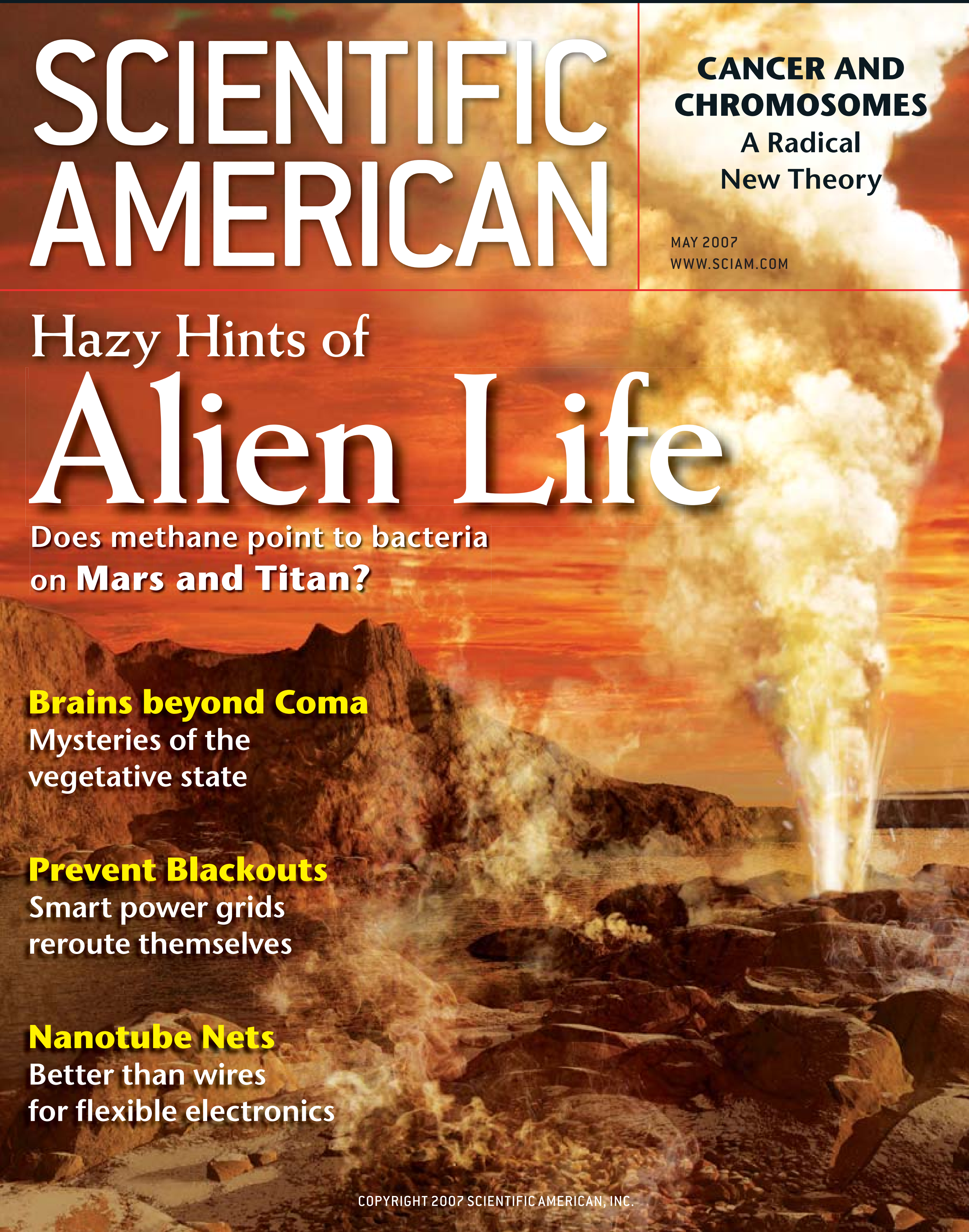
DO-IT-YOURSELF QUANTUM MECHANICS • LOST MAMMALS REWRITE HISTORY

SCIENTIFIC AMERICAN

MAY 2007
WWW.SCIAM.COM

CANCER AND CHROMOSOMES
A Radical New Theory



Hazy Hints of Alien Life

Does methane point to bacteria on **Mars and Titan?**

Brains beyond Coma
Mysteries of the vegetative state

Prevent Blackouts
Smart power grids reroute themselves

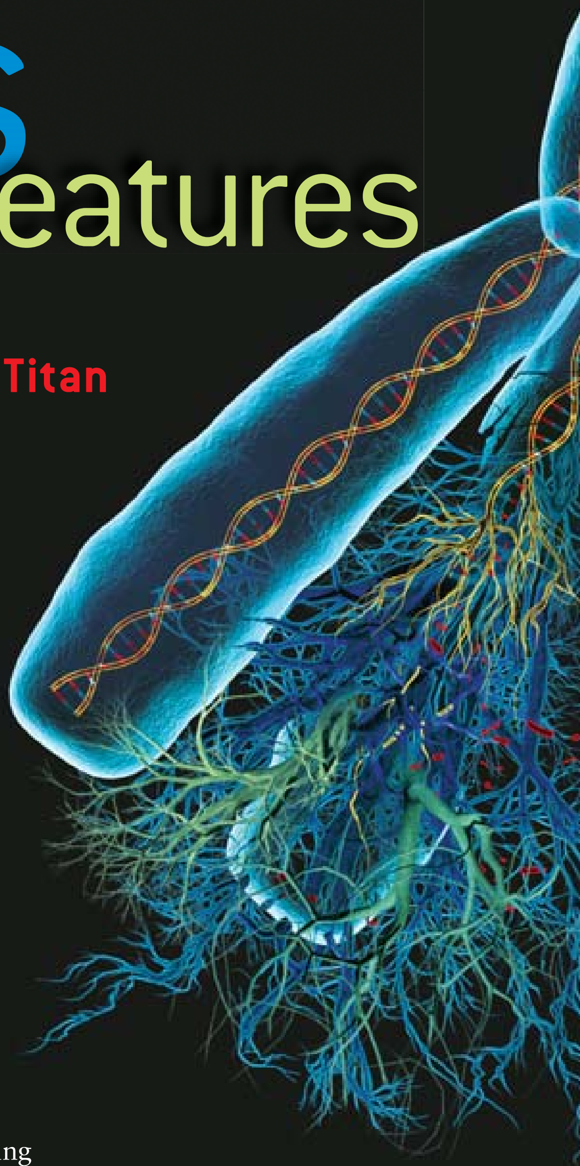
Nanotube Nets
Better than wires for flexible electronics

may 2007

contents

features

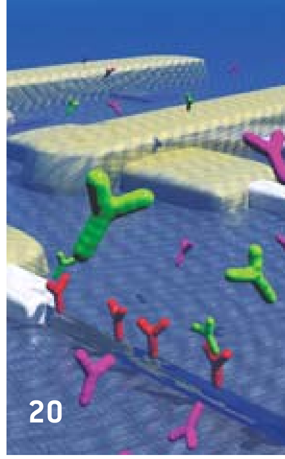
SCIENTIFIC AMERICAN Volume 296 Number 5

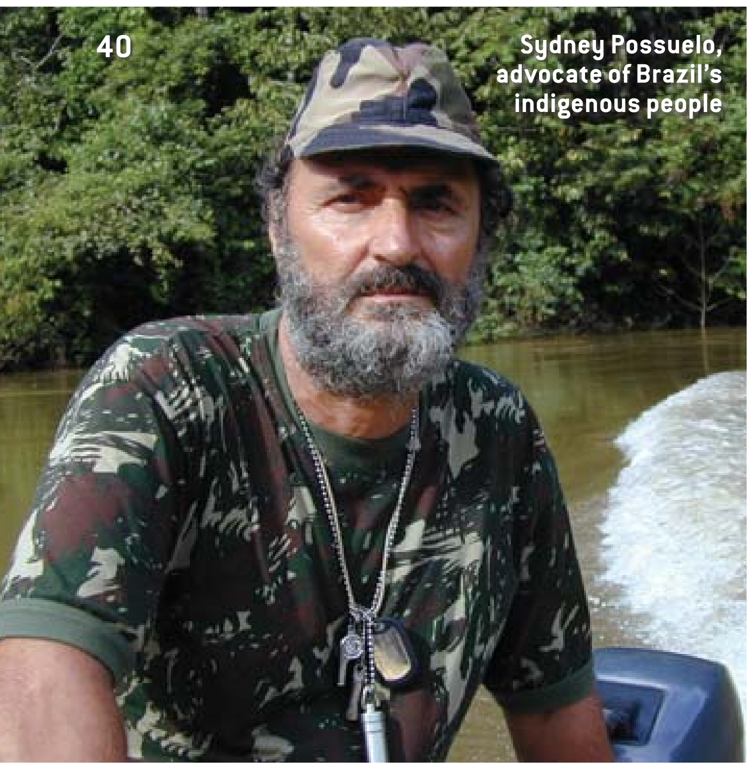


52 Do faulty chromosomes trigger cancer?

SCIENTIFIC AMERICAN Volume 296 Number 5

departments





Sydney Possuelo, advocate of Brazil's indigenous people

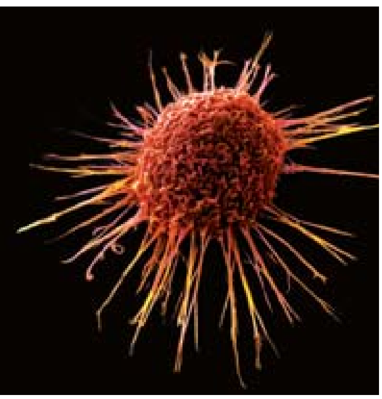
columns

Cover image by Ron Miller; photograph at left courtesy of Sydney Possuelo; illustration on page 5 by Kenn Brown.

Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2007 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced or transmitted in any form or by any means, electronic or mechanical, including photocopying and recording for public or private use, or by any information storage or retrieval system, without the prior written permission of the publisher. Periodicals postage paid at New York, N.Y., and at additional mailing offices. Canada Post International Publications Mail (Canadian Distribution) Sales Agreement No. 40012504. Canadian BN No. 127387652RT; QST No. Q1015332537. Publication Mail Agreement #40012504. Return undeliverable mail to Scientific American, P.O. Box 819, Stn Main, Markham, ON L3P 8A2. Subscription rates: one year $34.97, Canada $49 USD, International $55 USD. Postmaster: Send address changes to Scientific American, Box 3187, Harlan, Iowa 51537. **Reprints available:** write Reprint Department, Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111; (212) 451-8877; fax: (212) 355-0408. **Subscription inquiries: U.S. and Canada (800) 333-1199; other (515) 248-7684. Send e-mail to sacust@sciam.com** Printed in U.S.A.

When Pariahs Have Good Ideas

Even mentioning the name Peter Duesberg inflames strong feelings, both pro and con. After gaining fame in 1970 as the virologist who first identified a cancer-causing gene, in the 1980s he became the leading scientific torchbearer for the so-called AIDS dissidents who dispute that HIV causes the immunodeficiency disorder. To the dissidents, Duesberg is Galileo, oppressed for proclaiming scientific truth against biomedical dogma. A far larger number of AIDS activists, physicians and researchers, however, think Duesberg has become a crank who refuses to accept abundant proof that he is wrong. To them, he is at best a nuisance and at worst a source of dangerous disinformation on public health.



UNORTHODOX cancer theory bears consideration.

Readers may therefore be shocked to see Duesberg as an author in this month's issue. He is not here because we have misgivings about the HIV-AIDS link. Rather Duesberg has also developed a novel theory about the origins of cancer, one that supposes a derangement of the chromosomes, rather than of individual genes, is the spark that ignites malignant changes in cells. That concept is still on the fringe of cancer research, but laboratories are investigating it seriously. Thus, as wrong as Duesberg surely is about HIV, there is at least a chance that he is significantly right about cancer. We consider the case worthy of bringing to your attention, with the article beginning on page 52.

Thousands of scientific papers appear in technical journals every month; why do some rate more fame and journalistic attention? It helps for science news to have dramatic relevance to human affairs: Is there strong new hope for curing a disease, transforming the economy, building a better mousetrap? Alternatively, reporters and editors may gravitate toward new science that easily inspires the public's sense of wonder, as so many astronomy stories do. And reports that appear in certain major scientific journals tend to get more play because those publications have a self-fulfilling reputation for releasing the most noteworthy papers. (It doesn't hurt that those journals have particularly strong public relations departments, too.)

When we look at submitted manuscripts from scientists, we consider it a reassuring sign when the authors forthrightly acknowledge both their collaborators and their competitors and note potential conflicts of interest before we ask. If we see that they are describing the science of their rivals fairly, we can have more confidence that they are being similarly candid about their own work. (Still, the old nuclear disarmament treaty maxim applies: trust, but verify.) We typically steer away from controversial ideas too new to have much supporting evidence. Those that have lasted for years and accumulated some substantiation have earned consideration. Our judgments are imperfect, but they tend to mirror those of the scientific community.

Blots on a researcher's history often should bear on regard for his or her new work. Scientists who have intentionally published fraudulent papers, as the stem cell researcher Woo Suk Hwang so notoriously did two years ago, may be irredeemably tainted. But to dismiss a scientist solely for holding some wrong or controversial views risks sweeping away valuable nuggets of truth. We respect the opinions of any readers who may criticize our choice to publish Duesberg in this case but hope they will nonetheless evaluate his ideas about cancer on their own merits.

THE EDITORS editors@sciam.com

STEVE GSCHMEISSNER *Photo Researchers, Inc.*

How to Contact Us

EDITORIAL

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

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

SUBSCRIPTIONS

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

REPRINTS

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

PERMISSIONS

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

ADVERTISING

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

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

West/Northwest
818-827-7138
fax: 818-827-7139

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

On the Web

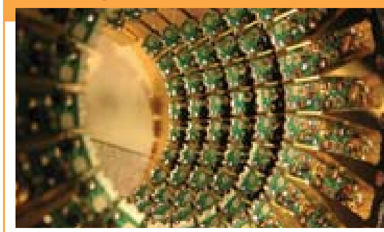
WWW.SCIAM.COM

UPDATED EVERY WEEKDAY

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

NEWS

First "Commercial" Quantum Computer Solves Sudoku Puzzles



Canadian quantum computing company banks on a long-shot form of quantum computation: adiabatic.

BLOG

The One Thing You Need to Know about Quantum Computers

Everything you may think you know about how quantum computers work—and what many science journalists explain in their reporting—is wrong.

SPECIAL REPORT

Happy International Polar Year!

Scientists from around the globe embarked on a two-year effort to explore and boost understanding of the earth's poles on March 1.

FACT OR FICTION?

Living People Outnumber the Dead

Booming population growth among the living, according to one rumor, outpaces the dead.

PUZZLE

Sci-Doku

Try your hand at Sci-Doku, a Sudoku puzzle that uses letters instead of numbers, with an added twist: a science-related clue accompanies each puzzle, and the answer is spelled out in one row or column of the solution.

Only at www.sciam.com

New Podcasts

Every weekday: *60-Second Science*
Every Wednesday: *Science Talk*

Listen at www.sciam.com/podcast

JASPAUL CHUNG AND MURRAY THOM *D-Wave Systems, Inc.*

Letters

EDITORS@SCIAM.COM

READERS RESPONDED with great passion to our January issue. Although Bill Gates's vision of the developing robotics industry instigated much debate, the majority of letters centered on Matthew L. Wald's article and its pessimistic view of corn-based ethanol fuel. Correspondents argued from all sides of this issue, although a large number echoed Joyce Ferree of Salida, Colo.: "Ethanol may not be the perfect fuel, but it is showing Americans that there are alternatives to foreign oil that stimulate rural economies across the U.S., improve our foreign trade balance and don't require our military to keep the supply coming.... Corn ethanol is not a 'bridge to nowhere' if we look at it as a solid foundation for a new industry to build on."

CORN-FED BEEF

In "Is Ethanol for the Long Haul?" Matthew L. Wald speculates that ethanol producers may turn to coal or even electricity for process heat instead of natural gas because that fuel has become expensive. Although a small minority of new ethanol plants use coal instead of natural gas, an encouraging trend appears to be a switch to renewable biomass for processing, which greatly reduces fossil-fuel use. Several new and existing ethanol plants have taken this path, yet Wald makes no mention of it.

Ryan Katofsky
Arlington, Mass.

Naysayers love to jab away at ethanol's alleged negative energy balance. The truth of the matter is that all energy conversions have a negative energy balance. All electric motors, for example, use more energy in electricity than they deliver in horsepower. In developing the energy balance figures cited by Wald, the study's author, Cornell University's David Pimentel, piled on extraneous bloated figures that other researchers have deemed insignificant.

Stan L. Simon
Kerkhoven, Minn.

Most critics and proponents of ethanol grown from corn ignore one major problem: corn is a nitrogen-hungry crop. In addition to the costs of manufacturing and applying chemical nitrogen fertilizer, the fertilization process itself is far from 100 percent efficient, especially when animal waste is used. The resulting pollution from the nitrogen and phosphorus not assimilated by the crop has economic and societal consequences that are real.

Lynton S. Land
Ophelia, Va.

Living near three ethanol plants, I know that it takes more than a gallon of water to produce a gallon of ethanol. Though touted as a "green" fuel in the Midwest, ethanol's effect on the further reduction of our water tables, lakes and rivers, when much of the country has been experiencing droughtlike conditions over the past several years, does not make the fuel green in my book.

Joy Ophoben
Rockville, Minn.

ROBOT RUMBLE

"A Robot in Every Home," by Bill Gates, was little more than a sales pitch for Microsoft. Gates already controls how most of the computers on the planet operate and clearly wants to do the same for the home robotics market. The diversity of robot software providers that he fears will enhance competition and result in better products, whereas a monopoly will yield the same security and reliability issues we currently experience with personal computer software.

Dan Gravatt
Lawrence, Kan.

As someone who has been working on requirements for robots for more than 30 years, I believe that Gates has made a mistake (as have most in the field) in thinking that the main obstacle to progress is a lack of solutions to problems. In truth, the key obstacle is a lack of understanding of what the problems are, how they differ and how they are related.

Most robotics engineers are attempting to design systems without thoroughly analyzing the requirements necessary to make them functional. One example is the need to characterize the many functions of vision. Too many people think of vision as simply the recognition of objects and ignore its role in controlling intricate actions, perceiving and understanding processes, seeing possibilities for future processes, and comprehending causal relations.

ARE ROBOTS ready for the big time?

We can deepen our understanding of requirements by examining in far more detail than roboticists typically do the actual variety of competences displayed by humans and other animals. But simple observation will not reveal the nature of such competences nor what problems need to be solved for a robot to achieve them. Discerning what is being achieved, even in apparently simple actions, often requires deep interdisciplinary analysis based in part on experience in artificial intelligence and philosophy.

Aaron Sloman
Honorary Professor of Artificial Intelligence and Cognitive Science
University of Birmingham, England

RUNAWAY RANGE?

I have recently seen a documentary that describes a theory not mentioned in "The Mississippi's Curious Origins," by Roy B. Van Arsdale and Randel T. Cox. In the film the missing piece of the Ouachita Mountains was believed to have been found ... in Argentina of all places! (The Argentine Precordillera region would constitute the missing piece.) Has this theory been discredited?

Mike Lehrter
Fort Thomas, Ky.

VAN ARSDALE AND COX REPLY: The idea that the southern part of North America rifted away and ended up as the Argentine Precordillera has been around for at least a couple of decades. It is largely accepted, based on fossil and lithologic affinities of sediments dated to the Cambrian period (0.5 billion years ago) found in Argentina and the Ouachita Mountains. These similarities suggest the two terranes (bodies of rock that exhibit distinctive geologic histories) were close together in Cambrian times (that is, they had just rifted apart). Subsequent plate motions put the rifted block in Argentina.

This idea is not in competition with the theory that the Ouachita Mountains were breached when the North American continent moved over the Bermuda hot spot. The Cambrian rifting episode created the Mississippi Valley graben (a depressed segment of the valley's crust bordered by faults) shown in our article.

PLANETARY PLACEMENT

In "What Is a Planet?" Steven Soter dodges a question he himself raises: "Isn't the definition of a planet really arbitrary?" The answer should be a simple yes. Exact definitions are undoubtedly necessary but are rarely perfect reflections of reality. Any classification or categorization of reality imposes arbitrary separations on spectra of experience or objects. What nonarbitrary sense can be read from the definition adopted by the International Astronomical Union: "[a body that] has cleared the neighborhood around its orbit"?

Matt Craver
via e-mail

SOTER REPLIES: Many definitions involve imprecise or arbitrary boundaries. But when nature reveals a large gap in some parameter that might have been continuous, it may be telling us something. By any measure of dynamical dominance, we find an enormous gap between eight bodies and everything else that orbits the sun. This gap provides a physically significant criterion and not an arbitrary place to draw the boundary defining planets. We may have to revise or reject the new definition when we learn more—particularly about other planetary systems—but for now it appears to describe what we know.

ERRATUM On page 46 of "Is Ethanol for the Long Haul?" by Matthew L. Wald, it is stated that a standard barrel (42 gallons) of ethanol contains 80,000 British thermal units (Btu) of energy. The text should have read that a gallon of ethanol contains this amount.

CLARIFICATION "Airborne Baloney," by Michael Shermer [Skeptic], states that it is unsafe to consume more than 10,000 units of vitamin A daily. Per the National Institutes of Health, this limit applies to adults consuming preformed vitamin A, which is derived from animal products. Provitamin A carotenoids, derived from plants, are considered safe in higher units.

Send letters to editors@sciam.com or to Scientific American, 415 Madison Ave., New York, NY 10017. Letters become the property of Scientific American *and may be edited for space and clarity.*

50, 100 & 150 Years Ago

FROM SCIENTIFIC AMERICAN

Automation Phobia ▪ Horseless Hearse ▪ Clean Weeds

MAY 1957

MACHINE BRAWN—"The word 'automation,' a journalistic coinage, has undeservedly become 'a source of fear,' according to the Earl of Halsbury, an English authority on industrial technology. In the British journal *Impact* he attempts to correct the impression created by journalists that automatic processes in industry 'will cause widespread unemployment.' Machines require trained men to maintain them. Displacements occur mainly in the ranks of the unskilled, who have a high rate of turnover anyway. Lord Halsbury is concerned for the man who is not going to be affected by automation—the coal miner, the stevedore and others 'who do the heavy laboring work for a society which does not know how to lighten their task.'"

MAY 1907

TURN FOR THE WORSE—"The Scientific American recently discussed the gyroscopic action of steam turbines in increasing the stresses in the frail hull structure of torpedo boats. An English naval architect proved that in the case of the British torpedo boat whose back was broken when she was plunging heavily into a head sea, the gyroscopic resistance to a change of plane of the revolving parts of the turbine may have amounted to several tons, and that these stresses, being unrecognized at the time the boat was designed, may have carried the total bending and wrenching stress beyond the limit of strength of the hull."

HEARSE SAY—"A lugubrious automobile novelty in the form of a motor-driven hearse has recently made its appearance in the streets of Paris. A nation that rejoices in the grim delights of trolley-car funerals is hardly justified in shaking its head at this latest application of the electric auto. No good reason can be advanced why a man should not be buried in a ceremonious and somber motor hearse, in these piping motor times."

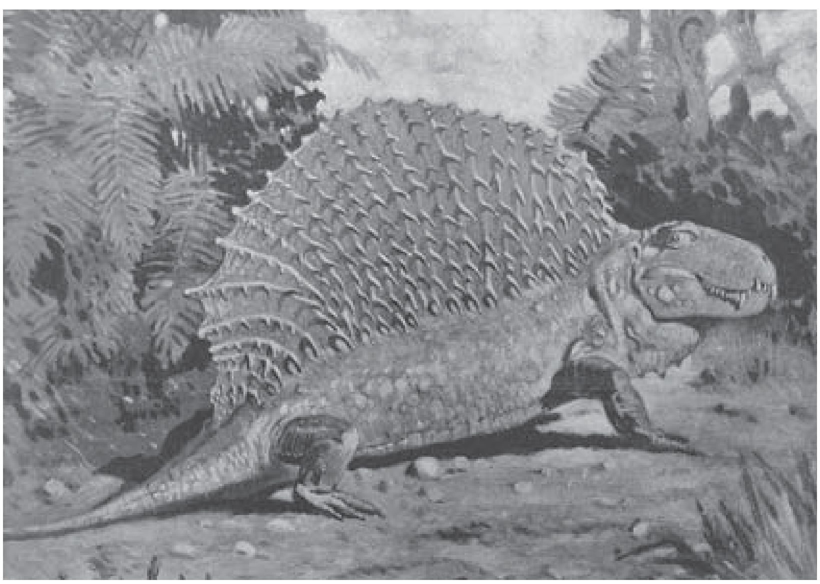
NAOSAURUS—"In presenting to public view the first and only complete mounted skeleton in the world of a remarkable fossil wonder, the carnivorous reptile *Naosaurus,* from the Permian of Texas, Prof. Henry F. Osborn, curator of the Department of Vertebrate Paleontology in the Museum of Natural History, New York, has made a noteworthy contribution to science. The reader's special attention is called to the splendid and ideal drawing, by Mr. Charles R. Knight. —Walter L. Beasley" [*Editors' note: The model was an incorrect composite of* Edaphosaurus *and* Dimetrodon *fossils.*]

MAY 1857

FOUL FUEL—"We believe that no particular use is made of the fluid petroleum, from the 'tar springs' of California, except as a lotion for bruises and rheumatic affections. It has a pungent odor, and although it can be made to burn with a pretty good light, its smell is offensive. This, perhaps, may be obviated by distilling it with some acid; we believe that this is not impossible in this age of advanced chemistry. If the offensive odor could be removed, a valuable and profitable business might be carried on in manufacturing burning fluid from it."

IT'S CLEANER GREENER—"'The tanks for water in India are covered with a green weed,' says the *India Annals of Medical Science,* 'and this, at the same time that it imparts a greenish hue to the water, possesses a remarkable power of filtering it, and rendering it wholesome, for where you have this green weed you also find small fish and infusoria, which preserve the water also. Sir Charles Napier, inspecting the hill districts of the Panjaub, observing the Bheestees drawing water from one of these tanks, ordered it to be immediately cleaned. The water soon became offensive, and was unfit for use until the same weed was replanted and had again covered the tanks.'"

SCIENTIFIC AMERICAN



NAOSAURUS—from a display mounted in 1907

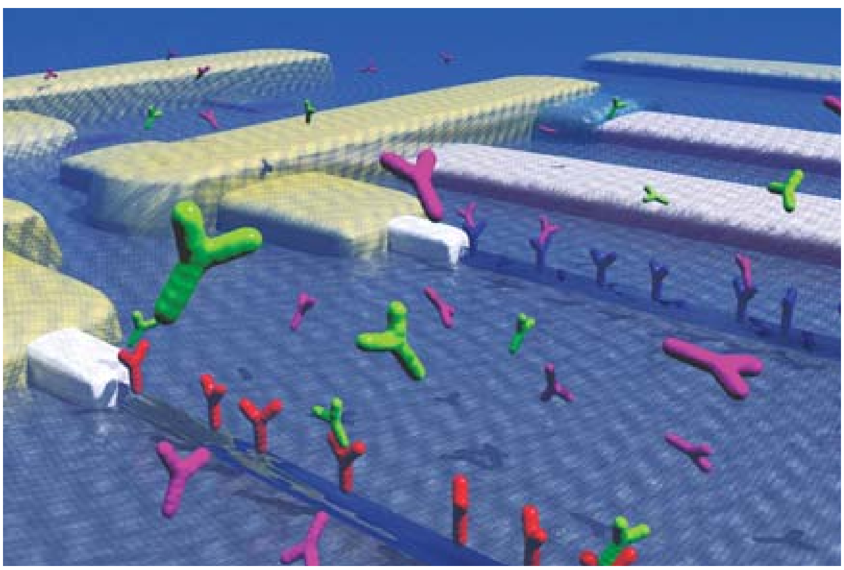
news SCAN

NANOTECH

Structured Settings

NOVEL TECHNIQUES EXPAND THE NANOTECH TOOL KIT **BY CHARLES Q. CHOI**

Nanoparticles are tantalizing construction blocks for researchers, capable of displaying properties of both tiny atoms and far bulkier conventional materials. They generally behave only like balls, however, which makes it hard to assemble them into solid structures other than those resembling displays of oranges in a grocery store. Now researchers have taken big steps in creating and using nanostructures that have eluded manipulation in the past.



FAST DIAGNOSES could be done via etched nanowires (*blue rods*), with bound antibodies (*blue and red Y shapes*) that detect other antibodies in solution. Rectangles are electrical contacts.

In the January 19 *Science,* materials scientist Francesco Stellacci of the Massachusetts Institute of Technology and his colleagues revealed a way to make nanoparticles act like links in a chain, capable of hooking together into strings of beads. The strategy was to take advantage of the so-called hairy ball theorem, which states that if a sphere is covered in hair, attempts to brush those strands to make them all lie flat will always leave behind two hairs standing up straight, each at opposite poles. (Imagine flattening hair on a globe just along the latitude lines; in the end, the hair on the poles will stick out.)

The investigators covered gold nanoparticles with a mix of two kinds of sulfurous molecular hairs. The points where hairs were supposed to stand up essentially became unstable defects on the nanoparticles' surfaces, making it easy to replace those hairs. Stellacci and his colleagues substituted these standouts with chemicals that behaved as handles, enabling the nanoparticles to hold onto one another.

"This really makes nanoparticles analogous to atoms—specifically, divalent atoms that have two chemical bonds. You can now form really interesting structures with them, just like you can make molecules from atoms," says Stanford University materials scientist Yi Cui. Stellacci mentions that his

MARK REED *Yale University*

group is now pursuing ways for nanoparticles to each have four bonds.

These nanostructures could be connected together with nanowires to help form advanced electronics. Researchers can make nanowires either by growing them from the bottom up or by etching them from the top down, akin to whittling toothpicks from a log. Bottom-up approaches face the challenge of integrating each tiny, floppy and often randomly dispersed strand into devices, whereas top-down methods, encompassing more traditional industrial techniques, "are like hacksaws," explains Yale University biomedical engineer Eric Stern. They leave behind rough-edged nanowires with poor electrical properties.

In the February 1 *Nature,* Yale researchers revealed an etching method that creates high-quality, smooth-walled nanowires. The trick was using the ammonium salt TMAH. It can etch silicon more slowly and more smoothly than other solvents used thus far. Yale biomedical engineer Tarek Fahmy adds that the technique is readily compatible with standard industrial semiconductor processes, which would help integrate nanowires into electronics.

These nanowires proved extraordinarily sensitive to their environment, changing in voltage on contact with molecules. They can detect the activation of a T cell immune response to a foreign compound within roughly 10 seconds by sensing the acid released by the cells. Detecting such activation usually takes several minutes, if not hours, to complete via conventional diagnostic assays using tagged antibodies. The investigators also show that nanowires bound with antibodies could detect as few as 60 cancer-linked molecules per cubic millimeter, as good as any current state-of-the-art sensor.

"You can imagine diagnosing patients right at the point of care with devices based off these nanowires, at the ER, the office or the battlefield," says Jonathan Schneck, an immunologist at Johns Hopkins University. "This represents a much more powerful tool than anything I've seen coming down the road in terms of speed of response."

Charles Q. Choi is a frequent contributor.

MOLECULAR CARRIERS

Ludwig Bartels of the University of California, Riverside, and his colleagues were the first to design a molecule that could move in a straight line on a flat surface. Now the investigators can make such walkers haul cargo. They coaxed the organic compound anthraquinone to carry and release two carbon dioxide molecules, as if it had a shopping bag in each hand. Understanding how to transport atoms and molecules in this manner could help engineers assemble objects or deliver materials.

Next, the researchers plan to make molecular carriers go around corners, manipulate their cargo or emit photons to convey their positions. The team may also "build 'muscles' into them that may be triggered by light," Bartels says.

GENETICS

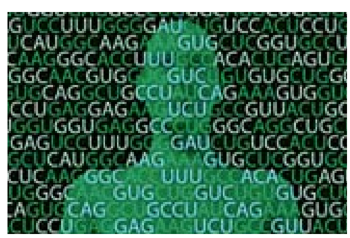
Codon Spell Check

SILENT MUTATIONS ARE NOT SO SILENT AFTER ALL BY CHRISTINE SOARES

As any diplomat will attest, nuances in language can dramatically affect the success of communication. But because the idiom of human cells was assumed to be much more literal than that of whole people, a recent discovery surprised even the scientists who made it. Researchers at the National Cancer Institute (NCI) identified a human protein that takes two distinct shapes depending on a subtle difference in its gene that should not make any difference at all. Their finding may explain variations in chemotherapy response among certain cancer patients and raises new questions concerning so-called silent mutations.

Chava Kimchi-Sarfaty and Michael M. Gottesman were examining a gene called *MDR1,* which is named for its association with multiple-drug resistance in tumor cells. Specifically, they looked at mutations called single-nucleotide polymorphisms (SNPs) in *MDR1.* Some did not change the amino acid sequence encoded by the gene, so the scientists assumed they could have no effect on the final protein. Still, those synonymous SNPs kept showing up in studies of cancer patients, sometimes seeming to influence drug response. "It drew our attention because there were so many reports, and often really contradicting each other," Kimchi-Sarfaty says, "so we started to explore."

Zeroing in on three particular synonymous mutations, the NCI investigators inserted versions of *MDR1* containing those SNPs into human and monkey cells. Because the gene encodes a cell membrane pump called a P-glycoprotein, which acts to siphon chemotherapies out of tumor cells, the team measured how well the pump was working by adding drugs to the test cells. They no-



CHANGES IN CODONS, represented by three letters specifying nucleotides, affect the body more than previously thought.

SLIM FILMS

news SCAN

ticed reduced function in cells carrying one of the SNPs and then found that the P-glycoprotein made by those cells was abnormally shaped. Further experiments revealed that the protein was also being manufactured more slowly than usual in the cells, which led the researchers to suspect that a phenomenon previously seen only in microbes might be at work. "It came to us that maybe we have a problem of codon usage, and we're really exhausting the system by expressing lots of *MDR1,*" Kimchi-Sarfaty recalls.

Codons are three-nucleotide sequences within genes that specify amino acids. With 64 possible triplet combinations of the four DNA bases to encode 20 amino acids, multiple codons can share the same amino acid meaning. Different organisms use certain codons more often than alternative synonyms, and such favored usage is reflected in the relative abundance of cellular translators called transfer RNA (tRNA). Each tRNA molecule recognizes a particular codon in a gene transcript and delivers a corresponding amino acid to the ribosome assembling a protein chain. When a rarer codon appears, fewer matching tRNAs are available in the cell, and the ribosome's operation can slow down.

The synonymous mutation in *MDR1,* the NCI investigators found, created a less common codon. Because the P-glycoprotein's assembly slowed, it folded improperly. The same problem can occur in microorganisms, such as *Escherichia coli* and yeast, when they attempt to produce a protein from an inserted gene containing codons the organisms rarely employ.

But researchers have also documented cases in which the use of an uncommon codon appears to slow protein manufacture in a way that maximizes efficient folding. Whether those examples reveal a new level of information about the correct final protein structure encoded in the gene sequence remains an open question. David Baker, who studies protein-folding dynamics at the University of Washington, thinks such rare situations are more likely to be just accidents that evolution has ignored. "There's probably no deeper level of meaning that would be selected for," he says. For instance, human P-glycoprotein is a large protein with multiple domains, Baker explains, "and for a membrane protein that folds and is inserted as it gets made, you can imagine that if you slow it down, the folding won't happen correctly."

Genome analyses tend to confirm that there is little evolutionary selection of codon bias in mammals, according to Laurence D. Hurst of the University of Bath in England, but codon selection is still sometimes detectable. Hence, codon usage could affect the accuracy of gene mutation rates estimated with the assumption that synonymous codons are "silent" in protein terms. Acknowledging the various mechanisms by which synonymous mutations may not always be neutral, Hurst wrote in a recent review, "will be important for understanding and potentially combating genetic disease."

Kimchi-Sarfaty, now at the U.S. Food and Drug Administration, has already moved on to exploring how polymorphisms and careful gene design could lead to better versions of human blood coagulation proteins. "People were definitely ignoring 'silent' mutations," she says, but she will no longer assume that "synonymous" translates into silence.

SUBTLETIES IN GENE-SPEAK

Making experimental or industrial amounts of drugs and other novel proteins often means inserting foreign genes into microorganisms such as bacteria and yeast. Designers of those genes definitely need to take codon biases into account, says Stephen J. Freeland of the University of Maryland. Because codon usage can affect the microbes' efficiency, Freeland's group is developing a database of synthetic genes and related software to aid in adapting designer genes to a target organism's preferences.

The discovery that synonymous codons could disrupt protein folding in humans illustrates why thinking of the genetic language as a "digital" one is too simplistic, Freeland says. "The idea that an RNA codon such as UUU 'means' phenylalanine just like 010101 'means' the letter A in ASCII," he adds, is one that 21st-century biologists "have to unlearn."

CONSUMER ELECTRONICS

Motion Pictures

HDTV BROADCASTERS PLAN TO GET INTO MOBILE VIDEO

BY MICHAEL ANTONOFF

Even as high-definition televisions are eclipsing entire walls in homes, tiny screens are proliferating on cell phones, pocket PCs and portable media players. The telecommunications industry has long considered supplying video to nomadic subscribers its next big opportunity after voice calls and text messaging, so wireless carriers have been upgrading their systems to third-generation, high-speed networks for that pur-

JAE C. HONG AP Photo

TV TO GO: Mobile video systems, such as Verizon's V Cast, may face competition from broadcasters.

pose. But thanks to new transmission standards, another industry plans to enter the field, one with much more experience in video delivery: digital TV (DTV) broadcasters.

Currently the market for mobile video seems largely untapped, as consumers have yet to become enamored with existing systems. Forrester Research, a technology and market research firm, found that only 1 to 3 percent of people surveyed in 2006 said they would care to access video on a handheld device. The chief reason is cost. Verizon Wireless subscribers, for example, pay an additional $15 per month for its so-called V Cast service. Moreover, a quality-of-service issue plagues conventional cellular systems because video hogs bandwidth: frame rates drop when too many callers are on at once. This problem has forced cellular providers to build higher-power transmitters and to use additional frequencies.

DTV stations may have some advantages. First, each station already has a six-megahertz chunk of spectrum for transmitting digital programs. In that chunk, broadcasters can squeeze in a high-definition channel, plus at least one other channel at a lesser resolution. Second, commercial TV is free, so subscriber resistance is a nonissue.

The big challenge for broadcasters will be adapting federally mandated DTV standards for mobile applications. In the U.S., the Advanced Television Systems Committee (ATSC) has deemed that broadcast antennas must beam signals in a format called vestigial sideband (VSB). But over-the-air reception has fared poorly in environments where the VSB signals reflect off buildings and other obstacles, producing multiple copies that compromise reception. Sophisticated electronics can cope with this multipath problem in stationary receivers, but reception on a TV in a vehicle zipping down the left lane presents a greater difficulty.

As a solution, engineers have proposed an advanced form of VSB called A-VSB that relies on two key components. One is the supplementary reference sequence, which helps the receiver stay locked onto the signal. The second is turbo coding, a mathematical technique for wrapping data in multiple layers of redundant error correction so that if the receiver misses a data bit, three or four others are in line to replace it. Although A-VSB does not specify a particular resolution, engineers have been experimenting with an image size of 320 by 240 pixels—good enough to provide moderate-to-sharp pictures on a small window on a notebook computer.

In January, Samsung, German transmitter manufacturer Rohde & Schwarz and a local Sinclair Broadcasting station demonstrated A-VSB during the 2007 Consumer Electronics Show in Las Vegas. The group fitted a bus with various TVs, including some prototype handheld models. Even as the bus approached speeds of 50 miles an hour, reception of the A-VSB picture remained uninterrupted; in contrast, a concurrently airing DTV program repeatedly disappeared.

Delivering A-VSB does not come without cost—in the demonstration, the reference sequence and turbo coding ate up about 28 percent of the digital transmission's data stream of 19.39 megabits a second. For the broadcaster, that might mean stealing bits from its other program streams and possibly reducing high-definition quality.

John Godfrey, vice president of government and public affairs for Samsung Information Systems America, expects ATSC-led field tests to begin in the spring. He believes that the cost to broadcasters of implementing A-VSB is incremental; a completed standard with rollouts of TV-capable cell phones or media players could begin in 2008.

As for what would be shown in mini form, a Sinclair official said it would likely be dedicated programming, such as national cable channels unleashed from their cables. Or it could be local traffic reports—the better to facilitate your commute home to watch the next episode of *Lost* in your living room.

Michael Antonoff writes about technology from New York City.

THE DISH ON MOBILE VIDEO

Besides broadcast stations and wireless carriers, satellite TV and radio operators are also well positioned to deliver mobile video. At this year's Consumer Electronics Show in Las Vegas, EchoStar Communications introduced MobileDISH, an in-car system that uses a special, roof-mounted antenna. Satellite radio operators Sirius and XM have talked about providing video to their customers for several years. A mobile video service geared toward children in the backseat will likely be available soon: the 2008 Dodge Magnum lineup, for instance, sports a factory-installed Sirius satellite radio with a streaming video option.

At the Las Vegas show, XM demonstrated how it could deliver video to a vehicle using its existing satellites and ground repeaters. Even so, XM spokesman David Butler acknowledges the challenges: "It's expensive, and it uses a significant amount of bandwidth."

PHYSICS

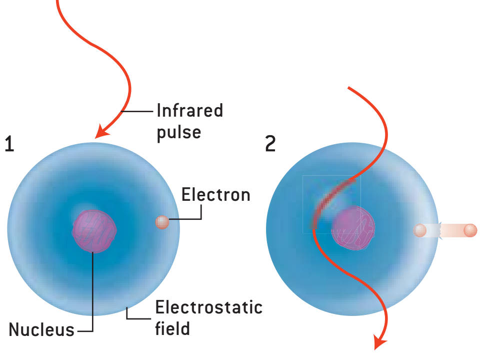
Escape from the Nucleus

IONIZATION VIA QUANTUM TUNNELING OBSERVED BY ALEXANDER HELLEMANS

To break free from an atom, the negatively charged electron typically has to absorb a high-energy photon, such as that from the ultraviolet (UV) or x-ray spectrum. The electron then gets excited enough to overcome the electrostatic attraction holding it to the positively charged nucleus and escapes, a process called ionization. A German-Dutch team has for the first time provided direct proof of an alternative mechanism. Powerful electric fields from a laser pulse can momentarily weaken the electrostatic bonds and enable the electron to quantum-mechanically tunnel away from the atom.

Leonid Keldysh, now at the Lebedev Physics Institute in Moscow, predicted the effect in 1964, and experiments have already proved that such unusual ionization can occur. But only with the advent of laser pulses lasting just a few hundred attoseconds can physicists observe the phenomenon. (One attosecond is a billionth of a billionth of a second.) Attosecond laser pulses have already made it possible to probe the motion of electrons in atoms and molecules, and improved versions will allow researchers to track electron movements that occur, for instance, during chemical reactions.

