRUCTING A GENETIC TNT DETECTOR

Drawing from interchangeable DNA parts (*in test tubes*), engineers could assemble slightly different circuits. One would glow red, but only when the TNT concentration is high. A second might fluoresce yellow at medium levels of TNT, and a third could glow green at low concentrations.

Engineers would insert the circuits into three separate bacterial cultures. In the soil over a mine (*below*), TNT tapers off in a circular gradient. So a mixture of the altered cells would produce a fluorescent bull's-eye centered on the mine.



A TNT DETECTOR IN ACTION

One design for a Goldilocks genetic circuit uses four interacting parts: a sensor, upper and lower thresholds, and an inverter. Each part has a distinctive behavior: the amount of protein output it produces varies as a function of the amount of input protein it receives. In the schematic below for a red-glowing circuit, the graphs illustrate how each part adjusts its output over the full range of TNT concentrations. (The geometric shapes reflect output levels when the TNT concentration is in the "sweet spot" of, say, 4 percent.)

SENSOR

sends out two signals that are roughly proportional to the level of TNT within the cell. Importantly, the two signals are unequal: at a TNT level of 4 percent, one of the genes in the sensor (*dark purple*) churns out only half as much protein (*squares*) as does the other gene (*light purple*).

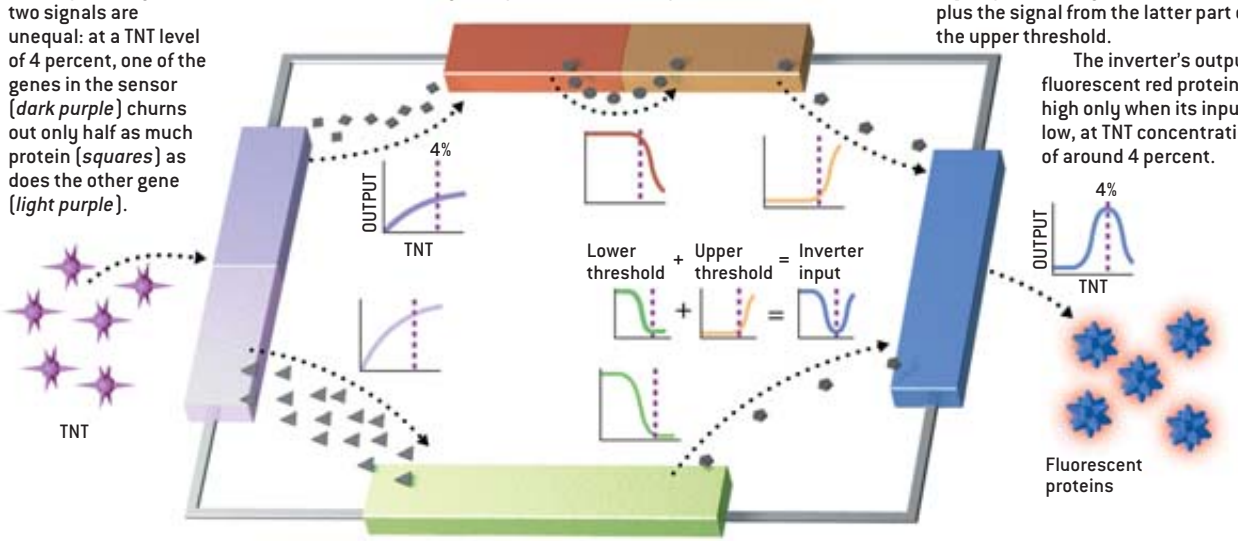
UPPER THRESHOLD

receives the weaker signal from the sensor. Output from the first gene in this part starts to fall dramatically but is still high when TNT levels are 4 percent. The next gene in the chain simply inverts whatever signal the first gene generates. So at 4 percent TNT concentration, the upper threshold sends very little protein to the next part (the inverter).

INVERTER

contains genes that express fluorescent proteins only when input signals from both thresholds are low. Its input (*pentagons*) is the sum of the protein signal produced by the lower threshold plus the signal from the latter part of the upper threshold.

The inverter's output—a fluorescent red protein—is high only when its input is low, at TNT concentrations of around 4 percent.



LOWER THRESHOLD

emits the inverse of its input signal (*triangles*), which is the protein that the sensor produces most prolifically. This part's output begins to fall steeply at TNT levels around 1 percent; by 4 percent TNT, the part produces almost no protein to send to the inverter.

BRYAN CHRISTIE DESIGN

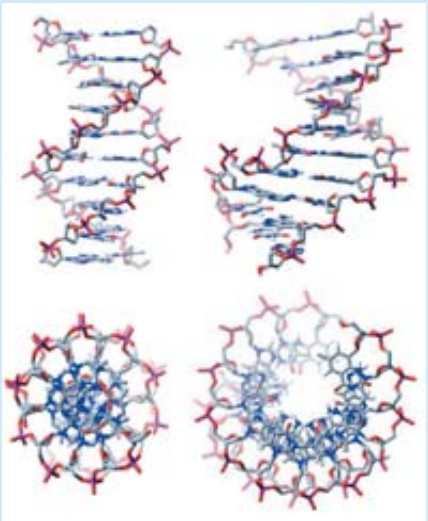
Life, but Not (Exactly) as We Know It

Life on earth has taken a tremendous range of forms, but all species arise from the same molecular ingredients: five nucleotides that form the building blocks for DNA and RNA, and 20 amino acids that serve as building blocks for proteins. (At least two additional natural amino acids are made by a few odd species.) These ingredients limit the chemical reactions that can happen inside cells and so constrain what life can do.

That constraint was eased in 2001, probably for the first time in more than three billion years. After years of trying, Lei Wang, Peter G. Schultz and their co-workers at the Scripps Research Institute in La Jolla, Calif., at last succeeded in adding to *Escherichia coli* bacteria all the genetic components the cells need to decode the three-nucleotide DNA sequence TAG into unnatural amino acids of various kinds.

It was a seminal step but an intermediate one, because amino acids by themselves have relatively few uses. The real goal is to modify cells so that they not only synthesize artificial amino acids but also string them together with natural amino acids to form proteins that no other form of life can make. Last year Schultz's group announced that it had done just that with *E. coli,* and in August the team reported its creation of a similarly talented form of yeast.

"The translational machinery [that reads RNA to make proteins] in yeast is very similar to the translational machinery of humans," points out T. Ashton Cropp, a biologist in Schultz's lab. "So far we have produced six kinds of unnatural amino acids in yeast," and the scientists have begun adapting the systems to work in human kidney cells and in roundworms, Cropp says. "We're very close to having a system that can make two different unnatural amino acids and put them in the same protein," he adds. "It is tricky because in order to do that, the cell has to decode a four-nucleotide DNA sequence," which, as far as anyone knows, no cell has ever done.



TWISTED LADDER OF DNA (*above left, seen in side view and top view*) may not be the only macromolecule capable of storing the blueprints for living organisms. Scientists are experimenting with semiartificial nucleic acids, such as xDNA (*at right*), that are more stable and thus less likely to suffer mutations.

"This advance could foster developments with inestimable biomedical potential," suggests Brian L. Davis of the Research Foundation of Southern California in La Jolla. He envisions white blood cells that could make novel proteins to destroy pathogens or cancerous cells more quickly. Cropp says the technology is already producing new research tools, such as proteins that include fluorescent amino acids or that change behavior when they are exposed to light. "It allows us to attach polymers to therapeutic proteins, which makes them work better as drugs," Cropp notes.

Synthetic biologists also have been avidly tinkering with unnatural forms of DNA. Steven A. Benner and his associates at the University of Florida developed a six-letter genetic alphabet more than a decade ago; it was recently used to create a rapid test for the SARS virus. "We're playing around with a variant called TNA, where ribose is replaced with a slightly simpler sugar," says Jack W. Szostak of Massachusetts General Hospital. TNA and xDNA, created by Eric T. Kool of Stanford University, are more stable than DNA. That may make them better suited as a medium for reprogramming cells. First, however, scientists will have to get them working inside living organisms. —*W.W.G.*

using multiple rounds of mutation followed by selection of those cells most fit for the desired task. But left unsupervised, evolution will tend to break genetic machines.

"I would like to make a genetically encoded device that accepts an input signal and simply counts: 1, 2, 3, … up to 256," Endy suggests. "That's not much more complex than what we're building now, and it would allow you to quickly and precisely detect certain types of cells that had lost control of their reproduction and gone cancerous. But how do I design a counter so that the design persists when the machine makes copies of itself that contain mistakes? I don't have a clue. Maybe we have to build in redundancy—or maybe we need to make the function of the counter somehow good for the cell."

Or perhaps the engineers will have to understand better how simple forms of life, such as viruses, have solved the problem of persistence. Synthetic biology may help here, too. Last November, Hamilton O. Smith and J. Craig Venter announced that their group at the Institute for Biological Energy Alternatives had re-created a bacteriophage (a virus that infects bacteria) called phiX174 from scratch, in just two weeks. The synthetic virus, Venter said, has the same 5,386 base pairs of DNA as the natural form and is just as active.

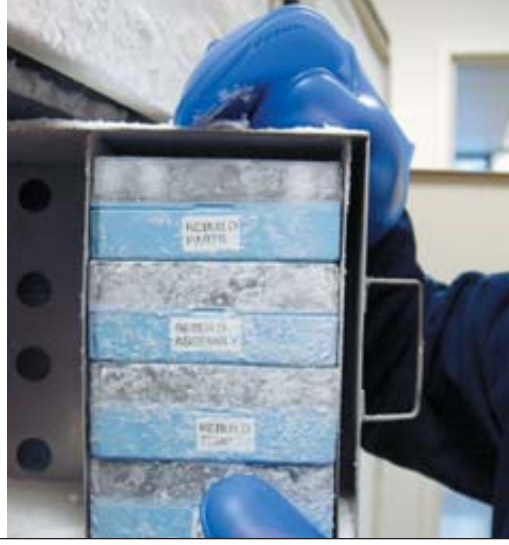
"Synthesis of a large chromosome is now clearly in reach," said Venter, who for several years led a project to identify the minimal set of genes required for survival by the bacterium *Mycoplasma genitalium*. "What we don't know is whether we can insert that chromosome into a cell and transform the cell's operating system to work off the new chromosome. We will have to understand life at its most basic level, and we're a long way from doing that."

Re-creating a virus letter-for-letter does not reveal much about it, but what if the genome were dissected into its constituent genes and then methodically put back together in a way that makes sense to human engineers? That is what Endy and colleagues are doing with the T7 bacteriophage. "We've rebuilt T7—not just resynthesized it but reengineered the genome and synthesized that," Endy reports. The scientists are separating genes that overlap, editing out redundancies, and so on. The group has completed about 11.5 kilobases so far and expects to finish the remaining 30,000 base pairs by the end of 2004.

REPRINTED WITH PERMISSION FROM H. LIU ET AL. IN *SCIENCE*, VOL. 302, PAGES 868–871; 2003. ©2003 AAAS

"The people in this class are happy and building nice, constructive things, as opposed to new species of virus or new kinds of bioweapons."

—Drew Endy, M.I.T.



Beta-Testing Life 2.0

SYNTHETIC BIOLOGISTS have so far built living genetic systems as experiments and demonstrations. But a number of research laboratories are already working on applications. Martin Fussenegger and his colleagues at ETH Zurich have graduated from bacteria to mammals. Last year they infused hamster cells with networks of genes that have a kind of volume control: adding small amounts of various antibiotics turned the output of the synthetic genes to low, medium or high. Controlling gene expression in this way could prove quite handy for gene therapies and the manufacture of pharmaceutical proteins.

Living machines will probably find their first uses for jobs that require sophisticated chemistry, such as detecting toxins or synthesizing drugs. Last year Homme W. Hellinga of Duke University invented a way to redesign natural sensor proteins in *E. coli* so that they would latch onto TNT or any other compound of interest instead of their normal targets. Weiss says that he and Hellinga have discussed combining his Goldilocks circuit with Hellinga's sensor to make land-mine detectors.

Jay Keasling, who recently founded a synthetic biology department at Lawrence Berkeley National Laboratory (LBNL), reports that his group has engineered a large network of wormwood and yeast genes into *E. coli*. The circuit enables the bacterium to fabricate a chemical precursor to artemisinin, a next-generation antimalarial drug that is currently too expensive for the parts of the developing world that need it most.

Keasling says that three years of work have increased yields by a factor of one million. By boosting the yields another 25- to 50-fold, he adds, "we will be able to produce artemisinin-based dual cocktail drugs to the Third World for about one tenth the current price." With relatively simple modifications, the bioengineered bacteria could be altered to produce expensive chemicals used in perfumes, flavorings and the cancer drug Taxol.

Other scientists at LBNL are using *E. coli* to help dispose of nuclear waste as well as biological and chemical weapons. One team is modifying the bacteria's sense of "smell" so that the bugs will swim toward a nerve agent, such as VX, and digest it. "We have engineered *E. coli* and *Pseudomonas aeruginosa* to precipitate heavy metals, uranium and plutonium on their cell wall," Keasling says. "Once the cells have accumulated the metals, they settle out of solution, leaving cleaned wastewater."

Worthy goals, all. But if you become a touch uneasy at the thought of undergraduates creating new kinds of germs, of private labs synthesizing viruses, and of scientists publishing papers on how to use bacteria to collect plutonium, you are not alone.

In 1975 leading biologists called for a moratorium on the use of recombinant-DNA technology and held a conference at the Asilomar Conference Grounds in California to discuss how to regulate its use. Self-policing seemed to work: there has yet to be a major accident with genetically engineered organisms. "But recently three things have changed the landscape," Endy points out. "First, anyone can now download the DNA sequence for anthrax toxin genes or for any number of bad things. Second, anyone can order synthetic DNA from offshore companies. And third, we are now more worried about intentional misapplication."

So how does society counter the risks of a new technology without also denying itself all the benefits? "The Internet stays up because there are more people who want to keep it running than there are people who want to bring it down," Endy suggests. He pulls out a photograph of the class he taught last year. "Look. The people in this class are happy and building nice, constructive things, as opposed to new species of virus or new kinds of bioweapons. Ultimately we deal with the risks of biological technology by creating a society that can use the technology constructively."

But he also believes that a meeting to address potential problems makes sense. "I think," he says, "it would be entirely appropriate to convene a meeting like Asilomar to discuss the current state and future of biological technology." This June, as leaders in the field meet to share their latest ideas about what can now be created, perhaps they will also devote some thought to what shouldn't. SA

W. Wayt Gibbs is senior writer.

MORE TO EXPLORE

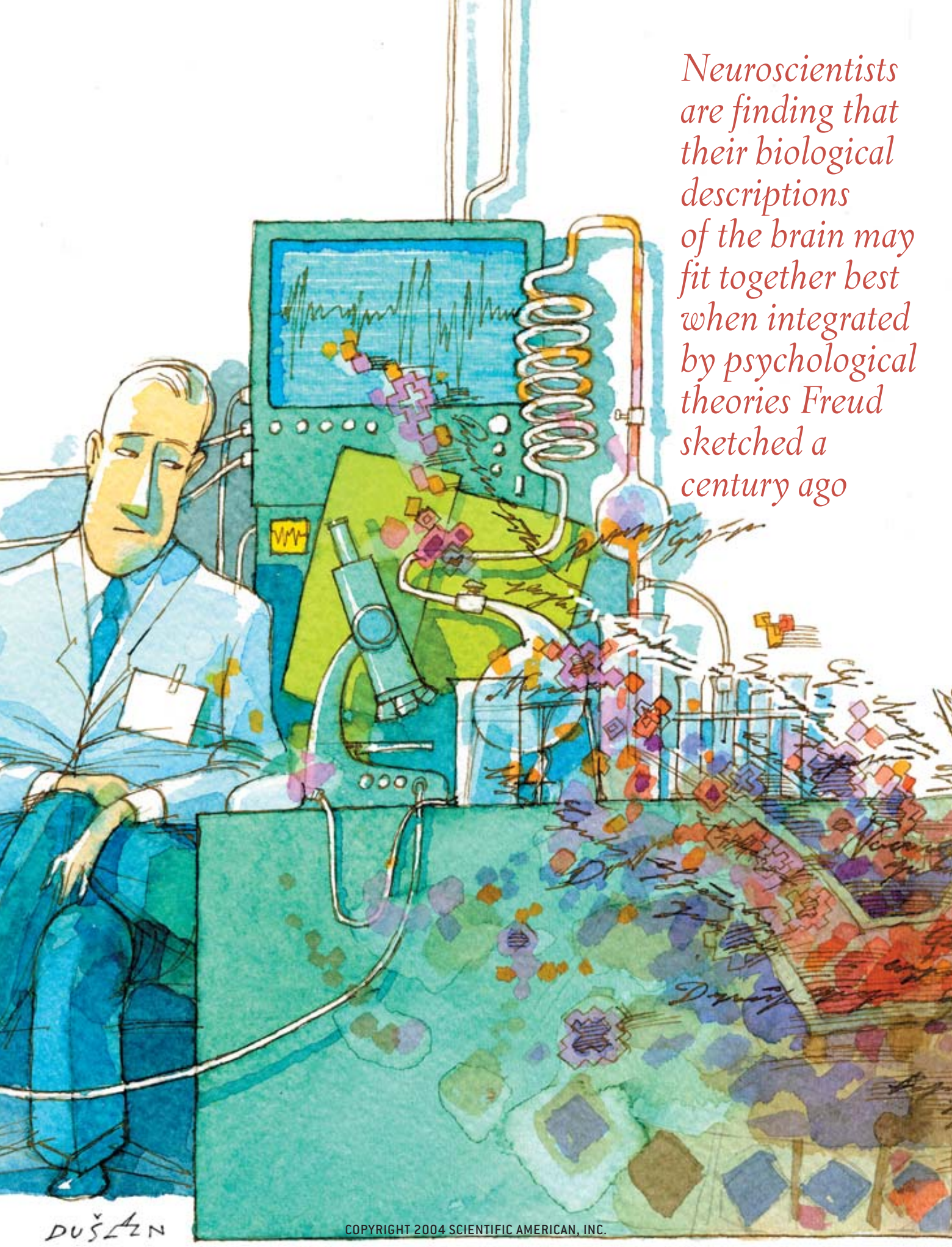
An Expanded Eukaryotic Genetic Code. Jason W. Chin et al. in *Science,* Vol. 301, pages 964–967; August 15, 2003.

