How Game Theory Could Stop Sports Doping (*page 82*)

SCIENTIFIC AMERICAN

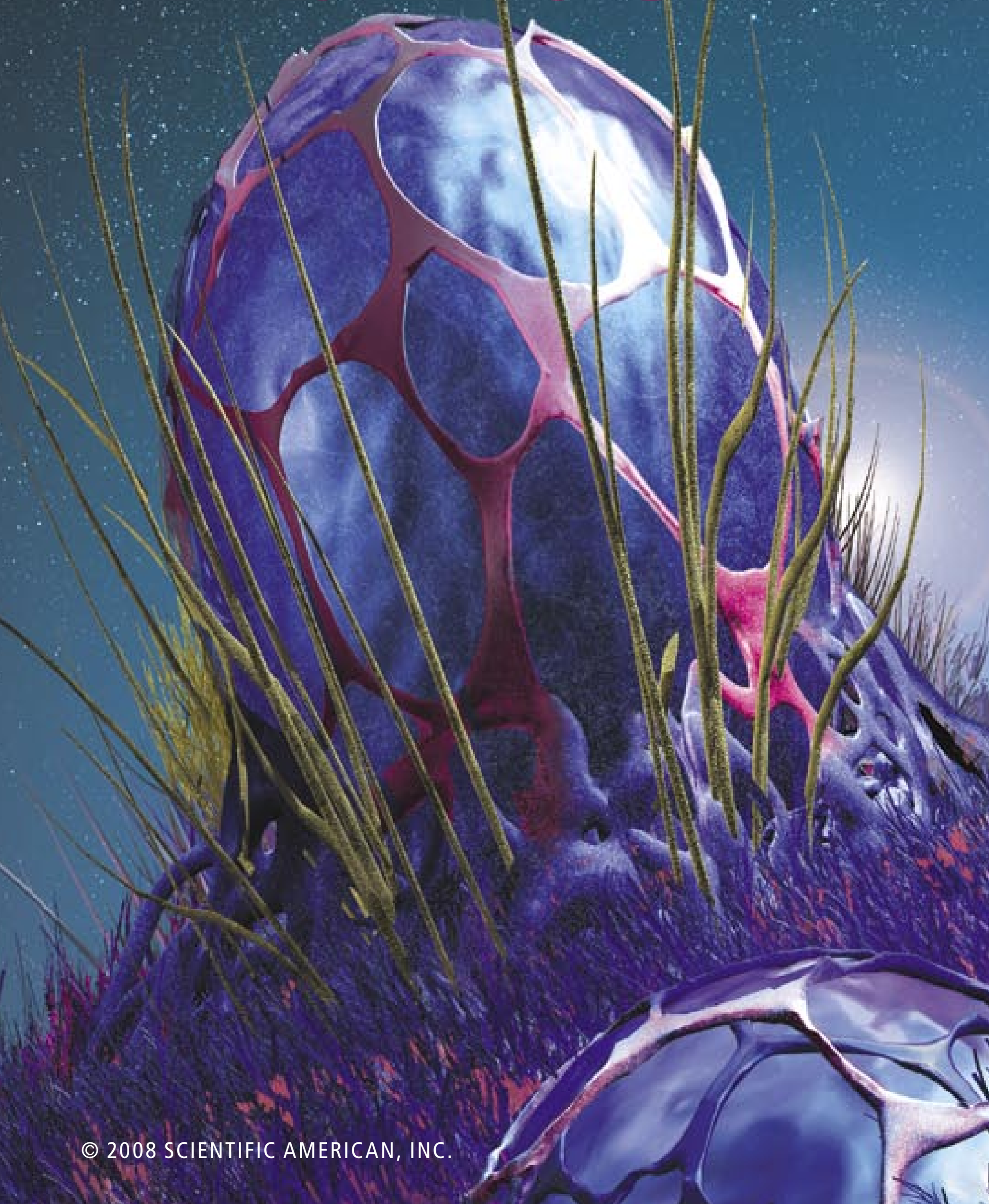
How to REGROW LIMBS

page 56

April 2008 $4.99 www.SciAm.com

The Shocking Colors of Alien Plants

Astronomers learn how to spot life on worlds where plants would not be green



Nuclear Smuggling

Why Scanning at Ports Can't Stop It

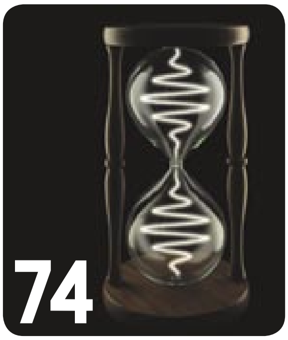
Laser Pulses

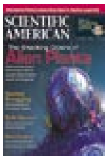
Making Atomic Clocks Better

Flat Carbon

Faster Than Silicon for Electronics

FEATURES

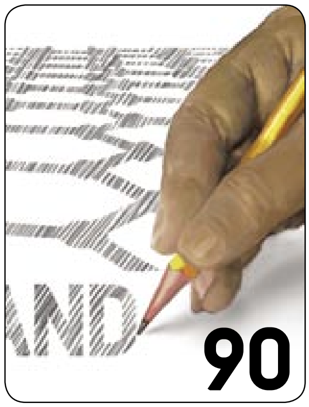


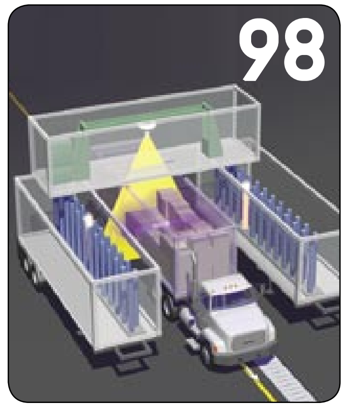


ON THE COVER

A shiny, blue-pigmented plant awaits the scorching assault of its supergiant F-type sun in this artist's conception by Kenn Brown. The conceptual image of the growing hand is by Jean-Francois Podevin.

MORE FEATURES





SCI AM

Doping and Baseball

Tell us how you would discourage players from using performance enhancers at **www.SciAm.com/ontheweb**

GO TO SCIAM.COM

OF ANTS, ELEPHANTS AND ACACIAS

Without large grazing herbivores to eat them, acacia trees suffer because of a shift in the ant populations they house.
More at **www.SciAm.com/ontheweb**

COURTESY OF TODD PALMER

In Focus

Grass Makes Better Ethanol Than Corn Does

Midwestern farms prove switchgrass could be the right crop for producing ethanol to replace gasoline.

News

What Are We Thinking When We Try to Solve Problems?

New research indicates what happens in the brain when we are faced with a dilemma.

Podcast

Life's U-Shaped Path of Happiness

A cross-cultural study of 74 countries finds that most people do indeed have a midlife crisis.

Weird Science

Looking at the Sun Can Trigger a Sneeze

This common reaction has a mysterious genetic underpinning.

Video

The Monitor

Scientific American's weekly video roundup of the newest and most fascinating in science.

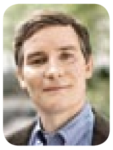
Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111. Copyright © 2008 by Scientific American, Inc. All rights reserved. No part of this issue may be reproduced or transmitted in any form or by any means, electronic or mechanical, 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.

DEPARTMENTS

BRIAN MARANAN PINEDA

Future Facts

Extraordinary developments spring from the intersection of science and fiction



The symbiotic relation between science and science fiction is so tight that in some instances it is hard to tell which inspired the other—or whether both grew from the same seed. Jules Verne's *20,000 Leagues under the Sea* stirred generations of submariners, and the U.S. Navy named its first nuclear sub, the USS *Nautilus,* in tribute to fictional Captain Nemo's vessel, although Verne himself was enthused by the submarine that Robert Fulton designed for the French in 1800. Legendary paleontologist George Gaylord Simpson wrote a short novel, *The Dechronization of Sam Magruder* (from which *Scientific American* posthumously published excerpts—see the September 1999 issue), to help the public understand scientists' reconstructions of life in the Cretaceous. Many roboticists credit Isaac Asimov's imaginary Three Laws of Robotics with influencing how they thought about their field.

GARDENS of Polaris IV

In this issue, at least two feature articles are all about cutting-edge science yet also describe the stuff of science fiction.

Nancy Y. Kiang, who studies how ecosystems and atmospheres interact for the NASA Goddard Institute for Space Studies, has rendered a service to future science-fiction writers everywhere with her article on "The Color of Plants on Other Worlds" (*page 48*). In most stories, alien flora receive only a sliver of the creative attention lavished on extraterrestrial beings or animals. That problem is particularly evident in science-fiction films, where, for understandable budgetary reasons, the plants on other planets often look distinctly like those found in the deserts near Los Angeles. (I have faint memories of reading decades ago that set designers on *Star Trek* sometimes created alien vegetation by turning potted plants upside down and exposing their roots.)

In reality, the equivalent of plants on other worlds may still possess something like stems and leaves as efficient structures for collecting solar energy, but whatever they use as a photosynthetic pigment may be a far cry from chlorophyll. Just as famed robber Willie Sutton quipped that he stole from banks "because that's where they keep the money," Earth's plants rely on chlorophyll because it efficiently absorbs red and blue light, and those are the parts of the spectrum where our sun invests most of its photonic energy.

Yet the sun's emissions are far different from those of many stars with planets; hence, any plants under their distant rays will need different color pigments. Kiang's essay is a fascinating guide to what we may someday find flourishing on those other worlds. (And let us recognize the spectacular job of artists Kenn Brown and Chris Wren in bringing that strange vegetation to life in their accompanying illustrations.)

Science fiction also sometimes holds out a promise that future medicine will be able to regenerate lost limbs. Biologists Ken Muneoka, Manjong Nan and David M. Gardiner are one team of investigators working to realize that dream, as they explain in their article, which begins on page 56. The encouraging results of their experiments to date should be inspiring to their fellow scientists and science-fiction writers alike. ■

JOHN RENNIE
editor in chief

FLYNN LARSEN (*Rennie*); KENN BROWN AND CHRIS WREN *Mondolithic Studios* (*landscape*)

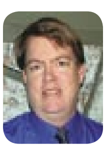
Among Our Contributors

THOMAS B. COCHRAN
is a nuclear physicist. He holds the Wade Greene Chair for Nuclear Policy at the Natural Resources Defense Council.

ANDRE K. GEIM
is a physicist at the University of Manchester in England. He directs the Manchester Center for Mesoscience and Nanotechnology.

NANCY Y. KIANG
is a biometeorologist at the NASA Goddard Institute for Space Studies. She is a member of the Virtual Planetary Laboratory, a team of the NASA Astrobiology Institute studying how to detect life on other worlds.

PHILIP KIM
is a physicist at Columbia University who focuses on quantum thermal and electrical transport processes in nanoscale materials.

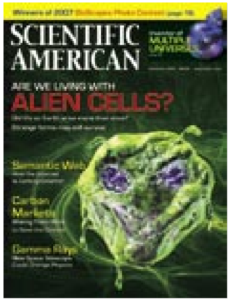


MATTHEW G. McKINZIE
is a nuclear physicist and senior scientist in the Nuclear Program at the Natural Resources Defense Council.

MICHAEL SHERMER
is a contributing editor of *Scientific American* and author of the *Skeptic* column. A serious amateur cyclist, he has competed in the Race Across America.

editors@SciAm.com

The Semantic Web ■ Many Worlds ■ Emissions Trading



DECEMBER 2007

"Given the complexities of a cap-and-trade system of carbon reduction, it would likely take years to get the necessary details agreed on."
—*John Burton* WASHINGTON, N.J.

A Question of Semantics

In "The Semantic Web in Action," Lee Feigenbaum, Ivan Herman, Tonya Hongsermeier, Eric Neumann and Susie Stephens describe the development of the Semantic Web, a set of formats and languages to find and analyze data on the World Wide Web easily. The problem with this system is that different people will not agree on exactly how to define all concepts. Any computer application that tries to standardize its ontology will necessarily distort what at least some people are trying to express.

I am also concerned that in conventional formal logic, if even one inconsistency exists it will be possible to draw *all* possible conclusions *and* their contradictions!

Robert W. Jones
Emporia State University

FEIGENBAUM AND HERMAN REPLY: *Discrepancies among ontologies do occur, but the Semantic Web does not rely on having one, all-encompassing ontology. Instead it is built up from small, like-minded communities that can find agreement on terms among themselves. Applications can therefore interact without attempting to achieve global consensus: a system that presents a retailer's wares to customers will harvest information from suppliers' databases (themselves likely to use heterogeneous formats) and map it onto the retailer's preferred ontology before being displayed to customers. Automated tax-return software takes bank data, conforming to individual banks' ontologies, and maps them onto the tax form. There is no requirement for global ontologies: instead an application need only map the terms relevant for a particular transaction into a common vocabulary. Of course, although agreement need only be local, adoption of existing vocabulary still facilitates data sharing and integration.*

As to the dangers of using conventional logic, "inference" in the Semantic Web can be characterized as discovering new relations. But the inferences on this network are done within a restricted, "guarded" subset of first-order logic. Ontological reasoning on the Semantic Web does not use the full power of first-order (or higher-order) logic and therefore avoids some of the dangerous conclusions that can come from an inferred inconsistency.

180 Degrees of Aberration?

In his profile of Hugh Everett and his many-worlds interpretation of quantum mechanics, "The Many Worlds of Hugh Everett," Peter Byrne states that Bryce S. DeWitt "swung around 180 degrees [on Everett's theory] and became its most devoted champion," thus implying that previously DeWitt had a negative opinion of the theory. In fact, DeWitt promoted Everett's work from the beginning. DeWitt was the acting editor for a section of the July 1957 issue of *Reviews of Modern Physics* containing proceedings from a physics conference that he and I had organized in January of that year (he was not the regular editor of the journal). DeWitt decided to include Everett's thesis in the proceedings even though Everett had not attended the conference. DeWitt had read the thesis at John Archibald Wheeler's request, but Wheeler had doubts about a work questioning Niels Bohr's understanding of quantum mechan-

➥ *continued on page 18*

➥ *continued from page 14*

ics. DeWitt found it "new and refreshing."

A few years later DeWitt discovered that Everett's paper had "slipped into instant obscurity" and resolved to start a publicity campaign. This campaign was not a 180-degree swing in his position but a continuation of his 1957 interest in Everett's work.

I am not writing as Bryce's widow but as Bryce S. DeWitt's colleague—a colleague who knew how to disagree with him when necessary. I wish to correct a misleading presentation of a historical fact.

Cecile DeWitt-Morette
Jane and Roland Blumberg Centennial Professor, Emerita
University of Texas at Austin

BYRNE REPLIES: *The historical record shows that DeWitt was initially troubled by Everett's theory. "I simply do not branch," he wrote in May 1957. He dropped that objection after Everett pointed out that the earth moves, even though we do not feel it.*

HUGH EVERETT theorized innumerable universes spawned by quantum effects.

Stagnating Strategy?

In "Enough Hot Air Already" [Perspectives], the editors say that because our legislators avoid carbon taxes as political suicide, we should adopt a national cap-and-trade market as an alternative. But in "Making Carbon Markets Work," David G. Victor and Danny Cullenward describe the process of allocating emission credits as "politically charged and corruption-prone." I would add to this the likelihood that given the complexities of a cap-and-trade system, it would take years to get the necessary details agreed on. Congress is still far from passing any bill since House Speaker Nancy Pelosi of California said in January 2007 that global-warming legislation would be a top priority. A cap-and-trade system requires acceptance of different credits by numerous industries and commercial establishments. Existing heavy emitters will insist on grandfather status; companies such as DuPont that have already cut emissions will want exemption from further cuts.

Maybe the step-by-step global approach described by Jeffrey D. Sachs in "Meaningful Goals for Climate Talks" [Sustainable Developments] should be applied to our domestic situation. There are several immediate steps that would make some cuts while Congress debates a carbon tax.

John Burton
Washington, N.J.

Speaking American

Although the researchers cited in "Use It or Lose It," by Nikhil Swaminathan [News Scan], claim to be investigating the English language, they appear to be studying the emerging American language specifically. It may surprise them to learn that in the country that produced *Beowulf*, the past tense form of the irregular verb "to slink" resolutely remains "slunk."

Perhaps research should be redirected into estimating how long it will be before the English and American languages can be considered separate, so that two great nations can get on with the daily business of misunderstanding each other properly?

Steve Green
Bristol, England

CLARIFICATIONS "The 2007 Nobel Prizes" [News Scan] should have noted another Nobel winner who wrote for *Scientific American:* Eric S. Maskin of the Institute of Advanced Study in Princeton, N.J. He shared the economics prize with Leonid Hurwicz of the University of Minnesota and Roger B. Myerson of the University of Chicago. All won for their development of mechanism design theory.

"Supersonic Pulse Power," by Steven Ashley [News Scan], should have stated that the firing rates of pulse detonation engines should occur more than twice as frequently as those in pulse jets because such engines will likely incorporate multiple combustion tubes.

Letters to the Editor

Scientific American
415 Madison Ave.
New York, NY 10017-1111
or editors@SciAm.com

Letters may be edited for length and clarity. We regret that we cannot answer all correspondence.

How to Contact Us

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

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: Permissions Department, Scientific American, 415 Madison Ave., New York, NY 10017-1111; www.SciAm.com/permissions or 212-451-8546 for procedures. 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.

ILLUSTRATION BY SEAN McCABE; COURTESY OF PRINCETON UNIVERSITY LIBRARY (*Everett*)

Hoyle on Matter ▪ Our Readers? ▪ Killing Beauty

Compiled by Daniel C. Schlenoff

APRIL 1958

THE MATTER—"Anti-matter may exist in our galaxy, but it cannot exceed about one part in 10,000,000 of ordinary matter if it is there. It is most unlikely that any of the stars in our galaxy can be made of anti-matter. Outside our galaxy, other galaxies in remote parts of the universe may consist entirely of anti-matter. The nearest approach to direct proof of the existence of such bodies is the presence of strong radio sources whose energy is difficult to explain by any known process but might be explained by the annihilation of anti-matter. On the other hand, if anti-matter does exist in the universe, we do not understand at present how the bulk of it became separated from matter. To explain this would apparently require a revolution in our thinking on cosmological problems. —Geoffrey Burbidge and Fred Hoyle"

APRIL 1908

SPA CHERNOBYL?—"In connection with the Austrian governmental establishment for the preparation of uranium products, there has been built in Joachimsthal, Bohemia, a laboratory for working up radio-active substances found in the tailings and by-products of the uranium minerals. There will also be erected a bathing establishment, where the radio-active mine water will be used for healing purposes."

PUBLIC MENACE—"It is surprising that the turnstile door [revolving door] has not long ago aroused a strong protest, on the ground that it constitutes a menace to public safety. We say this with due appreciation of the ingenuity of this device, and the success with which it accomplishes its desired end of preventing the inrush of cold air which accompanies the opening and shutting of doors of the ordinary hinged type. The menace of the door lies in the fact that people can pass through the rotating pockets of the door only one at a time. In the event of any accident that would cause a rush for the door, it would take an interminable time for the population of a hotel to file out."

U.S. MAGAZINES, according to *Puck*, 1908

VIEW FROM *PUCK*—"Our contemporary *Puck* has published, in one of its recent issues, illustrations which it considers typical of certain periodicals, among them the Scientific American. We reproduce the illustrations for the delectation of those readers of the Scientific American who wish to see themselves as *Puck* sees them. The coatless, bald-headed man of the dripping brow and intent mien is not the ideal reader that we have seen in our mind's eye; but his furious activity has our fullest approval."

APRIL 1858

WILD HORSES—"All kinds of theories have been formed in relation to the peculiar method of subduing the wild spirit of horses, so successfully practiced in Europe by Mr. Rarey, who is generally known as the 'American horse tamer.' It is well known that animals generally have an almost instinctive passion for certain odors, which appear to have a subduing influence over them. Mr. Rarey has intimated that his power over the horse is obtained solely through herbs or drugs which operate on the senses of smell and taste. We understand that Mr. Rarey has been challenged by D. Sullivan, also a horse tamer (grandson of the celebrated 'Sullivan, the Whisperer'), to a trial of his powers in Cork, Ireland."

FATAL ATTRACTION—"There is no custom so foolish and frivolous as that of painting the face, or endeavoring to obtain by artificial means an unnatural complexion; and this custom, which at first we are inclined to regard as simply childish, assumes the graver nature of a crime when we regard the means adopted to attain this silly end. For example: Arsenic is used in great quantities to produce a healthy look, ruby lips, and rotundity of form. Bismuth and antimony are also largely used in the manufacture of cosmetic articles, without which ladies do not consider their toilets complete. We would let every one know that some ladies actually, as well as figuratively, deal out to their admirers *killing* glances."

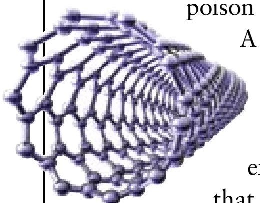
SCIENTIFIC AMERICAN, VOL. XCVIII, NO. 16; APRIL 18, 1908

Nanotubes ▪ Offshore Wind Power ▪ Synthetic Genome ▪ Stress at Work

Edited by Philip Yam

Nontoxic Nanotubes

Nanoparticles could perform promising biomedical tasks, such as ferrying drugs to diseased cells or detecting genetic anomalies [see "Less Is More in Medicine"; SciAm, September 2001]. But researchers fret that these tiny tools could poison the body.



A Stanford University team has done an experiment that may help allay those fears. The researchers injected about 20 micrograms of single-walled carbon nanotubes into mice. The rodents showed no signs of ill effects or any deleterious accumulation of the nanotubes in organs. In fact, after three months, the mice excreted the tubes in their feces and urine. The February 5 *Proceedings of the National Academy of Sciences USA* published the findings.

Seeking New Life

Constructing artificial organisms is a key goal of synthetic biology, because such customized creatures could be made to perform many useful functions [see "Synthetic Life"; SciAm, May 2004]. In January scientists at the J. Craig Venter Institute in Rockville, Md., reported making all 582,970 base pairs that constitute the genome of the bacterium *Mycoplasma genitalium*. Now the researchers just have to insert this man-made genome into a cell and see if the cell comes to life—a prospect likely to be achieved within a year, given their previous success with genome swapping among bacteria. To prove that the genome is artificial, the team encoded "watermarks" of amino acids whose single-letter abbreviations spell out the names of the institute and the researchers.

How Your Job Is Killing You

In the early 1970s a project tracking 18,000 male British civil servants found that the lowest-ranking white-collar workers had the highest rates of premature death. The results of the investigation, called the Whitehall Study, were a surprise, because it turned out that workload or responsibility had little

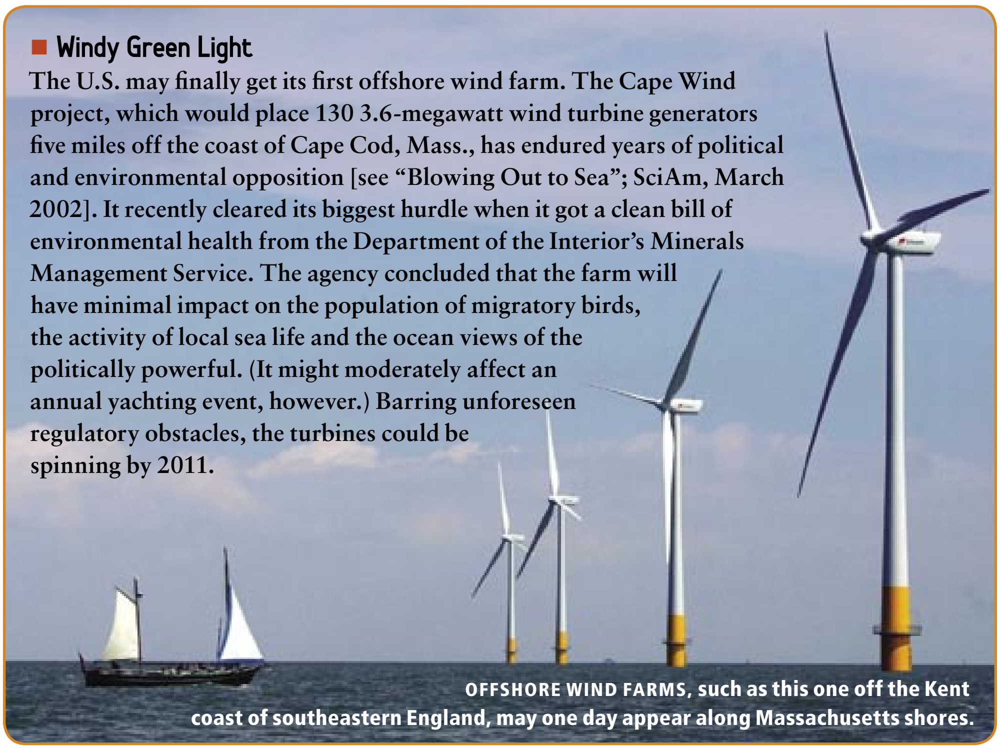
relation to stress levels. Rather it was how much control an employee had over the work he did and how he did it.

In January researchers—following up with the Whitehall II study, begun in the 1980s—unveiled fresh details about the mechanisms underlying the now firmly established links among low job control, stress and high cardiovascular disease. They found that fully a third of an individual's total risk for heart disease stemmed from stress-related unhealthy behaviors, such as poor diet, smoking and lack of exercise, as well as lifestyle-influenced conditions such as high blood pressure and blood glucose. The other two thirds of risk was attributable to direct biological wear and tear from living constantly in fight-or-flight mode.

Workers with the highest stress levels, for instance, had the lowest heart-rate variability, a measure of heart rhythms controlled by the body's autonomic nervous system. Chronic exposure to stress hormones weakens the heart's ability to respond to changing demands, and low heart-rate variability is associated with greater risk of heart attack and lower survival rates afterward. The *European Heart Journal* published the findings online January 23. —*Christine Soares*

Windy Green Light

The U.S. may finally get its first offshore wind farm. The Cape Wind project, which would place 130 3.6-megawatt wind turbine generators five miles off the coast of Cape Cod, Mass., has endured years of political and environmental opposition [see "Blowing Out to Sea"; SciAm, March 2002]. It recently cleared its biggest hurdle when it got a clean bill of environmental health from the Department of the Interior's Minerals Management Service. The agency concluded that the farm will have minimal impact on the population of migratory birds, the activity of local sea life and the ocean views of the politically powerful. (It might moderately affect an annual yachting event, however.) Barring unforeseen regulatory obstacles, the turbines could be spinning by 2011.



OFFSHORE WIND FARMS, such as this one off the Kent coast of southeastern England, may one day appear along Massachusetts shores.

LAGUNA DESIGN/PHOTO RESEARCHERS, INC. (*nanotube*); DAVID BEBBER, POOL *AP Photo* (*windmills*); ANDERSEN ROSS *Getty Images* (*man at desk*)

ECONOMICS

Volatility Kills

Recent economic growth in sub-Saharan Africa may not be sustainable BY GARY STIX

Despite gun battles in the capital of Chad, rioting in Kenya and galloping inflation in Zimbabwe, the economies of sub-Saharan Africa are, as a whole, in better shape than they were a few years ago. The World Bank has reported recently that this part of the continent experienced a respectable growth rate of 5.6 percent in 2006 and a higher rate from 1995 to 2005 than in previous decades. The bank has given a cautious assessment that the region may have reached a turning point.

An overriding question for developmental economists remains whether the upswing will continue so Africans can grow their way out of a poverty that relegates some 40 percent of the nearly 744 million in that region to living on less than a dollar a day. The optimism, when inspected more closely, may be short-lived because of the persistence of a devastating pattern of economic volatility that has lingered for decades.

The growth spurts have been interrupted by declines so fierce that these countries did not grow at all from 1975 to 2005. "In reality, African countries grow as fast as Asian countries and other developing countries during the good times, but afterward they see growth collapses," comments Jorge Arbache, a senior World Bank economist. "How to prevent collapses may be as important as promoting growth." If these collapses had not occurred, he observes, the level of gross domestic product for each citizen of the 48 nations of sub-Saharan Africa would have been a third higher.

The prerequisites to prevent the next crash are not in place, according to a World Bank study issued in January, *Is Africa's Recent Growth Robust?* The growth period that began in 1995, driven by a commodities boom spurred in particular by demand from China, may not be sustainable, because the economic fundamentals—new investment and the ability to stave off inflation, among other factors—are absent. The region lacks the necessary infrastructure that would encourage investors to look to Africa to find the next Bengaluru (Bangalore) or Shenzhen, a November report from the bank concludes.

AKINTUNDE AKINLEYE *Reuters*

BOON OR BANE? Financial bonanzas from oil and other natural resources in countries such as Nigeria can divert attention from instituting needed economic reform measures.

For sub-Saharan countries rich in oil and other resources, a boom period may even undermine efforts to institute sound economic practices. From 1996 to 2005, with growth accelerating, measures of governance—factors such as political stability, rule of law, and control of corruption—actually worsened, especially for countries endowed with abundant mineral resources, the January report notes.

Perhaps the most incisive analysis of the volatility question comes from Paul Collier, a longtime specialist in African economics at the University of Oxford and author of the recent book *The Bottom Billion.* He advocates a range of options that the U.S. and other nations could adopt when formulating policy toward African countries. They include revamped trade measures, better-apportioned aid and sustained military intervention in certain instances, to avert what he sees as a rapidly accelerating divergence of the world's poorest, primarily in Africa, from the rest of the world, even other developing nations such India and China.

Collier finds that bad governance is the main reason countries fail to take advantage of the revenue bonanza that results from a boom. Moreover, a democratic government, he adds, often makes the aftermath of a boom worse. “Instead of democracy disciplining governments to manage these resource booms well, what happens is that the resource revenues corrupt the normal functioning of democracy—unless you stop [them from] corrupting the normal function of democracy” with sufficient checks and balances, he said at a talk in January at the Carnegie Council in New York City.

Collier advocates that African nations institute an array of standards and codes to bolster governments, one which would substitute auctions for bribes in apportioning mineral rights and another which would tax export revenues adequately. He cites the Democratic Republic of the Congo, which took in $200 million from mineral exports in 2006 yet collected only $86,000 in royalties for its treasury. If a nation gets these points right, he argues, “it’s going to develop. If it gets them wrong, it won’t.”

To encourage reform, Collier recommends that the G8 nations agree to accept these measures as voluntary guidelines for multinationals doing business in Africa—companies, for instance, would only enter new contracts through auctions monitored by an international verification group. Such an agreement would follow the examples of the so-called Kimberley Process, which has effectively undercut the trade in blood diamonds, and the Extractive Industries Transparency Initiative, in which a government must report to its citizens the revenues it receives from sales of natural resources.

These measures, he says, are more important than elevating aid levels, an approach emphasized by economist Jeffrey D. Sachs of Columbia University and celebrity activists such as Bono. Collier notes that Angola receives tens of billions of dollars in oil revenue. “Whether it gets a few hundred million more or less in aid is really second-order,” he insists. Ultimately, hard-nosed prescriptions are likely to achieve more than the peregrinations to Darfur or Davos of Bono or Angelina Jolie.

ENERGY

Digital Diet

Computing industry gets serious about energy conservation BY DAVID BIELLO

Telecommuting, Internet shopping and online meetings may save energy as compared with in-person alternatives, but as the digital age moves on, its green reputation is turning a lot browner. E-mailing, number crunching and Web searches in the U.S. consumed as much as 61 billion kilowatt-hours last year, or 1.5 percent of the nation’s electricity—half of which comes from coal. In 2005 the computers of the world ate up 123 billion kilowatt-hours of energy, a number that will double by 2010 if present trends continue, according to Jonathan Koomey, a staff scientist at Lawrence Berkeley National Laboratory. As a result, the power bill to run a computer over its lifetime will surpass the cost of buying the machine in the first place—giving Internet and computer companies a business reason to cut energy costs, as well as an environmental one.

One of the biggest energy sinks comes not from the computers themselves but from the air-conditioning needed to keep them from overheating. For every kilowatt-hour of energy used for computing in a data center, another kilowatt-hour is required to cool the furnacelike racks of servers.

For Internet giant Google, this reality has driven efforts such as the installation of a solar array that can provide 30 percent of the peak power needs of its Mountain View, Calif., headquarters as well as increasing purchases of renewable energy. But to deliver Web pages within seconds, the firm must maintain hundreds of thousands of computer servers in cavernous buildings. "It's a good thing to worry about server energy efficiency," remarks Google's green energy czar Bill Weihl. "We are actively working to maximize the efficiency of our data centers, which account for most of the energy Google consumes worldwide." Google will funnel some of its profits into a new effort, dubbed RE<C (for renewable energy cheaper than coal, as Google translates it) to make sources such as solar-thermal, high-altitude wind and geothermal cheaper than coal "within years, not decades," according to Weihl.

In the meantime, the industry as a whole has employed a few tricks to save watts. Efforts include cutting down on the number of transformations the electricity itself must undergo before achieving the correct operating voltage; rearranging the stacks of servers and the mechanics of their cooling; and using software to create multiple "virtual" computers, rather than having to deploy several real ones. Such virtualization has allowed computer maker Hewlett-Packard to consolidate 86 data centers spread throughout the world to just three, with three backups, says Pat Tiernan, the firm's vice president of social and environmental responsibility.

The industry is also tackling the energy issue at the computer-chip level. With every doubling of processing power in recent years has come a doubling in power consumption. But to save energy, chipmakers such as Intel and AMD have shifted to so-called multicore technology, which packs multiple processors into one circuit rather than separating them. "When we moved to multicore—away from a linear focus on megahertz and gigahertz—and throttled down microprocessors, the energy savings were pretty substantial," says Allyson Klein, Intel's marketing manager for its Ecotech Initiative. Chipmakers continue to shrink circuits on the nanoscale as well, which "means a chip needs less electricity" to deliver the same performance, she adds.

POWER-HUNGRY: **Computer servers, such as these at the University of Texas at Houston, are growing in number and, consequently, in electricity usage. Vast server farms, such as those being established by Google, consume as much energy in cooling as they do in computing.**

MATTHEW LANDRY

With such chips, more personal computers will meet various efficiency standards, such as Energy Star compliance (which mandates that a desktop consume no more than 65 watts). The federal government, led by agencies such as NASA and the Department of Defense may soon require all their purchases to meet the Electronic Product Environmental Assessment Tool standard. And Google, Intel and others have formed the Climate Savers Computing Initiative, an effort to cut power consumption from all computers by 50 percent by 2010.

Sleep modes and other power management tools built into most operating systems can offer savings today. Yet about 90 percent of computers do not have such settings enabled, according to Klein. Properly activated, they would prevent a computer from leading to the emission of thousands of kilograms of carbon dioxide from power plants every year. But if powering down or unplugging the computer (the only way it uses zero power) is not an option, then perhaps the most environmentally friendly use of all those wasted computing cycles is in helping to model climate change. The University of Oxford's ClimatePrediction.net offers an opportunity to at least predict the consequences of all that coal burning.

How to Offset a Web Page's Carbon Footprint

CO_2Stats is a free tool that can be embedded into any Web site to calculate the carbon dioxide emissions associated with using it. That estimate is based on an assumption of 300 watts of power consumed by the personal computer, network and server involved—or 16.5 milligrams of CO_2 emitted every second of use. "The typical carbon footprint is roughly equivalent to 1.5 people breathing," says physicist Alexander Wissner-Gross of Harvard University, who co-created the Web tool. The founders then purchase enough renewable energy to account for the electricity used.

PLANETARY SCIENCE

Dropping Acid

Mars may have needed acid rain to stay wet BY SARAH SIMPSON

On Mars, signs of wetness keep pouring in: deeply carved river valleys, vast deltas and widespread remnants of evaporating seas have convinced many experts that liquid water may have covered large parts of the Red Planet for a billion years or more. But most efforts to explain how Martian climate ever permitted such clement conditions come up dry. Bitterly cold and parched today, Mars needed a potent greenhouse atmosphere to sustain its watery past. A thick layer of heat-trapping carbon dioxide from volcanoes probably shrouded the young planet, but climate models indicate time and again that CO_2 alone could not have kept the surface above freezing.

Now, inspired by the surprising discovery that sulfur minerals are pervasive in the Martian soil, scientists are beginning to suspect that CO_2 had a warm-up partner: sulfur dioxide (SO_2).

Like CO_2, SO_2 is a common gas emitted when volcanoes erupt, a frequent occurrence on Mars when it was still young. A hundredth or even a thousandth of a percent SO_2 in Mars's early atmosphere could have provided the extra boost of greenhouse warming that the Red Planet needed to stay wet, explains geochemist Daniel P. Schrag of Harvard University.

That may not sound like much, but for many gases, even minuscule concentrations are hard to maintain. On our home planet, SO_2 provides no significant long-term warmth because it combines almost instantly with oxygen in the atmosphere to form sulfate, a type of salt. Early Mars would have been virtually free of atmospheric oxygen, though, so SO_2 would have stuck around much longer.

"When you take away oxygen, it's a profound change, and the atmosphere works really differently," Schrag remarks. According to Schrag and his colleagues, that difference also implies that SO_2 would have played a starring role in the Martian water cycle—thus resolving another climate conundrum, namely, a lack of certain rocks.

Schrag's team contends that on early Mars, much of the SO_2 would have combined with airborne water droplets and fallen as sulfurous acid rain, rather than transforming into a salt as on Earth. The resulting acidity would have inhibited the formation of thick layers of limestone and other carbonate rocks.

Researchers assumed Mars would be chock-full of carbonate rocks because their formation is such a fundamental consequence of the humid, CO_2-rich at-

NASA/JPL/CORNELL

SULFUR MINERAL (*white*) that forms only in water was churned up by a Mars rover.

mosphere on Earth. Over millions of years, this rock-forming process has sequestered enough of the carbon dioxide spewed from earthly volcanoes to limit the buildup of the gas in the atmosphere. Stifling this CO_2-sequestration step on early Mars would have forced more of the gas to accumulate in the atmosphere—another way SO_2 could have boosted greenhouse warming, Schrag suggests.

Some scientists doubt that SO_2 was really up to these climatic tasks. Even in an oxygen-free atmosphere, SO_2 is still extremely fragile; the sun's ultraviolet radiation splits apart SO_2 molecules quite readily, points out James F. Kasting, an atmospheric chemist at Pennsylvania State University. In Kasting's computer models of Earth's early climate, which is often compared with that of early Mars, this photochemical destruction capped SO_2 concentrations at one thousandth as much as Schrag and his colleagues describe. "There may be ways to make this idea work," Kasting says. "But it would take some detailed modeling to convince skeptics, including me, that it is actually feasible."

Schrag admits that the details are uncertain, but he cites estimates by other researchers who suggest that early Martian volcanoes could have spewed enough SO_2 to keep pace with the SO_2 destroyed photochemically. Previous findings also indicate that a thick CO_2 atmosphere would have effectively scattered the most destructive wavelengths of ultraviolet radiation—yet another example of an apparently mutually beneficial partnership between CO_2 and SO_2 on early Mars.

Kasting maintains that an SO_2 climate feedback could not have made early Mars as warm as Earth, but he does allow for the possibility that SO_2 concentrations may have remained high enough to keep the planet partly defrosted, with perhaps enough rainfall to form river valleys.

Over that point, Schrag does not quibble. "Our hypothesis doesn't depend at all on whether there was a big ocean, a few lakes or just a few little puddles," he says. "Warm doesn't mean warm like the Amazon. It could mean warm like Iceland—just warm enough to create those river valleys." With SO_2, it only takes a little.