Ferenc Krausz of the Max Planck Institute for Quantum Optics in Garching, Germany, and his team describe their ionization experiment in the April 5 *Nature*. Targeting a gas of neon atoms, the group first used a 250-attosecond UV laser pulse to nudge one electron farther away from the nucleus. Almost simultaneously the physicists fired an infrared pulse 5,000 attoseconds long whose electric field oscillates only a few cycles. The field weakened the electrostatic force and enabled the loosened electron to tunnel out, as quantum particles can do when confronted with a thin barrier. By increasing the time between



TUNNEL TO FREEDOM: When an infrared photon strikes an atom (*1*), its electric field can distort and thereby weaken the atom's electrostatic field on one side, enabling the electron to tunnel away (*2*).

LUCY READING-IKKANDA

the UV and infrared pulses in small steps, the researchers found that the number of neon ions formed rose in parallel, clearly indicating that whenever the electric field of the infrared laser pulse reached a maximum, the rate of produced ions increased as well.

Keldysh's theory of strong-field ionization has become part of many other theories, and the result is "not really a dramatic surprise," Krausz admits. But critically, "the team has shown a new way to make measurements" of electron dynamics, comments physicist Paul Corkum of Canada's National Research Council in Ottawa. And the technique could probe poorly understood processes in which electrons exchange energy with one another.

As an example, Krausz cites the "shake up" process in atoms that occurs when an energetic x-ray photon kicks out an electron close to the nucleus. While flying away, this electron could impart some of its energy to another electron, which would become excited and move farther away from the nucleus. Hence, a small delay might exist between the absorption of the x-ray photon by the ejected electron and the repositioning of the second electron. The delay, Krausz remarks, "could be as little as 50 attoseconds; nobody really knows." The length of the delay is not exciting, he explains—rather the point would be whether a delay existed at all. A delay would mean the second electron got energy from the first and was not coincidentally and simultaneously excited by the x-ray photon.

Krausz claims he has now achieved 100-attosecond UV pulses, so he may soon solve that puzzle. As the lasers improve, answers to other questions are sure to follow in the coming years, if not in the coming attoseconds.

Alexander Hellemans is a science writer based in Bergerac, France.

DESIGN

Dialing In

MOBILE PHONES TARGET THE WORLD'S NONREADING POOR **BY SALLY LEHRMAN**

Those days are long gone when placing a telephone call meant simply picking up the receiver and asking the operator to patch you through. Modern cell phones require users to navigate a series of menus to find numbers, place calls or check messages. Even the most tech-savvy may take weeks to discover some of the more arcane multimedia functions. Imagine the difficulty for someone unable to read.

UNTAPPED POTENTIAL: Whereas some customers in India may have no problem with cell phones, nonliterate people may benefit from a new design.

M. LAKSHMAN AP Photo

That is the challenge for mobile communications companies aiming to branch out into developing countries. The prospects are alluring: according to the GSM Association, a global trade group, only about one third of China's vast population and about one tenth of India's use cell phones. But selling to poor rural areas is not likely to happen with a marketing version of "plug and play." Most potential buyers have little exposure to anything other than simple electronics. Reading through a series of hierarchical menus and pushing buttons for multiple purposes would be new concepts for such customers.

To come up with a suitable device, Motorola relied on a team of anthropologists, psychologists and designers to study how textually illiterate villagers use their aging televisions, tape players and phones. The researchers noticed that their subjects would learn each button's dedicated function. With something more complicated, such as an automated teller machine, users would memorize a set of behaviors in order, which allowed them to move through the machine's basic hierarchy without having to read the menu.

The research, which lasted three years, led Motorola to craft a cellular handset slimmed down to three essential activities: calling, managing numbers and simple text messaging. “A lot of the functions in a cell phone are not useful to anyone,” points out Gabriel White, who headed the interactive design team. The icon-based interface also required thought. Although some images worked fine—a clock with ringers to indicate the alarm-setting function—others had to be redrawn. A stack of coins was intended to indicate the remaining prepaid call value, but potential users mistook the original design as an illustration of a pile of food.

The culmination of Motorola’s efforts is the low-cost Motofone, which the company introduced to first-time mobile users in India and Pakistan late last year. At first glance, the nine-millimeter-thick phone looks much like other mobiles, except with fewer keys and no wording. It has a numeric keypad and separate buttons for call, end, menu and phone book. Instead of layer on layer of function lists, the menu button pulls up a set of icons on the screen, with each one representing just one activity. A navigation dial enables users to select the icons.

Not all cell phone companies believe that a design for nonliterate users should start from scratch. Nokia’s behavioral researchers noticed that “newbies” rely on friends and relatives to help them with basic functions. Rather than confronting the challenge of a completely new interface, Nokia chose to provide some audio menus in its popular 1100 model and a preview mode so that people could try out functions without the risk of changing anything important.

Mobile phones may even become tools for literacy, predicts BJ Fogg, who studies computer-human interaction at Stanford University. Phones might teach the alphabet or tell a story as users read along. “Imagine if it eventually could understand your weak points and drill you on those,” Fogg proposes. And soon enough, he declares, designs for illiterate users will lead to more straightforward, elegant phones for everyone.

Sally Lehrman is based in the San Francisco Bay Area.

MICROFLUIDICS

Microchips with Heart

NO BATTERIES NEEDED: PUMPS DRIVEN BY HEART CELLS

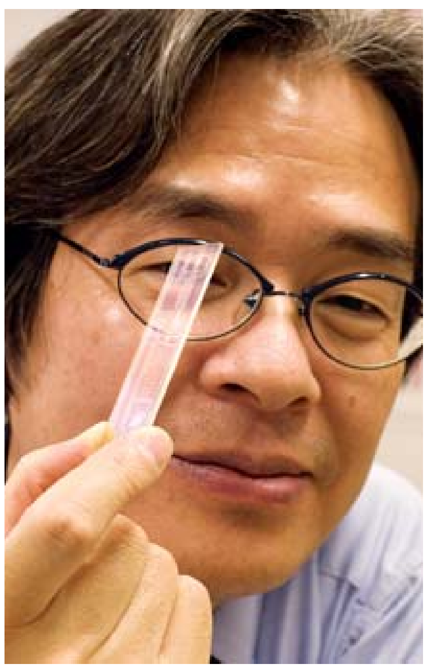
BY TIM HORNYAK

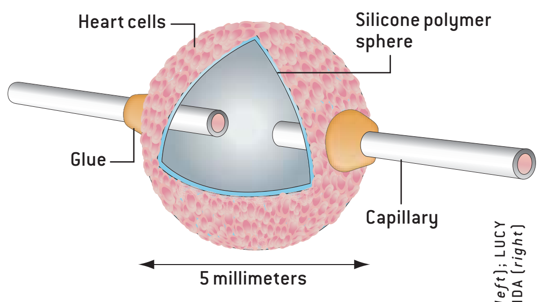
Bioengineers have eagerly explored tiny mechanisms that move microscopic or nanoscopic amounts of fluid. Already used in ink-jet printers, microfluidic devices hold out the promise of efficient and rapid medical assays requiring only minuscule sample volumes for results. Potential applications include lab-on-a-chip devices for on-site chemical analyses, engineered vital organs, and in vivo pharmacies that secrete drugs when needed. In terms of implants, one of the main problems is power. Microfluidics typically rely on large external gear and electricity supplies to run actuators that drive fluids through microchannels. Now a group of researchers has innovated a concept that instead relies on an old-fashioned kind of engine—a living heart.

The team, based in Japan, has transformed cultured rat heart muscle cells into a spherical, pulsating micropump five millimeters in diameter that requires no external power, wires or stimuli. The simple device consists of a hollow, flexible silicone sphere with tubes 400 microns wide at opposite sides—in structure, it resembles an earthworm's one-chamber heart. Rat heart muscle cells (cardiomyocytes) surround the sphere, and when they pulsate in sync, they drive fluid through the tubes and the sphere. Cell nutrients in the surrounding medium power the cells.

"In the future, this kind of microsystem may be applied to medical, clinical and biological applications," says project leader Takehiko Kitamori of the University of Tokyo. As examples, he points to "implanted chemical systems, such as drug delivery systems, or an insulin monitoring and delivery system on a wristwatch."

Kitamori has spent four years building the micropump. He collaborated with Teruo Okano of Tokyo Women's Medical University, who has developed cellular sheets that can bind to an impaired heart to support pumping. Forming a sphere out of the rat cells is a key technology of the research. First, the cells are cultured in a polystyrene dish coated with a heat-sensitive polymer that, when cooled, allows the cells to be harvested as a single sheet. The cell sheet, pulsating continuously, is





GETTING PUMPED: University of Tokyo's Takehiko Kitamori holds a kind of "lab on a chip" (*left*) that could be powered by his micropump, which is fashioned from rat heart cells grown over a flexible, hollow silicone polymer ball (*right*).

TIM HORNYAK (*left*); LUCY READING-IKKANDA (*right*)

then placed on a silicone sphere formed on a ball of sugar (the sugar is dissolved with water once the silicone solidifies). The cells need only an hour to stick to the silicone because it is coated with fibronectin, an adhesive glycoprotein involved in tissue repair and blood clotting. Epoxy glue attaches the two capillaries to the sphere.

The researchers do not know exactly why the heart muscle cells start beating simultaneously once one gets going, but the harmonized motion continued in a test of the pump. The team watched polystyrene tracking particles move through the pump for five days, with expected flow rate falling only slightly. Because replacing the culture medium seemed to cause beat-frequency fluctuations, future applications would necessitate a constant environment.

Kitamori would like to make the micropump more powerful and sophisticated—more like a fish heart—by adding cell layers and valves. Multiple chambers could also be incorporated by reshaping the silicone sphere. "This is chemical-energy-to-mechanical-energy conversion that can be used in the field, with no electrical power," says Kitamori, who wants to fashion chip-based model circulatory systems for direct observation of vascular and cardiac disease.

He also sees humans of the future having health-monitoring chip labs inside them performing dozens of functions. The pumps running those labs would be formed from the person's own myocardial cells, thereby eliminating immune rejection. In essence, such bioengineered humans would have not one heart but many. Of course, the miniature hearts would not replace the real thing. "We only need one heart for our lifetimes, and cardiac muscle is very tough," Kitamori laughs. "That's why we're applying this robust actuator to a liquid-control system."

Tim Hornyak, based in Tokyo, is author of Loving the Machine: The Art and Science of Japanese Robots.

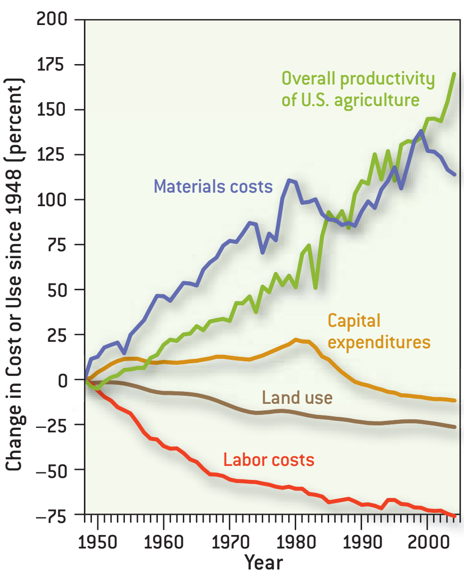
BY THE NUMBERS

Food Boom

AGRICULTURAL PRODUCTIVITY SUSTAINS THE U.S. ECONOMY **BY RODGER DOYLE**

One of the least appreciated but most remarkable developments of the past 60 years is the extraordinary growth of American agriculture. Farming now accounts for about one tenth of the gross domestic product yet employs less than 1 percent of all workers. It has accomplished this feat through exceptionally high growth in productivity, which has kept prices of food low and thereby contributed to rising standards of living. Furthermore, the exportable surplus has kept the trade deficit from reaching unsupportable levels. Agriculture not only has one of the highest rates of productivity growth of all industries, but this growth appears to have accelerated during the past two decades.

Over the period 1948 to 2004, total farm production went up by 166 percent. But as



RODGER DOYLE; SOURCE: U.S. DEPARTMENT OF AGRICULTURE; OVERALL PRODUCTIVITY TAKES INTO ACCOUNT THE CONTRIBUTION OF LABOR, CAPITAL, LAND, ENERGY AND MATERIALS

the chart shows, productivity per person improved so much that only one quarter as many hands were needed in 2004 as in 1948. Furthermore, the arable land used for farming dropped by one quarter over the 56-year period, and investment in heavy farm equipment and other capital expenditures decreased by 12 percent.

Several developments drove these changes, beginning with the replacement of the remaining horses by tractors immediately after World War II and with the expanding use of fertilizers and pesticides. Later came the adoption of hybrid seeds, genetic engineering of plants and improved livestock breeding. A key element was the U.S. Department of Agriculture's extension service. Operating through land-grant universities and other organizations, it educated farmers on biotechnology, pest management and conservation.

For many years, critics have claimed that modern agriculture is not sustainable, one of the major assertions being that it encourages erosion, which will eventually wash away most of the topsoil. Lost topsoil, the argument goes, is virtually irreplaceable because it takes up to 300 years for one inch of soil to form. Indeed, the fear of topsoil loss has dogged agriculture since the days of the 1930s Dust Bowl.

But a detailed study of two large areas, the Southern Piedmont and the Northern Mississippi Valley Loess Hills, showed that based on 1982 data, soil loss has dropped sharply from the very high rates of the 1930s. The conclusion of the analysis, which appeared in the July 14, 2000, issue of *Science*, was that if 1982 erosion rates continued for 100 years, crop yields would decline by only 2 to 4 percent. The study attributed the decrease in soil erosion to the USDA, which urged farmers after World War II to adopt conservation practices such as strip cropping, whereby alternating rows are planted, and leaving plant residues in the fields year-round to inhibit water runoff.

Despite being a robust contributor to the U.S. economy, modern agriculture is not without a dark side. Runoff of fertilizers, antibiotics and hormones degrade the environment and can upset the local ecology. If not grown properly, genetically modified crops could spread their DNA to conventional species. And the industrial approach to food has contributed to America's obesity epidemic as well as to sporadic but widespread *Escherichia coli* outbreaks.

Rodger Doyle can be reached at rodgerpdoyle@verizon.net

AMERICAN MONOCULTURE

The Food and Agriculture Organization (FAO) of the United Nations advocates the maintenance of a permanent or semipermanent organic soil cover. To achieve this condition, it recommends farming practices that include little or no tillage and varied crop rotation—measures that the U.S. follows to a limited extent. The FAO also insists that the future of agriculture depends on biodiversity. The U.S., however, has largely continued to practice monoculture: 82 percent of harvested acreage in 2005 was in four crops (corn, soybeans, hay and wheat), a rise from 78 percent in 1985.

FURTHER READING

Agricultural Productivity in the United States. Mary Ahearn, Jet Yee, Eldon Ball and Rich Nehring. U.S. Department of Agriculture Bulletin No. 740, January 1998.

U.S. Soil Erosion Rates—Myth and Reality. Stanley W. Trimble and Pierre Crosson in *Science*, Vol. 289, pages 248–250; July 14, 2000.

Productivity and Output Growth in U.S. Agriculture. Eldon Ball in *Agricultural Resources and Environmental Indicators.* Economic Research Service, U.S. Department of Agriculture, 2006. Available at www.ers.usda.gov/publications/arei/eib16/

Food and Agriculture Organization of the United Nations: www.fao.org

COGNITION

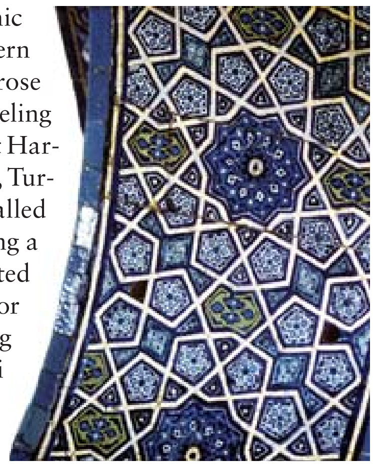
Rat-Think

Rats have a complex form of thought once known only in humans and other primates—specifically, metacognition, or the ability to mull over what one knows. Scientists at the University of Georgia gave rats a choice to either take a test or not. Passing the test resulted in a large food reward, failure yielded nothing, and opting out altogether led to a small reward. The test played noises of various durations that the rats had previously learned to classify as short (lasting two to 3.6 seconds) or long (4.4 to eight seconds). The more difficult it was to gauge the sound length—for instance, if the noise lasted 4.4 seconds—the more often rodents bailed out from the test. The findings, in the March 20 *Current Biology*, suggest the rats could judge whether they would pass or fail, hinting that metacognition may be more widespread than thought before. —*Charles Q. Choi*

PATTERNS

Medieval Quasicrystals

In designing architectural structures, medieval Islamic artisans hit on so-called quasicrystals, a complex pattern made famous by renowned mathematician Roger Penrose in the 1970s. After seeing hints of the pattern while traveling in Uzbekistan, Peter J. Lu, a physics graduate student at Harvard University, pored over photographs from Iran, Iraq, Turkey and Afghanistan. Some of the ornate tile work, called *girih,* could have only been accurately constructed using a set of five tiles—a bow tie, pentagon, diamond, elongated hexagon and large decagon—conclude Lu and co-author Paul J. Steinhardt of Princeton University. The tiling practice reached great sophistication with the Darb-i Imam shrine in Iran, which dates to 1453; it displays a symmetric pattern of pentagons and 10-sided stars. If extended indefinitely in all directions, this pattern would never repeat itself—the hallmark of a quasicrystal. The researchers describe their conclusions in the February 23 *Science.* —*JR Minkel*



TILE WORK on this 15th-century Turkish mosque exhibits a quasicrystal nature, in which the pattern never repeats.

PHYSICS

Spooky Action at a Greater Distance

The quantum link called entanglement keeps getting longer—in the latest demonstration, the distance spanned two of Spain's Canary Islands. When two photons are entangled, what happens to one instantaneously determines the fate of the other, no matter how far apart they are. Using a laser, Anton Zeilinger of the University of Vienna and his team created entangled pairs of photons on the island of La Palma and then fired one member of each pair to a telescope on Tenerife, 144 kilometers away. That distance is 10 times farther than entangled photons have ever flown through the air. Such photons might prove suitable for sending scrambled messages that cannot be decoded or secretly read. The team discussed its feat at the March meeting of the American Physical Society. —*JR Minkel*

Extended coverage of these News Scan briefs can be found at www.sciam.com/ontheweb

DATA POINTS: BIG BANG OF BITS

The expansion of the digital universe is accelerating, according to a study by IDC, an information and technology analysis firm, and EMC Corporation, a data management company. Powering this growth is the conversion of analog images, television broadcasts and voice calls to digital formats, as well as the storing and sharing of such files.

Number of bits in a byte: 8

Number of gigabytes in an exabyte: 1 billion

Exabytes of data created, captured and copied in 2006: 161

Equivalent data in stacks of books stretching from the earth to the sun: 12

Exabytes expected to be made in:

2007: 255

2010: 988

Available storage capacity today, in exabytes: 246

Number of bits stored on one square inch of a disk drive in:

1956: 2,000

2007: 100 billion

SOURCE: The Expanding Digital Universe, IDC White Paper, March 2007

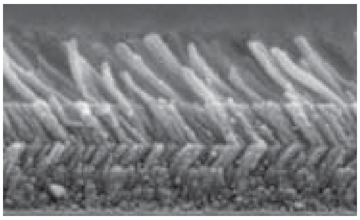
COURTESY OF W. B. DENNY; ILLUSTRATION BY MATT COLLINS

news SCAN

OPTICS

No Glare There

Researchers have created a coating that reflects almost no light, thanks to a refractive index close to that of air. A material's index of refraction refers to the speed of light in that substance, which in turn governs how much light bends when it passes into the substance. The new material consists of angled nanorods laid on top of a transparent semiconductor wafer. By stacking five layers of nanorods, each one less porous than the one below it, the researchers could vary the coating's refractive index in a series of steps, going from 2.03 at the bottom (closely matching the wafer) to 1.05 at the top (compared with 1.0 for air). Total reflection dropped to as little as 0.1 percent, note E. Fred Schubert and his group at Rensselaer Polytechnic Institute in the March *Nature Photonics*. The coating should prove useful in boosting the efficiency of LEDs and solar cells. —*JR Minkel*



LAYERED NANORODS, each with a different refractive index, nearly eliminate glare at several wavelengths.

MEMORY

Sweet (Scented) Dreams

A night on a bed of roses could sharpen your memory. The purpose of sleep remains a mystery, but one idea is that slumber replays new experiences to facilitate their inclusion in long-term memory storage. To explore this notion, volunteers in a rose-scented room played a computer version of the card game Memory, where they had to remember locations of paired cards that flashed on the screen for a few seconds at a time. Later, while they slept, researchers exposed them to the rose fragrance, because odors are well known for bringing back memories. Volunteers performed better at the game when given the scent during slow-wave sleep, often considered the deepest stage of sleep. Brain imaging during slow-wave sleep showed that the scent cue activated the hippocampus, a region linked with memory. Scientists at the University of Lübeck in Germany and their colleagues present their findings in the March 9 *Science*. —*Charles Q. Choi*

METEOROLOGY

Drought-Making Pollutants sa

Tiny particles resulting from fuel burning appear to be interfering with the formation of rain, snow and ice on one of China's mountains. As aerosols increase in the atmosphere, fewer droplets collide with one another to form raindrops. Atmospheric scientists studied records from the past 50 years at a meteorological observatory that sits at the peak of sacred Mount Hua. They not only found that visibility from the two-kilometer-high summit has, on average, declined from roughly 30 to 10 kilometers, but they also showed that rainfall has dropped by as much as 17 percent compared with the precipitation in neighboring areas. The findings explain the widely observed trend—from Canada to South Africa to Israel—of a decrease in highland precipitation compared with that in adjacent lowlands. Pollution is not just obscuring the view; it is also choking the mountain streams. The researchers report their work in the March 9 *Science*. —*David Biello*

DRY TIMES: Haze not only shrouds the view on Mount Hua but also discourages rain.

E. FRED SCHUBERT *Rensselaer Polytechnic Institute* (*top*); KAI LIU (*bottom*)

BRIEF POINTS

- **Prolactin, a hormone that increases during pregnancy, regenerates myelin destroyed in mouse models of multiple sclerosis (MS). The results may explain why symptoms of MS sometimes abate during pregnancy.**

  Journal of Neuroscience, February 21

- ***T. rex* and other meat-eating dinosaurs had small genomes, on par with modern birds, which have about 1.45 billion bases (humans have about three billion). The finding counters a hypothesis that birds evolved and maintained smaller genomes to conserve energy for flight.**

  Nature, March 8

- **Out-of-body sensations may stem from a disturbance in the brain's arousal system, in which it would fail to make a proper transition from REM sleep (when dreams often occur) to wakefulness.**

  Neurology, March 6

- **Asteroid surfaces that have absorbed warming sunlight can reemit the heat when they move into the shade, thereby creating recoil and affecting an asteroid's rotation.**

  Science Express, March 8

Page Intentionally Blank

Bush's Mistake and Kennedy's Error

Self-deception proves itself to be more powerful than deception By MICHAEL SHERMER

The war in Iraq is now four years old. It has cost more than 3,000 American lives and has run up a tab of $200 million a day, or $73 billion a year, since it began. That's a substantial investment. No wonder most members of Congress from both parties, along with President George W. Bush, believe that we have to "stay the course" and not just "cut and run." As Bush explained in a speech delivered on July 4, 2006, at Fort Bragg, N.C.: "I'm not going to allow the sacrifice of 2,527 troops who have died in Iraq to be in vain by pulling out before the job is done."

We all make similarly irrational arguments about decisions in our lives: we hang on to losing stocks, unprofitable investments, failing businesses and unsuccessful relationships. If we were rational, we would just compute the odds of succeeding from this point forward and then decide if the investment warrants the potential payoff. But we are not rational—not in love or war or business—and this particular irrationality is what economists call the "sunk-cost fallacy."

What happens when someone says, "I was wrong"?

The psychology underneath this and other cognitive fallacies is brilliantly illuminated by psychologist Carol Tavris and University of California, Santa Cruz, psychology professor Elliot Aronson in their book *Mistakes Were Made (But Not by Me)* (Harcourt, 2007). Tavris and Aronson focus on so-called self-justification, which "allows people to convince themselves that what they did was the best thing they could have done." The passive voice of the telling phrase "mistakes were made" shows the rationalization process at work. "Mistakes were quite possibly made by the administrations in which I served," confessed Henry Kissinger about Vietnam, Cambodia and South America.

The engine driving self-justification is cognitive dissonance: "a state of tension that occurs whenever a person holds two cognitions (ideas, attitudes, beliefs, opinions) that are psychologically inconsistent," Tavris and Aronson explain. "Dissonance produces mental discomfort, ranging from minor pangs to deep anguish; people don't rest easy until they find a way to reduce it." It is in that process of reducing dissonance that the self-justification accelerator is throttled up.

Wrongly convicting people and sentencing them to death is a supreme source of cognitive dissonance. Since 1992 the Innocence Project has exonerated 192 people total, 14 from death row. "If we reviewed prison sentences with the same level of care that we devote to death sentences," says University of Michigan law professor Samuel R. Gross, "there would have been *over 28,500 non-death-row exonerations in the past 15 years....*" What is the self-justification for reducing this form of dissonance? "You get in the system, and you become very cynical," explains Northwestern University legal journalist Rob Warden. "People are lying to you all over the place. Then you develop a theory of the crime, and it leads to what we call tunnel vision. Years later overwhelming evidence comes out that the guy was innocent. And you're sitting there thinking, 'Wait a minute. Either this overwhelming evidence is wrong, or I was wrong—and I couldn't have been wrong, because I'm a good guy.' That's a psychological phenomenon I have seen over and over."

What happens in those rare instances when someone says, "I was wrong"? Surprisingly, forgiveness is granted and respect is elevated. Imagine what would happen if George W. Bush delivered the following speech:

> This administration intends to be candid about its errors. For as a wise man once said, "An error does not become a mistake until you refuse to correct it." We intend to accept full responsibility for our errors.... We're not going to have any search for scapegoats ... the final responsibilities of any failure are mine, and mine alone.

Bush's popularity would skyrocket, and respect for his ability as a thoughtful leader willing to change his mind in the teeth of new evidence would soar. That is precisely what happened to President John F. Kennedy after the botched Bay of Pigs invasion of Cuba, when he spoke these very words. SA

Michael Shermer is publisher of Skeptic *(www.skeptic.com). His latest book is* Why Darwin Matters.

BRAD HINES

The Road to Clean Energy Starts Here

Realizing crucial energy technologies will take more than just research and development

By JEFFREY D. SACHS

The key to solving the climate change crisis is technology. To accommodate the economic aspirations of the more than five billion people in the developing countries, the size of the world economy should increase by a factor of four to six by 2050; at the same time, global emissions of greenhouse gases will have to remain steady or decline to prevent dangerous changes to the climate. After 2050, emissions will have to drop further, nearly to zero, for greenhouse gas concentrations to stabilize.

The overarching challenge is to make that transition at minimum cost and without economic disruption. Energy-saving technologies will play a pivotal role. Buildings can save energy at low capital cost, and often net overall savings, through improved insulation, efficient illumination and the use of heat pumps rather than home furnaces. Automobiles could, over time, reach 100 miles per gallon by a shift to plug-in hybrids, better batteries, lighter frames and other strategies. Of course, technologies such as heat pumps and plug-in hybrids partly reduce direct emissions by shifting from on-site combustion to electricity, so that low-emission power plants become paramount.

Some will impose real added costs; others will pay for themselves.

Low-emission electricity generation will be achieved in part through niche sources such as wind and biofuels. Larger-scale solutions will come from nuclear and solar power. Yet clean coal will be essential. New combustion techniques, combined with carbon capture and sequestration (CCS), offer the prospect of low- or zero-emission coal-fired thermal plants. The incremental costs of CCS may well be as low as one to three cents per kilowatt-hour.

All these technologies are achievable. Some will impose real added costs; others will pay for themselves as lower energy bills offset higher capital outlays. Some estimates suggest that, as of 2050, the world will have to negate around 30 billion tons of carbon dioxide emissions a year at a cost of roughly \$25 per ton, or \$750 billion annually. But with a world economy by then of perhaps \$200 trillion, the cost would be well under 1 percent of world income and perhaps under 0.5 percent, a true bargain compared with the costs of inaction.

Achieving these technological solutions on a large scale, however, will require an aggressive global technology policy. First, there will have to be market incentives to avoid emissions, in the form of either tradable permits or levies. A reasonable levy might be \$25 per ton of emitted carbon dioxide, introduced gradually over the next 10 to 20 years. Second, there will have to be ample government support for rapid technological change. Patents can help spur private market research and development (R&D), but public funding is required for basic science as well as for the public *demonstration* and the global *diffusion* of new technologies. In sum, we need a strategy sometimes described as RDD&D.

In the past two years, the Earth Institute at Columbia University has hosted a Global Roundtable on Climate Change, involving leading corporations from around the world. These companies, including many of the largest power producers, are ready to reduce carbon emissions. They know that CCS must be a high priority. A new Global Roundtable Task Force on CCS seeks to promote the required RDD&D.

Fortunately, the European Union has already pledged to build at least a dozen CCS demonstration projects in Europe by 2015. But we will also need such centers in the U.S., China, India, Australia, Indonesia and other highly significant coal-power producers. In the low-income countries, this will require a few billion dollars; that is where the RDD&D investments of the high-income countries will be essential.

The CCS Task Force aims to break ground on one or more demonstration plants by 2010 in every major coal region. By 2015 this crucial technology can be proved and added to the bid to avert climate disaster. This model of RDD&D won't stop there. Harnessing technology to achieve sustainable energy will involve much of the global economy for decades. SA

Jeffrey D. Sachs is director of the Earth Institute at Columbia University (www.earth.columbia.edu).

An expanded version of this essay is available at **www.sciam.com/ontheweb**

BRUCE GILBERT/EARTH INSTITUTE

Prime Directive for the Last Americans

Saving Amazonia's indigenous peoples means not meeting them, insists Sydney Possuelo—a policy of noninterference he hopes to extend, even if others hate it BY CLAUDIO ANGELO

The gray-bearded, balding man sips coffee in the kitchen of his apartment in Brasília, Brazil. With a blank stare, he ponders the future of the three things to which he has dedicated his life: "Everything dies at its own time. The forest dies, with it die the Indians, with them die the *sertanistas*." But at 67, Sydney Ferreira Possuelo acknowledges that the *sertanistas,* men who make a living out of protecting the isolated indigenous peoples in the Brazilian jungles, may be the first to go.

NICOLAS REYNARD *National Geographic Image Collection*

SYDNEY POSSUELO: NATIVE PROTECTION

- **Works to protect the way of life of isolated indigenous groups in the Amazon rain forest. (Possuelo is pictured above with a Korubo woman.)**
- **Perhaps four million indigenous people in 1,000 ethnic groups once lived in Brazil; today they number fewer than 734,000 in 220 tribes.**
- **On being fired from his state job: "I have never been one of those nice workers who agree with everything that the government does."**

Counting Possuelo, only five members remain. They travel to the remotest corners of the Amazon rain forest to track down isolated tribes and contact them before miners, cattle ranchers, loggers and others do, usually at gunpoint. Their job is to divert development around those tribes, so as to keep them from sharing the usual bleak fate of indigenous Brazilians. But it has been 20 years since the last *sertanista* was hired. And most of the old school—that is, people who have actual contact experience—have either died or retired.

Possuelo has done neither, although he was fired from his position with Brazil's National Foundation for Indigenous Peoples (FUNAI) last year after publicly criticizing the remarks of its president, who stated that there might be too much land already in the hands of native Brazilians. Today he coordinates efforts through the Instituto Indigenista Interamericano, a nongovernmental organization that is part of a multinational alliance dedicated to the protection of isolated tribes.

Possuelo had helped initiate a sea change in federal policy toward indigenous tribes by persuading the government that the natives should be left alone, to live exactly as they have since prehistoric times. The decision to make contact should be left to them. The prevailing wisdom throughout much of the 20th century and before had been "integration" into society. "The idea was that we should go fetch those peoples and share with them the benefits of civilization, since civilization is a good of all humankind that belongs to non-Indians and Indians alike," Possuelo says.

But such a bright positivist view conflicted with reality. After contact, many natives lost their lives to influenza and other diseases to which they had no natural defense. Not a few were expelled from their traditional territories to make room for highways, dams and cattle ranches. Many were murdered, in some cases with approval of government officials. But perhaps worst of all was the cultural dislocation.

When meeting a new group, Possuelo recounts, "you must keep an eye on the folks who are there. The guys fight you with bows and arrows, they kill you, they speak up to you, they assault you." Contact, however, changes all that: "One year later they are slack, emaciated, bowing their heads and begging for food and money by the roadside." A case in point: the Arara people, who had been called the scourge of the Transamazônica Highway for their attacks on work crews but today live as poor peasants fully dependent on federal aid. "You break down their health, their mythical universe, their work and their education system. They become outcasts, and many of them have been outcasts for 500 years," Possuelo says.

LAST OF A KIND: **Possuelo meets in peace with the Korubo, who have been known to kill intruders.**

The veteran *sertanista*'s career began in 1959, as a disciple of Orlando and Cláudio Villas Bôas, two of the three brothers credited with saving 15 tribes by creating the Xingu Indigenous Park, Brazil's first megareservation. With them, he journeyed into the Amazon for the first time in the early 1960s. In the 1970s he joined FUNAI and made his first solo contact with an isolated group, the Awá-Guajá of Maranhão.

By 1987 Possuelo decided to take the Villas Bôas strategy further, because even peaceful contact with such groups often destroyed their native culture and self-sufficiency. His idea: to avoid contact altogether. He convinced the government that it was more practical to demarcate indigenous lands and to guard the borders with armed agents than to provide for them indefinitely. "We do it for threatened animal species, why can't we do it for a unique ethnic group that has been there for thousands of years?" he asks. "That's good for the Indians, who will see their land and their tribal traditions respected, and good for the state, who will no longer be charged with genocide." As an added bonus, a study found that legally recognized indigenous lands in the Amazon had lower average deforestation rates (1.14 percent) than federal conservation units (1.47 percent).

Executing Possuelo's plan, however, means identifying the indigenous groups before miners, loggers and others find them—encounters that could lead to long-lasting, deadly skirmishes, as in the case of the Korubo people of the Javari Valley, which lies near the Peruvian border. First contacted in 1972, they violently resisted intruders and over the decades had clubbed to death many who ventured into their territories. In 1996 Possuelo managed to establish peaceful encounters with one Korubo clan, in part by singing aloud (enemies come in silence) and by offering gifts. Despite having known the clan for many years now, Possuelo still considers them unpredictable; he has kept anthropologists away, he has said, for their own safety. He himself will not bother to contact a larger Korubo group deeper in the jungle, feeling that they are remote enough that they will not meet outsiders.

The noninterference strategy has gained worldwide approval. The United Nations Declaration on the Rights of Indigenous Peoples, adopted in 2006, incorporates the right to nonassimilation. Despite that, the policy faces critics in Brazil, who see such protection as an additional burden to taxpayers that only postpones inevitable integration. "Of course, you're buying time," says Márcio Santilli, head of Brasília's Socio-Environmental Institute and a former president of FUNAI. "But buying time is central to mitigating the impacts of contact."

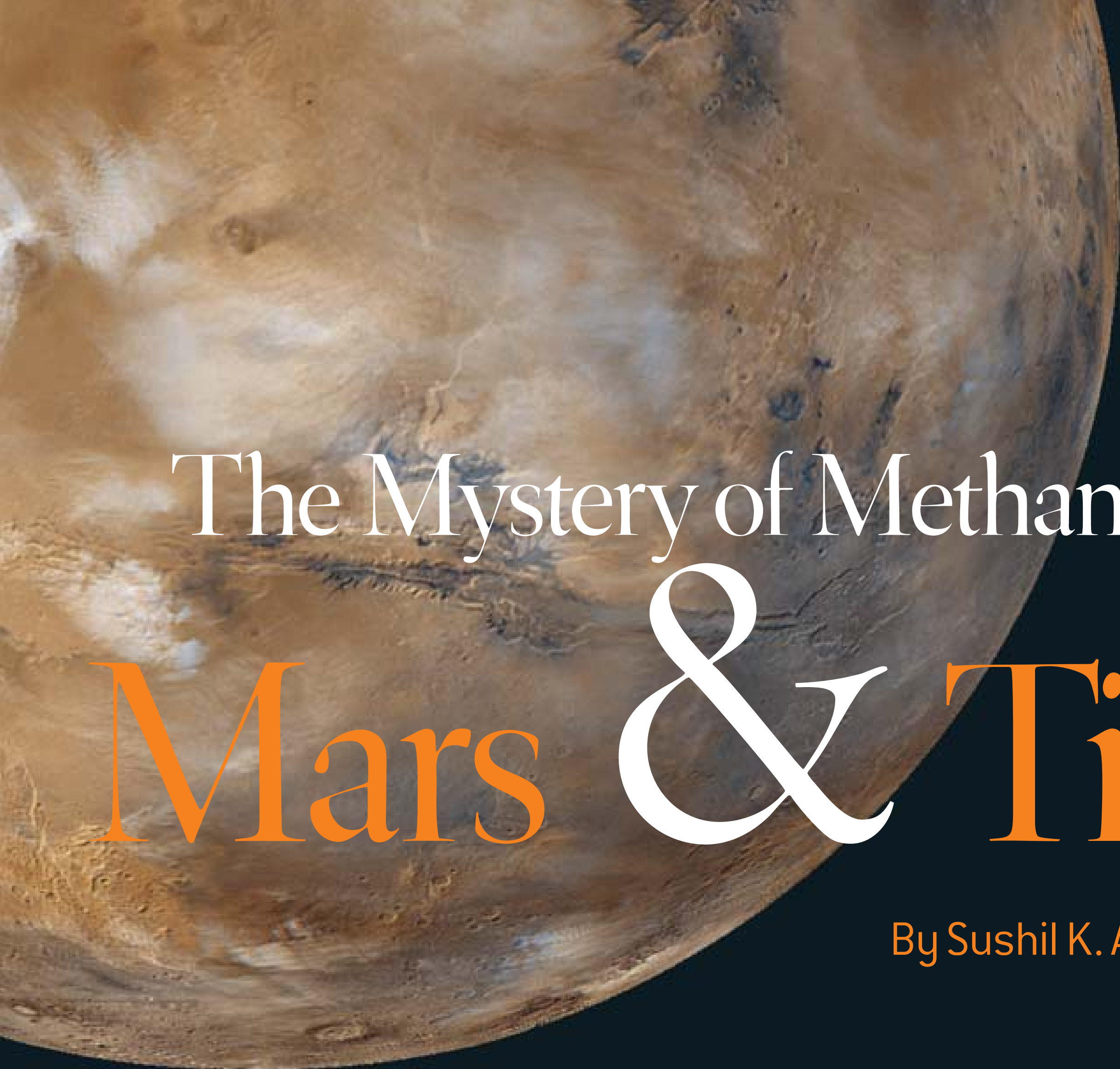
And time is running out for those Indians. Brazil has demarcated less than a dozen lands with isolated peoples. Possuelo and his colleagues have identified 22 isolated groups in the Amazon, although FUNAI believes that the true number might be closer to 68, based on aerial surveys and witness accounts. Some live in previously protected indigenous lands, some in areas that are right in the range of agricultural expansion.

In the other six South American countries with isolated peoples—Ecuador, Peru, Bolivia, Paraguay, Colombia and Venezuela—the prospect is even bleaker. "The issue is all but unknown to them," Possuelo states. "And it is no use to undertake protection policies here if indigenous peoples are not protected across the border and get killed on the other side."

In 2005, before leaving FUNAI, Possuelo organized a meeting in the Amazon city of Belém that gave birth to an alliance for the protection of isolated peoples in the seven countries. The alliance, consisting of state attorneys, environmental defense groups, anthropologists and indigenous organizations, called on their governments to identify and protect their indigenous charges. "The idea here is to raise public consciousness in the first place," he says. "Then we'll be able to talk about public policy." The alliance has been facing an unexpected challenge, though—from indigenous peoples themselves. "In Ecuador last year, contacted Indians killed 20 isolated Indians," says Marcelo dos Santos, who took over Possuelo's old FUNAI post.

Possuelo agrees that the isolated groups may not last much longer. But pointing at several figurines of Don Quixote that he keeps in his living room, he says he has not given up. "The fate of those peoples will depend ultimately on our choices," he argues. "Those peoples are the last Americans. We are indebted to them." SA

Claudio Angelo is science editor of the Brazilian daily newspaper Folha de S. Paulo. *A Q&A version of his interview with Possuelo is at www.sciam.com/ontheweb*



The Mystery of Methane on Mars & Titan

By Sushil K. Atreya

MARS has long been thought of as a possible abode of life. The discovery of methane in its atmosphere has rekindled those visions. The visible face of Mars looks nearly static, apart from a few wispy clouds (*white*). But the methane hints at a beehive of biological or geochemical activity underground.

Of all the planets in the solar system other than Earth, Mars has arguably the greatest potential for life, either extinct or extant. It resembles Earth in so many ways: its formation process, its early climate history, its reservoirs of water, its volcanoes and other geologic processes. Microorganisms would fit right in. Another planetary body, Saturn's largest moon Titan, also routinely comes up in discussions of extraterrestrial biology. In its primordial past, Titan possessed conditions conducive to the formation of molecular precursors of life, and some scientists believe it may have been alive then and might even be alive now.

To add intrigue to these possibilities, astronomers studying both these worlds have detected a gas that is often associated with living things: methane. It exists in small but significant quantities on Mars, and Titan is literally awash with it. A biological source is at least as plausible as a geologic one, for Mars if not for Titan. Either explanation would be fascinating in its own way, revealing either that we are not alone in the universe or that both Mars and Titan harbor large underground bodies of water together with unexpected levels of geochemical activity. Understanding the origin and fate of methane on these bodies will provide crucial clues to the processes that shape the formation, evolution and habitability of terrestrial worlds in this solar system and possibly in others.