Genetic Circuit Building Blocks for Cellular Computation, Communications, and Signal Processing. Ron Weiss et al. in *Natural Computing,* Vol. 2, No. 1, pages 47–84; 2003.

The M.I.T. Synthetic Biology Working Group: **syntheticbiology.org**

The M.I.T. Registry of Standard Biological Parts: **parts.mit.edu**

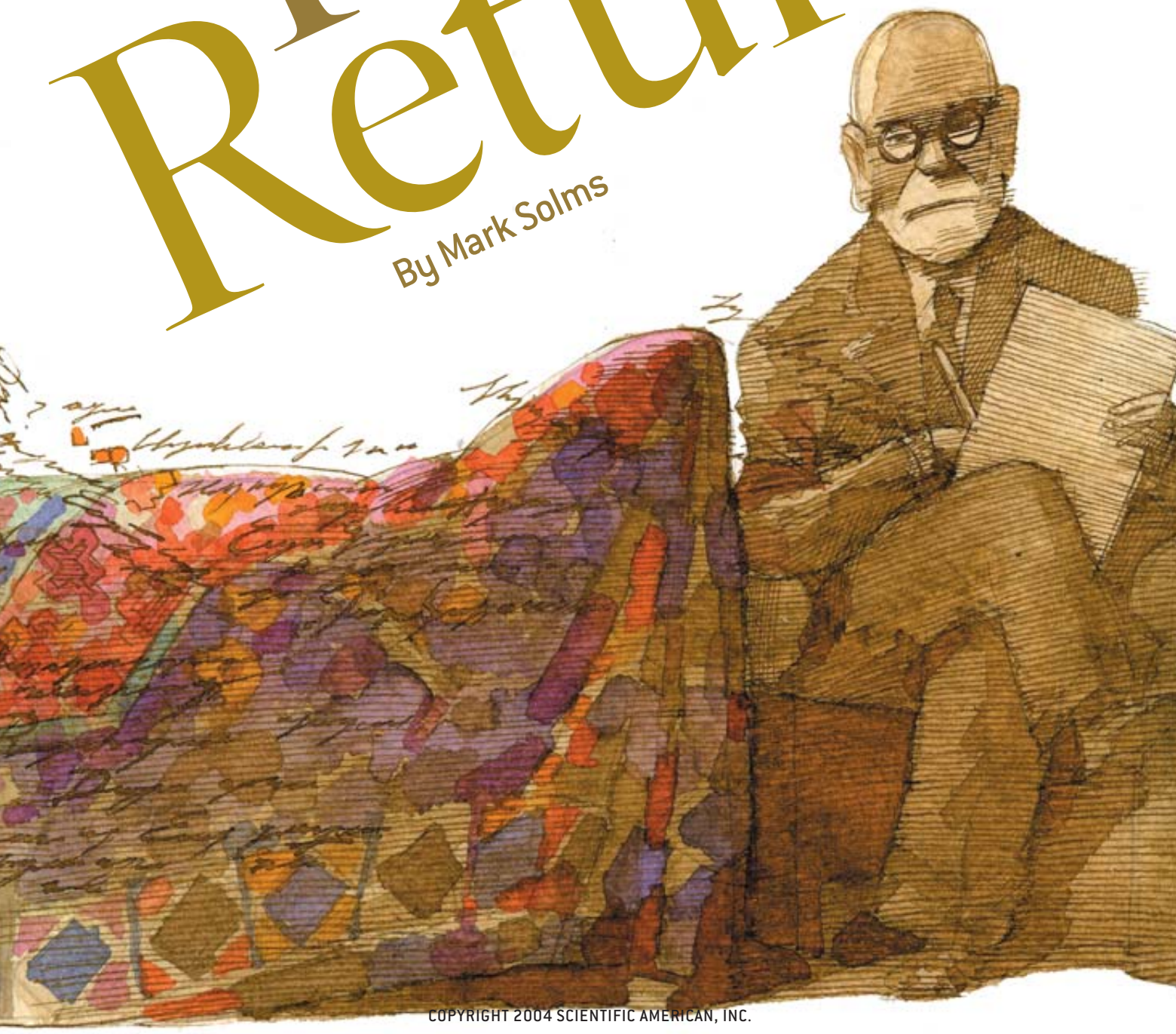
SAM OGDEN (*boxes of BioBricks*)



Neuroscientists are finding that their biological descriptions of the brain may fit together best when integrated by psychological theories Freud sketched a century ago

Freud Returns

By Mark Solms



YOUNG FREUD, circa 1891

FOR THE FIRST HALF OF THE 1900S, the ideas of Sigmund Freud dominated explanations of how the human mind works. His basic proposition was that our motivations remain largely hidden in our unconscious minds. Moreover, they are actively withheld from consciousness by a repressive force. The executive apparatus of the mind (the ego) rejects any unconscious drives (the id) that might prompt behavior that would be incompatible with our civilized conception of ourselves. This repression is necessary because the drives express themselves in unconstrained passions, childish fantasies, and sexual and aggressive urges.

Mental illness, Freud said until his death in 1939, results when repression fails. Phobias, panic attacks and obsessions are caused by intrusions of the hidden drives into voluntary behavior. The aim of psychotherapy, then, was to trace neurotic symptoms back to their unconscious roots and expose these roots to mature, rational judgment, thereby depriving them of their compulsive power.

As mind and brain research grew more sophisticated from the 1950s onward, however, it became apparent to specialists that the evidence Freud had provided for his theories was rather tenuous. His principal method of investigation was not controlled experimentation but simple observations of patients in clinical settings, interwoven with theoretical inferences. Drug treatments gained ground, and biological approaches to mental illness gradually overshadowed psychoanalysis. Had Freud lived, he might even have welcomed this turn of events. A highly regarded neuroscientist in his day, he frequently made remarks such as "the deficiencies in our description would presumably vanish if we were already in a position to replace the psychological terms by physiological and chemical ones." But Freud did not have the science or technology to know how the brain of a normal or neurotic personality was organized.

Overview/*Mind Models*

- For decades, Freudian concepts such as ego, id and repressed desires dominated psychology and psychiatry's attempts to cure mental illnesses. But better understanding of brain chemistry gradually replaced this model with a biological explanation of how the mind arises from neuronal activity.
- The latest attempts to piece together diverse neurological findings, however, are leading to a chemical framework of the mind that validates the general sketch Freud made almost a century ago. A growing group of scientists are eager to reconcile neurology and psychiatry into a unified theory.

By the 1980s the notions of ego and id were considered hopelessly antiquated, even in some psychoanalytical circles. Freud was history. In the new psychology, the updated thinking went, depressed people do not feel so wretched because something has undermined their earliest attachments in infancy—rather their brain chemicals are unbalanced. Psychopharmacology, however, did not deliver an alternative grand theory of personality, emotion and motivation—a new conception of "what makes us tick." Without this model, neuroscientists focused their work narrowly and left the big picture alone.

Today that picture is coming back into focus, and the surprise is this: it is not unlike the one that Freud outlined a century ago. We are still far from a consensus, but an increasing number of diverse neuroscientists are reaching the same conclusion drawn by Eric R. Kandel of Columbia University, the 2000 Nobel laureate in physiology or medicine: that psychoanalysis is "still the most coherent and intellectually satisfying view of the mind."

Freud is back, and not just in theory. Interdisciplinary work groups uniting the previously divided and often antagonistic fields of neuroscience and psychoanalysis have been formed in almost every major city of the world. These networks, in turn, have come together as the International Neuro-Psychoanalysis Society, which organizes an annual congress and publishes the successful journal *Neuro-Psychoanalysis*. Testament to the renewed respect for Freud's ideas is the journal's editorial advisory board, populated by a who's who of experts in contemporary behavioral neuroscience, including Antonio R. Damasio, Kandel, Joseph E. LeDoux, Benjamin Libet, Jaak Panksepp, Vilayanur S. Ramachandran, Daniel L. Schacter and Wolf Singer.

Together these researchers are forging what Kandel calls a "new intellectual framework for psychiatry." Within this framework, it appears that Freud's broad brushstroke organization of the mind is destined to play a role similar to the one Darwin's theory of evolution

DUSAN PETRICIC (*preceding pages*); BETTMANN/CORBIS

served for molecular genetics—a template on which emerging details can be coherently arranged. At the same time, neuroscientists are uncovering proof for some of Freud's theories and are teasing out the mechanisms behind the mental processes he described.

Unconscious Motivation

WHEN FREUD INTRODUCED the central notion that most mental processes that determine our everyday thoughts, feelings and volitions occur unconsciously, his contemporaries rejected it as impossible. But today's findings are confirming the existence and pivotal role of unconscious mental processing. For example, the behavior of patients who are unable to consciously remember events that occurred after damage to certain memory-encoding structures of their brains is clearly influenced by the "forgotten" events. Cognitive neuroscientists make sense of such cases by delineating different memory systems that process information "explicitly" (consciously) and "implicitly" (unconsciously). Freud split memory along just these lines.

Neuroscientists have also identified unconscious memory systems that mediate emotional learning. In 1996 at New York University, LeDoux demonstrated the existence under the conscious cortex of a neuronal pathway that connects perceptual information with the primitive brain structures responsible for generating fear responses. Because this pathway bypasses the hippocampus—which generates conscious memories—current events routinely trigger unconscious remembrances of emotionally important past events, causing conscious feelings that seem irrational, such as "Men with beards make me uneasy."

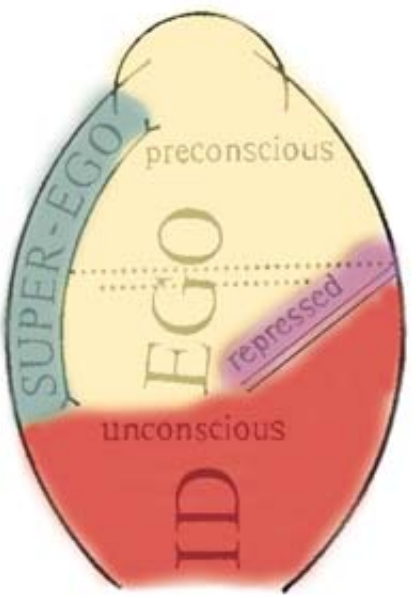
Neuroscience has shown that the major brain structures essential for forming conscious (explicit) memories are not functional during the first two years of life, providing an elegant explanation of what Freud called infantile amnesia. As Freud surmised, it is not that we forget our earliest memories; we simply cannot recall them to consciousness. But this inability does not preclude them from affecting adult feelings and behavior. One would be hard-pressed to find a developmental neurobiologist who does not agree that early experiences, especially between mother and infant, influence the pattern of brain connections in ways that fundamentally shape our future personality and mental health. Yet none of these experiences can be consciously remembered. It is becoming increasingly clear that a good deal of our mental activity is unconsciously motivated.

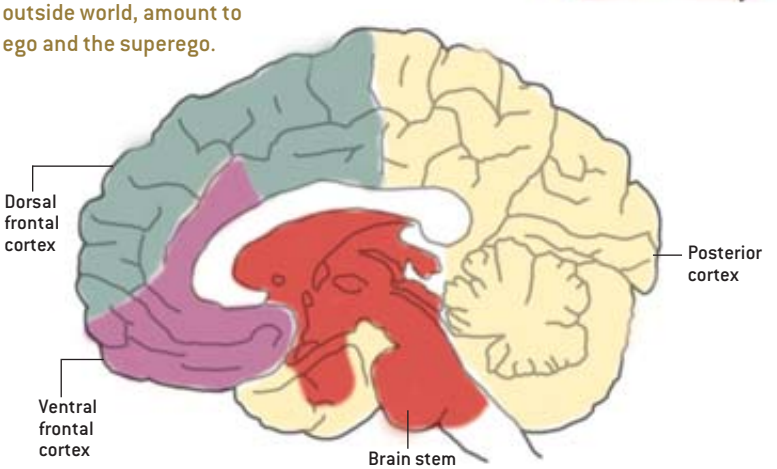
Repression Vindicated

EVEN IF WE ARE MOSTLY driven by unconscious thoughts, this does not prove anything about Freud's claim that we actively repress unpalatable information. But case studies supporting that notion are beginning to accumulate. The most famous one comes from a 1994 study of "anosognosic" patients by behavioral neurologist Ramachandran of the University of California at San Diego. Damage to the right parietal region of these people's brains makes them unaware of gross physical defects, such as paralysis of a limb. After artificially activating the right hemisphere of one such patient, Ramachandran observed that she suddenly became aware that her left arm was paralyzed—and that it had been paralyzed continuously since she had suffered a stroke eight days before. This showed that she was capable of recog-

MIND AND MATTER

Freud drew his final model of the mind in 1933 (*right; color has been added*). Dotted lines represented the threshold between unconscious and conscious processing. The superego repressed instinctual drives (the id), preventing them from disrupting rational thought. Most rational (ego) processes were automatic and unconscious, too, so only a small part of the ego (*bulb at top*) was left to manage conscious experience, which was closely tied to perception. The superego mediated the ongoing struggle between the ego and id for dominance. Recent neurological mapping (*below*) generally correlates to Freud's conception. The core brain stem and limbic system—responsible for instincts and drives—roughly correspond to Freud's id. The ventral frontal region, which controls selective inhibition, the dorsal frontal region, which controls self-conscious thought, and the posterior cortex, which represents the outside world, amount to the ego and the superego.





A. W. FREUD ET AL., BY ARRANGEMENT WITH PATERSON MARSH LTD., LONDON (*top*); OLIVER TURNBULL (*bottom; all coloring*)

nizing her deficits and that she had unconsciously registered these deficits for the previous eight days, despite her conscious denials during that time that there was any problem.

Significantly, after the effects of the stimulation wore off, the woman not only reverted to the belief that her arm was normal, she also forgot the part of the interview in which she had acknowledged that the arm was paralyzed, even though she remembered every other detail about the interview. Ramachandran concluded: "The remarkable theoretical implication of these observations is that memories can indeed be selectively repressed.... Seeing [this patient] convinced me, for the first time, of the reality of the repression phenomena that form the cornerstone of classical psychoanalytical theory."

Like "split-brain" patients, whose hemispheres become unlinked—made famous in studies by the late Nobel laureate Roger W. Sperry of the California Institute of Technology in the 1960s and 1970s—anosognosic patients typically rationalize away unwelcome facts, giving plausible but invented explanations of their unconsciously motivated actions. In this way, Ramachandran says, the left hemisphere manifestly employs Freudian "mechanisms of defense."

Durham neuropsychologist Aikaterini Fotopoulou recently studied a patient of this type in my laboratory. The man failed to recall, in each 50-minute session held in my office on 12 consecutive days, that he had ever met me before or that he had undergone an operation to remove a tumor in his frontal lobes that caused his amnesia. As far as he was concerned, there was nothing wrong with him. When asked about the scar on his head, he confabulated wholly implausible explanations: he had undergone dental surgery or a coronary bypass operation. In reality, he had indeed experienced these procedures—years before—and unlike his brain operation, they had successful outcomes.

Similarly, when asked who I was and what he was doing in my lab, he variously said that I was a colleague, a drinking partner, a client consulting him about his area of professional expertise, a teammate in a sport that he had not participated in since he was in college decades earlier, or a mechanic repairing one of his numerous sports cars (which he did not possess). His behavior was consistent with these false beliefs, too: he would look around the room for his beer or out the window for his car.

What strikes the casual observer is the wishful quality of these false notions, an impression that Fotopoulou confirmed objectively through quantitative analysis of a consecutive series of 155 of his confabulations. The patient's false beliefs were not random noise—they were generated by the "pleasure principle" that Freud maintained was central to unconscious thought. The man simply recast reality as he wanted it to be. Similar observations have been reported by others, such as Martin Conway of Durham and Oliver Turnbull of the University of Wales. These investigators are cognitive neuroscientists, not psychoanalysts, yet they interpret their findings in Freudian

Freud himself anticipated the day when neurological data would round out his psychological ideas.

BRAIN SCANS show the damage that causes disorders of psychological function, which Freud could study only clinically. A recent MRI image of a patient who confabulates grandiose stories of his life reveals a lesion (*arrow*) in the cingulate gyrus—part of the medial frontal lobe that serves functions Freud posited would normally prevent unconscious wishes from altering a person's rational self-image.

COURTESY OF OLIVER TURNBULL

Analogous phenomena have now been demonstrated in people with intact brains, too. As neuropsychologist Martin A. Conway of Durham University in England pointed out in a 2001 commentary in *Nature,* if significant repression effects can be generated in average people in an innocuous laboratory setting, then far greater effects are likely in real-life traumatic situations.

The Pleasure Principle

FREUD WENT EVEN further, though. He said that not only is much of our mental life unconscious and withheld but that the repressed part of the unconscious mind operates according to a different principle than the "reality principle" that governs the conscious ego. This type of unconscious thinking is "wishful"—and it blithely disregards the rules of logic and the arrow of time.

If Freud was right, then damage to the inhibitory structures of the brain (the seat of the "repressing" ego) should release wishful, irrational modes of mental functioning. This is precisely what has been observed in patients with damage to the frontal limbic region, which controls critical aspects of self-awareness. Subjects display a striking syndrome known as Korsakoff's psychosis: they are unaware that they are amnesic and therefore fill the gaps in their memory with fabricated stories known as confabulations.