Seeking Sulfites

If sulfur dioxide warmed early Mars, as a new hypothesis suggests, minerals called sulfites would have formed in standing water at the surface. No sulfites have yet turned up, possibly because no one was looking for them. The next-generation rover, the Mars Science Laboratory, is well equipped for the search. Scheduled to launch in 2009, the rover (*shown here in an artist's conception*) will be the first to carry an x-ray diffractometer, which can scan and identify the crystal structure of any mineral it encounters.

NASA/JPL/CALTECH

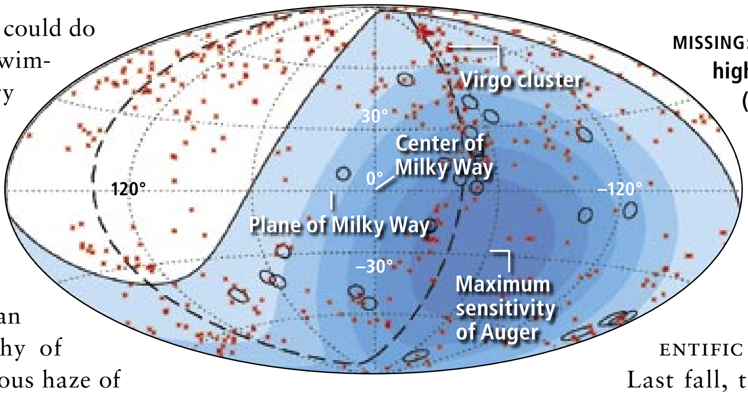
ASTRONOMY

Catching No Rays

Missing in action: ultraenergetic cosmic rays from the Virgo cluster BY GEORGE MUSSER

If you're being shot at, you could do worse than diving into a swimming pool. As the Discovery Channel show *MythBusters* once demonstrated, even a few feet of water can slow a bullet to nonlethal speeds. A similar drama plays itself out in the cosmos. To particles moving faster than a certain velocity, just shy of the speed of light, the tenuous haze of microwave radiation that fills space might as well be a dense sea. Passing through a few hundred million light-years of it should slow a particle to more moderate speeds.

So astronomers have long puzzled over why such particles, known as ultrahigh-energy cosmic rays (UHECRs), are hitting Earth. If they can punch their way through space without losing speed, something must be wrong with scientists' understanding of them, perhaps a failure of Einstein's special theory of relativity [see "Cosmic Power," by George Musser; Scientific American, January 1999].



MISSING: Sky map shows that the highest-energy cosmic rays (*black circles*) detected by the Auger Observatory tend to line up with galaxies less than 250 million light-years away (*red asterisks*). Strangely, the Virgo cluster lacks signs of such rays.

PIERRE AUGER OBSERVATORY

Last fall, though, the world's largest array of cosmic-ray detectors, the Pierre Auger Observatory in Argentina, matched the directions of incoming UHECRs with fairly nearby galaxies. Such an alignment would be unlikely if the particles came from more distant sources and barreled through the universe unimpeded. It looks like Einstein has dodged a bullet.

But what is it about these galaxies that makes them fire off ultrafast particles? On that front a new mystery has arisen. At the American Astronomical Society conference in January, Auger team member Vasiliki Pavlidou of the University of Chicago described how the observatory has seen not one single UHECR coming from the direction of the Virgo cluster of galaxies, a clump of at least 1,000 large galaxies about 60 million light-years away. You name it, Virgo has it: black holes, collapsing stars, dark matter. "Whatever your personal preference might be for the possible sources of ultrahigh-energy cosmic rays, Virgo was guaranteed to have plenty of those," Pavlidou says.

The lack of UHECRs from Virgo might just be a statistical fluke; another year's data might bring some to light. "We must be patient," says another Auger team member, Paul Sommers of Pennsylvania State University. "Virgo may appear." Cosmic rays are so rare, and the number of possible alignments with galaxies so large, that the team is wary of pinning a number to the statistical significance of their absence. Still, if you live near a Gatling gun, you are bound to hear at least a couple of bullets whiz by, no? "The Virgo deficit is already becoming uncomfortable," Pavlidou says.

Many researchers are taking the deficit as a much needed clue to what the ultrapowerful sources might be. Supermassive black holes have long been a prime suspect. These monsters gather swirling disks of matter around them, and as they munch away, they shoot out fast-moving jets of gas that flick particles to epic speeds. Most of the galaxies from which Auger detects UHECRs are thought to contain actively feeding holes. A pioneer of this hypothesis, Peter Biermann of the Max Planck Institute for Radio Astronomy in Bonn, Germany, attributes the Virgo deficit to strong magnetic fields in the outskirts of our own galaxy. These fields deflect incoming charged particles and could have caused particles from Virgo to be misattributed to some other source. The amount of deflection may indicate that UHECRs consist not of protons but of highly charged atomic nuclei, argues Susumu Inoue of the National Astronomical Observatory of Japan.

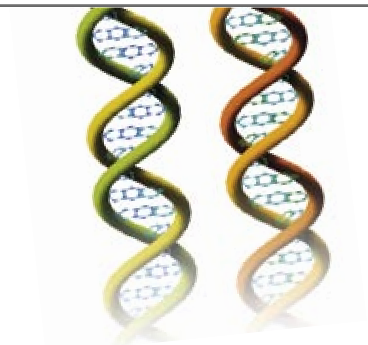
Glennys Farrar of New York University and her colleagues note that supermassive holes in Virgo tend to have undersize disks. Few generate the power required to sling UHECRs. Virgo may be an impressive clump of galaxies, but it is nothing special when it comes to supermassive black holes.

In other ways, too, Virgo could be an unexpectedly wimpy generator of ultra rays. Pavlidou says that the dense packing of galaxies could suppress various conceivable sources. For instance, interactions between galaxies could strip away gas, choke off star formation and reduce the number of exploding stars. Magnetic fields in Virgo itself may trap particles attempting to exit the cluster.

Researchers hope to corroborate the Virgo deficit using observatories in the Northern Hemisphere, which has a better view of the cluster. The so-called Telescope Array began operating in January in western Utah, and plans are afoot to build Auger North in southeastern Colorado. But one weak link in astronomers' knowledge isn't so much the cosmic-ray detection as the ordinary telescopic investigation of potential sources. Cosmic rays could help astronomers understand supermassive black holes and other phenomena. They have gone from novelty to tool.

A Really Fast Fact

Ultrahigh-energy cosmic rays are the fastest material objects known to science. If light takes 300 million years to reach us from a distant galaxy, these particles take 300 million years plus one microsecond.



PAUL MORRELL *Getty Images*

BIOLOGY

Double-Helix Double Up

Identical DNA strands recognize one another from afar BY CHARLES Q. CHOI

Talk about spooky action at a distance. Without any other molecules to guide them, double helices of DNA with identical sequences can recognize one another from a distance and even gather together.

That DNA bases attract is not a surprise, because base pairs are complementary like right- and left-handed gloves: adenine binds with thymine, cytosine with guanine. But when bound in a double-helix form, these bases are tucked away, hidden behind highly electrically charged strings of sugars and phosphates.

Nevertheless, scientists at Imperial College London and the U.S. National Institute of Child Health and Human Development found that double-stranded DNA with the same sequences were about twice as likely to come together as DNA with different sequences, from a distance of up to three nanometers. (Double-stranded DNA is about two nanometers wide.)

The researchers conjecture that the bases within each DNA cause a double helix to kink one way or the other. Although each DNA's electrically charged groups of sugars and phosphates repel those on other DNA double helices, identical molecules have matching curves. As such, whereas all double-stranded DNA molecules repel one another somewhat, the ridges and grooves of identical helices fit together better than with those of other DNA, making it easier for like to cluster with like.

This attraction might help gene fragments align properly before they get shuffled about, perhaps aiding the careful weaving of DNA that occurs during reproduction. It might also ward off some of the genetic errors that underlie cancer and aging. For more, strand yourself in the January 31 *Journal of Physical Chemistry B*.

Charles Q. Choi is a frequent contributor.

PHYSICS

Wipeout?

A hyped theory of everything sinks from sight BY GRAHAM P. COLLINS

"Surfer dude stuns physicists with theory of everything." So ran a *Daily Telegraph* headline last November. The story circulated and quickly achieved widespread notoriety (even my dentist asked me about it). The physics blogosphere carried long threads of comments attacking and defending the theory and then attacking the tone of the discussion. The shouting and acrimony have died down, and the mainstream physics community remains largely unconvinced that the theory can stay afloat. In the words of Marcus du Sautoy, a University of Oxford mathematician writing in the *Telegraph* in late January: "Unfortunately, the consensus, after investigation, is that it is impossible to use E8 in the way Lisi was hoping and produce a consistent model that reflects reality." Not everyone, of course, agrees.

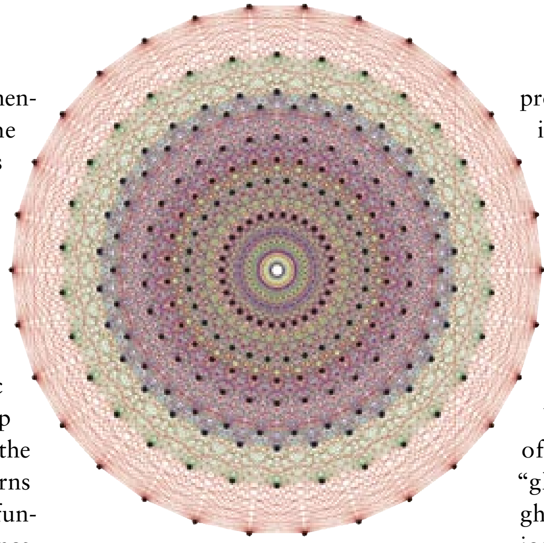
A. Garrett Lisi, the surfer dude in question, came up with his theory while dividing his time among surfing, snowboarding and speculating about physics. He has a Ph.D. in physics from the University of California, San Diego, but has held no academic affiliation since then. He presented his ideas at conferences and invited seminars months before the media furor. From the start, he has been quick to comment that the chances of his theory being correct are very small, but he considers string theory (the approach most favored by physicists) to be even less likely.

Taken at face value, the theory sounds like an incredible discovery. It is based on a remarkable mathematical structure called E8. With 248 dimensions, E8 is the largest, most complicated and most beautiful of five idiosyncratic objects known as the exceptional simple Lie groups. (The title of Lisi's paper, "An Exceptionally Simple Theory of Everything," is first and foremost a pun.) And although E8 has a vast number of dimensions, the physical universe described by the theory could have only the four dimensions we are familiar with and not the 10 or 11 of string theory.

E8 has come up before in physics, most notably in string theory, but Lisi's theory harkens back more to the early 1960s, when physicist Murray Gell-Mann noted that the zoo of subatomic particles then known could be organized into patterns that corresponded to features of another (and far more elementary) Lie group, SU(3). One of the patterns was missing a particle, and Gell-Mann predicted that a particle with certain properties should exist to fill that spot. Experimentalists soon discovered just such a particle.

Today the Standard Model of particle physics organizes all the known elementary particles into these patterns (or "representations"), but it takes a combination of three Lie groups to account for how the particles can interact via three fundamental forces (electromagnetism and the strong and weak nuclear forces). Lisi's insight was that he could place all these particles onto a representation of E8 with only a small number of spots left empty. This process is not just a matter of putting particles in nice-looking patterns in some arbitrary fashion; several properties, such as the electric charges of the particles, have to match up exactly with the relevant quantities in the representation. Furthermore, the patterns include particles that produce the four fundamental forces—including gravity. Hence the optimistic use of "theory of everything" in the title of Lisi's paper.

Closer examination, however, revealed a few Jurassic-size flies in the ointment. For instance, the theory combines the matter particles and the force-carrying particles, referred to in the trade as fermions and bosons, in a way that at first appears fundamentally inconsistent. Various "supersymmetric" theories (including superstring theory) do combine fermions and bosons as well—but only with a detailed mathematical underpinning that E8 does not provide. One way of stating the problem is that if the new theory really describes bosons and fermions, then the structure it places them in cannot possibly be a Lie group at all.



THEORY OF EVERYTHING? The E8 group is related to the symmetries of this eight-dimensional lattice (here, squashed to two dimensions)—and maybe those of the ultimate laws of physics.

JOHN STEMBRIDGE *University of Michigan at Ann Arbor*

Lisi argues that he is using a "mathematical trick" from what is called BRST theory, an established formalism used in string theory and quantum field theory, in which some bosons and fermions play the opposite of their usual roles (and are called "ghosts"). But in ordinary BRST theory, ghosts never manifest as detectable, physical particles, so it is unclear how they could do so consistently in the E8 theory.

Perhaps the longest public debate on the merits of Lisi's theory took place primarily between Jacques Distler of the University of Texas at Austin and Lee Smolin of the Perimeter Institute for Theoretical Physics in Ontario, the latter of whom had been widely quoted in the media with unqualified praise for the theory. (Smolin says he was quoted out of context.) Smolin had also quickly written a paper suggesting ways to correct certain flaws in the E8 proposal. For the particles in the E8 theory to represent the known particles properly, the combination of smaller groups used to form the Standard Model must be embedded inside E8 in just the right way. Distler had demonstrated in his blog that this is a mathematical impossibility. So far as he was concerned, the theory was dead and not worth trying to resuscitate. Yet argument raged on over details of Distler's proof and ultimately ended with neither side conceding. Lisi, incidentally, played very little part in these disputes.

Today the theory is being largely but not entirely ignored. Lisi, naturally, continues to work on it, as does Smolin. Lisi says that even if what Distler claims is true, it would only be true for the variant of E8 ("real E8") originally used in his paper and that another variant ("complex E8") would certainly work. Smolin argues that the press coverage gave the false impression that Lisi's proposal was a finished work. "In reality," he says, "almost every new theoretical proposal is first presented in a way that is flawed and incomplete, with open issues that need to be filled in.... While Lisi's proposal has exciting aspects, this is the case with it as well."

E8: Exceptional Complexity

Although A. Garrett Lisi's use of the mathematical group known as E8 to form a physics "theory of everything" remains controversial, other recent research into the group has been acclaimed as the scaling of a mathematical Mount Everest. German mathematician Wilhelm Killing first formulated E8 nearly 120 years ago, but it was only in January 2007 that a team of mathematicians completed a detailed map of E8's inner workings. The "map," a table of integers with more than 450,000 rows and columns, required 77 hours of supercomputer time to compute and occupied 60 gigabytes of disk space. For comparison, all the genetic information in a person's cells fills three gigabytes.

The program to compute the map took more than three years to write and was primarily the work of Fokko du Cloux of the University of Lyon 1 in France. He died from amyotrophic lateral sclerosis (Lou Gehrig's disease) just two months shy of seeing the project completed.

BIOFUELS

Cellulose Success

Firms seek greener ethanol from wood chips and agricultural waste BY STEVEN ASHLEY

Not too long ago many investors made the bet that renewable fuels from biomass would be the next big thing in energy. Converting corn, sugarcane and soybeans into ethanol or diesel-type fuels lessens our nation's dependence on oil imports while cutting carbon dioxide emissions. But already the nascent industry faces challenges. Escalating demand is hiking food prices while farmers clear rain-forest habitats to grow fuel crops. And several recent studies say that certain biofuel-production processes either fail to yield net energy gains or release more carbon dioxide than they use.

A successor tier of start-up ventures aims to avoid those problems. Rather than focusing on the starches, sugars and fats of food crops, many of the prototype bioethanol processes work with lignocellulose, the "woody" tissue that strengthens the cell walls of plants, says University of Massachusetts Amherst chemical engineer George W. Huber. Although the cellulose breaks down less easily than sugars and starches and thus requires a complex series of enzyme-driven chemical reactions, its use opens the industry to nonfood plant feedstocks such as agricultural wastes, wood chips and switchgrass. But no company has yet demonstrated a cost-competitive industrial process for making cellulosic biofuels.

PRAIRIE GRASS and other nonfood biomass could lead to greener ethanol.

BILL VARIE *Corbis*

So scientists and engineers are working on dozens of possible biofuel-processing routes, reports Charles Wyman, a chemical engineer at the University of California, Riverside, who is a founder of Mascoma Corporation in Cambridge, Mass., a leading developer of cellulosic ethanol processes. "There's no miracle process out there," he remarks. And fine-tuning a process involves considerable money and time. "The oil companies say that it takes 10 years to fully commercialize an industrial processing route," warns Huber, who has contributed some thermochemical techniques to another biomass start-up, Virent Energy Systems in Madison, Wis.

One promising biofuel procedure that avoids the complex enzymatic chemistry to break down cellulose is now being explored by Coskata in Warrenville, Ill., a firm launched in 2006 by high-profile investors and entrepreneurs (General Motors recently took a minority stake in it as well). In the Coskata operation, a conventional gasification system will use heat to turn various feedstocks into a mixture of carbon monoxide and hydrogen called syngas, says Richard Tobey, vice president of engineering and R&D. The ability to handle multiple plant feedstocks would boost the flexibility of the overall process because each region in the country has access to certain feedstocks but not others.

Instead of using thermochemical methods to convert the syngas to fuel—a process that can be significantly more costly because of the added expense of pressurizing gases, according to Tobey—the Coskata group chose a biochemical route. The group focused on five promising strains of ethanol-excreting bacteria that Ralph Tanner, a microbiologist at the University of Oklahoma, had discovered years before in the oxygen-free sediments of a swamp. These anaerobic bugs make ethanol by voraciously consuming syngas.

The "heart and soul of the Coskata process," as Tobey puts it, is the bioreactor in which the bacteria live. "Rather than searching for food in the fermentation mash in a large tank, our bacteria wait for the gas to be delivered to them," he explains. The firm relies on plastic tubes—filter-fabric straws as thin as human hair. The syngas flows through the straws, and water is pumped across their exteriors. The gases diffuse across the selective membrane to the bacteria embedded in the outer surface of the tubes, which permits no water inside. "We get efficient mass transfer with the tubes, which is not easy," Tobey says. "Our data suggest that in an optimal setting we could get 90 percent of the energy value of the gases into our fuel." After the bugs eat the gases, they release ethanol into the surrounding water. Standard distillation or filtration techniques could extract the alcohol from the water.

Coskata researchers estimate that their commercialized process could deliver ethanol at under $1 per gallon—less than half of today's $2-per-gallon wholesale price, Tobey claims. Outside evaluators at Argonne National Laboratory measured the input-output "energy balance" of the Coskata process and found that, optimally, it can produce 7.7 times as much energy in the end product as it takes to make it.

The company plans to construct a 40,000-gallon-a-year pilot plant near the GM test track in Milford, Mich., by the end of this year and hopes to build a full-scale, 100-million-gallon-a-year plant by 2011. Coskata may have some company by then; Bioengineering Resources in Fayetteville, Ark., is already developing what seems to be a similar three-step pathway in which syngas is consumed by bacteria isolated by James Gaddy, a retired chemical engineer at the University of Arkansas. Considering the advances in these and other methods, plant cellulose could provide the greener ethanol everyone wants.

Data Points

More Sleep, Less Fat

Lack of shut-eye contributes to the risk of obesity, according to many recent studies. In exploring childhood obesity, Johns Hopkins University researchers recently completed a meta-analysis of 11 studies that looked at children's sleep duration and their body mass. Not getting enough sack time, the scientists confirmed, disrupts hormone levels, which may lead to excessive weight gain.

Minimum sleep recommended for children:

Younger than 5 years: 11 hours

5–10 Years: 10 hours

Older than 10 years: 9 hours

Increase in odds of being overweight if the child sleeps less than the minimum recommended hours by:

1 Hour: 43 percent

1–2 Hours: 60 percent

More than 2 hours: 92 percent

SOURCE: Obesity, *February 2008*

ASTRONOMY

Mercury Rising

Pockmarked Mercury may superficially look like Earth's moon, but close-up images show that it is anything but moonlike. NASA's MESSENGER spacecraft transmitted these and other images during its first flyby of the scorched planet on January 14. Scientists last saw such detail when Mariner 10 zipped past 33 years ago, and thanks to better equipment and different lighting angles, MESSENGER has gleaned new information.

Equipped with 11 color filters, MESSENGER's eyes can see beyond what human vision can detect. In the false-color photograph (*top*), constructed from three different color-filtered images, younger craters—those no more than 500 million years old—reveal themselves with a faint bluish tinge. The craft has also spied many new "scarps," or faults, extending hundreds of kilometers. Such a scarp runs vertically along the right edge of the image (*bottom*), which encompasses a horizontal extent of about 200 kilometers. Scarps probably formed when Mercury's interior cooled and the planet shrunk, cracking the surface.

MESSENGER—which stands for *Mer*cury *s*urface, *s*pace *en*vironment, *geo*chemistry, and *r*anging—also determined that Mercury's giant Caloris basin, among the biggest impact craters in the solar system, spans 1,500 kilometers—nearly one third of the planet's diameter and 200 kilometers more than previous estimates. Inside the basin and unseen by Mariner 10 is a structure scientists are calling "the spider"—made of troughs radiating outward that probably mark the areas where the basin floor broke apart after its formation.

The craft will revisit Mercury this October and again in September 2009 before settling into an orbit around the planet in March 2011. It will continue to snap pictures and laser-map the topography, as well as take readings of the planet's magnetosphere.

—Philip Yam

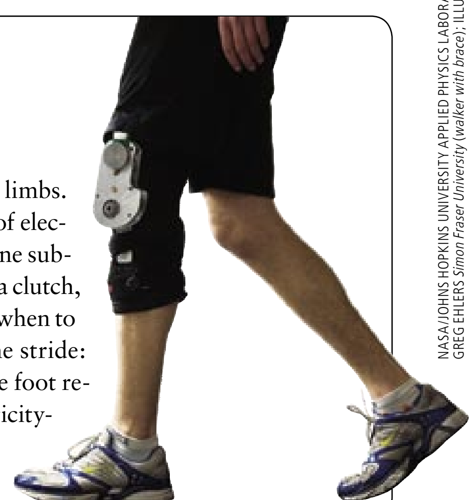
BIOENGINEERING

Knee-Jerk Power

SCI AM

A 1.6-kilogram knee brace worn while walking could power portable devices and prosthetic limbs. Developed by researchers at Simon Fraser University, the brace generated about five watts of electricity per person during a recent experiment, enough to run 10 cell phones concurrently. One subject generated 54 watts by running in place. To capture this energy, the brace relies on gears, a clutch, a generator and a computerized control system that monitors the knee's angle to determine when to engage and disengage power generation. The researchers targeted a particular part of the stride: halfway through the swing of the lower leg after it has left the ground through the time the foot returns to the ground. By tapping into unconscious muscle movements, this and other electricity-from-motion devices are more likely to be used. The researchers describe the brace in the February 8 *Science*.

—Larry Greenemeier



NASA/JOHNS HOPKINS UNIVERSITY APPLIED PHYSICS LABORATORY/CARNEGIE INSTITUTION OF WASHINGTON (*Mercury images*); GREG EHLERS *Simon Fraser University (walker with brace)*; ILLUSTRATION BY MATT COLLINS

In Brief

KISSING COUSINS HAVE MORE KIN

Icelandic women born between 1800 and 1824 who mated with a third cousin had more children and grandchildren (4.04 and 9.17, respectively) than women who mated with someone no closer than an eighth cousin (3.34 and 7.31). Those proportions held up a century later, when family size shrunk. Mating with a relative might reduce a woman's chance of having a miscarriage caused by an immunological incompatibility with her child. There is a limit to family closeness, however: couples that were second cousins or more closely related did not have as many children, probably because the kids inherited mutations that cut their lives short. —*Nikhil Swaminathan*

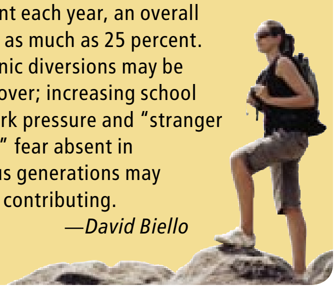
RUNNING DIALOGUE

New languages spin off from older ones with an initial burst of alterations to vocabulary before settling down and gradually changing over time, British researchers report. The group focused on three major language families: Bantu (Swahili and Zulu, for example), Indo-European (English, Latin), and Austronesian (Tagalog, Seediq). Some 10 to 33 percent of divergence between languages stemmed from key vocabulary changes at the time of language splitting.

This discrete evolutionary pattern occurs when a social group tries to forge a separate identity, the researchers say, citing as examples the sudden emergence of American English when Noah Webster published his dictionary in 1828 and, more recently, the development of black American English. —*Nikhil Swaminathan*

NOT GOING OUT TO PLAY

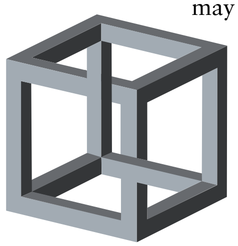
Americans are losing interest in going outdoors. Researchers analyzed trends in visits to parks and forests and in licenses for activities such as hunting and fishing. All peaked between 1981 and 1991 after 50 years of steady increase. But since then, they have been declining at roughly 1 percent each year, an overall drop of as much as 25 percent. Electronic diversions may be taking over; increasing school and work pressure and "stranger danger" fear absent in previous generations may also be contributing. —*David Biello*



PERCEPTION

Eye on the Illusion

Ambiguous images seem to flicker between two alternatives, as if the brain cannot quite make up its mind how to perceive them. The Necker cube, for instance, sometimes looks as if it is pointing into the page and sometimes appears to point out. In an experiment on six volunteers with different kinds of ambiguous visual and auditory stimuli, Christof Koch of the California Institute of Technology and his colleagues found that the pupils dilated around the time that perception shifted. The extent of the momentary dilation, which could be as much as one millimeter, also correlated with how long that particular perception lasted. (Pupils span about two millimeters under bright light.) Because the neurotransmitter norepinephrine controls the pupils, the compound may also play a role in rapid, unconscious decision making. Take a look at the study in the February 5 *Proceedings of the National Academy of Sciences USA*. —*Philip Yam*



IN AND OUT: Pupil size correlates with the perceptual shifts that occur while viewing ambiguous images such as this Necker cube.

NANOTECH

The New Black

Scientists have created the darkest material known: a carpet made of carbon nanotubes standing on end. This nanoforest can return less than 0.1 percent of the light that falls on it—a reflectance of less than one-third that of the previous record holder, a film of nickel-phosphorus alloy. (Normal black paint reflects back 5 to 10 percent.) The gaps between the nanotubes help to entrap light, and the nanotubes, grown on iron nanodots on top of a silicon wafer, are meshed together. The meshing forms an irregular surface that scatters light and thus both minimizes reflection and maximizes absorption, explain researchers at the Rensselaer Polytechnic Institute and Rice University. They suggest that the new material could boost the effectiveness of solar power cells, infrared sensors and astronomical detectors that soak up radiation from space. Absorb more in the February *Nano Letters*. —*Charles Q. Choi*

GEOCHEMISTRY

Lost City, Found Hydrocarbons

Oil, coal and other fuels rich in hydrocarbon come from ancient biological matter, according to conventional wisdom. Now scientists find evidence for an inorganic origin of hydrocarbons—specifically, at Lost City, a collection of natural rock towers atop a submerged mountain in the Atlantic Ocean. The ocean floor around Lost City is full of vents that go all the way to the mantle, and researchers at the Woods Hole Oceanographic Institution and their colleagues have found high concentrations of hydrocarbons in the fluid seeping from these cracks. The hydrocarbons do not carry the isotopic signature linked with life—instead the scientists conjecture that they were produced when carbon from mantle rock interacted with seawater vented from Lost City, which is loaded with hydrogen gas. The researchers suggest that many such fields exist and could have produced building blocks for life. Delve into the February 1 *Science* for the details. —*Charles Q. Choi*

Read More ...

News Scan stories with this icon have extended coverage on **www.SciAm.com/ontheweb**

LISA APPELBACHER (*cube*); DIGITAL VISION/GETTY IMAGES (*hiker*)

SciAm Perspectives

Let the Games Begin!

Nothing beats the excitement of honest, steroid-powered competition

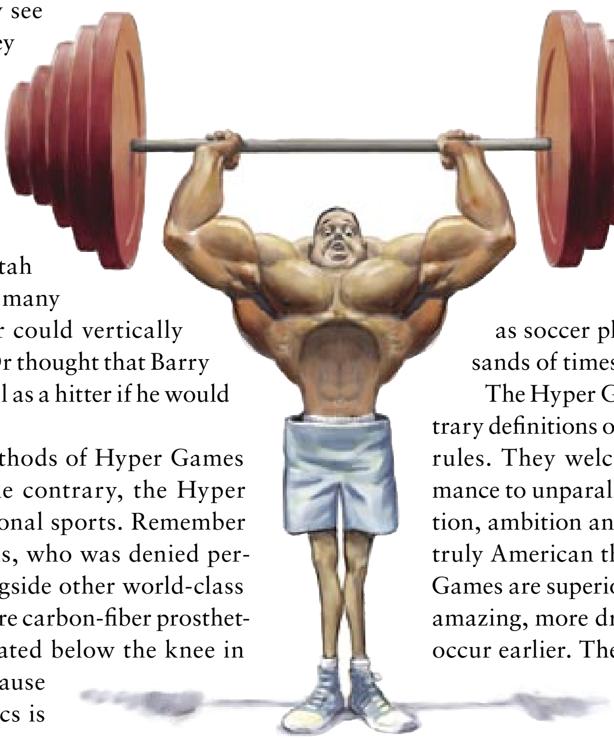
BY THE EDITORS

On page 82, *Scientific American*'s resident skeptic Michael Shermer writes about the doping scandals plaguing cycling, baseball and other sports, and he suggests how to curb those practices. Please ignore him. It would be a global tragedy if his meddling were to ruin the most eagerly awaited competitions of 2008.

No, not this summer's Olympics. Those will of course be modestly fun demonstrations of physical prowess. The pursuit of true excellence is cruel and unforgiving, however, which is why devotees of the *absolute best* in athletic achievement instead turn to the quadrennial Hyper Games.

"Winning isn't everything; it's the only thing," said Vince Lombardi, and those are the words that the elite competitors of the Hyper Games live and frequently die by. The Hyper Games honor those who can run fastest, jump highest, lift the most weight or otherwise excel—by any available means. Competitors are free to use steroids or other performance-enhancing drugs as they see fit, without stigma or penalty. They can also undergo surgical or genetic enhancements.

The Hyper Games are for any spectator who has looked at sprinters and wondered whether they could go faster if their leg muscles contained cheetah DNA. Or speculated about how many tons a determined weight lifter could vertically press before his spine snapped. Or thought that Barry Bonds might show some potential as a hitter if he would just put on some muscle.



MATT COLLINS

Traditionalists spurn the methods of Hyper Games athletes as cheating. But to the contrary, the Hyper Games are fairer than conventional sports. Remember Oscar "Blade Runner" Pistorius, who was denied permission in January to race alongside other world-class runners simply because his feet are carbon-fiber prosthetics? (Both his legs were amputated below the knee in infancy.) Judges ruled that because the springiness of his prosthetics is greater than that of flesh and bone, Pistorius should be barred. Apparently his prosthetics aren't defective enough for him to run against people who have feet. Yet judges routinely turn a blind eye to the natural biological advantages in muscle strength, tendon springiness, aerobic capacity and other traits that winning racers enjoy.

The Hyper Games recognize the hypocrisy of extolling the "purity of sport" when modern athletes aggressively seek every advantage they can find, on and off the playing field. They do not compete nude like classical Olympians: they wear running shoes, they box with gloves, they vault on poles of space-age materials. Swimmers shave off their body hair to reduce drag in the water. Baseball pitchers undergo "Tommy John" surgery that replaces a ligament in their elbow and supposedly gives them better fastballs. Top athletes and their coaches routinely use every available method of training, nourishing and otherwise grooming their bodies and psyches that is legal (or undetectable).

Critics sniff that the Hyper Games are hurtful to the athletes. But are conventional sports much better? Football players end up crippled with arthritis. Basketball players destroy their knees. Hockey players lose teeth. Boxers accumulate brain damage from uppercuts, as soccer players do from heading the ball thousands of times.

The Hyper Games do not inhibit athletes with arbitrary definitions of "fairness" or with patronizing safety rules. They welcome anyone willing to push performance to unparalleled heights, limited only by imagination, ambition and determination. They are thus more truly American than other sports. In short, the Hyper Games are superior to the Olympics in every way: more amazing, more dramatic, more memorable. They even occur earlier. The Olympics do not start until August. The opening ceremonies of the Hyper Games are on April Fools' Day. ■

Sustainable Developments

Keys to Climate Protection

Dramatic, immediate commitment to nurturing new technologies is essential to averting disastrous global warming

BY JEFFREY D. SACHS

Technology policy lies at the core of the climate change challenge. Even with a cutback in wasteful energy spending, our current technologies cannot support both a decline in carbon dioxide emissions and an expanding global economy. If we try to restrain emissions without a fundamentally new set of technologies, we will end up stifling economic growth, including the development prospects for billions of people.

Economists often talk as though putting a price on carbon emissions—through tradable permits or a carbon tax—will be enough to deliver the needed reductions in those emissions. This is not true. Europe's carbon-trading system has not shown much capacity to generate large-scale research nor to develop, demonstrate and deploy breakthrough technologies. A trading system might marginally influence the choices between coal and gas plants or provoke a bit more adoption of solar and wind power, but it will not lead to the necessary fundamental overhaul of energy systems.

For that, we will need much more than a price on carbon. Consider three potentially transformative low-emissions technologies: carbon capture and sequestration (CCS), plug-in hybrid automobiles and concentrated solar-thermal electricity generation. Each will require a combination of factors to succeed: more applied scientific research, important regulatory changes, appropriate infrastructure, public acceptance and early high-cost investments. A failure on one or more of these points could kill the technologies.

CCS, for example, depends on the ability to capture carbon dioxide at the power plant at low cost, transport it by pipeline over significant distances, and sequester it underground safely, reliably and durably. All these components are close to deployment, but each faces major challenges. Carbon capture is most promising for new types of coal-fired plants whose cost and reliability are yet to be proved. A vast new network of carbon dioxide pipelines would require major regulatory and policy support, with environmental and property-rights hurdles. The geologic sequestration of carbon dioxide at large scales must also be proved, carefully monitored and environmentally regulated. Early demonstration projects are likely to be many times more costly than later ones. Broad public acceptance and support will be crucial for the technology. Yet to date, the U.S. government has failed to get even one demonstration CCS power plant off the ground, and various private initiatives are currently stranded, all because of the lack of public support and financing.

Plug-in hybrid automobiles pose similar puzzles. Basic questions remain about the safety, reliability and durability of the batteries they require, as well as the need for extra investments in the power grid to support them. Solar-thermal power, which uses concentrated solar radiation in desert locations to boil water for the steam-turbine generation of electricity, also depends on solving a host of problems. Scientific challenges include the nighttime storage of power, and regulatory and financial obstacles include the installation of a new high-voltage, direct-current transmission grid to carry power over long distances from the desert to other locations.

The issues become even more complex when we consider that low-emissions technologies developed in the rich world will need to be adopted rapidly in poorer countries. Patent protection, while promoting innovation, could slow the diffusion of these technologies to low-income countries unless compensatory actions are taken.

All this technological innovation needs to start soon if we are to have a chance to stabilize carbon emissions at levels that avoid huge and potentially devastating global costs. By 2010 at the latest, the world should be breaking ground on demonstration CCS coal-fired plants in China, India, Europe and the U.S.; the wealthy nations should be helping to finance and build concentrated solar-thermal plants in states that border the Sahara; and highly subsidized plug-in hybrids should be rolling off the assembly line. Only these steps will enable us to peer much farther down the path of truly transformative change. ■

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

PHOTOGRAPH BY BRUCE GILBERT/EARTH INSTITUTE; ILLUSTRATION BY MATT COLLINS

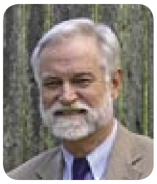
SCI AM An extended version of this essay is available at **www.SciAm.com/ontheweb**

Forum

The Economist Has No Clothes

Unscientific assumptions in economic theory are undermining efforts to solve environmental problems

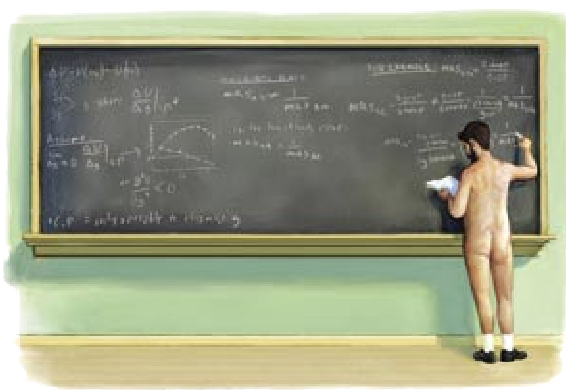
BY ROBERT NADEAU



The 19th-century creators of neoclassical economics—the theory that now serves as the basis for coordinating activities in the global market system—are credited with transforming their field into a scientific discipline. But what is not widely known is that these now legendary economists—William Stanley Jevons, Léon Walras, Maria Edgeworth and Vilfredo Pareto—developed their theories by adapting equations from 19th-century physics that eventually became obsolete. Unfortunately, it is clear that neoclassical economics has also become outdated. The theory is based on unscientific assumptions that are hindering the implementation of viable economic solutions for global warming and other menacing environmental problems.

The physical theory that the creators of neoclassical economics used as a template was conceived in response to the inability of Newtonian physics to account for the phenomena of heat, light and electricity. In 1847 German physicist Hermann von Helmholtz formulated the conservation of energy principle and postulated the existence of a field of conserved energy that fills all space and unifies these phenomena. Later in the century James Maxwell, Ludwig Boltzmann and other physicists devised better explanations for electromagnetism and thermodynamics, but in the meantime, the economists had borrowed and altered Helmholtz's equations.

The strategy the economists used was as simple as it was absurd—they substituted economic variables for physical ones. Utility (a measure of economic well-being) took the place of energy; the sum of utility and expenditure replaced potential and kinetic energy. A number of well-known mathematicians and physicists told the economists that there was absolutely no basis for making these substitutions. But the economists ignored such criticisms and proceeded to claim that they had transformed their field of study into a rigorously mathematical scientific discipline.



Strangely enough, the origins of neoclassical economics in mid-19th century physics were forgotten. Subsequent generations of mainstream economists accepted the claim that this theory is scientific. These curious developments explain why the mathematical theories used by mainstream economists are predicated on the following unscientific assumptions:

- The market system is a closed circular flow between production and consumption, with no inlets or outlets.
- Natural resources exist in a domain that is separate and distinct from a closed market system, and the economic value of these resources can be determined only by the dynamics that operate within this system.
- The costs of damage to the external natural environment by economic activities must be treated as costs that lie outside the closed market system or as costs that cannot be included in the pricing mechanisms that operate within the system.
- The external resources of nature are largely inexhaustible, and those that are not can be replaced by other resources or by technologies that minimize the use of the exhaustible resources or that rely on other resources.
- There are no biophysical limits to the growth of market systems.

If the environmental crisis did not exist, the fact that neoclassical economic theory provides a coherent basis for managing economic activities in market systems could be viewed as sufficient justification for its widespread applications. But because the crisis does exist, this theory can no longer be regarded as useful even in pragmatic or utilitarian terms because it fails to meet what must now be viewed as a fundamental requirement of any economic theory—the extent to which this theory allows economic activities to be coordinated in environmentally responsible ways on a worldwide scale. Because neoclassical economics does not even acknowledge the costs of environmental problems and the limits to economic growth, it constitutes one of the greatest barriers to combating climate change and other threats to the planet. It is imperative that economists devise new theories that will take all the realities of our global system into account. ■

Robert Nadeau teaches environmental science and public policy at George Mason University. His most recently published book is The Environmental Endgame *(Rutgers University Press, 2006).*

SCI AM An expanded version is available at **www.SciAm.com/ontheweb**

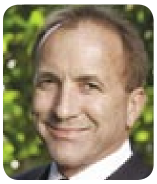
PHOTOGRAPH COURTESY OF ROBERT NADEAU; ILLUSTRATION BY MATT COLLINS

Skeptic

Wag the Dog

Emotions are as much a product of our evolutionary heritage as they are our environmental circumstances

BY MICHAEL SHERMER



The next time you come face to face with a dog wagging its tail, you can make a quick determination on whether to reach out and pet it or step back in deference: check the tail-wag bias. If the wagging tail leans to the dog's right, you're safe; if the tail leans to the dog's left, don't move.