Methane (CH_4) is abundant on the giant planets—Jupiter, Saturn, Uranus and Neptune—where it was the product of chemical processing of primordial solar nebula material. On Earth, though, methane is special. Of the 1,750 parts per billion by volume (ppbv) of methane in Earth's atmosphere, 90 to 95 percent is biological in origin. Grass-eating ungulates such as cows, goats and yaks belch out one fifth of the annual global methane release; the gas is a metabolic by-product of the bacteria in their guts. Other significant sources include termites, rice paddies, swamps,

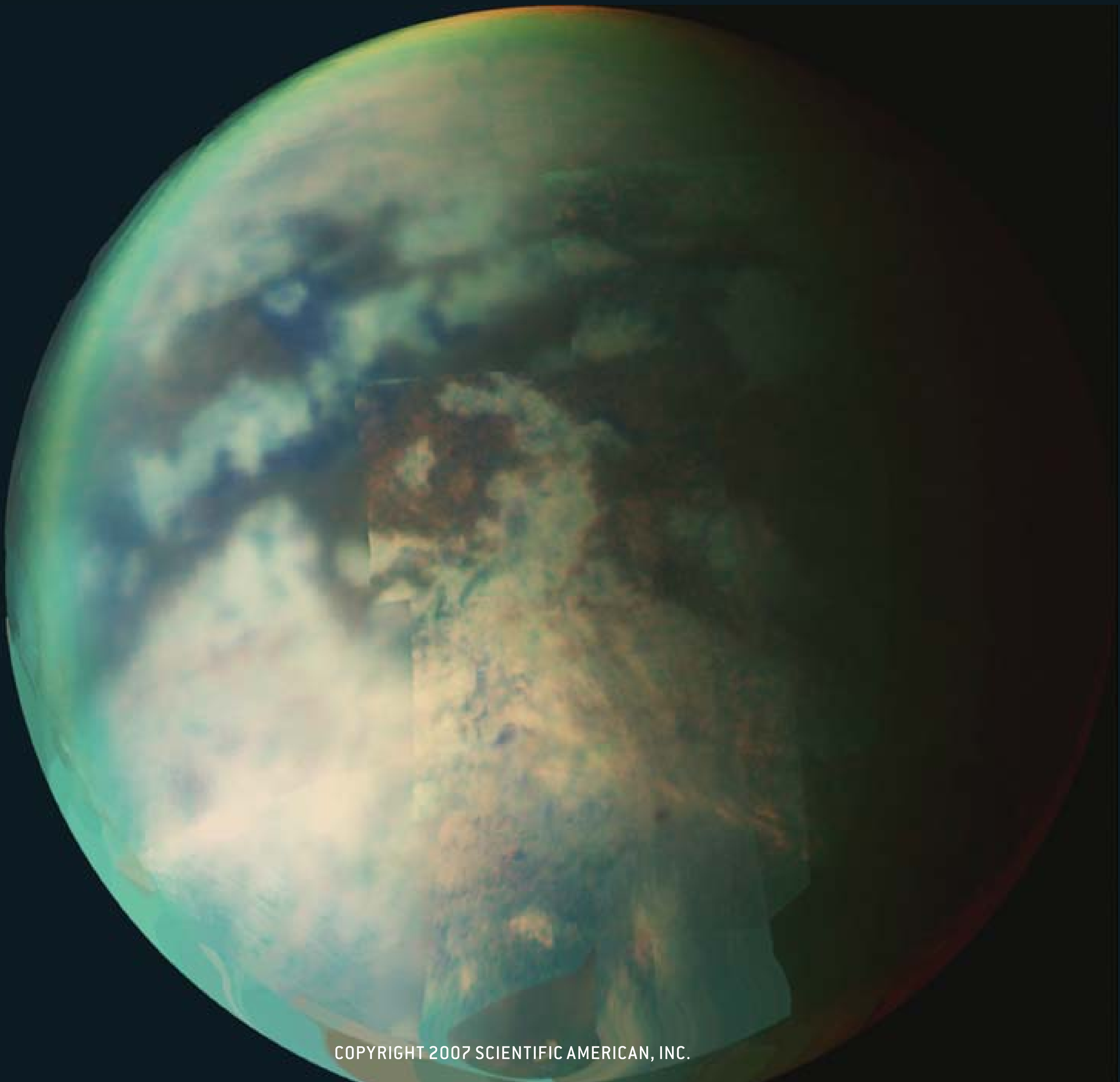
It might mean life,

it might mean unusual geologic activity; whichever it is, the

presence of methane in the atmospheres

of Mars and Titan is one of the most tantalizing puzzles in our solar system

TITAN, technically a satellite of Saturn but for all intents and purposes a full-fledged planet, has a nitrogen atmosphere denser than Earth's and a surface sculpted by tectonic activity and rivers of liquid methane. Where the methane comes from, no one knows for sure. The Cassini space probe took this composite infrared image last year.



leakage of natural gas (itself a result of past life) and photosynthetic plants [see "Methane, Plants and Climate Change," by Frank Keppler and Thomas Röckmann; Scientific American, February 2007]. Volcanoes contribute less than 0.2 percent of the total methane budget on Earth, and even they may simply be venting methane produced by organisms in the past. Abiotic sources such as industrial processes are comparatively minor. Thus, detection of methane on another Earth-like object naturally raises the prospect of life on that body.

In the Air

THAT IS WHAT HAPPENED with Mars in 2003 and 2004, when three independent groups of scientists announced the discovery of methane in the atmosphere of that planet. Using a high-resolution spectrograph at the Infrared Telescope Facility in Hawaii and at the Gemini South Telescope in Chile, a team led by Michael Mumma of the NASA Goddard Space Flight Center detected methane concentrations in excess of 250 ppbv, varying over the planet and perhaps over time. Vittorio Formisano of the Institute of Physics and Interplanetary Science in Rome and his colleagues (including me) analyzed thousands of infrared spectra collected by the Mars Express orbiter. We found methane to be much less abundant, ranging from zero to about 35 ppbv, with a planetary average of approximately 10 ppbv. Finally, Vladimir Krasnopolsky of the Catholic University of America and his colleagues, using the Canada-France-Hawaii Telescope, measured a planetary average of about 10 ppbv. They could not determine the variation over the planet because of poor signal and spatial resolution.

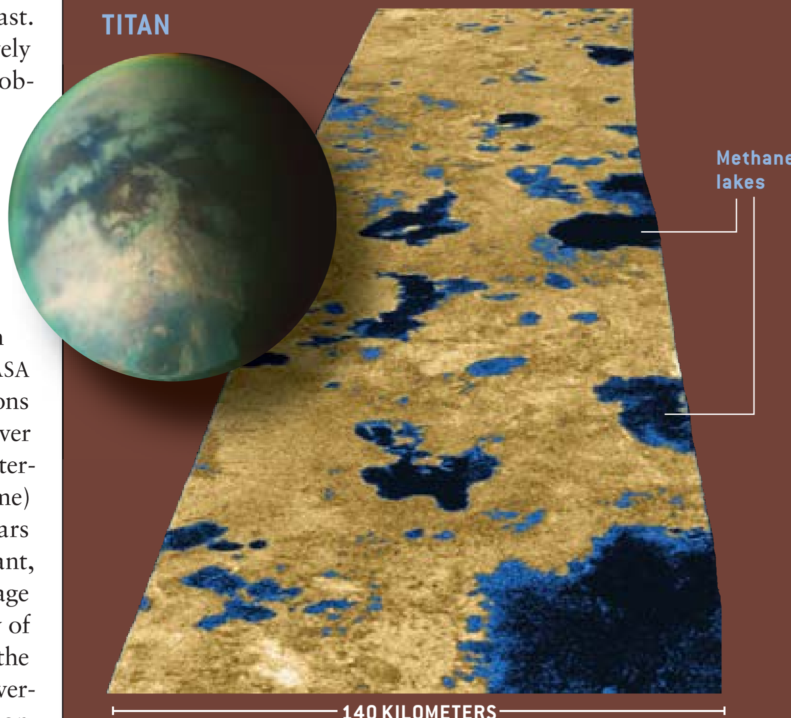
Mumma's team is now reanalyzing its data to try to determine why its value is the outlier. For now, I will take the 10 ppbv value as the most likely. It corresponds to a concentration of methane (in molecules per unit volume) that is only 40 millionths of the concentration in Earth's atmosphere. Nevertheless, even the barest presence of the gas demands an explanation.

Although astronomers detected methane on Titan as early as 1944, it was only the additional discovery of nitrogen 36 years later that generated the immense interest in this cold and distant moon [see "Titan," by Tobias Owen; Scientific American, February 1982]. Nitrogen is a key constituent of biological molecules such as amino acids and nucleic acids. A body with a nitrogen-methane atmosphere, where the ground-level pressure is one and a half times that of our home planet,

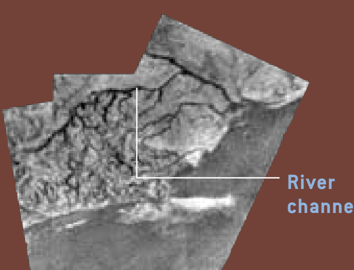
SNAPSHOTS FROM TITAN AND MARS

Astronomers discovered methane in Titan's atmosphere in the 1940s, but dense haze blocked their view of the surface. The Cassini-Huygens mission has now seen the extent to which methane shapes the terrain.

TITAN



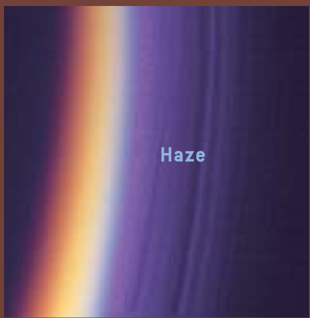
LARGE BODIES of liquid, probably methane, show up in these Cassini radar images of Titan's far northern hemisphere. Liquid appears dark (*colorized blue*) for the same reason that a wet road looks dark when you are driving at night: the smooth liquid surface reflects the headlight beams away from your eye. Conversely, dry, rough terrain appears bright (*tan*). This image has a resolution of 500 meters.



RIVER CHANNELS may have been carved by liquid methane flowing from a series of ridges (about 200 meters high) down to a lakebed (now dry). The pattern of tributaries suggests that the methane came from rainfall. The Huygens probe captured this image from an altitude of 6.5 kilometers as it descended through the atmosphere.

Overview/*Methane*

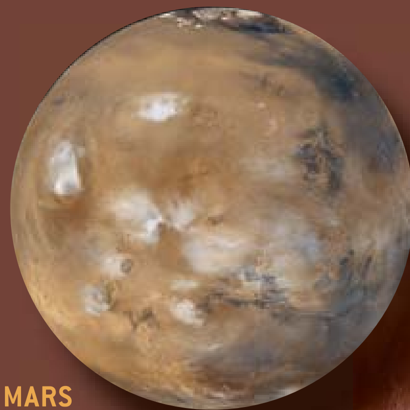
- Astronomers have been talking about life on Mars for a century or longer, but seldom with the benefit of hard data. That situation changed in 2003 with the discovery of methane in the atmosphere. Some ongoing process must pump it out to offset its steady destruction by sunlight.
- Researchers have narrowed the possibilities to two. The first is Martians—specifically, methane-belching bacteria like those in cows' guts on Earth. The second is a rock-water reaction called serpentinization, which occurs in the black smokers on Earth's seafloor. The latter possibility may seem like a letdown but would be a dramatic discovery in its own right. A new rover scheduled for launch in 2009 may be able to settle the issue.
- A similar debate swirls around Saturn's largest satellite, Titan. In 2005 the Huygens space probe showed that methane plays much the same role there as water does on Earth. The methane may come from geochemical reactions in a vast hidden ocean.



HAZE in the upper atmosphere of Titan consists of hydrocarbons created when sunlight strikes methane. The haze resembles urban smog. This image has a resolution of 700 meters.

TITAN'S SURFACE had never been glimpsed before the Huygens probe landed in January 2005. It may look like a boring field of rocks, but the "rocks" are in fact fist-size chunks of ice and, on close examination, show signs of erosion by flowing liquid—probably methane. As the probe warmed the soil, methane oozed out.

On the Red Planet, methane exists only at the parts-per-billion-by-volume (ppbv) level, so it cannot be seen as directly as on Titan. Various observed processes may destroy or create it.



MARS

DUST DEVILS, such as this one spied by the Mars Global Surveyor on May 21, 2002, rub dust grains together and thus create strong static-electric fields—which can tear apart water molecules and lead to the production of methane-destroying peroxides.

UNDERGROUND WATER FLOW in the Martian past would explain this whitish ridge (*arrow*) recently seen by the Mars Reconnaissance Orbiter. Water flowing through a rock fracture would have deposited minerals, much as hard water causes buildup in a household pipe. When surrounding rock eroded away, the minerals remained, leaving a ridge. Underground water might also have facilitated methane production.

METHANE DATA	EARTH	MARS	TITAN
Atmospheric concentration	1,750 ppbv	10 ppbv	5 percent
Molecular lifetime in atmosphere (years)	10	600	10 million
Production rate needed to sustain constant amount (tons per year)	515 million	125	25 million
Main sources	Cattle, termites, swamps, rice paddies, natural gas	Bacteria? Rock-water reactions in aquifer?	Rock-water reactions in underground ocean

NASA/JPL/UNIVERSITY OF ARIZONA (*Titan globe on opposite page and on preceding pages and ridge*); NASA/JPL/USGS (*lakes*); ESA/NASA/JPL/UNIVERSITY OF ARIZONA (*river and Titan surface*); NASA/JPL/MSSS (*Mars globe on this page and on preceding pages and dust devil*); NASA/JPL/SPACE SCIENCE INSTITUTE (*haze*)

may have the right ingredients for molecular precursors of life and, some have speculated, even life itself to form.

Methane plays a central, controlling role in maintaining Titan's thick nitrogen atmosphere. It is the source of hydrocarbon hazes, which absorb solar infrared radiation and warm the stratosphere by approximately 100 degrees Celsius, and of hydrogen, whose molecular collisions result in a 20-degree warming in the troposphere. If the methane ever ran out, temperatures would drop, nitrogen gas would condense into liquid droplets and the atmosphere would collapse. Titan's special character would change forever. Its smog and clouds would dissipate. The methane rain that seems to have carved its surface would stop. Lakes, puddles and streams would dry up. And, with its veil lifted, Titan's stark surface would lay bare and readily accessible to telescopes on Earth. Titan would lose its mystique and turn into just another satellite with thin air.

Could it be that methane on Mars and Titan has a biological origin, as on Earth, or does it have another explanation, such as volcanoes or impacts of comets and meteorites? Our understanding of geophysical, chemical and biological processes has helped narrow the field of possible sources on Mars, and many of the same arguments apply to Titan as well.

Split by Sunlight

THE FIRST STEP to answering the question is to determine the rate at which methane must be produced or delivered. That, in turn, depends on how fast the gas is being removed from the atmosphere. At altitudes of 60 kilometers and higher above the Martian surface, solar ultraviolet radiation splits methane molecules apart. Lower in the atmosphere, oxygen atoms and hydroxyl radicals (OH), which form when water molecules are broken apart by ultraviolet photons, oxidize methane. Without being resupplied, methane would gradually disappear from the atmosphere. The "lifetime" of methane—defined as the time it takes for the gas concentration to drop by a factor of the mathematical constant *e*, or roughly three—is 300 to 600 years, depending on the amount of water vapor, which undergoes seasonal changes, and on the strength of solar radiation, which varies during the solar cycle. On Earth, similar processes give methane a lifetime of about 10 years. On Titan, where solar ultraviolet radiation is much weaker and oxygen-bearing molecules are substantially less abundant, methane can last 10 million to 100 million years (which is still a short time in geologic terms).

Methane's lifetime on Mars is long enough for winds and diffusion to mix the gas into the atmosphere fairly uniformly. Thus, the observed variations of methane levels over the planet are puzzling. They may be a sign that the gas comes from localized sources or disappears into localized sinks. One possible sink is chemically reactive soil, which could accelerate the loss of methane. If such additional sinks operated, it would take an even larger source to maintain the observed abundance.

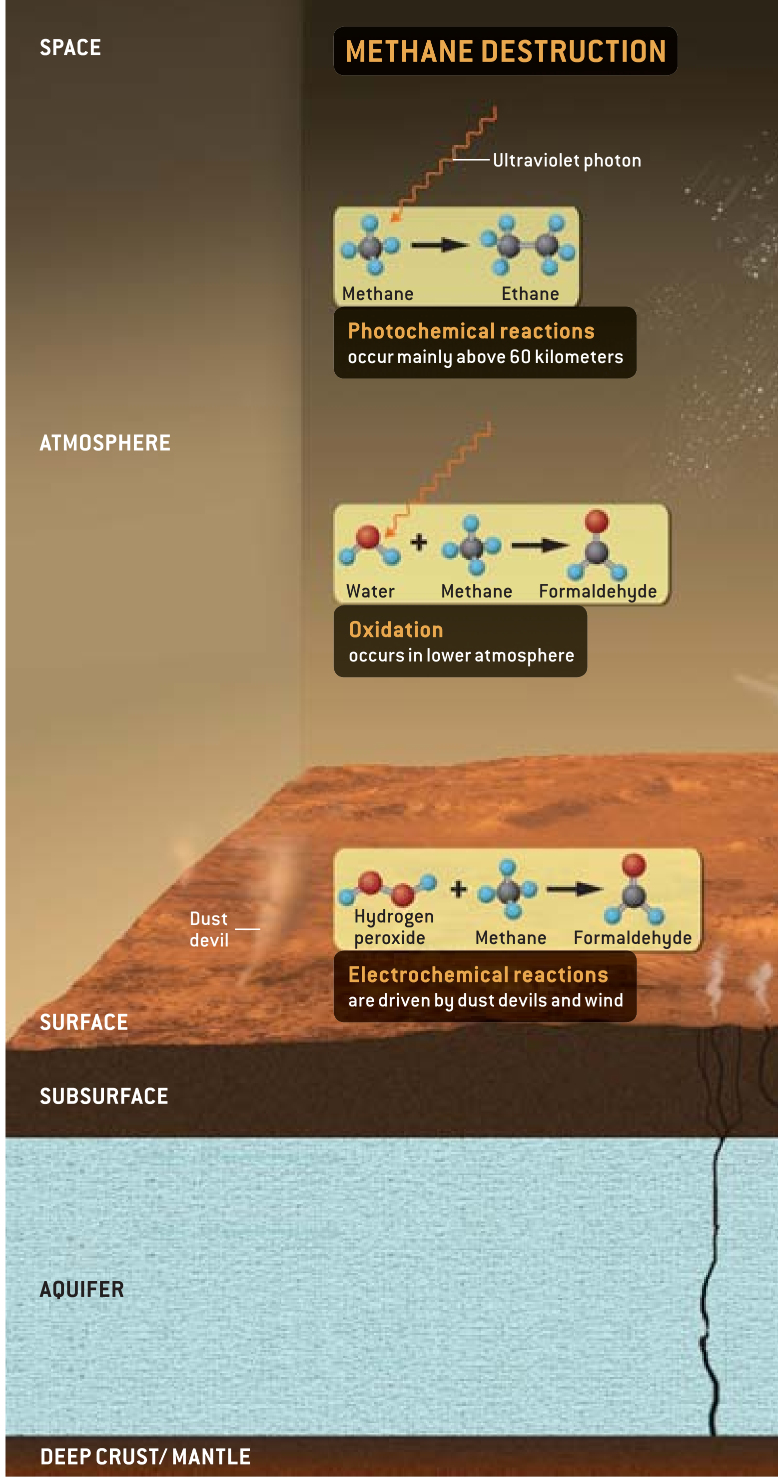
The next step is to consider potential scenarios for forming methane. The Red Planet is a good place to start because its methane abundance is so low. If a mechanism cannot explain even this small amount, it would be unlikely to account for Titan's much greater quantity. For a 600-year lifetime, a little over 100 metric tons of methane would have to be produced each year to maintain a constant global average of 10 ppbv. That is about a quarter-millionth the production rate on Earth.

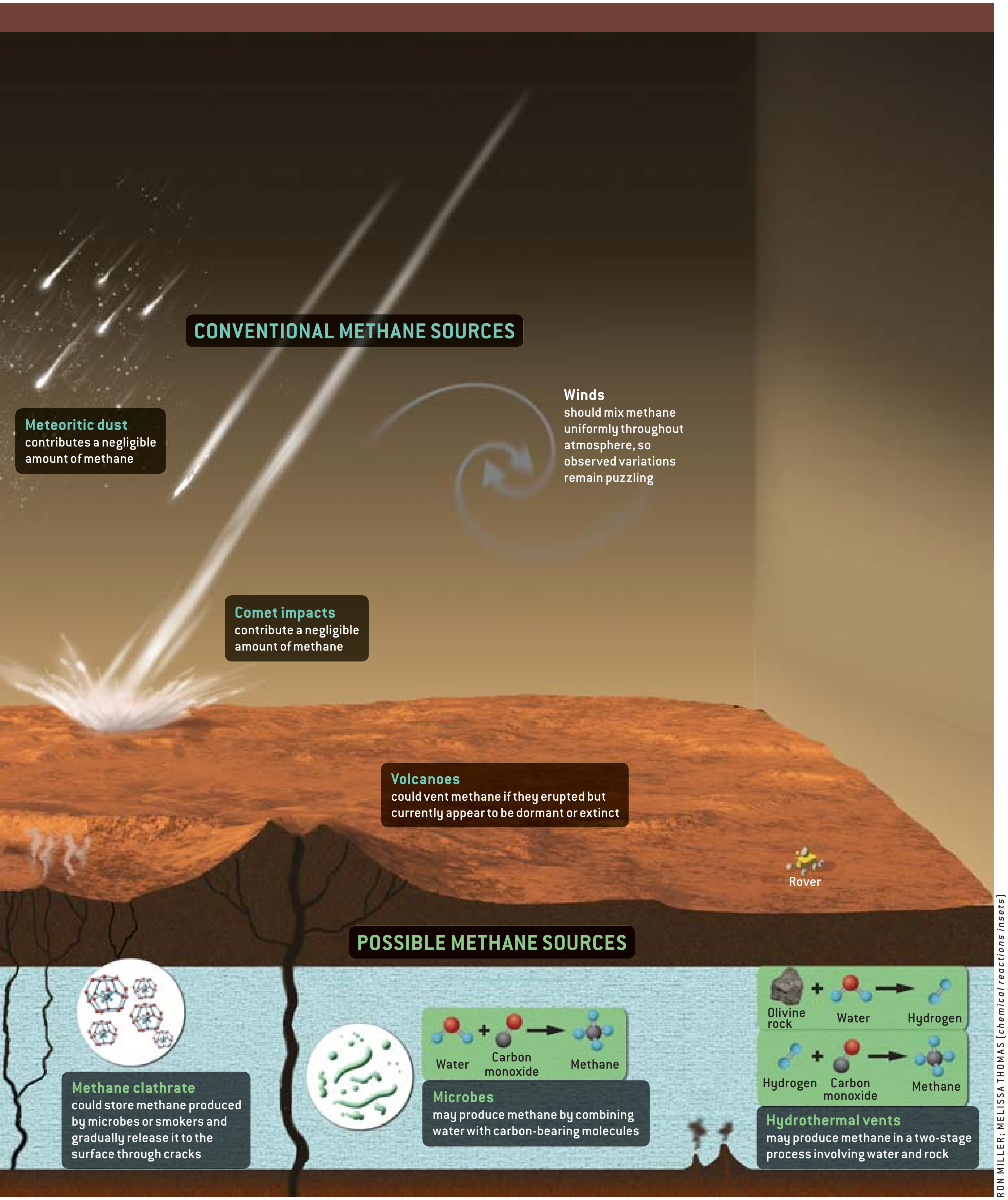
As on Earth, volcanoes are most likely not responsible. Martian volcanoes have been extinct for hundreds of millions of years. Furthermore, if a volcano had been responsible for the methane, it would also have pumped out enormous quantities of sulfur dioxide, and Mars's atmosphere is devoid of sulfur compounds. Extraplanetary contributions also appear minimal. Some 2,000 tons of micrometeoritic dust are estimated to reach the Martian surface every year. Less than 1 percent of their mass is carbon, and even this material is largely oxidized and hence an insignificant source of methane. Comets are about 1 percent methane by weight, but they strike Mars only once every 60 million years on average. Thus, the amount of methane delivered would be about one ton a year, or less than 1 percent of the required amount.

Could it be that a comet struck Mars in the recent past? It could have delivered a large amount of methane, and over time the abundance in the atmosphere would have declined to its present value. An impact of a comet 200 meters in diameter 100 years ago, or a comet 500 meters in diameter 2,000 years

METHANE ON MARS

By all rights, Mars should have zero methane. The gas is quickly cleansed from the air by chemical reactions driven by sunlight or weather patterns, and known geologic and astronomical processes cannot replenish it fast enough. Thus, the methane hints at unseen activity, such as black smokers or methane-creating microbes swimming in underground bodies of water.




CONVENTIONAL METHANE SOURCES
Meteoritic dust
contributes a negligible amount of methane
Winds
should mix methane uniformly throughout atmosphere, so observed variations remain puzzling
Comet impacts
contribute a negligible amount of methane
Volcanoes
could vent methane if they erupted but currently appear to be dormant or extinct
Rover
POSSIBLE METHANE SOURCES
Methane clathrate
could store methane produced by microbes or smokers and gradually release it to the surface through cracks
Water
Carbon monoxide
Methane
Microbes
may produce methane by combining water with carbon-bearing molecules
Olivine rock
Water
Hydrogen
Hydrogen
Carbon monoxide
Methane
Hydrothermal vents
may produce methane in a two-stage process involving water and rock
RON MILLER; MELISSA THOMAS (chemical reactions insets)

ago, could have supplied sufficient methane to account for the currently observed global average value of 10 ppbv. But this idea runs into a problem: the distribution of methane is not uniform over the planet. The time it takes to distribute methane uniformly vertically and horizontally is at most several months. Thus, a cometary source would result in a uniform methane distribution over Mars, contrary to observations.

Smoke in the Waters

THAT LEAVES US WITH two possible sources: hydrogeochemical and microbial. Either one would be fascinating. Hydrothermal vents, known as black smokers, were first discovered on Earth in 1977 on the Galápagos Rift [see "The Crest of the East Pacific Rise," by Ken C. Macdonald and Bruce P. Luyendyk; SCIENTIFIC AMERICAN, May 1981]. Since then, oceanographers have found them along many other mid-oceanic ridges. Laboratory experiments show that under the conditions prevailing at these vents, ultramafic silicates—rocks rich in iron or magnesium, such as olivine and pyroxene—can react to produce hydrogen in a process commonly referred to as serpentinization. In turn, reaction of hydrogen with carbon grains, carbon dioxide, carbon monoxide or carbonaceous minerals can produce methane.

The keys to this process are hydrogen, carbon, metals (which act as catalysts), and heat and pressure. All are available on Mars, too. The process of serpentinization can occur either at high temperatures (350 to 400 degrees C) or at milder ones (30 to 90 degrees C). These lower temperatures are estimated to occur within purported aquifers on Mars.

Although low-temperature serpentinization may be capable of producing the Martian methane, biology remains a serious possibility. On Earth, microorganisms known as methanogens produce methane as a by-product of consuming hydrogen, carbon dioxide or carbon monoxide. If such organisms lived on Mars, they would find a ready supply of nutrients: hydrogen (either produced in the serpentinization process or diffusing into the soil from the atmosphere) plus carbon dioxide and carbon monoxide (in the rocks or from the atmosphere).

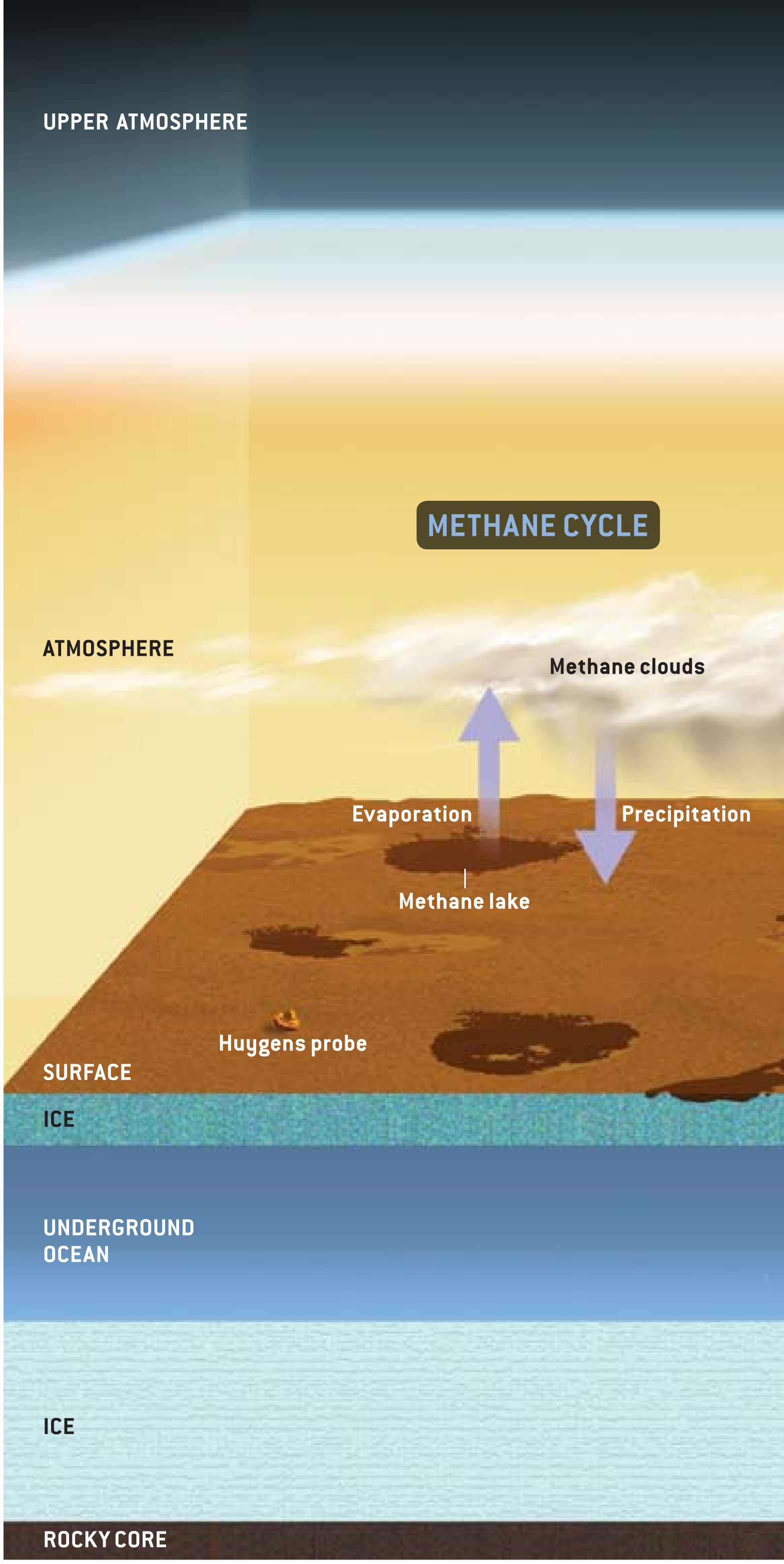
Once formed by either serpentinization or microbes, methane could be stored as a stable clathrate hydrate—a chemical

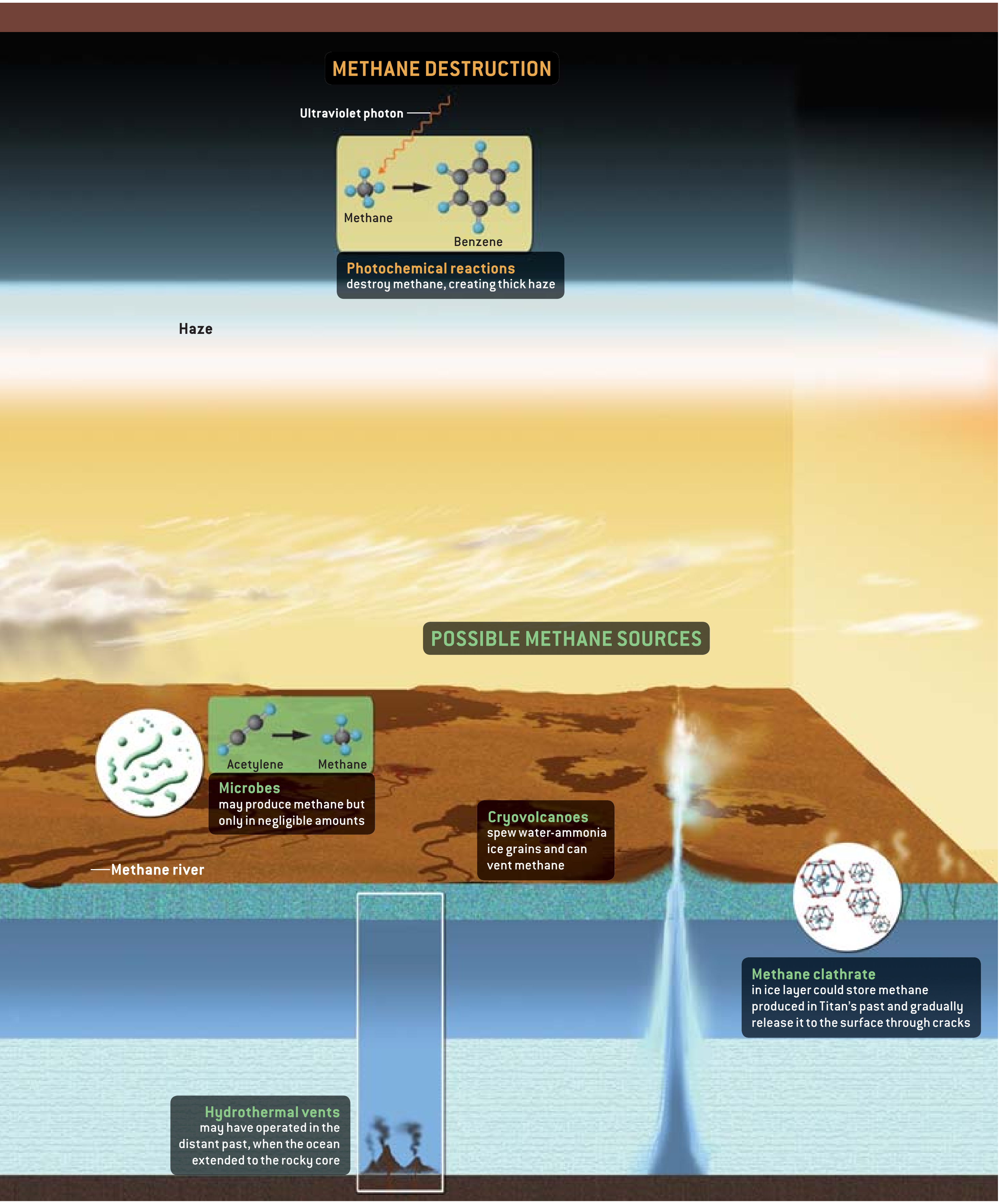
THE AUTHOR

SUSHIL K. ATREYA began his space career on the science teams of the Voyager missions to the giant planets, continuing with Galileo, Cassini-Huygens, Venus Express, Mars Express, Mars Science Laboratory (scheduled for launch in 2009) and Juno-Jupiter Polar Orbiter (2011). His research focuses on the origin and evolution of atmospheres and the formation of planetary systems. He is a professor at the University of Michigan at Ann Arbor, a Fellow of the American Association for the Advancement of Science, and a Distinguished Visiting Scientist at the Jet Propulsion Laboratory. When not exploring other worlds, Atreya likes to go hiking in the mountains and cycling. He expresses his gratitude to Chloé Atreya, Wesley Huntress, Paul Mahaffy, Inge ten Kate, Henry Pollack and Elena Adams for discussions and comments on the drafts of this article.

METHANE ON TITAN

Methane is to Titan what water is to Earth: a substance that cuts rivulets into the surface, pools in standing bodies, evaporates into the air and rains back down. As on Mars, chemical reactions deplete Titan's methane, so some geologic or biological activity must replenish it.





RON MILLER; MELISSA THOMAS (*chemical reactions insets*)

structure that traps methane molecules like animals in a cage—for later release to the atmosphere, perhaps by gradual outgassing through cracks and fissures or by episodic bursts triggered by volcanism. No one is sure how efficiently the clathrates would form or how readily they would be destabilized.

The Mars Express observations hint at greater methane concentrations over areas containing subsurface water ice. Either the geologic or biological scenario would explain this correlation. Aquifers below the ice would provide a habitat for creatures or a venue for the hydrogeochemical production of methane. Without more data, the biological and geologic possibilities appear equally likely.

A Titanic Ocean

AT FIRST GLANCE, one might think that Titan's methane would be easier to understand: the moon formed in the subnebula of Saturn, whose atmosphere contains huge amounts of the gas. Yet the data argue for production of methane on Titan rather than delivery of methane to Titan. The Huygens probe of the joint NASA and European Space Agency's Cassini-Huygens Mission found no xenon or krypton in the moon's atmosphere. Had the planetesimals that formed Titan brought methane, they would have brought these heavy noble gases as well. The absence of such gases indicates that methane most likely formed on Titan.

Therefore, the presence of methane on Titan is as mysterious as it is on Mars—in some respects more so because of its sheer quantity (5 percent by volume). A plausible source, as on Mars, is serpentinization at relatively low temperatures. Christophe Sotin of the University of Nantes in France and his colleagues have argued that Titan might sustain an underground ocean of liquid water [*see box on two preceding pages*]. Dissolved ammonia, acting as an antifreeze, would help to keep it from freezing solid. In their model, the ocean is 100 kilometers underneath Titan's surface and 300 to 400 kilometers deep. In the past, the decay of radioactive elements and the leftover heat from Titan's formation might have melted nearly all the body's ice—so the ocean might have extended all the way down to the rocky core.

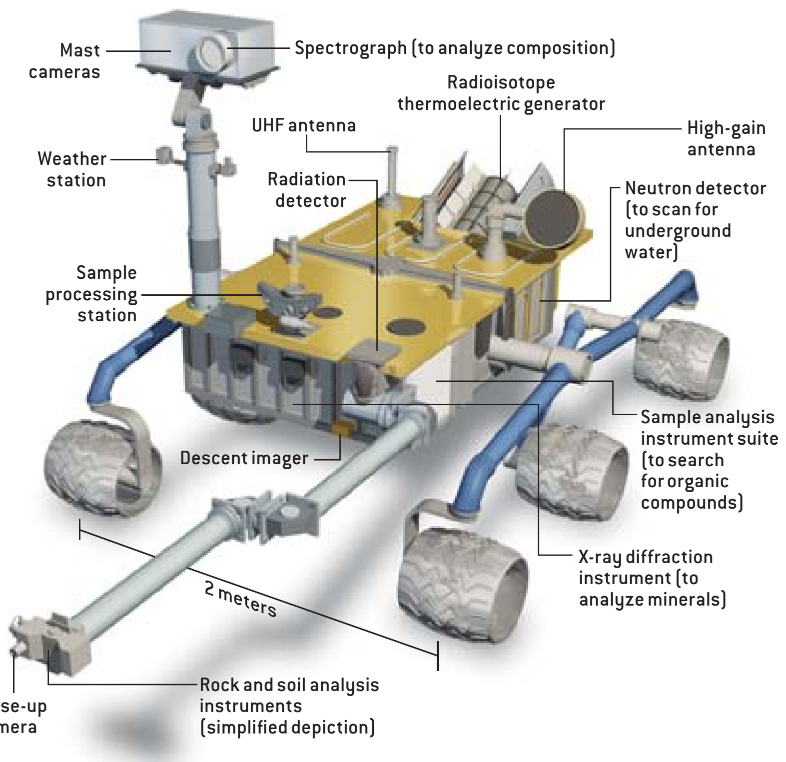
Under those conditions, reactions between the water and the rock would have liberated hydrogen gas, which in turn would have reacted with carbon dioxide, carbon monoxide, carbon grains or other carbonaceous material—producing methane. I estimate that this process would have been capable of explaining Titan's observed methane abundance. Once produced, methane could have been stored as a stable clathrate hydrate and released to the atmosphere either gradually, through volcanism, or in bursts, triggered by impacts.

An intriguing clue is the argon 40 gas detected by Huygens as it descended through Titan's atmosphere. This isotope forms by the radioactive decay of potassium 40, which is sequestered in the rocks deep in Titan's core. Because the radioactive half-life of potassium 40 is 1.3 billion years, the small amount of argon 40 in the atmosphere is evidence for slow release of gases from the interior. In addition, optical and radar images of the surface show signs of cryovolcanism—geyserlike eruptions of ammonia-water ice—which also indicates that material wells up from the interior. The surface appears relatively young and free of craters, which is a sign of resurfacing by material from the interior. The estimated resurfacing rate would release methane from the interior quickly enough to balance the photochemical loss. Methane on Titan plays the role of water on Earth, complete with liquid surface reservoirs, clouds and rain—a full-fledged methalogical cycle. Thus, a substantial body of evidence exists, even more so than for Mars, that methane stored in the interior would have no difficulty getting out to the surface and subsequently evaporating into the atmosphere.

Might biology also play a role in creating Titan's methane?

NASA'S NEXT ROVER

The next step to solving the problem of methane on Mars is NASA's Mars Science Laboratory, scheduled for launch in 2009. Its instruments should be able to measure the methane's isotopic makeup (a clue to its origin) and sift through soil looking for organic compounds (the composition of which would indicate which chemical or biological processes have created or destroyed methane).



DON FOLEY; SOURCE: NASA/JPL

Christopher McKay of the NASA Ames Research Center and Heather Smith of the International Space University in Strasbourg, France, as well as Dirk Schulze-Makuch of Washington State University and David Grinspoon of the Denver Museum of Nature and Science, have suggested that acetylene and hydrogen could serve as nutrients for methanogens even in the extreme cold of Titan's surface (–179 degrees C). This biogenic process differs from that employed by methanogens on Earth and their cousins, if any, on Mars in that no water is needed. Instead liquid hydrocarbons on Titan's surface serve as the medium.

Yet this hypothesis has a shortcoming. Huygens data rule out an underground source of acetylene; this compound must ultimately come from methane in the atmosphere. Thus, it seems like a circular argument: to produce methane (by microbes), one needs methane. Moreover, the sheer abundance of methane on Titan is so immense that methanogens would have to work in overdrive to produce it, severely depleting the available nutrients.

In view of these obstacles, a biological explanation for methane is much less attractive on Titan than on Mars. Nevertheless, the hypothesis of habitability bears investigating. Some scientists argue that this moon might have been or still be habitable. It receives enough sunlight to turn nitrogen and methane into molecules that are the precursors to biology. An underground water-ammonia brine, with some methane and other hydrocarbons thrown in, could be a friendly environment for complex molecules or even living organisms. In the distant past, when the young Titan was still cooling off, liquid water may even have flowed on the surface.

The Mars Science Laboratory rover will study solid and gaseous samples for **chemical signs of past or present life**.

Organic Food

ONE CRUCIAL MEASUREMENT that could help determine the sources of methane on Mars and Titan is the carbon isotope ratio. Life on Earth has evolved to prefer carbon 12, which requires less energy for bonding than carbon 13 does. When amino acids combine, the resulting proteins show a marked deficiency in the heavier isotope. Living organisms on Earth contain 92 to 97 times as much carbon 12 as carbon 13; for inorganic matter, the standard ratio is 89.4.

On Titan, however, the Huygens probe measured a ratio of 82.3 in methane, which is smaller, not larger, than the terrestrial inorganic standard value. This finding argues strongly against the presence of life as we know it. To be sure, some scientists suggest that life could have evolved differently on Titan than on Earth or that the inorganic isotope ratio may be different there.

No one has yet determined the carbon isotope ratio for Mars. This measurement is challenging when the concentration of the gas is so low (one billionth of that on Titan). NASA's Mars Science Laboratory (MSL) rover, scheduled to arrive at Mars in 2010, should be able to carry out precise measurements of the carbon isotopes in methane and possibly other organic materials. It will also study solid and gaseous samples for other chemical signs of past or present life, such as a very high abundance ratio of methane to heavier hydrocarbons (ethane, propane, butane) and chirality (a preference for either left-handed or right-handed organic molecules).