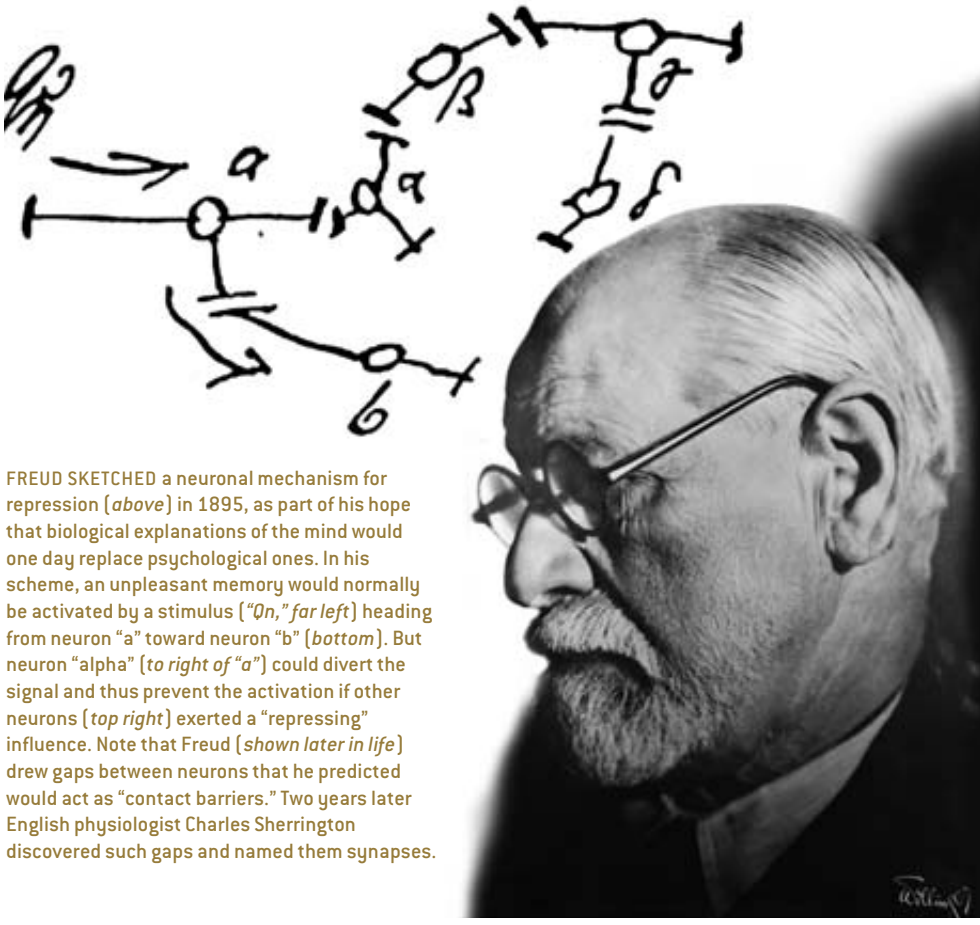
terms, claiming in essence that damage to the frontal limbic region that produces confabulations impairs cognitive control mechanisms that underpin normal reality monitoring and releases from inhibition the implicit wishful influences on perception, memory and judgment.

Animal Within

FREUD ARGUED THAT the pleasure principle gave expression to primitive, animal drives. To his Victorian contemporaries, the implication that human behavior was at bottom governed by urges that served no higher purpose than carnal self-fulfillment was downright scandalous. The moral outrage waned during subsequent decades, but Freud's concept of man-as-animal was pretty much sidelined by cognitive scientists.

Now it has returned. Neuroscientists such as Donald W. Pfaff of the Rockefeller University and Jaak Panksepp of Bowling Green State University believe that the instinctual mechanisms that govern human motivation are even more primitive than Freud imagined. We share basic emotional-control systems with our primate relatives and with all mammals. At the deep level of mental organization that Freud called the id, the functional anatomy and chemistry of our brains is not much different from that of our favorite barnyard animals and household pets.

Modern neuroscientists do not accept Freud's classification of human instinctual life as a simple dichotomy between sexuality and aggression, however. Instead, through studies of lesions and the effects of drugs and artificial stimulation on the brain, they have identified at least four basic mammalian instinctual circuits, some of which overlap. They are the "seeking" or "reward" system (which motivates the pursuit of pleasure); the "anger-rage" system (which governs angry aggression but not predatory aggression); the "fear-anxiety" system; and the "panic" system (which includes complex instincts such as those that govern social bonding). Whether other instinctual forces exist, such as a rough-and-tumble "play" system, is also being investigated. All these brain systems are modulated by specific neurotransmitters, chemicals that carry messages between the brain's neurons.



FREUD SKETCHED a neuronal mechanism for repression (*above*) in 1895, as part of his hope that biological explanations of the mind would one day replace psychological ones. In his scheme, an unpleasant memory would normally be activated by a stimulus (*"On," far left*) heading from neuron "a" toward neuron "b" (*bottom*). But neuron "alpha" (*to right of "a"*) could divert the signal and thus prevent the activation if other neurons (*top right*) exerted a "repressing" influence. Note that Freud (*shown later in life*) drew gaps between neurons that he predicted would act as "contact barriers." Two years later English physiologist Charles Sherrington discovered such gaps and named them synapses.

The seeking system, regulated by the neurotransmitter dopamine, bears a remarkable resemblance to the Freudian "libido." According to Freud, the libidinal or sexual drive is a pleasure-seeking system that energizes most of our goal-directed interactions with the world. Modern research shows that its neural equivalent is heavily implicated in almost all forms of craving and addiction. It is interesting to note that Freud's early experiments with cocaine—mainly on himself—convinced him that the libido must have a specific neurochemical foundation. Unlike his successors, Freud saw no reason for antagonism between psychoanalysis and psychopharmacology. He enthusiastically anticipated the day when "id energies" would be controlled directly by "particular chemical substances." Today treatments that integrate psychotherapy with psychoactive medications are widely recognized as the best approach for many disorders. And brain imaging shows that talk therapy affects the brain in similar ways to such drugs.

Dreams Have Meaning

FREUD'S IDEAS ARE also reawakening in sleep and dream science. His dream theory—that nighttime visions are partial glimpses of unconscious wishes—was discredited when rapid-eye-movement (REM) sleep and its strong correlation with dreaming were discovered in the 1950s. Freud's view appeared to lose all credibility when investigators in the 1970s showed that the dream cycle was regulated by the pervasive brain chemi-

THE AUTHOR

MARK SOLMS holds the chair in neuropsychology at the University of Cape Town in South Africa and an honorary lectureship in neurosurgery at St. Bartholomew's and the Royal London School of Medicine and Dentistry. He is also director of the Arnold Pfeffer Center for Neuro-Psychoanalysis of the New York Psychoanalytic Institute, a consultant neuropsychologist to the Anna Freud Center in London and a very frequent flier. Solms is editor and translator of the forthcoming four-volume series *The Complete Neuroscientific Works of Sigmund Freud* (Karnac Books). Solms thanks Oliver Turnbull, a senior lecturer at the University of Wales Center for Cognitive Neuroscience in Bangor, for assisting with this article.

BETTMANN/CORBIS (*photograph of Freud circa 1930*); A. W. FREUD ET AL., BY ARRANGEMENT WITH PATERSON MARSH LTD., LONDON (*drawing*)

cal acetylcholine, produced in a "mindless" part of the brain stem. REM sleep occurred automatically, every 90 minutes or so, and was driven by brain chemicals and structures that had nothing to do with emotion or motivation. This discovery implied that dreams had no meaning; they were simply stories concocted by the higher brain to try to reflect the random cortical activity caused by REM.

But more recent work has revealed that dreaming and REM sleep are dissociable states, controlled by distinct, though interactive, mechanisms. Dreaming turns out to be generated by a network of structures centered on the forebrain's instinctual-motivational circuitry. This discovery has given rise to a host of theories about the dreaming brain, many strongly reminiscent of Freud's. Most intriguing is the observation that others and I have made that dreaming stops completely when certain fibers deep in the frontal lobe have been severed—a symptom that coincides with a general reduction in motivated behavior. The lesion is exactly the same as the damage that was deliberately produced in prefrontal leukotomy, an outmoded surgical procedure that was once used to control hallucinations and delusions. This operation was replaced in the 1960s by drugs that dampen dopamine's activity in the same brain systems. The seeking system, then, might be the primary generator of dreams. This possibility has become a major focus of current research.

If the hypothesis is confirmed, then the wish-fulfillment theory of dreams could once again set the agenda for sleep research. But even if other interpretations of the new neurological data prevail, all of them demonstrate that "psychological" conceptualizations of dreaming are scientifically respectable again. Few neuroscientists still claim—as they once did with impunity—that dream content has no primary emotional mechanism.

Finishing the Job

NOT EVERYONE IS enthusiastic about the reappearance of Freudian concepts in the mainstream of mental science. It is not easy for the older generation of psychoanalysts, for example, to accept that their junior colleagues and students now can and must subject conventional wisdom to an entirely new level of biological scrutiny. But an encouraging number of elders on both sides of the Atlantic are at least committed to keeping an open mind, as evidenced by the aforementioned eminent psychoanalysts on the advisory board of *Neuro-Psychoanalysis* and by the many graying participants in the International Neuro-Psychoanalysis Society.

For older neuroscientists, resistance to the return of psychoanalytical ideas comes from the specter of the seemingly indestructible edifice of Freudian theory in the early years of their careers. They cannot acknowledge even partial confirmation of Freud's fundamental insights; they demand a complete purge [*see box on opposite page*]. In the words of J. Allan Hobson, a renowned sleep researcher and Harvard Medical School psychiatrist, the renewed interest in Freud is little more than unhelpful "retrofitting" of modern data into an antiquated theoretical framework. But as Panksepp said in a 2002 interview with *Newsweek* magazine, for neuroscientists who are enthusiastic about the reconciliation of neurology and psychiatry, "it is not a matter of proving Freud right or wrong, but of finishing the job."

If that job can be finished—if Kandel's "new intellectual framework for psychiatry" can be established—then the time will pass when people with emotional difficulties have to choose between the talk therapy of psychoanalysis, which may be out of touch with modern evidence-based medicine, and the drugs prescribed by psychopharmacology, which may lack regard for the relation between the brain chemistries it manipulates and the complex real-life trajectories that culminate in emotional distress. The psychiatry of tomorrow promises to provide patients with help that is grounded in a deeply integrated understanding of how the human mind operates.

Whatever undreamed-of therapies the future might bring, patients can only benefit from better knowledge of how the brain really works. As modern neuroscientists tackle once more the profound questions of human psychology that so preoccupied Freud, it is gratifying to find that we can build on the foundations he laid, instead of having to start all over again. Even as we identify the weak points in Freud's far-reaching theories, and thereby correct, revise and supplement his work, we are excited to have the privilege of finishing the job. SA

If scientists can reconcile neurology and psychology, then patients could receive more integrated treatment.

MORE TO EXPLORE

The Neuropsychology of Dreams. Mark Solms. Lawrence Erlbaum Associates, 1997.

Dreaming and REM Sleep Are Controlled by Different Brain Mechanisms. Mark Solms in *Behavioral and Brain Sciences,* Vol. 23, No. 6, pages 843–850; December 2000.

Freudian Dream Theory Today. Mark Solms in *Psychologist,* Vol. 13, No. 12, pages 618–619; December 2000.

Clinical Studies in Neuro-Psychoanalysis. K. Kaplan-Solms and M. Solms. Karnac Books, 2000.

The Brain and the Inner World. Mark Solms and Oliver Turnbull. Other Press, 2002.

The International Neuro-Psychoanalysis Society and the journal *Neuro-Psychoanalysis:* **www.neuro-psa.org**

FREUD RETURNS? LIKE A BAD DREAM

By J. Allan Hobson

Sigmund Freud's views on the meaning of dreams formed the core of his theory of mental functioning. Mark Solms and others assert that brain imaging and lesion studies are now validating Freud's conception of the mind. But similar scientific investigations show that major aspects of Freud's thinking are probably erroneous.

For Freud, the bizarre nature of dreams resulted from an elaborate effort of the mind to conceal, by symbolic disguise and censorship, the unacceptable instinctual wishes welling up from the unconscious when the ego relaxes its prohibition of the id in sleep. But most neurobiological evidence supports the alternative view that dream bizarreness stems from normal changes in brain state. Chemical mechanisms in the brain stem, which shift the activation of various regions of the cortex, generate these changes. Many studies have indicated that the chemical changes determine the quality and quantity of dream visions, emotions and thoughts. Freud's disguise-and-censorship notion must be discarded; no one believes that the ego-id struggle, if it exists, controls brain chemistry. Most psychoanalysts no longer hold that the disguise-censorship theory is valid.

Without disguise and censorship, what is left of Freud's dream theory? Not much—only that instinctual drives could impel dream formation. Evidence does indicate that activating the parts of the limbic system that produce anxiety, anger and elation shapes dreams. But these influences are not "wishes." Dream analyses show that the emotions in dreams are as often negative as they are positive, which would mean that half our "wishes" for ourselves are negative. And as all dreamers know, the emotions in dreams are hardly disguised. They enter into dream plots clearly, frequently bringing unpleasant effects such as nightmares. Freud was never able to account for why so many dream emotions are negative.

Another pillar of Freud's model is that because the true meaning of dreams is hidden, the emotions they reflect can be revealed only through his wild-goose-chase method of free association, in which the subject relates anything and everything that comes to mind in hopes of stumbling across a crucial connection. But this effort is unnecessary, because no such concealment occurs. In dreams, what you see is what you get. Dream content is emotionally salient on its face, and the close attention of dreamers and their therapists is all that is needed to see the feelings they represent.

Solms and other Freudians intimate that ascribing dreams to brain chemistry is the same as saying that dreams have no emotional messages. But the statements are not equivalent. The chemical activation-synthesis theory of dreaming, put forth by Robert W. McCarley of Harvard Medical School and me in 1977, maintained only that the psychoanalytic explanation of dream bizarreness as concealed meaning was wrong. We have always argued that dreams are emotionally salient and meaningful. And what about REM sleep? New studies reveal that dreams can occur during non-REM sleep, but nothing in the chemical activation model precludes this case; the frequency of dreams is simply exponentially higher during REM sleep.

Psychoanalysis is in big trouble, and no amount of neurobiological tinkering can fix it. So radical an overhaul is necessary that many neuroscientists would prefer to start over and create a neurocognitive model of the mind. Psychoanalytic theory is indeed comprehensive, but if it is terribly in error, then its comprehensiveness is hardly a virtue. The scientists who share this view stump for more biologically based models of dreams, of mental illness, and of normal conscious experience than those offered by psychoanalysis.

J. Allan Hobson, professor of psychiatry at Harvard Medical School, has written extensively on the brain basis of the mind and its implications for psychiatry. For more, see Hobson's book Dreaming: An Introduction to the Science of Sleep *(Oxford University Press, 2003).*

DUSAN PETRICIC

RETOOLING

the Global Positioning System

From hikers navigating with handheld locators to pilots landing in zero-visibility conditions, the Global Positioning System now serves more than 30 million users. See what's coming next

By Per Enge

During the next decade, the capabilities of the Global Positioning System (GPS) will take off. Not only will advanced GPS technology ensure far greater reliability and safety than is possible today, it also will provide much more accurate geolocation services: to within a meter. Underlying the improved capabilities is a series of system upgrades that include additional satellite signals, increased broadcast power, performance monitoring, guaranteed error bounds, smart antennas that can selectively direct and receive signals, and integration with television and cellular-phone networks.

When next-generation GPS becomes available, it will enable a broad range of exciting new applications. Geolocation coverage will extend from hiking trails and sea-lanes all the way downtown, indoors and into areas that are currently plagued with weak reception, such as under tree limbs. Businesses operating in industries such as air, sea and land transport, electric power, telecommunications, construction, mining, mapping and farming are likely to profit from the augmented services. So will geographers and earth scientists. Military users should benefit the most, as was intended by the original builders. With its greater dependability, enhanced GPS could ensure that an airplane lands automatically in zero-visibility weather, for example, or that a U.S. naval aircraft lands safely on a pitching carrier deck in the dark. In years to come, it may even guarantee the security of passengers in cars and trucks riding down automated highways.

AS GPS GETS ever more pervasive, it will also become sufficiently dependable for safety-critical applications such as automated guidance for airplanes and cars.

Sky's the Limit

GPS GOT ITS START when the U.S. Department of Defense launched the first Navstar satellite in 1978. Although the designers expected that civil and commercial applications would develop, their prime goal at the time was to allow an estimated 40,000 military users to navigate the land, sea and sky with high precision. Civilians began taking advantage of the positioning system during the 1980s. As the orbital constellation of GPS satellites approached the minimum 24 needed for continuous service in the early 1990s, mass-market uses soon multiplied.

Today some 30 million people regularly track their whereabouts using GPS. The receiving units assist in guiding road vehicles, ships and boats, as well as in fleet management for rental cars and buses, and recreational uses [see "A Walk in the Woods," by Mark Clemens, Technicalities; SCIENTIFIC AMERICAN, February]. Every month vendors ship more than 200,000 civilian receivers. In 2003 GPS equipment sales reached nearly $3.5 billion worldwide, and that annual market could grow to $10 billion after 2010, according to a recent survey conducted by

Frost & Sullivan, a market research firm. Those figures do not count revenues from satellite construction, launch and control segments or from GPS-related enterprises such as delivery-truck fleet management. Consumers, the study says, now account for slightly more than half the equipment sales; commercial customers make up 40 percent, and the armed forces take up the remainder (8 percent).

The American GPS Navstar satellites are not alone in orbit, however. Russian GLONASS navigation satellites share that physical and functional space, and in a few years so will the European Galileo constellation. The Russians built GLONASS during the cold war to compete with the U.S. military. Of late, though, GLONASS has fallen into disuse because its operators cannot afford to replenish satellites. The European Community's system is expected to enter service later this decade. The attraction is an anticipated booming end-user market, which will only grow as GPS receivers are added to automobiles and cell phones. Both the Europeans and Russians believe that they need their own satellite navigation systems to participate. The GPS and the Galileo management teams have recently concluded agreements on how the systems will interact.

Each time a GPS receiver locates itself on the planet's surface, it trilaterates (a cousin of triangulation) its precise distance from at least four GPS satellites using ranging signals broadcast from overhead [*see illustration on opposite page*]. In essence, the specially coded radio signals serve as invisible rulers that measure the path from the satellites to the receiver.

The accuracy of a typical $100 handheld receiver places the user within about five to 10 meters of his or her actual location. A more costly military GPS unit can find itself within five meters. Tandem observations using a receiver that gets error adjustments from a nearby stationary receiving device at known coordinates can achieve accuracies of half a meter. These tandem operations are called differential GPS.

Geolocation coverage will extend all the way downtown, indoors and under tree limbs.

Data Stream from Space

TO APPRECIATE WHERE GPS is going, it helps to first review its current operations. The precipitation of transmitted data from GPS satellites is like a light sprinkle. One GPS satellite radiates signals at 500 watts, which is the power of five incandescent lightbulbs. After traveling the 20,000 kilometers from space, the radio-ranging signals arrive at the earth's surface with power densities of only 10^{-13} watt per meter squared. For comparison, the power of a television signal received by a home set is one billion times as strong.

GPS satellites shower down two varieties of information. One type, the navigation message, consists of data bits that identify the satellite's orbital location and the time the transmission was sent. These spatial and temporal coordinates are developed by GPS's ground-control segment, which employs a network of GPS receivers at known reference points to calculate them. These values are beamed up to the satellite and packed into the navigation message for subsequent transmission to all users.