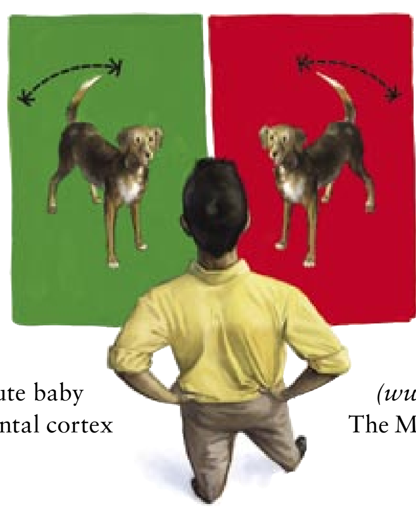
This tail-wagging bias was documented in a 2007 article in the journal *Current Biology* by Italian neuroscientist Giorgio Vallortigara and his veterinarian colleagues at the University of Bari. In an experiment, 30 mixed-breed dogs were each placed in a cage equipped with cameras that measured the asymmetrical bias (left or right) of tail wagging while the pooches were exposed to four stimuli: their owner, an unfamiliar human, a cat and an unfamiliar dominant dog. Owners elicited a strong right bias in tail wagging, and unfamiliar humans and the cat triggered a slight right bias. But the unfamiliar dominant dog (a large Belgian Shepherd Malinois) elicited a strong left bias in tail wagging. Why?

According to the researchers, because the left brain controls the right side of the body, and vice versa, the nerve signals cross the midline of the body and cause the dog's tail to wag more to the right when its left brain is experiencing a positive emotion. This left-right distinction may be explained by the fact that birds, fish and frogs show left-brain/right-brain differences in approach-avoidance behavior, with the left brain associated with positive approach feelings and the right brain associated with negative avoidance feelings. Closer to evolutionary home, when chimpanzees are experiencing negative emotions, they tend to scratch themselves on the left side of their bodies, and left-handed chimps, whose right brain is dominant, tend to be more fearful of novel stimuli than right-handed chimps.

In humans as well, experiments have revealed that the left brain is associated with positive emotions such as love, attachment, bonding and safety. For example, electroencephalogram (EEG) studies of the brains of subjects who report positive emotions or are shown a funny video clip experience an increase in activity in the left frontal cortex, whereas reports of negative emotions and unpleasant video clips coincide with an increase in activity in the right frontal cortex. In addition, brain scans of subjects who are viewing a photograph of a cute baby show increased activity in the same left frontal cortex area; subjects looking at a photograph of a grotesquely deformed baby show increased activity in the same right frontal cortex area. Finally, bombarding the left frontal cortex of the brain with a strong magnetic field elicits a positive mood in human subjects, and the reverse elicits a negative mood.

Why would the brain show such differences in neural networks associated with emotions? Employing evolutionary theory, I would like to suggest that emotions interact with our cognitive thought processes to guide our behaviors toward the goal of survival and reproduction. University of Southern California neuroscientist Antonio R. Damasio, for example, has demonstrated the vital role that emotions play in decision making. At low levels of stimulation, emotions appear to have an advisory role, interacting with the more reason-oriented cortical regions of the brain. At medium levels of stimulation, conflicts can arise between these cortical reason centers and the brain's deeper and evolutionarily older emotion centers. At high levels of stimulation, emotions can so overrun cortical cognitive processes that people can no longer reason their way to a decision and report feeling "out of control." But why should we have evolved emotions at all?

Emotions are evolutionary proxies for getting us to act in ways that lead to an increase in reproductive success. If we think of the feeling of hunger as a very basic emotion, for example, a little bit of hunger may be perceived as pleasant, motivating us to seek and find food, whereas too much hunger becomes an unpleasant emotion when it goes unmet. In this homeostatic model, emotions act as a feedback mechanism to alert the brain when the body is out of balance. Positive emotions help us build enduring personal resources, such as problem-solving skills, coordination and social resources. Negative emotions, in contrast, help to protect us. Fear causes us to pull back and retreat from risks. Disgust directs us to push out and expel that which is bad for us. Anger leads us to fight back or to signal displeasure at the violation of a social agreement. Jealousy leads us to guard our mates against intruders in pair-bonded relationships.



Such studies indicate that often the evolutionary tail wags the emotional dog. ■

Michael Shermer is publisher of Skeptic *(www.skeptic.com). His new book is* The Mind of the Market.

PHOTOGRAPH BY BRAD SWONETZ; ILLUSTRATION BY MATT COLLINS

Anti Gravity

Attack on the Clones

The fate of the world's favorite fruit

BY STEVE MIRSKY

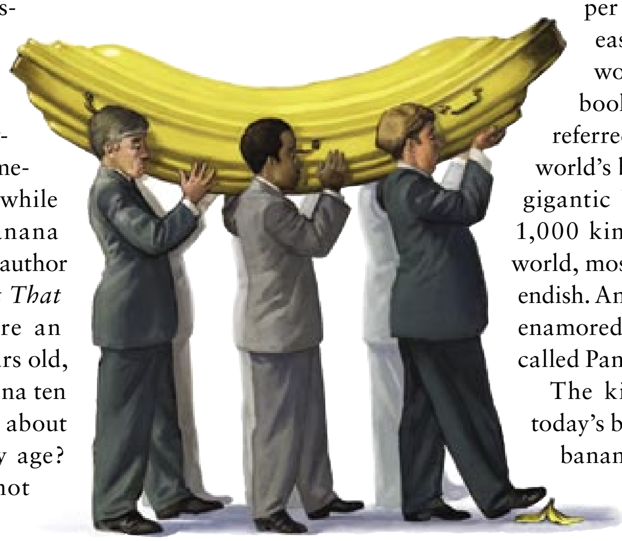
Where would we be without bananas? The silent-movie industry, founded on images of men in bowler hats being launched into the air by banana skins, might never have gotten off the ground, so to speak. Kids would have to pack drippy citrus into their lunch boxes. The band Bananarama could have been the more fetid Apricotarota. When Shakespeare "let slip the dogs of war," what do you think they slipped on?

I am banana-powered. When I was growing up, my daily breakfast carried the official name of "Rice Krispies, banana and milk." Nowadays I often tuck a banana into a pocket on my cycling shirt, for a midride potassium pick-me-up. In fact, I'm taking a short break to eat a banana right now.

Okay, I'm back. (I smeared a little peanut butter on the banana, something that doesn't work that well while biking.) What's my lifetime banana record? According to Dan Koeppel, author of *Banana: The Fate of the Fruit That Changed the World,* "If you are an average American, about forty years old, you're probably approaching banana ten thousand." So I'm probably up to about 15,000 bananas. (Because of my age? Because I'm not average? I'm not telling.)

While researching his book, Koeppel spent a week on a banana plantation in Honduras. This winter I found myself in a similar environment. On January 31, I left a message for myself on my digital voice recorder: "It's really fricking hot." The same heat that wilted me, however, contributed to the healthy development of hundreds of thousands of bananas growing all around me, just to the north of Honduras, on a banana plantation in Quiriguá, Guatemala. I found myself in Central America because I had been invited to speak on a *Scientific American*–sponsored cruise in the Caribbean. (Yes, tough job, someone has to do it.) One of the day trips available to cruisers was to the banana fields. And I wasn't going to say nah to bananas. (The previous day I observed howler monkeys in Belize, so bananas also completed a kind of cartoon symmetry.)

Our guide, Julio Cordova, informed us that this medium-size, 80-acre plantation and packing center fills five container trucks a day. Each truck carries 960 boxes. Each box holds perhaps a dozen hands. (What we call bunches are actually referred to as hands, with each banana a big yellow finger.) In the midst of the plantation, an assembly line of a few dozen workers takes apart huge bunches—the full banana assemblage on the tree—and converts them to the boxed, plastic-wrapped hands that will wind up on your table a week after being harvested. That work, on the "fricking hot day," truly is a tough job that someone has to do.



Some of the banana leaves showed signs of black Sigatoka, a potentially deadly fungus. Koeppel explains, however, that copper sulfate was found to cure the disease (sometimes at the expense of the workers' health). He also shares in his book these banana tidbits: what I just referred to as the banana tree is in fact the world's biggest herb; the fruit is actually a gigantic berry. And although more than 1,000 kinds of bananas exist around the world, most of us eat just one kind—the Cavendish. And the Cavendish, my fellow banana-enamored, is slowly dying. Another fungus, called Panama disease, is coming for it.

The killer has struck before. In fact, today's banana is a blander stand-in for the bananas our grandparents ate, a variety known as the Gros Michel, or "Big Mike." Koeppel explains: "It was larger, with … a creamier texture, and a more intense, fruity taste." But our favorite bananas are all clones of one another. (Notice how delectably seedless they are?) Which means they lack any genetic variability by which some individuals may be lucky enough to ward off a pathogen. Panama disease had wiped out Big Mike worldwide by the 1950s. The Cavendish took over and was thought to be invulnerable. But, Koeppel says, "the Cavendish had never actually been immune to the blight—only to the particular strain of the sickness that destroyed the Gros Michel."

No one wants the song lyric "Yes, we have no bananas" to be prophetic. So we are currently in a race against time to cure the disease, genetically modify the fruit or find a whole new banana variety. Because it is impossible to envision a world lacking a fruit with this kind of appeal. ■

PHOTOGRAPH BY FLYNN LARSEN; ILLUSTRATION BY MATT COLLINS

The Color of Plants on Other

Green aliens are so passé.

On other worlds, plants could be red, blue, even black

BY NANCY Y. KIANG

KEY CONCEPTS

- What color will alien plants be? The question matters scientifically because the surface color of a planet can reveal whether anything lives there—specifically, whether organisms collect energy from the parent star by the process of photosynthesis.
- Photosynthesis is adapted to the spectrum of light that reaches organisms. This spectrum is the result of the parent star's radiation spectrum, combined with the filtering effects of the planet's atmosphere and, for aquatic creatures, of liquid water.
- Light of any color from deep violet through the near-infrared could power photosynthesis. Around stars hotter and bluer than our sun, plants would tend to absorb blue light and could look green to yellow to red. Around cooler stars such as red dwarfs, planets receive less visible light, so plants might try to absorb as much of it as possible, making them look black.

—The Editors

The prospect of finding extraterrestrial life is no longer the domain of science fiction or UFO hunters. Rather than waiting for aliens to come to us, we are looking for them. We may not find technologically advanced civilizations, but we can look for the physical and chemical signs of fundamental life processes: "biosignatures." Beyond the solar system, astronomers have discovered more than 200 worlds orbiting other stars, so-called extrasolar planets. Although we have not been able to tell whether these planets harbor life, it is only a matter of time now. Last July astronomers confirmed the presence of water vapor on an extrasolar planet by observing the passage of starlight through the planet's atmosphere. The world's space agencies are now developing telescopes that will search for signs of life on Earth-size planets by observing the planets' light spectra.

Photosynthesis, in particular, could produce very conspicuous biosignatures. How plausible is it for photosynthesis to arise on another planet? Very. On Earth, the process is so successful that it is the foundation for nearly all life. Although some organisms live off the heat and methane of oceanic hydrothermal vents, the rich ecosystems on the planet's surface all depend on sunlight.

Photosynthetic biosignatures could be of two kinds: biologically generated atmospheric gases such as oxygen and its product, ozone; and surface colors that indicate the presence of specialized pigments such as green chlo-

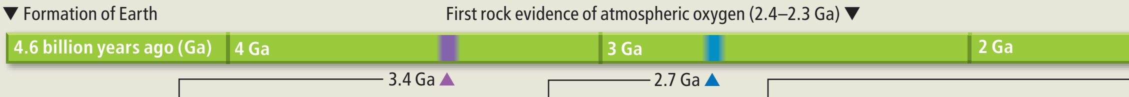
RED EARTH, GREEN EARTH, BLUE EARTH: Type M stars (red dwarfs) are feeble, so plants on an orbiting Earth-like world might need to be black to absorb all the available light (*first panel*). Young M stars fry planetary surfaces with ultraviolet flares, so any organisms must be aquatic (*second*). Our sun is type G (*third*). Around F stars, plants might get too much light and need to reflect much of it (*fourth*).

KENN BROWN AND CHRIS WREN *Mondolithic Studios*

Worlds

[TIMELINE OF PHOTOSYNTHESIS ON EARTH]

Photosynthesis evolved early in Earth's history. The rapidity of its emergence suggests it was no fluke and could arise on other worlds, too. As organisms released gases that changed the very lighting conditions on which they depended, they had to evolve new colors.



First photosynthetic bacteria
They absorbed near-infrared rather than visible light and produced sulfur or sulfate compounds rather than oxygen. Their pigments (possibly bacteriochlorophylls) were predecessors to chlorophyll.

Cyanobacteria
These ubiquitous bacteria were the first oxygen producers. They absorb visible light using a mix of pigments: phycobilins, carotenoids and several forms of chlorophyll.

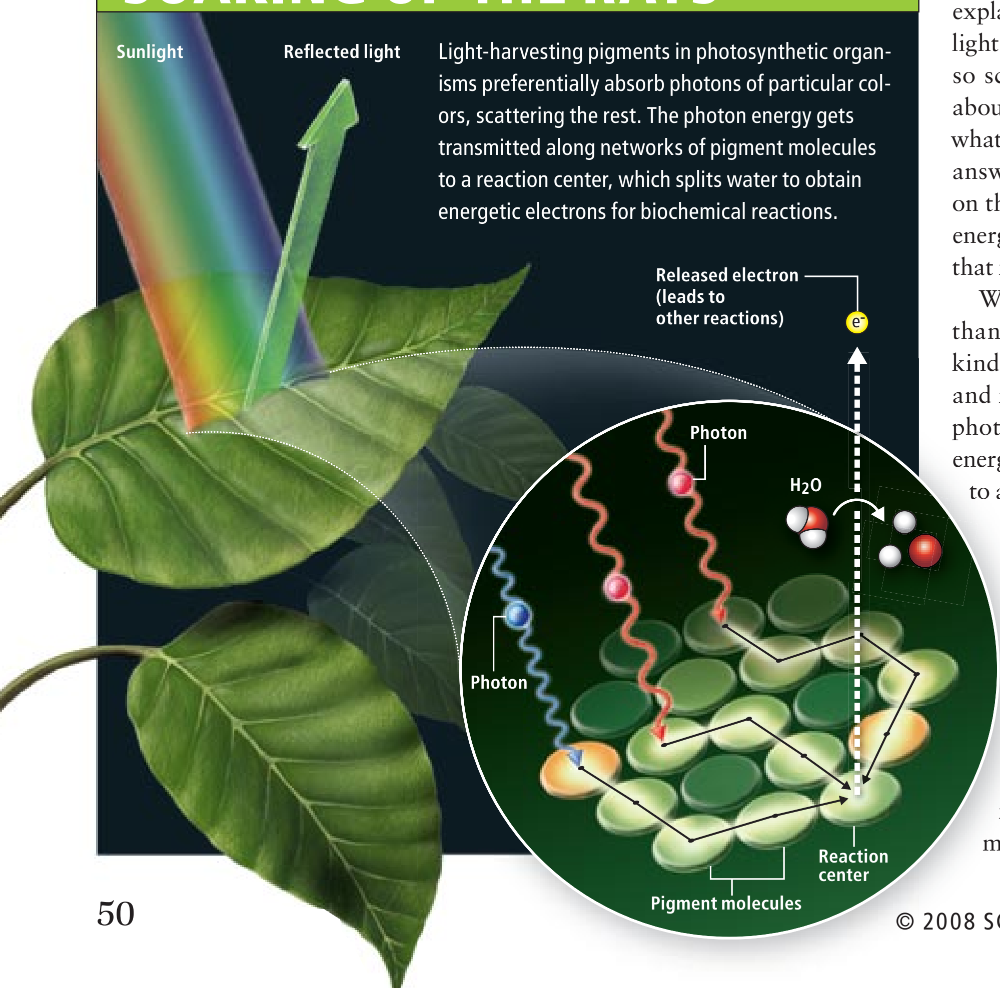
Red and brown algae
These organisms have more complex cellular structures than bacteria do. Like cyanobacteria, they contain phycobilin pigments as well as various forms of chlorophyll.

Green algae
Green algae do better than red and brown algae in the strong light of shallow water. They make do without phycobilins.

rophyll. The idea of looking for such pigments has a long history. A century ago astronomers sought to attribute the seasonal darkening of Mars to the growth of vegetation. They studied the spectrum of light reflected off the surface for signs of green plants. One difficulty with this strategy was evident to writer H. G. Wells, who imagined a different scenario in *The War of the Worlds:* "The vegetable kingdom in Mars, instead of having green for a dominant colour, is of a vivid blood-red tint." Although we now know that Mars has no surface vegetation (the darkening is caused by dust storms), Wells was prescient in speculating that photosynthetic organisms on another planet might not be green.

Even Earth has a diversity of photosynthetic organisms besides green plants. Some land plants have red leaves, and underwater algae and photosynthetic bacteria come in a rainbow of colors. Purple bacteria soak up solar infrared radiation as well as visible light. So what will dominate on another planet? And how will we know when we see it? The answers depend on the details of how alien photosynthesis adapts to light from a parent star of a different type than our sun, filtered through an atmosphere that may not have the same composition as Earth's.

[PHOTOSYNTHESIS 101]
SOAKING UP THE RAYS



Light-harvesting pigments in photosynthetic organisms preferentially absorb photons of particular colors, scattering the rest. The photon energy gets transmitted along networks of pigment molecules to a reaction center, which splits water to obtain energetic electrons for biochemical reactions.

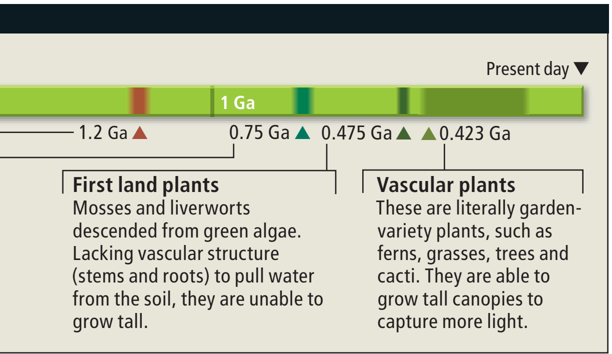
LISA APFELBACHER

Harvesting Light

In trying to figure out how photosynthesis might operate on other planets, the first step is to explain it on Earth. The energy spectrum of sunlight at Earth's surface peaks in the blue-green, so scientists have long scratched their heads about why plants reflect green, thereby wasting what appears to be the best available light. The answer is that photosynthesis does not depend on the total amount of light energy but on the energy per photon and the number of photons that make up the light.

Whereas blue photons carry more energy than red ones, the sun emits more of the red kind. Plants use blue photons for their quality and red photons for their quantity. The green photons that lie in between have neither the energy nor the numbers, so plants have adapted to absorb fewer of them.

The basic photosynthetic process, which fixes one carbon atom (obtained from carbon dioxide, CO_2) into a simple sugar molecule, requires a minimum of eight photons. It takes one photon to split an oxygen-hydrogen bond in water (H_2O) and thereby to obtain an electron for biochemical reactions. A total of four such bonds must be broken to create an oxygen molecule (O_2). Each of those photons is matched by at least one additional photon for



a second type of reaction to form the sugar. Each photon must have a minimum amount of energy to drive the reactions.

The way plants harvest sunlight is a marvel of nature. Photosynthetic pigments such as chlorophyll are not isolated molecules. They operate in a network like an array of antennas, each tuned to pick out photons of particular wavelengths. Chlorophyll preferentially absorbs red and blue light, and carotenoid pigments (which produce the vibrant reds and yellows of fall foliage) pick up a slightly different shade of blue. All this energy gets funneled to a special chlorophyll molecule at a chemical reaction center, which splits water and releases oxygen.

The funneling process is the key to which colors the pigments select. The complex of molecules at the reaction center can perform chemical reactions only if it receives a red photon or the equivalent amount of energy in some other form. To take advantage of blue photons, the antenna pigments work in concert to convert the high energy (from blue photons) to a lower energy (redder), like a series of step-down transformers that reduces the 100,000 volts of electric power lines to the 120 or 240 volts of a wall outlet. The process begins when a blue photon hits a blue-absorbing pigment and energizes one of the electrons in the molecule. When that electron drops back down to its original state, it releases this energy—but because of energy losses to heat and vibrations, it releases less energy than it absorbed.

The pigment molecule releases its energy not in the form of another photon but in the form of an electrical interaction with another pigment molecule that is able to absorb energy at that lower level. This pigment, in turn, releases an even lower amount of energy, and so the process continues until the original blue photon energy has been downgraded to red. The array of pigments can also convert cyan, green or yellow to red. The reaction center, as the receiving end of the cascade, adapts to absorb the lowest-energy available photons. On our planet's surface, red photons are both the most abundant and the lowest energy within the visible spectrum.

For underwater photosynthesizers, red photons are not necessarily the most abundant. Light niches change with depth because of filtering of light by water, by dissolved substances and by overlying organisms themselves. The result is a clear stratification of life-forms according to their mix of pigments. Organisms in lower water layers have pigments adapted to absorb the light colors left over by the layers above. For instance, algae and cyanobacteria have pigments known as phycobilins that harvest green and yellow photons. Nonoxygen-producing (anoxygenic) bacteria have bacteriochlorophylls that absorb far-red and near-infrared light, which is all that penetrates to the murky depths.

Organisms adapted to low-light conditions tend to be slower-growing, because they have to put more effort into harvesting whatever light is available to them. At the planet's surface, where light is abundant, it would be disadvantageous for plants to manufacture extra pigments, so they are selective in their use of color. The same evolutionary principles would operate on other worlds.

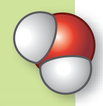
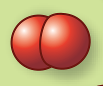
Just as aquatic creatures have adapted to light filtered by water, land dwellers have adapted to light filtered by atmospheric gases. At the top of

[THE AUTHOR]

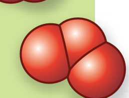
Nancy Y. Kiang is a biometeorologist at the NASA Goddard Institute for Space Studies in New York City. She specializes in computer simulations of interactions between ecosystems and the atmosphere, which help to regulate climate. In addition, she is a member of the Virtual Planetary Laboratory, a team of the NASA Astrobiology Institute studying how to detect life on other worlds. Kiang is also an independent filmmaker; her short film, *Solidarity,* has been making the festival rounds.

Biosignatures

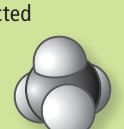
Aside from colors reflected by plants, these other features could be signs of life:

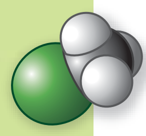
Oxygen (O_2) plus water (H_2O). Even on a lifeless world, light from the parent star naturally produces a small amount of oxygen in a planet's atmosphere by splitting water vapor. But the gas is quickly rained out, as well as consumed through oxidation of rocks and volcanic gases. Therefore, if a planet with liquid water has abundant oxygen, some additional source must be producing the gas. Oxygenic photosynthesis is the leading candidate.



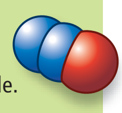
Ozone (O_3). In Earth's stratosphere, radiation splits apart oxygen, which then recombines to form ozone. Together with liquid water, ozone is a strong biosignature. Whereas oxygen can be detected at visible wavelengths, ozone can be detected at infrared wavelengths, which is easier for some telescopes.



Methane (CH_4) plus oxygen or seasonal cycles. Oxygen and methane are an awkward chemical combination that is hard to achieve without photosynthesis. A seasonal cycle of rising and falling methane concentrations is also a good sign of life. On a dead planet, methane levels are fairly constant, declining slightly over the long run as starlight splits the molecules.



Methyl chloride (CH_3Cl). On Earth this gas results from the burning of vegetation (mainly forest fires) and from the action of sunlight on plankton and seawater chlorine. Oxidation destroys it. But an M star's relatively weak radiation might allow the gas to build up to detectable amounts.

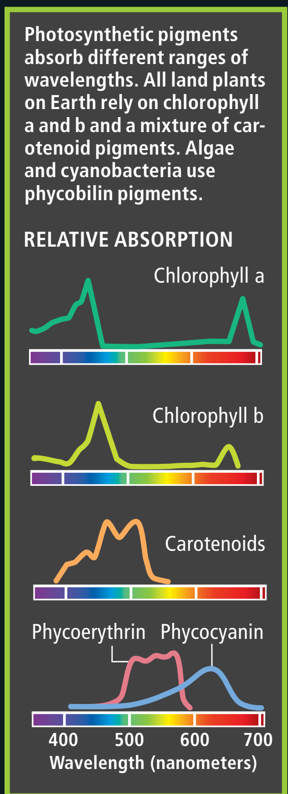


Nitrous oxide (N_2O). When plant matter decays, it releases nitrogen in the form of nitrous oxide. Abiotic sources of this gas, such as lightning, are negligible.

WALLACE ZANE (*Kiang*); LISA APFELBACHER (*molecules*)

Filtering Starlight

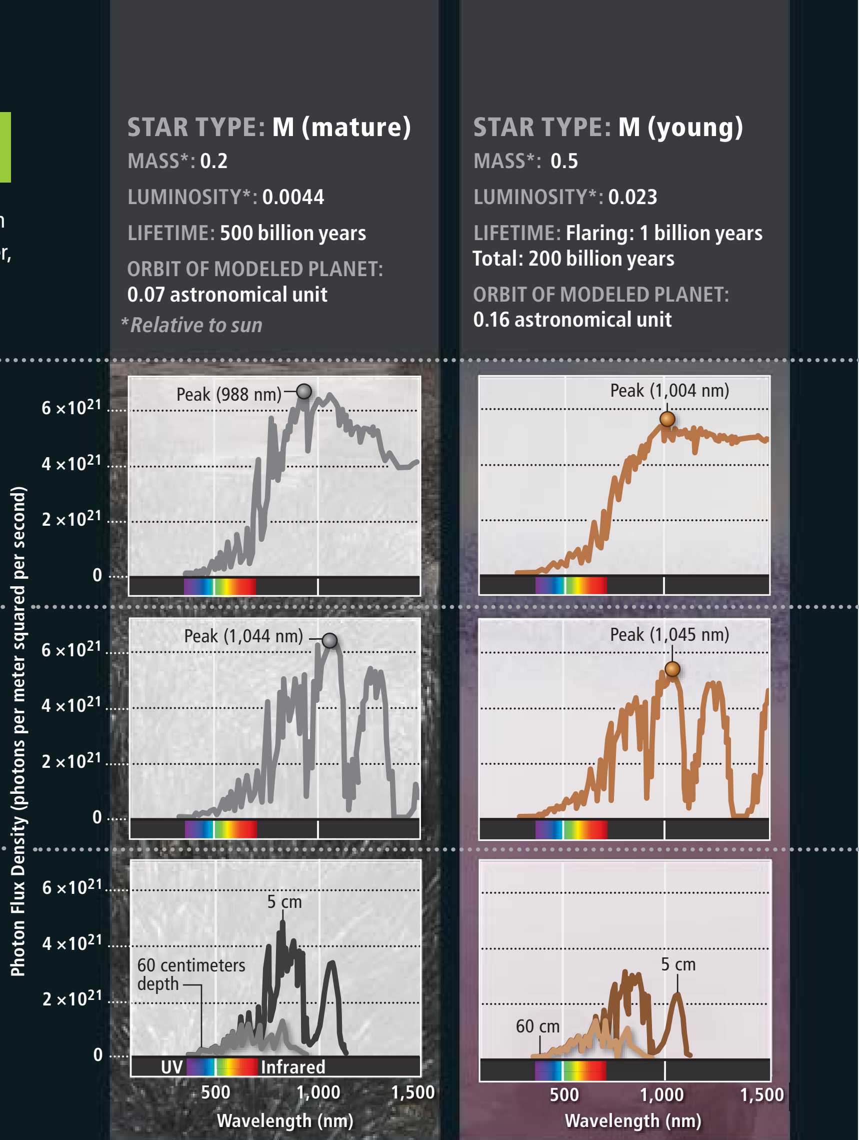
The color of plants depends on the spectrum of the star's light, which astronomers can easily observe, and filtering of light by air and water, which the author and her colleagues have simulated based on the likely atmospheric composition and life's own effects.



STARLIGHT
Before entering the atmosphere, starlight has a distinctive spectrum. The overall shape is determined by the surface temperature of the star, with a few dips produced by absorption in the star's own atmosphere.

SURFACE
Atmospheric gases absorb the starlight unevenly, shifting its peak color and introducing absorption bands—wavelengths that are screened out. These bands are best known for Earth (the G-star case).

UNDERWATER
Water tends to transmit blue light and absorb red and infrared light. The graphs shown here are for water depths of five and 60 centimeters. (The mature M-star case is for a low-oxygen atmosphere.)



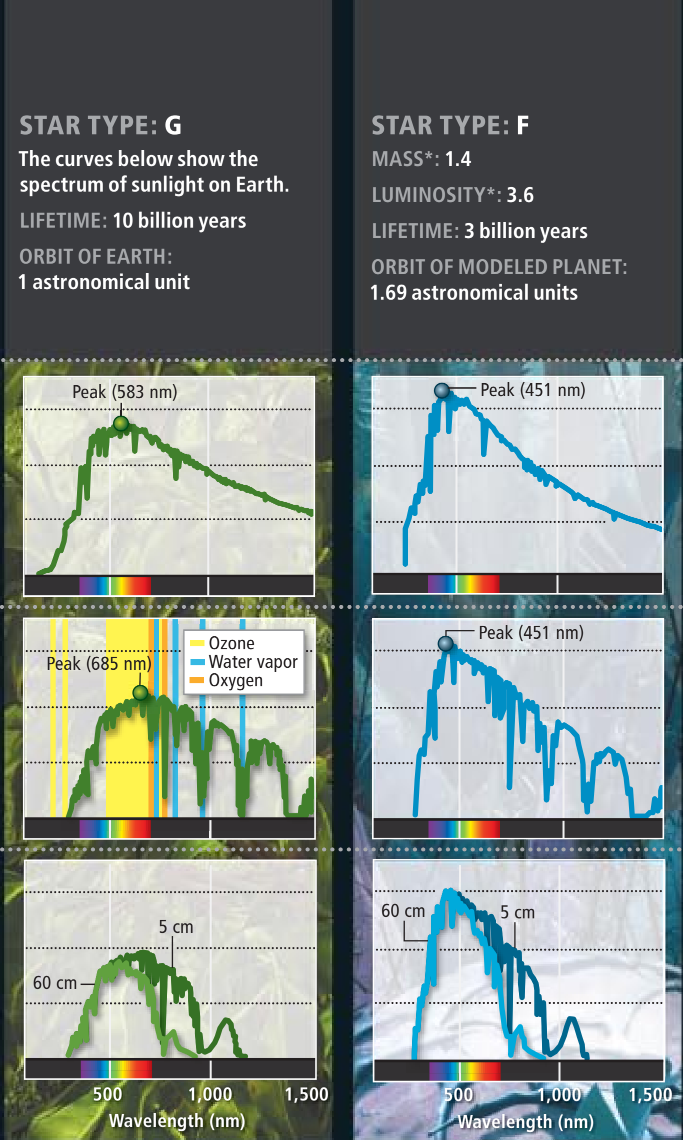
Earth's atmosphere, yellow photons (at wavelengths of 560 to 590 nanometers) are the most abundant kind. The number of photons drops off gradually with longer wavelength and steeply with shorter wavelength. As sunlight passes through the upper atmosphere, water vapor absorbs the infrared light in several wavelength ands beyond 700 nm. Oxygen produces absorption lines—narrow ranges of wavelengths that the gas blocks—at 687 and 761 nm. We all know that ozone (O_3) in the stratosphere strongly absorbs the ultraviolet (UV). Less well known is that it also absorbs weakly across the visible range.

Putting it all together, our atmosphere demarcates windows through which radiation can make it to the planet's surface. The visible radiation window is defined at its blue edge by the drop-off in the intensity of short-wavelength photons emitted by the sun and by ozone absorption of UV. The red edge is defined by oxygen absorption lines. The peak in photon abundance is shifted from yellow to red (about 685 nm) by ozone's broad absorbance across the visible.

Plants are adapted to this spectrum, which is determined largely by oxygen—yet plants are what put the oxygen into the atmosphere to begin with. When early photosynthetic organisms first appeared on Earth, the atmosphere lacked oxygen, so they must have used different pigments from chlorophyll. Only over time, as photosynthesis altered the atmospheric composition, did chlorophyll emerge as optimal.

The firm fossil evidence for photosynthesis dates to about 3.4 billion years ago (Ga), but earlier fossils exhibit signs of what could have been photosynthesis. Early photosynthesizers had to start out underwater, in part because water is a good solvent for biochemical reactions and in part because it provides protection against solar UV radiation—shielding that was essential in the absence of an atmospheric ozone layer. These earliest photosynthesizers were underwater bacteria that absorbed infrared photons. Their chemical reactions involved hydrogen, hydrogen sulfide or iron rather than water, so they did not produce oxygen gas. Oxygen-generating (oxy-

JEN CHRISTIANSEN



genic) photosynthesis by cyanobacteria in the oceans started 2.7 Ga. Oxygen levels and the ozone layer slowly built up, allowing red and brown algae to emerge. As shallower water became safe from UV, green algae evolved. They lacked phycobilins and were better adapted to the bright light in surface waters. Finally, plants descended from green algae emerged onto land—two billion years after oxygen had begun accumulating in the atmosphere.

And then the complexity of plant life exploded, from mosses and liverworts on the ground to vascular plants with tall canopies that capture more light and have special adaptations to particular climates. Conifer trees have conical crowns that capture light efficiently at high latitudes with low sun angles; shade-adapted plants have anthocyanin as a sunscreen against too much light. Green chlorophyll not only is well suited to the present composition of the atmosphere but also helps to sustain that composition—a virtuous cycle that keeps our planet green. It may be that another step of evolution will favor an organism that takes advantage of the shade underneath tree canopies, using the phycobilins that absorb green and yellow light. But the organisms on top are still likely to stay green.

Predicting alien plant colors takes experts ranging from astronomers to plant physiologists to biochemists.

Painting the World Red

To look for photosynthetic pigments on another planet in another solar system, astronomers must be prepared to see the planet at any of the possible stages in its evolution. For instance, they may catch sight of a planet that looks like our Earth two billion years ago. They must also allow that extrasolar photosynthesizers may have evolved capabilities that their counterparts here have not, such as splitting water using longer-wavelength photons.

The longest wavelength yet observed in photosynthesis on Earth is about 1,015 nm (in the infrared), in purple anoxygenic bacteria. The longest wavelength observed for oxygenic photosynthesis is about 720 nm, in a marine cyanobacterium. But the laws of physics set no strict upper limit. A large number of long-wavelength photons could achieve the same purpose as a few short-wavelength ones.

The limiting factor is not the feasibility of novel pigments but the light spectrum available at a planet's surface, which depends mainly on the star type. Astronomers classify stars based on color, which relates to temperature, size and longevity. Only certain types are long-lived enough to allow for complex life to evolve. These are, in order from hottest to coolest, F, G, K and M stars. Our sun is a G star. F stars are larger, burn brighter and bluer, and take a couple of billion years to use up their fuel. K and M stars are smaller, dimmer, redder and longer-lived.

Around each of these stars is a habitable zone, a range of orbits where planets can maintain a temperature that allows for liquid water. In our solar system, the habitable zone is a ring encompassing Earth's and Mars's orbits. For an F star, the habitable zone for an Earth-size planet is farther out; for a K or M star, closer in. A planet in the habitable zone of an F or K star receives about as much visible radiation as Earth does. Such a planet could easily support oxygenic photosynthesis like that on Earth. The pigment color may simply be shifted within the visible band.

M stars, also known as red dwarfs, are of special interest because they are the most abundant type in our galaxy. They emit much less visible radiation than our sun; their output peaks in the near-infrared. John Raven, a biologist at the University of Dundee in Scotland, and Ray Wolstencroft, an astronomer at the Royal Observatory, Edinburgh, have proposed that oxygenic photosynthesis is theoretically possible with near-infrared photons. An organism would have to use three or four near-infrared photons to split H_2O, rather than the two that suffice for Earth's plants. The photons work together like stages of a rocket to provide the necessary energy to an electron as it performs the chemical reactions.

M stars pose an extra challenge to life: when young, they emit strong UV flares. Organisms could avoid the damaging UV radiation deep underwater, but would they then be starved for light? If so, photosynthesis might not arise. As M stars age, though, they cease producing flares, at which point they give off even less UV radiation than our sun does. Organisms would not need a UV-absorbing ozone layer to protect them; they could thrive on land even if they did not produce oxygen.

In sum, astronomers must consider four scenarios depending on the age and type of star:

Anaerobic, ocean life. The parent star is a young star of any type. Organisms do not necessarily produce oxygen; the atmosphere may be mostly other gases such as methane.

Aerobic, ocean life. The parent star is an older star of any type. Enough time has elapsed for oxygenic photosynthesis to evolve and begin to build up atmospheric oxygen.

Aerobic, land life. The parent star is a mature star of any type. Plants cover the land. Life on Earth is now at this stage.

Anaerobic, land life. The star is a quiescent M star, so the UV radiation is negligible. Plants cover the land but may not produce oxygen.

Photosynthetic biosignatures for these different cases would clearly not be the same. From experience with satellite imagery of Earth, astronomers expect that any life in the ocean would be too sparsely distributed for telescopes to see. So the first two scenarios would not produce strong pigment biosignatures; life would reveal itself to us only by the atmospheric gases it produced. Therefore, researchers studying alien plant colors focus on land plants, either on planets around F, G and K stars with oxygenic photosynthesis or on planets around M stars with any type of photosynthesis.

Black Is the New Green

Regardless of the specific situation, photosynthetic pigments must still satisfy the same rules as on Earth: pigments tend to absorb photons that are either the most abundant, the shortest available wavelength (most energetic) or the longest available wavelength (where the reaction center absorbs). To tackle the question of how star type determines plant color, it took researchers from many disciplines to put together all the stellar, planetary and biological pieces.

Martin Cohen, a stellar astronomer at the University of California, Berkeley, collected data for an F star (sigma Bootis), a K star (epsilon Eridani), an actively flaring M star (AD Leo), and a hypothetical quiescent M star with a temperature of 3,100 kelvins. Antígona Segura, an astronomer at the National Autonomous University of Mexico, ran computer simulations of Earth-like planets in the habitable zone of these stars. Using models developed by Alexander Pavlov, now at the University of Arizona, and James Kasting of Pennsylvania State University, Segura studied the interaction between the stellar radiation and the atmosphere's likely constituents (assuming that volcanoes on these worlds emit the same gases they do on Earth) to deduce the planets' atmospheric chemistry, both for negligible oxygen and for Earth-like oxygen levels.

Using Segura's results, Giovanna Tinetti, a physicist at University College London, calculated the filtering of radiation by applying a

Plants on worlds around dim stars may need to harvest the full range of visible and infrared light. They might look black to our eyes.