Tied up with these issues is the question of why organics seem to be missing from the surface of Mars. Even in the absence of life, meteorites, comets and interplanetary dust particles should have delivered organics over the past four and a half billion years. Perhaps the answer lies in Martian dust devils and storms and ordinary saltation (the hopscotching of windblown dust grains). These processes generate strong static electric fields, which can trigger the chemical synthesis of hydrogen peroxide. Being a potent antiseptic, hydrogen peroxide would quickly sterilize the surface and scrub out the organics. The oxidant would also accelerate the loss of methane locally from the atmosphere, thus requiring a larger source to explain the abundances observed in the Martian atmosphere.

In summary, methane serves as the glue that holds Titan together in some mysterious ways. The presence of methane on Mars is equally intriguing, not the least because it evokes visions of life on that planet. Future exploration of both bodies will seek to determine whether they were ever habitable. Although life as we know it can produce methane, the presence of methane does not necessarily signify the existence of life. So planetary scientists must investigate thoroughly the sources, sinks and isotopic composition of this gas, along with other organic molecules and trace constituents in both gaseous and solid samples. Even if methane is found to have no connection to life, studying it will reveal some of the most fundamental aspects of the formation, climate histories, geology and evolution of Mars and Titan. SA

MORE TO EXPLORE

Detection of Methane in the Atmosphere of Mars. Vittorio Formisano, Sushil Atreya, Thérèse Encrenaz, Nikolai Ignatiev and Marco Giuranna in *Science,* Vol. 306, pages 1758–1761; October 28, 2004.

A Sensitive Search for SO_2 in the Martian Atmosphere: Implications for Seepage and Origin of Methane. Vladimir A. Krasnopolsky in *Icarus,* Vol. 178, No. 2, pages 487–492; November 2005.

Episodic Outgassing as the Origin of Atmospheric Methane on Titan. Gabriel Tobie, Jonathan I. Lunine and Christophe Sotin in *Nature,* Vol. 440, pages 61–64; March 2, 2006.

Titan's Methane Cycle. Sushil K. Atreya, Elena Y. Adams, Hasso B. Niemann, Jaime E. Demick-Montelara, Tobias C. Owen, Marcello Fulchignoni, Francesca Ferri and Eric H. Wilson in *Planetary and Space Science,* Vol. 54, No. 12, pages 1177–1187; October 2006.

Methane and Related Trace Species on Mars: Origin, Loss, Implications for Life, and Habitability. Sushil K. Atreya, Paul R. Mahaffy and Ah-San Wong in *Planetary and Space Science,* Vol. 55, No. 3, pages 358–369; February 2007.

Sushil K. Atreya's Web page: **www.umich.edu/~atreya**

Chromosomal Chaos and Cancer

Current wisdom on the role of genes in malignancy may not explain some features of cancer, but stepping back to look at the bigger picture inside cells reveals a view that just might

By Peter Duesberg

When I first began to study cancer as a young postdoctoral fellow in the early 1960s, it looked to leading scientists as though viruses could be the cause of most, if not all, malignancies. That idea was based on the discovery of several tumor- and leukemia-producing viruses that could infect a host cell and insert their own genetic material into its genome, sparking a cancerous transformation and proliferation of the cell. I was optimistic and naive enough to hope that if researchers could understand the exact molecular mechanisms by which such viruses caused cancer, we could develop vaccines to eliminate one of humanity's most dreaded diseases.

My own contribution to that pursuit came in 1970, when my colleagues, Michael Lai and Peter Vogt, and I managed to isolate a specific gene, *src,* which was suspected to be the tumor-initiating culprit in avian Rous sarcoma virus. Within a few years, more creative scientific minds than mine had followed this lead to a realization that a closely related gene was already present in the normal DNA of animals, including humans. And a new cancer model was born: it proposed that some triggering event, such as a mutation in a human cell's own version of *src,* could ignite tumorigenic powers like those possessed by its viral counterpart. The cancer-promoting potential of such a time bomb buried in our personal genomes earned it the title of "proto-oncogene." Once the mutation occurred, it would become a full-fledged oncogene.

The theory that mutations in certain key human genes are at the root of all cancers has dominated research for the past 30 years. Yet despite all the attempts of investigators, including myself, during that time to demonstrate that a handful of such oncogenes alone can transform normal cells into malignant ones, none

JEN CHRISTIANSEN

THE AUTHOR

PETER DUESBERG is a professor of molecular and cell biology at the University of California, Berkeley, where he arrived from Germany in 1964 as a research virologist. Within six years he had isolated the first true oncogene, from within the Rous sarcoma virus, and mapped the genetic structure of the entire virus. He proceeded to do the same for 10 more mouse and avian sarcoma and leukemia viruses and was elected to the National Academy of Sciences in 1986. By 1987 his work with retroviruses led him to conclude that HIV is merely a bystander and AIDS results from chemical exposures and malnutrition. His ongoing work with cancer viruses also persuaded him that mutations in individual genes are insufficient to cause the malignant transformations seen in cancer.

Editors' note: The author, Peter Duesberg, a pioneering virologist, may be well known to readers for his assertion that HIV is not the cause of AIDS. The biomedical community has roundly rebutted that claim many times. Duesberg's ideas about chromosomal abnormality as a root cause for cancer, in contrast, are controversial but are being actively investigated by mainstream science. We have therefore asked Duesberg to explain that work here. This article is in no sense an endorsement by SCIENTIFIC AMERICAN of his AIDS theories.

have succeeded. The oncogene model also ignores what a casual observer might perceive as a rather large elephant in the room: in every known instance of cancer, individual genes may well contain mutations, but entire chromosomes, which carry thousands of genes, are also severely scrambled—duplicated, broken, structurally rearranged or missing entirely. Growing evidence suggests that this chaos on the chromosomal level is not just a side effect of malignancy, as the prevailing model holds, but the direct cause and driving force of cancer.

With several colleagues in the U.S. and Europe, I have been investigating this possibility for more than a decade, and the recent work of many other researchers is also pointing to the conclusion that changes to the number and structure of entire chromosomes, rather than single genes, are sufficient to initiate and sustain malignancy. This view has important implications for cancer treatment and prevention, as well as for diagnosis of precancerous lesions when there may still be time to intervene. It also finally explains some characteristics of cancer cells and whole tumors that the gene mutation hypothesis leaves unresolved.

By looking at exceptions to the current rule, we hoped to find A BETTER RULE.

Overview/*Vicious Cell Cycles*

- Prevailing cancer theory blames mutations in important regulatory genes for upsetting normal controls on cells. But it does not give a primary role in carcinogenesis to the gross changes to whole chromosomes that are seen in all cancer cells.
- The author argues that these chromosomal mutations, which unbalance thousands of genes en masse, are sufficient to trigger cellular instability that leads to further chromosome disruption and to account for properties of malignant cells that cannot be explained by the activity of specific genes.

Invading Species

OUR RESEARCH GROUP arrived at a chromosomal theory of cancer in part by thinking about the basic biological features that make a normal human cell "normal," or even "human." Nature is extremely conservative with regard to chromosomes—the bound volumes of the encyclopedia of life—and the specific chromosome complements, or karyotypes, that define each species are final and stable for the duration of that species. Sexual reproduction also enforces conservation of a species-specific karyotype because embryonic development depends on absolute chromosomal conformity—cells containing altered or misallotted chromosomes are almost never viable. A rare exception, Down syndrome, illustrates the systemic damage that results from having just one extra copy of a relatively small chromosome, number 21, added to human cells.

Individual genes, in contrast, can be quite variable within a species. Single-nucleotide polymorphisms (SNPs), for example, which are mostly benign changes in the DNA sequence, are found scattered throughout every person's genome and passed from parent to child.

Cancer cells, and their less conspicuous precursors, are notorious violators of nature's karyotype stability laws, however. Whereas normal human cells are described as diploid, because they all have two copies of the chromosomes that define our species, the karyotypes of cells in solid tumors are *always* aneuploid—meaning that they have gained or lost whole chromosomes or segments of chromosomes [*see illustration on opposite page*]. As a result of this aneuploidy, the total DNA content of a cancer cell can rise to more than twice or fall to nearly half that of a normal diploid cell. This situation would cause the cell to produce wildly skewed dosages of the proteins encoded by the thousands of genes that were multiplied or lost. Such gross imbalances would inevitably disrupt the work of critical teams of enzymes, including those involved in repair or disposal of damaged DNA, and would destabilize

cellular structures and regulatory circuits. Indeed, the most complex coordinated squadron of proteins in a cell, and thus the most vulnerable, is the mitotic spindle apparatus that segregates chromosomes during cell division. Once aneuploidy is established, therefore, additional derangement of the chromosomes is likely.

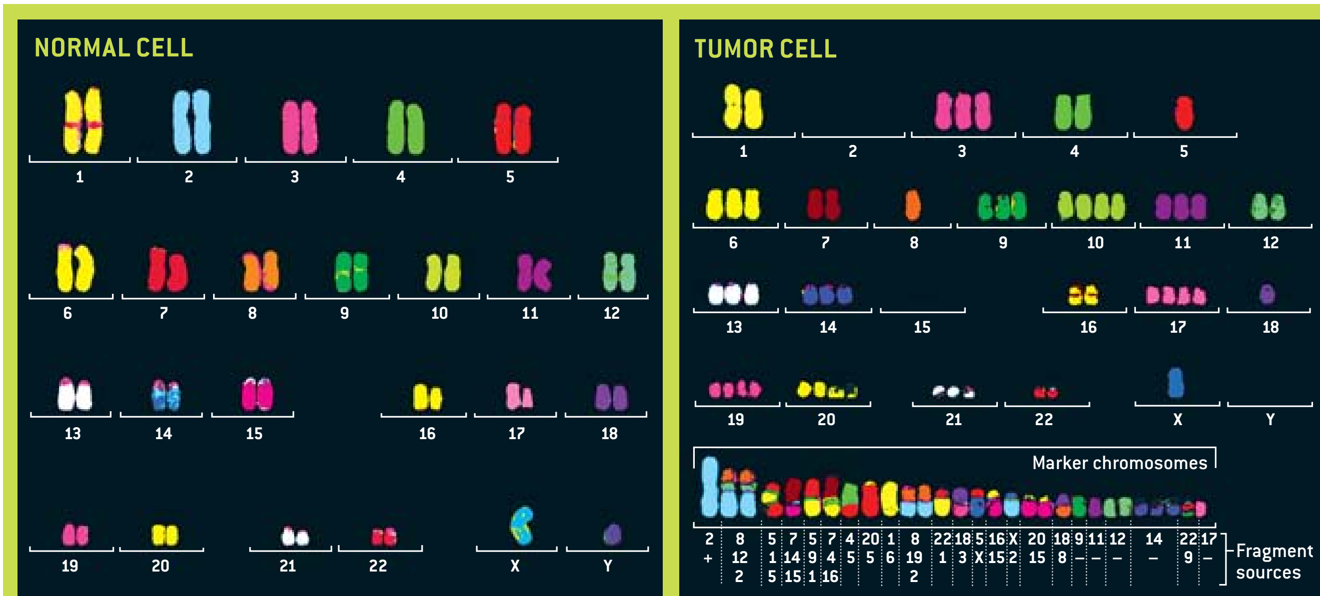
This could explain why even cancer cells within the same tumor can exhibit different combinations and alterations of their chromosomes, making each cell a kind of new species unto itself. Their inherent instability also sets individual cancer cells free to evolve new traits and behaviors (phenotypes), unlike the normal cells in an organism, which are destined to develop predetermined characteristics depending on the organ or tissue to which they belong and according to strict species-specific programming. With this license to change, an aneuploid cell can gradually drop more and more of its normal social responsibilities in a multicellular organism, multiplying ever more autonomously at the expense of other cells.

Collectively, cancerous cells tend to evolve from bad to worse. During this process, described as progression of carcinogenesis, the cells start developing their own exotic sizes, shapes, metabolisms and growth rates. Indeed, the very definition of malignancy includes the cells' acquisition of an unnatural ability to invade neighboring tissues and travel to distant organs, known as metastasis. The extreme variability of cancer cells and the enormous diversity of their phenotypes are primary reasons cancer has remained an intractable problem, scientifically and from the standpoint of treatment. Soon after a toxic drug is found to kill tumor cells, for instance, drug-resistant variant cells grow up in their place. Trying to tackle such entities with single drugs is like going up against an entire zoo of wild animals with a trap built only for crocodiles.

Yet scientists do know that in every case of cancer, the entire menagerie of malignant cells has arisen from a single unstable mother cell. This so-called clonal origin of a tumor is demonstrated by the presence in all or most of its cells of chromosomes whose rearrangements are so unique that they could not have arisen independently and so must all have been inherited from a common source. The challenge, therefore, is to find a theory that explains how one normal cell, out of the trillions that make up a human body, becomes chromosomally and phenotypically so unstable that it gives rise to a lethal cancer.

Until about 50 years ago, most cancer researchers saw the origins of cancer in aberrant chromosomes themselves. This idea had been advanced in the late 19th and early 20th centuries by two German scientists, David von Hansemann and Theodor Boveri. Von Hansemann, while studying cancer in Berlin, had discovered that all cancers contain abnormal chromosomes. Boveri, a biologist in Würzburg, had deduced that chromosomes are the vehicles for heritable information within cells by closely observing the steps of sea urchin embryo development. He reached his conclusion after witnessing the cellular degeneration that would follow if something went wrong during cell division and a single chromosome was broken or whole copies were unevenly distributed between two daughter cells. Boveri likened some of the resulting malformed embryos to tumors and, in 1914, predicted that gains or losses of specific chromosomes would initiate cancer.

That theory fell out of favor half a century later only because such patterns could not be detected in the chaotic and

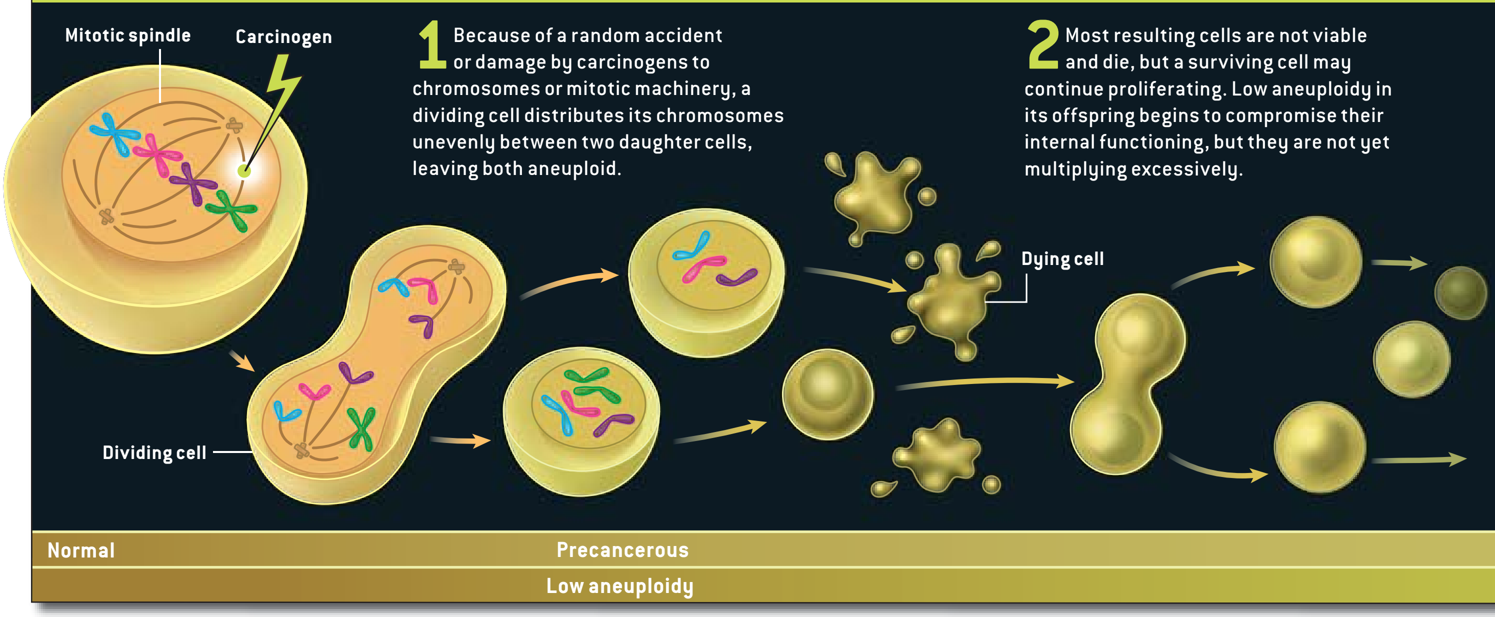


ORDER AND ANARCHY: A normal human cell's chromosome set, or karyotype (*left*), includes pairs of 23 standard chromosomes, whereas a tumor cell, like all cancers, exhibits the irregular karyotype (*right*) described as aneuploid: some whole chromosomes are missing, extra copies of others are present, and many have traded fragments. Among these are so-called marker chromosomes, whose distinctive structural rearrangements are seen in all the tumor's cells, indicating changes that originated in the cell that spawned the cancer. Numbers under each one specify the sources of its fragments; plus and minus signs identify those that are larger or smaller than usual.

FROM "THE CHROMOSOMAL BASIS OF CANCER," BY PETER DUESBERG ET AL., IN *CELLULAR ONCOLOGY*, VOL. 27; 2005, REPRINTED WITH PERMISSION OF IOS PRESS

HOW ANEUPLOIDY COULD CAUSE CANCER

Abnormal chromosome numbers in a cell create conditions that lead to further chromosome damage and disarray. With each new generation,



heterogeneous karyotypes of cancer cells, especially the most malignant ones, with the technology available at the time. In the absence of consistent chromosomal changes, the bewildering cancer karyotypes were interpreted as mere consequences of malignancy, which must itself be caused by another unseen factor.

All along, however, some evidence that was largely ignored continued to support an important role for aneuploidy in the genesis and progression of cancer. In fact, degree of aneuploidy is a feature that is still sometimes used by pathologists and doctors to assess the cancerous potential of abnormal cells biopsied from the cervix, prostate, liver, stomach, throat, breast and other tissues. Once an abnormal growth is defined as a cancer, or neoplasm, high cellular aneuploidy is also considered by some to be a sign of aggressiveness.

Moreover, new technologies have enabled researchers to reopen the question of whether specific chromosomal changes can be detected in cancer cells, and many investigators are beginning to find those long-sought patterns. Even oncogene enthusiasts studying the effects of aneuploidy in existing tumors have generated evidence that points to chromosome-level alterations, rather than discrete gene mutations, as the driver of malignancy [see "Untangling the Roots of Cancer," by W. Wayt Gibbs; Scientific American, July 2003]. Intrigued by the emergence of these data over the past decade, as well as by the enormous potential of chromosomal alteration to generate new cellular phenotypes—much as it produces new species in nature—my colleagues and I set ourselves the challenge of explaining the instability inherent in cancer cells, which is the source of the chaos that became the nemesis of the original chromosomal cancer theory.

Effects and Causes

OUR STRATEGY was to collect and analyze the features of carcinogenesis that are most inconsistent with the prevailing gene mutation theory—by looking at the exceptions to the current rule, we hoped to find a better rule. In the end, we identified six features of cancer that are inexplicable by gene mutation alone but that can be explained by chromosomal changes, and a coherent theory did emerge.

Cancer risk grows with age. Lamentably common, cancer afflicts about one in three people at some point in their lives, but mostly after the age of 50, which is when chances for malignancy soar. Thus, cancer is, by and large, a disease of old age. The gene mutation theory of cancer's origins, however, predicts that the disease should be quite common in newborns. If, as that hypothesis holds, about half a dozen mutations to critical genes were necessary to ignite malignancy, certainly some of those mutations would accumulate like SNPs over the course of generations in the genomes of many individuals. A baby could thus inherit three of six hypothetical colon cancer mutations from her mother, for example, and two from her father and be at extremely high risk of cancer from picking up the missing sixth mutation in any one of her billions of colon cells. Some babies might even be born with colon cancer from inheriting all six hypothetical colon cancer mutations from their parents. But colon cancer is never seen in children. Indeed, even laboratory mice intentionally engineered to carry an assortment of ostensibly carcinogenic mutations from birth can live and propagate happily, with no higher risk of developing tumors than normal lab mice.

Some proponents of the mutation theory maintain that except in rare cases of genetic predisposition to cancer—by which

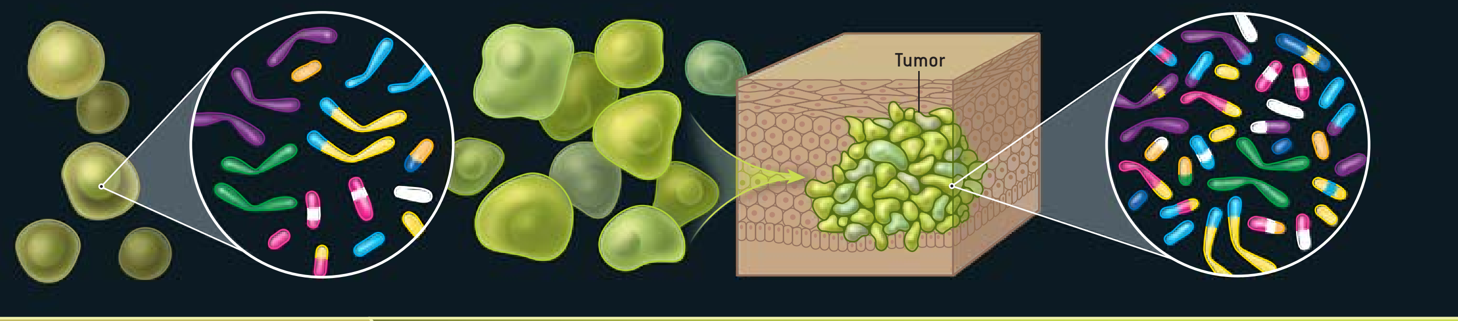
JEN CHRISTIANSEN

resulting cells grow increasingly unstable and develop ever more malignant traits.

3 Skewed dosages of proteins generated by the cells' irregular chromosome complements cause instability that further disrupts regulatory and DNA-maintenance processes. Additional chromosome breakage, structural rearrangements and duplication errors arise.

4 Cells begin exhibiting progressively more deviant traits as aneuploidy increases and their protein production grows more aberrant. These changes include atypical appearance and hyperproliferation, leading to formation of a tumor.

5 Malignant features, such as the ability to invade neighboring tissue or metastasize to distant locations and intrinsic resistance to drugs, may also arise as random effects of the internal chaos caused by the cells' escalating aneuploidy.



they mean the presence of inherited oncogenic mutations—the gene changes believed to cause malignancies must all be acquired after birth. That assumption implies a natural gene mutation rate over an individual's lifetime much higher than the norm, which is one change to a given gene in one out of every one million to 10 million cells (that is, once every 10^6 to 10^7 cell generations).

Interestingly, among the rare exceptions to cancer's age bias are children with congenital aneuploidy, as in Down syndrome, or with inherited chromosome instability syndromes, such as the disease known as mosaic variegated aneuploidy (MVA), which also causes severe mental retardation. Defects of mitotic spindle assembly in the cells of children with MVA produce random aneuploidies throughout their bodies, and nearly one third develop leukemia or unusual solid cancers.

Being born aneuploid, or prone to aneuploidy, clearly accelerates processes that lead to cancer. Indeed, the inherent instability of aneuploid cells would explain why most aneuploid embryos, as Boveri observed 100 years ago, would not be viable at all and thus why newborns are cancer-free and cancer is not heritable.

Carcinogens take a very long time to cause cancer. Numerous chemicals and forms of radiation have been shown to be carcinogenic in animals or established as the source of occupational or accidental cancers in humans. But even the strongest carcinogens at the highest survivable doses never cause cancer right away. Instead the disease emerges only after delays lasting years or even decades. In contrast, when substances known to cause gene mutations are administered to bacteria, the cells begin displaying new phenotypes within hours; in larger organisms such as flies, the effect is seen within days. A gene mutation scenario therefore does not explain why cells exposed to carcinogenic agents become cancer cells, much the way an undersea volcano becomes an island that appears only after many invisible eruptions.

Carcinogens, whether or not they cause gene mutations, induce aneuploidy. Scientists have looked for the immediate genetic effects of carcinogens on cells, expecting to see mutations in many crucial genes, but instead have found that some of the most potent carcinogens known induce no mutations at all. Examples include asbestos, tar, aromatic hydrocarbons, nickel, arsenic, lead, plastic and metallic prosthetic implants, certain dyes, urethane and dioxin. Moreover, the dose of carcinogen needed to initiate the process that forms malignant tumors years later was found to be less than one-thousandth the dose required to mutate any specific gene. In all cases, however, the chromosomes of cells treated with cancer-causing doses of carcinogens were unstable—that is, displaying higher than usual rates of breakage and disruption.

These findings suggest that carcinogens function as "aneuploidogens" rather than as mutagens. The gene mutation theory has never been able to explain how nonmutagenic carcinogens cause cancer. In fact, even the mutagenic carcinogens may actually work to trigger cancer by inducing aneuploidy through direct destruction or fragmentation of chromosomes. (Radiation, for example, induces mutations indirectly by first breaking the DNA strand. Cellular repair proteins attempt to fix the damage but may introduce errors or rearrangements in the nucleotide sequence.) Nonmutagenic carcinogens, such as aromatic hydrocarbons, can cause aneuploidy through a different mechanism. Those chemicals are known to destroy the microtubule polymers within a cell that normally pull dupli-

cate chromosomes apart symmetrically during mitosis [*see box on two preceding pages*]. Thus, the common denominator of all carcinogens seems to be their initiation of random aneuploidy.

Patterns of aneuploidy are seen in different tumors. If aneuploidy is only a side effect of cancer, then chromosomal changes in the cancers of different people should be random. But a pair of chromosome-painting technologies, known as comparative genomic hybridization and fluorescent in situ hybridization, have begun to reveal signature patterns amid the chaos in cancer cells. These techniques enable scientists to tag and track bits and pieces of chromosomes with colored DNA-specific probes to build a picture of all the chromosomal segments gained, lost or rearranged in a given cell.

Many researchers have begun to find evidence of "nonrandom" aneuploidies—specific chromosomal changes shared by most cancers of a certain kind, such as neoplasms of the breast or cervix, as predicted by the original chromosomal cancer theory. In just one example reported last year, scientists at Karolinska University Hospital in Sweden examined cells from 10 patients with Burkitt's lymphoma and found frequent translocations (swapping of chromosome fragments) involving chromosomes 3, 13 and 17, as well as specific losses or gains in segments of chromosomes 7 and 20.

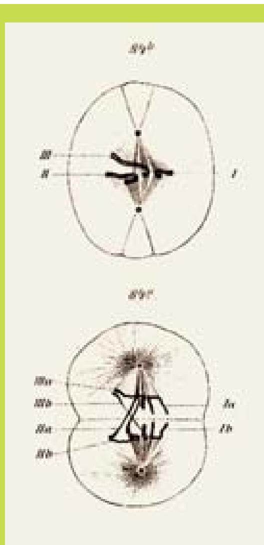
Because such patterns in cancer cells of different individuals are specific to the type of tissue in which the cancer originates, they may well represent essential chromosome changes needed to overcome programmed constraints on the original mother cell's development. These changes would thus represent the minimum alterations needed for a viable aneuploid cell of that kind to start down the road toward becoming neoplastic.

In addition to such cancer-type-specific patterns, chromosome changes indicative of cancer stage, metastatic potential, and even drug resistance have also been identified by several research teams. For instance, the Karolinska group noted that translocations in one segment of chromosome 17 and gains on parts of 7 and 20 were associated with drug resistance.

As scientists have continued to work out the exact functional effects of such cancer-specific aneuploidies, many analyses of the amounts of different gene transcripts and proteins manufactured by cancer cells have also shown that the proteins encoded by specific purported oncogenes are actually often generated at the same levels as in normal cells. Among several interesting recent studies of this kind, one by a team of U.S. and Israeli researchers set out to assess the protein levels in colon cancer cells and did find that large numbers of proteins were being overproduced or underproduced—in amounts corresponding to the total DNA content changes within the cell. With greater aneuploidy, the protein imbalances increased, and so did the cells' cancerous progression. This example strongly supports the idea that by raising or reducing a cell's normal dosage of thousands of genes at once, aneuploidy produces malignant phenotypes.

Gratuitous traits do not contribute to the cancer's survival. Some of the most common and dreadful characteristics of cancer do not offer any competitive survival advantages to a tumor. Examples of these include intrinsic resistance to drugs the tumor has never encountered before and metastasis, which does not help tumor cells compete with normal cells at their site of origin. Individual gene mutations, which are rare to begin with, would only be selectively conserved in tumor cells if the mutation were advantageous, so the chances of an untreated tumor becoming drug-resistant through random gene mutations in its cells are practically zero. Because chromosomes are much larger and can harbor thousands of genes, however, they could be selectively retained for their contribution of some cancer-specific phenotype, and numerous unselected traits would be carried along as well. Indeed, the evidence of specific chromosomal changes associated with drug resistance or metastasis supports this possibility. And collectively, cancer cells can evolve all kinds of new traits very rapidly.

Aneuploid cells reshuffle their chromosomes MUCH FASTER than mutation can alter their genes.



UNEQUAL CHROMOSOME distribution during division of a sea urchin embryo cell is depicted in 1888 drawings by Boveri, who attributed the problem observed in this sequence to the "defective" action of "spindle fibers."

COURTESY OF KURT STUEBER (*Boveri drawings*)

ARGUMENT FOR ANEUPLOIDY?

In the early 20th century German biologist Theodor Boveri advanced a theory that cancer stems from chromosomal disruption based on his studies of sea urchin development (*left*). Boveri predicted that most aneuploid cells would not be viable, but certain chromosome losses, gains or rearrangements could give a cell and its descendants the ability to transform and multiply in ways that would lead to cancer. Those signature alterations would therefore be seen in most cancers of the same tissue type. Technology that permits researchers to track specific pieces of DNA is beginning to reveal chromosome changes common to tumors and even precancerous cells of the same kind in different people (*right*). Because such patterns are unlikely to arise randomly if aneuploidy is only an effect, rather than a cause, of cancerous cell transformation, they could support the new chromosomal theory of cancer.

Cancer cells morph much faster than genes. Cancer cells generate new phenotypes and lose old ones exceedingly fast. Given a normal mutation rate—and many studies have demonstrated that in more than 90 percent of cancers, the cellular gene mutation rate is not accelerated—the odds are very long once again that a particular gene will mutate to generate a new phenotype as rapidly as is observed in cancer cells.

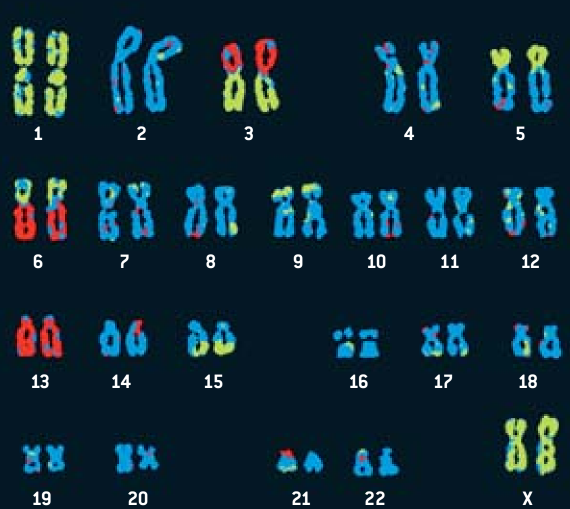
To test the hypothesis that whole-karyotype alterations could achieve such speedy change, our group recently examined the chromosomal variations present in highly aneuploid human breast and colon cancer cells, as well as the speed with which the cells acquired resistance to toxic drugs. Karyotype changes were observed in these cells at rates of at least one in 100 generations, and drug-resistance-specific changes at a rate of one in 1,000 to one in 100,000 cell generations. In other words, aneuploid cells reshuffle their chromosomes and phenotypes much faster than mutation can alter their genes.

These and similar experiments also demonstrated that the more aneuploid the cancer cells were, the quicker new chromosomal alterations appeared. This pattern strongly supports a conclusion that the chromosomal instability seen in cancer cells is catalyzed by aneuploidy itself. Once this vicious cycle is under way, the fact that every cell would be randomly generating its own new phenotypes could explain an observation made decades ago by Leslie Foulds of the Royal Cancer Hospital in London that "no two tumors are exactly alike ... even when they originate from the same tissue ... and have been induced experimentally in the same way." Such individuality is yet another hallmark of cancer that cannot be explained by the activity or inactivity of specific genes, which would be expected to have consistent effects each time and in each cell.

Because every one of the extraordinary features of carcinogenesis that cannot be explained by the mutation theory is associated with chromosomal alterations, we have proposed a revised chromosomal theory of cancer that takes this inherent instability into account.

FROM "ADVANCED-STAGE CERVICAL CARCINOMAS ARE DEFINED BY A RECURRENT PATTERN OF CHROMOSOMAL ABERRATIONS REVEALING HIGH GENETIC INSTABILITY AND A CONSISTENT GAIN OF CHROMOSOME ARM 3Q," BY K. HESELMEYER ET AL., IN *GENES, CHROMOSOMES AND CANCER*, VOL. 19, NO. 4; 1997, REPRINTED WITH PERMISSION OF WILEY-LISS, INC., A SUBSIDIARY OF JOHN WILEY & SONS, INC.

CHROMOSOME CHANGES in cervical cancer cells from 30 patients showed some consistent patterns when areas where genetic material was lost (*red*) or added (*green*) were highlighted, as in the image from one case shown here. Specific gains in chromosome 3 were seen in 23 cases and appeared in tumors progressing to invasive carcinoma. Additions to chromosome 1 and losses in chromosomes 2, 3 and 4 were found in a third to half of all the cases.



Deadly Dynamic

WHEN CANCER IS VIEWED as a chromosomal disease, carcinogens, rare genetic syndromes and accidental mitotic errors can initiate carcinogenesis by inducing random aneuploidies. Because aneuploidy unbalances thousands of genes and their protein output, it sets the stage for still more aneuploidy. This dynamic self-catalyzing condition becomes a steady source of variation from which, in classical Darwinian terms, selection of selfish chromosome combinations eventually gives rise to viable and competitive neoplastic cells. These are effectively new cell species, albeit parasitic ones, with unstable karyotypes.

Once cancer progression is under way, random chromosome reshuffling can rapidly generate gratuitous traits that include lethal properties such as drug resistance and metastasis. Thus, the prospects for success in treating tumors with individual drugs, particularly those targeted to single genes, are not good. Some investigators have recently proposed fighting fire with fire in the form of treatments that accelerate chromosomal DNA damage and aneuploidy, with the goal of making tumor cells so unstable that they are no longer viable. This might work in a very small, well-contained tumor, but it is a technique that could prove difficult to control.

The slow progression from early aneuploidy to aggressive neoplasm does, however, offer a wide window of opportunity for doctors to detect and surgically remove a potential cancer before it enters a neoplastic stage. Early cancers could also be distinguished from morphologically identical benign tumors by checking for aneuploidy. In more advanced tumors, chromosome patterns associated with drug resistance or metastatic potential could be used to guide treatment choices.

Finally, screening for chromosome-damaging substances in foods, drugs and the environment could significantly improve cancer prevention by identifying aneuploidy-inducing potential carcinogens. Today I am still optimistic enough to hope that by coming full circle, cancer scientists will eventually arrive at a basic understanding of this fearsome disease, yielding effective prevention, management and even cures. SA

MORE TO EXPLORE

Debate Surges over the Origins of Genomic Defects in Cancer. Jean Marx in *Science*, Vol. 297, pages 544–546; July 26, 2002.

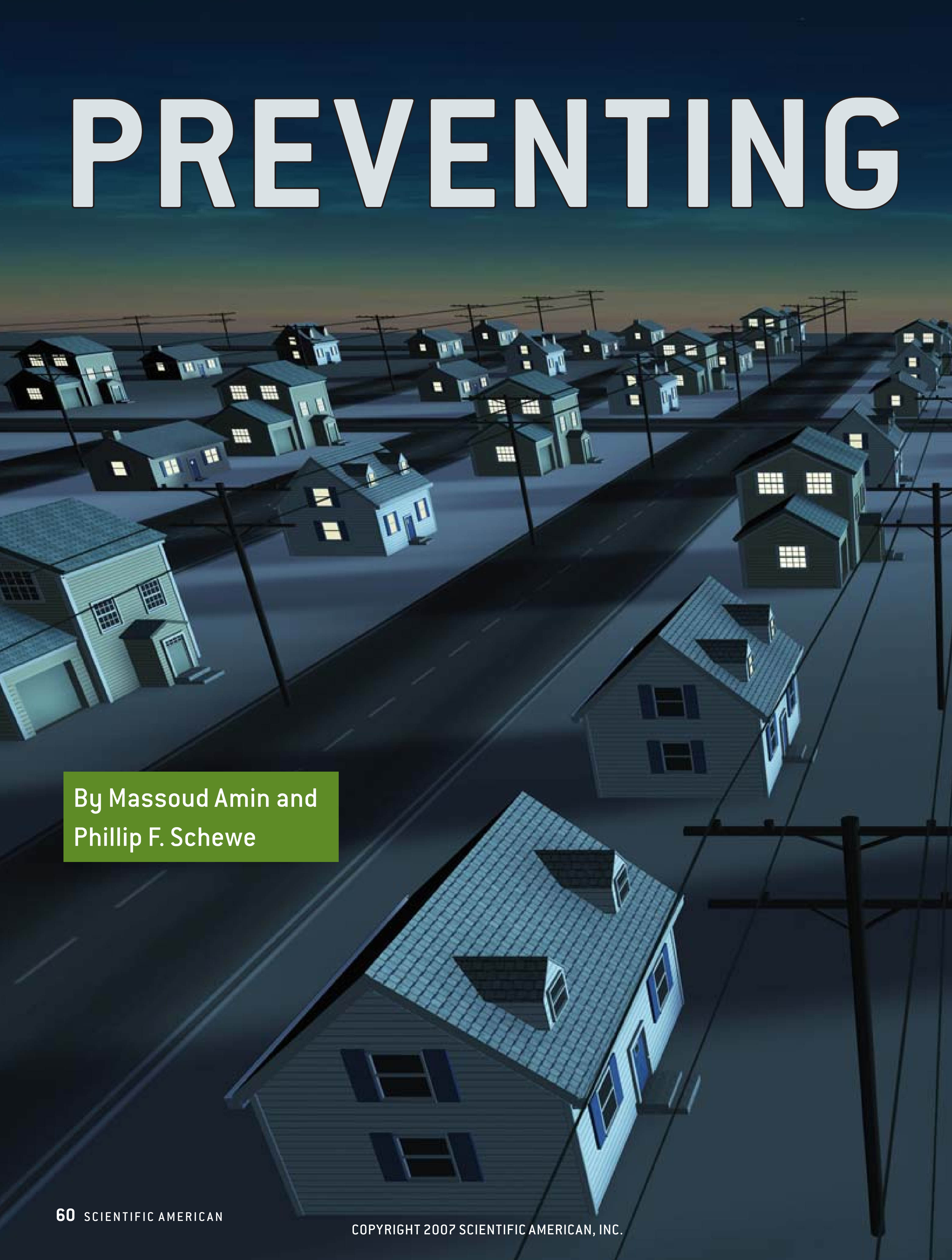
The Sigmoidal Curve of Cancer. Roberto P. Stock and Harvey Bialy in *Nature Biotechnology*, Vol. 21, pages 13–14; January 2003.

The Chromosomal Basis of Cancer. Peter Duesberg et al. in *Cellular Oncology*, Vol. 27, Nos. 5–6, pages 293–318; 2005.

Cancer Drug Resistance: The Central Role of the Karyotype. Peter Duesberg et al. in *Drug Resistance Updates* (in press).

PREVENTING

By Massoud Amin and Phillip F. Schewe



BLACKOUTS

A smarter power grid that automatically responds to problems could reduce the rising number of debilitating blackouts

RANDY LYHUS

August 14, 2003, was a typical warm day in the Midwest. But shortly after 2:00 P.M. several power lines in northern Ohio, sagging under the high current they were carrying, brushed against some overgrown trees and shut down. Such a disturbance usually sets off alarms in a local utility's control room, where human operators work with controllers in neighboring regions to reroute power flows around the injury site.

On this day, however, the alarm software failed, leaving local operators unaware of the problem. Other controllers who were relaying, or "wheeling," large amounts of power hundreds of miles across Ohio, Michigan, the northeastern U.S. and Ontario, Canada, were oblivious, too. Transmission lines surrounding the failure spot, already fully taxed, were forced to shoulder more than their safe quota of electricity.

To make matters worse, utilities were not generating enough "reactive power"—an attribute of the magnetic and electric fields that move current along a wire. Without sufficient reactive power to support the suddenly shifting flows, overburdened lines in Ohio cut out by 4:05 P.M. In response, a power plant shut down, destabilizing the system's equilibrium. More lines and more plants dropped out. The cascade continued, faster than operators could track with the decades-old monitoring equipment that dots most of the North American power grid, and certainly much faster than they could control. Within eight minutes 50 million people across eight states and two Canadian provinces had been blacked out. The event was the largest power loss in North American history.

The 2003 disaster was a harbinger, too. Within two months, major blackouts occurred in the U.K., Denmark, Sweden and Italy. In September 2003 some 57 million Italians were left in the dark because of complications in transmitting power from France into Switzerland and then into Italy. In the U.S., the annual number of outages affecting 50,000 or more customers has risen for more than a decade.

In addition to inconvenience, blackouts are causing major economic losses. The troubles will get worse until the entire transmission system that moves power from generating plants to neighborhood substations is overhauled. More high-voltage lines must be built to catch up with the rising demand imposed by ever more air conditioners, computers and rechargeable gadgets.

But perhaps even more important, the power grid must be made smarter. Most of the equipment that minds the flow of electricity dates back to the 1970s. This control system is not good enough to track disturbances in real time—as they happen—or to respond automatically to isolate problems before they snowball. Every node in the power grid should be awake,

responsive and in communication with every other node. Furthermore, the information that operators receive at central control stations is sparse and at least 30 seconds old, making it impossible for them to react fast enough to stop the large cascades that do start. A self-healing smart grid—one that is aware of nascent trouble and can reconfigure itself to resolve the problem—could reduce blackouts dramatically, as well as contain the chaos that could be triggered by terrorist sabotage. It would also allow more efficient wheeling of power, saving utilities and their customers millions of dollars during routine operation. The technology to build this smart grid largely exists, and recent demonstration projects are proving its worth.

Overwhelmed by Progress

THE TRANSMISSION SYSTEM has become vulnerable to blackouts because of a century-long effort to reduce power losses. As power moves through a wire, some of it is wasted in the form of heat. The loss is proportional to the amount of current being carried, so utilities keep the current low and compensate by raising the voltage. They have also built progressively longer, higher-voltage lines to more efficiently deliver power from generation plants to customers located far away. These high-voltage lines also allow neighboring utilities to link their grids, thereby helping one another sustain a critical balance between generation supply and customer demand.