The second type of information GPS satellites emit is a set of ranging codes, a unique patterned sequence of digital pulses. These transmissions do not carry data in the traditional sense. The codes are, in fact, designed to help the receiver measure the arrival time of the incoming signal, a key for precisely determining location. Engineers emphasize the distinct nature of these ranging signals by saying these so-called pseudo-random noise (PRN) codes are made up of a series of "chips" rather than bits.

Each PRN code sequence is like the musical notes in a song. Let's say a particular song was played by both the satellite and the receiver at exactly the same time. The user would hear both versions of the song (or PRN code), but the satellite's rendition would be delayed by the time it takes for sound to make its way from orbit to the earth's surface. If the user timed the delay between when each song version reached a specific note with a stopwatch, he or she could then tell how long it took for sound to traverse the distance from space. By multiplying the resulting number of seconds by the speed of sound, the user could calculate the range to the satellite.

GPS performs an analogous procedure when a receiver monitors a PRN code being broadcast from a satellite. By aligning the received ranging code sequence (the sequence of musical notes) with a replica of the unique PRN code sequence for that satellite stored in the receiver, the device can estimate the delay in the arrival time of that satellite's radio ranging signal. The receiver then multiplies the time delay by the speed of light and so can determine the distance to the satellite.

Overview/*Enhanced GPS*

- More than 30 million people rely on the Global Positioning System (GPS) regularly, and that number will soon multiply as receivers find their way into more cellular phones and automobiles.
- Enhanced geolocation accuracy will result when new signals become available for civilian and military uses. The first of these signals will appear when improved satellites are launched in 2005, and the second will come into operation a few years later.
- GPS integrity machines will guarantee GPS reliability by providing valid error bounds in real time. A range of institutional, operational and technical activities will harden GPS transmissions against radio-frequency interference.

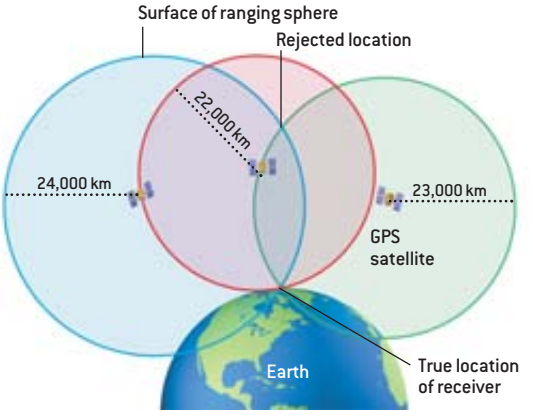
TOM DRAPER DESIGN (*preceding pages*); KEVIN MARTY *AP Photo/Imperial Valley Press* (*background*); FROM LEFT TO RIGHT: CORBIS; LEIF SKOOGFORS *Corbis*; ALAN SCHEIN *Corbis*; CORBIS; RUSSELL MUNSON *Corbis*; PATRICK DURAND *Corbis Sygma*

HOW GPS WORKS

The Global Positioning System (GPS) is a radio-based worldwide navigation system comprising two dozen satellites and associated ground stations. Using a cousin of triangulation called trilateration, GPS calculates the coordinates of a terrestrial location by measuring the distance to at least four satellites. Several factors combine for accurate geolocation.

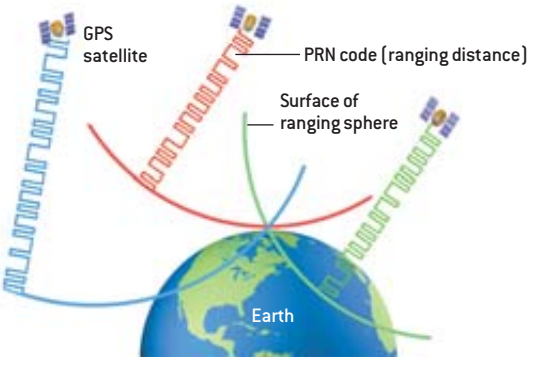
INTERSECTING SPHERES

Say that a GPS receiver measures the distance to one satellite as being 22,000 kilometers. The receiver must then sit on the surface of a ranging sphere centered on the satellite with a radius of 22,000 kilometers. Suppose the receiver also estimates the distance to two other satellites as being 23,000 and 24,000 kilometers, respectively. The receiver's location must therefore be at the intersection of those three ranging spheres. Geometry states that three spheres can mutually intersect at no more than two points. Only one of those positions will be close enough to the earth to be the receiver's position.



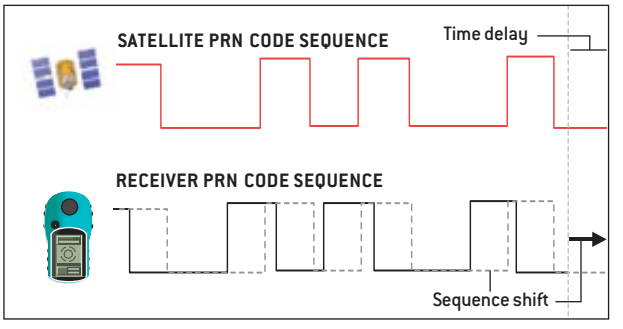
TIMING SIGNALS

Gauging the distance to a satellite requires timing how long it takes for a satellite signal to arrive at a receiver. Velocity multiplied by travel time equals the distance crossed. Radio signals move at the speed of light, or roughly 300,000 kilometers per second. The problem is measuring the travel time. The pseudo-random noise (PRN) code, a complicated sequence of digital data, helps to accomplish this task. Each code is unique to a satellite, which ensures that the receiver does not confuse the signals.



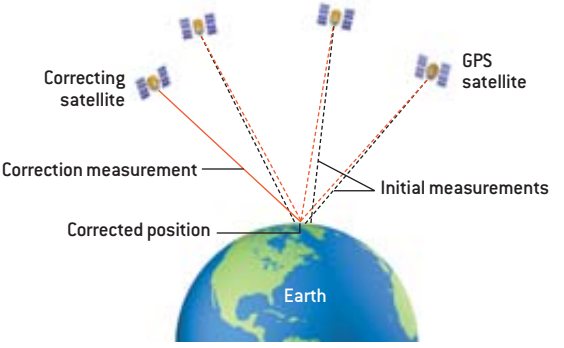
SHIFTING THE CODE

To understand how PRN codes aid in measuring distance, consider this analogy: suppose that both the satellite and the receiver started playing the same song—the PRN code—at the same time. The broadcast traveling from space would be slightly delayed compared with the receiver's reference song. By measuring how long it took for a given note in the song, or segment of the PRN code, from the satellite to arrive, the system could determine travel time. Multiply that time by the speed of light, and the result is the distance to the satellite.



SYNCHRONIZING CLOCKS

Time is kept almost perfectly onboard GPS satellites, each of which carries an atomic clock, but GPS receivers must make do with a cheap, much less accurate quartz clock. The resulting timing flaws mean that the initial three ranging measurements do not intersect exactly (*black dashed lines*). To synchronize the clocks in orbit with those on the earth and thus compensate for the timing error, GPS must make a fourth satellite measurement. This reading determines a single correction factor that will bring the endpoints of the first three ranging measurements into coincidence at the true location of the receiver (*marked in red*).



ALFRED T. KAMAJIAN

Thus, receivers measure range using a virtual ruler that each satellite extends to the earth. The codes provide tick marks on the radio ruler, whereas the navigation message describes the satellite's location, which is analogous to the end point of the ruler. If the GPS unit could incorporate a perfect clock, then three range measurements would allow the receiver to solve for its three-dimensional position—latitude, longitude and altitude. With perfect clocks, a single measurement would place it on the surface of a sphere with the prescribed radius from the satellite. Two GPS readings would locate the user on the intersection of two similar spheres, and three measurements would place the user at a unique point defined by the three spheres. Hence, the receiver would solve three equations for three unknowns: longitude, latitude and altitude.

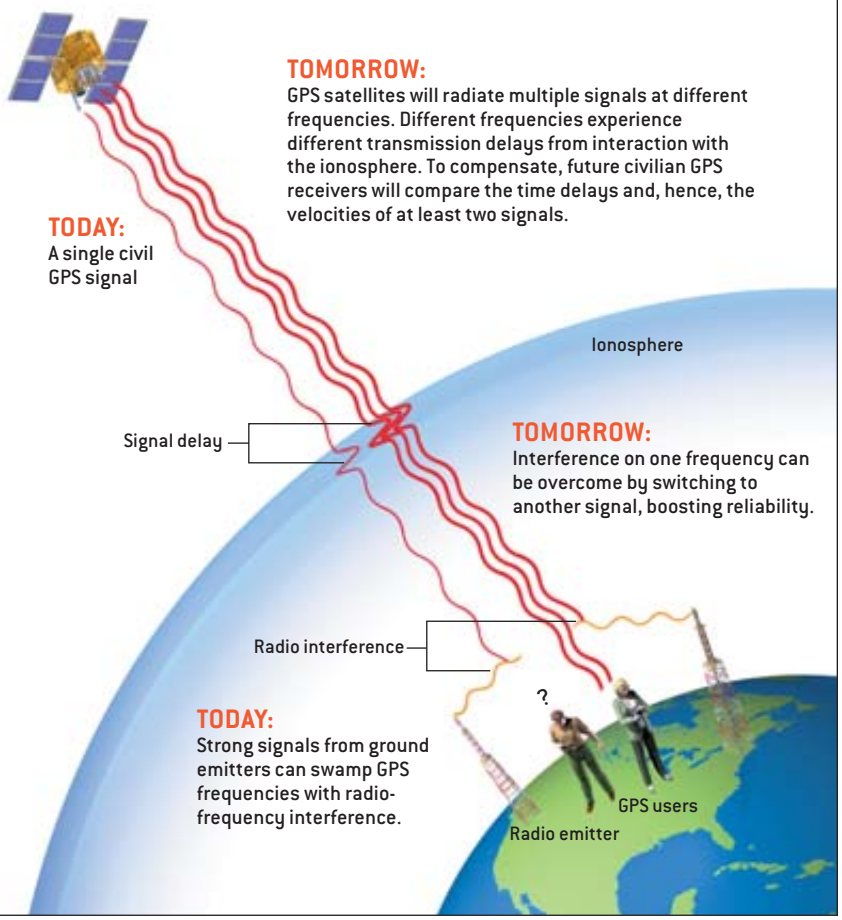
In fact, perfect clocks do not exist, so GPS receivers must also solve for a fourth unknown: the offset between the receiver's internal inexpensive clock and GPS network time. GPS time is controlled to within one billionth of a second by atomic clocks, but the receiver clock might be subject to an error of a second or more per day. One can convert time error to distance error by multiplying by the speed of light (300,000 kilometers per second). This offset adds an unknown number to the distance gauged to each satellite, explaining why the length measurements are called pseudo-range measurements. Fortunately, the time offset is the same for all satellites, so a fourth satellite reading allows the receiver to solve four equations for the four unknowns: longitude, latitude, altitude and time.

Because mobile users change position rapidly, current GPS receivers also monitor the Doppler shift of the incoming signals—that is, motion-caused shifts of the signal's wavelengths. If the user is traveling away from the satellite, the wavelength appears longer. If the user is moving toward the satellite, the arriving wave gets shorter. Each satellite is analogous to a train passing a person (the receiver). As the train approaches, its whistle rises in pitch, but as the train moves away, the pitch becomes lower. Monitoring these wave shifts allows these devices to estimate the user's velocity directly and more accurately.

It is notable that GPS receivers accomplish the complex geolocation task without transmitting any signals. Nevertheless, those receivers destined for installation in future cell phones will be quite cheap, costing less than $5 apiece.

MULTIPLE SIGNALS FOR BETTER RELIABILITY

Finding GPS coordinates requires precise estimation of the distances from geolocation satellites to a receiver—a calculation that depends on the time it takes the signal to travel from orbit [*see box on preceding page*]. Charged particles in the ever changing ionosphere, however, slow the signals, which creates a timing error. Advanced GPS will correct for ionospheric effects and signal disruption from competing broadcasts.



To Pierce the Ionosphere

TRANSMITTERS ONBOARD GPS satellites broadcast their information through standard radio-frequency (RF) waves. The RF carrier is the classic sinusoid; its frequency counts the number of cycles (each peak and valley) per second. Current GPS technology employs two frequency bands—L1 and L2—that fall in the microwave portion of the radio spectrum. L1 is commonly referred to as the civil signal, even though the armed forces also share this resource. It is available to everyone and supports the vast majority of today's civilian applications. L2 serves the military primarily. The public is permitted to use the L2 signal, but without knowledge of the military PRN codes. This knowledge gap makes civilian application of L2 fragile. Civilian receivers, for example, have difficulty using the L2 signal from satellites that are sitting low in the sky or are obscured by even minor obstructions, such as trees. Moreover, L2 receivers are expensive because they require special signal-processing techniques to ac-

ALFRED T. KAMAJIAN

cess the L2 signal when the PRN codes are not known.

For these reasons, the vast majority of civilian units use only the L1 signal. By so doing, they typically achieve an accuracy of five to 10 meters, an error range largely caused by charged particles in the earth's ionosphere, which extends from about 70 kilometers above the ground out to 1,300 kilometers or more. This conductive shell slows the transmission of radio waves from the GPS satellites much as water in a glass bends or diffracts the view of an immersed pencil. Depending on conditions, it can delay the arrival of transmission from one to 10 meters or more.

To compensate, some users employ differential GPS, or D-GPS. The technique involves two GPS receivers: a roving unit and a reference unit that is placed at a known location. The reference device transmits the differences between its measurements and the computed ranges to the roving receiver, which then uses the data to correct its reported location. D-GPS works best when the mobile receiver stays relatively close to the reference receiver. At ranges of less than 100 kilometers, the ionospheric errors cancel out almost completely because the radio beam from the satellite to the reference receiver passes through the same atmospheric obstacles that the signal from the satellite to the mobile receiver did.

Sharper, Stronger Signals

STARTING IN 2005, GPS satellites will begin to broadcast new signals that will boost the robustness of services and help fine-tune their positioning accuracy by eliminating the ionospheric errors [*see illustration on opposite page*]. Two military signals will be added to the L1 and L2 bands, and another civilian signal will supplement the L2 band. The current signals will continue to operate to ensure that existing receivers will work well into the future. By around 2008, a further round of improved GPS satellites will begin to emit even more civil signals in a third frequency band called L5. (L3 and L4 carry nonnavigation information for the military.) The new L5 signals will be four times as powerful as today's.

The extra signals will enable a single receiver to calculate the transmission delay caused by the ionosphere, reducing errors. L1 signals traveling through the uneven ionospheric layer would show a delay different from that seen in signals sent through L5, for example. Future receivers could thus first compare the delay in the signals received from L1 and L5. They could then use this calculation to estimate the electron density of the ionosphere and compensate for its effects. This is the calculation that some costly civilian GPS receivers try to make using the current civil signal at L1 and the military signal at L2. Because the civil signals will employ publicly known codes, the operational frailties currently associated with dual-frequency signal processing will disappear. This means that dual- or even triple-frequency receivers will be the rule for consumer and commercial users.

Operators of D-GPS units will also benefit from the new signals. Remember that D-GPS accuracy degrades as the user moves away from the reference receiver, because the radio beam from the satellite to the user pierces the ionosphere at a point that is increasingly distant from where the reference beam traversed the plasma layer. With multiple frequencies, the roving receiver would be able to evaluate the ionosphere autonomously and the D-GPS corrections could be used to mitigate the other (smaller) errors. Future D-GPS users will be able to achieve accuracies of 30 to 50 centimeters.

The most demanding users of today's GPS, including surveyors, scientists and farmers, need centimeter- and even millimeter-level accuracy. Such accuracy requires an advanced form of D-GPS that goes beyond the application of PRN codes, a technique that digs under these codes and measures the arrival time of the carrier waves that transport GPS signals from orbit.

The radio-frequency waves that carry the GPS signals are sinusoidal microwaves. An individual cycle has a wavelength—the distance from one peak to the next—of 19 centimeters. A receiver can measure this arrival time with a precision of about 1 percent. This resolution corresponds to a travel distance of one or

Overcoming GPS Signal Interference

GPS radio emissions are very weak, so users depend on a quiet radio spectrum; even low-power radio-frequency sources can interfere with GPS operations. The U.S. Federal Communications Commission has mandated that the GPS bands must be kept quiet—only nature's radio noise is present.

Despite these efforts, safety-critical users, such as air traffic controllers, are still subject to signal loss from accidental interruption or malevolent jamming of GPS broadcasts. Thankfully, these users have access to a growing number of defenses against interference. Airborne users, for example, can rely on backup navigation systems based on inertial measurements, Loran-C, or distance-measuring equipment. Military GPS applications frequently make use of "smart," beam-steering antennas that selectively null out interfering signals (by suppressing reception from certain directions) without appreciably degrading GPS signal strength. In the not too distant future, consumer applications may well augment GPS reliability with range measurements to the antennas of nearby television stations or cellular-phone base stations. —*P.E.*

THE AUTHOR

PER ENGE is professor of aeronautics and astronautics at Stanford University, where he is Kleiner-Perkins, Mayfield, Sequoia Capital Professor in the School of Engineering. Enge also serves as associate chair of the department and director of the GPS Research Laboratory, where he works on GPS integrity machines that provide error bounds for GPS data in real time and harden systems against radio-frequency interference. He has received the Kepler, Thurlow and Burka awards from the Institute of Navigation (ION) for his research. He has been made fellow of the Institute for Electrical and Electronics Engineers and the ION.
The author gratefully acknowledges the support of the Federal Aviation Administration, the U.S. Navy and NASA. He also extends his thanks to his Stanford colleagues and to the larger aviation navigation community.

two millimeters. This is the level of accuracy high-end users need, yet the carrier-wave measurements are ambiguous—that is, the receiver cannot tell which cycle is which. Unless it can uniquely identify which individual cycle it is tracking, the measurement can contain an error equal to any number of whole wavelengths.

This difficulty resembles measuring distance using the fine tick marks on a ruler. Unlike the coarse tick marks, the fine marks are very close together and therefore precise, but because they are not individually marked, they are ambiguous. Fortunately, a special procedure unambiguously links the rough, 30-centimeter-scale resolution of standard D-GPS to the desired fine, two-millimeter-scale resolution of the carrier wavelengths. This process generates an intermediate-length measurement scale with the right-size resolution to connect them. The computational bridge that spans the measurements is built on this intermediate scale, as I will explain.