DETAIL OF F-STAR FOLIAGE

KENN BROWN AND CHRIS WREN *Mondolithic Studios* model developed by David Crisp of the Jet Propulsion Laboratory in Pasadena, Calif. (This is one of the models enlisted to calculate how much light reaches the solar panels of the Mars rovers.) Interpreting these calculations required the combined knowledge of five of us: microbial biologist Janet Siefert of Rice University, biochemists Robert Blankenship of Washington University in St. Louis and Govindjee of the University of Illinois at Urbana-Champaign, planetary scientist Victoria Meadows of the University of Washington, and me, a biometeorologist at the NASA Goddard Institute for Space Studies.

We found that the photons reaching the surface of planets around F stars tend to be blue, with the greatest abundance at 451 nm. Around K stars, the peak is in the red at 667 nm, nearly the same as on Earth. Ozone plays a strong role, making the F starlight bluer than it otherwise would be and the K starlight redder. The useful radiation for photosynthesis would be in the visible range, as on Earth.

Thus, plants on both F- and K-star planets could have colors just like those on Earth but with subtle variations. For F stars, the flood of energetic blue photons is so intense that plants might need to reflect it using a screening pigment similar to anthocyanin, giving them a blue tint. Alternatively, plants might need to harvest only the blue, discarding the lower-quality green through red light. That would produce a distinctive blue edge in the spectrum of reflected light, which would stand out to telescope observers.

The range of M-star temperatures makes possible a very wide variation in alien plant colors. A planet around a quiescent M star would receive about half the energy that Earth receives from our sun. Although that is plenty for living things to harvest—about 60 times more than the minimum needed for shade-adapted Earth plants—most of the photons are near-infrared. Evolution might favor a greater variety of photosynthetic pigments to pick out the full range of visible and infrared light. With little light reflected, plants might even look black to our eyes.

Pale Purple Dot

The experience of life on Earth indicates that early ocean photosynthesizers on planets around F, G and K stars could survive the initial oxygen-free atmosphere and develop the oxygenic photosynthesis that would lead ultimately to land plants. For M stars, the situation is trickier. We calculated a "sweet spot" about nine meters underwater where early photosynthesizers could both survive UV flares and still have enough light to be productive. Although we might not see them through telescopes, these organisms could set the stage for life at the planet's surface. On worlds around M stars, land plants that exploited a wider range of colors would be nearly as productive as plants on Earth.

For all star types, an important question will be whether a planet's land area is large enough for upcoming space telescopes to see. The first generation of these telescopes will see the planet as a single dot; they will lack the resolution to make maps of the surface. All scientists will have is a globally averaged spectrum. Tinetti calculates that for land plants to show up in this spectrum, at least 20 percent of the surface must be land that is both covered in vegetation and free from clouds. On the other hand, oceanic photosynthesis releases more oxygen to the atmosphere. Therefore, the more prominent the pigment biosignature, the weaker the oxygen biosignature, and vice versa. Astronomers might see one or the other, but not both.

If a space telescope sees a dark band in a planet's reflected light spectrum at one of the predicted colors, then someone monitoring the observations from a computer may be the first person to see signs of life on another world. Other false interpretations have to be ruled out, of course, such as whether minerals could have the same signature. Right now we can identify a plausible palette of colors that indicate plant life on another planet; for instance, we predict another Earth to have green, yellow or orange plants. But it is currently hard to make finer predictions. On Earth, we have been able to determine that the signature of chlorophyll is unique to plants, which is why we can detect plants and ocean phytoplankton with satellites. We will have to figure out unique signatures of vegetation for other planets.

Finding life on other planets—abundant life, not just fossils or microbes eking out a meager living under extreme conditions—is a fast-approaching reality. Which stars shall we target, given there are so many out there? Will we be able to measure the spectra of M-star planets, which tend to be very close to their stars? What wavelength range and resolution do the new telescopes need? Our understanding of photosynthesis will be key to designing these missions and interpreting their data. Such questions drive a synthesis of the sciences in a way that is only beginning. Our very ability to search for life elsewhere in the universe ultimately requires our deepest understanding of life here on Earth. ■

PLANET FINDERS

The European Space Agency (ESA) plans to launch Darwin in about a decade to measure the spectra of Earth-size extrasolar planets. NASA's Terrestrial Planet Finder would do the same, when the agency can fund it. ESA's COROT, launched in December 2006, and NASA's Kepler, scheduled for 2009, seek the slight dimming caused by Earth-like planets as they pass in front of their stars. NASA's SIM PlanetQuest would look for a telltale wobbling of the star.

TERRESTRIAL PLANET FINDER

COURTESY OF NASA/JPL/CALTECH

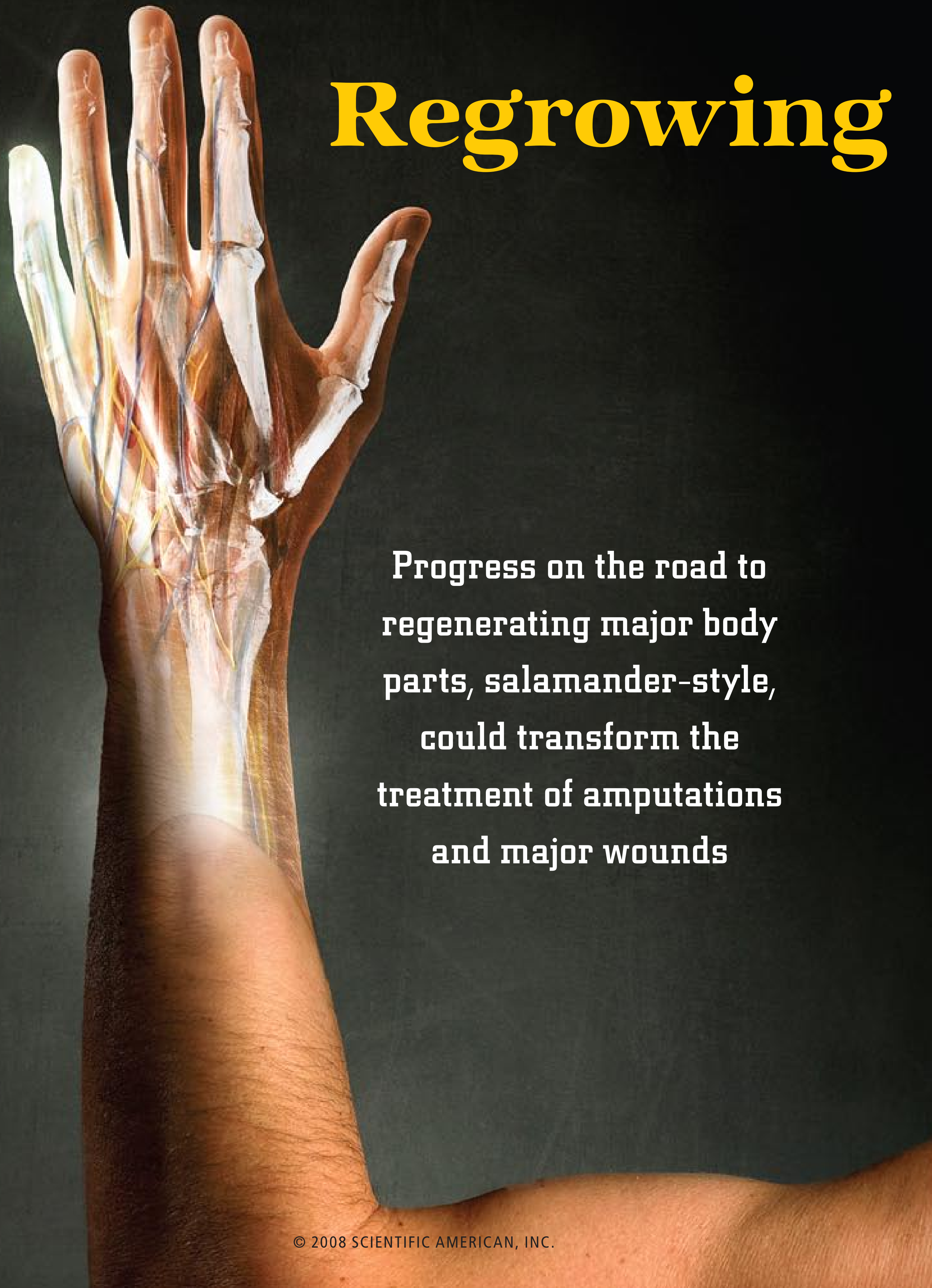
➡ MORE TO EXPLORE

Spectral Signatures of Photosynthesis II: Coevolution with Other Stars and the Atmosphere on Extrasolar Worlds. Nancy Y. Kiang, Antígona Segura, Giovanna Tinetti, Govindjee, Robert E. Blankenship, Martin Cohen, Janet Siefert, David Crisp and Victoria S. Meadows in *Astrobiology,* Special Issue on M Stars, Vol. 7, No. 1, pages 252–274; February 1, 2007. **http://pubs.giss.NASA.gov/docs/2007/2007_Kiang_etal_2.pdf**

Water Vapour in the Atmosphere of a Transiting Extrasolar Planet. Giovanna Tinetti, Alfred Vidal-Madjar, Mao-Chang Liang, Jean-Philippe Beaulieu, Yuk Yung, Sean Carey, Robert J. Barber, Jonathan Tennyson, Ignasi Ribas, Nicole Allard, Gilda E. Ballester, David K. Sing and Franck Selsis in *Nature,* Vol. 448, pages 169–171; July 12, 2007. **www.arxiv.org/abs/0707.3064**

Virtual Planetary Laboratory: **http://vpl.astro.washington.edu**

Astrobiology magazine: **www.astrobio.net**



Regrowing

Progress on the road to regenerating major body parts, salamander-style, could transform the treatment of amputations and major wounds

Human Limbs

By Ken Muneoka, Manjong Han and David M. Gardiner

A salamander's limbs are smaller and a bit slimier than those of most people, but otherwise they are not that different from their human counterparts. The salamander limb is encased in skin, and inside it is composed of a bony skeleton, muscles, ligaments, tendons, nerves and blood vessels. A loose arrangement of cells called fibroblasts holds all these internal tissues together and gives the limb its shape.

Yet a salamander's limb is unique in the world of vertebrates in that it can regrow from a stump after an amputation. An adult salamander can regenerate a lost arm or leg this way over and over again, regardless of how many times the part is amputated. Frogs can rebuild a limb during tadpole stages when their limbs are first growing out, but they lose this ability in adulthood. Even mammalian embryos have some ability to replace developing limb buds, but that capacity also disappears well before birth. Indeed, this trend toward declining regenerative capacity over the course of an organism's development is mirrored in the evolution of higher animal forms, leaving the lowly salamander as the only vertebrate still able to regrow complex body parts throughout its lifetime.

Humans have long wondered how the salamander pulls off this feat. How does the regrowing part of the limb "know" how much limb is missing and needs to be replaced? Why doesn't the skin at the stump form a scar to seal off the wound as it would in humans? How can adult salamander tissue retain the embryonic potential to build an entire limb from scratch multiple times? Biologists are closing in on the answers to those questions. And if we can understand how the regeneration process works in nature, we hope to be able to trigger it in people to regenerate amputated limbs, for example, and transform the healing of other major wounds.

The human body's initial responses to such a serious injury are not that different from those of a salamander, but soon afterward the human and amphibian wound-healing strategies diverge. Ours results in a scar and amounts to a failed regeneration response, but several signs indicate that humans do have the potential to rebuild complex parts. The key to making that happen will be tapping into our latent abilities so that our own wound healing becomes more salamanderlike. For this reason, our research first focused on the experts to learn how it is done.

Lessons from the Salamander

When the tiny salamander limb is amputated, blood vessels in the remaining stump contract quickly, so bleeding is limited, and a layer of skin cells rapidly covers the surface of the amputation site. During the first few days after injury, this so-called wound epidermis transforms into a layer of signaling cells called the apical epithelial cap (AEC), which is indispensable for successful regeneration. In the meantime, fibroblasts break free from the connective tissue meshwork and migrate across the amputation surface to meet at the center of the wound. There they proliferate to form a blastema—an aggregation of stemlike cells that will serve as progenitors for the new limb [*see box on next two pages*].

Many years ago studies in the laboratory of our colleague Susan V. Bryant at the University of California, Irvine, demonstrated that the cells in the blastema are equivalent to the cells in the developing limb bud of the salamander embryo. This discovery suggested that the construction of a limb by the blastema is essentially a recapitulation of the limb formation that took place during the animal's original development. An important implication of this insight was that the same genetic program is involved in both situations, and because humans make limbs as embryos, in principle we should already have the necessary programming to regenerate them as adults, too. It seemed, therefore, that all scientists needed to do was figure out how to

KEY CONCEPTS

- The gold standard for limb regeneration is the salamander, which can grow perfect replacements for lost body parts throughout its lifetime. Understanding how can provide a road map for human limb regeneration.
- The early responses of tissues at an amputation site are not that different in salamanders and in humans, but eventually human tissues form a scar, whereas the salamander's reactivate an embryonic development program to build a new limb.
- Learning to control the human wound environment to trigger salamanderlike healing could make it possible to regenerate large body parts.

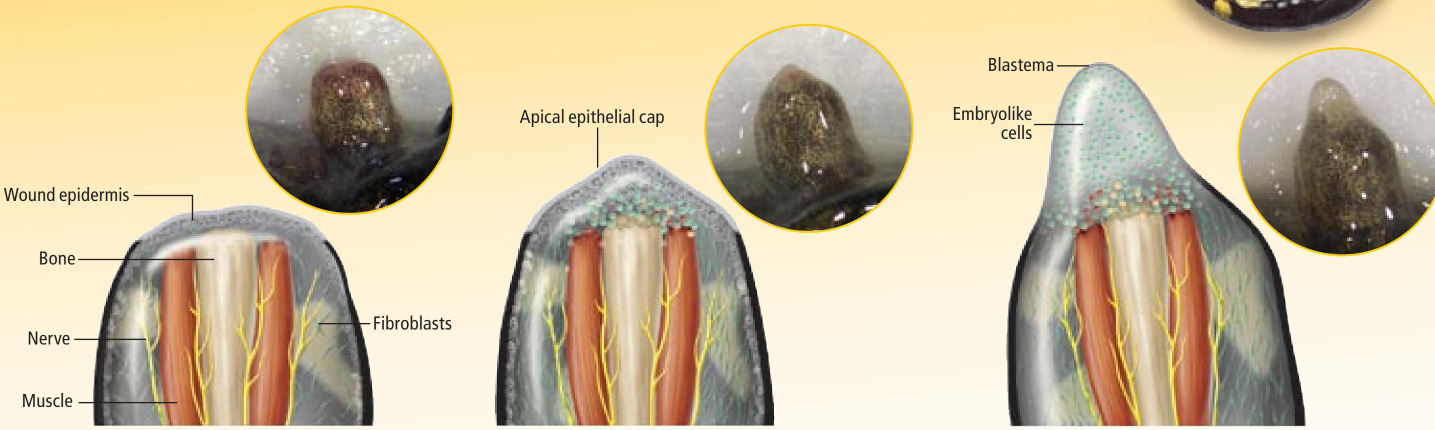
—The Editors

AARON GOODMAN

[THE GOLD STANDARD]

PERFECT REGENERATION

Salamanders are the only vertebrates able to regrow lost limbs, as well as many other body parts, throughout their lifetimes—and they can do it repeatedly. Studies of how a limb forms on the salamander have revealed that the process begins with rapid wound closure and a rush of cells from stump tissues to the amputation site. The next stages involve reversion of those cells to an embryonic state and their building of a new limb following the same steps as in embryonic development.



WOUND CLOSURE
Within hours of a leg amputation, epidermal skin cells migrate across the wound to seal it, forming a wound epidermis.

HEALING SIGNALS
Epidermal cells form a ridge known as an apical epithelial cap, which generates critical signals that guide the behavior of other cells. Fibroblasts and muscle cells start migrating toward the wound site.

RETURN TO THE WOMB
Cells that migrated to the wound revert to a less specialized embryonic state and begin dividing to populate the bud of a new limb, called a blastema.

induce an amputated limb to form a blastema.

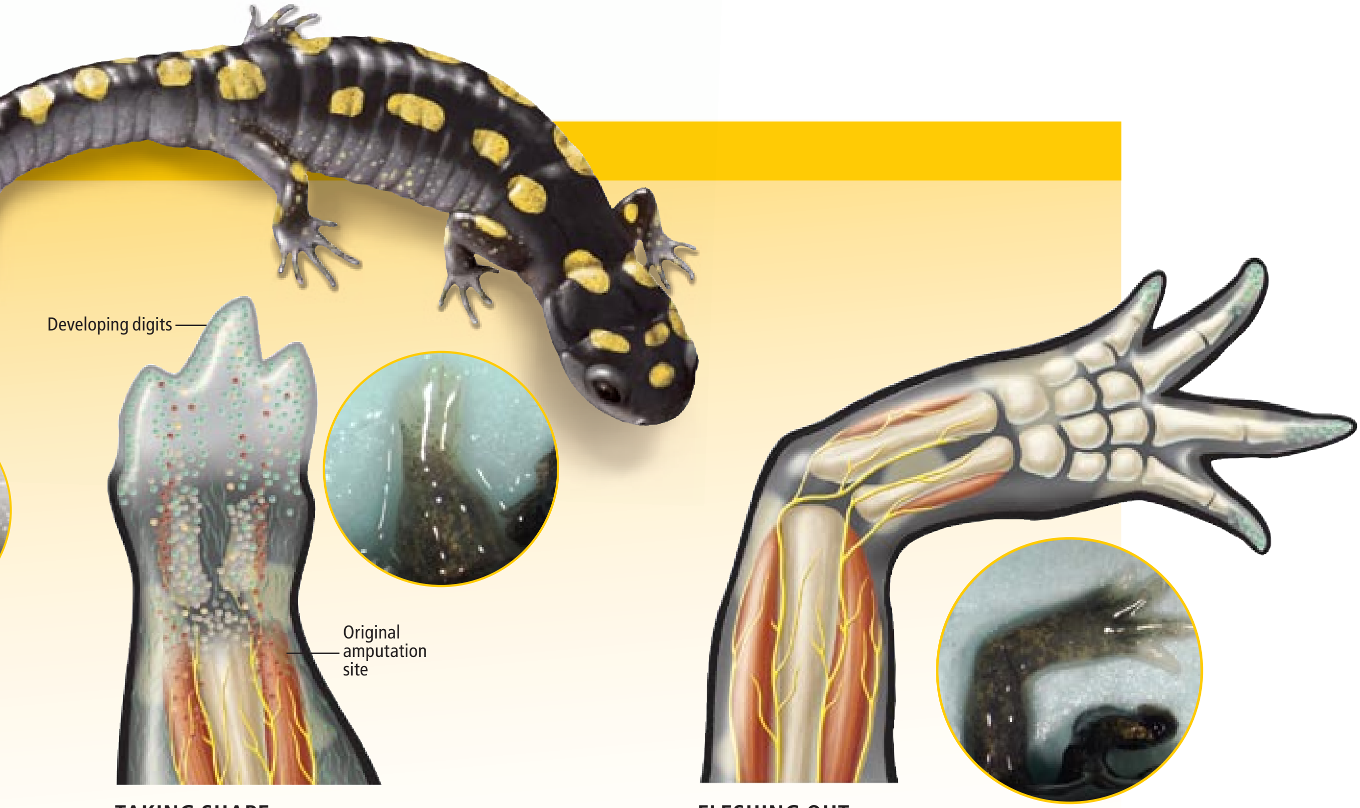
One of us (Gardiner)—working with Tetsuya Endo of U.C. Irvine a few years ago—took a minimalist approach to answering the basic question of how to make a blastema. Instead of studying amputation sites on the salamander, where a blastema would naturally form, we looked at simple wounds on the side of a salamander limb, which would normally heal just by regenerating the skin. Our idea was that such wounds are similar to the site of an amputated mammalian limb that fails to generate a new limb. If we could get an entire limb to grow where a simple wound-healing response would typically occur, then we could further dissect the regeneration process.

After we made a small incision in the salamander leg, epidermal cells migrated to cover and seal the wound, as they would at an amputation site, and fibroblasts from the dermis layer of the skin also moved in to replace the missing skin. But if we carefully deviated a nerve to the wound site, we could induce those fibroblasts to form a blastema instead. Marcus Singer of Case Western Reserve University had already demonstrated more than half a century ago that innervation was required for a regeneration response, but our experiments clarified that unknown factors provided by the nerve were influencing regeneration by altering the behavior of resident fibroblasts.

These induced blastemas never progressed to the later stages of regeneration to form a new limb, however. One more ingredient was needed. The key to inducing a blastema that produced a new limb was to graft a piece of skin from the opposite side of the limb to the wound site, which allowed fibroblasts from opposite regions of the limb to participate in the healing response. The resulting accessory limb was, of course, growing out at an abnormal location, but it was anatomically normal [*see box on pages 62 and 63*]. So the basic recipe for making a blastema seemed relatively simple: you need a wound epidermis, nerves and fibroblasts from opposite sides of the limb. With this minimal view of limb regeneration in mind, we began to focus on understanding the roles of the individual ingredients.

We knew that the epidermis is derived from one of three layers of primitive cells within an early developing embryo, the ectoderm, which is also well known to provide signals that control the outgrowth of limbs from limb buds on the embryo. Ectoderm cells gather in the bud to form an apical ectodermal ridge (AER), which transiently produces chemical signals that guide

DAVID M. GARDINER (*photographs*); ALICE Y. CHEN (*limb illustrations*); LISA APFELBACHER (*salamander*)



TAKING SHAPE
As the blastema grows, it begins to form the outline of the new limb, including the tip that will become the foot. The embryonic cells give rise to new tissues by proliferating and differentiating into bone, muscle, fibroblasts, and so on.

FLESHING OUT
As its internal anatomy and outline become more mature, the limb lengthens to fill in the missing segment between the original amputation plane and the toes.

the migration and proliferation of the underlying limb bud cells.

Although some of the critical signals from the epidermis have not yet been identified, members of the family of fibroblast growth factors (FGFs) are involved. The AER produces a number of FGFs that stimulate the underlying cells of the limb bud to produce other FGFs, fueling a feedback circuit of signaling between the AER and limb bud cells that is essential for the outgrowth of a limb. A similar feedback circuit spurred by the AEC is thought to function in the same way during limb regeneration, and Hiroyuki Ide of Tohoku University in Japan discovered that the progressive loss of regenerative ability in frog tadpoles is associated with a failure to activate the FGF circuit. By treating older nonregenerating tadpole limbs with FGF10, he was able to jump-start this signaling circuit and stimulate partial regeneration of amputated limbs.

The excitement this result inspired was tempered, however, by the fact that the induced regenerates were abnormal, consisting of irregularly placed limb parts, which raises the important issue of how regeneration is controlled so that all the appropriate anatomical structures that are lost when the limb is amputated are accurately replaced. It turns out that the other primary cellular players, the fibroblasts, carry out this function.

Location, Location, Location

Recall from the minimalist accessory-limb experiments that the presence of fibroblasts *per se* was not sufficient for regeneration because fibroblasts are present at the simple wound site that does not make a new limb. It was the fibroblasts from the opposite side of the limb that proved essential. That discovery illustrates the importance of cellular position in triggering a regeneration response. In an embryo, the sequence of events in limb development always begins with formation of the base of the limb (the shoulder or hip) and is followed by progressive building of more distal structures until the process terminates with the making of fingers or toes. In salamander regeneration, on the other hand (or foot), the site of amputation can be anywhere along the length of the limb and regardless of where the wound is located, only those parts of the limb that were amputated regrow.

This variable response indicates that cells at the amputation wound edge must "know" where they are in relation to the entire limb. Such positional information is what controls the cellular and molecular processes leading to

[THE AUTHORS]

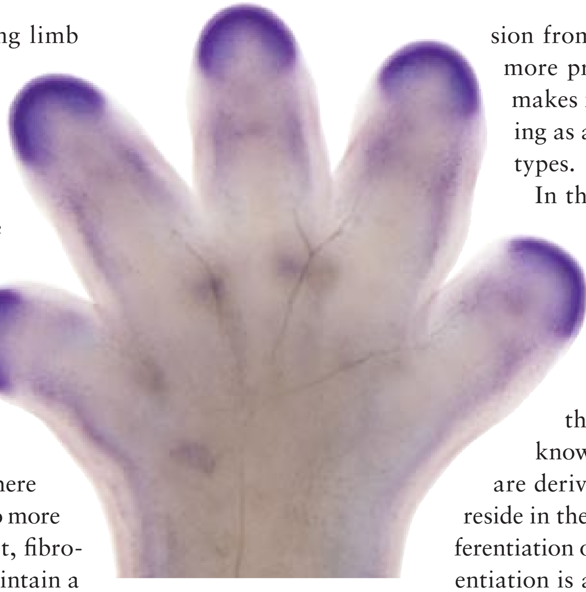
Ken Muneoka, Manjong Han and **David M. Gardiner** are part of a multi-institution research team working toward regenerating a mammalian limb. Their group, led by Muneoka, is one of only two to have received a multimillion-dollar grant from the Defense Advanced Research Projects Agency to pursue human limb regeneration. Muneoka is a professor and Han an assistant research professor in the department of cell and molecular biology at Tulane University. Gardiner is a research biologist in the department of cell and molecular biology at the University of California, Irvine, where Muneoka also completed a postdoctoral fellowship. Muneoka unexpectedly found himself working there once again after Hurricane Katrina forced him to relocate his family and his Tulane research staff to Irvine for five months.

the perfect replacement of the missing limb parts, and it is encoded in the activity of various genes. Examining which genes are at work during these processes helps to reveal the mechanisms controlling this stage of regeneration.

Although a large number of genes are involved during embryonic development in educating cells about their position in the limb, the activity of a gene family called *Hox* is critical. In most animals, cells in the developing limb bud use the positional code provided by *Hox* genes to form a limb, but then they "forget" where they came from as they differentiate into more specialized tissues later on. In contrast, fibroblasts in the adult salamander limb maintain a memory of this information system and can reaccess the positional *Hox* code in the process of limb regeneration.

During regeneration the fibroblasts bring this information with them as they migrate across the wound to initiate blastema formation, and once in the blastema, cells are able to "talk" to one another to assess the extent of the injury. The content of this crosstalk is still largely a mystery, but we do know that one outcome of the conversation is that the regenerating limb first establishes its boundaries, including the outline of the hand or foot, so that cells can use their positional information to fill in the missing parts between the amputation plane and the fingers or toes.

Because muscle and bone make up the bulk of a limb, we are also interested in understanding where the raw material for those tissues originates and what mechanisms control their formation. When the regenerative response is initiated, one of the key early events involves a poorly understood process called dedifferentiation. The term is typically used to describe a cell's reversion from a mature specialized state to a more primitive, embryonic state, which makes it capable of multiplying and serving as a progenitor of one or more tissue types.



MOUSE PAW produces a growth factor called BMP4 (*purple stain*) during fetal digit development that is also essential for natural digit-tip regeneration in mice.

In the field of regeneration, the word was first used by early scientists who observed under the microscope that the salamander stump tissues, particularly the muscle, appeared to break down and give rise to proliferative cells that formed the blastema. We now know that those muscle-associated cells are derived from stem cells that normally reside in the muscle tissue and not from dedifferentiation of muscle. Whether or not dedifferentiation is actually happening in the case of every tissue type within a regenerating limb has yet to be proved, although it is clear that a variation of this theme does occur during regeneration. Fibroblasts that enter the blastema and become primitive blastemal cells have the ability to differentiate into skeletal tissues (bone and cartilage) as well as to redifferentiate into the fibroblasts that will form the interstitial meshwork of the new limb, for instance.

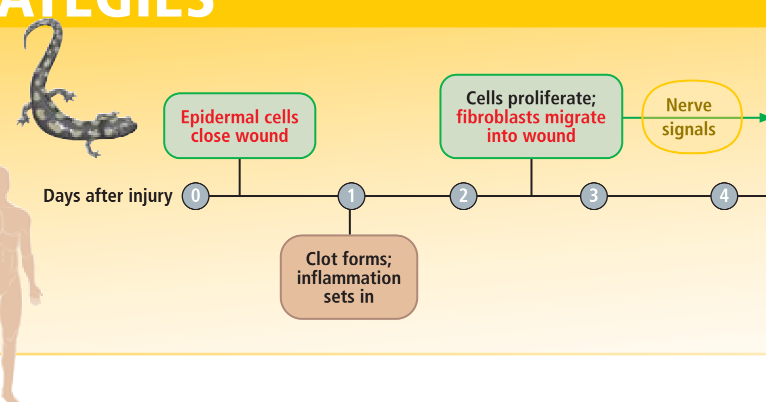
Returning to another of the central cellular players in blastema formation, the epidermal cells, we can also pinpoint moments in the regeneration process when it seems these cells are making a transition to a more embryonic state. A number of genes active in the embryonic ectoderm are critical for limb development, including *Fgf8* and *Wnt7a*, but as the ectoderm of the embryo differentiates to form the multilayered epidermis of the adult, these genes are turned off. During regeneration in the adult, the epidermal cells that migrate across the amputation wound and establish a wound epidermis initially begin to display gene activity, such as production of

KEN MUNEOKA (*photograph*); ALICE Y. CHEN (*illustrations*)

[THE HUMAN CHALLENGE]

DIVERGENT HEALING STRATEGIES

In mammals and salamanders, cell responses to a severe injury such as an amputation are similar in some ways (*red type*), although the pace of healing is much slower in mammals and environmental signals that promote regeneration are not present. Salamanders close a wound within hours but do not form scars. Instead cells respond to signals from one another and from the wound environment to begin a rebuilding process within days. Mammalian wound healing aims to just seal off the damaged area, first with a scab and then with a scar. By the time human epidermal cells close a wound, the salamander is already forming a blastema in preparation for regeneration.



wound-healing keratin proteins, that is not specifically related to regeneration. Later the wound epidermal cells activate *Fgf8* and *Wnt7a,* the two important developmental genes. For practical purposes, then, the essential definition of dedifferentiation—as it pertains to the epidermis and other cell types—is the specific reactivation of essential developmental genes.

Thus, our studies of salamanders are revealing that the regeneration process can be divided into pivotal stages, beginning with the wound-healing response, followed by the formation of a blastema by cells that revert to some degree to an embryonic state, and finally, the initiation of a developmental program to build the new limb. As we move toward the challenge of inducing limb regeneration in humans, we rely on these insights to guide our efforts. Indeed, the hardest things to discover in science are those that do not already occur, and limb regeneration in humans fits snugly into this category, although that does not mean humans have no natural regenerative capacity.

Potential at Our Fingertips

One of the most encouraging signs that human limb regeneration is a feasible goal is the fact that our fingertips already have an intrinsic ability to regenerate. This observation was made first in young children more than 30 years ago, but since then similar findings have been reported in teenagers and even adults. Fostering regeneration in a fingertip amputation injury is apparently as simple as cleaning the wound and covering it with a simple dressing. If allowed to heal naturally, the fingertip restores its contour, fingerprint and sensation and undergoes a varying degree of lengthening. The success of this conservative treatment of fingertip amputation injuries has been documented in medical journals thousands of times. Interestingly, the alternative protocol for such injuries typically included operating to suture a skin flap over the amputation wound, a "treatment" that we now know will inhibit regeneration even in the salamander because it interferes with formation of the wound epidermis. The profound message in these reports is that human beings have inherent regenerative capabilities that, sadly, have been suppressed by some of our own traditional medical practices.

It is not easy to study how natural human fingertip regeneration works because we cannot go around amputating fingers to do experiments, but the same response has been demonstrated in both juvenile and adult mice by several researchers. In recent years two of us (Muneoka and Han) have been studying the mouse digit-tip regeneration response in more detail. We have determined that a wound epidermis does form after digit-tip amputation, but it covers the regenerating wound much more slowly than occurs in the salamander. We have also shown that during digit-tip regeneration, important embryonic genes are active in a population of undifferentiated, proliferating cells at the wound site, indicating that they are blastema cells. And indirect evidence suggests that they are derived from fibroblasts residing in the interstitial connective tissues and in bone marrow.

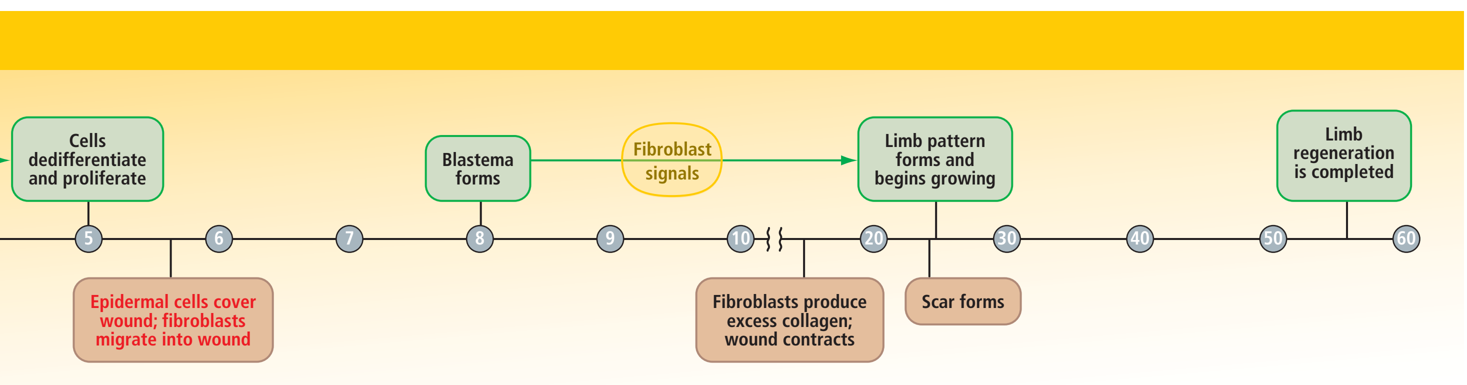
To explore the roles of specific genes and growth factors during the mouse-digit regeneration response, we developed a tissue culture that serves as a model for fetal mouse-digit regeneration. With it, we found that if we experimentally depleted a growth factor called bone morphogenetic protein 4 (BMP4) from the fetal amputation wound, we inhibited regeneration. In addition, we have shown that a mutant mouse lacking a gene called *Msx1* is unable to regenerate its digit tips. In the fetal digit tip, *Msx1* is critical to the production of BMP4, and we were able to restore the regener-

FAST FACTS

Young American alligator photographed in Louisiana is regenerating its amputated tail.

- The tail has about the same diameter as a human limb, suggesting that the ability to regenerate an appendage is not limited by the size of the amputation wound surface.

- Regrowing an adult human limb also might not take as long as it took to grow the first time. In salamanders, a poorly understood phenomenon known as catch-up allows the regenerating limb to go through a phase of rapid growth, resulting in a final limb that is appropriately scaled to the rest of the animal.



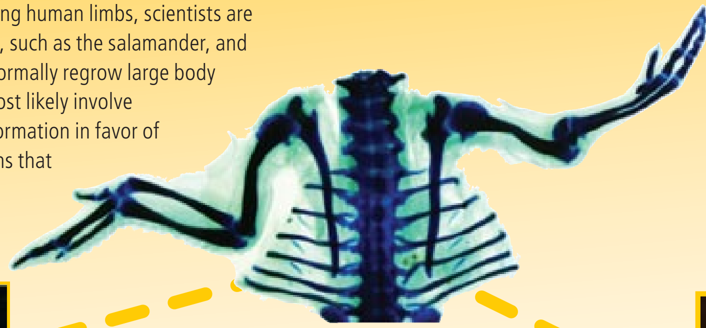
CAROL BURDSAL (*photograph*); FROM "LIMB REGENERATION IN HIGHER VERTEBRATES: DEVELOPING A ROADMAP," BY MANJONG HAN ET AL., IN *ANATOMICAL RECORD PART B: THE NEW ANATOMIST,* VOL. 287B, ISSUE 1; NOVEMBER 2005

[GETTING THERE]

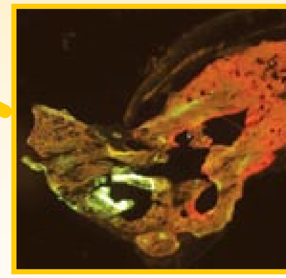
THE ROAD TO REGENERATION

Taking a step-by-step approach toward the goal of regrowing human limbs, scientists are learning how to control the process in natural regenerators, such as the salamander, and how to trigger similar mechanisms in animals that do not normally regrow large body parts. Tapping the regenerative potential in humans will most likely involve redirecting our wound-healing responses away from scar formation in favor of a limb-building program similar to the biological instructions that first create our limbs during fetal development.

▲ **REDIRECTED WOUND HEALING**
Causing a new limb to grow from the site of an incision on front of the leg of an axolotl established the basic requirements in salamanders for triggering a limb-regeneration response where normally only simple wound healing would occur.



▲ **NONREGENERATING VERTEBRATE**
A normal ankle and foot grew from the "elbow" of a chicken embryo's wing (*above right*) after leg tissue was grafted into the wing bud earlier in the chick's development. Regrowth of the amputated leg segment in an animal that does not naturally regenerate shows that limb-building programs can be reactivated when the wound environment is permissive.



DAVID M. GARDINER (*axolotl*); KEN MUNEOKA (*chicken and mouse digit*)

ation response by adding BMP4 to the wound in the *Msx1*-deficient mouse, confirming BMP4's necessity for regeneration.

Studies by Cory Abate-Shen and her colleagues at the Robert Wood Johnson Medical School have also demonstrated that the protein encoded by *Msx1* inhibits differentiation in a variety of cell types during embryonic development. That link to the control of differentiation suggests that the protein plays a role in the regeneration response by causing cells to dedifferentiate. Although *Msx1* is not active during the early dedifferentiation stages of salamander limb regeneration, its sister gene *Msx2* is one of the first genes reactivated during regeneration and very likely serves a similar function.

The Human Challenge

The idea of regenerating a human limb may still seem more like fantasy than a plausible possibility, but with insights such as those we have been describing, we can evaluate in a logical stepwise manner how it might happen. An amputated human limb results in a large and complex wound surface that transects a number of different tissues, including epidermis, dermis and interstitial connective tissue, adipose tissue, muscle, bone, nerve and vasculature. Looking at those different tissue types individually, we find that most of them are actually very capable of regenerating after a small-scale injury.

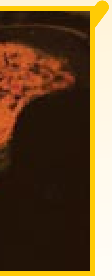
In fact, the one tissue type within a limb that lacks regenerative ability is the dermis, which is composed of a heterogeneous population of cells, many of which are fibroblasts—the same cells that play such a pivotal role in the salamander regeneration response. After an injury in humans and other mammals, these cells undergo a process called fibrosis that "heals" wounds by depositing an unorganized network of extracellular matrix material, which ultimately forms scar tissue. The most striking difference between regeneration in the salamander and regenerative failure in mammals is that mammalian fibroblasts form scars and salamander fibroblasts do not. That fibrotic response in mammals not only hampers regeneration but can be a very serious medical problem unto itself, one that permanently and progressively harms the functioning of many organs, such as the liver and heart, in the aftermath of injury or disease.