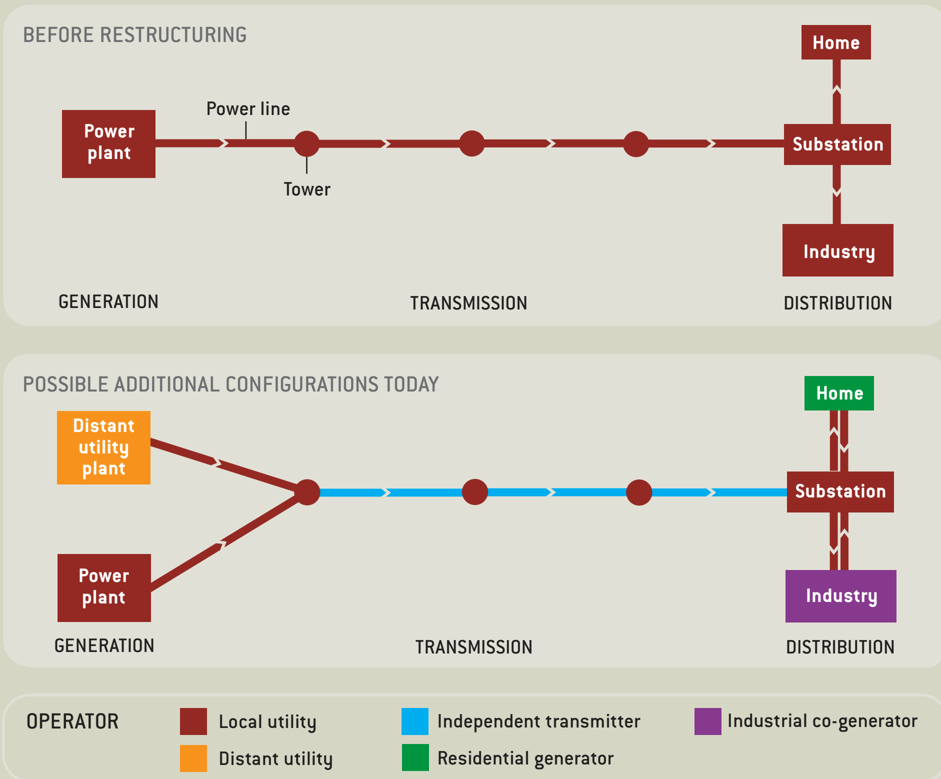
Such interconnectedness entails certain dangers, however, including the possibility that a shutdown in one sector could rapidly propagate to others. A huge 1965 blackout in the Northeast prompted utilities to create the North American Electric Reliability Council—now called the North American Electric Reliability Corporation (NERC)—to coordinate efforts to improve system reliability. Similar bodies, such as Europe's Union for the Coordination of Transmission of Electricity, exist around the world.

Why, then, had the U.S. grid become vulnerable enough to fail massively in 2003? One big reason is that investment in upgrading the transmission system has been lacking. Sharply rising fuel prices in the 1970s and a growing disenchantment with nuclear power prompted Congress to pass legislation intended to allow market competition to drive efficiency improvements. Subsequent laws

THE PROBLEM: TOO MANY PLAYERS, TOO LITTLE INVESTMENT

Restructuring of the utility industry (*below*) because of deregulation is a major reason why U.S. blackouts are increasing (*opposite page, bottom graph*). With a single company no longer in charge of the action in a given region, the power grid is not being upgraded or expanded to keep pace with growing demand (*opposite page, top panels*).

DEREGULATION HAS FRAGMENTED CONTROL OF THE POWER SYSTEM

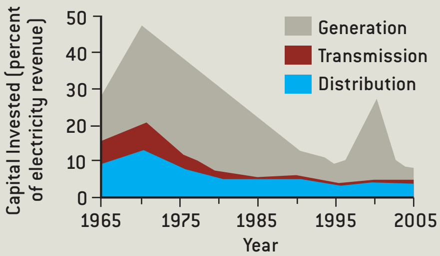


JEN CHRISTIANSEN; SOURCES: UNIVERSITY OF MINNESOTA, EDISON ELECTRIC INSTITUTE, NORTH AMERICAN ELECTRIC RELIABILITY CORP., U.S. DEPARTMENT OF ENERGY

Overview/*Smart Grid*

- Demand for electricity has increased steadily for decades, yet transmission lines that transport power from generation plants to customers have not been added or upgraded at the same pace. As a result, the grid has become overloaded, making it more prone to blackouts, which have risen in number and severity and cost the U.S. more than $70 billion in annual economic losses.
- Even with more lines, a self-healing smart grid that can sense local problems early, and automatically fix or isolate them before they snowball, is needed to prevent the cascading power failures that cause blackouts.
- Digital controllers and real-time communications devices must be placed on every transmission line, substation, power plant and utility operations center.
- The operations centers also need updated computers and software that enable human controllers to manually take over the automated smart grid if a blackout does somehow begin. And controllers require better training to know how to react quickly.

MONEY SPENT ON U.S. TRANSMISSION LINES IS DECLINING

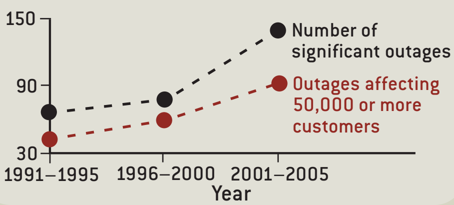


TRANSMISSION CAPACITY IS FALLING BEHIND GROWTH IN DEMAND

20% INCREASE IN **DEMAND** FOR ELECTRICITY (1999–2009)

7% INCREASE IN TRANSMISSION **CAPACITY** (1999–2009)

RESULT: LARGE BLACKOUTS ARE GROWING IN NUMBER AND SEVERITY



have instigated a sweeping change in the industry that has come to be called restructuring. Before restructuring began in earnest in the 1990s, most utilities conducted all three principal functions in their region: generating power with large plants, transmitting it over high-voltage lines to substations, then distributing it from there to customers over lower-voltage lines. Today many independent producers sell power near and far over transmission lines they do not own. At the same time, utilities have been selling off parts of their companies, encouraged by the Federal Energy Regulatory Commission to further promote competition. Gradually the transmission business has become a confusing mixture of regulated and unregulated services, with various companies controlling fragmented pieces.

Investors have found generation, now largely deregulated, to be attractive. But because the transmission system has been only partially deregulated, uncertainty over its fate makes investors wary. (Deregulation of distribution is still in its infancy.) Meanwhile, even though wheeling occurred in the past, since the 1990s much larger amounts of power have been moved over great distances. As a result, massive transfers are flowing over transmission lines built mostly by utilities for local use decades ago.

Proposed federal legislation might encourage more investment, but even if transmission capacity is added, blackouts will still occur. The entire power grid has to be refurbished, because the existing control technology—the key to quickly sensing a small line failure or the possibility of a large instability—is antiquated. To remain reliable, the grid will have to operate more like a fighter plane, flown in large part by autonomous systems that human controllers can take over if needed to avert disaster.

A Need for Speed

MODERN WARPLANES are so packed with sophisticated gear that pilots rely on a network of sensors and automatic controls that quickly gather information and act accordingly. Fortunately, the software and hardware innovations required to fly the power grid in a similar fashion and to instantly reroute power flows and shut down generation plants are at hand.

Reconfiguring a widely interconnected system is a daunting challenge, though. Most power plants and transmission lines are overseen by a supervisory control and data acquisition (SCADA) system. This system of simple sensors and controllers provides three critical functions—data acquisition, control of power plants, and alarm display—and allows operators who sit at central control stations to perform certain tasks, such as opening or closing a circuit breaker. SCADA monitors the switches, transformers and pieces of small hardware, known as programmable logic controllers and remote terminal units, that are installed at power plants, substations, and the intersections of transmission and distribution lines. The system sends information or alarms back to operators over telecommunications channels.

SCADA technology goes back 40 years, however. Much of it is too slow for today's challenges and does not sense or control nearly enough of the components around the grid. And although it enables some coordination of transmission among utilities, that process is extremely sluggish, much of it still based on telephone calls between human operators at the utility control centers, especially during emergencies. What is more, most programmable logic controllers and remote terminal units were developed before industry-wide standards for interoperability were established; hence, neighboring utilities often use incompatible control protocols. Utilities are operating ever closer to the edge of the stability envelope using 1960s-era controls.

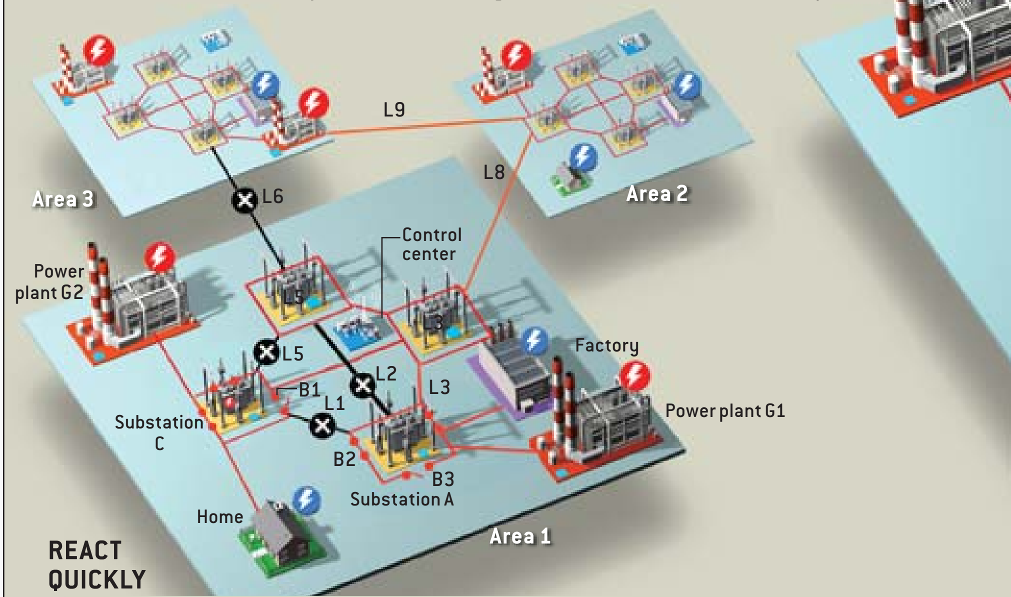
The Self-Healing Smart Grid

THE RESULT is that no single operator or utility can stabilize or isolate a transmission failure. Managing a modern grid in real time requires much more automatic monitoring and far greater interaction among human operators, computer systems, communications networks and data-gathering sensors that need to be deployed everywhere in power plants and substations. Reliable operation also requires multiple, high-data-rate, two-way communications links among all these nodes, which do not exist today, plus powerful computing facilities at the control center. And intelligent processors—able to automatically reconfigure power flows when precursors to blackouts are sensed—must be distributed across the network.

Flying the grid begins with a different kind of system design. Recent research from a variety of fields, including nonlinear dynamical systems, artificial intelligence, game theory and software engineering, has led to a general theory of how to design complex systems that adapt to changing conditions. Mathematical and computational techniques developed for this young discipline are providing new tools for grid engineers. Industry working groups, including a

THE SOLUTION: A SMART GRID THAT HEALS ITSELF

Imagine that a thunderstorm knocks out power lines L5 and L6. This occurrence would typically cause a chain reaction of line faults that would black out Area 1. But a smart grid would isolate and correct the problem as depicted below. The action begins as a look-ahead computer at the control center simulates corrective actions in less than half a second and sends instructions to control computers around the grid.



REACT QUICKLY

0.04 second later
The loss of L5 and L6 causes a fault in line L1. Control computers tell circuit breakers B1 and B2 to open to isolate the fault, but B2 becomes stuck in the closed position.

0.1 second
Power generator G1 automatically accelerates to meet demand from the loss of G2 caused by problems on lines L5 and L1. G1 also accelerates to attempt to keep line voltage throughout Area 1 at the required 60 hertz (cycles per second).

0.4 second
The control computer-simulator in substation A tells breaker B3 to open to protect the substation against damage from excessive current flow through it. B3 opens, shutting down line L2. G1 accelerates further to compensate.

0.5 second
The control center shuts down generator G1 to prevent damage to it from excessive acceleration.

group run by one of us (Amin) while at the Electric Power Research Institute (EPRI) in Palo Alto, Calif., have proposed complex adaptive systems for large regional power grids. Several utilities have now deployed, at a demonstration scale, smart remote terminal units and programmable controllers that can autonomously execute simple processes without first checking with a human controller, or that can be reprogrammed at a distance by operators. Much wider implementation is needed.

A self-healing smart grid can best be built if its architects try to fulfill three primary objectives. The most fundamental is real-time monitoring and reaction. An array of sensors would monitor electrical parameters such as voltage and current, as well as the condition of critical components. These measurements would enable the system to constantly tune itself to an optimal state.

The second goal is anticipation. The system must constantly look for potential problems that could trigger larger disturbances, such as a transformer that is overheating. Computers would assess trouble signs and possible consequences. They would then identify corrective actions, simulate the effectiveness of each action, and present the most useful responses to human operators, who could then quickly implement corrective action by dispatching the grid's many automated control features. The industry calls this capability fast look-ahead simulation.

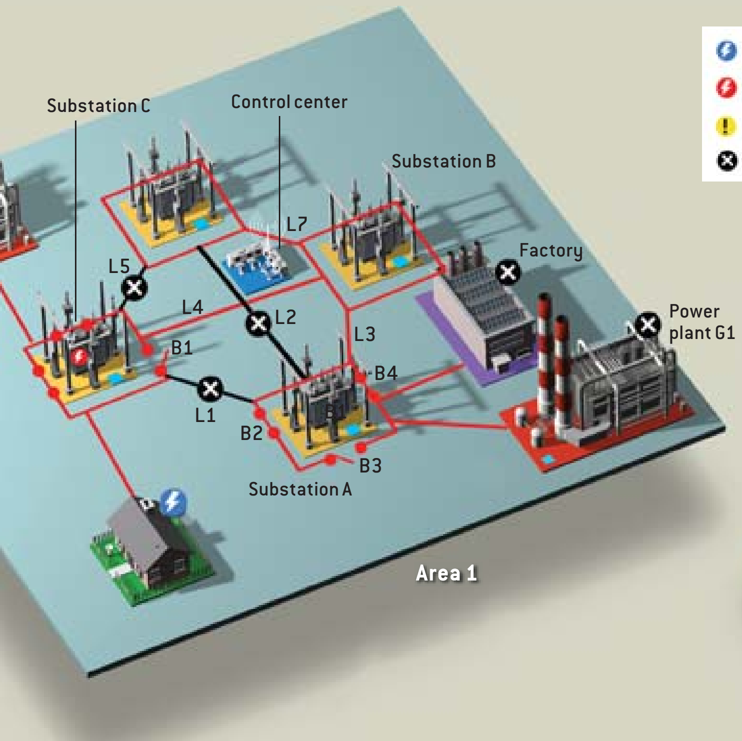
The third objective is isolation. If failures were to occur, the whole network would break into isolated "islands," each of which must fend for itself. Each island would reorganize its power plants and transmission flows as best it could. Although this might cause voltage fluctuations or even small outages, it would prevent the cascades that cause major blackouts. As line crews repaired the failures, human controllers would prepare each island to smoothly rejoin the larger grid. The controllers and their computers would function as a distributed network, communicating via microwaves, optical fibers or the power lines themselves. As soon as power flows were restored, the system would again start to self-optimize.

To transform our current infrastructure into this kind of self-healing smart grid, several technologies must be deployed and integrated. The first step is to build a processor into each switch, circuit breaker, transformer and bus bar—the huge conductors carrying electricity away from generators. Each transmission line should then be fitted with a processor that can communicate with the other processors, all of which

THE AUTHORS

MASSOUD AMIN and *PHILLIP F. SCHEWE* have extolled the virtues of a smart grid for years. Amin is professor of electrical and computer engineering at the University of Minnesota and is director of the school's Center for the Development of Technological Leadership. When he was with the Electric Power Research Institute in Palo Alto, Calif., he led the development of more than 20 advanced technologies and devised the foundations of a "self-healing grid," a term he coined. Schewe is chief science writer at the American Institute of Physics and edits the institute's weekly newsletter, *Physics News Update*.

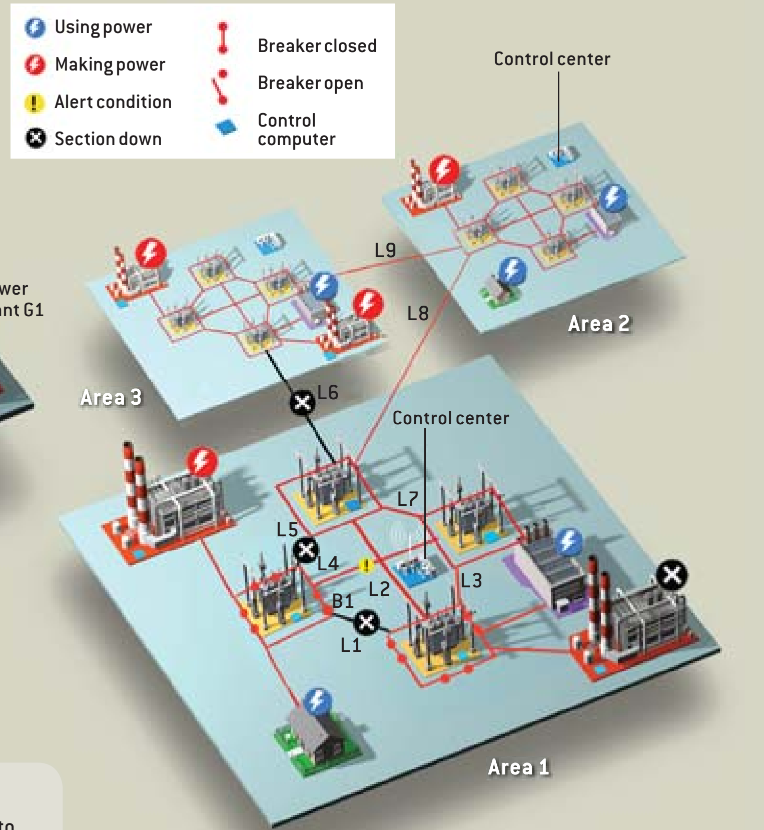
DON FOLEY; SOURCE: ELECTRIC POWER RESEARCH INSTITUTE



MAKE DECISIONS

0.6 second
The control computer in substation B would typically shut down line L3 to reduce demand if generator G1 were accidentally lost, but because it was stopped deliberately, computers across Area 1 communicate and decide instead to shut down a big factory, lowering demand considerably. This action reduces the mismatch between generation and demand so critical functions such as streetlights and hospitals can stay powered.

10 seconds
After several seconds, however, the substation B computer detects that the voltage there is beginning to oscillate beyond safe tolerances because the mismatch is still significant, threatening to damage equipment on lines L3, L4 and L7. Rather than shutting down those lines (the old-fashioned response), the area computers change control of generator G2 to manual, advising human operators at the Area 1 control center to raise generation or reduce load. They do some of both.



RETURN TO NORMAL

60 seconds
Lines L3, L4 and L7 have been spared, but L4 is becoming overloaded. Human operators at the control center communicate via satellite to operators in the Area 2 control center, asking for help. Operators in Area 2 send power over line L8; they also instruct the control computers in their sector to modify power flows slightly to compensate for the sudden export. Once road crews fix damaged lines L5 and L6, the computers will bring L1 and power plant G1 back into service. Power in the three areas returns to normal flow.

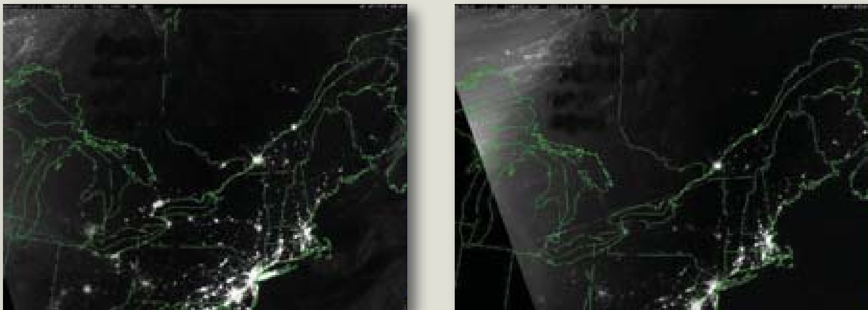
would track the activity of their particular piece of the puzzle by monitoring sensors built into their systems.

Once each piece of equipment is being monitored, the millions of electromechanical switches currently in use should be replaced with solid-state, power-electronic circuits, which themselves must be beefed up to handle the highest transmission voltages: 345 kilovolts and beyond. This upgrade from analog to digital devices will allow the entire network to be digitally controlled, the only way real-time self-monitoring and self-healing can be carried out.

A complete transition also requires digitization of the small, low-voltage distribution lines that feed each home and business. A key element is to replace the decades-old power meter, which relies on turning gears, with a digital meter that can not only track the current going into a building but also track current sent back out. This will allow utilities to much better assess how much power and reactive power is flowing from independent producers back into the grid. It will also allow a utility to sense very local disturbances, which can provide an earlier warning of problems that may be mounting, thereby improving look-ahead simulation. And it will allow utilities to offer customers hour-by-hour rates, including incentives to run appliances and machines during off-peak times that might vary day to day, reducing demand spikes that can destabilize a grid. Unlike a meter, this digital energy portal would allow network intelligence to flow back and forth, with consumers responding to variations in pricing. The portal is a tool for moving

THE HUMAN FACTOR

If a local blackout begins to escalate beyond a smart grid's ability to automatically keep it in check, human operators in regional control rooms could attempt to cut off the chain reaction. To do so, they would need complete and up-to-the-second information about the network, consistent computer protocols, predetermined response procedures and solid training. Each of these prerequisites was lacking when the massive 2003 U.S. blackout began to snowball, as dialogue during the first minutes of the event shows (*portions appear below*). The recorded conversations, published by the North American Electric Reliability Council, took place between reliability controllers in neighboring regions who were trying to help one another balance power flows that were heading out of control.



SATELLITE IMAGES show the Northeast the night before (*left*) and after the 2003 blackout (*right*).

NATIONAL OCEANIC AND ATMOSPHERIC ADMINISTRATION

POOR OPERATOR TRAINING, LACK OF REAL-TIME DATA

AEP operator: "What do you have on the Sammis-Star [line]?"
PJM operator: "I'm sorry? Sammis-Star, okay, I'm showing 960 on it and it's highlighted in blue.... Tell me what that means on your machine."
AEP: "Blue? Normal.... I mean—that's what's on it?"
PJM: "960, that's what it says."
AEP: "That circuit just tripped. South Canton-Star."
PJM: "Did it?"
AEP: "It tripped and reclosed ..."

INCONSISTENT COMPUTER PROTOCOLS, LACK OF CONTINGENCY PLANS

PJM: "I'm still seeing flow on both those lines. Am I looking at state-estimated data?"
AEP: "Probably."
PJM: "Yeah, it's behind, okay. You're able to see raw data?"
AEP: "Yeah; it's open. South Canton-Star is open.... We have more trouble ... more things are tripping. East Lima and New Liberty tripped out. Look at that.... Oh, my gosh, I'm in deep ..."
PJM: "You and me both, brother. What are we going to do?"

INCOMPLETE INFORMATION, LACK OF REAL-TIME DATA

PJM: "... it looks like they lost South Canton-Star 345 line. I was wondering if you could verify flows on the Sammis-Star line for me."
MISO operator: "Well, let's see what I've got. I know that First Energy lost their Juniper line, too."
PJM: "Did they?"
MISO: "And they recently have got that under control here."
PJM: "And when did that trip? That might have ..."
MISO: "I don't know yet...."
PJM: "And right now I am seeing AEP systems saying Sammis to Star is at 1378...."
MISO: "Let me see. I have got to try and find it here, if it is possible.... I see South Canton-Star is open, but now we are getting data of 1199, and I am wondering if it just came after."
PJM: "Maybe it did."

PJM: Pennsylvania–New Jersey–Maryland
AEP: American Electric Power
MISO: Midwest Independent System operator

beyond the commodity model of electricity delivery into a new era of energy services as diverse as those in today's dynamic telecommunications market.

The EPRI project to design a prototype smart grid, called the Complex Interactive Networks/Systems Initiative, was conducted from 1998 to 2002 and involved six university research consortia, two power companies and the U.S. Department of Defense. It kicked off several subsequent, ongoing efforts at the U.S. Department of Energy, the National Science Foundation, the DOD and EPRI itself to develop a central nervous system for the power grid. Collectively, the work shows that the grid can be operated close to the limit of stability, as long as operators constantly have detailed knowledge of what is happening everywhere. An operator would monitor how the system is changing, as well as how the weather is affecting it, and have a solid sense of how to best maintain a second-by-second balance between load (demand) and generation.

As an example, one aspect of the EPRI's Intelligrid program is to give operators greater ability to foresee large-scale instabilities. Current SCADA systems have a 30-second delay or more in assessing the isolated bits of system behavior that they can detect—analogous to flying a plane by looking into a foggy rearview mirror instead of the clear airspace ahead. At EPRI, the Fast Simulation and Modeling project is developing faster-than-real-time, look-ahead simulations to anticipate problems—analogous to a master chess player evaluating his or her options several moves ahead. This kind of grid self-modeling, or self-consciousness, would avoid disturbances by performing what-if analyses. It would also help a grid self-repair—adapt to new conditions after an outage, or an attack, the way a fighter plane reconfigures its systems to stay aloft even after being damaged.

Who Should Pay

TECHNOLOGICALLY, the self-healing smart grid is no longer a distant dream. Finding the money to build it, however, is another matter.

RESEARCHERS at the Pacific Northwest National Laboratory sit in a simulated regional control center and test prototype software that would provide human operators with real-time grid information, necessary to stop a nascent blackout before it expands.

PACIFIC NORTHWEST NATIONAL LABORATORY

The grid would be costly, though not prohibitively so given historic investments. EPRI estimates that testing and installation across the entire U.S. transmission and distribution system could run $13 billion a year for 10 years—65 percent more than the industry is currently investing annually. Other studies predict $10 billion a year for a decade or more. Money will also have to be spent to train human operators. The costs sound high, but estimates peg the economic loss from all U.S. outages at $70 to $120 billion a year. Although a big blackout occurs about once a decade, on any given day 500,000 U.S. customers are without power for two hours or more.

Unfortunately, research and development funding in the electric utility industry is at an all-time low, the lowest of any major industrial sector except for pulp and paper. Funding is a huge challenge because utilities must meet competing demands from customers and regulators while being responsive to their stakeholders, who tend to limit investments to short-term returns.

Other factors must be considered: What terrorism threat level is the industry responsible for and what should government cover? If rate increases are not palatable, then how will a utility be allowed to raise money? Improving the energy infrastructure requires long-term commitments from patient investors, and all pertinent public and private sectors must work together.

Government may be recognizing the need for action. The White House Office of Science and Technology Policy and the U.S. Department of Homeland Security recently declared a "self-healing infrastructure" as one of three strategic thrusts in their National Plan for R&D in Support of Critical Infrastructure Protection. National oversight may well be needed, because the current absence of coordinated decision making is a major obstacle. States' rights and state-level public utility commission regulations essentially kill the motivation of any utility or utility group to lead a nationwide effort. Unless collaboration can be created across all states, the forced nationalization of the industry is the only way to achieve a smart grid.

At stake is whether the country's critical infrastructures can continue to function reliably and securely. At the very least, a self-healing transmission system would minimize the impact of any kind of terrorist attempt to "take out" the power grid. Blackouts can be avoided or minimized, sabotage can be contained, outages can be reduced, and electricity can be delivered to everyone more efficiently.

Had a self-healing smart grid been in place when Ohio's local line failed in August 2003, events might have unfolded very differently. Fault anticipators located at one end of the sagging transmission line would have detected abnormal signals and redirected the power flowing through and around the line to isolate the disturbance several hours before the line would have failed. Look-ahead simulators would have identified the line as having a higher-than-normal probability of failure, and self-conscious software along the grid and in control centers would have run failure scenarios to determine the ideal corrective response. Operators would have approved and implemented the recommended changes. If the line somehow failed later anyway, the sensor network would have detected the voltage fluctuation and communicated it to processors at nearby substations. The processors would have rerouted power through other parts of the grid. The most a customer in the wider area would have seen would have been a brief flicker of the lights. Many would not have been aware of any problem at all. SA

MORE TO EXPLORE

Technical Analysis of the August 14, 2003, Blackout: What Happened, Why, and What Did We Learn? North American Electric Reliability Council, 2004. Available at **www.nerc.com/~filez/blackout.html**

Toward a Smart Grid: Power Delivery for the 21st Century. Massoud Amin and Bruce F. Wollenberg in *IEEE Power and Energy Magazine,* Vol. 3, No. 5, pages 34–41; September/October 2005.

The Grid: A Journey through the Heart of Our Electrified World. Phillip F. Schewe. Joseph Henry Press, 2007.

SOUTH AMERICA'S MISSING MAMMALS

Startling fossil discoveries in the Chilean Andes reveal an unexpected menagerie of unique mammals that once roamed South America. The finds also overturn long-held wisdom about the continent's geologic history.

By John J. Flynn, André R. Wyss and Reynaldo Charrier

At the edge of a sprawling grassland, a pair of hoofed grazers resembling horses, an antelopelike notoungulate and a ground sloth feed quietly, oblivious to their impending doom. Equally unaware are the chinchilla and the tiny, mouselike marsupial nibbling seeds nearby. Suddenly, one of the jagged, snow-covered volcanoes on the horizon explodes catastrophically, sending a flood of muddy ash down its steep slopes. Soon after, this roiling slurry bursts across the flatter lowlands, entombing the unsuspecting animals in its path.

As devastating as this volcanic torrent was for the creatures it buried, it would become a boon for paleontology. Tens of millions of years after the mammals' untimely deaths, the exhuming forces of mountain building and subsequent erosion exposed remnants of their fossilized skeletons to the light of day high in the Andes Mountains of central Chile. Our team discovered the first of these bones in 1988 while searching for dinosaur remains in an alpine valley of the Tinguiririca River, near the border with Argentina. The initial find-

UNUSUAL MAMMALS that once populated South America go about their business in this artist's reconstruction, unaware that a violent flood of muddy ash from a nearby volcano will soon spell their doom.

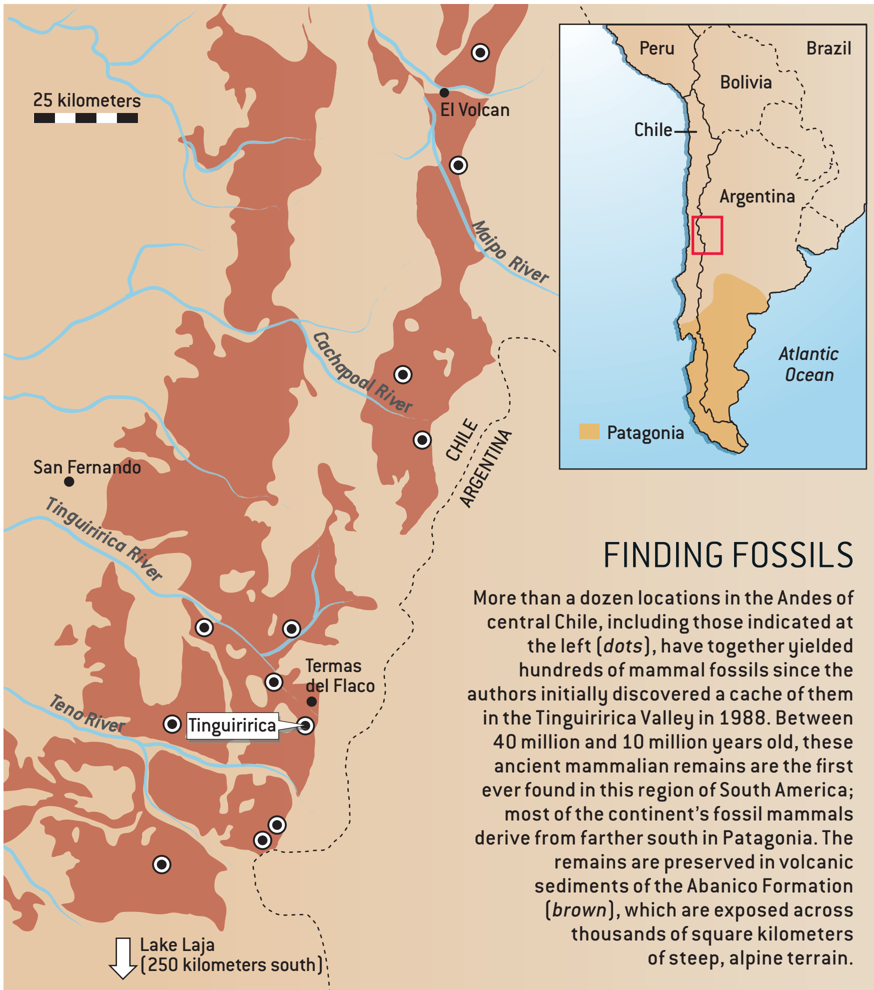
RAÚL MARTIN

ing of mammal bones proved so fruitful that we have returned to the region nearly every year since. So far we have uncovered more than 1,500 fossils of ancient mammals from dozens of sites in the central Chilean Andes.

Painstaking laboratory analysis of our growing collection has yielded major revelations about the history of South America's ancient mammals. To our astonishment, the Chilean fossils range from 40 million to 10 million years old—much younger than anything we expected to find there. Indeed, many of the specimens represent the only mammal remains from segments of that time interval found anywhere in South America. Some of these unique fossils illuminate a previously opaque period in the history of the continent's native mammal lineages; others help to resolve long-standing debates about the origins of key immigrant groups. Together they have revised understanding of when certain ecosystems—and the mountains themselves—appeared in this part of the world.

Tantalizing Discovery

MOST OF WHAT scientists know about South America's ancient mammals is based on clues unearthed in the continent's far southern reaches, mainly Patagonia. Those regions have abundant outcrops of typical fossil-bearing rocks—shale, sandstone and other hardened sediments from rivers and their floodplains. Before our first visit to Chile, researchers had not searched systematically for land animal fossils in that country's mountainous areas, because most rocks there are volcanic. (The standard assumption is that lava and other erupted materials are too hot and violent to preserve organic remains.)



FINDING FOSSILS

More than a dozen locations in the Andes of central Chile, including those indicated at the left (*dots*), have together yielded hundreds of mammal fossils since the authors initially discovered a cache of them in the Tinguiririca Valley in 1988. Between 40 million and 10 million years old, these ancient mammalian remains are the first ever found in this region of South America; most of the continent's fossil mammals derive from farther south in Patagonia. The remains are preserved in volcanic sediments of the Abanico Formation (*brown*), which are exposed across thousands of square kilometers of steep, alpine terrain.

We decided to take a chance that the Tinguiririca Valley might harbor fossils when we learned of a report about dinosaur footprints. The rocks were the right age—geologists then assumed that most rocks along the main spine of the Chilean Andes dated back at least 65 million to 100 million years, to the latter part of the Mesozoic era, when dinosaurs reigned supreme. We knew that any sediments preserving footprints might also contain bony remnants of the track makers. If we were extremely lucky and kept our eyes close to the ground, we might even find a fossil of one of the dinosaurs' tiny mammalian contemporaries, which were no bigger than shrews.

On the last day of a one-week reconnaissance trip in 1988, our team of four split up to prospect the precipitous slopes flanking each side of the Tinguiririca River. Almost immediately, the pair working north of the river reached the layer of ancient sediments that bore the dinosaur tracks, then continued up the valley in search of more potential fossil-bearing deposits. To their dismay, however, the only fossils they recovered were from fish, ammonites and other ocean-dwelling creatures—no reptiles or mammals. Meanwhile the team members working on the south side of the river were having a similarly frustrating day. Late in the af-

Overview/*Fossils Galore*

- The more than 1,500 fossil mammal specimens discovered in the Andes of central Chile include a remarkable array of new species as well as the continent's oldest known rodent remains.
- From 40 million to 10 million years old, these new fossils bridge gaps in the known history of South America's unusual mammals.
- Some of the fossils document the existence of a 32-million-year-old grassland that predates such ecosystems elsewhere in the world by some 15 million years.

LUCY READING-IKKANDA

ternoon, though, their spirits soared when they spied a few fossilized scraps of bone and teeth eroding out of a large patch of reddish-brown volcanic sediments nearly 1,000 meters above the valley floor. A closer look revealed that the fossils were land-dwelling vertebrates about the size of a small horse.

At first, we tried to shoehorn these fragments into the prevailing view about the age of the rocks—animals this size must be peculiar dinosaurs or other odd Mesozoic beasts. But the complex, differentiated teeth with the high-crowned, flat-topped and multifaceted molars that are unique to some mammals told a different story. These mammals were clearly too large—and too advanced—to have lived anytime before about 50 million years ago. Apparently geologists had been a long way off in their estimate of the age of these rocks. Indeed, later analyses confirmed that the new fossils came from well within the Cenozoic era, the ongoing period of the earth's history that began when the nonavian dinosaurs went extinct 65 million years ago. (Birds are now known to be theropods, thereby representing a living group of dinosaurs.)

Island Menagerie

FINDING FOSSILS of any kind would have been great news for us. That they were mammals—and unexpectedly recent ones at that—was more than enough motivation to focus our next field season on that single area. We returned to the Tinguiririca Valley in the austral summer of 1989, after the high mountain snows melted enough for local authorities to reconstruct the small access road, which washes out almost every spring. This time we reached the fossil site on a clear, sunny January morning with a crew of seven scientists and complete expedition gear. Quickly unloading the pack animals, we set up camp near a small creek and started hunting.

To our delight, exquisite shards of bones and teeth popped into view within minutes of our beginning to comb the hillside. Protruding out both ends of one potato-size nodule of rock was a skull that was unmistakably mammalian, as indicated by, among other characteristics, its single pair of lower jawbones. (The lower jaws of reptiles are made up of many separate bones.) Later, we would formally describe that creature as a new species of notoungulate, a diverse group of hoofed herbivores that ranged from rabbit- to hippo-size and became extinct less than one million years ago. This new species probably looked a bit like an antelope. The teeth we found the previous year turned out to be from a rhinolike notoungulate. In all, during our first three seasons of fieldwork in Tinguiririca, we would turn up more than 300 specimens, including marsupials, an early sloth, armadillos and a chinchillalike rodent.

The full significance of our finds would take years to unfold, but we knew right away that we were onto something big. These new fossils clearly had much to say about the history of South America's distinctive living mammals—sloths, monkeys, anteaters and chinchillas among them. The ancestors of these animals, including many of the new creatures we were finding in the Tinguiririca Valley, were among those that evolved when South America was an island continent. For most of the past 80 million years, following the breakup of the supercontinent Pangaea and its southern portion, Gondwana, tectonic plate motions kept South America separated from other landmasses. This period of isolation fostered the evolution of native mammals uniquely adapted to that island's conditions and every bit as bizarre as those indigenous to modern

Cast of Characters

During five seasons of exploring the Tinguiririca Valley of central Chile, our team discovered fossils of 25 species of mammals. These ancient animals, most of them new to science, lived about 32 million years ago. Known collectively as the Tinguiririca fauna, the more than 300 individual specimens feature the oldest South American rodents, including an ancient chinchillalike form that is currently unnamed (*1*). Fossils of a new species of horselike *Eomorphippus* (*2*) and *Santiagorothia chiliensis* (*3*) are the oldest representatives of two subgroups of notoungulates, hoofed herbivores that became extinct in the past 20,000 years. The fauna also includes South America's greatest diversity of the notoungulates known as archaeohyracids—*Archaeotypotherium tinguiriricaense* (*4*) among them. Unique to Tinguiririca are *Pseudoglyptodon chilensis* (*5*), the closest known relative of sloths, and shrewlike *Klohnia charrieri* (*6*), the only marsupial of its kind. *—J.J.F., A.R.W. and R.C.*

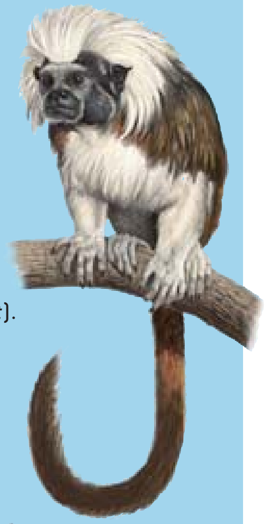
RAÚL MARTIN

African Immigrants

New World monkeys and caviomorph rodents (the group encompassing today's capybaras, chinchillas and their kin) were not original inhabitants of South America. Rather they arrived before 25 million years ago, while South America was an island. Some researchers have argued that the animals' ancestors made the relatively short water crossing from North America, but new fossils discovered in central Chile strongly suggest that both groups are more closely related to predecessors in Africa. Presumably the original colonists traveled from Africa to South America on floating islands of vegetation or by some other rare dispersal mechanism that scientists may never be able to identify.

▼ MONKEY: *Chilecebus carrascoensis*

TINY SKULL (five centimeters long and 20 million years old) belonged to a small monkey that weighed a kilogram or less and may have resembled a modern marmoset (*right*).



▼ RODENT: *Unnamed new species*

JAW FRAGMENT (two centimeters long) comes from a 32-million-year-old rodent and is one of the two oldest rodent fossils ever found in South America. This one may have looked like a modern agouti (*right*).

islands such as Australia (like the platypus and koala) and Madagascar (famous for its lemurs). Unusual predecessors of South America's modern groups include hopping marsupials; saber-toothed marsupial "pseudocats"; armadillo cousins equipped with massive, spike-studded tail clubs; bear-size rodents; sloths as big as elephants; and sloths that swam in the sea.

Knowledge of the ancestors of South America's living mammals had been gleaned from earlier fossil discoveries in Patagonia and elsewhere, but critical information about many of these forebears remained elusive. Paleontologists knew, for instance, that sloths and anteaters got their start prior to 40 million years ago, as did several exotic lineages that are now extinct (including certain marsupials and notoungulates). But no fossils representing the transition to the second phase of South American mammal history—from about 40 million to 30 million years ago—had ever been discovered. Most exciting for us during those first few years, then, was the growing realization that the animals we found at Tinguiririca lived during this period of previously unknown history.

Paleontologists had suspected that during this mysterious gap in the fossil record many of the continent's unique lineages underwent explosive diversification. Indeed, our collection includes the earliest record of several groups of notoungulates and represents at least 25 mammal species, nearly all of them new to science [*see box on preceding page*]. The era also saw the arrival of rodents and primates, neither of which were among South America's original mammal inhabitants.

Seafaring Rodents

ONE OF THE MOST significant of our Tinguiririca discoveries was a fossil of the earliest known South American rodent, a find that lends powerful evidence to a debate over the origins of today's living capybaras and chinchillas. Known as caviomorph rodents, these creatures and their immediate kin make up South America's most ancient rodent lineage (and are distinct from the younger rodent line of rats, mice and related creatures that arrived from the north about 3.5 million years ago, when the Isthmus of Panama first reconnected the two Americas). Paleontologists agreed that the first caviomorph rodents arrived sometime within the broad time span between 55 million and 25 million years ago, while South America was still an island. A few younger caviomorph fossils hinted that the predecessors came from Africa, but many researchers found it easier to imagine that the immigrant rodents made the shorter trip from North America, possibly via a chain of Caribbean islands.