The challenge is best understood by analogy. As noted previously, PRN codes can be likened to a song's musical note sequence, where every note is distinct and identifiable. The carrier-wave measurements are akin to the song's drumbeat, which encompasses many beats per note. If one listens to the drumbeat alone, it is hard to say what part of the song one is hearing. The key is to use the song's notes to identify which drumbeat is which. For GPS, this is a tough task: the starting time of each note (or PRN code chip) can be determined with an accuracy of just 30 centimeters. Each drumbeat (the carrier wavelength) lasts for just 19 centimeters. Separated by only 19 centimeters, these beats are too close together to discern—one cannot tell them apart with the 30-centimeter accuracy of the PRN codes.

To identify an individual beat, one needs an additional drummer, one that beats at a slower rate. Advanced GPS receivers create this slower beat by multiplying the L1 carrier and the L2 carrier to produce what is known as a beat frequency. This operation also has a musical analogy. When two tones are played simultaneously on an instrument, the listener hears the original tones but also perceives a new tone corresponding to the difference in the two original frequencies—the beat frequency. Because the new frequency equals the difference frequency, it is necessarily lower in pitch than either of the two original tones. Lower frequencies mean longer wavelengths. In GPS the wavelength of the difference frequency is 85 centimeters, and the system can measure that with a resolution of about eight millimeters. This wavelength is sufficiently long to be resolved to the 30-centimeter acuracy of the receiver's code measurements. Thus, expensive receivers that employ this technique can meet the requirements of top-end users.

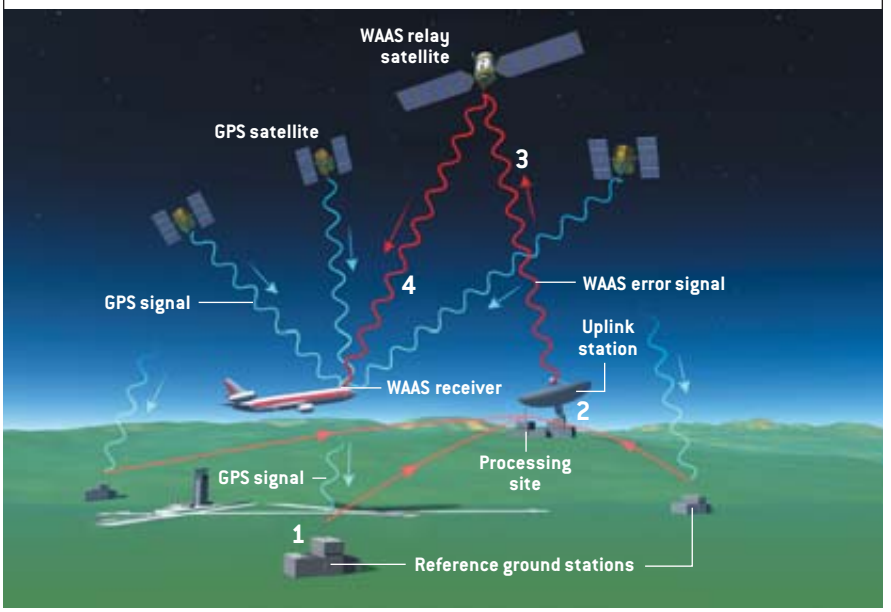
With the soon-to-be-added GPS signals, this computational bridge from the PRN code chips to the underlying carrier cycles will get even stronger: Civilian receivers will have access to public codes on the L2 signal as well as on the brand-new L5 signal. Receivers will be able to process a trio of beat notes (L1 minus L2, L1 minus L5, and L2 minus L5), which will provide several paths from the PRN code chips to the carrier cycles and, thus, ultrahigh geopositioning accuracies.

Flight-Ready GPS

TO SEE SOME of the real-world implications of improved GPS, consider the Federal Aviation Administration's new flight guidance technology, a function in which reliability is clearly critical. The innovative systems, parts of which are already online, will allow pilots to employ GPS to guide aircraft right down to the runway even when severe weather creates conditions of zero visibility. Completing this task safely and surely entails more than mere navigation accuracy. It also requires two guarantees. First, pilots must know the maximum size of their possible

FLYING ON A WING AND A WAAS

Flying safety is greatly improved when pilots know their airplane's position precisely. The wide-area augmentation system (WAAS), designed by the U.S. Federal Aviation Administration, improves the accuracy and integrity of safety-critical GPS signals. WAAS offers accuracy of one to two meters in the horizontal axis and two to three meters in the vertical throughout most of the U.S. The system starts with a network of 25 reference ground stations that are placed at known locations (*1*). Each station compares its GPS satellite reading with its confirmed map coordinates and develops corrections for all the satellites in view, which are then transferred to one of two master processing sites (*2*). From there, correction data are uplinked to geostationary relay satellites (*3*), which in turn broadcast to WAAS receivers (*4*) that decode the geopositioning corrections in real time.



ALFRED T. KAMAJIAN

positioning error (that is, an error bar) for all circumstances. When maneuvering for final approach, for example, a pilot can tolerate location errors that are no larger than 10 meters. Second, users need a guarantee that their navigation system will suffer no breaks in service.

The FAA has developed two D-GPS-based systems to provide these real-time error bounds for location data. These systems include networks of reference receivers that monitor GPS measurements continuously but operate independently from the ground-control segment.

The wide-area augmentation system (WAAS), which began operating in 2003, relies on a nationwide network of sensor stations to measure GPS performance [*see illustration on opposite page*]. These monitors resemble the reference stations for D-GPS, and indeed, WAAS does produce corrections to improve accuracy. In addition, however, it compares positioning corrections from multiple stations to generate the error bounds that are crucial for guiding aircraft. It then employs geostationary satellites to relay its performance guarantees to pilots. If needed, WAAS can adjust the transmitted error range within seven seconds. The system pinpoints the location of aircraft flying at altitude and helps to steer aircraft that are descending toward airports down to an altitude of 300 feet. Engineers in Europe, China, Japan, India, Australia and Brazil are working on similar systems.

Where WAAS leaves off, local systems take over to shepherd aircraft on the lower segments of their landing paths. In time, the local-area augmentation system (LAAS) will enable completely automatic landings in zero visibility. Because the system serves only aircraft near an airport, it uses a short-range radio system to send its corrections and error bounds. LAAS is closely related to the Joint Precision Approach and Landing System (JPALS), a developmental system that will guide aircraft onto the pitching and rolling decks of aircraft carriers. During final approach, naval aviators must control the altitude of their aircraft relative to a moving deck to within a single meter to make sure that the drag hook hanging off the rear fuselage catches the capture cable.

Navy engineers are attempting to make carrier landings easier and safer through JPALS, which places the D-GPS reference receiver on the aircraft carrier. It should enter trials later this year. Both LAAS and JPALS are dual-frequency systems—two GPS frequencies are required to ensure accuracy during these most demanding of aircraft operations. JPALS will be able to use the military signals that are available on L1 and L2 today.

Even though the aforementioned improvements will make GPS all but ubiquitous, the U.S. government has begun planning the next round of further improvements to satellite navigation technology, known as GPS III. The driving forces behind the upgrade are to gain even better reliability and accuracy, to ensure more resistance to interference and jamming, and to foster the adoption of alternative geolocation services as well as new, more sophisticated GPS-enabled applications such as intelligent highway and traffic safety systems. As a result, industrial competitors for the eventual multibillion-dollar program—Boeing and the partnership of Lockheed Martin and Spectrum Astro—have announced that they will vie for the contracts. Initial launch of a GPS III satellite may occur early next decade. SA

GPS during Wartime

In recent years, civilian GPS users have begun to vastly outnumber the military users for whom the system was originally developed. The civil signal is available free to anyone with a GPS receiver. But because the U.S. armed forces and its allies rely on GPS for navigation and weapons targeting, military use takes priority when martial conflict threatens. In regions of the world where warfare is occurring, the U.S. could jam local GPS operations by transmitting strong radio signals with frequencies that lie right in the center of the bands, swamping the weak GPS signals. Sanctioned military use, however, continues because the military signal broadcasts are far enough offset from the center of the civil bands to be unaffected. Under these circumstances, any adversary's utilization of the military signals is impossible, because the military codes are secret. Enemy jamming of the military GPS signals is likely to be short-lived, given that allied armed forces can rapidly detect and destroy such systems, as was demonstrated in the recent war in Iraq. Civilian GPS use would continue outside the conflict area because any jamming signal would lose power far away from the jamming emitter. —*P.E.*

GPS will guide aircraft right down to the runway, even in conditions of zero visibility.

MORE TO EXPLORE

Global Positioning System: Theory and Applications Set. Edited by Bradford W. Parkinson, James J. Spilker, Jr., Penina Axelrad and Per Enge. American Institute of Aeronautics and Astronautics, 1996.

Proposed New L5 Civil GPS Codes. J. J. Spilker, Jr., and A. J. Van Dierendonck in *Navigation: Journal of the Institute of Navigation,* Vol. 48, No. 3, pages 135–143; Fall 2001.

Global Positioning System: Signals, Measurements, and Performance. Pratap Misra and Per Enge. Ganga-Jamuna Press, 2001. Available from Navtech Seminars.

The Rosum Television Positioning Technology. M. Rabinowitz and J. J. Spilker in *Proceedings of the 59th Annual Meeting of the Institute of Navigation;* 2003. **www.ion.org**

GPS World. Monthly magazine published by Advanstar Communications: **www.gpsworld.com**

Galileo's World. Quarterly magazine published by Advanstar Communications from 1999 to 2002.

Institute of Navigation: **www.ion.org**

GPS Research Laboratory at Stanford University: **www.stanford.edu/group/GPS/**

The Transit of Venus

When Venus crosses the face of the sun this June, scientists will celebrate one of the greatest stories in the history of astronomy

By Steven J. Dick

"We are now on the eve of the second transit of a pair, after which there will be no other till the twenty-first century of our era has dawned upon the earth, and the June flowers are blooming in 2004.... What will be the state of science when the next transit season arrives God only knows."

BLACK DOT in front of the sun is Venus, photographed during the last transit in 1882.

—U.S. Naval Observatory astronomer William Harkness, 1882

SPECIAL COLLECTIONS, VASSAR COLLEGE LIBRARIES

June 8, 2004, will dawn just like any other day, but around the world many lucky individuals will witness an extraordinarily rare astronomical event. Properly situated observers equipped with suitable filters for their eyes, binoculars or telescopes will be able to see the planet Venus silhouetted against the sun, a black dot moving across the fiery disk for almost six hours. The entire transit of Venus will be visible in most of Asia, Africa and Europe. People in Australia will see only the opening stages of the transit before the sun sets there, and Venus will be three quarters through its crossing by the time the sun rises over the eastern coasts of the U.S. and South America. Those unlucky souls on the western coast of the U.S. and in southwestern South America will miss the event completely [*see illustration on page 104*].

A transit of Venus is not nearly as spectacular as a solar eclipse, caused by the passage of the moon between Earth and the sun. Although Venus is three and a half times as large as the moon, it is so much farther away from Earth that it appears as a speck against the sun, with only about 3 percent of the sun's diameter. So why are scientists, educators and amateur astronomers so excited about the upcoming transit? Partly because it is such a rare phenomenon—astronomers have observed a transit of Venus only five times before, with the last one occurring on December 6, 1882. If sky watchers miss the 2004 transit, they will have another chance in 2012, but after that they will have to pass the baton to their descendants in the year 2117.

Overview/*A Wondrous Transit*

- A transit of Venus occurs when the planet passes directly in front of the sun (as observed from Earth). Usually only four transits happen every 243 years.
- Because a transit of Venus is barely visible to the naked eye, astronomers have observed the phenomenon only five times before. In the 18th and 19th centuries, scientists tried to use the transit to gauge the distance between Earth and the sun.
- Professional and amateur astronomers are eagerly awaiting this year's transit. The observations could be useful to researchers who are preparing a spacecraft designed to detect planets in other solar systems.

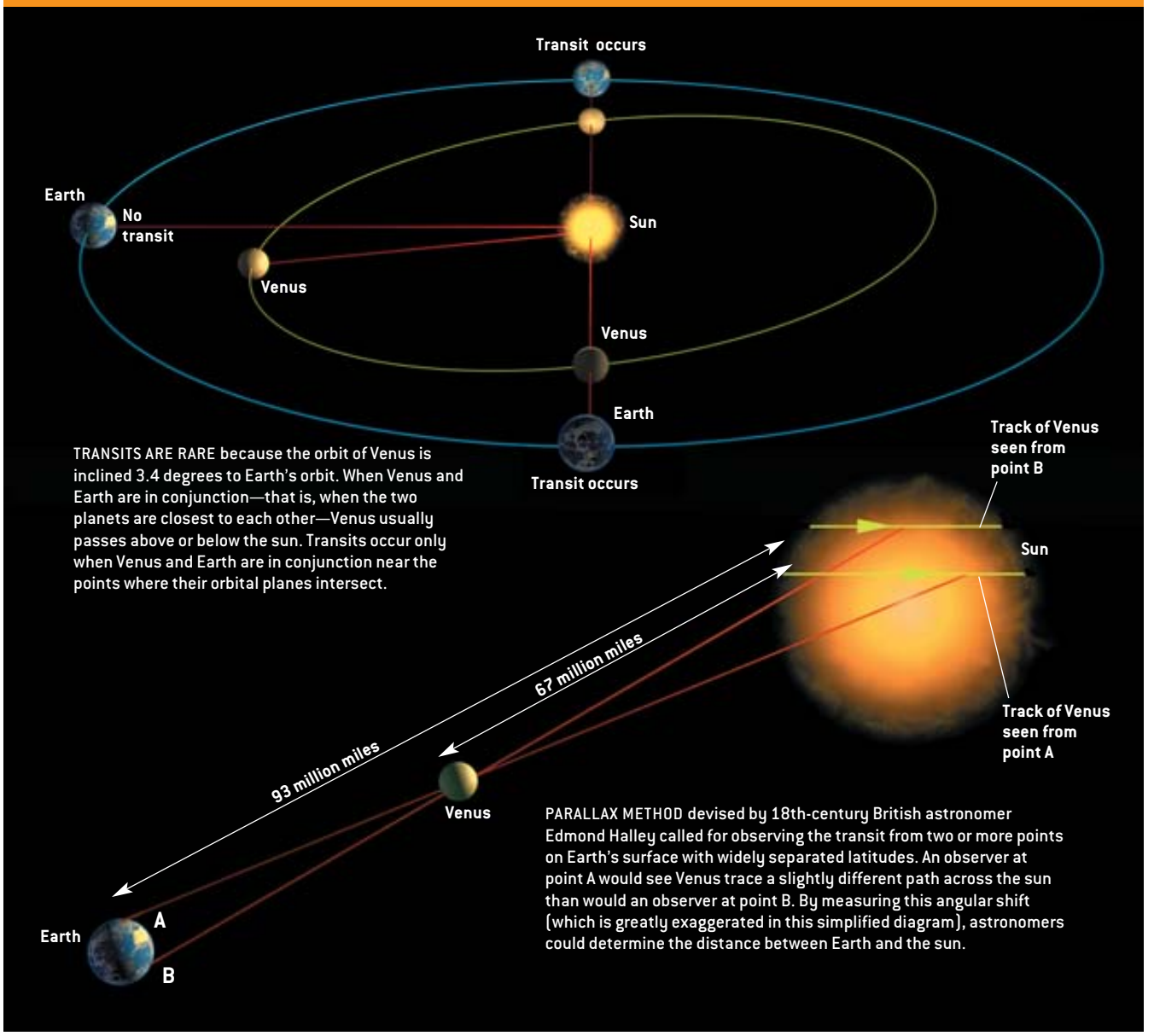
Another part of the transit's appeal is the colorful history of the efforts to observe it in the 17th, 18th and 19th centuries. The story has all the ingredients of a scientific thriller: international rivalry, mysterious observational effects and controversial results bearing on one of the most confounding problems in the history of astronomy. In addition, the phenomenon is of great interest to current researchers because the transit of Venus may shed light on a hot topic in modern astronomy: the detection of planets in other solar systems.

From Kepler to Captain Cook

THE OCCURRENCE OF any planetary transit is a simple matter of geometry—the planet in question must pass between the observer and the sun. From Earth, one may see transits of Mercury and Venus. From Mars, one may view a transit of Earth as well. (Arthur C. Clarke's famous short story "Transit of Earth" was inspired by the realization that an observer on Mars on May 11, 1984, would have seen Earth cross the face of the sun.) Such events are relatively rare because the orbits of the planets are not in the same plane as the ecliptic, the sun's path in the sky as seen from Earth. The orbit of Venus, for example, is inclined 3.4 degrees to Earth's orbit, so even when Venus is in the same direction as the sun ("in conjunction," as astronomers say), most of the time it is too far above or below the ecliptic to cross the sun's face [*see top illustration in box on opposite page*]. For a similar reason, the moon does not eclipse the sun once a month as it orbits Earth; it generally passes above or below the ecliptic.

A transit of Venus takes place only when Earth and Venus are in conjunction near the points where their two orbital planes intersect. As a result, transits typically recur only four times every 243 years. The intervals between transits follow a predictable pattern: one transit is generally followed by another eight years later; the next transit occurs after 105.5 years and the next after another eight years; and the cycle begins again after another 121.5 years. Why do the transits usually occur in pairs separated by eight years? Because Venus takes 224.7 days to travel around the sun, 13 Venusian years are almost exactly equal to eight Earth years. Eight years after the first transit of a pair, Venus and Earth return to almost the same positions in their orbital dance, so they will still be aligned roughly with the sun. The sun's angular diameter—how big it appears in the sky—is about half a degree, which allows a little leeway; if the first

THE GEOMETRY OF TRANSIT



ALFRED T. KAMAJIAN

transit took place near one edge of the solar limb, the next one will be close to the opposite edge. Occasionally, though, only a single transit occurs because one of the pair is a near miss. There was just one transit of Venus in the 14th century, and this will be the case again on December 18, 3089.