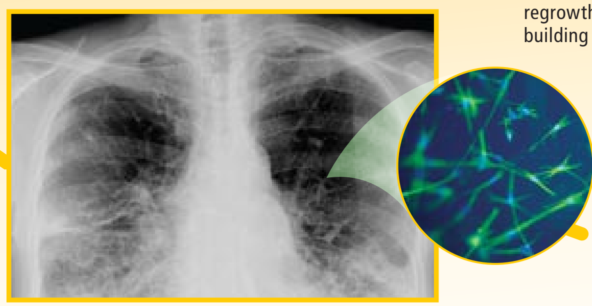
Studies of deep wounds have shown that at least two populations of fibroblasts invade an injury during healing. Some of these cells are fibroblasts that reside in the dermis, and the others are derived from circulating fibroblast-like stem cells. Both types are attracted to the wound by signals from immune cells that have also rushed to the scene. Once in the wound, the fibroblasts migrate and proliferate, eventually producing and modifying the extracellular matrix of the area. This early process is not that dissimilar to the regeneration response in a salamander wound, but the mammalian fibroblasts produce an excessive amount of matrix that becomes abnormally cross-linked as the scar tissue matures. In contrast, salamander fibroblasts stop producing matrix once the normal architecture has been restored.

An exception to this pattern in mammals does exist, however. Wounds in fetal skin heal without forming scars—yielding perfect skin

◀ HUMAN POTENTIAL
Natural human regeneration of amputated fingertips has been well documented, including the recent case of Lee Spievak. His middle finger, about an inch of which was severed by a model airplane propeller, is shown after complete healing. The injury was treated with a protein powder that might have aided regeneration by acting as a scaffold for regrowing tissues.

▼ REBUILDING LIMBS
Most human tissues are individually able to regenerate, which suggests that regrowing complex body parts is a realistic goal. Regenerating whole limbs will require changing the signals cells receive in the wound environment so that brakes on regrowth are removed and our innate limb-building programs are reactivated.





◀ MAMMALIAN DIGIT TIP
New bone (*red stain*) growing from the site where a mouse's digit tip was amputated (*green stain*) illustrates the regenerative potential present in mammals. The authors have also shown that a blastema forms at the site where a mouse digit will regenerate.

▲ BLOCKING FIBROSIS
Fibroblast cells (*inset*) that form scar tissue at a wound site also cause organ-scarring diseases, such as pulmonary fibrosis, which constricts breathing. Learning to prevent scarring in amputation wounds as a prelude to regeneration should also yield treatments for unwanted fibrosis in other body tissues.

regeneration and indicating that the switch to a fibrotic response arises with the developmental maturation of the skin. Although this difference could reflect a change in the biology of the fibroblasts, it is more likely a result of altered signaling from the extracellular wound environment modulating the behavior of the fibroblasts, which in turn suggests that therapeutically modifying those signals could change the healing response. At the same time, the fact that limb amputations during fetal stages of development do not result in regeneration of the limb reminds us that scar-free wound healing is likely to be necessary but not sufficient for regeneration.

To advance our understanding of what it will take to induce limb regeneration in people, we are continuing our work with mice. Our research group has already described a natural blastema in a mouse amputation injury, and our goal within the next year is to induce a blastema where it would not normally occur. Like the accessory-limb experiments in salamanders, this achievement would establish the minimal requirements for blastema formation. We hope that this line of investigation will also reveal whether, as we suspect, the blastema itself provides critical signaling that prevents fibrosis in the wound site.

If we succeed in generating a blastema in a mammal, the next big hurdle for us would be coaxing the site of a digit amputation to regenerate the entire digit. The complexity of that task is many times greater than regenerating a simple digit tip because a whole digit includes joints, which are among the most complicated skeletal structures formed in the body during embryonic development. Developmental biologists are still trying to understand how joints are made naturally, so building a regenerated mouse digit, joints and all, would be a major milestone in the regeneration field. We hope to reach it in the next few years, and after that, the prospect of regenerating an entire mouse paw, and then an arm, will not seem so remote.

Indeed, when we consider all that we have learned about wound healing and regeneration from studies in various animal models, the surprising conclusion is that we may be only a decade or two away from a day when we can regenerate human body parts. The striking contrast between the behavior of fibroblasts in directing the regeneration response in salamanders versus the fibrotic response leading to scarring in mammals suggests that the road to successful regeneration is lined with these cells. Equally encouraging is the recent discovery by Howard Y. Chang and John L. Rinn of Stanford University that adult human fibroblasts, like salamander fibroblasts, retain a memory of the spatial coordinate system used to establish the body plan early in the embryo's development. Given that such positional information is required for regeneration in salamanders, its existence in human fibroblasts enhances the feasibility of tapping into and activating developmental programs necessary for regeneration.

Now, as we watch a salamander grow back an arm, we are no longer quite as mystified by how it happens. Soon humans might be able to harness this truly awesome ability ourselves, replacing damaged and diseased body parts at will, perhaps indefinitely. ■

➥ MORE TO EXPLORE

Limb Regeneration. Panagiotis A. Tsonis. Cambridge University Press, 1996.

Ontogenetic Decline of Regenerative Ability and the Stimulation of Human Regeneration. David M. Gardiner in *Rejuvenation Research*, Vol. 8, No. 3, pages 141–153; September 1, 2005.

Limb Regeneration in Higher Vertebrates: Developing a Roadmap. Manjong Han et al. in *Anatomical Record Part B: The New Anatomist*, Vol. 287B, Issue 1, pages 14–24; November 2005.

Appendage Regeneration in Adult Vertebrates and Implications for Regenerative Medicine. Jeremy P. Brockes and Anoop Kumar in *Science*, Vol. 310, pages 1919–1923; December 23, 2005.

Regenerative Biology and Medicine. David L. Stocum. Academic Press, 2006.

AL BEHRMAN *AP Photo (Spievak)*; LIVING ART ENTERPRISES, LLC *Photo Researchers, Inc. (x-ray)*; L. Y. LEUNG AND D. J. TSCHUMPERLIN *Harvard School of Public Health (fibroblast inset)*; JIM WATSON *AFP/Getty Images (amputee)*

Reclaiming the Aral Sea

KEY CONCEPTS

- The Aral Sea in Central Asia was the fourth-largest lake on the planet in 1960. By 2007 it had shrunk to 10 percent of its original size. Widespread, wasteful irrigation of the deserts along the Amu and Syr rivers, which feed the Aral, cut the freshwater inflow to a trickle.
- The sea has shriveled into three major residual lakes, two of which are so salty that fish have disappeared. The once thriving fishing fleets have disappeared, too. Former shore towns have collapsed. Vast seabeds lie exposed and dried; winds now blow salts and toxic substances across populated areas, causing significant health problems.
- Nevertheless, a dam built in 2005 has helped the northernmost lake expand quickly and drop substantially in salinity. Fish populations and wetlands are returning—and with them signs of economic revival. The two big southern lakes could become dead seas, however, unless the Amu river, which once fed them, is substantially reengineered, a project requiring tens of billions of dollars and difficult political agreements.
- Other lakes worldwide are beginning to suffer similar fates, chief among them Lake Chad in Central Africa and the Salton Sea in Southern California. Lessons learned about the Aral's demise and partial resurrection could benefit these regions. *—The Editors*

Recklessly starving the world's fourth-largest lake to irrigate crops turned rich waters into a barren wasteland. Now the northern part, at least, is coming back

By Philip Micklin and Nikolay V. Aladin

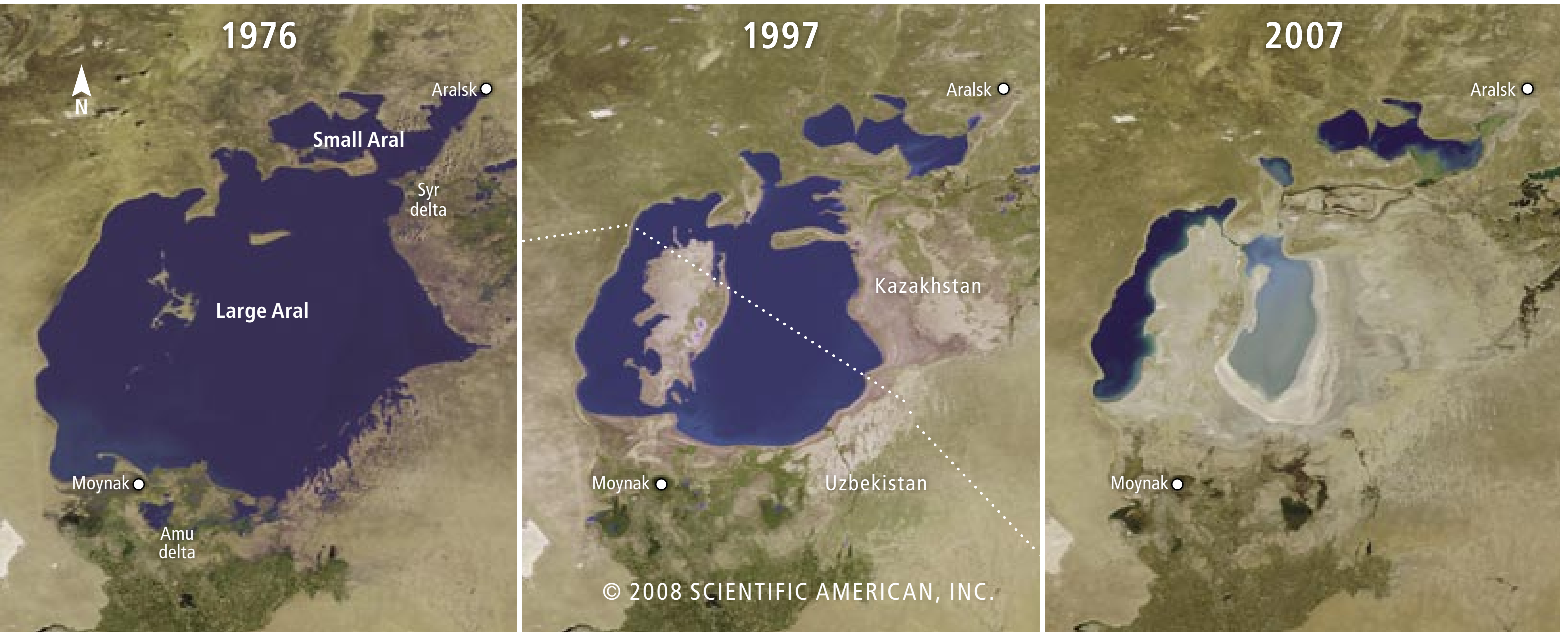
MATTHIEU PALEY *Corbis* (*waterless sea*); WORLD SAT (*satellite images*); MAPPING SPECIALISTS (*all maps*); LUCY READING-IKKANDA (*all illustrations*)

COLLAPSE

The Aral Sea gets almost all its water from the Amu and Syr rivers. Over millennia the Amu's course has drifted away from the sea, causing it to shrink. But the lake always rebounded as the Amu shifted back again. Today heavy irrigation for crops such as cotton and rice siphons off much of the two rivers, severely cutting flow into their deltas and thus into the sea. Evaporation vastly outpaces any rainfall, snowmelt or groundwater supply, reducing water volume and raising salinity.

The Soviet Union hid the sea's demise for decades until 1985, when leader Mikhail Gorbachev revealed the great environmental and human tragedy. By the late 1980s the sea's level had dropped so much that the water had separated into two distinct bodies: the Small Aral (north) and the Large Aral (south). By 2007 the south had split into a deep western basin, a shallow eastern basin and a small, isolated gulf. The Large Aral's volume had dropped from 708 to only 75 cubic kilometers (km^3), and salinity had risen from 14 to more than 100 grams per liter (g/l). The 1991 dissolution of the Soviet Union divided the lake between newly formed Kazakhstan and Uzbekistan, ending a grand Soviet plan to channel in water from distant Siberian rivers and establishing competition for the dwindling resource.

▼ SHRINKING LAKE has receded 100 kilometers from this former shoreline near Moynak, Uzbekistan.



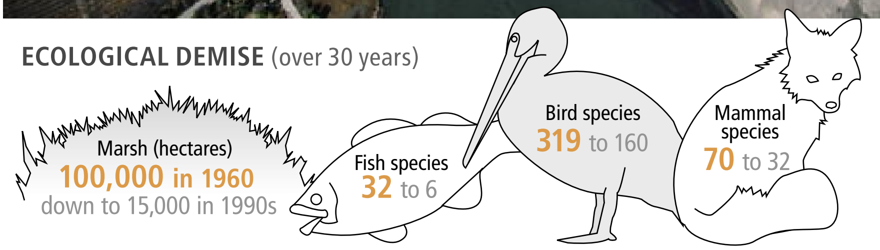
ROCK BOTTOM

Desiccation of the Aral Sea has wrought severe consequences. Greatly reduced river flows ended the spring floods that sustained wetlands with freshwater and enriched sediment. Fish species in the lakes dropped from 32 to six because of rising salinity and loss of spawning and feeding grounds (most survived in the river deltas). Commercial fisheries, which caught 40,000 metric tons of fish in 1960, were gone by the mid-1980s; more than 60,000 related jobs were lost. The most common remaining lake occupant was the Black Sea flounder (*kambala* in Russian), a saltwater fish introduced in the 1970s, but by 2003 it had disappeared from the southern lakes because salinity was more than 70 g/l, double that of a typical ocean.

Shipping on the Aral also ceased because the water receded many kilometers from the major ports of Aralsk to the north and Moynak in the south; keeping increasingly long channels open to the cities became too costly. Groundwater levels dropped with falling lake levels, intensifying desertification. By the mid-1990s meager stretches of halophytes (plants tolerant of saline soils) and xerophytes (those tolerant of dry conditions) struggled where lush expanses of trees, bushes and grasses had once flourished on the banks. Only half the number of native mammal and bird species could be found in the area. The climate also changed up to 100 kilometers beyond the original shoreline: today summers are hotter, winters are colder, humidity is lower (so rainfall is less), the growing season is shorter and drought is more common.

▼ **FREIGHT LINES** that once transported manufactured goods and foods from the thriving port city of Aralsk lie in ruin, as does the city's economy.

▼ **VAST DRAINAGE BASIN** (*top*) provides almost no water to the Aral Sea because irrigation channels, as shown in the photograph below, siphon water from the Amu and Syr rivers for hundreds of kilometers through several countries. Among other results, animal and plant species have disappeared (*bottom*).



KAZUYOSHI NOMACHI *Corbis* (*aerial irrigation*); PAUL HOWELL *Sygma/Corbis* (*train*)

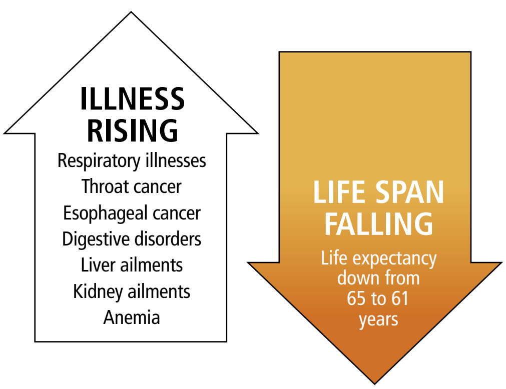
TOXIC SUBSTANCES

The receding sea has exposed and dried 54,000 square kilometers of seabed, which is choked with salt and in some places laced with pesticides and other agricultural chemicals deposited by runoff from area farming. Strong windstorms blow salt, dust and contaminants as far as 500 km. Winds from the north and northeast drive the most severe storms, seriously impacting the Amu delta to the south—the most densely settled and most economically and ecologically important area in the region. Airborne sodium bicarbonate, sodium chloride and sodium sulfate kill or retard the growth of natural vegetation and crops—a cruel irony given that irrigating those crops starves the sea.

Health experts say the local population suffers from high levels of respiratory illnesses, throat and esophageal cancer, and digestive disorders caused by breathing and ingesting salt-laden air and water. Liver and kidney ailments, as well as eye problems, are common. The loss of fish has also greatly reduced dietary variety, worsening malnutrition and anemia, particularly in pregnant women.

Vozrozhdeniya Island also poses a unique problem. When it was far out to sea, the Soviet Union used Vozrozhdeniya as a testing ground for biological weapons; anthrax, tularemia, brucellosis, plague, typhus, smallpox, and botulinum toxin were tried on horses, monkeys, sheep, donkeys and laboratory animals. But as a result of receding waters, Vozrozhdeniya united with the mainland to the south in 2001. Health experts fear that weaponized organisms have survived and could reach civilization via fleas on infected rodents or that terrorists might gain access to the organisms.

HEALTH TOLL (on area population)



▲ **REFUSE and pesticides once dumped into Aralsk harbor are now exposed (*top*). Terrific windstorms (*middle*) blow the toxic substances and massive quantities of sand and dried salts across the region, killing crops and sickening local people.**

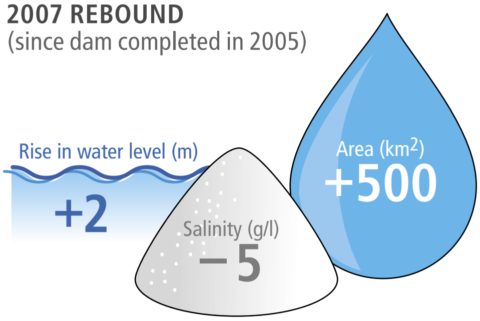
▶ **SOVIET PATROL BOAT, once used to guard bioweapons activity on Vozrozhdeniya Island, is now grounded on the dried seabed.**

ANTOINE GYORI AGP/*Corbis* (*sea debris*); JACQUES DESCLOITRES *MODIS Land Rapid Response Team, NASA GSFC* (*aerial dust storm*); PHILIP MICKLIN (*boat*)

HOPE FOR THE NORTH

Returning the entire Aral Sea to its 1960s state is unrealistic. The annual inflow from the Syr and Amu rivers would have to be quadrupled from the recent average of 13 km^3. The only means would be to curtail irrigation, which accounts for 92 percent of water withdrawals. Yet four of the five former Soviet republics in the Aral Sea basin (Kazakhstan is the exception) intend to expand irrigation, mainly to feed growing populations. Switching to less water-intensive crops, such as replacing cotton with winter wheat, could help, but the two primary irrigating nations, Uzbekistan and Turkmenistan, intend to keep cotton to earn foreign currency. The extensive irrigation canals could be greatly improved; many are simply cuts through sand, and they allow enormous quantities of water to seep away. Modernizing the entire system could save 12 km^3 a year but would cost at least $16 billion. The basin states do not have the money or the political will.

Kazakhstan has nonetheless tried to partially restore the northern Aral. In the early 1990s it constructed an earthen dike to block outflow to the south that was uselessly lost to evaporation, but a catastrophic failure in April 1999 destroyed it. The effort demonstrated that water level could be raised and salinity lowered, however, prompting Kazakhstan and the World Bank to fund an $85-million solution. The key element was a much heftier, 13-km earthen dike with a gated concrete dam for water discharge, completed in November 2005. Heavy runoff from the Syr River in the ensuing winter jump-started the Small Aral's recovery. The water rose from 40 to 42 meters—the intended design height—in only eight months. Area increased by 18 percent, and salinity has dropped steadily, from roughly 20 to about 10 g/l today. Fishers are once again catching several species in substantial numbers—most important, the highly prized pike perch (known as *sudak* in Russian) and *sazan* (a type of carp).



▼ **GATED DAM (*top*) and a 13-kilometer dike completed by Kazakhstan in 2005 saved the Small Aral by stopping outflow into dying channels that led nowhere. Since then water levels and fish populations have risen (*bottom*). The gates, which can release excess water to control lake level, were finished in November 2005—water, two meters deep, had returned by the following summer.**

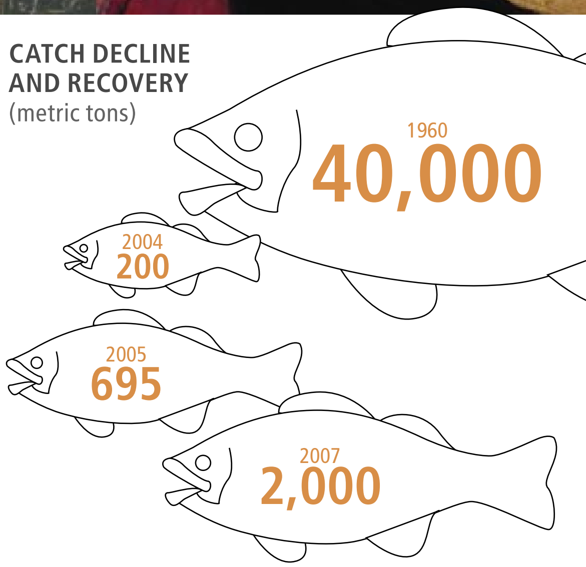
PHILIP MICKLIN (*top*); IGOR PLOTNIKOV (*bottom*)

PROSPERITY RISING

We expect salinities in the Small Aral to settle at three to 14 g/l, depending on location. At these levels many more indigenous species should return, although the saltwater *kambala* would disappear from most places. Further restoration is possible. For example, if irrigation improvements raised the average annual inflow from the Syr to 4.5 km^3, which is entirely feasible, the lake's level could stabilize at about 47 meters. This change would bring the shoreline to within eight kilometers of Aralsk, the former major port city, close enough to allow redredging of an earlier channel that connected the city to the receding waters. The channel would give large commercial fishing vessels access to the sea, and shipping could restart. Marshlands and fish populations would improve even more because of a further reduction in salinity. Outflow to the southern lakes could also increase, helping their restoration [*see map on next page*]. Such a plan would require a much longer and higher dike, as well as reconstruction of the gate facility, and it is not clear that Kazakhstan has the means or desire to pursue it. The country is, however, now discussing more modest proposals to bring water closer to Aralsk.

SYLVAIN LIECHTI

FISH have returned to the Small Aral in rapidly increasing numbers, providing livelihoods for fishers from surrounding villages (*top* and *middle*). A processing plant in Aralsk has also reopened (*bottom*), boosting the local economy.



SPECIES CAUGHT
(autumn 2007, most to least)

1. Carp
2. Aral bream
3. Aral roach
4. Pike perch
5. Flounder
6. Aral asp
7. Aral white-eye bream
8. Crucian carp
9. Sabre fish
10. Pike
11. Perch
12. Rudd
13. Aral shemaya
14. Wels
15. Snakehead
16. Turkestan ide

LONG SHOT FOR THE SOUTH

The Large Aral faces a difficult future; it continues to shrink rapidly. Only a long, narrow channel connects the shallow eastern basin and the deeper western basin, and this could close altogether. If countries along the Amu make no changes, we estimate that at current rates of groundwater in and evaporation out, an isolated eastern basin would stabilize at an area of 4,300 square kilometers (km^2). But it would average only 2.5 meters deep. Salinity would exceed 100 g/l, possibly reaching 200 g/l; the only creatures that could live in it would be brine shrimp and bacteria.

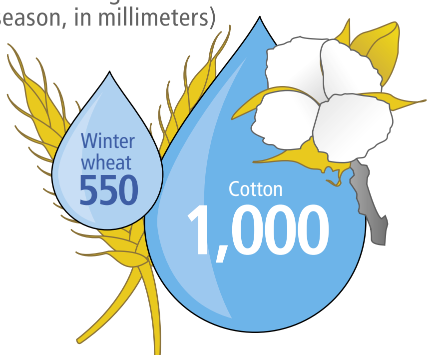
The western basin's fate depends on groundwater inflow, estimates for which are uncertain. One of us (Aladin) has noted numerous freshwater springs on the western cliffs. Our most reliable calculations indicate that the basin would settle at about 2,100 km^2. The lake would still be relatively deep, reaching 37 meters in spots, but salinity would rise well above 100 g/l.

Large-scale engineering could partially rehabilitate the western basin. One early plan recently updated by one of us (Micklin) could help [*see map at right*]. It has received little evaluation, so costs are unknown, but it would be very expensive. It would require only modestly increased flow through the Amu, however, which could be attained with relatively reasonable irrigation improvements in the river's drainage basin. Rehabilitating wetlands is also a major goal.

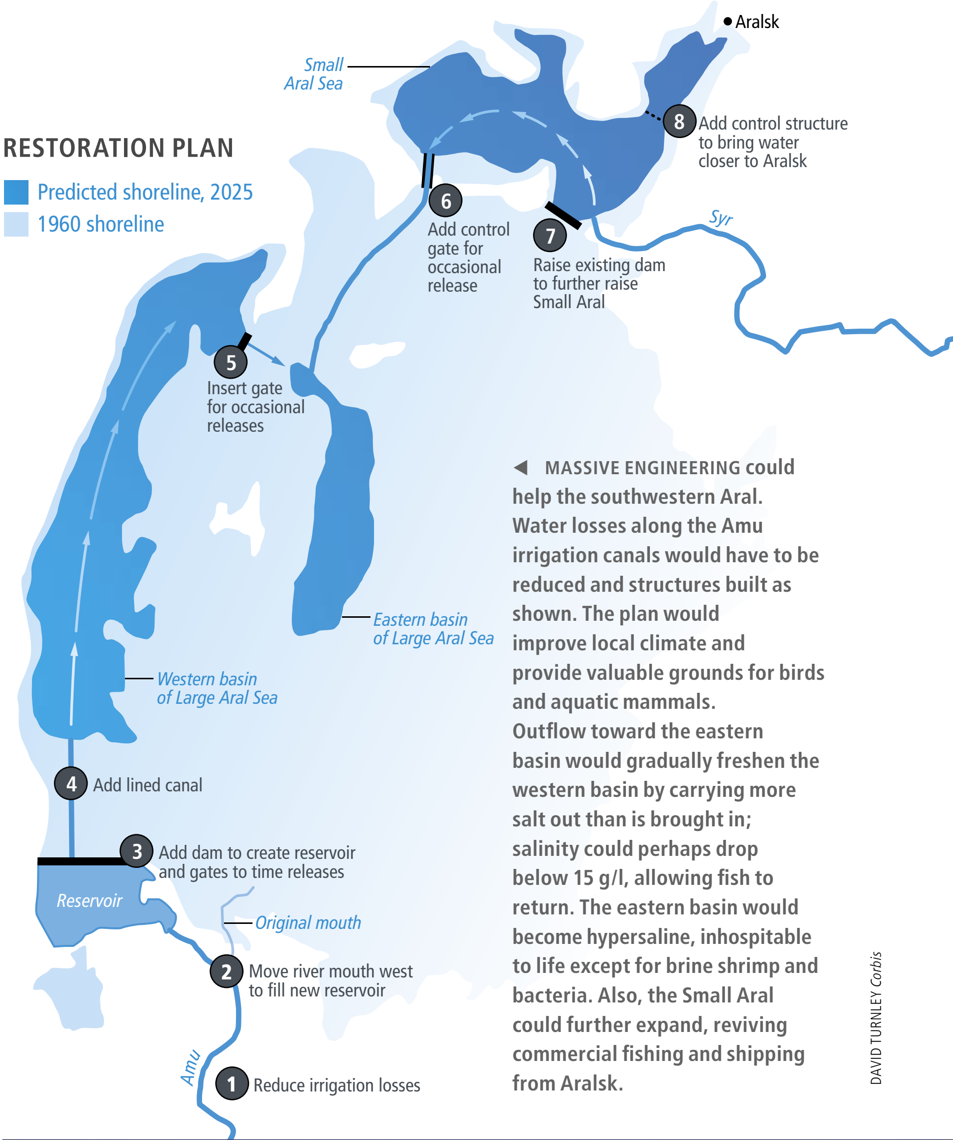
The Soviet Union began such work in the late 1980s, and Uzbekistan has continued this effort with the help of international donors. Biodiversity, fisheries and natural filtering of wastewater by aquatic vegetation have marginally improved, but there is no quick fix. The Aral has been desiccated for more than 40 years; sustainable, long-term solutions will require not only major investments and technical innovations but fundamental political, social and economic change.

CROPS' EFFECT ON WATER USAGE

(rainfall and irrigation needed per season, in millimeters)



RESTORATION PLAN



◀ **MASSIVE ENGINEERING could help the southwestern Aral. Water losses along the Amu irrigation canals would have to be reduced and structures built as shown. The plan would improve local climate and provide valuable grounds for birds and aquatic mammals. Outflow toward the eastern basin would gradually freshen the western basin by carrying more salt out than is brought in; salinity could perhaps drop below 15 g/l, allowing fish to return. The eastern basin would become hypersaline, inhospitable to life except for brine shrimp and bacteria. Also, the Small Aral could further expand, reviving commercial fishing and shipping from Aralsk.**

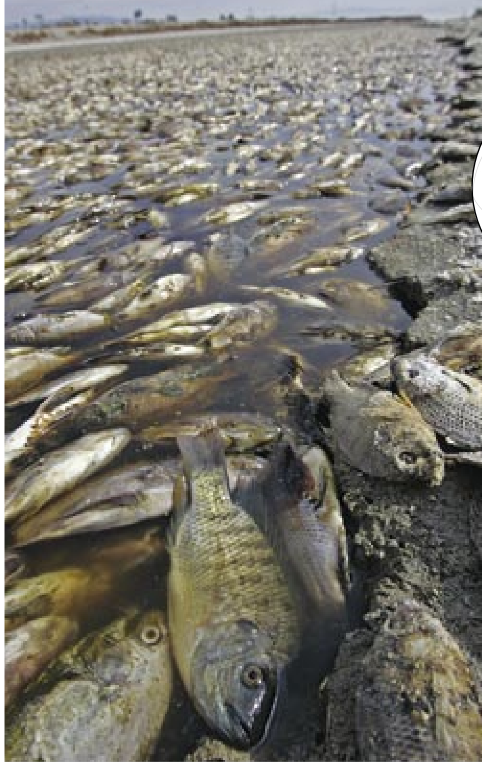
DAVID TURNLEY *Corbis*

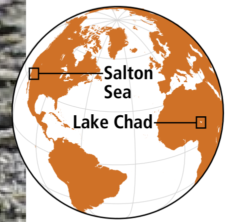
▲ **COTTON consumes much of the region's irrigation. A switch to less thirsty crops such as winter wheat could spare water badly needed to revive the Aral Sea, but countries rely on cotton sales for foreign currency.**

GLOBAL IMPLICATIONS

Until recently, many observers considered the Aral Sea a lost cause. Progress in the north, however, convincingly demonstrates that sizable parts of the remnant sea can be made ecologically and economically productive. The Aral story illustrates the enormous capacity of modern, technological societies to wreak havoc on the natural world and their own people, and yet the story also demonstrates the great potential for restoring the environment. Other water bodies around the world are beginning to suffer Aral-esque fates, notably Lake Chad in Central Africa and the Salton Sea in Southern California. We hope the lessons learned will be heeded elsewhere. Among them are:

- Humans can quickly wreck the natural environment, but repairing it is a long, arduous process. Planners must cautiously evaluate the consequences of large-scale interference in natural systems before starting any action, which the Soviet Union did not do.
- Avoidance of serious problems at present is no guarantee for the future. Widespread irrigation took place in the Aral Sea basin for many centuries and did not seriously hurt the sea before the 1960s, but further expansion pushed the region's hydrologic system beyond the point of sustainability.
- Beware of quick fixes for complex environmental and human problems. Major cuts in cotton growing could send more water to the sea but would damage national economies, raise unemployment and contribute to social unrest. Sustainable solutions require not only money and innovation but must be politically, socially and economically practical.
- The natural environment is amazingly resilient, so do not abandon hope or efforts to save it. Many pundits wrote off the Aral Sea as doomed, but substantial parts of it are now being ecologically restored. ■





▲ DEAD TILAPIA choke the shore of California's Salton Sea (*above*), growing ever saltier because of botched irrigation. Various plans are being considered to freshen the lake before all the fish are gone. Africa's Lake Chad (*below*) has shrunk to one-tenth its size since the 1960s because of widespread irrigation. Farmers, herders and dwellers from the four border countries often fight violently over the remaining water (*bottom right*, *blue*), now only 1.5 meters deep.

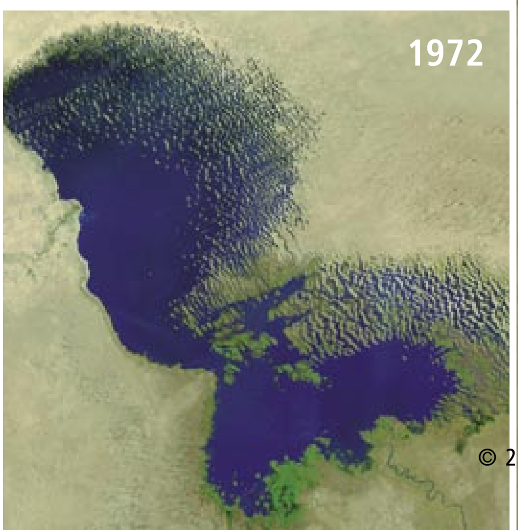
[THE AUTHORS]

Philip Micklin and **Nikolay V. Aladin** have conducted several on-site investigations of the Aral Sea over the past decade. Micklin is professor emeritus of geography at Western Michigan University. Aladin is head of the Brackish Water Laboratory at the Russian Academy of Sciences's Zoological Institute in St. Petersburg.

➡ MORE TO EXPLORE

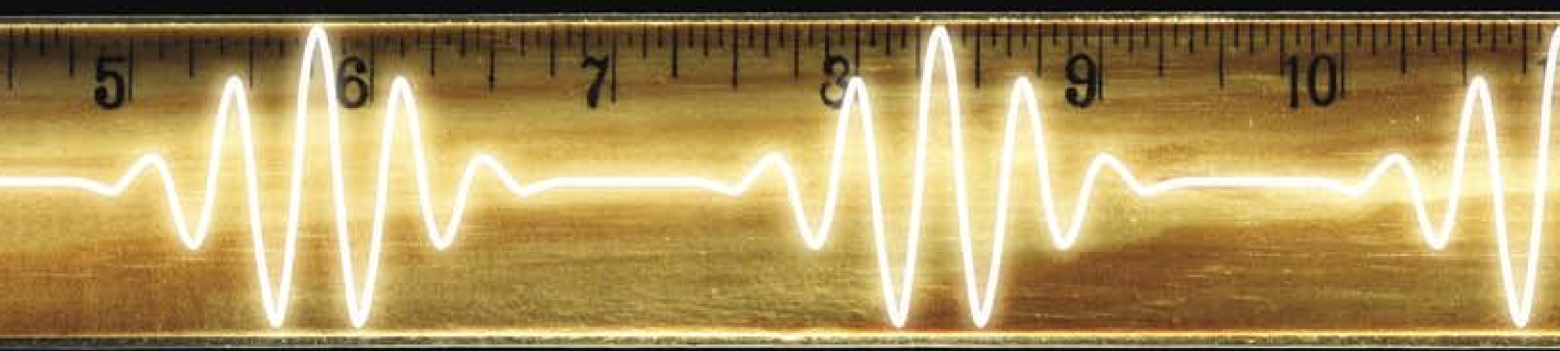
Hydrobiology of the Aral Sea. Edited by Nikolay V. Aladin et al. *Dying and Dead Seas: Climatic vs. Anthropic Causes.* NATO Science Series IV: Earth and Environmental Sciences, Vol. 36. Kluwer, 2004.

The Aral Sea Disaster. Philip Micklin in *Annual Review of Earth and Planetary Sciences,* Vol. 35, pages 47–72; 2007.



WORLD SAT (*aerial Lake Chad*); JAY CALDERON *The Desert Sun* (*dead tilapia*)

RULERS



A revolutionary kind of laser light called an optical frequency comb makes possible a more precise type of atomic clock and many other applications

KEY CONCEPTS

- A new kind of laser light—called an optical frequency comb—can be used to measure frequencies of light and intervals of time more precisely and easily than ever before.
- The comb is made of a train of evenly spaced, ultrashort laser pulses with a spectrum that looks like tens of thousands of "teeth."
- Applications include a more precise atomic clock, ultrasensitive chemical detectors, laser control of chemical reactions, higher-capacity telecommunications using optical fibers and improved lidar (light detection and ranging).

—*The Editors*

In the blink of an eye, a wave of visible light completes a quadrillion (10^{15}) oscillations, or cycles. That very large number presents both opportunities and a challenge. The opportunities promise numerous applications both inside and outside of laboratories. They go to the heart of our ability to measure frequencies and times with extremely high precision, a skill that scientists rely on for some of the best tests of laws of nature—and one that GPS systems, for instance, depend on. The challenge has centered on the impossibility of manipulating light with the techniques that work so well for electromagnetic waves of much lower frequencies, such as microwaves.

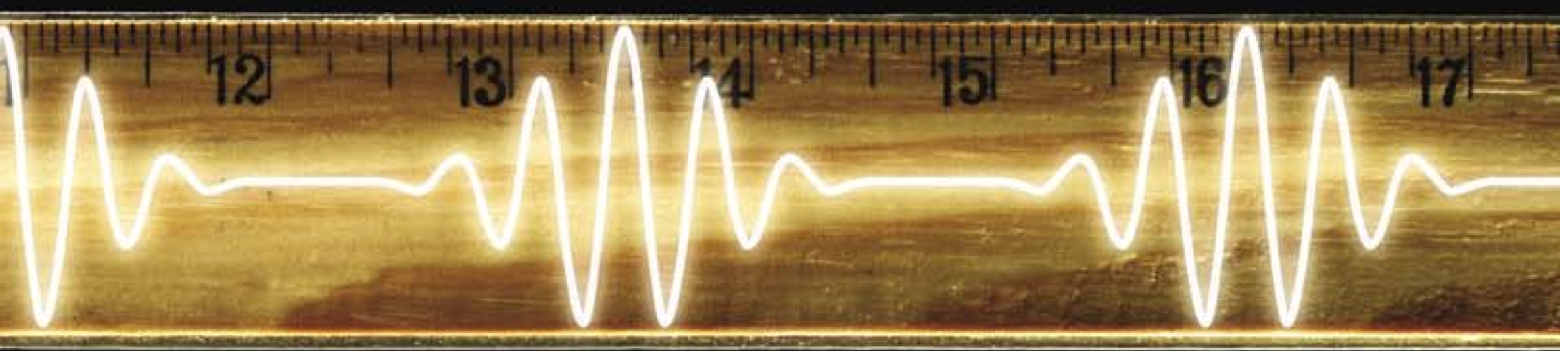
Now, thanks to a decade of revolutionary advances in laser physics, researchers have at hand technologies that can unlock the latent potential that visible light's high frequencies previously kept us from realizing. In particular, scientists have developed the tools to exploit a type of laser light known as an optical frequency comb. Like a versatile ruler of light with tens or hundreds of thousands of closely spaced "tick marks," an optical frequency comb provides exquisitely precise measurements of light. Such a comb can form a bridge spanning the huge frequency gap from microwaves to visible light: very precise microwave measurements can, with an optical comb, produce equally exact data about light.

Myriad applications are in the pipeline. Optical combs will enable a new generation of more precise atomic clocks, ultrasensitive chemical detectors and the means to control chemical reactions using lasers. The combs could greatly boost the sensitivity and range of lidar (*l*ight *d*etection *a*nd *r*anging)—and also provide a vast increase in the number of signals traveling through optical fiber [*see box on page 76*].

Combs will greatly simplify the task of measuring optical frequencies with extremely high precision. In the 20th century such a measurement would have required a team of Ph.D.s running rooms full of single-frequency lasers. Today a graduate student can achieve similar results with a simple apparatus using optical frequency combs. The new optical atomic clocks also spring from this simplification. Much as a pendulum in a grandfather clock requires gears to record its swings and slowly turn the clock's hands, an optical atomic clock uses an optical frequency comb to count the oscillations of light and convert them into a useful

DAVID EMMITE

OF LIGHT



By Steven Cundiff, Jun Ye and John Hall

LASER PULSES can form a kind of ruler of light, which scientists use to measure the frequencies of other lasers with great precision.

electronic signal. In just the past year, researchers have used optical combs to surpass the cesium-based atomic clocks that have been the best system available for decades.