To help settle the debate, we compared the anatomical details of the Tinguiririca animal to rodent remains found elsewhere in the world. Most informative was the shape of the tiny teeth still rooted in the lower jaw (the upper jaw and molars have yet to be found). That shape implied that the Tinguiririca animal's upper molars had five distinct crests—as did the upper molars of African rodents from the same period. In contrast, North America's ancient rodent species had only four crests on their upper molars. These comparisons strongly suggest that the Tinguiririca rodent is more closely related to animals in Africa. The absence of plausible caviomorph forebears in older North American fossil beds also supports the out-of-Africa theory.

Presumably the original caviomorph colonists traveled from Africa to South

RAÚL MARTIN (*illustrations*); JOHN WEINSTEIN *Field Museum of Natural History* (*monkey skull*); MARLENE DONNELLY *Field Museum of Natural History* (*rodent jaw*)

America on floating logs or other rafts of vegetation—scientists' best guess for how various unusual plants and animals made their way to many geographically isolated regions [see "Madagascar's Mesozoic Secrets," by John J. Flynn and André R. Wyss; SCIENTIFIC AMERICAN, February 2002]. The idea of such an incredible transoceanic journey may seem far-fetched, but it is more plausible in the context of the global environment before about 32 million years ago. At that time, the South Atlantic was only about 1,400 kilometers at its narrowest point—half as wide as it is today—and east-to-west ocean currents in the tropics were strengthening sporadically.

These conditions would have permitted a journey in about two weeks, and the animals may have gone into torpor (inactivity and greatly reduced metabolism during times of stress). Moreover, sea level was dropping at the time (because of the formation of ice sheets on and around Antarctica), so one or more volcanic "stepping stone" islands, now submerged, may have made the crossing easier.

Emerging Ecosystems

USING A NEW and extremely precise dating method that analyzes trace amounts of argon gas trapped inside crystals of the fossil-bearing rocks, we determined that the rodents and other Tinguiririca mammals were between 33 million and 31.5 million years old. Growing Antarctic ice sheets and other phenomena indicate that global climate at that time was becoming cooler and drier. Knowing that a major shift in climate occurred just as the Tinguiririca mammals were thriving led us to test whether the animals and their environment might have been responding to those changes.

Several lines of indirect evidence have enabled us to reconstruct the habitat of the Tinguiririca mammals even though we have never found plant fossils from the same rocks. Our early analysis of their teeth told us that the Tinguiririca animals must have lived in a very different ecosystem than their immediate forebears did. Most of the older South American mammals known from about 65 million to 34 million years ago were herbivores that browsed on typical forest greenery, such as tree leaves and herbs. (Indeed, plant fossils do confirm that lush forests probably covered much of the continent during that period.) Mammals that eat these soft foods, including humans, typically have low-crowned teeth with a thin cap of protective enamel covering each tooth only to the gum line.

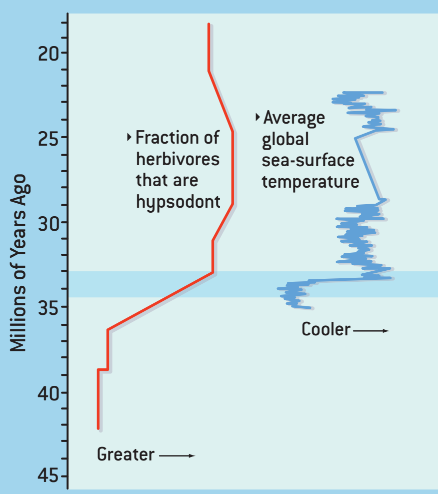
In sharp contrast, most of the Tinguiririca herbivores possessed extremely high crowned teeth with enamel extending past the gum line nearly to the root tips, a condition known as hypsodonty. The additional enamel (much

THE AUTHORS

JOHN J. FLYNN, ANDRÉ R. WYSS and *REYNALDO CHARRIER* together have explored the fossil history entombed in the Chilean Andes for the past 20 years. Flynn is chairman and Frick curator of the division of paleontology at the American Museum of Natural History in New York City and dean of the museum's new Richard Gilder Graduate School. Wyss is professor of earth science at the University of California, Santa Barbara. Charrier is professor of geology at the University of Chile in Santiago. The authors thank the many dedicated fossil preparators, the National Museum of Natural History in Santiago, the National Monuments Council of Chile, the U.S. National Science Foundation, NASA, and the National Commission for Scientific and Technological Research of Chile for their exceptional support of this research.

Teeth Tell All

Most of the 32-million-year-old fossil herbivores discovered in Chile's Tinguiririca Valley have high-crowned, or hypsodont, cheek teeth like those modern cattle use to grind gritty, fibrous plants. These specialized molars suggest that the Tinguiririca region was a dry grassland when those animals lived. Earlier, much of South America was heavily forested, and a much smaller proportion of its herbivore species were hypsodont (*red line in graph*). Global climate change (*dark blue line*) to cooler and thus drier conditions around 34 million years ago (*light blue bar*) may explain why the lush forests dried out.



▲ **HIGH-CROWNED TEETH**
The broken right lower jaw of a sheep-size Tinguiririca herbivore displays the abrasion-resistant teeth typical of grassland grazers. The crowns, which are coated with protective enamel, extend well below the gum line toward the root tips—providing more material to wear down as the teeth grow.

▲ **LOW-CROWNED TEETH**
In the right lower jaw of this as yet unnamed ancient Chilean herbivore, protective enamel covers only the portions of the crowns that extend above the gum line. Such teeth are ill suited to a gritty diet; they usually belong to animals that browse on leaves or other soft food.

JOHN J. FLYNN, ANDRÉ R. WYSS AND REYNALDO CHARRIER (*high-crowned teeth*); ALLISON SMITH *American Museum of Natural History* (*low-crowned teeth*)

harder than the dentin in tooth interiors) makes hypsodont teeth more wear-resistant than low-crowned teeth are. The Tinguiririca herbivores almost certainly evolved such teeth in response to abrasive particles within the foods they ate, as did cows, antelopes, horses and other animals that eat gritty grasses in open prairies and savannas elsewhere in the world. It is also significant that two thirds of species of the entire Tinguiririca fauna were hypsodont. The proportion of hypsodont taxa relative to other dental types generally increases with the amount of open habitat, and the Tinguiririca level of hypsodonty surpasses even that observed for mammals living in modern, open habitats such as the Great Plains of central North America.

The findings implied that the Tinguiririca herbivores grazed in open grasslands rather than in forests, but the teeth are not alone in pointing to this conclusion. Our former graduate student Darin Croft, now a professor at Case Western Reserve University, provided two other independent deductions about the annual rainfall and vegetation that dominated the ancient Tinguiririca ecosystem. Statistical analyses of the number of species within different body-size categories and their ecological attributes—so-called cenogram and macroniche analyses—revealed that the Tinguiririca animals most closely resembled modern fauna that live in dry grasslands with patchy woodlands, such as parts of the savanna of Africa or the Caatingas and Chaco habitats of South America.

Our conclusion that the ancient Tinguiririca habitat was open, relatively dry and contained abundant grasses came as a bit of a surprise, considering that all previous evidence suggested that the first open grasslands on other continents did not appear until 18 million years ago. The apparent emergence of the Tinguiririca grasslands some 15 million years earlier could have resulted from the global trend toward aridity and cooling around that time—perhaps accentuated by the rain shadow cast by the rising Andes Mountains. Grasslands would have been better suited to a cooler, drier climate than the lush forests of previous millennia were. So far, though, suspicion that grasslands resulted from global cooling needs more scrutiny. Further testing a direct causal connection is an avenue of future research.

▲ FOSSIL-CONTAINING LAYERS, pushed upward and tilted almost vertically, are spectacular evidence of the tectonic forces that have been squeezing central Chile for millions of years. These steep slopes are located near Lake Laja, about 300 kilometers south of the Tinguiririca Valley.

PHILLIP GANS *University of California, Santa Barbara*

Late to Rise

Making the fossil mammals of the Chilean Andes all the more significant are the ways they have substantially advanced understanding of the geologic history of this segment of the Andes. The unexpectedly young age of the ancient mammals overturns prevailing assumptions about the age of their host rocks, indicating that this part of the Andes formed at least 70 million years later than previously thought. Most of the rocks making up the spine of the main range of the central Chilean Andes, first named and described by Charles Darwin, were long believed to date back at least 100 million years, and the initial phases of uplift were thought to be similarly ancient. Dating these fossil-bearing deposits enabled our team to calculate, for the first time, precisely when this part of the Andes began rising: between 15 million and 18 million years ago. The rise continued episodically until the present and still goes on today. It is now clear that extensive, sediment-filling basins formed during volcanic phases in the range's history, both shaping the area's unique ancient ecosystems and providing a means of preserving its spectacular fossil mammal record. *—J.J.F., A.R.W. and R.C.*

No Monkeying Around

HAVING UNCOVERED such a rich trove of paleontological and environmental information at several sites in the Tinguiririca Valley, we began to consider whether this part of the Andes was a paleontological "one-hit wonder." Soon after, an unexpected obstacle led us to an answer. In the spring of 1994 reconstruction of the road leading into the Tinguiririca Valley was especially slow, but we did not learn of this impediment until arriving on the scene. Turning frustration into an opportunity, we began explorations outside the valley.

We searched other major valleys sporting outcrops of the same volcanically derived sedimentary rocks, which are exposed across thousands of square kilometers of mountainous terrain.

HERBIVORE SKULLS, as first seen in the field, are preserved in a coarse matrix of fossilized ash and mud from ancient volcanic eruptions. Even before researchers removed the bones from the rock, they identified both specimens as types of extinct hoofed herbivores known as notoungulates. The high-crowned teeth typical of grass-eaters are particularly notable on the specimen at the left. The other skull, which comes from an older leaf-eating notoungulate called a notopithecine, was recently discovered near the Teno River.

ANDRÉ R. WYSS *(left)*; DARIN CROFT *Case Western Reserve University (right)*

Over the ensuing years of fieldwork, we determined that fossil mammals were indeed not restricted to the Tinguiririca Valley and that the volcanic torrent that inundated that ancient landscape was not a one-time, isolated cataclysm. Rather, seen over a time frame of millions of years, such devastating events occurred with some frequency. Each time, countless layers of material from additional eruptions buried the older deposits (and the bones they contained) ever more deeply. Ultimately, this layered stack of sediment (now transformed into rock) and lavas measured upward of three kilometers thick. Later, collision of converging tectonic plates squeezed this unruly pile skyward.

Our ongoing analyses of the numerous faunas, which range between 10 million and 40 million years in age, are revealing fresh insights into the region's history. One of our most significant newer finds—from a site about 100 kilometers north of Tinguiririca, in the drainage basin of the Cachapoal River—is the most complete skull of an early New World monkey yet discovered. The five-centimeter-long skull, with both eye sockets and every tooth in the upper jaw intact, came from a petite monkey weighing about one kilogram at the most. Named *Chilecebus carrascoensis*, this creature resembled modern New World monkeys, such as marmosets and tamarins. As with the caviomorph rodents, experts had long debated whether New World monkeys originated in North America or Africa. But anatomical details of the *Chilecebus* skull and teeth argue for its common heritage with a group of primates originating in Africa. Like the caviomorph rodents, it seems the ancestors of *Chilecebus* somehow made the Atlantic crossing from Africa.

Beginning with the Tinguiririca fauna and culminating with the New World monkey and other ongoing discoveries across central Chile, volcanic deposits once ignored for fossils have turned out to contain superbly preserved bones and are now recognized as a premier archive of South American mammal evolution. Over the years, we have developed a keen sense of what auspicious rocks look like, sometimes even being able to spot them from many kilometers away. These fossils are hard-won, though, given the precipitous topography and the remoteness of many of the localities. Some sites lie within a few kilometers of gravel roads or dirt trails, but most can be reached only by long approaches on foot, horse or even helicopter. We jokingly refer to "Andy's Rule" (dubbed in honor of Wyss), which says that the difficulty of getting there is proportional to the quantity and quality of fossils we will find.

Together the Cenozoic fossils of the Chilean Andes are helping to elucidate mammalian evolution and environmental transformations in South America, a continent whose long history of isolation represents a splendid natural experiment for investigating large-scale evolutionary phenomena. SA

sa Learn the story behind "Andy's Rule"—photographs included—at *www.sciam.com/ontheweb*

MORE TO EXPLORE

Splendid Isolation: The Curious History of South American Mammals. George Gaylord Simpson. Yale University Press, 1980.

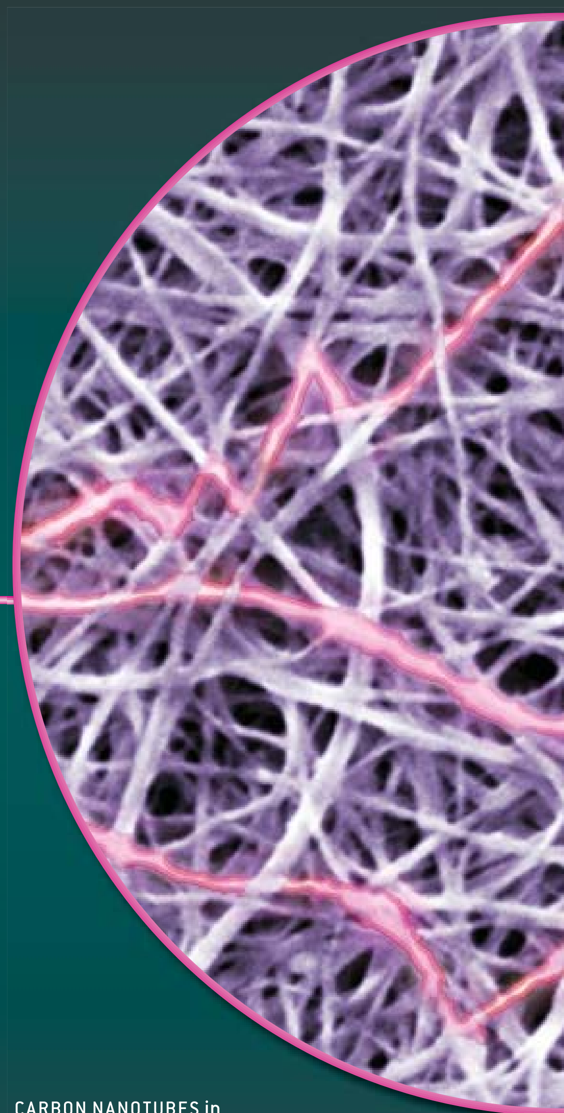
Cenozoic Environmental Change in South America as Indicated by Mammalian Body Size Distributions (Cenograms). Darin A. Croft in *Diversity and Distributions,* Vol. 7, No. 6, pages 271–287; November 2001.

The Tinguiririca Fauna, Chile: Biochronology, Paleoecology, Biogeography, and a New Earliest Oligocene South American Land Mammal "Age." J. J. Flynn, A. R. Wyss, D. A. Croft and R. Charrier in *Palaeogeography, Palaeoclimatology, Palaeoecology,* Vol. 195, No. 3/4, pages 229–259; June 15, 2003.

For more details about the mammal lineages that inhabited South America when it became geographically isolated, visit **www.dcpaleo.org/Research/SAMammals/SAMammals.html**

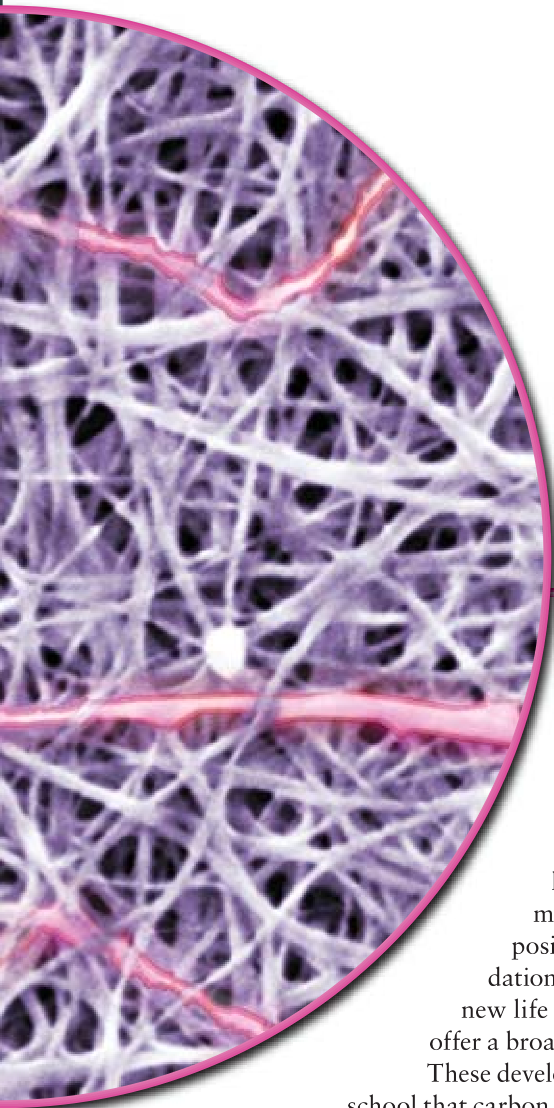
CARBON NANONETS

SPARK NEW ELECTRONICS



CARBON NANOTUBES in a weblike mesh ensure multiple alternative pathways for electrons (*pink highlights*), providing surefire electrical conduction. The entire field of view is 0.7 micron in diameter.

JOHN YORSTON *Carl Zeiss SMT* AND ANDREW PAUL LEONARD *APL Microscopic*; COLORIZED BY JEAN-FRANCOIS PODEVIN



Random networks of tiny carbon tubes could make possible low-cost, flexible devices such as "electronic paper" and printable solar cells

By George Gruner

In many classic science-fiction stories, alien life is based on silicon—the substance at the core of modern electronics technology—rather than on carbon, the fundamental building block of earthly biology. Scientists have even speculated that they might someday create silicon life-forms. Instead the opposite is starting to happen: carbon is serving as the foundation for electronic devices—and in the process is breathing new life into the quest for inexpensive, flexible products that offer a broad range of capabilities.

These developments may surprise those of us who learned in high school that carbon, in its familiar incarnations of diamond and graphite, does not conduct electricity well. During the past 15 years, however, researchers have discovered new forms of carbon: very small structures comprising a few hundred to 1,000 atoms, through which electrons travel with ease. Of particular interest is the carbon nanotube, a molecule that resembles rolled-up chicken wire, only the "wire" is a sheet of carbon atoms that is 100 million times as small as the version used for chicken coops.

Investigators have found that random networks of carbon nanotubes—called nanonets—can perform a variety of basic electronic functions. Using novel chemistry, researchers can make such networks mimic the conductive properties of metals such as copper or the less conductive characteristics of semiconductors such as silicon. These innovations have paved the way for this single material to assume different roles in electronic devices.

Further, engineers can construct such carbon-based devices by employing simple fabrication methods. Researchers can dissolve the tubes in a liquid and

spray the resulting solution to form thin layers on, say, flexible plastic sheets. They can also lay or print these materials on other layers that have various electronic functions; for example, substances that emit light when a voltage is applied.

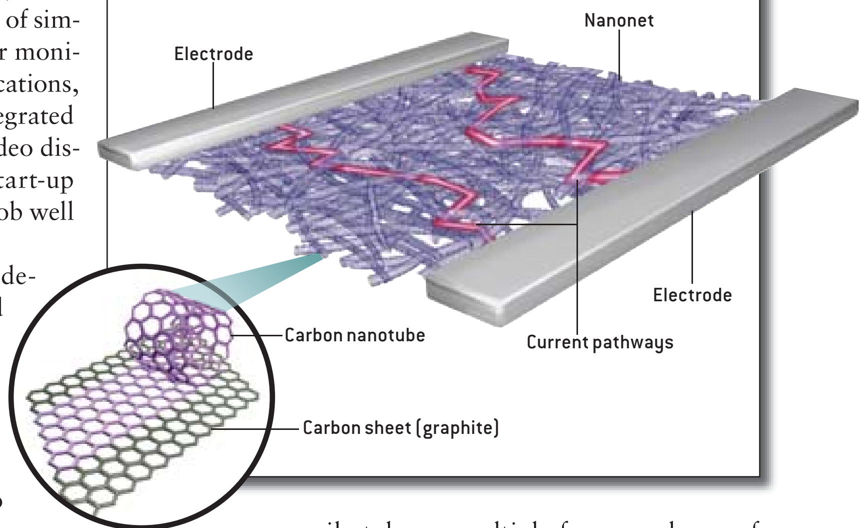
It takes little imagination to see how this kind of straightforward system could form the basis of many extremely cheap but handy products: "electronic paper" that could display information on sheets that roll up like conventional newsprint; chemical sensors; wearable electronic devices; solar cells that could be printed onto rooftop tiles; or scads of simple radio-frequency identification (RFID) sensors for monitoring warehouse or store inventories. For such applications, the expensive, lightning-fast processing power of integrated chips like Intel's Pentium processors or Samsung's video displays is not needed; rather R&D laboratories and start-up firms are racing to find technologies that can do the job well enough at low cost [*see table on page 83*].

Such exciting applications would make severe demands on today's electronic materials: they would need to be conductive, flexible, lightweight, transparent (at least for some applications, such as solar cells and displays) and inexpensive. But most conductors are metals, the majority of which are not transparent, whereas, as a rule, thin films of materials that are transparent, such as diamond, are insulators (they do not transmit electricity). Light can pass through one special class of metals, called metal oxides, however. The best known, indium tin oxide, is frequently used where engineers need see-through electrodes. But metal oxides are costly. They are also heavy and brittle, and their manufacture requires high processing temperatures and multibillion-dollar fabrication facilities.

Another alternative is an unusual category of plastics known as conducting polymers. Although common plastic substances are insulators, chemists have in recent decades managed to convert some polymers into semiconductors and even full-fledged conductors. Polymers can be produced using room-temperature techniques. Lightweight and flexible, they can easily take on multiple forms and are, of course, dirt cheap [see "Next Stretch for Plastic Electronics," by Graham P. Collins; SCIENTIFIC AMERICAN, August 2004]. On the downside, weak bonds hold together the atoms in most plastics. The bonds can break rather readily, which leads most polymers to degrade over time. Consider just how useful a solar cell would be if it failed after only a few warm, sunny days.

FROM NANOTUBES TO NANONETS

Thin networks of carbon nanotubes, or nanonets, can serve as electronic devices. Each nanotube is a one-atom-thick sheet of carbon, called graphite, rolled into a cylinder with a diameter of about a nanometer (*inset*)—approximately 50,000 times as small as the width of a human hair. Electric current passes through the interconnected tubes from one electrode to another.



Overview/*Nanonet Electronics*

- Carbon nanotubes—minuscule cylinders of rolled-up carbon sheets—conduct electricity well, which could make them useful for many exciting electronic applications. But manufacturing products that use single tubes is expensive and suffers from significant reliability problems.
- Random networks of many carbon nanotubes, called nanonets, enable numerous basic electronic functions at low cost. The durable nature of nanonets also makes them suitable for portable devices.
- Carbon nanonets should find use in sensors, solar cells, electronic paper, and flexible touch screens and displays within a few years.

A Better Wire

ENTER CARBON-BASED NANOWIRES. Carbon nanotubes were first discovered several decades ago, but no one realized their value at the time. Then, in 1991, Japanese chemist Sumio Iijima of NEC Corporation rediscovered them. These tiny tubes of carbon have a diameter of around one nanometer—about the same as a strand of a DNA molecule [*see box above*]. The electrical conductivity of the tubes is comparable to that of copper and surpasses that of any polymer by several orders of magnitude. They can also carry more than 100 times more current than the best metals. Carbon nanotubes are, in addition, physically robust: they can be bent easily, they do not react with most chemicals and they resist damage from day-to-day use.

Manufacturers make nanotubes by reducing coal into its component atoms using the heat of an electric arc or a laser, which creates a so-called carbon plume. They then add catalysts to the plume, which promotes the formation of various types of carbon molecules. This relatively straightforward procedure produces what is essentially soot—carbon molecules in many forms, including spherical structures called

JEAN-FRANCOIS PODEVIN

buckyballs as well as other "fullerenes" and carbon nanotubes. Fabricators must then laboriously separate the nanotubes out of the mixture. Techniques focus on separating out only the long, nearly perfect specimens that have a single "chicken wire" wall (rather than multiple, concentric walls). Suitable nanotubes are thus currently quite pricey, but makers are confident that costs will drop significantly if market demand rises enough to justify high-volume manufacturing facilities.

When a single nanotube is employed to build a transistor, the voltage-activated switch that is the workhorse of modern electronics [see "Nanotubes for Electronics," by Philip G. Collins and Phaedon Avouris; Scientific American, December 2000], the resulting device can easily outperform the transistors on the silicon chips in today's computers. But single carbon nanotubes will not replace silicon and copper in the foreseeable future. The main obstacle lies in their manufacturability, which is one of the most vexing problems afflicting nanotechnology's commercialization. Current devices based on a single nanotube can take days to make, because they often must be assembled by hand. Another difficulty is performance variability. Nanotubes come in slightly different shapes and forms, which affects their electrical attributes.

From One Wire to a Network

Although individual tubes differ from one to the next, researchers realized that this variation could be averaged out by using many tubes together—any shortcomings present in some of the tubes could be compensated for by better-performing counterparts. The simplest example is a random network of nanotubes [*see box on opposite page*]. Just as an interstate highway system can offer alternative routes when you encounter a traffic jam on one roadway, so, too, can a random assembly of electrically conducting nanotubes—a nanonet—speed the transmission of electrons by providing alternative pathways. Investigators soon established that these nearly two-dimensional random networks offered interesting properties in their own right.

First, the nanonet's many pathways and connections guarantee good electrical conductance between one electrode and another, despite possible manufacturing flaws. A good analogy is the freeway system that serves the Los Angeles metropolitan area. No one would want to attempt to traverse the City of Angels by hiking cross-country or driving the slow, stoplight-strewn surface roads; instead travelers take the freeway. The same concept applies to the nanonet, which allows electrons to jump on the tubes and move around on what is essentially a nanoscale freeway system. The multiple avenues provided by these networks also afford a considerable resistance to failure, or fault tolerance; if one route breaks from use, others are there to take up the slack.

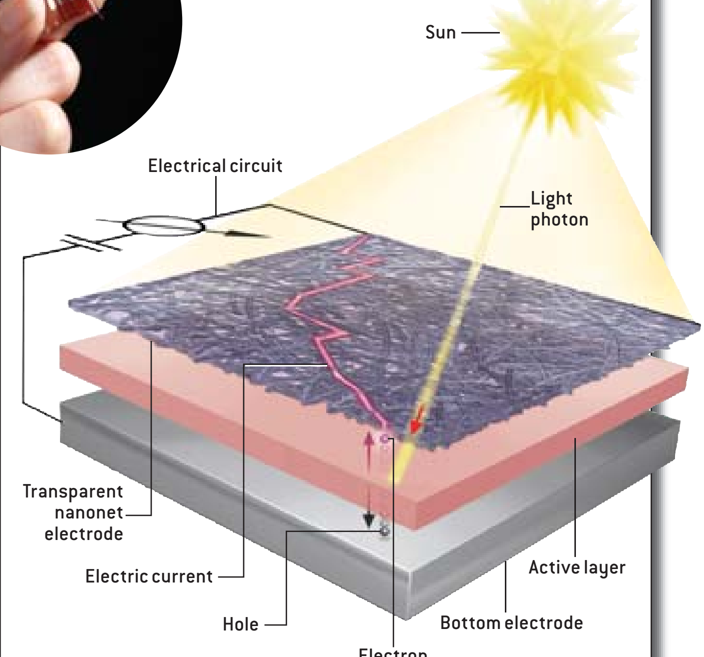
A conductive nanonet is in fact a simple example of the concept of percolation, which describes how objects, materials or electric currents move through a random medium. Imagine dropping pickup sticks on a tabletop one at a time. With only a few sticks, the chances of finding a connected pathway (by going from one stick to the next) from one end of the table to the other are slim. In fact, below a certain critical density of sticks, the odds drop to zero. But as the number of sticks increases, the pile will eventually surpass that critical density, the so-called percolation transition, where at first one and then more and more pathways form. If the pickup-stick approach were applied to copper wires on the tabletop, at some point the network would achieve electrical conduction across the table as well—with the current dependent on the density of the copper wires. Theorists studied this concept some time ago, and my group at the University of California, Los Angeles, was able to map out such a transition in networks of nanotubes.

Nanonets can be, in addition, highly transparent—an advantage in applications that require light transmission. Just as the freeway pavement covers only a small fraction of the natural terrain, a web of long and skinny wires allows passage of most of the incident light—a fraction that approaches 100 percent for what can be considered one-dimensional nanowires.

Finally, much like a spiderweb, a network of nanowires typically is more robust than the same material in undifferentiated bulk, which often tends to break when bent. These

CHEAPER SOLAR CELLS

Transparent and flexible carbon nanonet electrodes would be significantly cheaper than the rigid indium tin oxide electrodes used in today's solar cells (*shown in exploded view*). The carbon nanonets would conduct electric charges (negative electrons and positive "holes") dislodged by light photons from an active, charge-creating, layer of a semiconductor to an external circuit. A sample printed nanonet solar cell shows its bendability (*inset*).



JEAN-FRANCOIS PODEVIN; BRAD SWONETZ *Redux* (*solar cell, Stanford University*)

characteristics make the nanonet architecture eminently suited for applications in which resistance to day-to-day use and misuse is important. Think about how many times you have dropped your cell phone or iPod.

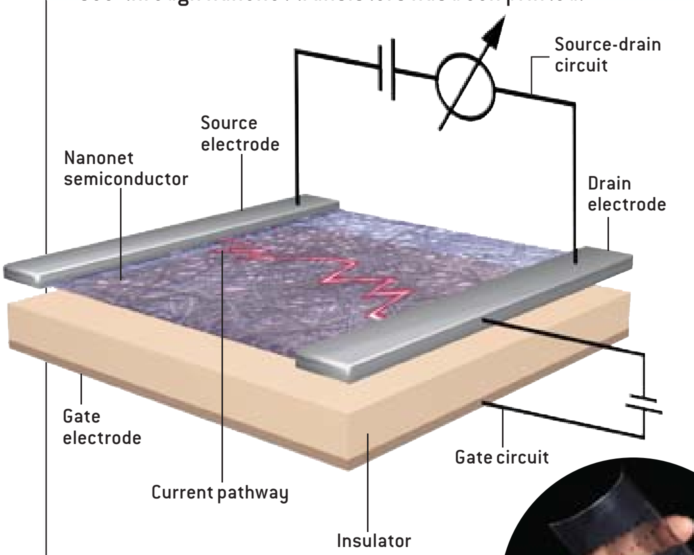
Weaving a Nanonet

THESE PERFORMANCE BENEFITS augur well for the technology's potential in real-world applications, but any new replacement material must, of course, be more than competitive in terms of function and cost with current materials. Nanotube films initially made a couple of years ago—by my team, by a group led by physicist Siegmar Roth at the Max Planck Institute for Solid State Research in Stuttgart, Germany, and by one at the University of Texas at Austin—were not up to the task. Finding the optimal processing routes and the most advantageous way to deposit the tubes onto surfaces were not trivial problems.

Clearly, one cannot fabricate thin films of such networks by merely throwing down a tube at a time like playing pickup sticks; another strategy is needed. One might, for instance, dissolve the tubes in a solvent (water, alcohol, organic liquids) and then spray the resulting fluid onto a surface, but that is not as easy as it sounds. When mixed in a liquid, the tubes tend to bundle together, requiring a chemical additive to keep them apart. Some agents, called surfactants (soaps), accomplish this job by completely surrounding the tubes. But surfactants, if they remain on the tubes after they are sprayed onto a surface, impede the flow of electrons between tubes (blocking the freeway ramps, so to speak). Through steady trial-and-error efforts with innumerable solvents, surfactants and processing procedures, however, researchers have created simple (room-temperature) avenues to make such thin films of nanotube networks. At the moment, a method pioneered by my team and a group led by chemist Andrew Rinzler at the University of Florida yields films that have the lowest electrical resistance and thus the best operating performance to date among nanonet-based devices.

TRANSPARENT TRANSISTORS

Carbon nanonet films, tailored to perform like semiconductors, can serve as the basis for field-effect transistors—the building blocks of computers, cellular phones and other digital devices. This switchlike mechanism (*shown in exploded view*) uses a small electric voltage provided by the gate electrode to greatly boost the current in the source-drain circuit. In the inset, a technician bends a plastic sheet onto which an array of see-through nanonet transistors has been printed.



As researchers experimented with the conductivity of the tubes, they learned that the material could be transparent, a property that is important for applications such as displays and solar cells. The discovery that carbon nanonets are transparent to light came about as a by-product of research on their conductivity. The first indication that nanonets could be clear arose in 2001, when my former postdoctoral associate Leonardo Degiorgi and his group at the Swiss Federal Institute of Technology in Zurich, along with physicist David Tanner and his co-workers at the University of Florida, studied their optical characteristics. To measure nanonet conductivity precisely, they fabricated thick films: these were too deep to transmit light, but their data led them to conclude that a thinner film would be both transparent and a good conductor. After these groups made this determination, Rinzler's team (collaborating with Katalin Kamarás and her colleagues at the Central Institute of Physics in Budapest) and mine at U.C.L.A. followed up with direct measurements of a nanonet film's optical transparency. Today scientists can fabricate tailor-made films with different levels of transparency and electrical conductivity by changing the thickness of the films.

Nanonet Transistors

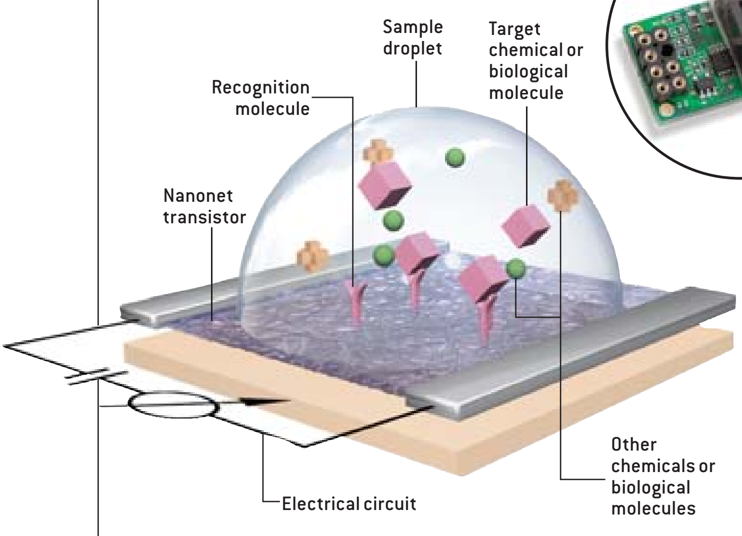
RESEARCHERS SOON TURNED their attention from making nanonet conductors to nanonet semiconductors that could serve as the basis for transistors. A transistor requires materials whose conductivity changes greatly in response to only small incremental inputs, such as altering an electric field [*see box at left*].

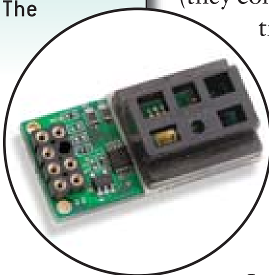
The notion that carbon nanotube networks could serve as the backbone of thin-film field-effect transistors emerged around seven years ago. Thereafter, progress was relatively rapid—with advances in creating nanonets on flexible substrates, demonstrating the transparency of the devices, coming in short order. Working in parallel, my R&D group at Nanomix, a start-up firm in Emeryville, Calif., where I served as chief scientist, and a research team at the Naval Research Laboratory in Washington, D.C., led by materials scientist Eric Snow, produced nanonet transistors in 2003. But these devices were formed on rigid glass substrates at processing temperatures of 900 degrees Celsius—too hot for use with flexible plastic substrates that melt at

JEAN-FRANCOIS PODEVIN; BRAD SWONETZ *Redux* (*transistor, Unidym*)

LOW-COST SENSORS

A nanonet device can become a sensor with the addition of "recognition molecules" that react with a target chemical or biological molecule. When the target binds to a recognition molecule, it alters the sensor's electrical output. Such devices can detect many chemicals, including a blood-borne cancer marker called prostate specific antigen (PSA) and, soon, microorganisms such as the anthrax bacterium, a potential bioweapon. Arrays of nanonet sensors, each with different recognition molecules, could cheaply detect specific genes or proteins for medical purposes. The inset shows a nanonet-based detector chip (*black*) on a printed circuit board.





120 degrees C. Nanomix researchers Keith Bradley and Jean-Christophe Gabriel, in collaboration with my U.C.L.A. team, manufactured the first flexible nanotube network transistors on plastic in 2003. Soon afterward, my colleagues and I at U.C.L.A., working with Roth's group at the Max Planck Institute, managed to fabricate devices that were also transparent, making them suitable for applications such as portable visual displays. Physicist John Rogers and his colleagues at the University of Illinois achieved similar success only a few months later. Although these field-effect transistors operated at fast rates—the key metric for such devices—other necessary characteristics, such as low-voltage function, were lacking. The goal was to run the devices at voltages less than those standard batteries can provide to save power, but this feat was attained only recently by Rogers and by chemist Tobin Marks of Northwestern University, who employed specially made polymers to insulate the devices' conductive parts.

Nanonets in Action

CARBON NANONETS CAN OFFER distinct advantages in many portable products, a conclusion that becomes more obvious when one compares them with some of the current contenders for these applications, including films composed of organic or polymeric metals and some semiconductors. For these uses, electronic materials must exhibit good electrical conductance (otherwise, applied current heats them up, resulting in power losses) and high optical transparency (because the viewer of a display, for example, needs to see the layers that lie underneath).

Such substances will enable the development of what people variously call printed, plastic, disposable or macroelectronic products. One example is the photovoltaic cell. Typical solar cells made of single-crystal silicon have excellent performance (they convert as much as 18 percent of incoming light into electricity) but are bulky, heavy and costly to manufacture. Instead imagine a razor-thin solar cell that, though less efficient (converting only 5 or 6 percent of incoming light), is significantly cheaper to fabricate and offers the potential for easy mass production of large-area systems, both of which can compensate for the material's lower performance levels [*see box on page 79*].

In a solar cell, incoming sunlight dislodges electrons and their positively charged counterparts, called holes, in the middle layer of the device. The electrons then migrate to one electrode, power some electrical load and return to the holes via another electrode to complete the circuit. Several companies are working to perfect a cell's active (charge-creating) layers using advanced polymers and other substances that are transparent and flexible. Together with Michael McGehee's materials science group at Stanford University and physical chemist Niyazi Serdar Sariciftci of the University of Linz in Austria, my U.C.L.A. team has produced flexible, proof-of-concept solar cells with nanonet electrodes that exhibit performance comparable to that of indium tin oxide electrodes.

Also under consideration are nanonet-based films that would lie at the heart of an inexpensive, flexible and lightweight touch screen or visual display. A touch screen, for instance, consists of two sheets of electrodes separated by insulating spacers. When a finger touches the top sheet at some point, the electrodes there meet, completing an electrical circuit specific for that location that is formed by a pattern of smooth, thin layers of conductive materials that have been imprinted on the bottom sheet. In collaboration with Richard Kaner's group at U.C.L.A., my team has fabricated and tested proof-of-concept devices based on nanonets.

Nanonets also work in light-emitting diodes, which resemble photovoltaics that run in reverse so that they create

THE AUTHOR

GEORGE GRUNER, a Distinguished Professor of Physics at the University of California, Los Angeles, investigates fundamental issues relating to nanoscale phenomena. He has served as the chief technology officer at Nanomix in Emeryville, Calif., and he is founder of Unidym in Menlo Park, Calif.; both companies concentrate on developing novel applications of nanotechnology. Gruner's interests include skiing, classical music and technology assessment—not necessarily in that order. His research is supported by the National Science Foundation.

JEAN-FRANCOIS PODEVIN; NANOMIX "SENSATION" (*sensor*)

light when electricity passes between the electrodes. In collaboration with Marks's group at Northwestern, my team has recently demonstrated proof-of-concept light-emitting diodes with excellent performance (sufficient to meet the requirements for use in televisions, for example), as has a research group at the University of Montreal led by Richard Martel.

Transistors made from nanonets will, in addition, find use in printed electronics. Tests indicate that the operating speed of carbon nanonets lags somewhat behind that of crystalline silicon, from which most integrated chips are fabricated, but their conductivity and durability advantages over polymers make them attractive to device manufacturers. Although nanotube films cannot yet work in laptop computers or television sets, they are competitive in many other products—especially those that require a material that is cheap, flexible, lightweight, environmentally friendly and resistant to abuse. The first such application is expected to be large-area visual displays, called active-matrix displays. The transistors in a display must run rapidly so that the images can be readily refreshed.

Of course, the kind of portable devices that will use these displays will need power sources as well—cheap, lightweight, razor-thin and disposable batteries and supercapacitors. Nanonets could also play an important role in such power devices, serving not only as electrodes but as high-surface-area components for collecting electric charge to store it for later discharge.

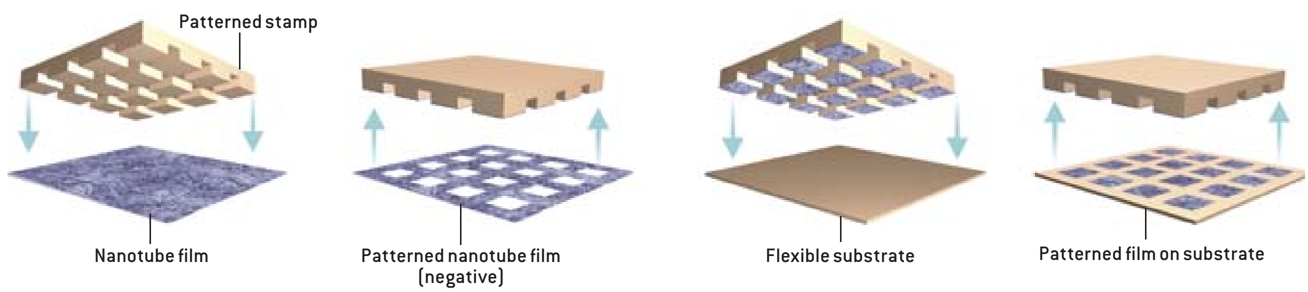
Many Pathways to Take

THE NASCENT CARBON NANONET industry has only just begun to perfect this fledgling technology. There is little doubt that the recent feasibility studies that I have described will soon be followed by working prototypes and eventually products based on those new devices. Today the industry is at the stage where the silicon chip business was half a century ago. The nanotubes are improving steadily, and researchers are successfully sorting those that conduct electricity as well as

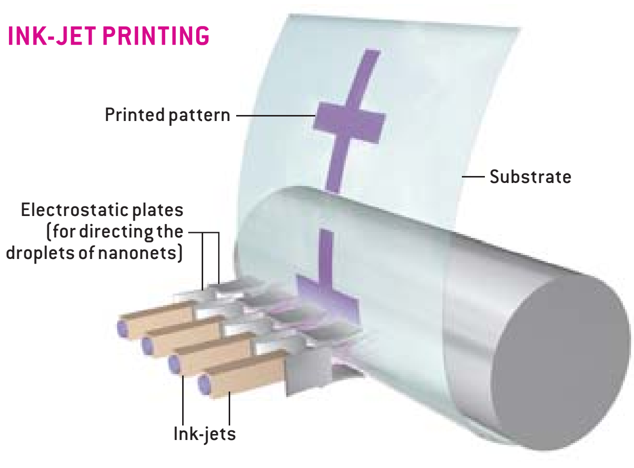
PRINTING A NANONET

Prospective makers of products based on carbon nanonets are developing several inexpensive ways to "print" an engineered pattern of the material onto a flexible polymer surface to produce, for example, an electronic circuit. The simplest method resembles using an ink pad and a stamp (*top*). A patterned stamp comes into contact with a nanonet layer, parts of which stick to the stamp's bottommost surface. The primed stamp presses down on the surface of a substrate, printing the nanonet pattern onto it. Manufacturers are also working on two mass-production techniques, including the use of standard ink-jets (*bottom left*) to spray a liquid containing dispersed nanotubes onto substrates, and a variant of offset printing, in which a nanonet solution substitutes for ink (*bottom right*).