Because a transit of Venus is barely visible to the naked eye, for most of history humans went about their lives oblivious to such events. The first to predict a planetary transit was 17th-century German astronomer Johannes Kepler, whose Rudolphine Tables provided what was then the most accurate guide to planetary motion. Kepler determined that Mercury would cross the sun on November 7, 1631, followed by Venus on December 6 of the same year. Kepler did not live to see if his predictions were correct; he died in 1630. But the transit of Mercury was observed by at least three people, most notably French natural philosopher Pierre Gassendi, who left a detailed account. Gassendi estimated Mercury's apparent diameter to be about 20 arc seconds—that is, about 1/180 of a degree—which was in itself a considerable scientific advance. The transit of Venus, on the other hand, was not visible in Europe, and although Kepler had spread the word around the world, no one is known to have observed it.

English astronomer Jeremiah Horrocks (1618–1641) realized that another transit of Venus would occur on December 4, 1639. (Horrocks listed the date as November 24 because England did not adopt the Gregorian calendar until 1752.) He set up a small telescope in his home in Much Hoole, near Liverpool; by projecting the light from the telescope onto a sheet of paper, he was able to view an enlarged image of the sun. He saw nothing unusual until noon, when he reluctantly had to dash off, possibly to attend a church service. When he returned shortly after three o'clock, he found Venus already on the face of the sun! Although Horrocks was only able to observe the early stages of the

"I then beheld a most agreeable spectacle … a spot of unusual magnitude and of a perfectly circular shape…."
—English astronomer Jeremiah Horrocks, 1639

transit for about 30 minutes before sunset, he estimated Venus's apparent diameter at about one arc minute, three times the diameter Gassendi had measured for Mercury. From Manchester, 25 miles southeast of Much Hoole, Horrocks's friend William Crabtree used a similar telescope to glimpse Venus in transit just before sunset. As far as we know, Horrocks and Crabtree were the only two humans to witness the event.

The 1761 and 1769 transits of Venus were the subject of much more serious observations. By this time British Astronomer Royal Edmond Halley, best known for his famous comet, had detailed a method of using the transit of Venus to determine the distance between Earth and the sun (now known as the astronomical unit). If scientists observed the transit from two or more points on Earth's surface with widely separated latitudes, each observer would see Venus trace a slightly different path across the sun [*see bottom illustration in box on preceding page*]. Because each path takes the form of a chord—a straight line connecting two points at the edge of the sun's disk—astronomers could measure the angular shift between the paths by comparing the durations of the transits. This angular shift, called the parallax of Venus, would provide a measure of the distance between Earth and Venus because the two quantities are inversely proportional to each other. To see how this method works, hold a finger in front of your face and view it alternately with one eye and then the other. The apparent shift in the finger's position as you open and close your eyes is greater when the finger is close to your face than when it is far away.

Although Mercury has 13 or 14 transits every century, it was not a suitable candidate for either Halley's parallax method or the variations devised by later astronomers. Because Mercury is so far from Earth, the angular shifts were too small to be measured accurately. Even with the much closer Venus, the observations were tricky; it was crucial to know the exact geographic positions of the observing stations and to accurately time the four "contacts" between Venus and the sun. (The first and second contacts occur at ingress, when Venus's disk touches the sun's from first the outside and then the inside; the third and fourth contacts occur at egress.) But the potential payoff from the observations would be enormous. Astronomers already knew from Kepler's laws of planetary motion the relative distances of all the planets from the sun, so they could determine the solar parallax from the parallax of Venus. And this measurement, in turn, would allow scientists to estimate not only the distance between Earth and the sun but also the scale of the entire solar system.

Unfortunately, the results from the 1761 transit were not as good as expected: the measured values of the solar parallax ranged from 8.3 to 10.6 arc seconds. But the observations in 1769 yielded a narrower range—from 8.43 to 8.8 arc seconds—which put the estimate of the astronomical unit between about 93 million and 97 million miles. Among the observers in 1769 was David Rittenhouse, the preeminent scientist in the American colonies, who fainted from excitement after peering through his telescope. The first voyage of British explorer Captain James Cook on the *Endeavour* was mounted in large part to observe the transit while exploring the South Pacific. Cook and his crew did so successfully from an area still known as Point Venus in Tahiti and from two nearby locations. But Cook reported an ominous problem that also plagued other observers: attempts to determine the exact times of contact of Venus with the sun were frustrated because the limbs of the two bodies appeared to cling together for several seconds [*see right illustration on opposite page*]. Cook speculated that this phenomenon, which became known as the black-drop effect, was caused by "an atmosphere or dusky cloud round the body of the planet."

When in 1824 German astronomer Johann Franz Encke analyzed the results of both 18th-century transits, he settled on a value of 8.58 arc seconds for the solar parallax, which corresponded to a mean distance for the sun of 95.25 million miles. Thirty years later, however, Danish astronomer Peter Andreas Hansen argued, based on perturbations of the moon's motion caused by the sun's gravity, that the sun must be considerably closer. This claim gained further support in 1862, when measurements of the parallax of Mars—determined by comparing the planet's position in the sky from two widely separated observation points—gave estimates between 91 million and 92.5 million miles for the astronomical unit. Thus, on the eve of the 19th-century transits of Venus, the distance to the sun was still a value of considerable uncertainty. British Astronomer Royal

THE AUTHOR

STEVEN J. DICK is chief historian for NASA. For 25 years he was an astronomer and historian of science at the U.S. Naval Observatory, the institution that led the American transit of Venus expeditions in 1874 and 1882. He is author of *The Biological Universe, Life on Other Worlds* and, most recently, *Sky and Ocean Joined: The U.S. Naval Observatory, 1830–2000* (Cambridge University Press, 2003). The latter includes a detailed historical chapter on the transits of Venus. He is past president of the History of Astronomy Commission of the International Astronomical Union and now serves as chairman of its Transit of Venus Working Group.

ASTLEY HALL

George B. Airy said at midcentury that determining the solar parallax was "the noblest problem in astronomy." Agnes Mary Clerke, a 19th-century astronomy historian, wrote that the solar parallax was "the standard measure for the universe … the great fundamental datum of astronomy—the unit of space, any error in the estimation of which is multiplied and repeated in a thousand different ways, both in the planetary and sidereal systems."

Desperately Seeking Parallax

BY 1857 AIRY HAD FORMULATED a general plan for observing the 1874 transit of Venus, and by 1870 Britain was constructing the necessary instruments. Similar plans were under way in other parts of the scientific world. As the much anticipated event approached, no fewer than 26 expeditions were launched from Russia, 12 from Britain, eight from the U.S., six each from France and Germany, three from Italy and one from Holland. "Every country which had a reputation to keep or to gain for scientific zeal was forward to co-operate in the great cosmopolitan enterprise of the transit," Clerke wrote. The colorful history of these expeditions would take a book to describe; each has its own story, and each met varying degrees of success or failure.

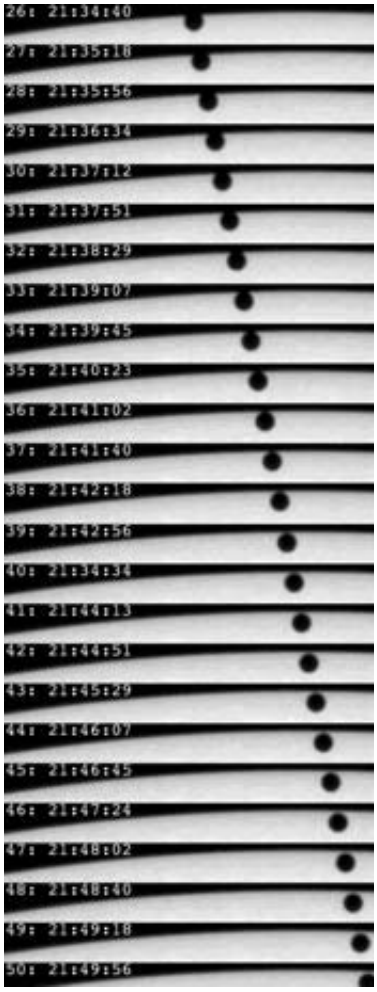
Simon Newcomb of the U.S. Naval Observatory—the leading astronomical institution in America at that time—urged the National Academy of Sciences to take up the problem. Congress formed an American Transit of Venus Commission, in which Newcomb and other Naval Observatory astronomers played a prominent role. The commission outfitted a total of eight expeditions for the 1874 event—three to the Northern Hemisphere and five to the Southern. In all, Congress appropriated the munificent sum of $177,000, equivalent to more than $2 million today.

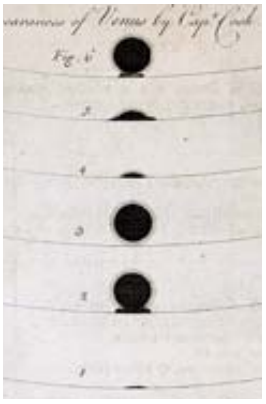
Each expedition was equipped with elaborate instrumentation. To visually observe the moments of contact between Venus and the sun, the researchers employed a refractor telescope with a five-inch-wide lens made by Alvan Clark and Sons, the premier telescope maker in 19th-century America. The scientists were also able to photograph the sun using a photoheliograph, an instrument that had been invented only two decades before. Sunlight was directed through a fixed horizontal telescope, which had a focal length of 40 feet, by a slowly turning mirror that kept the sun's image stationary. The telescope produced images of the sun four inches in diameter, allowing astronomers to track Venus's movement precisely across the solar disk.

SCIENTIFIC AMERICAN, which avidly followed the progress of the expeditions, reported in its September 26, 1874, issue that the ship *Swatara* carrying the U.S. observation parties bound for the Southern Hemisphere had made the trip from New York to Brazil in a speedy 35 days. The Europeans, for the most part, opted for a different photographic setup: smaller telescopes with shorter focal lengths. Their equipment was designed to yield high-quality photographs, but because their images would be smaller than those of the American teams, measuring Venus's position against the sun would be more difficult.

When the transit finally took place on December 9, 1874, bad weather stymied many of the expeditions. Worse, when the astronomers analyzed the visual contact observations, they soon found that the results were no better than those recorded in the 18th century. Around the world the problem was the same. William Harkness, the U.S. Naval Observatory astronomer who led the observation party at Hobart Town on the Australian island of Tasmania, stated that "the black drop, and the atmospheres of Venus and the Earth, had again produced a series of complicated phenomena, extending over many seconds of time, from among which it was extremely difficult to pick out the true contact."

The photographic observations were thus all the more important, but here again disappointment was widespread. Harkness recalled that "it soon began to be whispered about that those taken by European astronomers were a failure." The official British report declared that "after laborious measures and calculations it was thought best to abstain from publishing the results of the photographic measures as comparable with those deduced from telescopic view." As Harkness noted, it was impossible to determine accurately Venus's position against the sun because the researchers could not pinpoint the boundary of the solar disk: "However well the sun's limb on the photograph appeared to the naked eye to be defined, yet on apply-





BLACK-DROP EFFECT was observed by British explorer Captain James Cook during the 1769 transit of Venus. A drawing based on Cook's observations (*above*) shows the limb of Venus clinging to the sun's perimeter, making it impossible to determine the exact moment of contact. Cook speculated that the cause was an atmosphere around Venus. But in 1999 the Transition Region and Coronal Explorer (TRACE) spacecraft recorded a similar phenomenon during a transit of Mercury, which has no atmosphere (*left*). The cause of the black drop is still controversial.

GLENN SCHNEIDER, JAY M. PASACHOFF AND LEON GOLUB *Lockheed Martin Solar and Astrophysics Laboratory and Smithsonian Astrophysical Observatory (left)*; LINDA HALL LIBRARY OF SCIENCE, ENGINEERING AND TECHNOLOGY (*right*)

ing to it a microscope it became indistinct and untraceable, and when the sharp wire of the micrometer was placed on it, it entirely disappeared." The French did publish their results, but with wide error bars.

All hope focused on the American expeditions, which had returned with about 220 measurable photographic plates taken with the long-focus photoheliographs. A value for the solar parallax of 8.883 arc seconds was published in 1881, on the eve of the next transit. But the results were sufficiently ambiguous that many astronomers, including Newcomb, argued that the transit of Venus was not a good method for determining the astronomical unit. Harkness, though, never lost faith, and with additional congressional appropriations, the U.S. mounted eight more expeditions to observe the transit of 1882. After analyzing the photographs of this transit for almost a decade, Harkness concluded that the best estimate of the solar parallax was 8.809 arc seconds, yielding a sun-Earth distance of 92,797,000 miles, with a probable error of 59,700 miles. The actual average distance, now measured precisely through spacecraft observations and other techniques, is 92,955,859 miles. (The corresponding value of the solar parallax is 8.794148.)

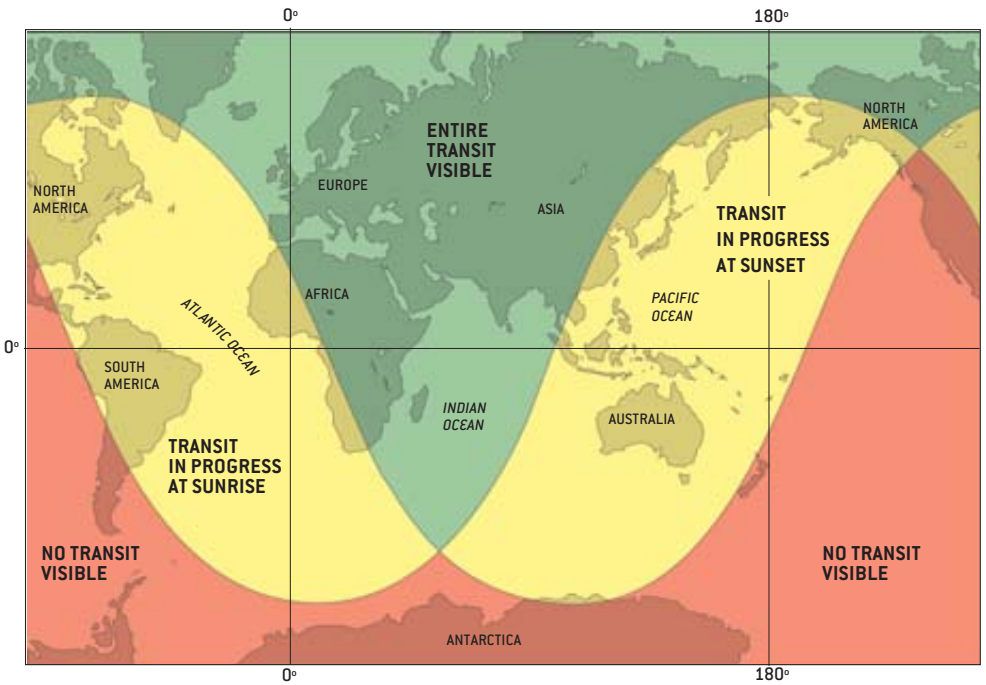
How important were the transit of Venus observations to the history of astronomy? Although Newcomb, whose system of astronomical constants would be used internationally for most of the 20th century, adopted a value for the solar parallax that was quite close to Harkness's, he gave the transits of Venus a very low weight compared with other methods for estimating the constant. In his opinion, the black-drop effect and other errors greatly impaired the transit's use for determining the astronomical unit.

Interestingly, the cause of the black-drop effect is still a subject of much controversy. Eighteenth- and 19th-century astronomers attributed it to a variety of sources, including the atmospheres of both Earth and Venus. But when scientists used the Transition Region and Coronal Explorer (TRACE) spacecraft to observe the 1999 transit of Mercury—a planet with no atmosphere, viewed from a satellite far above Earth's atmosphere—they still saw a weak black-drop effect [*see left illustration on preceding page*]. Although this finding does not preclude atmospheric effects strengthening the black drop, the underlying cause must be something else.

The TRACE team (led by Glenn Schneider of the University of Arizona's Steward Observatory, Jay M. Pasachoff of Williams College–Hopkins Observatory and Leon Golub of the Smithsonian Astrophysical Observatory) concluded that the black drop is caused partly by optical smearing between the planetary and solar disks. To see a similar phenomenon, hold your thumb and forefinger very close together and view the narrow gap against a bright background; a dark "ligament" will appear between them even when they are not touching. In addition, solar limb darkening—the lessening of brightness at the sun's edge—also contributes significantly to the black drop. The TRACE researchers suggested that the effect might be mitigated using new techniques for the upcoming Venus transit.

Safety First!

ALTHOUGH TRANSITS OF VENUS are no longer important for determining the astronomical unit, this year's event is sure to be one of the most widely observed in astronomical history. It is possible to see Venus against the sun without magnification and certainly with binoculars or a small telescope, but Fred Espenak of the NASA Goddard Space Flight Center warns that viewers must take the same precautions that are employed for a solar eclipse. Looking at the sun through a telescope without



BEST PLACES to observe the June 8, 2004, transit of Venus are in Europe, Africa and Asia (*left*). Sky watchers in Australia and the eastern U.S. will be able to see only parts of the transit; people in the western U.S. will miss the event completely. To avoid eye damage, observers *must* use appropriate solar filters when viewing the transit (*above*). More information about safety precautions can be found on the Web at www.transitofvenus.org/safety.htm

NINA FINKEL (*left*); CHUCK BUETER (*right*)

"When the last transit season occurred the intellectual world was awakening from the slumber of ages. . . ."

—William Harkness, 1882

an appropriate filter can cause instant eye damage and permanent blindness.

One of the safest ways to see the transit is to project the image of Venus and the sun onto a piece of paper. Using time-honored occultation techniques, amateur astronomers can make useful observations of the timings of contacts, which they should send (along with their geographic coordinates) to the Mercury/Venus Transit Section of the American Association of Lunar and Planetary Observers. After the 1882 transit, many skygazers sent reports to the U.S. Naval Observatory that are still preserved in the National Archives.

The 2004 Observer's Handbook of the Royal Astronomical Society of Canada has gone to the trouble of detailing the average frequency of cloud cover at the time of the transit for locations around the globe. According to the handbook, the best observation spots are in Iraq, Saudi Arabia and Egypt, with the optimal site being Luxor, Egypt, which has a 94 percent chance of clear weather based on historical records. For this reason, at least one cruise line is heading for the Nile.

Just as the 1882 transit stirred the celestial interests of the young George Ellery Hale and Henry Norris Russell—two astronomical pioneers of the 20th century—perhaps the 21st-century transits of Venus will also encourage young people to study astronomy. Hoping to make the most of the educational opportunities, NASA's Office of Space Science is sponsoring a wide array of events designed to involve students and the general public. A consortium of European institutions has made similar plans. What is more, the "Transit of Venus March," composed by the legendary American composer John Philip Sousa after the 1882 transit, has been resurrected and is being performed with increasing frequency after going unplayed for more than 100 years.