In some respects, the scene-changing advent of optical combs is similar to the leap forward that resulted from the invention of the oscilloscope about 100 years ago. That device heralded the modern age of electronics by allowing signals to be displayed directly, which facilitated development of everything from television to the iPhone. Light, however, oscillates 10,000 times faster than the speed of the fastest available oscilloscopes. With optical combs, the same capability to display the waveform is becoming available for light.

Optical frequency comb applications require exquisite control of light across a broad spectrum of frequencies. This level of control has been available for radio waves for a long time but is only now becoming possible for light. An analogy to music helps in understanding the required level of control. Before the development of combs, lasers could produce a single color, like a single optical tone. They were akin to a violin with only one string and no fingerboard, capable of playing only one note (ignore for the moment that musical notes are much richer than pure tones). To play even a simple piece would require many different instruments, each painstakingly tuned. Each violin would require its own musician, just as every single-frequency laser requires its own operator.

In contrast, one operator can use an optical comb to cover the entire optical spectrum, not merely like a pianist at a piano but like a keyboardist playing an electronic synthesizer that can be programmed to mimic any musical instrument or even an entire orchestra. Comb technology, in effect, enables symphonies of hundreds of thousands of pure optical tones.

Anatomy of a Comb

Optical frequency combs are generated by devices called mode-locked lasers, which create ultrashort pulses of light. To understand the important features of such pulses, begin by imagining the light wave of the other chief kind of laser, a continuous-wave (CW) laser. Ideally, such a wave would be an endless stream of perfectly regular oscillations (representing the light wave's electric field), every wave crest and trough having the same amplitude and arriving at an unchanging rate. A pulse from a mode-

[APPLICATIONS]

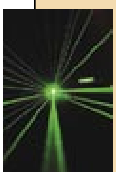
COMB TECHNOLOGIES

OPTICAL ATOMIC CLOCKS

The most accurate and precise clocks ever made, optical atomic clocks have already surpassed the microwave-based systems that have been the standard since 1967. They will play a central role in space navigation, communications between satellites, exquisitely sensitive tests of fundamental physics and other measurements.

CHEMICAL SENSORS

Researchers have demonstrated ultrasensitive chemical detectors based on optical combs and are now developing prototypes for commercial instruments. Comb-based sensors will let security screeners rapidly identify hazards such as explosives or dangerous pathogens. Doctors will diagnose illnesses by detecting chemicals in a patient's breath.



SUPERLASERS

With frequency combs, the outputs of many lasers can be stitched together to form a single stream of pulses whose light is as organized ("coherent") as light from a single laser. Eventually it should be possible to control the electromagnetic spectrum coherently all the way from radio waves to x-rays.

TELECOMMUNICATIONS

Optical combs will increase the number of signals that can be sent down a single optical fiber by orders of magnitude, requiring only one comb instead of a large number of individual lasers. Interference between the channels will be reduced. Secure communications, in particular, will benefit from the use of combs.

DESIGNER CHEMISTRY

Scientists are already investigating how to use lasers' coherent light to control chemical reactions. Optical combs will make this technique more predictable and reliable and will be instrumental in developing a new class of so-called ultracold chemical reactions. One day the combs will manipulate biological reactions, which are far more complicated than other chemical reactions.

LIDAR

Laser radar, or lidar (light detection and ranging), uses laser light to determine the position, velocity and characteristics of distant objects. By generating waveforms with custom-designed shapes, optical frequency combs are expected to boost lidar's sensitivity and range by orders of magnitude.

locked laser, in contrast, is a short series of wave crests and troughs whose amplitude rises from zero to a maximum and then falls back to zero [*see box on opposite page*]. The shortest pulses, with durations of less than 10 femtoseconds, contain just a few full oscillations of the light wave. The general outline of the pulse—its overall rise and fall—is called its envelope. One can think of the pulse as being like the earlier continuous wave (the "carrier wave"), with that wave's amplitude multiplied by the changing height of the envelope.

The carrier wave consists of light of one pure frequency. A plot of its spectrum would have a single spike at that frequency, representing the presence of that frequency alone. You might expect that the pulse you are imagining would also consist of light only at that frequency—after all, it is just the single-frequency carrier wave with its amplitudes changed—but that is not how waves and spectra work. Instead the pulse is made up of light of many frequencies all traveling together. The frequencies form a small, continuous band centered on the carrier frequency. The shorter the pulse, the broader the spread of frequencies.

Two additional features of the pulses emitted by mode-locked lasers are keys to the development of optical frequency combs. First, shifting the envelope a little relative to the carrier wave results in slightly different pulses. The peak of the pulse envelope may occur at the same time as a crest of the carrier, but it may also be shifted to any other stage of the oscillation. The amount of displacement is called the phase of the pulse.

Second, mode-locked lasers emit trains of pulses at a very regular rate, called the repetition rate. The frequency spectrum of such a train of pulses does not form a continuum spread on each side of the carrier frequency but rather breaks into many discrete frequencies. Plotted, the spectrum looks like the teeth of a hair comb, spaced at precisely the laser's repetition rate.

A typical repetition rate is around one gigahertz (a billion cycles per second), somewhat slower than modern computer processors. An optical comb that spanned the visible spectrum would have 400,000 teeth if they were spaced at one gigahertz. Scientists can measure repetition rates in the gigahertz (microwave) range very accurately using high-speed photodiodes, which detect each pulse in turn—and an optical comb would appear to leverage that accuracy up to visible wavelengths. Why not, then, use the teeth of the frequency comb as reference points to measure against?

There is, however, a catch. It relates to the phase. Everything is fine if the phase of every pulse in the train is exactly the same, because in that case the comb teeth will be precisely at integer multiples of the repetition rate. Thus, you would know the teeth positions once you had measured the laser's repetition rate.

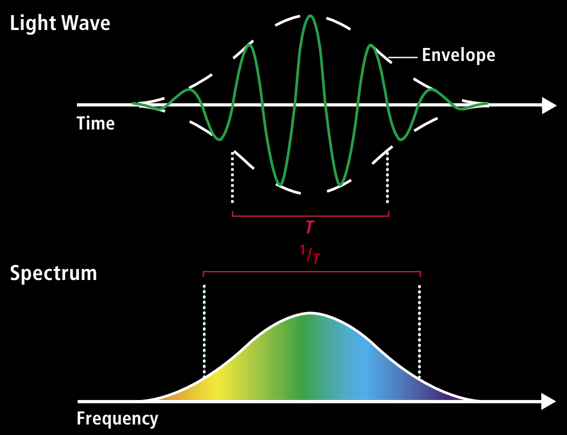
But it usually happens that the phase changes from one pulse to the next by some unpredictable but fixed amount [*see box on page 79*]. In that case, the comb teeth are shifted in frequency away from the exact integer multiples of the repetition rate by an amount called the offset fre-

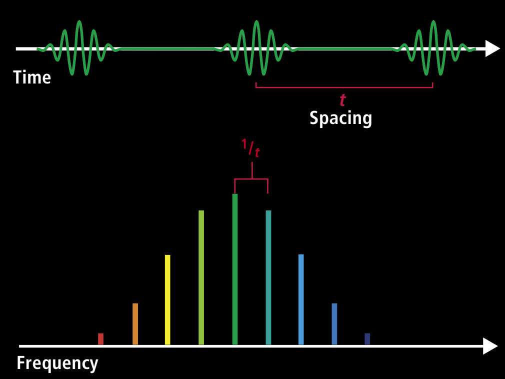
SORA/CORBIS (*clock*); EMILY HARRISON (*illustration on clock*); ROBERT KOHLHUBER *iStockPhoto* (*laser*); CONEYL JAY *Getty Images* (*optical fiber*); STEVE DIBBLEE *iStockPhoto* (*flask*)

[WHAT IT IS]

A COMB OF LIGHT

An optical frequency comb consists of a series of laser pulses that are virtually identical and spaced at very regular intervals, making them useful for precise measurements. The light is called a comb because, unlike a single pulse, the series has a spectrum made up of evenly spaced spikes.





SINGLE PULSE
Although the electric field (*top, green*) of a laser pulse oscillates at regular intervals, such a pulse does not consist of light of one pure frequency. The rise and fall of the wave's envelope (*dashed lines*) can occur only if the light is actually composed of a band of frequencies (*bottom*). The shorter the pulse (*top,* T), the greater the spectral width (*bottom,* $^{1}/_{T}$). The frequencies in a one-femtosecond pulse span about half of the visible spectrum, not counting the low-intensity tails.

MULTIPLE PULSES
You can be forgiven for expecting a train of pulses (*top*) to have the same spectrum as a single pulse. In actuality, the train's spectrum is split into spikes like the teeth of a comb (*bottom*), meaning that the light consists of a series of discrete frequencies and not a continuous band of them. If a pulse occurs every *t* nanoseconds, the frequency comb teeth are spaced by $^{1}/_{t}$ gigahertz. Researchers can thus determine the comb spacing very accurately by measuring the rate at which the laser emits pulses.

LISA APFELBACHER

quency. To know the frequencies of the comb teeth, one must measure that frequency as well as the repetition rate. Measuring the offset frequency was a barrier to progress with optical combs. This barrier fell resoundingly in 2000. It took the combined efforts of scientists from two separate branches of laser research and the discovery of a new material.

Converging Disciplines

For most of the past 40 years, ultrafast-laser researchers—those who focus on making and using the shortest pulses—largely ignored the pulse phase and the theoretical comblike spectrum of an ideal series of pulses. Their experiments typically only depended on the intensity of individual pulses, in which case the phase has no effect. Although the members of the ultrafast community often measured the spectrum of their mode-locked lasers, they rarely did so with sufficient resolution to observe the underlying comb spectrum; instead the lines would blend together and look like a continuous band of frequencies.

High-resolution measurements were the domain of specialists in precision spectroscopy and optical frequency metrology, wherein highly stable CW lasers reigned as the preferred tools. As mentioned earlier, a CW laser sends out a steady stream of light at a precise frequency, and its spectrum looks like one sharp spike. Not many researchers in the metrology community were cognizant of the workings of mode-locked lasers, and those who did know about them were skeptical that such lasers could produce a well-defined comb spectrum in practice. They expected that modest fluctuations in the timing or the phase of the pulses would wash it out.

But a few researchers, most notably Theodor W. Hänsch of the Max Planck Institute for Quantum Optics in Garching, Germany, had faith that mode-locked lasers could one day be a useful tool for high-precision spectroscopy and metrology. In the 1970s, while a faculty member at Stanford University, Hänsch used mode-locked dye lasers (which have a colorful liquid dye as the medium where the laser light is generated) to do a series of measurements that established the basic concept of the comb spectrum and its offset frequency. These seeds then lay dormant for almost 20 years until laser technologies had advanced enough for further progress with combs to be practical.

In the late 1980s Peter Moulton, then at Schwartz Electro-Optics in Concord, Mass., developed titanium-doped sapphire as a laser gain medium with a large bandwidth. Wilson

[THE AUTHORS]

Steven Cundiff, Jun Ye and John Hall bring different backgrounds to their collaboration on developing and applying femtosecond optical frequency combs. Hall has been a leader in precision measurement using ultrastable continuous-wave (CW) lasers for more than four decades. In 2005 he shared the Nobel Prize in Physics for his work, including development of comb techniques. Ye began his career about 15 years ago with a focus on ultrastable CW lasers, but since the advent of comb techniques he has been making significant contributions to the broad field of ultrafast science. Cundiff worked in ultrafast science, mainly spectroscopy but also on mode-locked lasers, before teaming up with Hall and Ye 10 years ago. All three are fellows at JILA, a joint institute between the National Institute of Standards and Technology and the University of Colorado at Boulder.

COMBINED SOUND of two tuning forks, one of them slightly out of tune, produces the phenomenon known as beating: the sound's volume oscillates up and down at a rate that is the beat frequency—the difference in frequency of the two forks. Beating of light waves is used in many laser measurements, including those involving optical combs.

DAVID EMMITE

THE STANDARD SECOND

Optical frequency combs will one day be used for the official standard of time.

- Today the standard is based on the frequency of microwave radiation absorbed by cesium atoms to excite them between two specific "hyperfine" energy states.

- One second is defined to be the time it takes for such light to oscillate precisely 9,192,631,770 times.

- An optical standard would use light emitted or absorbed by some chosen atom or ion at roughly 60,000 times the cesium frequency.

Sibbett of the University of St. Andrews in Scotland pioneered its use in mode-locked lasers in the early 1990s. Within only a few years, titanium-sapphire lasers were routinely generating pulses shorter than 10 femtoseconds, corresponding to only three cycles of light [see "Ultrashort-Pulse Lasers: Big Payoffs in a Flash," by John-Mark Hopkins and Wilson Sibbett; Scientific American, September 2000].

With these titanium-sapphire lasers available, Hänsch dusted off his 20-year-old idea of optical frequency combs. He performed a series of experiments in the late 1990s that demonstrated the latent potential of mode-locked lasers. In one measurement, he showed that comb lines at opposite ends of the output spectrum are well defined with respect to one another. The comb teeth were revealed to be like the marks engraved on a steel ruler and not like lines drawn along a rubber band. In another experiment, he measured the frequency of an optical transition in cesium atoms (a change in their state that absorbs or emits light at a precise frequency) using a mode-locked laser to span the difference in frequency between two CW lasers. His results inspired a group of us to undertake serious research in this arena.

At JILA, a joint institute between the National Institute of Standards and Technology (NIST) and the University of Colorado at Boulder, we were in a unique position to take the technological advances in two branches of laser physics and run with them. JILA has a strong tradition in optical frequency metrology and precision spectroscopy, largely built on the ultrastable CW laser technology developed over 40 years by one of us (Hall). In 1997 another one of us (Cundiff) joined JILA, bringing expertise in mode-locked lasers and short-pulse techniques. It took many hallway and lunch table conversations before we surmounted our conceptual divide and decided to join forces, along with a pair of postdoctoral fellows: Scott Diddams, now at NIST, and David Jones, now at the University of British Columbia. The third of us (Ye) joined the fun at JILA in the summer of 1999, just as the revolution began in earnest; he soon led the way to finding applications for the new frequency combs.

Magic Fiber

As impressive as Hänsch's results were, we knew that his motivation was to dispose of most of his complex apparatus. The techniques to accomplish this simplification, however, required that a mode-locked laser produce an enormous bandwidth, preferably an octave. (An octave is a factor of two in frequency, whether it be in music, electronics or optics.) Although titanium-sapphire lasers produced impressive bandwidth at the time, they could not yet yield an octave of light.

The final puzzle piece fell into place at the 1999 Conference on Lasers and Electro-Optics where Jinendra Ranka of Bell Laboratories presented a paper on a new kind of optical fiber known as microstructure fiber. In this medium, micrometer-size airholes in the fiber guide light along its core. The fiber's properties allow pulses at the frequencies produced by a titanium-sapphire laser to travel along it without being stretched (as occurs in ordinary fiber and most other optical media). The lack of stretching keeps the pulse intensity high, which in turn leads to much greater spectral broadening than occurs in ordinary optical fiber [see "The Ultimate White Light," by Robert R. Alfano; Scientific American, December 2006]. The results are visually stunning. The output of a titanium-doped sapphire laser is in the near-infrared, just beyond the limits of human vision. It appears as a faint red color to the eye. Spectral broadening in microstructure fiber converts that faint red to visible wavelengths, causing the fiber to glow with successive colors of the rainbow.

In the fall of 1999 we managed to acquire some of this magic fiber. The timing could not have been more perfect. We had just completed a series of experiments demonstrating the use of a titanium-sapphire laser to span a gap nearly three times wider than Hänsch's initial demonstration. We already had an operating setup into which we could almost drop the new microstructure fiber. Within two weeks of receiving the express package from Bell Laboratories, we had done a proof-of-principle experiment showing that the spectral broadening in the microstructure fiber preserved the frequency comb structure in the original laser pulse.

The importance of an octave-spanning spectrum is that it allows the offset frequency to be measured directly as a radio frequency, thus surmounting the aforementioned barrier to using combs to measure other frequencies. There are several specific methods of determining the offset frequency given an octave-spanning spectrum, many of which can be traced to techniques employed in radio engineering for measuring frequencies before high-speed counters were readily available. (Counters do the job by simply counting how many oscillations occur in a radio wave per unit of time but cannot keep up with the much higher frequencies that light has.) We will now describe the simplest and most versatile of the methods for measuring the offset frequency—self-referencing.

The key idea is that an octave-spanning spectrum enables scientists to compare the frequencies of two comb lines at opposite ends of the spectrum with each other. If the offset frequency is zero, then each line at the low-frequency end of the spectrum has a corresponding line with exactly twice its frequency at the high-frequency end. Any deviation from this exact ratio turns out to be precisely the offset frequency [*see box below*]. The scheme is called self-referencing because one is comparing the comb's light against itself.

ULTIMATE BREATHALYZER

Optical frequency combs may be used to rapidly detect molecules in a person's breath that could signal a variety of conditions:

METHYLAMINE: Liver and renal diseases

AMMONIA: Renal failure

ETHANE: Some forms of cancer

CARBON ISOTOPE RATIOS: Presence of *Helicobacter pylori*

[MAKING IT USEFUL]

"CALIBRATING" THE COMB

The comb's teeth are offset slightly because of a subtle effect that changes their frequencies. Before scientists can use the combs to measure the light of another laser, they first have to correct for this offset.

THE PROBLEM

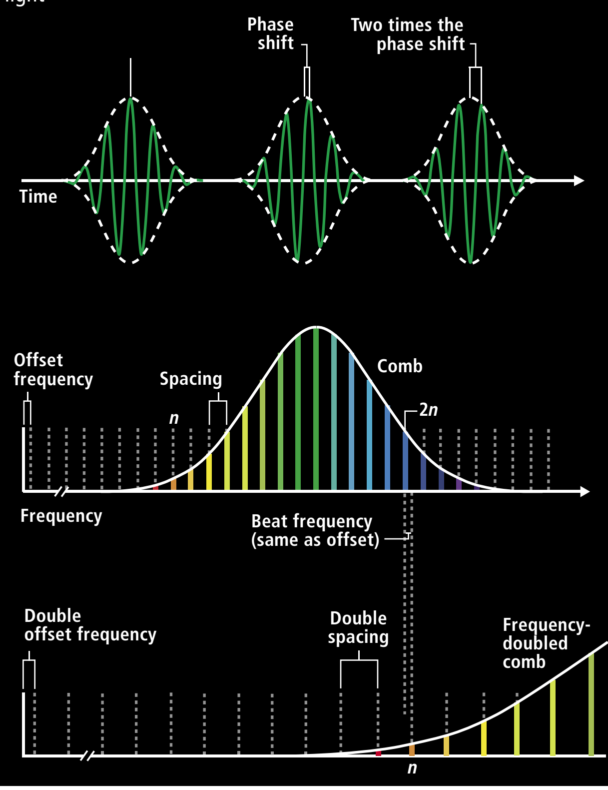
With each successive pulse, the alignment of the highest amplitude of the pulse wave relative to the maximum of the envelope changes, an effect called phase shift.

HOW THE COMB CHANGES

Phase shift moves the optical frequency comb's teeth by an amount called the offset frequency. The comb teeth will be at the offset frequency plus integer multiples of the comb spacing. A technique called self-referencing can determine the offset frequencies; it relies on the optical comb's spanning a full octave—that is, running all the way from one frequency (*red, line* n) to twice that frequency (*violet, line 2*n).

SOLUTION: COMPARE COMBS

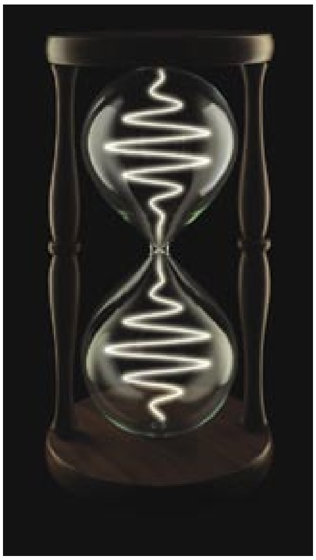
Researchers send part of the comb light through a crystal that produces comb lines at double the original frequencies (and some other lines not shown here). Because the doubled low-frequency lines differ from the original high-frequency lines by the offset frequency, combining them produces beating at just that frequency, which is measured. The researchers now know the precise frequencies of their comb lines.



LISA APFELBACHER

TIME NUTS

While scientists develop clocks based on optical combs, traditional atomic clocks have become a tool for amateur science. At www.leapsecond.com/great2005, Tom Van Baak describes his family's road trip up Mount Rainier in Washington State with three atomic clocks to observe gravitational time dilation (for their trip, 22 nanoseconds) predicted by general relativity.



Self-referencing is carried out in practice by passing some of the laser light through a so-called second-harmonic generation crystal, which doubles the light's frequency. Thus, one can split off the light that forms the lower-frequency end of the comb using a mirror that only reflects longer-wavelength light but passes shorter wavelengths, then send it through the doubling crystal, and finally direct both it and the light of the higher-frequency end of the comb onto the same photodetector. The combined light oscillates in intensity—it "beats"—in just the same fashion as the combined sound of a tuned and a mistuned note beats. In both cases, the frequency of the beats equals the amount of mistuning. For the light pulses, the beats have the same frequency as the comb's offset frequency because every doubled low-end line will be mistuned by that amount from a high-end line. In electronics and optics, this procedure of combining signals to get the beat frequency is called heterodyne detection.

Redefining Time

The simplicity of optical frequency metrology based on optical frequency combs can only be appreciated in comparison to techniques used prior to their development. Briefly, these techniques consisted of frequency multiplication chains, where each link in the chain consisted of an oscillator that had a multiple of the frequency of the previous link. The first link in the chain was a cesium clock, a kind of atomic clock used as the international time standard that defines the second. The cesium clock is based on nine-gigahertz microwaves absorbed by cesium atoms. To reach all the way from nine gigahertz to the frequency of visible light (a factor of at least 40,000) required about a dozen stages. Each stage used a different technology, including lasers for visible light. Running these chains was resource- and personnel-intensive; just a few in the world were built, and measurements were made only intermittently. In addition, in practice the many links in the chain impaired the accuracy of the ultimate optical frequency measurement.

Once stabilized optical frequency combs were invented, it was much easier to precisely measure the frequency of a CW laser. As with a frequency chain, comb-based frequency measurements still must be referenced to a cesium clock. As we will now see, a cesium clock's ability to measure frequencies up to about nine gigahertz is all that you need to use an optical comb to determine the frequency of a laser line. Several pieces of information involving the comb are needed. First, as we discussed earlier, the comb's offset frequency and the spacing of its lines must be measured. From those two numbers the frequencies of all the comb's lines can be calculated. Next, the unknown laser light is combined with the comb's light to get the beat frequency (that is, the difference in frequency) between it and the nearest comb line.

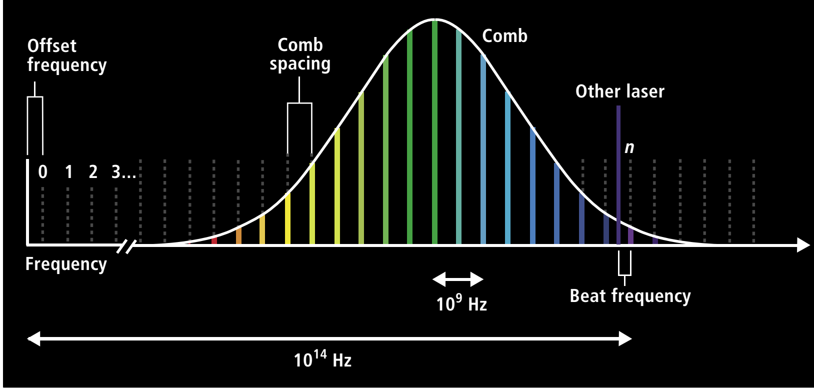
These three frequencies are all within the microwave range that can be measured extremely accurately using a cesium clock. Recall that the comb's line spacing is the same as the repetition rate of the pulses producing the comb. Most mode-locked lasers operate at a repetition rate of 10 gigahertz or less, making that quantity easy to measure against the cesium clock. Both the offset frequency and the beat frequency are also within range to be measured by the cesium clock because they must be smaller than the comb spacing.

Two further pieces of data must be determined: to *which* comb line was the unknown laser light closest and on *which side* of the line? Commercial wave meters can measure an optical line's frequency to within less than one gigahertz, which is good enough to answer those two questions. In the absence of such a wave meter, you can systematically vary the repetition rate and the offset frequency to monitor how the beat frequency changes in response.

[USING IT]

MEASURING LIGHT

To determine the frequency of another laser (*violet*), physicists combine its light with optical comb light and measure the beat frequency generated with the closest comb line (*n*). They can learn which comb tooth the laser is nearest from approximate knowledge of the laser's frequency obtained by standard, less precise techniques. Thus, by measuring three frequencies in the gigahertz range—offset frequency, comb spacing and beat frequency—researchers can determine the frequency of light very precisely in the 100-terahertz, or 10^{14}-hertz, range.



DAVID EMMITE (*hourglass*); LISA APFELBACHER (*graph*)

With enough of those data points, you can work out where the line must be.

The simplicity of optical combs has not only increased how often scientists around the world make these extremely precise frequency measurements but also greatly decreased the uncertainty in those measurements. Such benefits may one day lead to an optical time standard replacing the present microwave cesium-based one. With this in mind, groups at NIST led by James C. Bergquist and at JILA led by Ye have been measuring frequencies relative to clocks that use light and a comb to produce the output signal. Already the uncertainties in measurements using the best of these clocks are smaller than those in measurements using the very best cesium standards. It is an exciting time, with many laboratories around the world poised to build optical frequency standards that can surpass what has been the primary frequency standard for many decades. Measurements by Leo Hollberg's group at NIST, as well as by other groups elsewhere, suggest that the intrinsic limit of the optical comb is still a couple of orders of magnitude better than the uncertainty in current optical frequency measurements.

Higher and Higher

Adopting an optical time standard remains years in the future, however. Metrologists must first carefully evaluate numerous atomic and ionic optical transitions before selecting the one that seems to be the best for a standard.

In addition to the many practical applications of combs, fundamental comb research continues apace on many fronts. For example, Ye's group can use a single comb to detect very sensitively many different transitions of atoms and molecules all at once. Thus, the whole range of energy states of an atom can be analyzed in one measurement. Alternatively, this technique can be applied to detect many trace species in a sample.

Comb technology has already had a large impact on studies of how atoms and molecules respond to the strong electric fields obtainable in intense, ultrashort light pulses. Much of this work has been led by a collaborator of Hänsch's, Ferenc Krausz, who is now at the Max Planck Institute for Quantum Optics. Among other achievements, his group has used the response of electrons to measure the electric field of a laser's ultrashort pulses and display the waveform, much like displaying a radio-frequency wave on an oscilloscope. Krausz used optical combs to stabilize the pulses' phase to have an unchanging waveform from pulse to pulse.

Another very active area of research is the quest to push comb techniques to higher frequencies of the electromagnetic spectrum. (Producing lower-frequency combs, including combs that run from microwaves all the way to visible light, is straightforward.) In 2005 Ye's group at JILA and Hänsch's group in Garching generated a precise frequency comb in the extreme ultraviolet (not far below x-rays in frequency). Scientists are using this extended comb to study the fine structure of atoms and molecules with extreme ultraviolet laser light.

In the space of a few short years, optical frequency combs have gone from being a research problem studied by a small number of scientists to being a tool to be used across a broad gamut of applications and fundamental research. We have only begun to explore the full potential of these rulers of light. ■

GRANDFATHER CLOCK relies on gears of different sizes to turn the steady swinging of its pendulum into the very slow, very precise movements of its hands. Optical combs do something similar by acting as "gears" that turn the high frequencies of optical light into lower frequencies that can be measured—and they are used for time measurements, to boot.

DAVID EMMITE

➡ MORE TO EXPLORE

Time Measurement at the Millennium. James C. Bergquist, Steven R. Jefferts and David J. Wineland in *Physics Today*, Vol. 54, No. 3, pages 37–42; 2001.

Optical Frequency Combs. National Institute of Standards and Technology. Online at **www.nist.gov/public_affairs/newsfromnist_frequency_combs.htm**

Frequency Combs. Max Planck Institute for Quantum Optics. Online at **www.mpq.mpg.de/~haensch/comb/research/combs.html**

Sr Lattice Clock at 1×10^{-16} Fractional Uncertainty by Remote Optical Evaluation with a Ca Clock. A. D. Ludlow et al. in *Science Express;* posted online February 14, 2008.

The Doping Dilem

Game theory helps to explain the pervasive abuse of drugs in cycling, baseball and other sports

By Michael Shermer

For a competitive cyclist, there is nothing more physically crushing and psychologically demoralizing than getting dropped by your competitors on a climb. With searing lungs and burning legs, your body hunches over the handlebars as you struggle to stay with the leader. You know all too well that once you come off the back of the pack the drive to push harder is gone—and with it any hope for victory.

I know the feeling because it happened to me in 1985 on the long climb out of Albuquerque during the 3,000-mile, nonstop transcontinental Race Across America. On the outskirts of town I had caught up with the second-place rider (and eventual winner), Jonathan Boyer, a svelte road racer who was the first American to compete in the Tour de France. About halfway up the leg-breaking climb, that familiar wave of crushing fatigue swept through my legs as I gulped for oxygen in my struggle to hang on.

To no avail. By the top of the climb Boyer was a tiny dot on the shimmering blacktop, and I didn't see him again until the finish line in Atlantic City. Later that night Jim Lampley, the commentator for *ABC's Wide World of Sports,* asked what else I might have done to go faster. "I should have picked better parents," I deadpanned. We all have certain genetic limitations, I went on, that normal training cannot overcome. What else could I have done?

Plenty, and I knew it. Cyclists on the 1984 U.S. Olympic cycling team had told me how they had injected themselves with extra blood before races, either their own—drawn earlier in the season—or that of someone else with the same blood type. "Blood doping," as the practice is called, was not banned at the time, and on a sliding moral scale it seemed only marginally distinguishable from training at high altitude. Either way, you increase the number of oxygen-carrying red blood cells in your body. Still, I was already 30 years old and had an academic career to fall back on. I was racing bikes mostly to see how far I could push my body before it gave out. Enhancing my performance artificially didn't mesh well with my reasons for racing.

But suppose I had been 20 and earning my living through cycling, my one true passion,

JOE ZEFF DESIGN, INC.

ma

with no prospects for some other career. Imagine that my team had made performance-enhancing drugs part of its "medical program" and that I knew I could be cut if I was not competitive. Finally, assume I believed that most of my competitors were doping and that the ones who were tested almost never got caught.

That scenario, in substance, is what many competitive cyclists say they have been facing since the early 1990s. And although the details differ for other sports such as baseball, the overall doping circumstances are not dissimilar [*see box on pages 86 and 87*]. Many players are convinced that "everyone else" takes drugs and so have come to believe that they cannot remain competitive if they do not participate. On the governance side, the failure of Major League Baseball to make the rules clear, much less to enforce them with extensive drug testing throughout the season, coupled with its historical tendency to look the other way, has created an environment conducive to doping.

Naturally, most of us do not want to believe that any of these stellar athletes are guilty of doping. But the convergence of evidence leads me to conclude that in cycling, as well as in baseball, football, and track and field, most of the top competitors of the past two decades have been using performance-enhancing drugs. The time has come to ask not if but why. The reasons are threefold: first, better drugs, drug cocktails and drug-training regimens; second, an arms race consistently won by drug takers over drug testers; and third, a shift in many professional sports that has tipped the balance of incentives in favor of cheating and away from playing by the rules.

Gaming Sports

Game theory is the study of how players in a game choose strategies that will maximize their return in anticipation of the strategies chosen by the other players. The "games" for which the theory was invented are not just gambling games such as poker or sporting contests in which tactical decisions play a major role; they also include deadly serious affairs in which people make economic choices, military decisions and even national diplomatic strategies. What all those "games" have in common is that each player's "moves" are analyzed according to the range of options open to the other players.

The game of prisoner's dilemma is the classic example: You and your partner are arrested for a crime, and you are held incommunicado in separate prison cells. Of course, neither of you wants to confess or rat on the other, but the D.A. gives each of you the following options:

1. If you confess but the other prisoner does not, you go free and he gets three years in jail.
2. If the other prisoner confesses and you do not, you get three years and he goes free.
3. If you both confess, you each get two years.
4. If you both remain silent, you each get a year.

KEY CONCEPTS

- An alarming number of sports—baseball, football, track and field, and especially cycling—have been shaken by doping scandals in recent years.
- Among the many banned drugs in the cycling pharmacopoeia, the most effective is recombinant erythropoietin (r-EPO), an artificial hormone that stimulates the production of red blood cells, thereby delivering more oxygen to the muscles.
- Game theory highlights why it is rational for professional cyclists to dope: the drugs are extremely effective as well as difficult or impossible to detect; the payoffs for success are high; and as more riders use them, a "clean" rider may become so noncompetitive that he or she risks being cut from the team.
- The game theory analysis of cycling can readily be extended to other sports. The results show quantitatively how governing bodies and antidoping agencies can most effectively target efforts to clean up their sports.

—The Editors

In the 1990s something happened to alter the matrix for the doping game: r-EPO.

The table below, called the game matrix, summarizes the four outcomes:

PRISONER'S DILEMMA		MY OPPONENT'S STRATEGY: COOPERATE (remain silent)	MY OPPONENT'S STRATEGY: DEFECT (confess)
MY STRATEGY	COOPERATE (remain silent)	One year in jail (High Payoff)	Three years in jail (Sucker Payoff)
MY STRATEGY	DEFECT (confess)	No jail time (Temptation Payoff)	Two years in jail (Low Payoff)

With those outcomes, the logical choice is to defect from the advance agreement and betray your partner. Why? Consider the choices from the first prisoner's point of view. The only thing the first prisoner cannot control about the outcome is the second prisoner's choice. Suppose the second prisoner remains silent. Then the first prisoner earns the "temptation" payoff (zero years in jail) by confessing but gets a year in jail (the "high" payoff) by remaining silent. The better outcome in this case for the first prisoner is to confess. But suppose, instead, that the second prisoner confesses. Then, once again, the first prisoner is better off confessing (the "low" payoff, or two years in jail) than remaining silent (the "sucker" payoff, or three years in jail). Because the circumstances from the second prisoner's point of view are entirely symmetrical to the ones described for the first, each prisoner is better off confessing no matter what the other prisoner decides to do.

Those preferences are not only theoretical. When test subjects play the game just once or for a fixed number of rounds without being allowed to communicate, defection by confessing is the common strategy. But when testers play the game for an unknown number of rounds, the most common strategy is tit-for-tat: each begins cooperating by remaining silent, then mimics whatever the other player does. Even more mutual cooperation can emerge in many-person prisoner's dilemma, provided the players are allowed to play long enough to establish mutual trust. But the research shows that once defection by confessing builds momentum, it cascades throughout the game.

In cycling, as in baseball and other sports, the contestants compete according to a set of rules. The rules of cycling clearly prohibit the use of performance-enhancing drugs. But because the drugs are so effective and many of them are so difficult (if not impossible) to detect, and because the payoffs for success are so great, the incentive to use banned substances is powerful. Once a few elite riders "defect" from the rules (cheat) by doping to gain an advantage, their rule-abiding competitors must defect as well, leading to a cascade of defection through the ranks. Because of the penalties for breaking the rules, however, a code of silence prevents any open communication about how to reverse the trend and return to abiding by the rules.

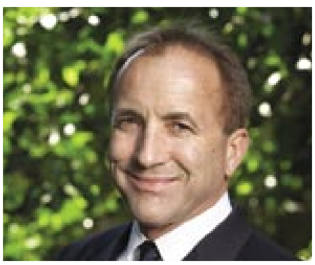
It was not ever thus. Many riders took stimulants and painkillers from the 1940s through the 1980s. But doping regulations were virtually nonexistent until Tom Simpson, a British rider, died while using amphetamines on the climb up Mont Ventoux in the 1967 Tour de France. Even after Simpson's death, doping controls in the 1970s and 1980s were spotty at best. With no clear sense of what counted as following the rules, few perceived doping as cheating. In the 1990s, though, something happened to alter the game matrix.

The EPO Elixir

That "something" was genetically engineered recombinant erythropoietin: r-EPO. Ordinary EPO is a hormone that occurs naturally in the body. The kidneys release it into the bloodstream, which carries it to receptors in the bone marrow. When EPO molecules bind to those receptors, the marrow pumps out more red blood cells. Chronic kidney disease and chemotherapy can cause anemia, and so the development of the EPO substitute r-EPO in the late 1980s proved to be a boon to chronically anemic patients—and to chronically competitive athletes.

Taking r-EPO is just as effective as getting a blood transfusion, but instead of hassling with bags of blood and long needles that must be poked into a vein, the athlete can store tiny ampoules of r-EPO on ice in a thermos bottle or hotel minifridge, then simply inject the hormone under the skin. The effect of r-EPO that matters most to the competitor is directly measurable: the hematocrit (HCT) level, or the percentage by volume of red blood cells in the blood. More red blood cells translate to more oxygen carried to the muscles. For men, the normal HCT percentage range is in the mid-40s. Trained endurance athletes can naturally sustain their HCT in the

[THE AUTHOR]



Michael Shermer (www.michaelshermer.com) is a contributing editor for *Scientific American*, where he writes the regular Skeptic column (*page 44*). He is also a serious amateur cyclist, and in 1982 he was one of only four riders to compete in the first Race Across America (then called the Great American Bike Race). He finished third in that race, completing the 3,000-mile journey in just 10 days, 19 hours and 54 minutes. He is the author of a number of books, including two on the subject of cycling. His most recent book is *The Mind of the Market: Compassionate Apes, Competitive Humans, and Other Tales from Evolutionary Economics* (Times Books, 2007). He is the founder and publisher of *Skeptic* magazine (www.skeptic.com).

BRAD SWONETZ (*Shermer*)

high 40s or low 50s. EPO can push those levels into the high 50s and even the 60s. The winner of the 1996 Tour de France, Bjarne Riis, was nicknamed Mr. 60 Percent; last year he confessed that he owed his extraordinary HCT level to r-EPO.

The drug appears to have made its way into professional cycling in the early 1990s. Greg LeMond thinks it was 1991. Having won the Tour de France in 1986, 1989 and 1990, LeMond set his sights on breaking what would then have been a record of five Tour de France victories, and in the spring of 1991 he was poised to take his fourth. "I was the fittest I had ever been, my split times in spring training rides were the fastest of my career, and I had assembled a great team around me," LeMond told me. "But something was different in the 1991 Tour. There were riders from previous years who couldn't stay on my wheel who were now dropping me on even modest climbs."

LeMond finished seventh in that Tour, vowing to himself that he could win clean the next year. It was not to be. In 1992, he continued, "our [team's] performance was abysmal, and I couldn't even finish the race." Nondoping cyclists were burning out trying to keep up with their doping competitors. LeMond recounted a story told to him by one of his teammates at the time, Philippe Casado. Casado learned from a rider named Laurent Jalabert, who was racing for the

[THE MEDICINE SHOW]

DRUGS IN SPORTS

The choice of drugs for "performance enhancement" that now confronts the athlete is a bewildering array of highly specialized products for various desired effects. In virtually every case, however, the doping athlete runs the risk of side effects that range from the merely inconvenient to the life-threatening. This table is only illustrative; no attempt has been made to compile complete lists. The drugs named in italics are specific examples in a category.