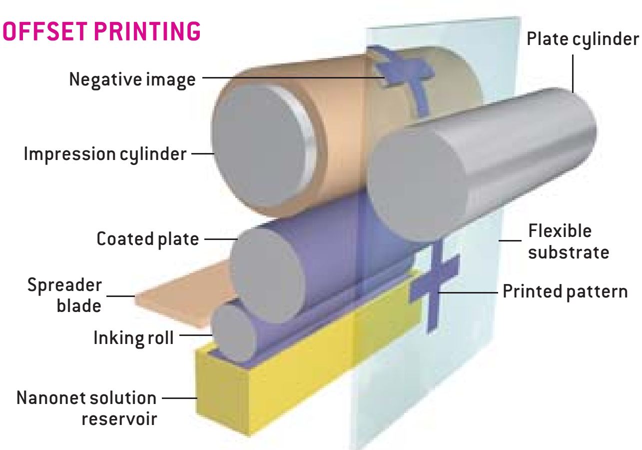
STAMP PRINTING



INK-JET PRINTING



OFFSET PRINTING



JEAN-FRANCOIS PODEVIN

Carbon Nanonet Research and Product Development

Many R&D organizations are producing or working to develop carbon nanotube materials, carbon nanonet films and the electronic devices that incorporate them. A new technology typically passes through the following sequence of developmental stages: concept, R&D, proof of concept, prototype, product development and production.

ORGANIZATION	PRODUCT FOCUS	STATUS
HIGH-GRADE MATERIALS FOR ELECTRONICS		
CarboLex, Lexington, Ky. (www.carbolex.com)	Electric arc– and chemical vapor deposition (CVD)–based fabrication	Production
Carbon Nanotechnologies, Houston (www.cnanotech.com)	CVD- and carbon monoxide–based fabrication	Production
Carbon Solutions, Riverside, Calif. (www.carbonsolution.com)	Electric arc–based fabrication	Production
SouthWest NanoTechnologies, Norman, Okla. (www.swnano.com)	CVD-produced specialty nanotubes	Production
Thomas Swan, Consett, England (www.thomas-swan.co.uk)	High-volume CVD-based fabrication	Production
TRANSPARENT FILMS		
Battelle Memorial Institute, Columbus, Ohio (www.battelle.org)	Transparent coatings	R&D
Eikos, Franklin, Mass. (www.eikos.com)	Conducting ink	Product development
Eastman Kodak, Rochester, N.Y. (www.kodak.com)	Transparent optical coatings	R&D, prototype
Unidym, Menlo Park, Calif. (www.unidym.com)	Films for touch screens, solar cells, light-emitting diodes	Product development
DEVICES		
DuPont, Wilmington, Del. (www.dupont.com)	Transparent electronics	R&D
IBM, Armonk, N.Y. (www.ibm.com)	Computer-compatible transistors and interconnects	R&D
Intel, Santa Clara, Calif. (www.intel.com)	Interconnects	R&D
Motorola, Schaumburg, Ill. (www.motorola.com)	Biological and chemical sensors	Prototype
Nanomix, Emeryville, Calif. (www.nano.com)	Chemical and biological sensors	Product development, R&D
Nantero, Woburn, Mass. (www.nantero.com)	Novel memory technology	Prototype
Samsung, Seoul, South Korea (www.samsung.com)	Displays	R&D
Unidym (*see above*)	Printed electronics for displays	Proof of concept

metals do from those that are semiconductors, which will further better device performance. Meanwhile investigators have made progress on a process that resembles silicon doping, in which special molecules are attached to the tubes to finely alter their electrical properties. Many observers believe that it is only a matter of time before such films exceed the performance of traditional metals and start to make inroads into silicon-based digital electronics technology.

Carbon nanonets have just recently left the realm of science fiction and entered that of practical reality. Like silicon, this budding technology is highly unlikely to lead to artificial life anytime soon, but it has every chance of enabling innovative products that in the not too distant future will improve our everyday lives. SA

MORE TO EXPLORE

Carbon Nanotubes. Mildred Dresselhaus, Gene Dresselhaus, Peter Eklund and Riichiro Saito in *Physics World,* Vol. 11, pages 33–38; January 1998.

Carbon Nanotubes and Related Structures. Peter J. F. Harris. Cambridge University Press, 2001.

Nanotubes for Electronics. Philip G. Collins and Phaedon Avouris in *Scientific American,* Vol. 283, No. 6, pages 62–69; December 2000.

Carbon Nanotubes—The Route Toward Applications. R. H. Baughman, A. A. Zakhidov and W. A. de Heer in *Science,* Vol. 297, pages 787–792; August 2, 2002.

Carbon Nanotube Films for Transparent and Plastic Electronics. George Gruner in *Journal of Materials Chemistry,* Vol. 16, No. 35, pages 3533–3539; 2006.

Carbon Nanotube Transistors for Biosensing Applications. G. Gruner in *Analytical and Bioanalytical Chemistry,* Vol. 384, pages 322–335; 2006.

Eyes Open, Brain

By Steven Laureys

Recent progress in medical care has greatly increased the number of people who survive acute brain damage. Doctors can save the lives of many patients who suffer trauma to the brain (often after a road accident) or a lack of oxygen (for example, after a cardiac arrest or drowning), but if the damage is severe, the victim will slip into a coma. Individuals in this condition do not open their eyes; at best, they will show some reflex movements of the limbs. Coma rarely lasts longer than two to five weeks. Those who regain consciousness typically do so within days. Others will die, and still others will awaken from their coma but remain unconscious, entering what is called the vegetative state.

Even for experts, the vegetative state is a very disturbing condition. It illustrates how the two main components of consciousness can become completely dissociated: wakefulness remains intact, but awareness—encompassing all thoughts and feelings—is abolished. By wakefulness, I mean that patients in a vegetative state have sleep/wake cycles. At the times when they seem to be awake, their eyes open and sometimes wander. At other times they keep their eyes shut and appear to be asleep, although they may open them and stir when touched or spoken to. These patients usually can breathe without technical assistance and can make a variety of spontaneous movements—such as grinding teeth, swallowing, crying, smiling, grasping another's hand, grunting or groaning—but these motions are always reflexive and not the result of purposeful behavior. Typically patients will not fix their eyes on anything for a sustained period, but in rare instances they may briefly follow a moving object or turn fleetingly toward a loud sound.

Shut

New brain-imaging techniques are giving researchers a better understanding of patients in the vegetative state

HALFWAY WORLD: People in the vegetative state are awake without being aware. They can open their eyes and breathe without assistance, but they cannot obey commands or make purposeful movements.

Many people in a vegetative state regain consciousness in the first month after a brain injury. After a month, however, the patient is said to be in a persistent vegetative state (PVS) and the probability of recovery diminishes as more time passes. Every year in the U.S. at least 14,000 new victims of acute brain damage remain vegetative one month after their injury. In 1994 the Multi-Society Task Force on PVS (a team of 11 researchers from various institutions) concluded that the chances of recovery are close to zero if the patient shows no signs of awareness one year after a traumatic brain injury or three months (later revised to six) after brain damage from lack of oxygen or other causes. For these very long-standing cases, the task force proposed the name "permanent vegetative state."

The study of PVS was thrust into the public spotlight in 2005 as politicians debated the case of Terri Schiavo, a Florida woman who had been in a vegetative state since 1990. (Her parents and husband disagreed over whether she could ever recover; the courts ultimately allowed doctors to remove Schiavo's feeding tube, and she died of dehydration 13 days later.) The controversy highlighted the importance of developing more effective ways to determine whether a patient is permanently vegetative or in a more hopeful condition. Scientists have recently studied the use of brain-imaging techniques to identify signs of awareness that could be hidden in an unresponsive patient. If physicians had access to reliable methods of detecting conscious awareness, they might be able to better distinguish between brain-damaged patients who have a chance of recovery and those who face bleaker prospects. At the same time, this research may shine a new light on the nature of consciousness itself.

A Difficult Diagnosis

IN PATIENTS WHO RECOVER from the vegetative state, the first signs of consciousness are often minimal and appear gradually. The patient may start making deliberate, nonreflexive movements but remain unable to express and communicate thoughts and feelings. To classify these cases, doctors have created a clinical category called the minimally conscious state. Like the vegetative state, the minimally conscious state may be a transient condition on the way to further recovery of consciousness, or it may be chronic and sometimes permanent. An important difference, though, is that patients who have remained in the minimally conscious state for years still have a chance of recovery. In a much publicized case, Terry Wallis, an Arkansas man who had been in a minimally conscious state since a road accident in 1984, started talking in 2003. Wallis also regained some ability to move his limbs, although he cannot walk and still needs around-the-clock care.

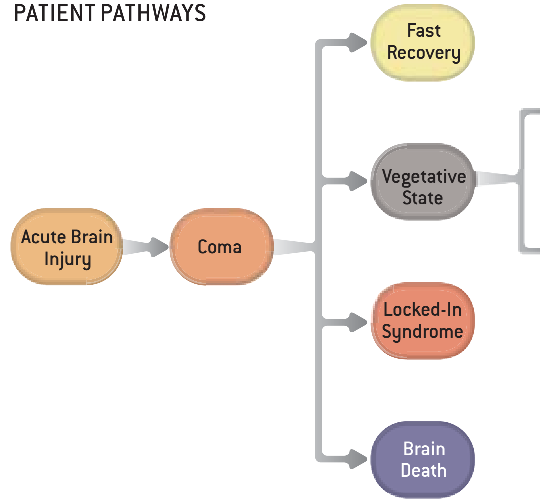
Making the distinction between the vegetative and minimally conscious states is challenging and requires repeated examinations by well-trained physicians who have experience with such patients. Clinicians making the diagnosis of a vegetative state will base their judgment on the absence of behavioral signs of consciousness. Put simply, if a patient looks awake (meaning he or she has his or her eyes open), repeatedly fails to obey commands (such as "pinch my hand" or "look downward") and makes only reflexive movements, the doctor will conclude that the patient is in a vegetative state.

In the early 1990s, however, studies led by Nancy Childs of the Healthcare Rehabilitation Center in Austin, Tex., and Keith Andrews of the Royal Hospital for Neurodisability in London showed that more than one third of the patients initially diagnosed as being in a vegetative state in fact showed some signs of awareness when carefully examined. To obtain a more reliable diagnosis, doctors need to employ standardized clinical tests that gauge a patient's responses to a wide variety of auditory, visual and tactile stimuli. Examples of such tests include the Coma Recovery Scale, developed by Joseph Giacino of the JFK Johnson Rehabilitation Institute in Edison, N.J., and the Sensory Modality Assessment Rehabilitation Technique, created by Helen Gill-Thwaites, also at the Royal Hospital for Neurodisability. The diagnostic superiority of these specialized consciousness scales is beyond doubt, but they are much more time-consuming than a routine neurological examination

AWAKE BUT NOT AWARE

After brain injury that leads to coma, a patient's progress may follow one of several paths (*left*). If the patient does not die or quickly recover, he or she will most likely transition to the vegetative state. (In rare cases, a person may develop locked-

PATIENT PATHWAYS

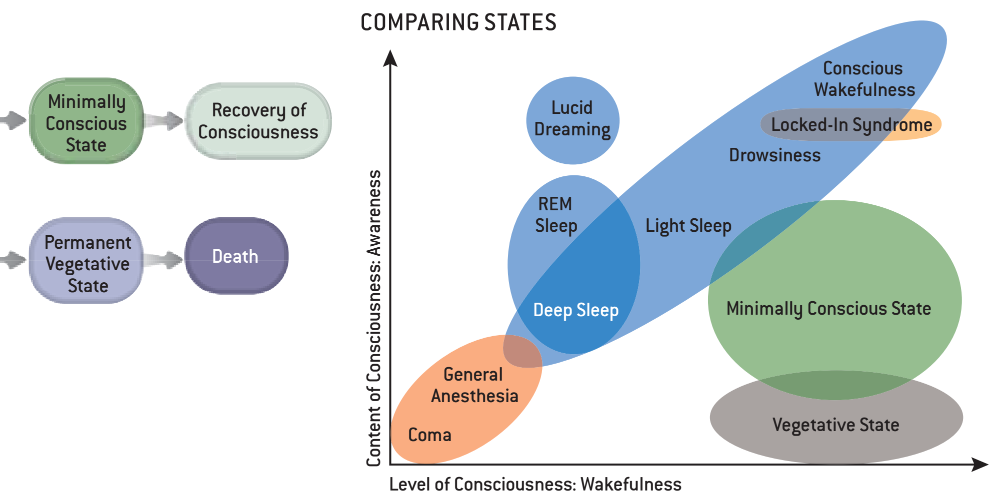


Overview/*The Vegetative State*

- Every year thousands of Americans who have suffered brain injuries transition from coma to a vegetative state. If patients remain in this state for more than a year, their chances of regaining consciousness are close to zero.
- Researchers are trying to develop brain-imaging techniques for diagnosing the vegetative state, which would allow doctors to better determine which brain-damaged patients have a chance of recovery.
- Functional neuroimaging of vegetative patients has revealed new clues to the mechanisms of consciousness, but more research is needed before physicians can use this tool for diagnosis and prognosis.

JEAN-FRANCOIS PODEVIN (*preceding pages*)

in syndrome, a complete paralysis of the body's voluntary muscles.) The patient may then evolve to the minimally conscious state and often to further recovery of consciousness or remain in the vegetative state permanently. Compared with people in other mental states (*right*), vegetative patients have a high level of wakefulness—unlike comatose individuals, they have sleep/wake cycles—but none of the awareness that characterizes normal conscious wakefulness.



or a simpler test such as the Glasgow Coma Scale.

But conscious awareness is a subjective experience that is inherently difficult to measure in another being. Might even the most careful assessment miss signs of awareness in acutely brain-damaged, noncommunicative patients? Over the past decade investigators have struggled to find an objective test that could confirm or dispute a clinical diagnosis of the vegetative state. Structural imaging of the brain—employing either magnetic resonance imaging (MRI) or x-ray computed tomographic (CT) scanning—can help doctors visualize the extent of brain damage, but such techniques cannot detect signs of consciousness. Recent studies, though, indicate that examining MRI images of traumatic brain injuries may help physicians predict whether a patient will emerge from the vegetative state. For example, patients with damage to certain regions, such as the brain stem and the corpus callosum (the band of tissue connecting the cerebral hemispheres), appear to have a poorer chance of recovery.

Furthermore, investigations using a new technique called MRI diffusion tensor imaging—which gauges the integrity of the brain's white matter, the neuronal axons that carry nerve impulses—have increased understanding of the mechanisms underlying recovery from the vegetative state. A team led by Nicholas Schiff of Cornell University, for instance, recently used diffusion tensor imaging to show the regrowth of axons in the brain of Wallis, the patient who emerged from a minimally conscious state after 19 years of silence.

Another widely used tool is the electroencephalogram (EEG), which measures the brain's electrical activity. EEG results can reveal a patient's state of wakefulness because electrical activity slows during nondreaming sleep. For patients in a coma, the instrument can confirm the clinical diagnosis of brain death (when the EEG becomes isoelectric—that is, a flat line). The EEG is not as reliable, though, for measuring changes in awareness. For patients in the vegetative state, the device can neither confirm their diagnoses nor predict their chances of recovery. My colleagues and I at the University of Liège in Belgium have demonstrated that minimally conscious patients show an electrical brain response called the P300 potential when hearing their own names, but not when hearing other first names. Some patients in a chronic vegetative state, however, demonstrated similar P300 responses, so the technique appears to have no diagnostic use.

A Consciousness Region?

PERHAPS THE MOST promising method for investigating the vegetative state is functional neuroimaging. Studies employing positron-emission tomography (PET) have shown that the metabolic activity of the brain—as measured by its consumption of glucose—decreases in the vegetative state to less than half of normal values. These experiments were first performed in the late 1980s by a group led by Fred Plum of Cornell and later repeatedly confirmed by several European groups, including ours. But some patients who recover from the vegetative state do so without substantial changes in overall brain metabolism, as reported by our laboratory in the late 1990s.

Moreover, we observed that some fully conscious and healthy volunteers have global brain metabolism values comparable to those seen in some patients in a vegetative state. And Schiff has reported that a few vegetative patients had a close to normal cortical metabolism. Hence, measuring overall levels of energy consumption in the brain cannot indicate the presence of awareness.

Our group could, however, identify areas in the brain that appear to be particularly important for the emergence of awareness. Comparing patients in a vegetative state with a large cohort of healthy

THE AUTHOR

STEVEN LAUREYS leads the coma research group at the Cyclotron Research Center at the University of Liège in Belgium and is head of clinics at the department of neurology at Sart Tilman Liège University Hospital. He obtained his M.D. from the Free University of Brussels (Flemish) in 1993 and his Ph.D. from the University of Liège in 2000. His work is supported by the Belgian National Fund for Scientific Research, the European Commission and the Mind Science Foundation. Laureys recently published *The Boundaries of Consciousness: Neurobiology and Neuropathology* (Elsevier, 2006).

MELISSA THOMAS

volunteers, we found a significant lack of metabolic activity in the widespread network of polymodal associative cortices (located in the frontal and parietal lobes of the brain) that are involved in the cognitive processing of sensory information [*see box below*]. We showed that awareness is also related to the so-called cross talk, or functional connectivity, within this frontoparietal network and with centers deeper in the brain, notably the thalamus. In our vegetative patients, the long-distance connections between one cortex and another, as well as the links between the cortices and the thalamus, appeared to be disrupted. Moreover, recovery from the vegetative state is paralleled by a functional restoration of the frontoparietal network and its connections.

Unfortunately, patients in a minimally conscious state seem to show a rather similar cerebral dysfunction. As a result, PET measurements of brain metabolism cannot distinguish between the vegetative and minimally conscious states when the patient is at rest. The technique did reveal differences, though, when analyzing changes in brain function induced by external stimuli such as pain and spoken words.

We studied pain perception by electrically stimulating the hand (which was experienced as painful by healthy control subjects) and using PET to gauge cerebral blood flow, which is another marker for neural activity. In both the vegetative patients and the controls, the results indicated activity in the brain stem, the thalamus and the primary somatosensory cortex, which receives sensory information from the peripheral nerves. In the vegetative patients, however, the rest of the brain failed to respond. The small cortical region that did show activity (the primary somatosensory cortex) was isolated and disconnected from the rest of the brain, in particular from the networks believed to be critical for conscious pain perception. (These results can reassure family members and caregivers that patients in a vegetative state do not perceive pain in the same way that healthy people do.)

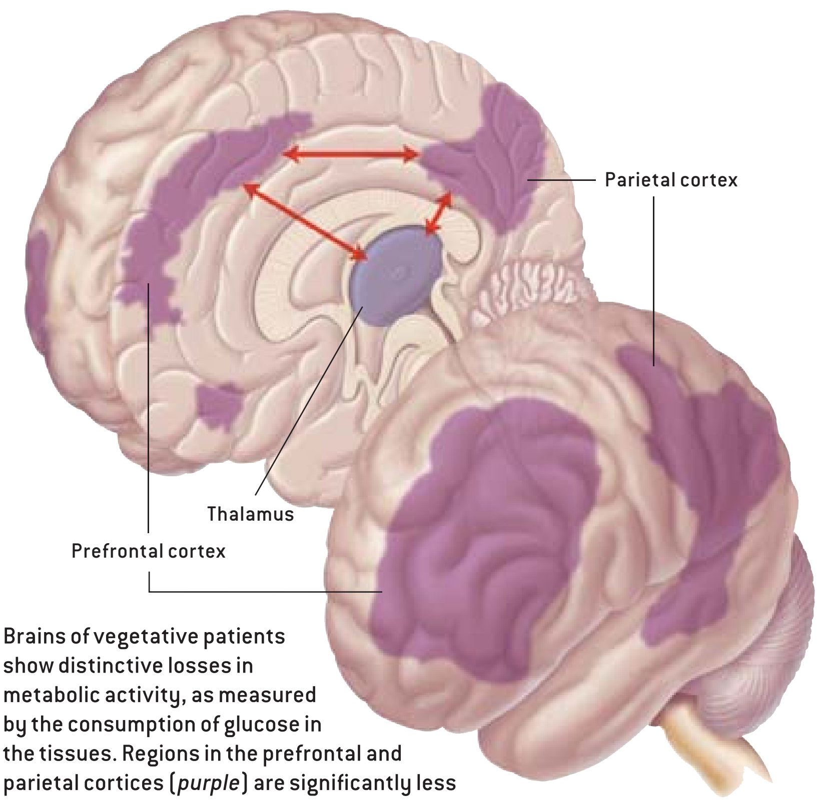
The PET studies found a similar pattern when we talked to vegetative patients. Just as for somatosensory stimulation, the activity was limited to the lowest cortical centers (in this case, the primary auditory cortices) while higher-order polymodal areas failed to respond and remained functionally disconnected. This level of cerebral processing is considered insufficient for auditory awareness. In minimally conscious patients, however, auditory stimuli can trigger large-scale, higher-order cortical activity normally not observed in the vegetative state. Schiff was the first to use functional MRI (fMRI) on patients in the minimally conscious state and showed that their language networks were activated during the presentation of a personally meaningful story read by a familiar voice. When played backward, the story did not provoke such a response, whereas it did so in healthy control subjects.

Similarly, in 2004 our group reported that auditory stimuli with emotional power (such as infant cries and the patient's own name) induce a much more widespread activation in minimally conscious patients than meaningless noise does. These results indicate that content does matter when talking to a minimally conscious patient. But to develop this technique into a diagnostic tool, we knew we had to show that complex auditory stimuli never activate large-scale networks in vegetative patients.

Tennis in the Brain

THIS HYPOTHESIS FACED its toughest test last year when a team led by Adrian Owen of the University of Cambridge, in collaboration with Melanie Boly from our group, studied a 23-year-old woman who had suffered a traumatic injury to the front of her brain in a traffic accident. She remained comatose for more than a week and then evolved to a vegetative state. She spontaneously opened her eyes but never responded

VIEWING THE DAMAGE

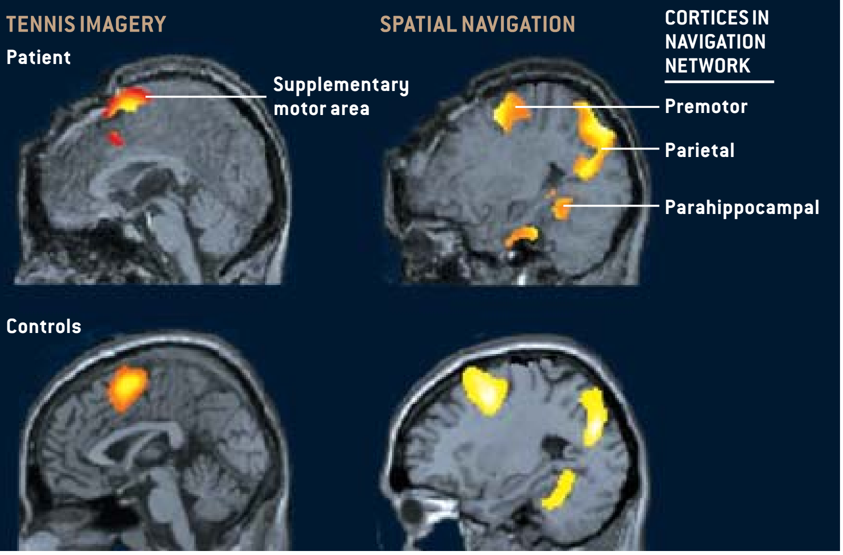


Brains of vegetative patients show distinctive losses in metabolic activity, as measured by the consumption of glucose in the tissues. Regions in the prefrontal and parietal cortices (*purple*) are significantly less active in vegetative patients. The dysfunction could be a result of cortical damage or disruption of the links (*red arrows*) between the cortices and the thalamus. These connections appear to be crucial to conscious awareness.

ALICE Y. CHEN

SIGNS OF RECOVERY

Functional neuroimaging of a 23-year-old brain-damaged woman (*top images*) showed activity in a motor area when she was asked to imagine playing tennis and activation of a spatial navigation network when she was asked to envision walking through her house. Similar responses were seen in healthy control subjects (*bottom images*), indicating that the woman was beginning the transition from the vegetative to the minimally conscious state.



to any verbal or nonverbal commands.

Five months after the accident Owen and his colleagues studied the woman's brain using fMRI. During the scanning, the researchers played recordings of spoken sentences—for example, "There was milk and sugar in his coffee"—along with noise sequences that acoustically matched those sentences. The spoken sentences triggered activity in her superior and middle temporal gyri, brain regions involved in understanding speech and the meaning of words; the same pattern was observed in healthy control subjects. These results possibly indicated conscious linguistic processing in the vegetative woman, but not necessarily. Studies of healthy patients have shown that such processing may also occur during sleep and even under general anesthesia.

To clarify whether the patient was consciously responding to language, the investigators conducted a second study in which they asked her to perform mental imagery tasks. When she was asked to imagine playing a game of tennis, the fMRI scans showed activity in the supplementary motor area of her brain, just as it did in the control subjects. When she was asked to imagine walking through the rooms of her house, the scans showed activation of the network involved in spatial navigation—the premotor, parietal and parahippocampal cortices. Again, the response was indistinguishable from that seen in the healthy subjects. Despite the clinical diagnosis that the patient was in a vegetative state, she understood the tasks and repeatedly performed them and hence must have been conscious.

The first question raised by these spectacular results was whether the patient was misdiagnosed. Although repeated expert assessments had confirmed that she was in the vegetative state at the time of the study, the examinations revealed that her eyes briefly fixed on objects. This finding is sometimes observed in the vegetative state, but it is atypical and should prompt physicians to search for other signs of awareness. During another examination about six months after the study, the patient was able to fix her eyes on an object for a sustained period (more than five seconds) and could track her own mirror image; both these signs herald a transition to the minimally conscious state. Currently the patient is still minimally conscious, sometimes obeying commands but not communicating.

Given her young age and the cause and duration of the vegetative state, we knew from the start that the patient's chances of recovery were not zero but about one in five. Thus, the results of the study should not be misinterpreted as evidence that all patients in a chronic vegetative state may actually be conscious. In fact, we have not observed any similar signs of awareness in functional scans of more than 60 other vegetative patients studied at the University of Liège. The most likely explanation of the results is that our 23-year-old patient was already beginning the transition to the minimally conscious state at the time of the experiment. Indeed, a recent study by Di Haibo of Zhejiang University in China and his colleagues confirmed that the activation of higher-level brain regions during fMRI tests can predict recovery to the minimally conscious state.

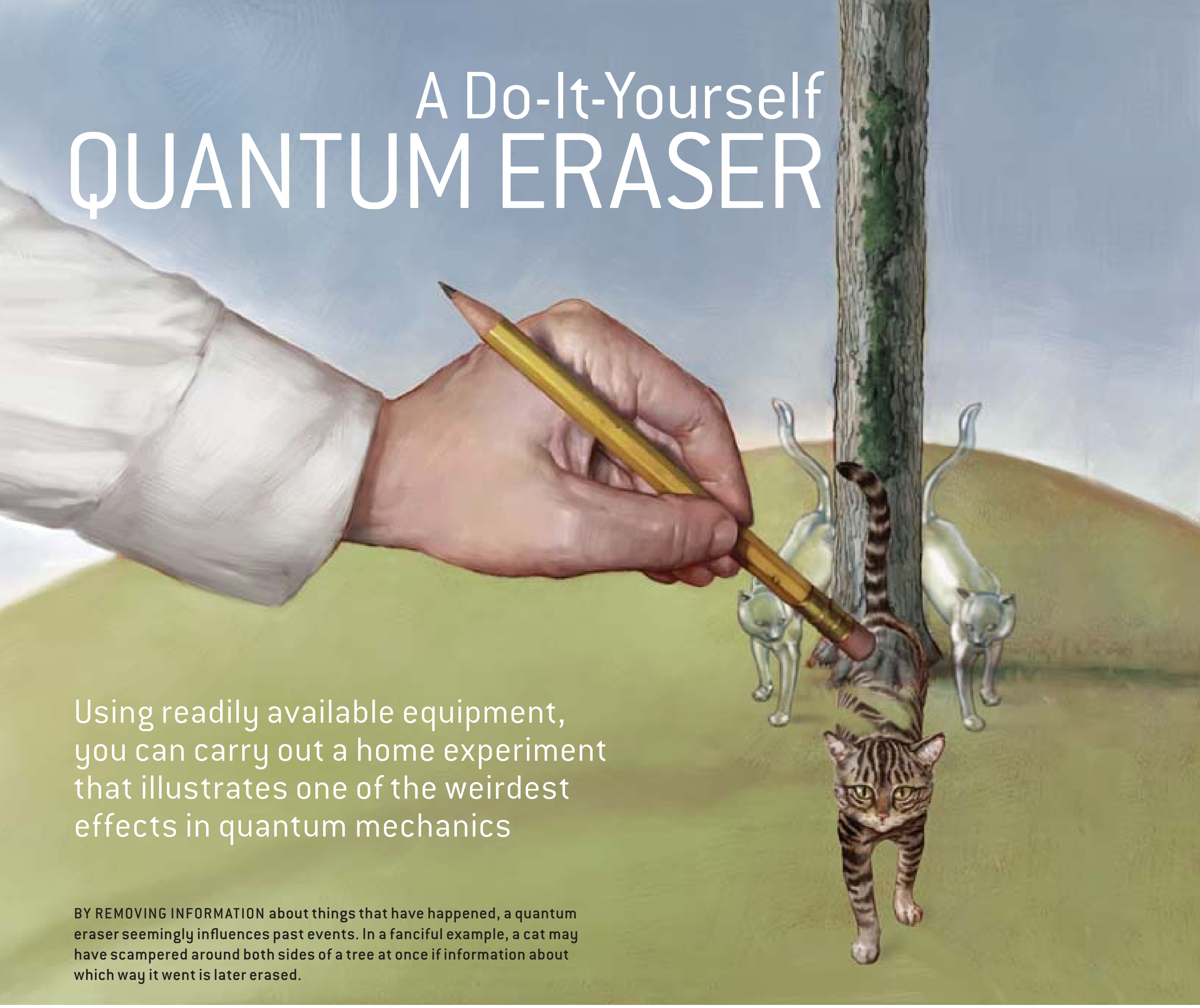
These findings underline the point that ascertaining consciousness is a tricky business. We have learned much from new imaging techniques that measure neural activity in brain-damaged patients, but more research is needed before scientists can use functional neuroimaging to confirm a diagnosis of the vegetative state and to help in the prognosis and treatment of this devastating medical condition. For the time being, doctors must continue to rely on thorough clinical examinations when making their difficult therapeutic decisions. SA

MORE TO EXPLORE

The Vegetative State: Medical Facts, Ethical and Legal Dilemmas. B. Jennett. Cambridge University Press, 2002.

Science and Society: Death, Unconsciousness and the Brain. Steven Laureys in *Nature Reviews Neuroscience,* Vol. 6, No. 11, pages 899–909; November 2005.

Detecting Awareness in the Vegetative State. A. M. Owen, M. R. Coleman, M. Boly, M. H. Davis, S. Laureys and J. D. Pickard in *Science,* Vol. 313, page 1402; September 8, 2006.



A Do-It-Yourself QUANTUM ERASER

Using readily available equipment, you can carry out a home experiment that illustrates one of the weirdest effects in quantum mechanics

BY REMOVING INFORMATION about things that have happened, a quantum eraser seemingly influences past events. In a fanciful example, a cat may have scampered around both sides of a tree at once if information about which way it went is later erased.

MATT COLLINS

BY RACHEL HILLMER AND PAUL KWIAT

Notoriously, the theory of quantum mechanics reveals a fundamental weirdness in the way the world works. Commonsense notions at the very heart of our everyday perceptions of reality turn out to be violated: contradictory alternatives can coexist, such as an object following two different paths at the same time; objects do not simultaneously have precise positions and velocities; and the properties of objects and events we observe can be subject to an ineradicable randomness that has nothing to do with the imperfection of our tools or our eyesight.

Gone is the reliable world in which atoms and other particles travel around like well-behaved billiard balls on the green baize of reality. Instead they behave (sometimes) like waves, becoming dispersed over a region and capable of crisscrossing to form interference patterns.

Yet all this strangeness still seems remote from ordinary life. Quantum effects are most evident when tiny systems are involved, such as electrons held within the confines of an atom. You might know in the abstract that quantum phenomena underlie most modern technologies and that various quantum oddities can be demonstrated in laboratories, but the only way to see them in the home is on science shows on television. Right? Not quite.

On pages 92 and 93, we will show you how to set up an experiment that illustrates what is known as quantum erasure. This effect involves one of the oddest features of quantum mechanics—the ability to take actions that change our basic interpretation of what happened in past events.

Before we explain what we mean by that and outline the experiment itself, we do have to emphasize one caveat in the

interest of truth in advertising. The light patterns that you will see if you conduct the experiment successfully *can* be accounted for by considering the light to be a classical wave, with no quantum mechanics involved. So in that respect the experiment is a cheat and falls short of fully demonstrating the quantum nature of the effect.

Nevertheless, the individual photons that make up the light wave are indeed doing the full quantum dance with all its weirdness intact, although you could only truly prove that by sending the photons through the apparatus and detecting them one at a time. Such a procedure, unfortunately, remains beyond the average home experimenter. Still, by observing the patterns in your experiment and by thinking about what they mean in terms of the individual photons, you can get a firsthand glimpse of the bizarre quantum world.

If you want to go straight to the home experiment, it is detailed on the next two pages. The discussion that follows here (and continues on page 94) delves into the science of quantum erasers in general. This explanation will help you understand what the do-it-yourself eraser demonstrates, but you might want to come back to it after seeing what that specific kind of eraser does.

What a Quantum Eraser Erases

ONE OF THE STRANGE FEATURES of quantum mechanics is that the behavior that something exhibits can depend on what we try to find out about it. Thus, an electron can behave like a particle or like a wave, depending on which experimental setup we subject it to. For example, in some situations particlelike behavior emerges if we ascertain the specific trajectory that an electron has followed and wavelike behavior transpires if we do not.

A standard demonstration of this duality relies on what is called a two-slit experiment (your do-it-yourself quantum eraser is similar to this experiment in that it involves two pathways, but not two slits). A source emits particles, such as electrons, toward a screen that has two slits they can pass through. The particles ultimately arrive at a second screen where each one produces a spot. Where each particle lands is to some extent random and unpredictable, but as thousands of them accumulate, the spots build up into a definite, predictable pattern. When the conditions are right for the particles to behave as waves, the result is an interference pattern—in this case a series of fuzzy bars, called fringes, where most of the particles land, with very few hitting the gaps between them.

The particles will generate the interference pattern only if each particle could have traveled through either of the two slits, and there is no way of ascertaining which slit each one passed through. The two pathways are then said to be indistinguishable and each particle acts as if it actually traveled through both slits. According to the modern understanding of quantum mechanics, interference occurs when indistinguishable alternatives are combined in this way.

When two or more alternatives coexist, the situation is called a superposition. Erwin Schrödinger highlighted the oddity of quantum superpositions in 1935, when he proposed his now infamous concept of a cat that is simultaneously alive and dead, sealed inside a hermetic box where it cannot be observed. When quantum interference happens, something in the experiment is like a kind of Schrödinger's cat. But instead of being alive and dead at the same time, the cat may be walking by a tree, passing on both sides of it simultaneously.

Schrödinger's cat ceases to be in a superposition as soon as we look inside its box: we always see it to be either alive or dead, not both (although some interpretations of quantum mechanics have it that *we* become in a superposition of having seen a dead or a live cat). If a spotlight is shining near the tree, we see the quantum cat go one way or the other. Similarly, we can add a measurement tool to watch each particle as it passes the slits. One could imagine having a light shining on the slits so that as each particle comes through we can see a flash of light scatter from where the particle went. The flash makes the two alternative pathways distinguishable, which destroys the superposition, and the particles arrive at the final screen not in a pattern of fringes but in one featureless blob. Experiments analogous to this scenario have been conducted, and, as predicted by quantum mechanics, no interference pattern builds up.

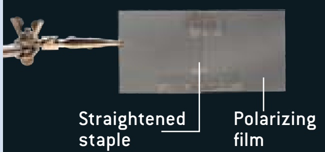
We need not actually "do the looking." We do not have to

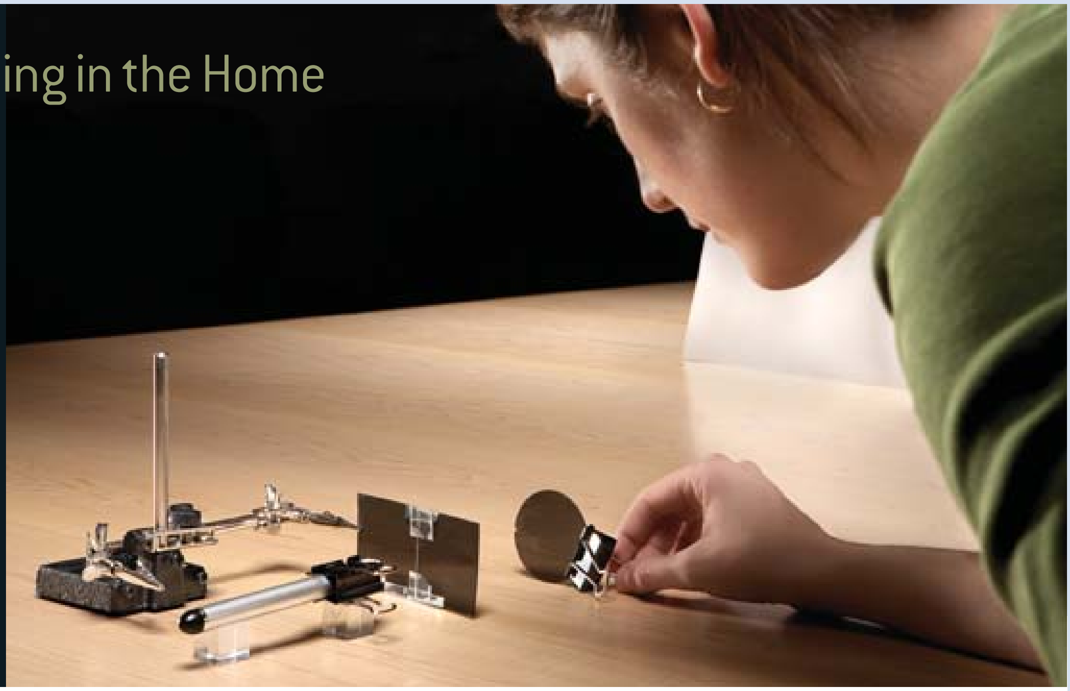
What you will need for the experiment

- **A very dark room.**
- **Polarizing film.** Plain gray, high-quality film ("experimental grade") gives the best results; avoid film tinted with a color (see www.sciam.com/ontheweb for some places that sell film). You need to cut it into six squares, each about two inches on a side. The box on page 94 describes what polarizers do to photons.
- **A laser,** such as a laser pointer. If yours emits polarized light, align its polarization at 45 degrees from the vertical. If your laser is not polarized, include a polarizer at 45 degrees immediately after the laser at every step. Use a rubber band to keep the laser turned on.
- **A thin, straight piece of wire,** such as from an unused twist tie or a straightened staple. The thinner the better.
- **Some tinfoil and a pin** to poke a hole in it. The light that goes through the pinhole will expand outward, forming a narrow, conical beam. The pinhole makes the patterns dimmer but may improve the results if the room is dark enough.
- **Some stands** to hold the laser and polarizers in place. These could be as low-tech as cereal boxes.
- **A screen** to display the final patterns. The bare wall will do if it is plain enough; otherwise use a sheet of paper.

Quantum Erasing in the Home

The steps presented here outline how to see quantum erasure in action. See www.sciam.com/ontheweb for a fuller description and additional information, such as the basics of how waves interfere and produce fringes.

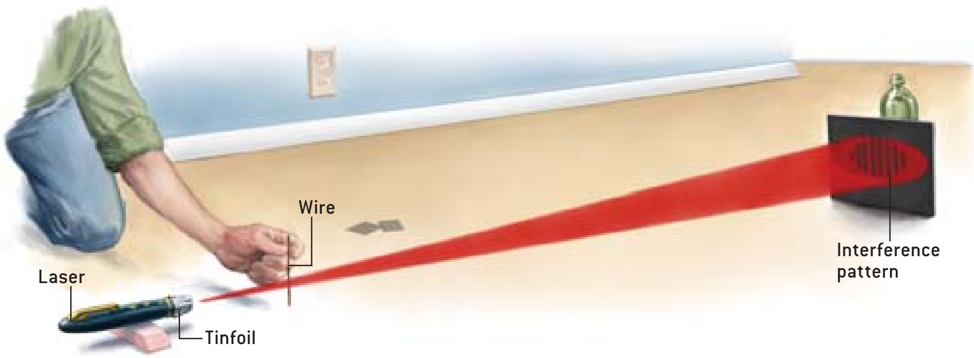


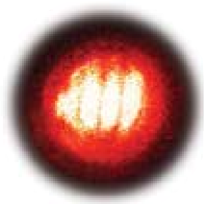


1 SEEING INTERFERENCE

- Wrap the tinfoil around the business end of the laser and put a pinhole in it to let through some of the light beam.
- Set up the laser so it shines on the screen from at least six feet away. It should produce a circular spot of light on the screen.
- Position the wire vertically and centered in the light.

WHAT HAPPENS: As shown, you should see an interference pattern consisting of a row of fringes (bright and dark bands). The interference pattern arises because light passing on the left of the wire is combining, or "interfering," with light passing on the right-hand side. If you hold a piece of paper just after the wire, you will see a lobe of light on each side of the shadow of the wire. The lobes expand and largely overlap by the time they reach the screen. For each individual photon arriving at the screen in the overlap region, it is impossible to tell whether it went on the left or the right side of the wire, and the combination of the two ways it went causes the fringes. Although you are looking at trillions of photons, each of them is interfering only with itself.