Extrasolar Transits

MEANWHILE PROFESSIONAL astronomers around the world will be celebrating the transit even as they study it. While scientists are training ground-based telescopes and spacecraft instruments at the sun, the International Astronomical Union will hold a meeting near the place where Horrocks viewed the 1639 transit. The IAU's Working Group on Transits of Venus is encouraging the placement of memorial plaques at the sites of previous transit observations.

The transit of Venus still intrigues researchers because it offers a rare opportunity to develop techniques for detecting and characterizing planets in other solar systems. Most of the 120 extrasolar planets discovered to date have been found because their gravity causes small periodic motions in the stars around which they orbit. In 1999, however, astronomers announced the first detection of a planet by measuring a diminution of the light from a star as the planet passed between it and Earth. Located 153 light-years from our solar system, the planet reduced the light from its star by 1.7 percent during the three-hour transit. Unlike conventional planet-finding techniques, transit observations allow astronomers to determine the orbital plane of the extrasolar planet, from which its mass can be deduced. And because the amount of light diminution indicates the size of the planet, scientists can estimate the body's density.

Now NASA is planning to use a spacecraft to find other extrasolar planets by observing their transits. Scheduled for launch in 2007, the Kepler probe will monitor 100,000 sunlike stars over four years. Because photometers can detect tiny decreases in the brightness of a star, the craft will be able to discover planets as small as Earth. Observations of this year's transit of Venus could help researchers calibrate the instruments for making these breakthroughs.

Thus, the story of the transit of Venus has come full circle from Kepler the man to Kepler the spacecraft. Newcomb, Harkness and their contemporaries would surely be amazed at the progress of astronomy since the last transit in 1882. And what will be the state of science and civilization when Venus again approaches the sun in 2117? It is quite possible that by that time a transit of Earth will have been observed from Mars, as foretold by Arthur C. Clarke. If there are humans on Mars on November 10, 2084, they will see their home planet move slowly across the face of the sun, a black dot against a brilliant background. It will surely be a poignant moment and another milestone in the history of planetary transits and human exploration. SA

MORE TO EXPLORE

June 8, 2004: Venus in Transit. Eli Maor. Princeton University Press, 2000.

The Transit of Venus: The Quest to Find the True Distance of the Sun. David Sellers. Magavelda Press, 2001.

The Transits of Venus. William Sheehan and John Westfall. Prometheus, 2003.

More information on the transits of Venus can be found at **www.transitofvenus.org** and **sunearth.gsfc.nasa.gov/eclipse/transit/transit.html**

NASA's educational and public outreach plans are detailed at **sunearth.gsfc.nasa.gov/sunearthday/2004/index_vthome.htm**

European plans are described at **www.eso.org/outreach/eduoff/vt-2004/index.html**

U.S. NAVAL OBSERVATORY LIBRARY

LASER EYE SURGERY

Clear Favorite

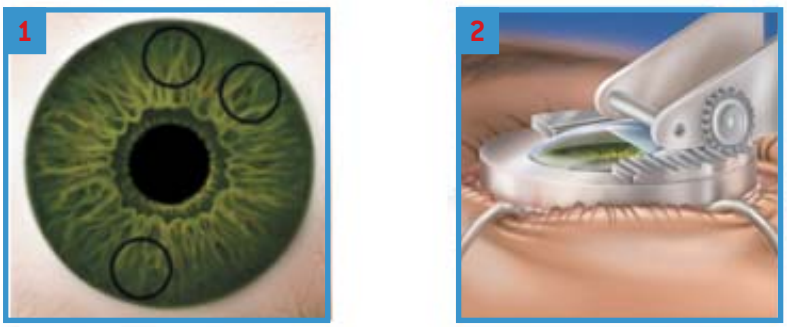
Since excimer laser eye surgery was approved by the U.S. Food and Drug Administration in 1995, it has soared in popularity. Last year more than 1.5 million nearsighted, farsighted or astigmatic people underwent the procedure to eliminate the need to wear eyeglasses or contact lenses.

Several laser-correction schemes exist, but laser-assisted in situ keratomileusis (Lasik) is by far the frontrunner. The procedure reshapes the cornea by vaporizing cells so that light focuses onto the retina properly. Up to 8 percent of patients develop minor complications, among them poorer night vision and visual distractions such as glare or halos, which may disappear after a few months or can be improved with a second treatment. Less than 1 percent develop severe conditions such as infection or scarring.

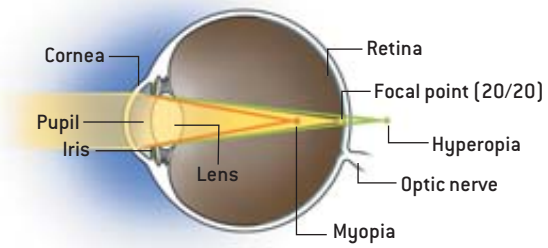
Fully corrected vision may not last forever, though. Ophthalmologists have only 10 years of data. Most of the early patients "appear to retain their full correction, but a few began to regress after eight or five or even three years," says Douglas D. Koch, an ophthalmology professor at the Baylor College of Medicine. Regression is usually mild and caused by natural changes in the eye. In most cases, a laser fix can be repeated, but each surgery thins the cornea, which should not be trimmed to less than 250 microns. Any thinner, Koch says, and the cornea may develop an irregular curvature because it cannot support itself.

Competition has pushed prices down to $1,000 per eye. Cheaper discount providers have sprung up, but ophthalmology associations worry that patients might be misled or receive poor care. (The FDA offers advice at www.fda.gov/cdrh/lasik) Other procedures include photorefractive keratectomy (PRK) and laser epithelial keratomileusis (Lasek), which avoid certain Lasik side effects such as dry eyes but may involve more initial discomfort and recuperation time.

The latest advance is wavefront-guided Lasik. It allows a surgeon to ablate specific points on each person's eye instead of implementing a generalized fix, as is done with standard Lasik. Wavefront technology has been shown to provide better vision than regular Lasik, but it can increase the cost by $400 or more for each eye treated. —*Mark Fischetti*



LASIK SURGERY begins with anesthetic drops that numb the eye. A surgeon then places registration marks on the cornea (*1*). A suction ring immobilizes and pressurizes the eye so it can be cut cleanly by a motorized blade (*2*) that slices into the cornea, creating a flap about eight millimeters in diameter and 0.15 millimeter thick. (In a new procedure, a laser makes the cut.) The flap is pulled back, exposing the stroma. A laser vaporizes cells to a certain depth (*3*), reshaping the cornea in 60 seconds or less. The laser emits pulses of 193-nanometer ultraviolet light to ablate cells to an accuracy of 0.25 micron. The surgeon repositions the flap (*4*), which rebonds naturally.



CLEAR VISION occurs when the cornea focuses light rays exactly on the retina. In myopia (nearsightedness), the cornea is too steep or the eyeball is too long; although diverging rays coming from close objects converge at the retina, parallel rays from distant objects convene too early. Vaporizing the center of the cornea to flatten it fixes the problem. In hyperopia (farsightedness), the cornea is too flat or the eyeball is too short; parallel rays from distant objects focus behind the retina, and diverging rays from near objects are even farther behind. Vaporizing a ring of cells gives the cornea the needed, steeper slope.

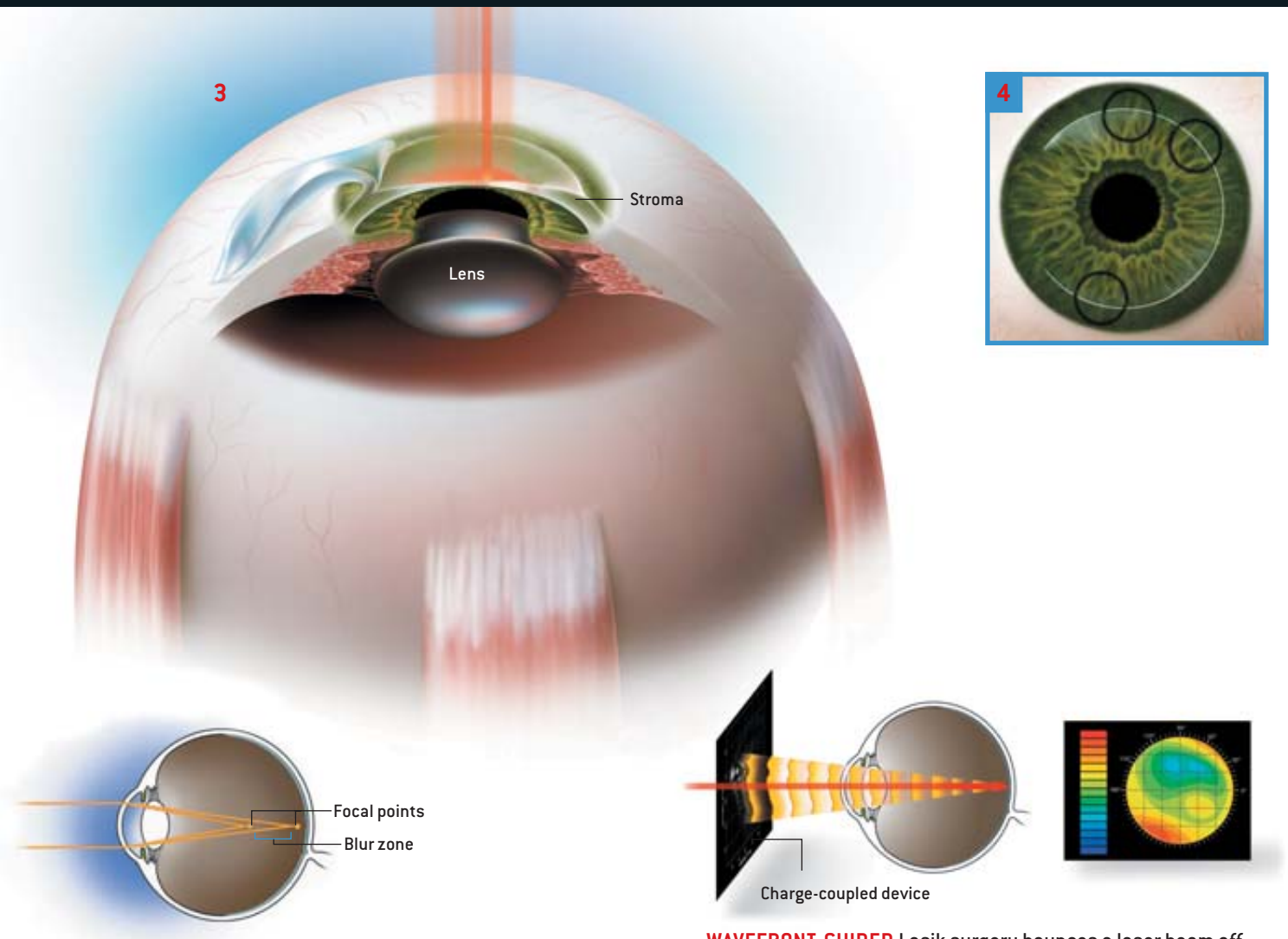
CONNIE FUNKHOUSER BALEK *Precision Graphics*

DID YOU KNOW . . .

➤ BETTER ONE: Eye doctors determine prescriptions with the subjective, decades-old process of sliding different glass lenses in front of a patient's eyes and asking if a chart of letters looks "better with lens one or better with lens two." Laser wavefront sensors approved to guide Lasik surgery are being adapted for more objective measurement. They sample numerous points on the eye, leading to diagnoses that are 50 times as accurate.

➤ SUPER-VISION: Good vision is labeled 20/20—a person sees objects 20 feet away as they should appear (at 20/40, the person must stand at 20 feet to see what normal eyes see at 40 feet). But the density of light-sensing cones in the retina would allow 20/8 vision (more than twice as sharp) if every cornea aberration could be eliminated. Advanced wavefront-guided lasers recently approved could approach that goal. "They are finding distortions we didn't know existed," says Daniel Durrie, director of refractive surgery at Durrie Vision in Overland Park, Kan., "and they can tell surgical lasers how to correct them." Super-vision might be possible—unless the procedure creates unforeseen distractions such as distorted color perception.

➤ HELLO, READING GLASSES: Tiny muscles push and pull the eye's crystalline lens to bring objects into focus. As people age, the lens loses elasticity, making it difficult to zoom in on small objects close at hand. By age 45 virtually everyone has this degradation, which stabilizes in another 10 to 20 years when the lens simply loses all flex. The condition is called presbyopia—"old eye." It cannot be prevented.



ASTIGMATISM (blurry vision) results when the cornea has uneven regions of curvature, which focus rays at multiple points. Smoothing the surface helps to bend rays uniformly.

WAVEFRONT-GUIDED Lasik surgery bounces a laser beam off the retina and senses the reflections on a charge-coupled device. Software maps the distorted rays caused by ocular aberrations (*such as blue region, right*) as small as 0.05 micron and directs the laser to vaporize specific points on the stroma to compensate for each error. In regular Lasik, the surgeon measures the cornea with traditional instruments and the laser ablates a standard, symmetric region to provide a good but generalized correction.

Send topic ideas to workingknowledge@sciam.com

In the Land of the Dreamtime

VISITING PETROGLYPHS, PRISTINE MARSHES AND THE DEEP PAST IN THE VAST WILD OF AUSTRALIA'S KAKADU NATIONAL PARK **BY W. WAYT GIBBS**

In astronomy, distance equals time— the farther we peer into space, the deeper we see into the past. As I watch the ocher rays of a setting sun add meaty color to skeletal figures painted eons ago on a giant rock called Ubirr, it occurs to me that geography can work that way, too. If I were to locate my house on a globe and spin the sphere round to the antipodal point, my finger would hover close to Kakadu National Park, a primeval jumble of wetlands, cliffs and forests punctuated by huge boulders that bear some of the oldest and most impressive Aboriginal artworks known. A journey of more than 24 hours at jet speed brought me only to Darwin, still a half-day's drive from the World Heritage Area in the remote Australian outback.

As I head east from Darwin, signs of civilization evaporate. The FM radio stations fade out, then the AM, until the radio offers just static. Near the park entrance, enormous termite mounds rise from the grasslands. Some, just a few paces off the road, stand six meters high. It is July—the dry midwinter—so the nests are mostly quiet now, but come wet season the mounds will erupt, spewing black clouds of winged insects.

At the bridge over the South Alligator River, signs warn travelers to beware the crocs that stalk its banks. Sound advice, I decide, and turn out instead at a trailhead for a three-kilometer path around the Mamukala wetlands. An observation blind opens onto the stunning 25-kilometer-long marsh; I was not aware that pristine wetlands of this magnitude still existed anywhere. As I walk back to the car, an intimidating wallaroo engages me in a staring contest. It is as long as I am tall. But I don't blink, and it hops off.

TERMITE MOUNDS as tall as houses tower over the grasslands of Kakadu.

MICHAEL S. YAMASHITA *Corbis*

Most of Kakadu's almost 20,000 square kilometers (nearly five million acres) are inaccessible by car—especially during the rainy summer months from November through April—and even the landscape near the roads seems virtually untouched by humanity. That is an illusion, ranger Alex Dudley points out the next afternoon as he guides a walking tour at the base of Nourlangie Rock. "This is not a wilderness," Dudley asserts. "This place has been home to people for 50,000 years."

The local culture holds that Nayuhyunggi, the "first people," arrived in Kakadu during the Dreamtime, a creation period when supernatural beings emerged from deep in the earth. Some of these creation ancestors, Dudley explains, ended their journeys by transferring themselves onto rock walls, leaving impressions that perceptive artists enhanced with natural pigments. Aboriginal people who live here say that creation ancestors still rest in the southern part of the park. Travelers to that area, called the "sickness country," are urged to tread with great caution so as not to wake the sleeping immortals.

As he speaks, Dudley turns and gestures at the wide, shallow cave behind him. Every smooth vertical surface is covered in drawings of red, white, yellow and orange, spanning an impressive

range of styles. A few are simple outlines of pigment blown over an outstretched hand: the Aboriginal "Kilroy was here." Others are "x-ray" renderings of skeletal fish diving for bottom. Still others portray sticklike figures dancing, their heads triangles, their tongues wagging.

"We don't know a lot about the stories behind these paintings," Dudley says wistfully. The Gagadju people who drew most of the art here have since died off. "But some of the paintings in this area are quite new, and the artists have told us that the stories they represent have many, many layers. You see, there is no such thing as free knowledge in Aboriginal society. You are only given the full details of the story once you have passed several levels of initiation."

Dudley clambers up the boulders. The sheer face of Nourlangie Rock rises 100 meters behind him. There are no fossils in the sandstone here, he notes, because it formed before macroscopic life evolved, nearly two billion years ago. "The freshwater billabongs that are such a big part of Kakadu today are relative newcomers; they formed around 1,500 years ago," he adds. It is a testament to the adaptability of life that such a galaxy of species has coalesced within the wetlands in such a short time. For at least part of the year, the park is home to 280 species of birds, which makes life an adventure for the 46 species of freshwater fish.

Out on a 300-meter-long catwalk over the Yellow Water Billabong, an excited gaggle of birdwatchers track a sea eagle circling overhead with a fish in its beak, with another eagle in hot pursuit. The first eagle delivers his meal to a mate waiting in a nest the size of a queen mattress, then turns and upbraids his pursuer. Binoculars are passed around so that all can enjoy a glimpse of a jabiru, an odd kind of stork that looks like a pelican on stilts.

Otherworldly animals are everywhere. I drove by a frilled lizard sunning in the road, neck wing fanned to its fullest. Eleven varieties of dragons and monitors live in Kakadu. Here the grasshoppers are blood-red or mustard-yellow. "We've got the world's most venomous spider, tree frog and octopus," Dudley boasts. "The pythons in these parts stretch from here …," he takes four paces, "… to here." Crocodiles lurk even in the freshwater ponds, and scorpions scuttle over the rocks. And yet, Dudley avers, "you are more likely to be killed here by a European honeybee than by any native animal."

Toward sundown, I zip over to Ubirr

PAUL A. SOUDERS *Corbis (left)*; GREG MILES *Environmental Media (right)*

ABORIGINAL PAINTINGS of creation ancestors decorate Nourlangie Rock; some are 20,000 years old. The image above depicts Namarrgon (*at right*), who creates lightning and thunder with axes on his limbs. Just a few kilometers away, Jabiru storks can be seen hunting frogs in Anbangbang Billabong.