	DRUG OR TREATMENT	COMMON SIDE EFFECTS
ENHANCE OXYGEN TRANSFER	*Recombinant erythropoietin (r-EPO)*	Blood clots, heart attack, stroke
	Transfusion of blood and blood-related products (blood doping)	Allergic reactions, kidney damage, infection (HIV, hepatitis), circulatory overload, blood clots, metabolic shock
BUILD MUSCLE AND BONE	Anabolic steroids: *Androstenedione, testosterone, Winstrol (stanozolol)*	Liver damage, stunted growth, breast development in men, changes in attitude, chloracne on upper body
	Beta$_2$ agonists (especially by mouth or by injection): *Albuterol*	Tremors, headaches, rapid heartbeat
	Hormone modulators: *Human growth hormone (HGH)*	Allergic reactions; can lead to diabetes and acromegaly (soft-tissue swelling, growth of jaw, hands and feet, skull expansion)
	Antiestrogenic agents: *Tamoxifen*	Diverse, such as cancer, liver disease, blood clots
CONTROL WEIGHT	Diuretics (for increasing the rate of urine excretion): *Acetazolamide*	Dehydration, muscle cramps, kidney stones, electrolyte imbalance
MASK DRUG USE	Diuretics (for diluting urine)	See Control Weight (*above*)
	Epitestosterone (to foil a standard test for testosterone doping)	None known
STIMULATE QUICK ENERGY	*Amphetamine* Amphetamine-like agents: *Ephedrine, Ritalin*	Increased heart rate, increased blood pressure, hallucinations, weight loss, tremors
	Epinephrine	Heart irregularity, high blood pressure
AID IN RELAXING	Beta blockers: *Acebutolol, alprenolol, atenolol*	Low blood pressure, slow heart rate
MASK PAIN	Narcotics: *Morphine, oxycodone*	Increased pain threshold, leading to failure to recognize injury; dependence; risk of life-threatening overdose

BRYAN CHRISTIE DESIGN; SOURCES: "2008 GUIDE TO PROHIBITED SUBSTANCES AND PROHIBITED METHODS OF DOPING," EDITED BY LARRY D. BOWERS ET AL. U.S. ANTI-DOPING AGENCY (WWW.USADA.ORG/DRO/RESOURCE/RELATED.ASPX); "THE WORLD ANTI-DOPING CODE: THE 2008 PROHIBITED LIST, INTERNATIONAL STANDARD," WORLD ANTI-DOPING AGENCY (WWW.USADA.ORG/GO/PROHIBITEDLIST)

Spanish cycling team ONCE, that Jalabert's personal doping program was entirely organized by the ONCE team. That program, LeMond said, included r-EPO, which LeMond refused to take, thereby consigning himself to another DNF ("did not finish") in 1994, his final race.

Some who did go along with the pressure to dope paid an even higher price. Casado, for instance, left LeMond's team to join one that had a doping program—and died suddenly in 1995 at age 30. Whether his death resulted directly from doping is not known, but when HCT reaches around 60 percent and higher, the blood becomes so thick that clots readily form. The danger is particularly high when the heart rate slows during sleep—and the resting heart rates of endurance athletes are renowned for measuring in the low 30s (in beats per minute). Two champion Dutch riders died of heart attacks after experimenting with r-EPO. Some riders reportedly began sleeping with a heart-rate monitor hooked to an alarm that would sound when their pulse dropped too low.

Trapped in an Arms Race

Just as in evolution there is an arms race between predators and prey, in sports there is an arms race between drug takers and drug testers. In my opinion, the testers are five years away from catching the takers—and always will be. Those who stand to benefit most from cheating will always be more creative than those enforcing the rules, unless the latter have equivalent incentives. In 1997, because there was no test for r-EPO (that would not come until 2001), the Union Cycliste International (UCI), the sport's governing body, set an HCT limit for men of 50 percent. Shortly afterward, riders figured out that they could go higher than 50, then thin their blood at test time with a technique already allowed and routinely practiced: injections of saline water for rehydration. Presto change-o.

Willy Voet, the *soigneur,* or all-around caretaker, for the Festina cycling team in the 1990s, explained how he beat the testers in his tell-all book, *Breaking the Chain:*

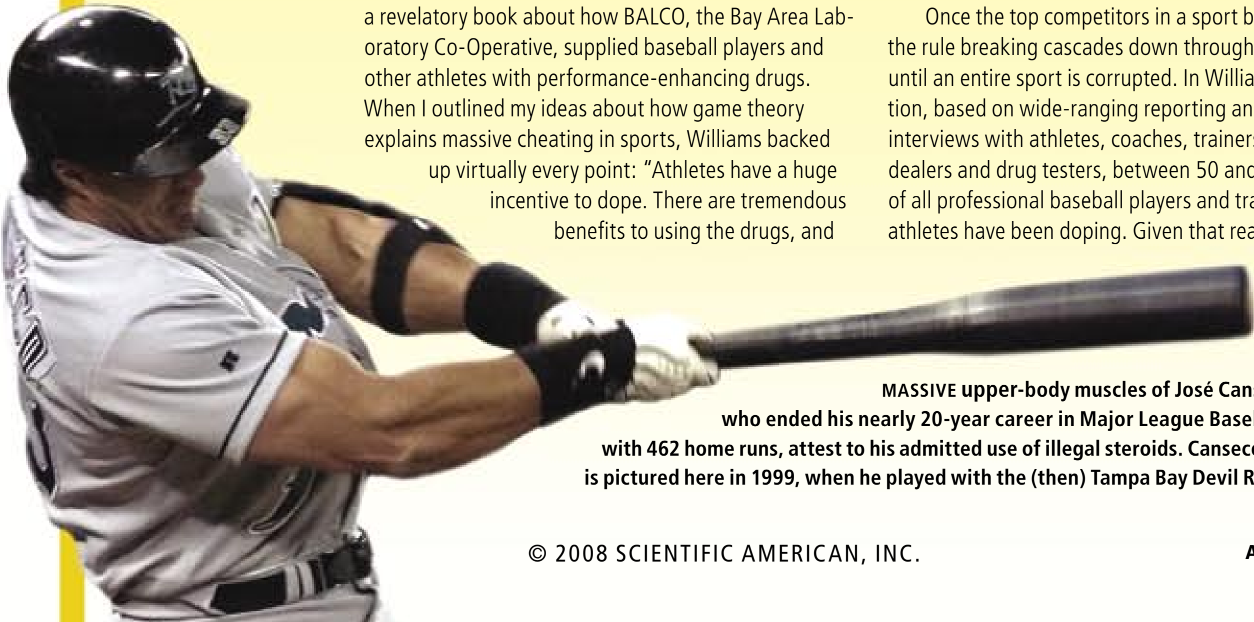
> Just in case the UCI doctors arrived in the morning to check the riders' hematocrit levels, I got everything ready to get them through the tests.... I went up to the cyclists' rooms with sodium drips.... The whole transfusion would take twenty minutes, the saline diluting the blood and so reducing the hematocrit level by three units—just enough.
>
> This contraption took no more than two minutes to set up, which meant we could put it into action while the UCI doctors waited for the riders to come down from their rooms.

How did the new rules of the doping game change the players' strategies? I put the question

Gaming Baseball: Why Players Dope

A game theory model of doping in cycling applies to other sports as well, particularly baseball. For expert insight, I spoke with Lance Williams, an investigative reporter with the *San Francisco Chronicle* and co-author (with Mark Fainaru-Wada) of *Game of Shadows,* a revelatory book about how BALCO, the Bay Area Laboratory Co-Operative, supplied baseball players and other athletes with performance-enhancing drugs. When I outlined my ideas about how game theory explains massive cheating in sports, Williams backed up virtually every point: "Athletes have a huge incentive to dope. There are tremendous benefits to using the drugs, and there is only a small chance that you will get caught. So depending on your sport and where you are in your career, the risk is often worth it. If you make the team, you'll be a millionaire; if you don't, you'll probably go back to driving a delivery truck."

Once the top competitors in a sport begin cheating, the rule breaking cascades down through the ranks until an entire sport is corrupted. In Williams's estimation, based on wide-ranging reporting and numerous interviews with athletes, coaches, trainers, drug dealers and drug testers, between 50 and 80 percent of all professional baseball players and track-and-field athletes have been doping. Given that reality, Williams



MASSIVE upper-body muscles of José Canseco, who ended his nearly 20-year career in Major League Baseball with 462 home runs, attest to his admitted use of illegal steroids. Canseco is pictured here in 1999, when he played with the (then) Tampa Bay Devil Rays.

KEVIN FRAYER *AP Photo*

directly to Joe Papp, a 32-year-old professional cyclist currently banned after testing positive for synthetic testosterone. Recalling the day he was handed the “secret black bag,” Papp explained how a moral choice becomes an economic decision: “When you join a team with an organized doping program in place, you are simply given the drugs and a choice: take them to keep up or don’t take them and there is a good chance you will not have a career in cycling.”

When Papp came clean, professional cycling slapped him with a two-year ban. But the social consequences were far worse than that. “The sport spit me out,” he lamented to me. “A team becomes a band of brothers,… but with a team of dopers there’s an additional bond—a shared secret—and with that there is a code of silence. If you get busted, you keep your mouth shut. The moment I confessed I was renounced by my friends because in their mind I put them at risk. One guy called and threatened to kill me if I revealed that he doped.”

Papp was never a Tour-caliber cyclist, however, so perhaps the game matrix—with its implications for the rider’s own cycling career—is different at the elite level. Not so, as I learned from another insider. “For years I had no trouble doing my job to help the team leader,” said Frankie Andreu, who was the *superdomestique,* or lead pacer, supporting Lance Armstrong throughout much of the 1990s. “Then, around 1996, the speeds of the races shifted dramatically upward. Something happened, and it wasn’t just training.” Andreu resisted the temptation as long as he could, but by 1999 he could no longer do his job: “It became apparent to me that enough of the *peloton* [the main group of riders in a cycling race] was on the juice that I had to do something.” He began injecting himself with r-EPO two to three times a week. “It’s not like Red Bull, which gives you instant energy,” he explained. “But it does allow you to dig a little deeper, to hang on to the group a little longer, to go maybe 31.5 miles per hour instead of 30 mph.”

The Doping Difference

One of the subtle benefits of r-EPO in a brutal three-week race like the Tour de France is not just boosting HCT levels but keeping them high. Jonathan Vaughters, a former teammate of Armstrong’s, crunched the numbers for me this way: “The big advantage of blood doping is the ability to keep a 44 percent HCT over three weeks.” A “clean” racer who started with a 44 percent HCT, Vaughters noted, would expect to end up at 40 percent after three weeks of racing because of natural blood dilution and the breakdown of red blood cells. “Just stabilizing [your HCT level] at 44 percent is a 10 percent advantage.”

Scientific studies on the effects of performance-enhancing drugs are few in number and are usually conducted on nonathletes or recreational ones, but they are consistent with Vaughters’s assessment. (For obvious reasons, elite athletes who dope are disinclined to disclose their data.) The consensus among the sports physiologists I interviewed is that r-EPO improves performance by at least 5 to 10 percent. When it is mixed in with a brew of other drugs, another 5 to 10 percent boost can be squeezed out of the human engine. In events decided by differences of less than 1 percent, this advantage is colossal.

Italian sports physiologist Michele Ferrari, as knowledgeable on doping as he is controversial (because of his close affiliation with elite athletes who have tested positive for doping or been accused of same), explains it this way: “If the volume of [red blood cells] increases by 10 percent, performance [the rider’s net gain in output of useful kinetic energy] improves by approximately 5 percent. This means a gain of about 1.5 seconds per kilometer for a cyclist pedaling at 50 kilometers per hour in a time trial, or about eight seconds per kilometer for a cyclist climbing at 10 kph on a 10 percent ascent.”

In the Tour de France, those numbers imply

explained, for many athletes not only is doping not cheating, it is necessary. To illustrate his point, Williams cited Charlie “the Chemist” Francis, coach of the sprinter and (briefly) the 1988 Olympic gold medalist in the 100-meter run, Ben Johnson, who was busted for doping and stripped of his medals. The doping was “completely self-defensive,” Francis told Williams. “It was cheat or lose.”

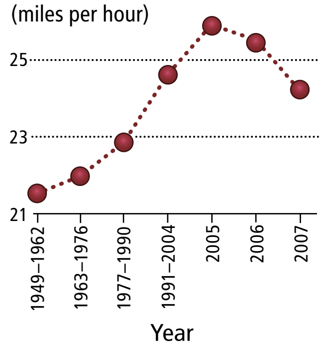
How can leagues and governing bodies change the incentives in the game matrix for baseball? Williams suggested stronger penalties against both individual athletes and entire teams, along with stiffer legal sanctions. Ironically, Williams and Fainaru-Wada personally felt legal heat for refusing to reveal their sources to federal authorities: “We stood to serve more time behind bars [both faced 18-month sentences] than any of the steroid dealers convicted in the BALCO conspiracy and any of the athletes we reported on, including [longtime San Francisco Giants slugger] Barry Bonds.” You know a system is corrupt when the messenger is shot and the gun makers walk. —***M.S.***

WHAT A DIFFERENCE A DRUG MAKES

Average speeds of the winners of the Tour de France took a quantum leap upward beginning in 1991, when recombinant erythropoietin, or r-EPO, apparently came into wide use in cycling.

To control for yearly variance caused by course changes and weather, the author averaged the speeds over 14-year periods going backward and forward in time from 1991; the averages are plotted here.

Average Tour de France Speeds
(miles per hour)



- As the graph shows, in the period 1991–2004 the winners’ average speed jumped substantially (8 percent) over the corresponding speed in the period 1977–1990, an increase that cannot be accounted for by improvements in equipment, nutrition or training.
- Even after 2004, when stricter controls were put in place, doping seems to have continued unabated: as the graph shows, the 2005 and 2006 Tour races were the fastest ever.
- Perhaps the massive disqualification of dopers in the 2007 race finally made a difference: the average speed of that race was below the 1991–2004 average.

JEN CHRISTIANSEN

[THE DOPING GAME]

PAYOFFS THAT MAKE CHEATERS INTO LOSERS

Game Assumptions: Current Competition

- Value of winning the Tour de France: $10 million
- Likelihood that a doping rider will win the Tour de France against nondoping competitors: 100%
- Value of cycling professionally for a year, when the playing field is level: $1 million
- Cost of getting caught cheating (penalties and lost income): $1 million
- Likelihood of getting caught cheating: 10%
- Cost of getting cut from a team (forgone earnings and loss of status): $1 million
- Likelihood that a nondoping rider will get cut from a team for being noncompetitive: 50%

FAVORS CHEATING

		MY OPPONENT'S STRATEGY	
		CASE I COOPERATE (abide by rules)	CASE II DEFECT (cheat with drugs)
MY STRATEGY	COOPERATE (abide by rules)	$1 million (High Payoff)	–$0.4 million (Sucker Payoff)
MY STRATEGY	DEFECT (cheat with drugs)	$8.9 million (Temptation Payoff)	$0.8 million (Low Payoff)

Why do cyclists cheat? The game theory analysis of doping in cycling (*below*), which is closely modeled on the game of prisoner's dilemma, shows why cheating by doping is rational, based solely on the incentives and expected values of the payoffs built into current competition. (The expected value is the value of a successful outcome multiplied by the probability of achieving that outcome.) The payoffs assumed are not unrealistic, but they are given only for illustration; the labels "high," "temptation," "sucker" and "low" in the matrices correspond to the standard names of strategies in prisoner's dilemma. It is also assumed that if competitors are playing "on a level playing field" (all are cheating, or all are rule-abiding), their winnings will total $1 million each, without further adjustment for a doping advantage.

—***Peter Brown, staff editor***

CASE I

My opponent abides by the rules (he "cooperates"). I have two options:

High Payoff

I abide by the rules (I "cooperate," too). The playing field is level.

Value of competing for one year: **$1 million**

Since I am not cheating, I expect no penalties: **$0**

Total expected High Payoff: **$1 million**

Temptation Payoff

I cheat by doping (I "defect").

Expected value of winning the Tour de France (if I do not get caught cheating): $10 million × 90% = **$9.0 million**

Expected penalty for cheating (if I do get caught): $1 million × 10% = **–$0.1 million**

Total expected Temptation Payoff: **$8.9 million**

Because $8.9 million is greater than $1 million, my incentive in case I is to cheat.

CASE II

My opponent cheats by doping (he "defects"). Again, I have two options:

Sucker Payoff

I abide by the rules (I "cooperate"). I can earn the average winnings for a competitive racer only if my opponent gets caught cheating and is disqualified.

Expected value of competing for one year: $1 million × 10% = **$0.1 million**

Expected cost of getting cut from a team: $1 million × 50% = **–$0.5 million**

Total expected Sucker Payoff: **–$0.4 million**

Low Payoff

I also cheat by doping (I "defect"). The playing field is level.

Expected value of competing for one year (if I do not get caught): $1 million × 90% = **$0.9 million**

Expected penalty for cheating (if I do get caught): $1 million × 10% = **–$0.1 million**

Total expected Low Payoff: **$0.8 million**

My incentive in case II is also to cheat.

Game Assumptions: After Reforms

- New, higher cost of getting caught cheating (penalties and lost income): $5 million
- New, higher likelihood of getting caught cheating: 90%
- Consequent new, lower likelihood that a nondoping rider will get cut from a team for being noncompetitive: 10%

FAVORS PLAYING BY RULES

		MY OPPONENT'S STRATEGY	
		CASE I COOPERATE (abide by rules)	CASE II DEFECT (cheat with drugs)
MY STRATEGY	COOPERATE (abide by rules)	$1 million (High Payoff)	$0.8 million (Sucker Payoff)
MY STRATEGY	DEFECT (cheat with drugs)	–$3.5 million (Temptation Payoff)	–$4.4 million (Low Payoff)

To reform cycling and encourage cyclists to play by the rules, the expected values of the doping game must be changed. Here a stricter drug-testing regimen dramatically increases the likelihood of getting caught, and the penalty for doping is raised from temporary to lifetime suspension of the rider.

CASE I

My opponent abides by the rules (he "cooperates"). I have two options:

High Payoff

I abide by the rules (I "cooperate," too). The playing field is level.

Value of competing for one year: **$1 million**

Since I am not cheating, I expect no penalties: **$0**

Total expected High Payoff: **$1 million**

Temptation Payoff

I cheat by doping (I "defect").

Expected value of winning the Tour de France (if I do not get caught cheating): $10 million × 10% = **$1.0 million**

Expected penalty for cheating (if I do get caught): $5 million × 90% = **–$4.5 million**

Total expected Temptation Payoff: **–$3.5 million**

Because earning $1 million is better than losing $3.5 million, my incentive in case I has changed to abiding by the rules.

CASE II

My opponent cheats by doping (he "defects"). Again, I have two options:

Sucker Payoff

I abide by the rules (I "cooperate"). I can earn the average winnings for a competitive racer only if my opponent gets caught cheating and is disqualified.

Expected value of competing for one year: $1 million × 90% = **$0.9 million**

Expected cost of getting cut from a team: $1 million × 10% = **–$0.1 million**

Total expected Sucker Payoff: **$0.8 million**

Low Payoff

I also cheat by doping (I "defect"). The playing field is level.

Expected value of competing for one year (if I do not get caught): $1 million × 10% = **$0.1 million**

Expected penalty for cheating (if I do get caught): $5 million × 90% = **–$4.5 million**

Total expected Low Payoff: **–$4.4 million**

My incentive in case II has also changed to playing by the rules.

that a cyclist who boosts his HCT by 10 percent will cut his own time by 75 seconds in a 50-kilometer (31-mile) time trial, a race typically decided by a few seconds. On any of the numerous 10-kilometer (six-mile) climbs in the Alps and the Pyrenees, on grades as steep as 10 percent, that same blood difference would gain the rider a whopping 80 seconds per climb. If any of the top cyclists are on the juice, their erstwhile competitors cannot afford to give away such margins. That is where the game matrix kicks into defection mode.

Nash Equilibrium

In game theory, if no player has anything to gain by unilaterally changing strategies, the game is said to be in a Nash equilibrium. The concept was identified by mathematician John Forbes Nash, Jr., who was portrayed in the film *A Beautiful Mind*. To end doping in sports, the doping game must be restructured so that competing clean is in a Nash equilibrium. That is, the governing bodies of each sport must change the payoff values of the expected outcomes identified in the game matrix [*see box on opposite page*]. First, when other players are playing by the rules, the payoff for doing likewise must be greater than the payoff for cheating. Second, and perhaps more important, even when other players are cheating, the payoff for playing fair must be greater than the payoff for cheating. Players must not feel like suckers for following the rules.

In the game of prisoner's dilemma, lowering the temptation to confess and raising the payoff for keeping silent if the other prisoner confesses increases cooperation. Giving players the chance to communicate before they play the game is the most effective way to increase their cooperation. In sports, that means breaking the code of silence. Everyone must acknowledge there is a problem to be solved. Then drug testing must be done and the results communicated regularly and transparently to all until the test results are clean. That will show each player that the payoff for playing fair is greater than the payoff for cheating, no matter what the other players do.

Here are my recommendations for how cycling (and other sports) can reach a Nash equilibrium in which no one has any incentive to cheat by doping:

- Grant immunity to all athletes for past (pre-2008) cheating. Because the entire system is corrupt and most competitors have been doping, it accomplishes nothing to strip the winner of a title after the fact when it is almost certain that the runners-up were also doping. With immunity, retired athletes may help to improve the antidoping system.
- Increase the number of competitors tested—in competition, out of competition, and especially immediately before or after a race—to thwart countermeasures. Testing should be done by independent drug agencies not affiliated with any sanctioning bodies, riders, sponsors or teams. Teams should also employ independent drug-testing companies to test their own riders, starting with a preseason performance test on each athlete to create a baseline profile. Corporate sponsors should provide additional financial support to make sure the testing is rigorous.
- Establish a reward, modeled on the X prizes (cash awards offered for a variety of technical achievements), for scientists to develop tests that can detect currently undetectable doping agents. The incentive for drug testers must be equal to or greater than that for drug takers.
- Increase substantially the penalty for getting caught: one strike and you're out—forever. To protect the athlete from false positive results or inept drug testers (both exist), the system of arbitration and appeals must be fair and trusted. Once a decision is made, however, it must be substantive and final.
- Disqualify all team members from an event if any member of the team tests positive for doping. Compel the convicted athlete to return all salary paid and prize monies earned to the team sponsors. The threat of this penalty will bring the substantial social pressures of "band of brothers" psychology to bear on all the team members, giving them a strong incentive to enforce their own antidoping rules.

That may sound utopian. But it can work. Vaughters, who is now director of the U.S. cycling team Slipstream/Chipotle, has already started a program of extensive and regular in-house drug testing. "Remember, most of these guys are athletes, not criminals," he says. "If they believe the rest are stopping [the doping] and feel it in the speed of the *peloton*, they will stop, too, with a great sigh of relief."

Hope springs eternal. But with these changes I believe the psychology of the game can be shifted from defection to cooperation. If so, sports can return to the tradition of rewarding and celebrating excellence in performance, enhanced only by an athlete's will to win. ■

One way to discourage cheating is to raise the penalty: one strike, and you're out—forever.

JOE ZEFF DESIGN, INC.

➥ MORE TO EXPLORE

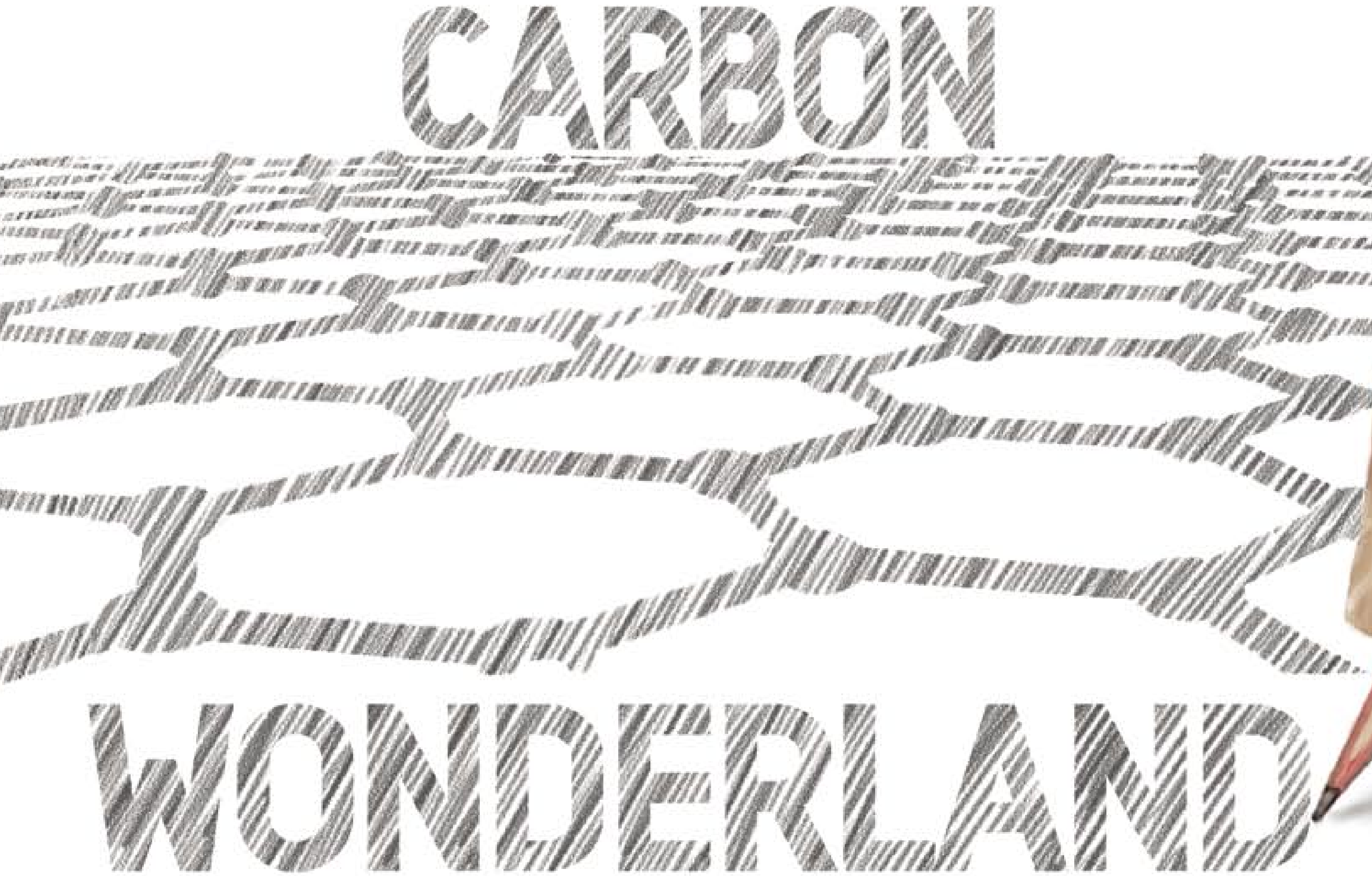
Breaking the Chain: Drugs and Cycling: The True Story. Willy Voet. Random House UK, 2002.

Behavioral Game Theory: Experiments in Strategic Interaction. Colin F. Camerer. Princeton University Press, 2003.

From Lance to Landis: Inside the American Doping Controversy at the Tour de France. David Walsh. Ballantine Books, 2007.

Rough Ride: Behind the Wheel with a Pro Cyclist. Paul Kimmage. Random House UK, 2007.

The Tour de France: A Cultural History. Christopher S. Thompson. University of California Press, 2008.



Graphene, a newly isolated form of carbon, provides a rich lode of novel fundamental physics and practical applications

BY ANDRE K. GEIM AND PHILIP KIM

Consider the humble pencil. It may come as a surprise to learn that the now common writing instrument at one time topped the list of must-have, high-tech gadgets. In fact, the simple pencil was once even banned from export as a strategic military asset. But what is probably more unexpected is the news that every time someone scribes a line with a pencil, the resulting mark includes bits of the hottest new material in physics and nanotechnology: graphene.

Graphene comes from graphite, the "lead" in a pencil: a kind of pure carbon formed from flat, stacked layers of atoms. The tiered structure of graphite was discerned centuries ago, and so it was natural for physicists and materials scientists to try splitting the mineral into its constituent sheets—if only to study a substance whose geometry might turn out to be so elegantly simple. Graphene is the name given to one such sheet. It is made up entirely of carbon atoms bound together in a network of repeating hexagons within a single plane just one atom thick.

For years, however, all attempts to make graphene ended in failure. The most popular early approach was to insert various molecules between the atomic planes of graphite to wedge the planes apart—a technique called chemical exfoliation. Although graphene layers almost certainly detached from the graphite at some transient stage of the process, they were never identified as such. Instead the final product usually emerged as a slurry of graphitic particles—not much different from wet soot. The early interest in chemical exfoliation faded away.

Soon thereafter experimenters attempted a more direct approach. They split graphite crystals into progressively thinner wafers by scraping or rubbing them against another surface. In spite of its crudeness, the technique, known as

EVERYDAY PENCIL MARKS include minute quantities of graphene, one of the hottest "new" materials in science and engineering.

micromechanical cleavage, worked surprisingly well. Investigators managed to peel off graphite films made up of fewer than 100 atomic planes. By 1990, for example, German physicists at RWTH Aachen University had isolated graphite films thin enough to be optically transparent.

A decade later one of us (Kim), working with Yuanbo Zhang, then a graduate student at Columbia University, refined the micromechanical cleavage method to create a high-tech version of the pencil—a "nanopencil," of course. "Writing" with the nanopencil yielded slices of graphite just a few tens of atomic layers thick [*see box on page 93*]. Still, the resulting material was thin graphite, not graphene. No one really expected that such a material could exist in nature.

That pessimistic assumption was put to rest in 2004. One of us (Geim), in collaboration with then postdoctoral associate Kostya S. Novoselov and his co-workers at the University of Manchester in England, was studying a variety of approaches to making even thinner samples of graphite. At that time, most laboratories began such attempts with soot, but Geim and his colleagues serendipitously started with bits of debris left over after splitting graphite by brute force. They simply stuck a flake of graphite debris onto plastic adhesive tape, folded the sticky side of the tape over the flake and then pulled the tape apart, cleaving the flake in two. As the experimenters repeated the process, the resulting fragments grew thinner [*see box on page 95*]. Once the investigators had many thin fragments, they meticulously examined the pieces—and were astonished to find that some were only one atom thick. Even more unexpectedly, the newly identified bits of graphene turned out to have high crystal quality and to be chemically stable even at room temperature.

The experimental discovery of graphene led to a deluge of international research interest. Not only is it the thinnest of all possible materials, it is also extremely strong and stiff. Moreover, in its pure form it conducts electrons faster at room temperature than any other substance. Engineers at laboratories worldwide are currently scrutinizing the stuff to determine whether it can be fabricated into products such as supertough composites, smart displays, ultrafast transistors and quantum-dot computers.

In the meantime, the peculiar nature of graphene at the atomic scale is enabling physicists to delve into phenomena that must be described by relativistic quantum physics. Investigating such

KEY CONCEPTS

- Graphene is a one-atom-thick sheet of carbon that stacks with other such sheets to form graphite—pencil "lead." Physicists have only recently isolated the material.
- The pure, flawless crystal conducts electricity faster at room temperature than any other substance.
- Engineers envision a range of products made of graphene, such as ultrahigh-speed transistors. Physicists are finding the material enables them to test a theory of exotic phenomena previously thought to be observable only in black holes and high-energy particle accelerators.

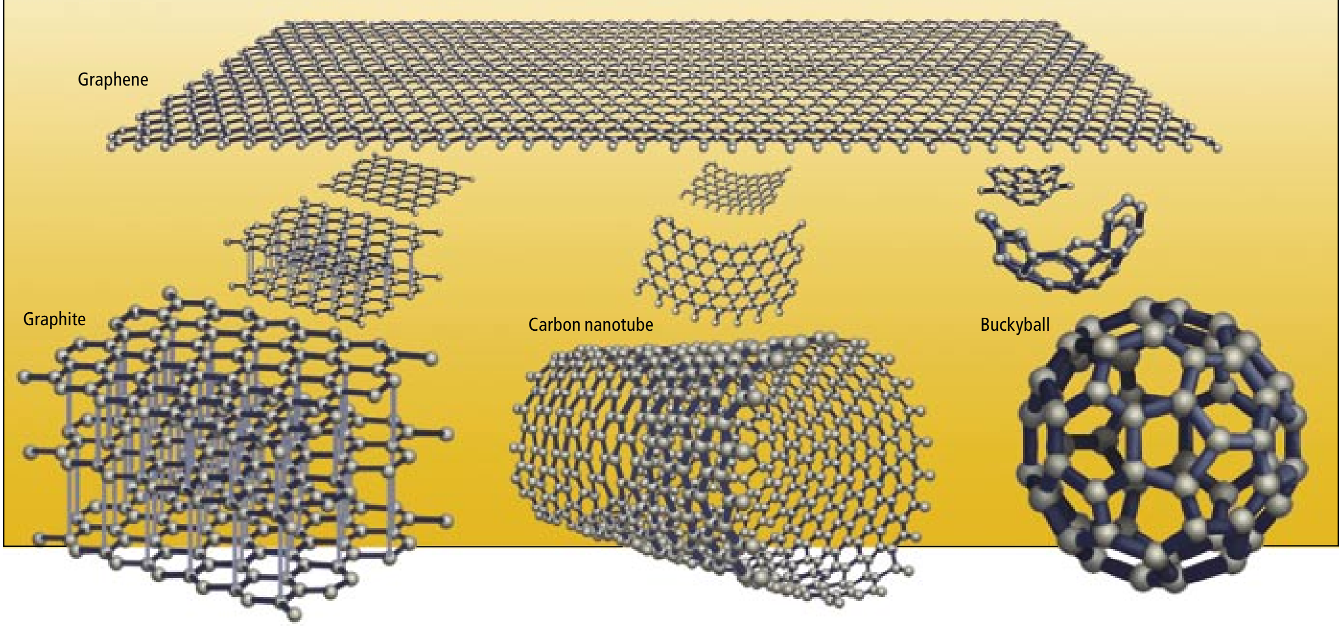
—The Editors

MATT COLLINS

[MOLECULAR FORMS]

THE MOTHER OF ALL GRAPHITES

Graphene (*below, top*), a plane of carbon atoms that resembles chicken wire, is the basic building block of all the "graphitic" materials depicted below. Graphite (*bottom row at left*), the main component of pencil "lead," is a crumbly substance that resembles a layer cake of weakly bonded graphene sheets. When graphene is wrapped into rounded forms, fullerenes result. They include honeycombed cylinders known as carbon nanotubes (*bottom row at center*) and soccer ball–shaped molecules called buckyballs (*bottom row at right*), as well as various shapes that combine the two forms.



MATT COLLINS

phenomena, some of the most exotic in nature, has heretofore been the exclusive preserve of astrophysicists and high-energy particle physicists working with multimillion-dollar telescopes or multibillion-dollar particle accelerators. Graphene makes it possible for experimenters to test the predictions of relativistic quantum mechanics with laboratory benchtop apparatus.

Meet the Graphene Family

Given how widespread the pencil is today, it seems remarkable that what became known as graphite did not play a role in ancient literate civilizations such as those of China or Greece. Not until the 16th century did the English discover a large deposit of pure graphite, then called *plumbago* (Latin for "lead ore"). Its utility as a marker was immediately apparent, though, and the English wasted no time in making it into an easy-to-use substitute for quill and ink. The pencil soon became all the rage among the European intelligentsia.

But it was not until 1779 that Swedish chemist Carl Scheele showed that *plumbago* is carbon, not lead. A decade later German geologist Abraham Gottlob Werner suggested that the substance could more appropriately be called graphite, from the Greek word meaning "to write." Meanwhile munitions makers had discovered another use for the crumbly mineral: they found it made an ideal lining in casting molds for cannonballs. That use became a tightly guarded military secret. During the Napoleonic Wars, for instance, the English Crown embargoed the sale to France of both graphite and pencils.

In recent decades graphite has reclaimed some of its once lofty technological status, as investigators have explored the properties and potential applications of several previously unrecognized molecular forms of carbon that occur in ordinary graphitic materials. The first of them, a soccer ball–shaped molecule dubbed the buckyball, was discovered in 1985 by American chemists Robert Curl and Richard E. Smalley, along with their English colleague Harry Kroto. Six years later Sumio Iijima, a Japanese physicist, identified the honeycombed, cylindrical assemblies of carbon atoms known as carbon nanotubes. Although nanotubes had been reported by many investigators in earlier decades, their importance had not been appreciated. Both the new molecular forms were classified as fullerenes. (That name and the term "buckyball" were coined in honor of the visionary U.S. architect and engineer Buckminster Fuller, who investigated those shapes before the carbon forms themselves were discovered.)

The discovery of graphene has led to a deluge of international research interest.

Molecular Chicken Wire

Graphite, the fullerenes and graphene share the same basic structural arrangement of their constituent atoms. Each structure begins with six carbon atoms, tightly bound together chemically in the shape of a regular hexagon—what chemists call a benzene ring.

At the next level of organization is graphene itself, a large assembly of benzene rings linked in a sheet of hexagons that resembles chicken wire [*see box on opposite page*]. The other graphitic forms are built up out of graphene. Buckyballs and the many other nontubular fullerenes can be thought of as graphene sheets wrapped up into atomic-scale spheres, elongated spheroids, and the like. Carbon nanotubes are essentially graphene sheets rolled into minute cylinders. And as we mentioned earlier, graphite is a thick, three-dimensional stack of graphene sheets; the sheets are held together by weak, attractive intermolecular forces called van der Waals forces. The feeble coupling between neighboring graphene sheets is what enables graphite to be broken so easily into minuscule wafers that make up the mark left on paper when someone writes with a pencil.

With the benefit of hindsight, it is clear that fullerenes, despite going unnoticed until recently, have been close at hand all along. They occur, for instance, in the soot that coats every barbecue grill, albeit in minute quantities. Just so, bits of graphene are undoubtedly present in every pencil mark—even though they, too, long went undetected. But since their discovery, the scientific community has paid all these molecules a great deal of attention.

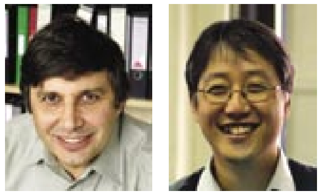
Buckyballs are notable mainly as an example of a fundamentally new kind of molecule, although they may also have important applications, notably in drug delivery. Carbon nanotubes combine a suite of unusual properties—chemical, electronic, mechanical, optical and thermal—that have inspired a wide variety of innovative potential applications. Those innovations include materials that might replace silicon in microchips and fibers that might be woven into lightweight, ultrastrong cables. Although graphene itself—the mother of all graphitic forms—became part of such visions just a few years ago, it seems likely that the material will offer even more insights into basic physics and more intriguing technological applications than its carbonaceous cousins.

Exceptional Exception

Two features of graphene make it an exceptional material. First, despite the relatively crude ways it is still being made, graphene exhibits remarkably high quality—resulting from a combination of the purity of its carbon content and the orderliness of the lattice into which its carbon atoms are arranged. Investigators have so far failed to find a single atomic defect in graphene—say, a vacancy at some atomic position in the lattice or an atom out of place. That perfect crystalline order seems to stem from the strong yet highly flexible interatomic bonds, which create a substance harder than diamond yet allow the planes to bend when mechanical force is applied. The flexibility enables the structure to accommodate a good deal of deformation before its atoms must reshuffle to adjust to the strain.