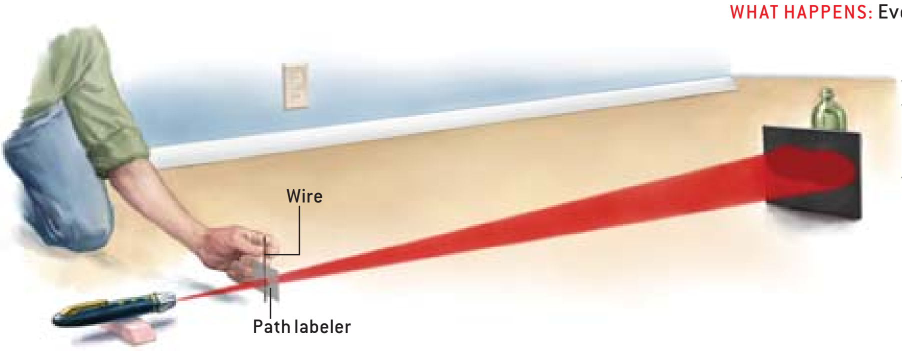


INTERFERENCE seen is captured in this photograph. The size and other features of the patterns depicted in the diagrams are exaggerated.

2 LABELING THE PATH

- Take two polarizers and rotate one of them so that their axes are perpendicular; you have done this correctly if when you overlap the film temporarily, no light goes through the overlap region.
- Tape them together side by side with no gap or overlap. Do the taping along the top and bottom so the tape will not block the light. We will call this the path labeler.
- Position the labeler in the beam so that its join is right behind the wire. Attaching the wire to the labeler might be easiest. Wire and labeler will not be moving for the rest of the experiment. We will say that the left-hand polarizer produces vertically polarized light (V), and the right-hand one horizontally polarized (H). It does not matter if we have these labels reversed.

WHAT HAPPENS: Even though the light is again passing on both sides of the wire, the fringes should be gone. If a photon reaches the screen by passing to the left of the wire, it arrives V-polarized; if to the right of the wire, H-polarized. Thus, the labeler has made available the information about which way each photon went, which prevents the interference.

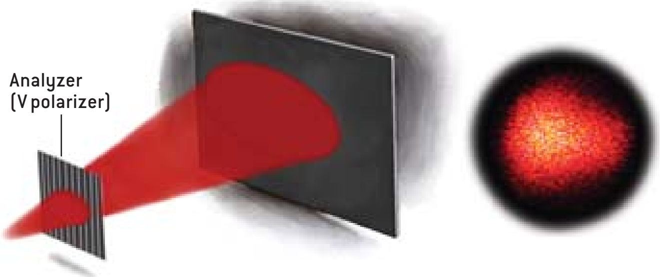


RACHEL HILLMER (*photographs of light patterns*); BRIAN MARANAN PINEDA (*laser setup*); MATT COLLINS (*drawings of setups and cat*)

3 SELECTING THE LEFT-PASSING PHOTONS

- Position a third polarizer (the "analyzer") between the labeler and the screen in the V orientation.

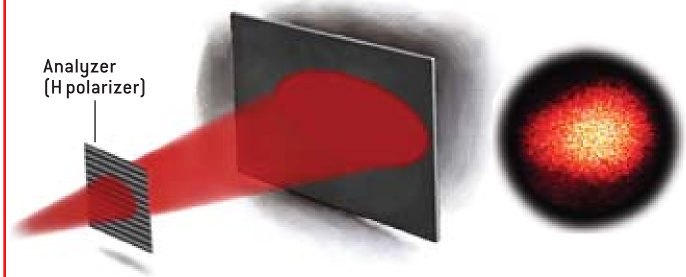
WHAT HAPPENS: The analyzer will block all the right-passing photons (which became H-polarized at the labeler) and will let through all the left-passing ones. The pattern will be nearly the same as in the previous step—just dimmer and not extending quite so far on the right, because it is only the left lobe of light. With the analyzer, you are accessing the information that the labeler made available: you know that all the photons hitting the screen passed to the left of the wire.



4 SELECTING THE RIGHT-PASSING PHOTONS

- Put the analyzer in the H orientation.

WHAT HAPPENS: The H analyzer blocks the left-hand lobe of light and lets through only the right-hand lobe. If you could measure intensities of light (or numbers of photons) at the screen, you would find that the light in step 2 was just the sum of the light in steps 3 and 4. Notice that the fringes were missing from step 2 even though you were not ascertaining the polarization of the photons; it was enough that you could have done so, as in steps 3 and 4.



5 ERASING THE PATH INFORMATION

- Rotate the polarizer 45 degrees clockwise from V, an orientation we call diagonal (D).

WHAT HAPPENS: The fringes reappear! Why? The polarizer is erasing the information about which side each photon used. Now each left-passing V photon has a 50 percent chance of getting through it to the screen, as does each right-passing H photon. In both cases, the photons that get through become D-polarized, so there is no way to tell which way each photon went. Once again, each photon apparently goes both ways at once and interferes with itself.

6 THE ANTI-ERASER

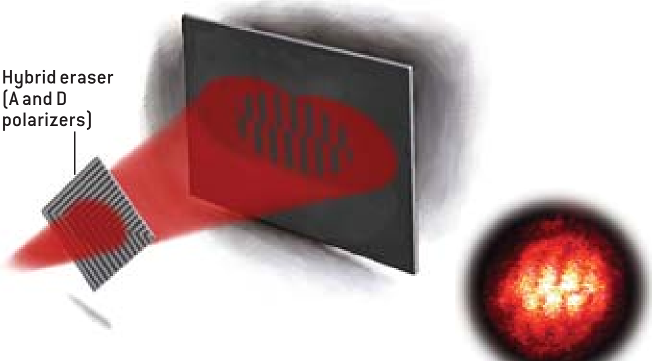
- Rotate the polarizer 45 degrees counterclockwise from V ("antidiagonal" or "A").

WHAT HAPPENS: Again there are fringes—everything said in step 5 applies to an A-polarized eraser as well. But if you look very closely, you will see that the fringes are shifted slightly in the two cases. The A fringes are bright where the D ones are dark, and vice versa. If you could add up the intensities, or numbers of photons, for the D and A erasers, the sum would again be the shape from step 2, with no interference visible.

7 BOTH ERASERS AT ONCE

- Cut in half horizontally a D-oriented and an A-oriented polarizer.
- Join the top half of the D with the bottom half of the A.
- Put the hybrid analyzer in place.

WHAT HAPPENS: D fringes appear in the top half of the light and A fringes in the bottom half. The pattern looks a bit like misaligned teeth and makes clearer how the dark and bright fringes of each eraser correspond.



CONCLUSION

Think about what the photons were doing in each of the steps.

- In some steps (*3* and *4*), each photon went on one side or the other of the wire (no interference), but in others (*1, 5, 6* and *7*), they seemingly went on both sides at once (producing interference).
- Our interpretation of what the photons did at the wire depends on what they encountered *later on* in the setup—be it an analyzer or an eraser or nothing but the screen.
- Steps 6 and 7 revealed that the "which way?" information can be erased in more than one way, to produce either the original interference pattern or the inverse of it.

What polarizers do to photons

Polarizing film has an axis (in our diagrams we depict its direction with lines on the film), and the film allows passage of light that is oscillating parallel to the axis. You can think of light as being like a wave on a rope held between two people; the wave can make the rope move up and down or side to side or at any angle in between. The angle of the oscillation is the polarization of the wave.

Polarizing film is like a screen of parallel bars that the rope passes through: it lets through waves polarized parallel to it unhindered, blocks perpendicular ones completely and allows waves on other angles to get through with reduced amplitude. Most important, the wave (if any) that comes out the other side of a polarizer is polarized parallel with the polarizer's transmission axis.

The quantum description of what happens to light going through a polarizing film sounds only slightly different: The light is made up of individual particles called photons, and like a wave, the photons can each have a direction of oscillation. A photon will get through every time when it hits a polarizer with the transmission axis parallel to the photon's polarization. A perpendicular polarizer blocks the photon every time. At a 45-degree angle, the photon has a 50 percent chance of getting through (the exact probability varies as the angle is varied). Most important, when a photon does go through a polarizer, on the other side it will be polarized parallel with the polarizer's transmission axis.

Light can also be unpolarized, which means the photons making up the light have random polarizations. That is another case in which half the photons will get through a polarizer, and, as always, those that do so become polarized parallel with the polarizer.

You can see how polarizers work by putting two of them together. As you rotate one of the polarizers, you can see through them clearly when their axes are aligned, barely at all when they are perpendicular and to some extent at other angles. Photons that make it through the first polarizer are polarized by it, and then their probability of getting through the second one depends on the angle between their polarization and the second polarizer's axis.

An interesting effect happens if two polarizers are perpendicular and a third one is inserted between them on an angle (45 degrees is best): adding the third polarizer allows some light to get through, even though you might expect it to be an additional obstacle for the light. See if you can explain why that happens (the answer is at www.sciam.com/ontheweb). The do-it-yourself quantum eraser also relies on a polarizer at 45 degrees changing what the light does.

detect the light flashes and ascertain which way each particle went. It suffices that the information is available in the flashes and *could* have been observed in that way.

Now we finally get to the quantum eraser. The eraser is something that can erase the information indicating which path each particle has followed, thereby restoring the indistinguishability of the alternatives and restoring interference.

How might an eraser do that? Imagine that the "flash of light" that scatters from each particle is a single photon. For the photon to reveal the "which path?" information of the particle, it must be possible (even if only in principle) to tell which slit the photon came from. That means we must be able to measure the position of where each photon scattered accurately enough to tell the slits apart. Heisenberg's uncertainty principle, however, tells us that if we instead measure the momentum of each photon with great accuracy, then the photons' positions become less well defined. So if we pass the photons through a lens that makes their momentum information available, the information about their positions is erased. When that happens, the two paths the particles can follow are again indistinguishable and interference is restored.

THE AUTHORS

RACHEL HILLMER and *PAUL KWIAT* are both at the University of Illinois, where Hillmer is an undergraduate student in Kwiat's laboratory. Kwiat is Bardeen Chair in Physics, and his research includes the phenomena of quantum interrogation and quantum erasure, optical implementations of quantum information protocols and nonquantum studies of swing dancing. Hillmer is working on novel ways to encode quantum information in light.

We have omitted one last tricky detail, but we will come back to that. First, stop and think a bit more about what is happening in the erasing process we just described, because that is where the weirdness lies. When we detect the position where one of the photons scattered, we learn which slit its corresponding particle went through, which means the particle *did* go through one slit or the other, not both. If we instead detect the photon's momentum, however, we cannot know which slit the particle went through. What is more, when we do many momentum measurements and see an interference pattern, we infer that in those cases the particles went through both slits (interference would be impossible otherwise).

In other words, the answer to the question, "Did the particle go through one slit or both slits?" depends on what we do with its corresponding photon long *after* the particle has gone through. It is almost as if our actions with the photons influence what has happened in past events. We can find out which slit the particle went through, or with our quantum eraser we can delete that information from the universe.

Strangest of all, we can decide which measurement to make *after* the particle has passed through the slits—we can have the apparatus for both alternative measurements in place, with a switch that we flick one way or the other just before each pho-

BRIAN MARANAN PINEDA

ton arrives. Physicists call this variation a delayed-choice experiment, an idea introduced by John A. Wheeler of the University of Texas at Austin in 1978 that extends a scenario that Niels Bohr and Albert Einstein used in their arguments about quantum mechanics and the nature of reality in 1935.

At this point, some particularly clever readers will be worrying about a fundamental problem that seems to undermine what we have just described: Why can't we delay the choice of our photon measurement until after we have seen if the particles form an interference pattern? We could, in fact, arrange to do just that by having the final screen not too far from the slits and the photon detector much farther away. So what would happen if we saw the particles form fringes but then chose to do photon position measurements that should prevent such fringes from forming? Wouldn't we have created a paradox? Surely we would not expect the already registered interference pattern to vanish! Similar reasoning suggests we could use the delayed-choice effect to transmit messages instantaneously over arbitrary distances (thereby circumventing the speed of light).

That tricky detail that we omitted earlier is what saves the day: to see the interference of the particles after applying the quantum eraser, we first have to divide them into two groups and observe the groups separately. One group will display the original pattern of fringes; the other will display the inverse of that pattern, with particles landing on what were originally the dark bands and avoiding the places where the bright fringes were. The two groups combined fill in all the gaps, hiding the interference.

The paradox is avoided because we need data from the photon measurement to know which group each particle belongs to. Thus, we cannot observe the fringes until after we have done the photon measurements, because only then do we know how to split the particles into groups. In the home experiment, dividing particles into groups is done for you automatically because one group gets blocked by a polarizing filter, and you can therefore see the interference pattern of the group that gets through with your own eyes. In the final step you can see the interference patterns of the two groups right next to each other.

From a practical standpoint, the inability to send messages faster than the speed of light and create a paradox is perhaps disappointing, but physicists and logicians consider it to be a very good feature. SA

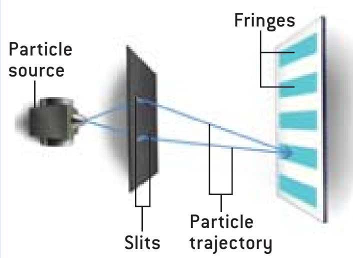
For more discussion about quantum erasers, go to ***www.sciam.com/ontheweb***, where you will find:

- A list of cutting-edge interference and quantum eraser experiments carried out in recent years.
- A short discussion of what quantum erasers have to do with how the ordinary world we are familiar with emerges from the weird underlying quantum reality.
- More information about delayed-choice experiments and the impossibility of superluminal messages.
- A few other related experiments you can do at home.

How a Quantum Eraser Works

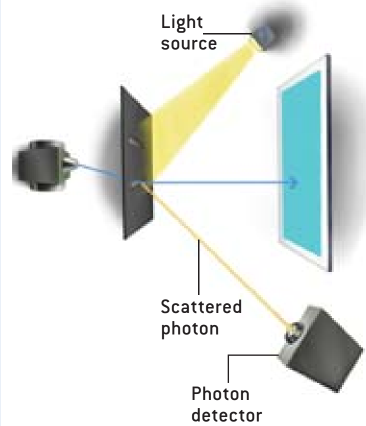
How quantum particles behave can depend on what information about them can possibly be accessed. A quantum eraser eliminates some information and thereby restores the phenomenon of interference. The eraser's action is most easily understood by considering a "double-slit" experiment (*below*).

CREATING QUANTUM INTERFERENCE



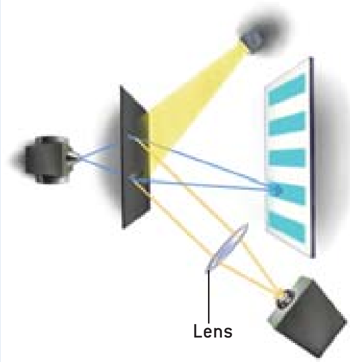
Particles sent through two slits generate bands (called fringes) on a detector screen when large numbers arrive at some regions (*blue*) and very few arrive at other regions (*white*). This interference pattern arises only if each particle could have traveled through both slits to arrive at the screen (*arrows*).

PREVENTING INTERFERENCE



The fringes do not appear if the particles interact with something that could thereby be used to ascertain each particle's location at the slits. For example, a photon of light (*yellow line*) might scatter from the particle and reveal that it went through the right-hand slit. The photon need not be detected—all that matters is that the "which slit?" information in principle could be determined if it were to be detected.

ERASER RESTORES INTERFERENCE



A quantum eraser erases the "which slit?" information. If the particle scatters a photon, a lens could make it impossible to ascertain which slit the photon came from. In that case, the corresponding particle apparently goes through both slits, as before, and fringes can be observed. The strangest feature of this quantum erasing is that the behavior of the particle at the slits seemingly depends on what the photon encounters *after* the particle has passed through the slit(s).

MATT COLLINS

WORKINGKNOWLEDGE

DIRECTIONAL SOUND

Psst . . . Hey, You

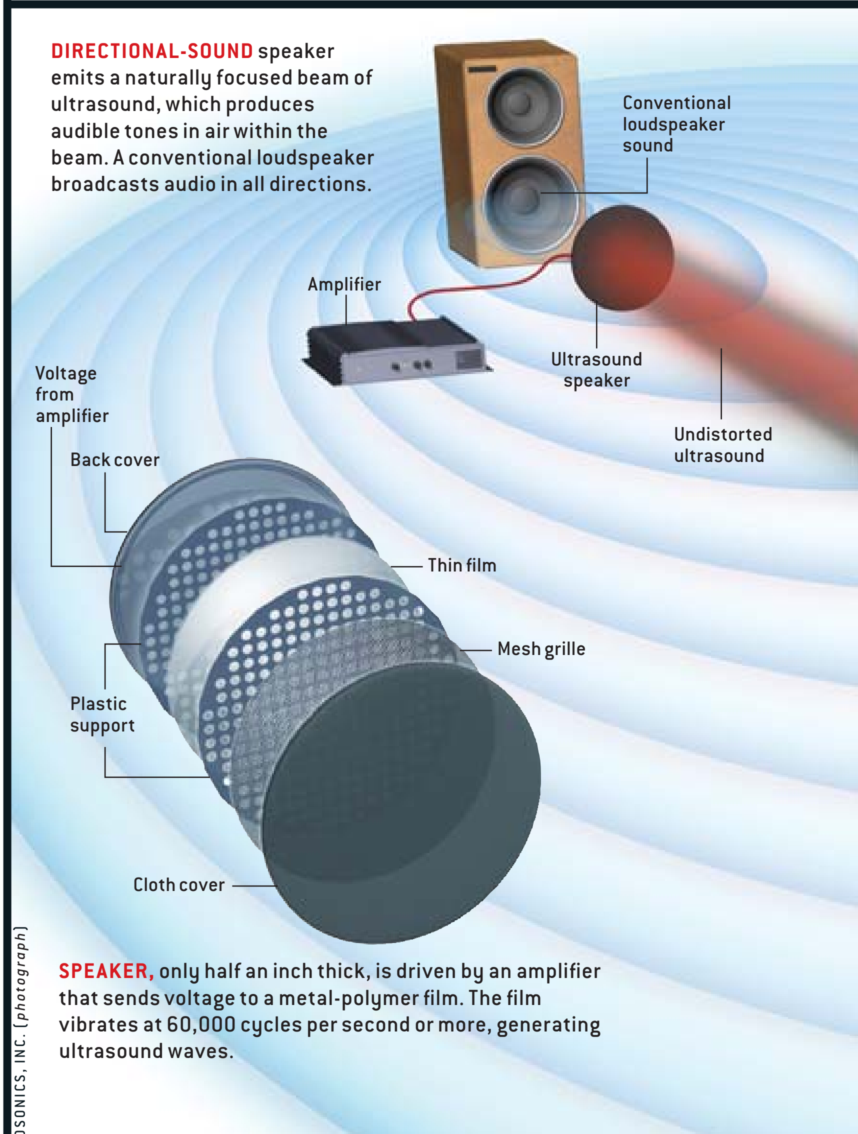
You are walking down a quiet grocery store aisle when suddenly a voice says: "Thirsty? Buy me." You stop in front of the soda display, but no one is next to you, and shoppers a few feet away do not seem to hear a thing.

At that moment, you are standing in a cylinder of sound. Whereas a loudspeaker broadcasts sound in all directions, the way a lightbulb radiates light, a directional speaker shines a beam of waves akin to a spotlight. The beam consists of ultrasound waves, which humans cannot hear, but which can emit audible tones as they interact with air. By describing these interactions mathematically, engineers can coax a beam to exude voice, music or any other sound.

Military and sonar researchers tried to harness the phenomenon as far back as the 1960s but only managed to generate highly distorted audible signals. In 1998 Joseph Pompei, then at the Massachusetts Institute of Technology, published algorithms that cut the distortion to only a few percent. He then designed an amplifier, electronics and speakers to produce ultrasound "that is clean enough to generate clean audio," Pompei says. He trademarked the technology Audio Spotlight and started Holosonics, Inc., in Watertown, Mass., in 1999. Rival inventor Woody Norris markets a competing product called HyperSonic Sound from his American Technology Corporation in San Diego.

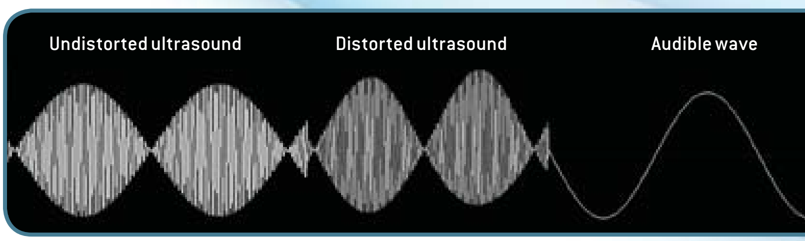
Pompei's speakers are installed in company lobbies, and above exhibits at the Boston Museum of Fine Arts and Walt Disney World's Epcot Center, among other locations. Narrations inform visitors standing in front of artifacts or video screens without filling the rooms with noise. Department stores have tried the arrangement for retail displays, and automakers are experimenting with them so passengers can hear only their own music or movies. A speaker above a recliner in the living room would allow Dad to hear the television while other family members read on the couch in peace.

Detractors say that in certain situations headphones can provide similar benefits, and note random problems, such as unwanted reflections off a car seat. But the primary obstacle to wider deployment is cost: systems run from $600 to $1,000 or more. If the price drops, consumers are more likely to consider buying the gear . . . or encounter it while shopping. —*Mark Fischetti*



DIRECTIONAL-SOUND speaker emits a naturally focused beam of ultrasound, which produces audible tones in air within the beam. A conventional loudspeaker broadcasts audio in all directions.

SPEAKER, only half an inch thick, is driven by an amplifier that sends voltage to a metal-polymer film. The film vibrates at 60,000 cycles per second or more, generating ultrasound waves.



UNIFORM ULTRASOUND waves (*left*) produce density fluctuations in air, which distort the waves (*center*), causing them to emit an audible tone (*right*).

5W INFOGRAPHICS (*illustration*); HOLOSONICS, INC. (*photograph*)

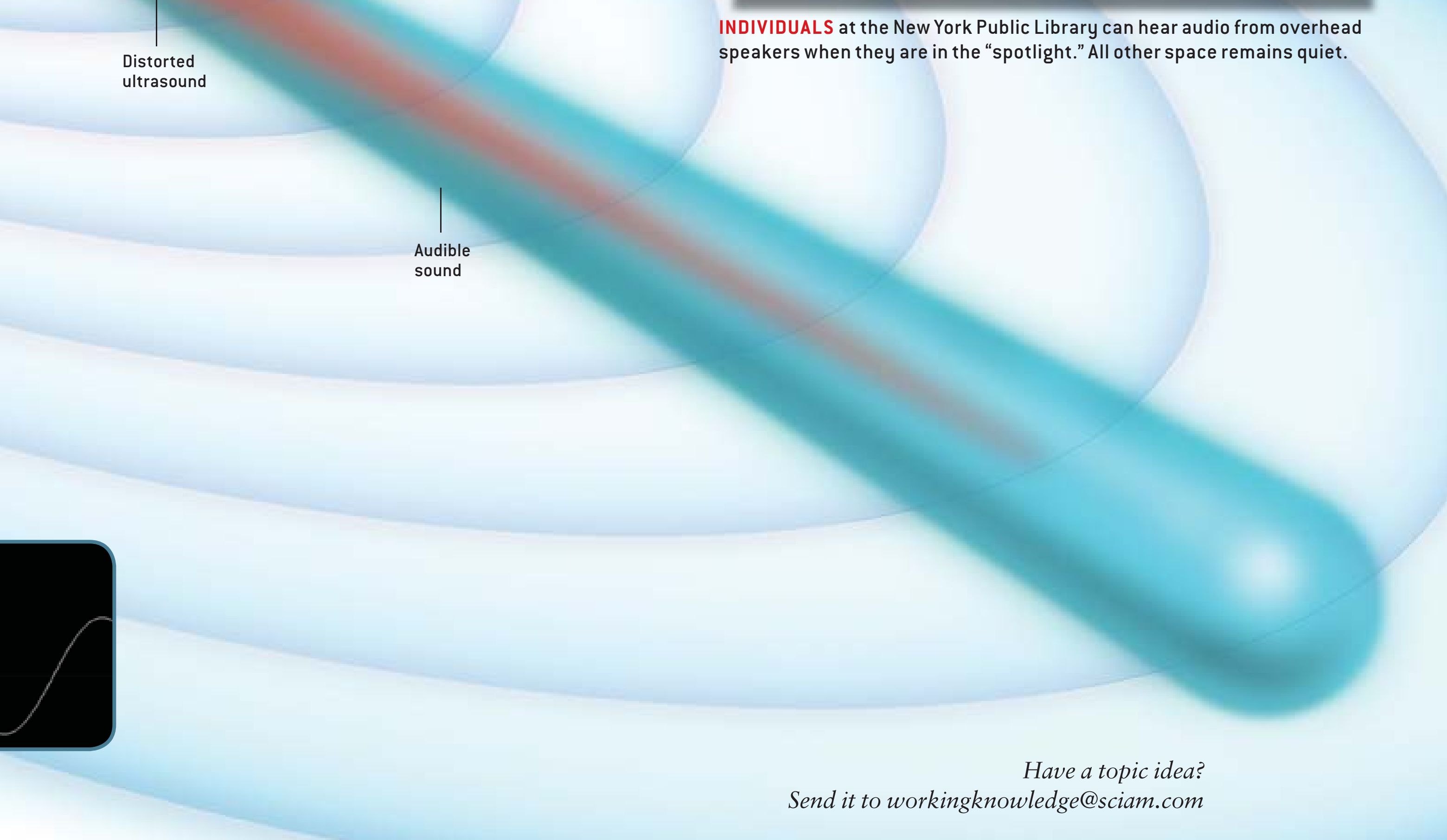
DID YOU KNOW . . .

BOUNCED: Ultrasound waves remain in a tight column when they reflect off a hard, smooth surface. Police teams could bounce a beam off a building at the end of an alley or off a distant window inside a warehouse to flush out suspects, who would run away from the sound—and right into the officers' waiting arms.

BATS NOT DOGS: Certain animals can detect the ultrasound noise behind audible directed sound. The ultrasound speakers emit frequencies from 40,000 to 80,000 cycles a second, or hertz (Hz). Humans typically hear frequencies between 20 and 20,000 Hz. Dogs can hear up to 40,000 Hz or so, mice up to 90,000, and bats, porpoises and beluga whales up to 100,000 Hz or higher.

BONUS: Middle ear bones limit human hearing to below 20,000 Hz. But researchers have applied ultrasound up to 200,000 Hz to the skulls of volunteers, some of whom report "hearing" sounds; the skull may be distorting vibrations that reach the cochlea.

MURDERER: Last November, Court TV used Audio Spotlight in major bookstores to market its new *Murder by the Book* television series. When a patron, milling about, unwittingly stepped near a poster about the show, a voice from nowhere whispered, "Hey, you—over here. Don't turn around. Do you hear me? Do you ever think about murder? Committing the ultimate crime? I do. All the time. I get paid to think about it. I'm a best-selling crime writer . . . "



INDIVIDUALS at the New York Public Library can hear audio from overhead speakers when they are in the "spotlight." All other space remains quiet.

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

The Interplay of Art and Science

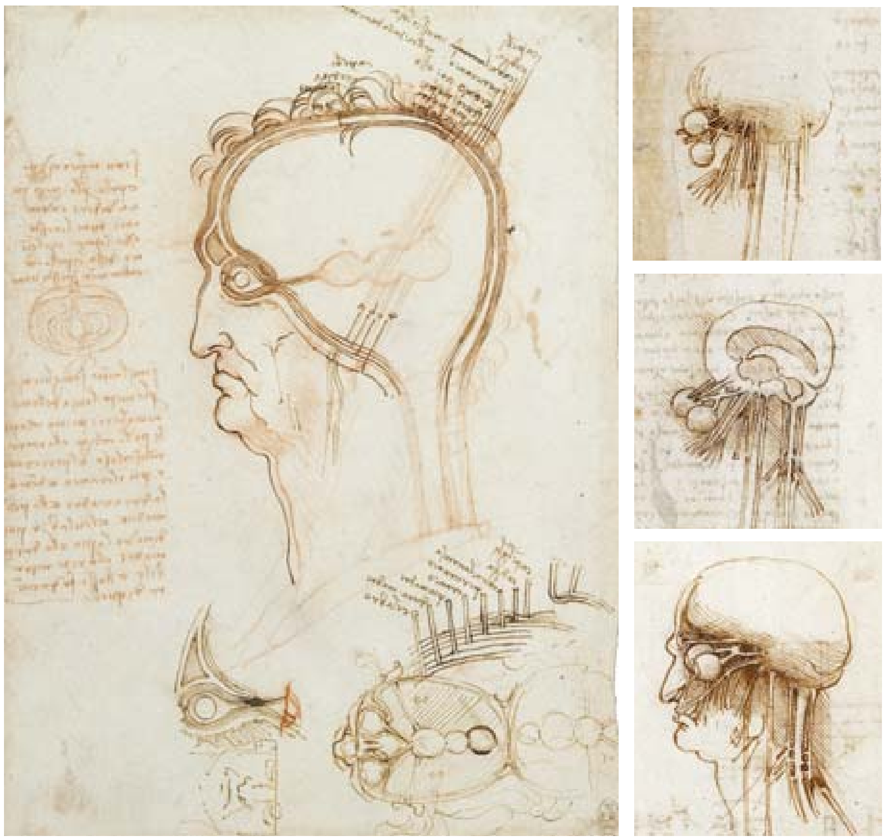
TWO WAYS OF VIEWING THE WORLD MEET IN THE VISUAL REALM BY BETTYANN HOLTZMANN KEVLES

LEONARDO DA VINCI: EXPERIENCE, EXPERIMENT AND DESIGN
by Martin Kemp
Princeton University Press, 2006 ($60)

SEEN/UNSEEN: ART, SCIENCE, AND INTUITION FROM LEONARDO TO THE HUBBLE TELESCOPE
by Martin Kemp
Oxford University Press, 2006 ($45)

Published almost simultaneously, these very different books present a double view of Martin Kemp's original and often brilliant approach to the connection between science and art. *Leonardo* focuses on a single genius; *Seen/Unseen* pulls back the lens to investigate the nature of creativity thematically, using profiles of extraordinary artists/scientists over a span of 500 years. Kemp is intrigued by visual works that combine the skills of artist and scientist, often, but not always, in the same person; he calls himself a "historian of the visual."

Kemp's ideas may be familiar to readers of *Nature* who have followed his science and culture column. Currently professor of art history at the University of Oxford, he has produced a veritable library of books and articles about Leonardo. Away from his desk he organizes exhibitions, most recently one at the Victoria and Albert Museum in London; the beautiful, coffee table–size *Leonardo da Vinci: Experience, Experiment and Design* accompanied this exhibition. Kemp's text dovetails seamlessly with the often life-size reproductions of Leonardo's drawings, sketches,



THROUGH HIS STUDIES and drawings, Leonardo became aware of the joining and subsequent divergence of the nerves from the eye in the optic synapse. (Circa 1500; pen, ink and red chalk.)

THE ROYAL COLLECTION, © 2007 HER MAJESTY QUEEN ELIZABETH II

jottings and notes, providing insight into the artist's philosophy, mastery of geometry, and compassion.

Leonardo da Vinci has fascinated generations of artists, art critics and novelists. But few have been as absorbed as Kemp, who has walked in the master's footsteps through Italy and France. Kemp's name is synonymous with Leonardo scholarship. In a February *New York Times* account of the announcement of the possible discovery of a missing Leonardo masterpiece, *The Battle of Anghiari*—walled off but retrievable in the Palazzo Vecchio—the reporter gave weight to the claim by noting that Kemp strongly supported the investigation.

Leonardo is based on a "theme sheet," a miraculously recovered piece of folded paper (reproduced across two pages in the book) on which the artist began sketching and writing in 1490.

Kemp explains the value of paper half a millennium ago—which meant that every inch was used—and he interprets the apparent chaos of sketches as a record of Leonardo's "laboratory for thinking." He tells us that there are about 6,000 surviving pages of the artist's notes and sketches.

Yet even for the keenly interested, Leonardo's notes, written in a mirror script, are difficult to decipher, and their dispersion on this theme sheet may not be deliberate. A finely drawn portrait of an old man merges with a delicate pair of trees, and the rest of the drawings, notes and designs bear no apparent relation to these sketches or to one another. The jottings, combined with references to the well-known paintings and models for machines, reveal a man of his time at home with geometry, though probably not mathematically prepared for the scientific revolution that was a century in the future. Ahead of his time in a visual way, however, Leonardo observed nature closely and drew "notes" from what he saw, from an autopsied heart to floodwaters and debris, believing that all knowledge had to be confirmed by observing natural phenomena. He must have had remarkable eyesight to capture motion as he did, as well as a rare ability to see mentally in three dimensions, one of the themes of Kemp's second book, *Seen/Unseen*.

Although Leonardo is not the heart of *Seen/Unseen,* he is the touchstone for much of Kemp's thinking and one of the "nodal" figures, men (mostly) who as both artists and scientists are the focus of each of its chapters. In "Wholes and Parts," for instance, Kemp identifies Charles Darwin as a man with one foot in the dynamic view of nature of the early 19th century, where men such as his grandfather Erasmus Darwin and artist George Stubbs saw "an animate nature in continual flux," and the other foot in the Victorian era, in which "Natural Selection itself represents the emergence of the 'fittest' theory into a very specific environment" that was familiar with the earlier worldview and thus receptive to his much more radical theory.

In this same chapter Kemp leaps to the 20th century and Richard Dawkins's "selfish genes" and James Lovelock's Gaia hypothesis and wonders about a self-regulating biosphere. Kemp likes to move back and forth across the centuries, explaining how scientific ideas morph and mutate in different historical eras—a sort of dialogue across time that he relishes. In "Growth and Form," he pays homage to D'Arcy Wentworth Thompson's classic 1917 book *On Growth and Form,* explaining Thompson's references to Albrecht Durer's method of proportional transformation

and connecting this approach to form to the visual mathematics of fractals, chaos theory and machine-made images.

Sometime in the 19th century, according to Kemp, the invention of the camera and its offspring technologies, cinema and radiography, liberated artists from having to reproduce nature with conventional perspective, freeing them to move in different directions. Scientists, in contrast, embraced photography and happily wed cameras to their microscopes and telescopes.

Kemp is still in the process of exploring the explosion of visual culture in the late 20th and early 21st centuries. We humans are, as he illustrates in various ways throughout this rich discussion, visual creatures. We need to *see* in order to understand. Biologists, he reminds us, needed an electron microscope to see DNA and, from that glimpse, to construct a model. He lauds the contribution of physicist Richard Feynman, who devised the "Feynman diagrams" (which are not included in the text), because he thought, and taught, with images. Kemp applauds the visual reconstructions of contemporary planetary scientists who capture images of planets, such as Venus, while taking the liberty of coloring them with familiar, Earth-like tones. Why, he asks, bother to map Venus, where we will never go? Because, he replies, we are explorers, and explorers want to map their discoveries.

Kemp acknowledges that his reconsideration of scientists and artists across the centuries is sometimes difficult, and he asks his readers to bear with him. Some may be left sensing that this is not his final effort to explain visual culture, which is not the study of art or of science but an inquiry into human creativity that can be *seen*. In his own ongoing inquiry, Kemp is ambitious and enlightening but occasionally difficult to follow. He questions the validity, interpretation and ultimately the use of computerized, machine-made images extracted, for example, from PET and fMRI brain scans. Suddenly he fears the technology he has been describing. "The more technological the image looks, the more it exudes… authority," he writes, but a computer is, nonetheless, a man-made tool that "seems to promise a non-human precision." And it would be a mistake to put the tool makers in the privileged position of deciding how their tools should be used.

Kemp's thoughts on art are self-consciously different in kind from other recent efforts to link art and science that either concentrate on individual artists and scientists or focus on a particular scientific breakthrough. None approaches the world as Kemp does. *Seen/Unseen* is a glimpse into his own thought process the way his *Leonardo* is a glimpse into the mind of an almost mythical genius. Kemp offers us a way of considering how artists and scientists have intuited visual truths in the past, reminds us that the past and the present are connected, and warns us against the potential tyranny of the newest digitized images that, though often beautiful and beguiling, are still man-made and not infallible. SA

Bettyann Holtzmann Kevles, author of Naked to the Bone: Medical Imaging in the Twentieth Century, *teaches a course on the history of science, invention and the visual arts at Yale University.*

Ale's Well with the World

A YEASTY DISCUSSION OF THE SCIENCE OF MALT AND HOPS **BY STEVE MIRSKY**

Free beer. And thus I found my way to a lecture in late February at the New York Academy of Sciences by renowned beer maven Charlie Bamforth. A man for whom the word "avuncular" was coined, the British Bamforth has three decades of brewing expertise under his belt. Over his belt is what he insists is "a sausage belly, not a beer belly."

"Beer is the basis of modern static civilization," began Bamforth, Anheuser-Busch Endowed Professor of Brewing Science at the University of California, Davis. "Because before beer was discovered, people used to wander around and follow goats from place to place. And then they realized that this grain [barley] could be grown and sprouted and made into a bread and crumbled and converted into a liquid which gave a nice, warm, cozy feeling. So gone were the days that they followed goats around. They stayed put while the grain grew and while the beer was brewed. And they made villages out of their tents. And those villages became towns, and those towns became cities. And so here we are in New York, thanks to beer." Another syllogism ended his address: "He who drinks beer sleeps well. He who sleeps well cannot sin. He who does not sin goes to heaven. The logic is impeccable."

In between came numerous nuggets about what Bamforth insists is "the world's favorite beverage," some of which were at the expense of another popular adult drink. While discussing the simple chemical equation of fermentation, by which a molecule of sugar is converted to two molecules of ethanol and two of carbon dioxide (along with some energy), Bamforth noted that "if the sugars are from grapes, you produce a fine beverage called wine. If the sugars are from grain, it's a superior beverage called beer."

Bamforth decried beer's sometimes dicey image: "Beer is perceived as a bad-boy drink, and beer has been too often marketed in all sorts of strange ways—flatulent horses [an infamous ad during the 2005 Super Bowl, overshadowed by the even more infamous Janet Jackson "wardrobe malfunction"] and men behaving badly. And wine is perceived somehow as being superior and speaking to a higher quality of life. Really, it's unfair. Beer is more consistent; it's produced with more devotion and care; it's at least as healthy." Plus, human feet are conspicuously absent from beer making.

But fetid feet pale when an ale becomes stale: tinted bottles or talented chemists must be employed to keep beer from becoming "sunstruck." Light converts certain beer bitter acids to the dreaded 3-methyl-2-butene-1-thiol (MBT) compound, a close relative of the malodorous thiols produced by *Mephitis mephitis,* the striped skunk. In his book *Beer: Tap into the Art and Science of Brewing* (Oxford University Press, 2003), Bamforth notes that some people can smell MBT at levels as low as 0.4 part per trillion. "These poor people," he writes, "would have been able to detect a tenth of a gram of MBT distributed throughout the balloon of the airship *Graf Zeppelin II.*" Eau, the humanity.

So, does the brewski master have a favorite beer? "It depends on where I am," Bamforth explains. "If I'm in an old pub with a ceiling about my height and there's a roaring log fire, cask ale from England is sublime. I wouldn't have an American-style lager. If I'm at a Sacramento River Cats AAA baseball game and it's 100 degrees outside, I could kill for a Bud. I'm not going to drink a Guinness. So it's horses for courses." As long as those horses aren't flatulent.

Answering the charge that beer is empty calories, Bamforth points out that beer is actually rich in B vitamins, except for thiamine. "A famous doctor came up to me," he remembers, "and said, 'Is it true that if we could just boost the levels of thiamine, beer would be a meal in itself?' I answered, 'Even I wouldn't say that. You need a few pretzels.' " SA

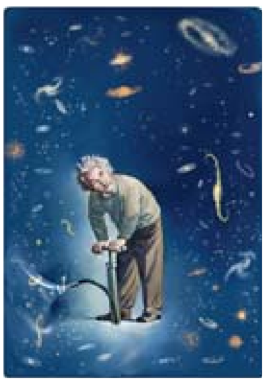
ILLUSTRATION BY MATT COLLINS; PHOTOGRAPH BY FRANK VERONSKY

Where is the universe expanding to?

—A. Kenny, Canisbay, Scotland

Astrophysicist Alexander Kashlinsky of the NASA Goddard Space Flight Center offers this explanation:

The evolution of the universe is described by the physics of general relativity discovered by Albert Einstein during the early 20th century. In general relativity, space and time are merged into one continuum and the universe can be represented as a four-dimensional spacetime grid. When viewed from this perspective, the universe's expansion does not push it into new territory—rather the spacetime grid itself is expanding.



In prerelativistic, Newtonian physics (which describes celestial bodies as moving according to the laws discovered by Isaac Newton), space and time are absolute, with time no more than a parameter in the equations of motion. Gravity, meanwhile, is seen as a force of attraction between massive bodies, but its source is a mystery.

The physics of general relativity is conceptually distinct—even if its equations of motion can be reduced to Newtonian equations in many practical cases. In general relativity, the properties of the spacetime grid are uniquely specified (via gravity) by the bodies inhabiting it. Gravity curves the spacetime continuum, and general relativity thus describes gravitational interactions as manifestations of that curvature. Objects under gravity's influence "fall" from less curved parts of spacetime to more curved parts.

According to Einstein's general relativity equations, the spacetime containing matter cannot remain stationary and must either expand or contract. Galaxies, then, are not strictly moving away from one another but rather are attached to the fixed grid on the expanding fabric of spacetime, thereby giving the impression of moving away from one another. As an analogy, imagine placing dots on the surface of a balloon, then inflating it. The distances between the dots—which represent galaxies—will increase, so if you live in one of these dots, you will interpret the others as receding from you. In reality the dots remain in the same positions, with respect to the two coordinates (latitude and longitude) on the surface of the balloon, and it is the fabric of the balloon that is actually expanding.

In the framework of general relativity with only four dimensions, the question posed here does not have an answer, because it implies some other coordinate grid outside spacetime. Because spacetime is linked to matter, there is no *outside* to the surface of the balloon—it is all the spacetime that is available.

What is "junk" DNA, and what is it worth?

—A. Khajavinia, Isfahan, Iran

Wojciech Makalowski, a Pennsylvania State University biology professor and genomics researcher, replies:

All animals have a large excess of junk DNA—genetic material that does not code for the proteins used to build bodies and catalyze chemical reactions within cells. In our genetic blueprint, for instance, only about 2 percent of DNA actually codes for proteins.

In 1972 the late geneticist Susumu Ohno coined the term "junk DNA" to describe all noncoding DNA sections, most of which consist of repeated segments scattered randomly throughout the genome. Typically sections of junk DNA come about through transposition, or movement of sections of genetic material to different positions in the genome. As a result, most of these regions contain multiple copies of so-called transposons—sequences that literally copy or cut themselves out of one part of the genome and reinsert themselves somewhere else.

In the early 1990s interest in junk DNA, and especially in repetitive elements, began to grow; many biologists now regard such repetitions as genomic treasures. It appears that these transposable sequences increase the ability of a species to evolve by serving as hot spots for genetic recombination and by providing important signals for regulating gene expression. As such, repetitive elements are hardly "junk" but rather are integral components of our genomes.

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

MATT COLLINS