Rock. Some visitors pause on the trail to gaze, mystified, at intricate images of sorcerers traced on an overhanging shelf 10 meters above the ground. But most scramble straight up Ubirr's flank, over cool gray rock the texture of a crocodile's back, up 250 meters to reach its flat top.

A sweeping, stunning panorama unfurls below. I spin around slowly to take it in. To the west lie two placid billabongs and a swampy forest. To the south roll green woodland hills. In the east is stone country, grassland dominated by layered pillarlike boulders. And to the north, a landscape unlike any I have ever seen, where improbably stacked slabs of gold-gray sandstone rise from reeds dotted with egrets in a marsh ringed by prehistoric-looking palms and dense eucalyptus.

As the sun sinks into the horizon, its rays inflaming lacy clouds, the confluence of worlds takes on a supernatural glow. For a long moment, the past seems present, and the Dreamtime seems as plausible as the big bang. Then the sun disappears, and the moment passes. Spontaneously, the crowd applauds.

Kakadu National Park is 170 kilometers east of Darwin and a similar distance north of Katherine. For maps and information, see www.deh.gov.au/parks/kakadu or call the Bowali Visitor Center at +61-8-8938-1120. SA

AI at the Inception

A 25TH-ANNIVERSARY EDITION OF A CLASSIC CHRONICLES THE FLEDGLING SCIENCE OF ARTIFICIAL INTELLIGENCE **BY HENRY FOUNTAIN**

MACHINES WHO THINK: A PERSONAL INQUIRY INTO THE HISTORY AND PROSPECTS OF ARTIFICIAL INTELLIGENCE
By Pamela McCorduck
A K Peters, Natick, Mass., 2004 ($19.95)

The review you are reading was written by a human, not a machine. This fact would no doubt disappoint some of the pioneers of artificial intelligence, who would have thought that by the 21st century a computer would be able to read a book, consider it in the context of other knowledge and express some thoughtful opinions about it.

On the other hand, the human who wrote this review was aided in researching and preparing it by telecommunications and computer networks, including the Internet, that owe a big part of their existence—and even more of their smooth functioning—to theories and concepts that arose from artificial-intelligence research.

The enormous, if stealthy, influence of AI bears out many of the wonders foretold 25 years ago in *Machines Who Think*, Pamela McCorduck's groundbreaking survey of the history and prospects of the field. A novelist at the time (she has since gone on to write and consult widely on the intellectual impact of computing), McCorduck got to the founders of the field while they were still feeling their way into a new science. Her novelist's eye for detail and ear for style formed a book that this magazine's review of the first edition described as "delicious."

When *Machines Who Think* was first published in 1979, it was an up-to-the-moment history. But in a digital world, that moment was an eternity ago, so McCorduck has appended a 30,000-word afterword to bring the reader up-to-date. The original text has been wisely left unaltered (including a few passages that now seem quaint, such as the explanation of the difference between hardware and software).

Her story begins long before the advent of computing, in ancient thinking about the human need to make something in our own image. McCorduck sees AI research as the continuation of a long tradition of thought, encompassing everything from the Ten Commandments' prohibition against idols to Mary Shelley and her Frankenstein monster.

But the book, like the field, really doesn't begin to take off until computing machines—mechanical at first, then eventually digital—enter the picture. McCorduck details the thoughts of theorists such as Alan Turing (who believed machine intelligence was possible) and John von Neumann (who didn't) and devotes considerable space to work on chess- and checkers-playing machines, which was the early public face of AI. She notes seminal events, particularly the Dartmouth Conference, a 1956 workshop where much of the groundwork for future research was laid by such men as Marvin Minsky, John McCarthy and two upstarts who would be hugely influential, Alan Newell and Herbert Simon.

Newell and Simon were in large part responsible for a shift in thinking away from the idea that machine intelligence must mimic the brain physically, an approach that drew parallels between neurons and digital devices, and toward the view that it should simulate human thought processes—what became known as the information-processing model. McCorduck shows how this idea developed over the years, how problems that were first seen as "impossibly nonme-

NURSEBOT PEARL, currently being developed by a team from Carnegie Mellon University and elsewhere, is a fusion of many AI technologies—speech understanding, computer vision, dialogue management, embedded sensors, mobility, planning, scheduling, and so on. The purpose of the robot is to help older people stay in their homes several years longer than they might otherwise be able to do.

NURSEBOT PROJECT

chanical" were solved and how these solutions "slowly began to be brought into the domain of ordinary computational processes."

That slow infusion of AI into everyday computing picked up speed after 1979, and in the afterword McCorduck gives a taste of these advances and of recent research in robotics, natural-language processing and other fields that are, in essence, AI spin-offs. This part of the book feels sketchy, and the author acknowledges that it is not meant as a definitive survey of the field's past 25 years. But the reader is left wanting more.

Still, taken together, the original and the afterword form a rich and fascinating history. Along the way, McCorduck introduces us to some interesting characters, not the least of whom are the naysayers. She devotes a chapter to Hubert Dreyfus, the philosopher who in the 1960s became a thorn in the side of researchers with his public pronouncements about the futility of their work (they had the last laugh, however, when a machine beat him at chess). And she writes about those thinkers, most recently the technologist Bill Joy, for whom the great hopes of AI have been replaced by great fears, of machines that might rule rather than rival humans.

The book is described as a "personal inquiry," and now, as then, McCorduck leaves little doubt as to where her personal allegiance lies. From the title to the very last sentence, she is a believer in what she calls a "heroic enterprise." She may admit that researchers have a long way to go, but she dismisses the doubters as well: AI, she writes, is "neither the field of dreams nor the field of nightmares portrayed." Were she to produce a 50th-anniversary edition in 2029, she might be somewhat surprised, but surely very pleased, to see it reviewed by a machine who thinks. SA

Henry Fountain is a writer and editor at the New York Times, *specializing in science and technology.*

REVIEWS

THE EDITORS **RECOMMEND**

FOR LOVE OF INSECTS

By Thomas Eisner. Harvard University Press, 2003 ($29.95)

ANT ENTANGLED **by the bristles of a millipede. An even closer look (*right*) shows the "grappling hook" part of one of the bristles fastened to an ant's hair.**

Among the many wondrous tales that Eisner relates in this memoir of his research on insects is that of a tiny millipede (a polyxenid) that defends itself by coating its attacker—usually an ant—with bristles. Scanning electron micrographs taken by Maria Eisner, co-worker and wife of Thomas Eisner, show how the entangling mechanism works. The bristle tips are grappling hooks that become fastened to the ant's hairs. To make matters worse, barbs on the bristle shafts cross-link the bristles, creating a loose meshwork that muzzles the ant and strings its legs together. After observing an attack, Eisner wrote that the ants "attempted to clean themselves, but in so doing seemed only to aggravate their plight. They wiped antennae with forelegs, drew appendages through the mouthparts, or stroked legs against one another, but they usually succeeded only in further entangling themselves.... Many lost their footing and fell to the side, without ever recovering.... The polyxenids, without exception, survived the encounters."

Unlike the polyxenids, most of the insects Eisner has studied use chemicals to defend themselves. In fact, his discoveries of these defenses, beginning in the 1950s just after he earned his doctorate from Harvard University, helped to found a new field of biology, chemical ecology. He has, ever since, been busy making new discoveries about these surprising strategies in the field and in laboratory experiments at Cornell University, where he is J. G. Schurman Professor of Chemical Ecology. The findings he describes are intriguing—all the more so in that they provide the scaffolding on which we see at work the mind of one of our most distinguished scientists and naturalists.

Exquisitely illustrated with photographs, most taken by Eisner, who is widely admired for his photography, the book is written in a style that is conversational, witty and graphic. Beautiful to look at and beautiful to read.

The books reviewed are available for purchase through www.sciam.com

MARIA EISNER

Jump Snatch

BY DENNIS E. SHASHA

Let's imagine a game involving a tic-tac-toe board and several circular counters that can be placed in the grid's squares. Assume the following simple rules: a counter can be jumped if it lies between another counter and an empty square (along any vertical, horizontal or diagonal line). When a player uses one counter to jump another, the latter is removed from the board, as in the game of checkers.

In the solitaire version of this game, your goal is to have only one counter left on the grid after some number of jumps. Consider the starting configuration shown in illustration A. Is there any way to ensure that only one counter will be left at the end of the game? Illustrations B, C and D show a solution.

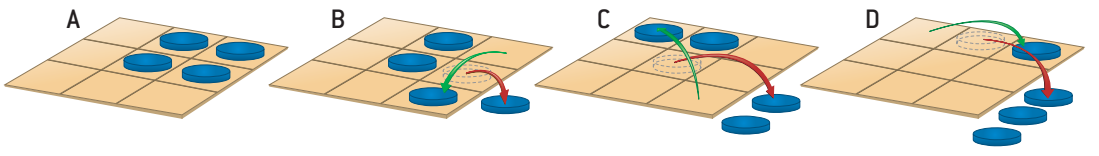
Your first challenge is to answer two questions about the game: To guarantee having only one counter at the end, what is the minimum number of squares that you must leave empty at the start, and where should they be on the grid? And if the tic-tac-toe board is four-by-four instead of three-by-three, how many squares must you leave empty, and where should they be?

Now let's consider a two-player version of the game, which I call Jump Snatch. To start this version, put counters on all the squares of the three-by-three grid. The first player, called the Snatcher, removes a counter from any square. The second player, called the Jumper, then makes a jump if he can and has the option of making additional jumps if they are possible. When the Jumper is finished, the Snatcher attempts to make his own jumps, and the moves alternate until one player wins the game by making a jump that leaves only one counter on the board. If, at the start of a move, a player faces a configuration in which no jump is possible, he must slide a counter to the center square; if this move is also not possible, he may slide any noncenter counter to any neighboring square.

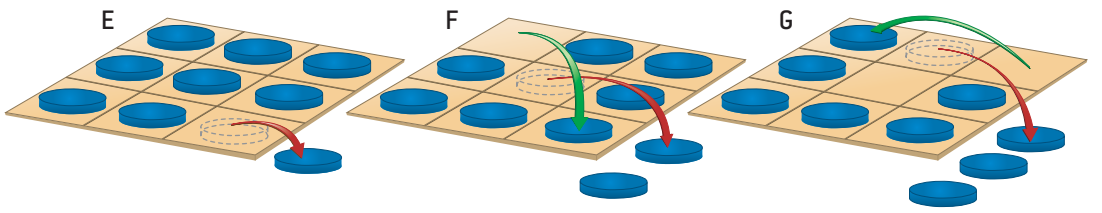
Illustrations E, F and G show the first three moves of a Jump Snatch game. Who will win this contest, assuming optimal strategy on both sides? (The answers are on page 16. This Puzzling Adventures column will be the last to run in the magazine, but future installments will continue to appear on *Scientific American*'s Web site: www.sciam.com) SA

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

SOLITAIRE VERSION



TWO-PLAYER GAME



Snatcher removes one counter

Jumper makes one jump

Snatcher makes one jump

Answers to Last Month's Puzzle:
The Bluffhead problems have these solutions:

1. Jordan has a king, and the others have lower cards.

2. David and Jordan have kings, and Caroline has a lower card.

3. Caroline has a king.

4. When Caroline says, "I lose," in the second round, we know Jordan has a queen and the other two have lower cards. David says, "I lose," and Jordan says, "I win."

5. Jordan has a 6, Caroline has a 5, and David has a 4 or 5.

Web Solution
For a full answer to last month's problem, visit www.sciam.com

JOHNNY JOHNSON

Television Coverage

A MODEST PROPOSAL FOR SMALL SCREENING IN MEDICINE **BY STEVE MIRSKY**

In January a Romanian woman underwent surgery in a Bucharest hospital to remove a 175-pound tumor. News reports quoted a plastic surgeon at the hospital as having delivered the startling revelation that "the lack of the tumor really suits her."

Of course, 175-pound tumors don't grow overnight. And the woman had apparently tried for years to raise money for the operation. The Discovery Channel finally forked over the funding, in exchange for film rights.

Finding money for medical treatment can also be a problem in the U.S. This past February saw the release of the *Economic Report of the President,* which noted that more than 43 million people in this country lack health insurance. The report also stressed that "U.S. markets provide incentives to develop innovative health care products and services that benefit both Americans and the global community."

Keeping those sentiments and the Romanian tumor case in mind, one solution becomes obvious. Uninsured patients, who have not appreciated that their diseases are in fact valuable market commodities, could sell their conditions to television programs, which would pay for medical treatment.

In that spirit, here are some suggestions for the fall lineup of new series: SA

NEW PROGRAMS GUIDE

Everybody Loves Radiology. A dysfunctional family is crammed into a magnetic resonance imager to see who can stay in the longest. The last one left gets scanned and treated if the MRI finds anything funny.

American Eye Doc. Glaucoma patients do cost-benefit analyses of getting their meds either through pharmacies or from a guy called Spliffy the Bongmeister.

E.R.R. An attorney has complete access to a public hospital's medical records for one hour to find the best malpractice case. The patient, if living, then gets to choose: sue, settle or a "do-over" at a private hospital.

Just Don't Shoot Me. Twelve unrestrained four-year-olds are put into a room with a pediatrician who has 11 doses of DTP vaccine.

The Simple Life-Threatening Emergency. A full checkup is the prize as uninsured contestants attempt to use the Heimlich maneuver to dislodge a foreign object from Paris Hilton.

Let's Make a Drug Deal. The audience watches as a patient with multiple preexisting conditions gets to choose pharmaceutical treatment for only one.

American Choppers. A panel of judges rates octogenarians as they eat corn on the cob and bob for apples, with the winner receiving a full set of new dentures.

Barely Live with Regis and Kelly. One lucky audience member gets medical care—if he or she is sitting in the seat with the same number as the one chosen at random by a caller.

N.Y.P.D. Code Blue. A police car has 30 minutes to get a patient with chest pains from Manhattan's Lower East Side to the Upper West Side, with treatment guaranteed if they make it. **Warning:** May contain nude images of Dennis Franz, which should be viewed only by contestants from *American Eye Doc.*

Dr. Timothy Johnson's Jackass. Johnson, the ABC News medical editor, personally treats kids injured jumping off their roofs, riding shopping carts down hills or imitating pro wrestlers.

CSI: Bethesda. An uninsured elite forensics team tries to determine how a private pharmaceutical company got a proprietary interest in a product created through publicly funded research at the National Institutes of Health.

The Price Is Nuts. People who actually have insurance but still can't afford their 50 percent co-pay on mental health care try to guess the cost of an hour-long session with their nearest competent therapist without going over.

Parasite Island. Six 18- to 34-year-olds dine at an all-you-can-eat discount sushi bar and then evaluate the proposition in the *Economic Report of the President* that some young people "may remain uninsured because they are young and healthy and do not see the need for insurance."

PHOTOGRAPH BY FRANK VERONSKY

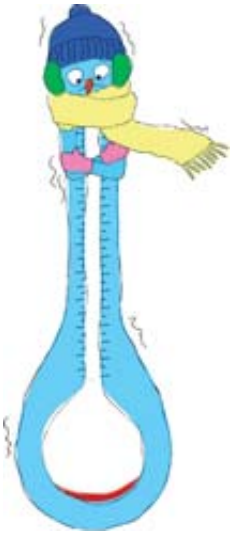
ASK THE EXPERTS

How are temperatures close to absolute zero achieved and measured?

Wolfgang Ketterle of the Massachusetts Institute of Technology, who won the Nobel Prize in Physics in 2001 for his work with ultracold atoms, explains:

First, let me introduce the scientific meaning of temperature: it is a measure of the energy content of matter. When air molecules are hot, they move fast and have high kinetic energy. The colder the molecules are, the lower their velocities and the less energy they have. Absolute zero corresponds to zero kelvins (–273 degrees Celsius or –460 degrees Fahrenheit).

Cooling requires extracting energy from an object and depositing that energy somewhere else. By combining laser cooling and evaporative cooling, scientists have been able to achieve temperatures in clouds of atomic gases below one nanokelvin (one billionth of a kelvin). The current record, described by our group in the September 12, 2003, issue of *Science,* is 450 picokelvins (half a billionth of a kelvin).



In laser cooling, the target atoms scatter laser light. An incoming laser photon is absorbed and then reemitted in a different direction. On average, the color of the scattered photon is slightly shifted to the blue relative to the laser light. That is, a scattered photon has a slightly higher energy than does an absorbed photon. Because total energy is conserved, the difference in photon energy is extracted from the atomic motion—the atoms slow down.

As an atomic cloud becomes denser and colder, the cooling effect becomes dominated by other processes, which still result in some trembling motion of the atoms. The processes include energy release from collisions between atoms and the random recoil kicks in light scattering. At this point, however, the atoms are cold enough to be confined by magnetic fields. We choose atomic species that have an unpaired electron and therefore a magnetic moment. These atoms behave like little bar magnets. External magnetic fields levitate the atoms against gravity and keep them together; in effect, the fields form invisible walls that contain the atoms in a magnetic cage.

Evaporative cooling can then selectively remove the most energetic atoms from the system. In a magnetic trap, the most energetic atoms can move farther against the pull of the magnetic forces and can reach regions with higher magnetic fields than can the colder atoms. When the atoms encounter those higher magnetic fields, they get into resonance with radio waves or microwaves, which changes the magnetic moment in such a way that the atoms escape from the trap.

How do we measure very low temperatures of atoms? One way is simply to look at the extension of the cloud. The larger the cloud, the more energetic its atoms must be, because they can move farther against the magnetic forces. Another method is to measure the atoms' kinetic energy. The magnetic trap is switched off. In the absence of magnetic forces, the atoms fly away, and the cloud expands ballistically. The cloud size increases with time, and this increase is a direct way to observe the velocity of the atoms and, hence, their temperature. When a smaller cloud is observed after a fixed time of expansion, that change indicates the achievement of lower temperature.

If heat rises, why is air cooler at higher elevations?

Paul B. Shepson, professor of atmospheric chemistry at Purdue University's School of Science, provides this answer:

In the earth's atmosphere, pressure, which is related to the number of molecules per unit volume, decreases exponentially with altitude. Therefore, if a parcel of air from the surface rises (because of wind flowing up the side of a mountain, for example), it undergoes an expansion, from higher to lower pressure. When air expands, it cools. This phenomenon is familiar to everyone—stick your finger on the valve of a car tire and let some air escape. It is not cool inside the tire, but as the air comes out it expands and thus cools. SA

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

SARA CHEN