The quality of its crystal lattice is also respon-

[THE AUTHORS]

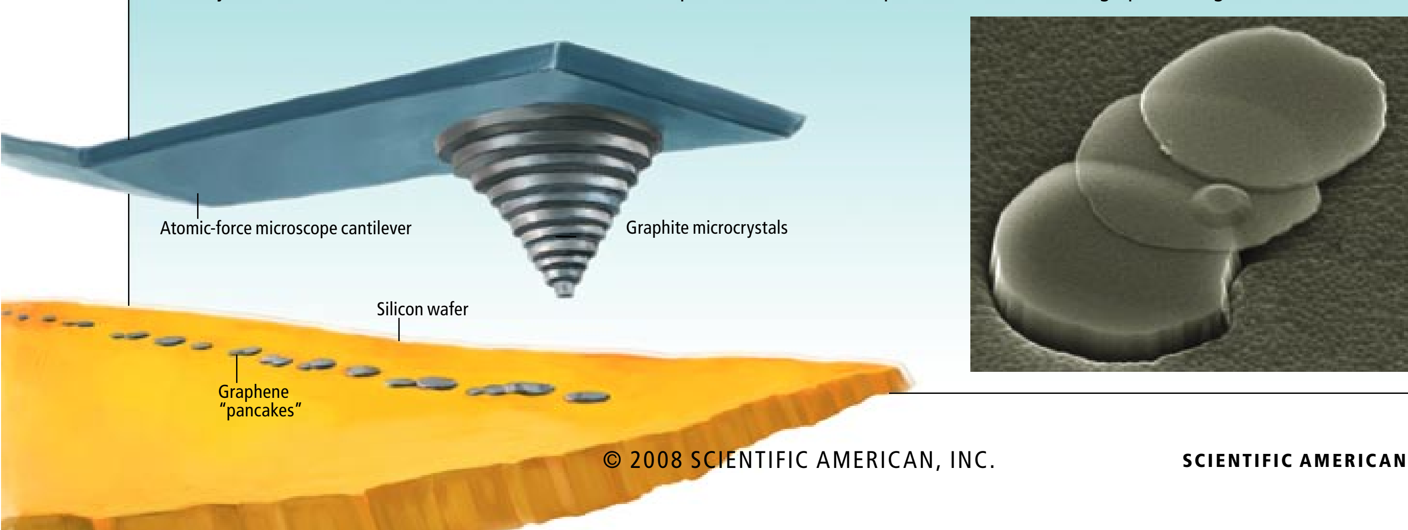


Andre K. Geim (*left*) **and Philip Kim** (*right*) **are condensed matter physicists who in recent years have investigated the nanoscale properties of one-atom-thick, "two-dimensional" crystalline materials. Geim is a fellow of the Royal Society and Langworthy Professor of Physics at the University of Manchester in England. He also directs the Manchester Center for Mesoscience and Nanotechnology. Geim received his Ph.D. from the Institute of Solid State Physics in Chernogolovka, Russia. Kim, a fellow of the American Physical Society who received his doctoral degree from Harvard University, is associate professor of physics at Columbia University. His research focuses on quantum thermal and electrical transport processes in nanoscale materials.**

[GRAPHENE IN THE MAKING]

MARK OF THE NANOPENCIL

Making graphitic samples that approach the thickness of single-layer graphene has taken considerable effort. One way is to attach a graphite microcrystal to the cantilever arm of an atomic-force microscope and scratch the tip of the microcrystal across a silicon wafer (*left*). This "nanopencil" deposits thin graphene "pancakes" onto the wafer (*right*). The samples in the electron micrograph are magnified 6,000×.



MATT COLLINS (*drawings*); KOSTYA NOVOSELOV *University of Manchester* (*micrograph*); COURTESY OF ANDRE K. GEIM (*Geim*); COURTESY OF PHILIP KIM (*Kim*)

Interpreting quantum electrodynamics never comes without a good deal of wrestling with ordinary intuition.

sible for the remarkably high electrical conductivity of graphene. Its electrons can travel without being scattered off course by lattice imperfections and foreign atoms. Even the jostling from the surrounding carbon atoms, which electrons in graphene must endure at room temperature, is relatively small because of the high strength of the interatomic bonds.

The second exceptional feature of graphene is that its conduction electrons, besides traveling largely unimpeded through the lattice, move much faster and as if they had far less mass than do the electrons that wander about through ordinary metals and semiconductors. Indeed, the electrons in graphene—perhaps "electric charge carriers" is a more appropriate term—are curious creatures that live in the weird world where rules analogous to those of relativistic quantum mechanics play an important role. That kind of interaction inside a solid, so far as anyone knows, is unique to graphene. Thanks to this novel material from a pencil, relativistic quantum mechanics is no longer confined to cosmology or high-energy physics; it has now entered the laboratory.

Big Bang in Carbon Flatland

To appreciate the weird behavior of the electric charge carriers in graphene, it may be useful to compare it with the way ordinary electrons travel in an ordinary conductor. The "free" electrons that make up an electric current in, say, a metal are not really free; they do not act exactly like electrons moving in a vacuum. Electrons, of course, carry a negative charge, and so when they move through a metal they leave a charge deficit in the metal atoms from which they originate. Thus, when electrons move through the lattice, they interact with the electrostatic fields it creates, which push and pull them to and fro in a complex way. The end result is that the moving electrons act as if they had a different mass than ordinary electrons do—their so-called effective mass. Physicists call such charge carriers quasiparticles.

These charged, electronlike particles move much slower than the speed of light through the conducting metal. There is no need, therefore, to apply the corrections of Einstein's theory of relativity to their motions; that theory becomes important only at speeds approaching that of light. Instead the interactions of quasiparticles in a conductor can be described either by the familiar classical physics of Newton or by "ordinary" (that is, nonrelativistic) quantum mechanics.

As electrons travel through the chicken-wire web of carbon atoms in graphene, they, too, act as if they were a kind of quasiparticle. Astonishingly, however, the charge-carrying quasiparticle in graphene does not act much like an elec-

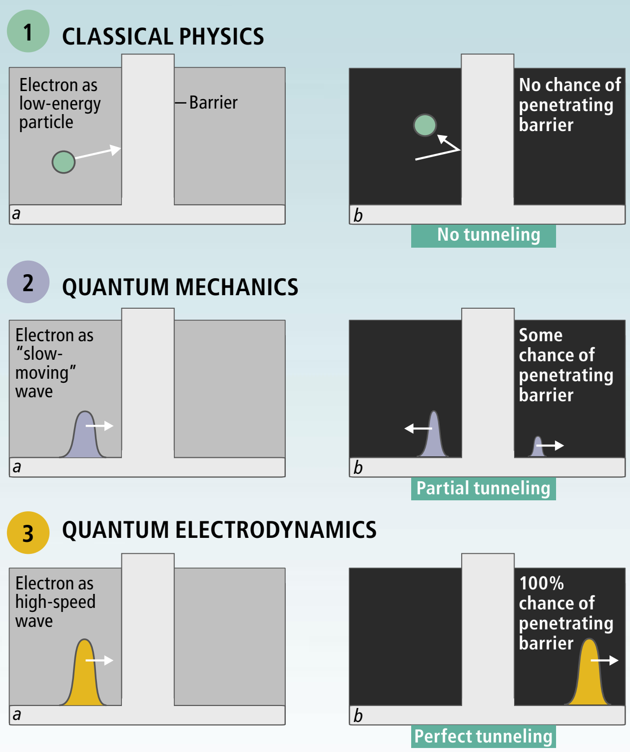
Quantum Electrodynamics Enters the Lab

Electrons move virtually unimpeded through the highly regular atomic structure of graphene, reaching such great speeds that their behavior cannot be described by "ordinary" quantum mechanics. The theory that applies instead is known as relativistic quantum mechanics, or quantum electrodynamics (QED), a theory whose distinctive (and weird) predictions were thought, until now, to be observable only in black holes or high-energy particle accelerators. With graphene, though, physicists can test one of the weirdest predictions of QED in the laboratory: "perfect quantum tunneling."

In classical, or Newtonian, physics, a low-energy electron (*green ball in 1a*) acts like an ordinary particle. If its energy is not enough to carry it over the top of a potential-energy barrier, it remains trapped on one side of the barrier (*1b*) as surely as a truck out of gas in a valley remains stranded on one side of a hill.

In the ordinary quantum-mechanical picture, an electron acts in some contexts like a wave that spreads out in space. The wave represents, roughly, the probability of finding the electron at a particular point in space and time. When this "slow-moving" wave approaches a potential-energy barrier (*blue wave in 2a*), it penetrates the barrier in such a way that there is some probability, neither 0 nor 100 percent, that the electron will be found on the far side of the barrier (*2b*). In effect, the electron tunnels through the barrier.

When a high-speed electron wave in graphene (*orange wave in 3a*) comes to a potential-energy barrier, QED makes an even more startling prediction: the electron wave will subsequently be found on the far side of an energy barrier with 100 percent probability (*3b*). The observation that graphene conducts electricity so well seems to confirm that prediction.



DANIELA NAOMI MOLNAR

tron at all. In fact, its closest analogue is another elementary particle, the nearly massless neutrino. Of course, the neutrino, in accord with its name, is electrically neutral (in Italian, *neutrino* means "little neutral one"), whereas the quasiparticle in graphene carries the same electric charge as the electron. But because the neutrino travels at nearly the speed of light, no matter what its energy or momentum, it must be described in terms of the theory of relativity. Similarly, a quasiparticle in graphene always moves at a high constant speed, albeit about 300 times slower than the speed of light. In spite of its scaled-down speed, its behavior closely parallels the relativistic behavior of the neutrino.

The relativistic nature of the quasiparticles in graphene renders ordinary, nonrelativistic quantum mechanics useless in describing how they act. Physicists must reach for a more complex framework in their arsenal of theories: relativistic quantum mechanics, which is now known as quantum electrodynamics. That theory has its own language, and central to that language is the probabilistic equation named after English physicist Paul A. M. Dirac, who first wrote his equation down in the 1920s. Accordingly, theorists sometimes describe electrons moving within graphene as massless Dirac quasiparticles.

Particles from "Nothing"

Unfortunately, interpreting quantum electrodynamics never comes without a good deal of wrestling with ordinary intuition. One must become familiar, if never quite comfortable, with phenomena that seem paradoxical. The paradoxes of quantum electrodynamics often arise from the fact that relativistic particles are always accompanied by their Bizarro-world alter egos: antiparticles. The electron, for instance, pairs with an antiparticle called the positron. Its mass is exactly the same as that of the electron, but its electric charge is positive. A particle-antiparticle pair can appear under relativistic conditions because it costs little energy for an extremely fast-moving, high-energy object to create a pair of "virtual particles." Oddly, the pair emerges directly from nothing—from the vacuum.

Why that happens is a consequence of one of the many versions of Heisenberg's uncertainty principle in quantum mechanics: roughly speaking, the more precisely an event is specified in time, the less precise is the amount of energy associated with that event. Consequently, on extremely short timescales, energy can take on almost any value. Because energy is equivalent to mass, according to Einstein's famous formula $E = mc^2$, the energy equivalent to the mass of a particle and its antiparticle can appear out of nothing. For example, a virtual electron and a virtual positron can suddenly pop into existence by "borrowing" energy from the vacuum, provided the lifetimes of the virtual particles are so short that the energy deficit is paid back before it can be detected.

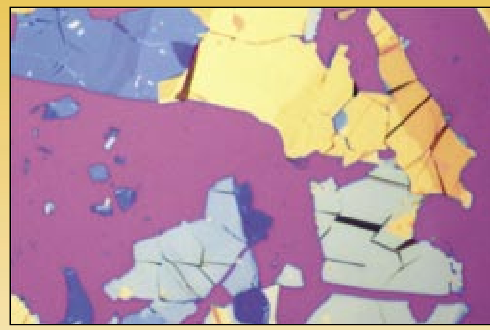
The intriguing dynamism of the vacuum in quantum electrodynamics leads to many peculiar effects. The Klein paradox is a good example. It describes circumstances in which a relativistic object can pass through any potential-energy barrier, no matter how high or how wide [*see box on opposite page*]. A familiar kind of potential-energy barrier is an ordinary rise in the landscape that surrounds a valley. A truck leaving the valley gains potential energy as it

D.I.Y. Graphene

1 Work in a clean environment; stray dirt or hair plays havoc with graphene samples.

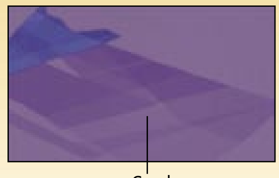
2 Prepare a wafer of oxidized silicon, which helps you see graphene layers under a microscope. To smooth out the surface to accept the graphene and to clean it thoroughly, apply a mix of hydrochloric acid and hydrogen peroxide.

3 Attach a graphite flake to about six inches of plastic sticky tape with tweezers. Fold the tape at a 45-degree angle right next to the flake, so that you sandwich it between the sticky sides. Press it down gingerly and peel the tape apart slowly enough so that you can watch the graphite cleaving smoothly in two.



4 Repeat the third step about 10 times. This procedure gets harder to do the more folds you make.

5 Carefully lay the cleaved graphite sample that remains stuck to the tape onto the silicon. Using plastic tongs, gently press out any air between the tape and sample. Pass the tongs lightly but firmly over the sample for 10 minutes. With the tongs, keep the wafer planted on the surface while slowly peeling off the tape. This step should take 30 to 60 seconds to minimize shredding of any graphene you have created.


Graphene

6 Place the wafer under a microscope fitted with a 50× or 100× objective lens. You should see plenty of graphite debris: large, shiny chunks of all kinds of shapes and colors (*upper image*) and, if you're lucky, graphene: highly transparent, crystalline shapes having little color compared with the rest of the wafer (*lower image*). The upper sample is magnified 115×; the lower 200×.

—JR Minkel, online news reporter

SCI AM

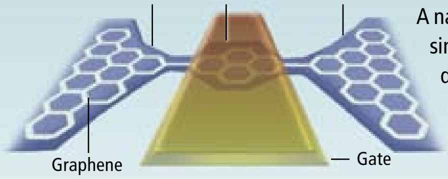
A slide show and description of how to make graphene with plastic sticky tape, the technique that JR Minkel outlines in the box above, are at **www.SciAm.com/ontheweb**

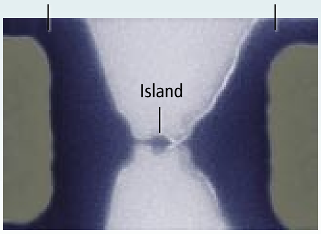
PETER BLAKE *Graphene Industries Ltd.*

[APPLICATIONS]

GRAPHENE-BASED TECHNOLOGY

Graphene has been available for too short a time for engineers to have developed any products that use it, but the list of prospective graphene-based technologies is long. Two examples include:





◄ SINGLE-ELECTRON TRANSISTORS

A nanoscale graphene plane can be formed into a single-electron (or quantum-dot) transistor. The diagram (*upper left*) shows schematically how two electrodes, a "source" and a "drain," are connected by an "island" of conducting material, or quantum dot, that is only 100 nanometers across. The island, which appears in the center of an electron micrograph of such a device (*lower left*)—shown here magnified 40,000×—is too small to accommodate more than one new electron at a time; any second electron is kept away by electrostatic repulsion. An electron from the source tunnels quantum mechanically to the island, then leaves by tunneling on to the drain. The voltage applied to a third electrode called the gate (not shown in the electron micrograph) controls whether a single electron can enter or exit the island, thereby registering either a 1 or a 0.

COMPOSITE MATERIALS ►

Two or more complementary materials can often be combined to obtain the desirable properties of both. Typically a bulk matrix and a reinforcement are used: think of a fiberglass boat hull made of plastic infused with strong glass fibers. Investigators are testing the physical properties of composites fabricated from polymers reinforced with graphene-based materials such as graphene oxide, a chemically modified version of graphene that is stiff and strong. Unlike graphene, graphene oxide "paper" (*right, inset*) is relatively easy to make and may soon find its own useful applications in laminated composites (*right, background*). The scale bar is one micron long.

In the long run, one can envision entire integrated circuits carved out of a single graphene sheet.

travels uphill, at the expense of the energy released by the fuel its engine burns. From the top of the hill, though, the truck can coast down the other side with the engine off and the transmission in neutral. The potential energy it gained by climbing is converted back into the energy of motion as the truck rolls downhill.

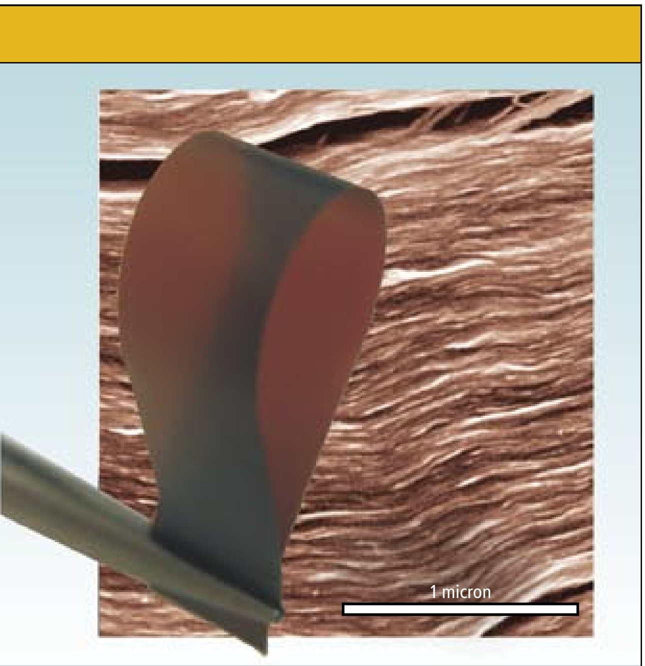
Testing the Bizarre

Particles, too, can readily move "downhill" on their own, from relatively high regions of potential energy to relatively low ones. If a "hillside" of high potential energy surrounds a particle in an energy "valley," however, the particle is no less stuck than a truck out of gas in a real valley. There is one big caveat to that conclusion, which occurs in ordinary, nonrelativistic quantum mechanics. A second version of Heisenberg's uncertainty principle states that it is impossible to know the exact position of a particle. Accordingly, physicists describe the position of a particle probabilistically. A strange consequence is that even though a low-energy particle might seem to be "trapped" by a high barrier, there is some probability that the particle will later be found outside that barrier. If it is, its ghostly passage through the energy barrier is called quantum tunneling.

In nonrelativistic quantum tunneling, the probability that a low-energy particle will tunnel through a high potential-energy barrier varies, but it can never be 100 percent. The probability of quantum tunneling shrinks as the barrier gets higher and thicker. The Klein paradox completely changes the character of quantum tunneling, however. It states that relativistic particles should tunnel through barrier regions of high energy and broad expanse with 100 percent probability. At a barrier the particles just pair up with their antiparticle twins, which experience the world in an upside-down, topsy-turvy fashion whereby real-world hills are seen as antiparticle valleys. After traveling readily through the odd, antiworld valley of the barrier, the antiparticles convert back into particles at the other side and emerge unimpeded. Even to many physicists, this prediction of quantum electrodynamics seems deeply counterintuitive.

Such an outlandish prediction cries out for testing, yet it has long remained unclear whether the Klein paradox could be tested at all, even in principle. The massless Dirac quasiparticles in graphene have now come to the rescue. In graphene, the Klein paradox becomes a routine effect with readily observable consequences. As charge-carrying, massless Dirac quasiparticles move within a graphene crystal across which a voltage, or potential-energy difference, has been applied, experimenters can measure the material's electrical conductivity. Perfect (100 percent probability) tunneling accounts for the lack of additional resistance that one would expect from the extra barriers and boundaries. Investigators are now measuring the flow of such tunneling particles through potential barriers of varying heights. Physicists expect that graphene will also

DANIELA NAOMI MOLNAR (*illustration*); LEONID PONOMARENKO *University of Manchester* (*micrograph*)



FROM "PREPARATION AND CHARACTERIZATION OF GRAPHENE OXIDE-BASED PAPER," BY DMITRIY A. DIKIN ET AL., IN *NATURE*, VOL. 448; 2007

help demonstrate many of the other oddball effects predicted by quantum electrodynamics.

2-D or Not 2-D

It is too early to fully assess the many possible technological applications of graphene. But more than a decade of research on carbon nanotubes—rolled-up graphene—gives graphene a huge head start. It is not unreasonable to think that nearly every useful role envisaged for nanotubes is also open to their flat cousin. High-tech industries are penciling in some commercial applications, and some are already placing bets on its promise. Meeting the demand for such applications will call for graphene output on an industrial scale, and many technology research teams are hard at work developing improved production techniques. Although graphene powder can already be made in industrial quantities, sheet graphene is still difficult to make and currently ranks as probably the most expensive material on the planet. Today a micromechanically cleaved graphene crystallite smaller than the thickness of a human hair can cost more than $1,000. Groups in Europe and at several U.S. institutions—the Georgia Institute of Technology, the University of California, Berkeley, and Northwestern University among them—have grown graphene films on silicon carbide wafers similar to the ones common in the semiconductor industry.

In the meantime, engineers worldwide are striving to exploit the highly desirable physical and electronic properties unique to graphene [*see box on opposite page and at left*]. Its high surface-to-volume ratio, for instance, should make it handy in manufacturing tough composite materials. The extreme thinness of graphene could also lead to more efficient field emitters—needlelike devices that release electrons in the presence of strong electric fields.

The properties of graphene can be finely tuned by applying electric fields, which could make it possible to build improved superconducting and so-called spin-valve transistors, as well as ultrasensitive chemical detectors. Finally, thin films fabricated from overlapping patches of graphene show great promise in serving as transparent and conducting coatings for liquid-crystal displays and solar cells. The list is far from exhaustive, but we expect that some niche applications could reach the market in only a few years.

Reprieve for Moore's Law?

One engineering direction deserves special mention: graphene-based electronics. We have emphasized that the charge carriers in graphene move at high speed and lose relatively little energy to scattering, or colliding, with atoms in its crystal lattice. That property should make it possible to build so-called ballistic transistors, ultrahigh-frequency devices that would respond much more quickly than existing transistors do.

Even more tantalizing is the possibility that graphene could help the microelectronics industry prolong the life of Moore's law. Gordon Moore, a pioneer of the electronics industry, pointed out some 40 years ago that the number of transistors that can be squeezed onto a given area doubles roughly every 18 months. The inevitable end of that continuing miniaturization has been prematurely announced many times. The remarkable stability and electrical conductivity of graphene even at nanometer scales could enable the manufacture of individual transistors substantially less than 10 nanometers across and perhaps even as small as a single benzene ring. In the long run, one can envision entire integrated circuits carved out of a single graphene sheet.

Whatever the future brings, the one-atom-thick wonderland will almost certainly remain in the limelight for decades to come. Engineers will continue to work to bring its innovative byproducts to market, and physicists will continue to test its exotic quantum properties. But what is truly astonishing is the realization that all this richness and complexity had for centuries lain hidden in nearly every ordinary pencil mark. ■

➡ MORE TO EXPLORE

Electrons in Atomically Thin Carbon Sheets Behave Like Massless Particles. Mark Wilson in *Physics Today,* Vol. 59, pages 21–23; January 2006.

Drawing Conclusions from Graphene. Antonio Castro Neto, Francisco Guinea and Nuno Miguel Peres in *Physics World,* Vol. 19, pages 33–37; November 2006.

Graphene: Exploring Carbon Flatland. A. K. Geim and A. H. MacDonald in *Physics Today,* Vol. 60, pages 35–41; August 2007.

The Rise of Graphene. A. K. Geim and K. S. Novoselov in *Nature Materials,* Vol. 6, pages 183–191; 2007.

Andre K. Geim's Mesoscopic Physics Group at the University of Manchester: **www.graphene.org**

Philip Kim's research group at Columbia University: **pico.phys.columbia.edu**

DETECTING NUCLEAR SMUGGLING

Radiation monitors at U.S. ports cannot reliably detect highly enriched uranium, which onshore terrorists could assemble into a nuclear bomb

By Thomas B. Cochran and Matthew G. McKinzie

KEY CONCEPTS

- Existing radiation portal monitors, as well as new advanced spectroscopic portal machines, cannot reliably detect weapons-grade uranium hidden inside shipping containers. They also set off far too many false alarms.
- So-called active detectors might perform better, but they are several years off and are very expensive.
- The U.S. should spend more resources rounding up nuclear smugglers, securing highly enriched uranium that is now scattered overseas, and blending down this material to low-enriched uranium, which cannot be fashioned into a bomb.

—The Editors

Customs inspectors at a pier in New York City send a sealed cargo container just taken off a ship from Istanbul through a radiation scanner. A dozen new tractors seem to be inside. Although the detector senses no radiation, the inspectors open the container anyway. Their handheld units show no radiation either, so they allow the container to leave. A private hauler drives it to a small Midwestern city. There terrorist cell members remove what was their final shipment of highly enriched uranium, concealed as 10 metal washers in the tractor engines, together weighing two kilograms. Months later an improvised nuclear device with a yield of one kiloton is detonated in Los Angeles. The blast, fire and airborne radioactivity kill more than 100,000 people. Virtually all shipping into the U.S. is halted, precipitating a financial crisis. Military operations commence in the Middle East after forensics and intelligence efforts trace the plot to cells in Pakistan and Iran.

Are these terrible events far-fetched? Twice in recent years the two of us helped an ABC News team that smuggled a soda can–size cylinder of depleted uranium through radiation detectors at U.S. ports. The material did not pose a danger to anyone, but it did emit a radiation signature comparable to that of highly enriched uranium (HEU), which can be assembled into a nuclear bomb. As you read this article, the Bush administration and the U.S. Congress are likely considering spending billions of dollars for additional detectors at ports and other border crossings—detectors that would also fail to reliably spot our cylinder or a similar amount of HEU.

A crude nuclear device constructed with HEU poses the greatest risk of mass destruction by terrorists. In the aftermath of the September 11 attacks, the U.S. government sought to prevent the smuggling of nuclear weapons and materials. The U.S. Department of Homeland Security instituted what it called a "layered defense," built largely around costly radiation detectors.

Why focus on detection? The sheer number of cargo containers entering the U.S. is staggering. Containers come in different sizes, so the number is counted as the equivalent of standard, 20-foot containers, or "twenty-foot equivalent units" (TEUs). More than 42 million TEUs entered American ports in 2005. By 2007 Homeland Security had deployed hundreds of radiation portal monitors. It also asked Congress for additional, advanced machines but in October

JULIE JACOBSON *AP Photo;* PHOTOILLUSTRATION BY SCIENTIFIC AMERICAN

backed off to perform further testing on those units, in light of software problems. Although some federal officials and government contractors claim that the technology will be effective, an analysis we have conducted shows that the machines will not reliably reveal HEU. Instead the government must place a much higher priority on efforts to identify and eliminate or secure known stocks of HEU, stopping the potential problem at its source.

Easy to Hide

To wreak havoc, terrorists could steal, purchase or be given a fully assembled nuclear weapon, but that scenario is not likely. Intact nuclear weapons are generally under greater physical security than the fissile material needed to build one. A more probable route is to illicitly obtain this material—which is now scattered among many civil, military and space power facilities worldwide—and then to smuggle it into the U.S. and assemble a bomb. Two fissile materials are of primary concern: plutonium and HEU.

Less plutonium than HEU is needed to achieve a given explosive yield, but crafting a plutonium weapon requires far more complex engineering. Plutonium is also easier to detect if shipped among cargo. HEU is easier to handle, to form into a crude explosive device and is much harder to detect in a cargo container. Furthermore, a greater amount of HEU exists in more dispersed and less secured places. According to the International Atomic Energy Agency, 275 confirmed incidents involving nuclear material and criminal intent occurred globally between January 1993 and December 2006. Four involved plutonium, but 14 involved HEU. More than 40 countries harbor HEU, with the highest risk of theft being from facilities in Russia, other former Soviet states and Pakistan. And a recent Harvard University study concluded that U.S.-funded security work had not been completed at 45 percent of nuclear sites of concern in countries once part of the Soviet Union.

Obtaining nuclear material is no small challenge, but the ABC News exercises showed that smuggling it into the U.S. can be a straightforward matter. In the summer of 2002 an ABC News unit successfully slipped a lead-lined steel pipe containing a 6.8-kilogram (15-pound) cylinder of depleted uranium (DU) past U.S. Customs and Border Protection by placing it inside a standard cargo container. This material is unsuitable for a weapon, but its chemical properties are

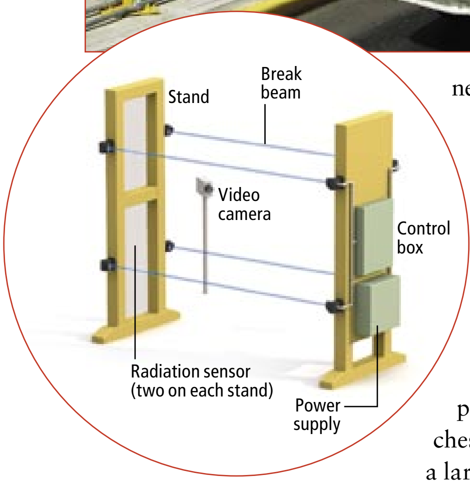
URANIUM IN A HAYSTACK

20
Length, in feet, of a typical shipping container. The volume of international shipping is measured in "twenty-foot equivalent units," or TEUs.

297 million
Number of TEUs shipped worldwide in 2005.

42 million
TEUs entering U.S. ports that same year.

6,500
TEUs arriving at the Port of New York and New Jersey on a light day; up to 13,000 on a busy day.



CURRENT PORTAL MONITORS wait to detect radiation as containers slowly pass by. Inset: Sensors on both sides can pick up radiation; a video camera identifies the vehicle.

FRANK FRANKLIN II *AP Photo* (*top*); GEORGE RETSECK (*inset*)

nearly identical to those of HEU. Our organization, the Natural Resources Defense Council (NRDC), prepared the shielded cylinder. The ABC News crew placed the pipe in an ordinary suitcase and carried it on passenger trains from Vienna to Istanbul—a route chosen to simulate a terrorist journey. The news crew saw no radiation detection equipment along the way.

On reaching Istanbul, the journalists placed the suitcase inside an ornamental chest, packed alongside crates of huge vases in a large metal shipping container that left Istanbul by ship on July 10. When the container arrived at Staten Island in New York, Customs officials, part of Homeland Security, targeted it as high risk, in part because of its origin, and flagged it for more thorough screening. The machine and its operators failed to sense the uranium. ABC News aired its story on September 11, 2002, a year after the 9/11 attacks.

ABC News repeated the experiment a year later. This time the suitcase was placed inside a teak trunk that was loaded with other furniture within a container in Jakarta. The container arrived at the Port of Long Beach, Calif., on August 23. As before, this shipment was targeted and screened by Customs personnel but was allowed to proceed by truck on September 2. ABC News announced that officials once again failed to detect depleted uranium. After the story aired, Customs seized our cylinder, "disposed" of it and placed one of us (Cochran) on an air-travel watch list for several months.

Inadequate Detectors

The Department of Homeland Security began installing first-generation detector systems—known as radiation portal monitors (RPMs)—in late 2002. Despite the ABC News results, more than 800 of these machines and their successors have been placed at manned ports of entry, land border crossings, airports, seaports, and international mail and courier facilities. These machines are so-called scintillation detectors that count neutrons and gamma rays but do not measure their total energy. The instruments can therefore gauge the intensity of detected radiation but cannot measure the characteristic radiation spectrum, or signature, of a source.

To reveal radioactive material, the radiation must first be detected, but then the counts must also be discernible from those produced by harmless radioactive substances in the cargo—everything from bananas, brazil nuts and white potatoes to cat litter, aircraft parts, glass and concrete. This natural background radiation can vary significantly depending on a container's contents. The inability of RPMs to measure a source's characteristic radiation spectrum, however, leads inevitably to false alarms from background sources, and the false-alarm rate for the current monitors is problematically high—as high as several hundred a day at certain facilities. When an alarm is triggered, Customs agents must run containers through further scans or inspect them by hand, adding considerable shipping delays and cost.

The further testing is done by a gamma-ray imaging system called VACIS, which produces an x-ray scan of the container contents. Customs officers are also equipped with small, pagerlike personal radiation detectors. Although all these sensors were present during the second ABC experiment, none of them detected the concealed uranium.

Given the RPMs' deficiencies, Homeland Security announced in 2006 that it would acquire hundreds of second-generation radiation detectors—advanced spectroscopic portal (ASP) machines—with the price tag for total hardware alone exceeding $1 billion. The system has both gamma-ray and neutron detectors but can also perform gamma-ray spectroscopy to show a radiation signature; when applied together, the techniques are intended to lower false alarms by identifying a source as harmless radioactive cargo. In August 2007 President George W. Bush signed into law the Implementing Recommendations of the 9/11 Commission Act of 2007, which mandates that within five years *all* maritime cargo be scanned *before* it is loaded into vessels in foreign ports heading to the U.S. Many more detectors would have to be deployed.

This move is ill-advised, however, because even the ASP machines are not dependable. Their ability to reliably sense shielded HEU was not demonstrated during classified trials at the U.S. Department of Defense's Nevada test site. Moreover, the ASP machines failed to function properly, when installed at the Port Authority of New York and New Jersey, as a result of software problems. Indeed, in November 2007 the *Washington Post* revealed that Homeland Security itself had questioned the machines' effectiveness. According to the newspaper, in September 2006 auditors at the U.S. Government Accountability Office alleged that officials had greatly exaggerated the tools' capabilities. Another investigation by the accountability office a year later found that officials had overseen compromised tests of the ASP system. After petitioning Congress to allocate funds for more ASPs, in October 2007 Homeland Security chief Michael Chertoff decided to postpone certifying the new ASPs and further production until problems were resolved. When this issue of *Scientific American* reaches newsstands, Homeland Security will likely be presenting Congress with new contractor performance data on the ASP detectors. Legislators will have to decide whether to continue to fund such acquisitions in light of the machines' checkered history.

A Fair Surrogate

Homeland Security was not fond of the ABC exercises and asserted publicly that if NRDC's slug of depleted uranium had been HEU, inspectors would have identified and intercepted it. We disagree. Our analyses show that when even lightly shielded with lead and steel, depleted and highly enriched uranium have similarly weak radiation signals and would be equally hard to detect by either generation of monitor.

To compare DU with HEU, we calculated the radiation particle count and detector dose rate at various distances from shielded and unshielded samples of both and then compared these quantities with background radiation. We performed our calculations using the standard radiation analysis software employed by Los Alamos National Laboratory and common uranium radiation data supplied by Lawrence Livermore National Laboratory.

Some HEU contains extremely small concentrations of the isotope uranium 232, which is not found in natural uranium but is produced when HEU is irradiated in nuclear reactors. In the U.S. and Russia most HEU was enriched from uranium that was recovered from military fuel and thus is contaminated with trace amounts of uranium 232. Even minuscule quantities, less than one part per billion (ppb), have a telltale radiation signal.

Two significant factors can reduce an RPM's ability to detect the signal of either pure or contaminated HEU: shielding that absorbs radiation, and distance between the source and detector. When not shielded, the calculated dose rate for pure and contaminated HEU is greater than for DU. But much of the dose in both cases arises from the lower-energy part of the gamma-ray spectrum, which is readily absorbed by shielding. If the source is wrapped in a one-millimeter-thick layer of lead, the radiation dose rate from HEU with no uranium 232 is *less* than

RADIATION COMPARED

The slug of depleted uranium (DU) that we prepared and that ABC News smuggled past portal monitors is shown below. It is a good substitute for testing whether the monitors could detect highly enriched uranium (HEU) because the radiation signal from uncontaminated DU, when shielded inside a cargo container, is actually greater than that from uncontaminated HEU.

CALCULATED DOSE RATE
(microrad/hour)
two centimeters away, when wrapped in three millimeters of lead

DU: 1,500
HEU: 100

CALCULATED DOSE RATE
(microrad/hour)
two meters away, for same samples

DU: 1.0
HEU: 0.1

NATURAL BACKGROUND: 2–10

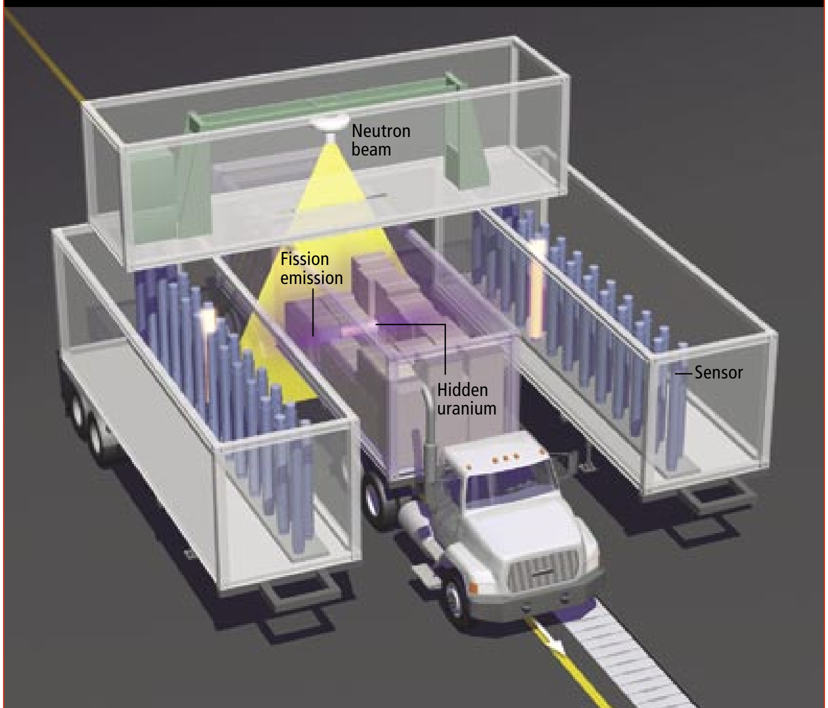
DEPLETED URANIUM cylinder, shielded in lead, was slipped past portal detectors.

COURTESY OF MATTHEW G. MCKINZIE

[NEW TECHNOLOGY]

ACTIVE DETECTION

Active detectors being developed by several laboratories could distinguish highly enriched uranium in cargo containers from other naturally radioactive materials. At issue is whether the incoming detector beam could harm people who illegally stow away inside containers and whether the public would balk if containers held food, even though imparted radiation would decay in less than a minute.



In a concept from Lawrence Livermore National Laboratory, a detector sends a low-power neutron or gamma-ray beam (*yellow*) into a passing container; if highly enriched uranium is present (*purple*), it will undergo a brief fission and emit neutrons and high-energy gamma rays, which have a short half-life decay. Sensors (*blue rods*) would distinguish each of these signals.

the dose rate of similarly shielded DU. For HEU contaminated with 2 ppb of uranium 232 and encased in the same shielding, the gamma-ray dose rate is about equal to that from DU. Public data indicate that roughly half of Russian HEU may have a uranium 232 concentration less than 0.2 ppb, and all HEU produced in Pakistan and Iran is likely free of the isotope.

For both DU and HEU, the emitted radiation decreases with distance from the source. We calculated that for lightly shielded HEU (covered by one-millimeter-thick lead) with no uranium 232, the dose rate is less than 5 percent of the background radiation at two meters; for HEU contaminated with 0.2 ppb the dose rate is less than 10 percent of the background radiation. In a typical portal monitor the distance from a standard shipping container's center to a detector on either side is greater than two meters. Thus, neither RPMs nor the newer ASPs would distinguish most shielded HEU from Russia if any were placed near the center of a container.

Our work demonstrates that DU is a good surrogate for HEU in testing Homeland Security's ability to detect smuggled nuclear weapons material. Had our test slug been HEU, the RPMs would not have detected it, and neither would have the next-generation ASPs.

Enough for a Bomb

To make matters worse, it is conceivable that terrorists could smuggle HEU into the U.S. by shipping pieces smaller than the slug we used. The question, then, is whether small pieces of HEU could be assembled into a crude bomb with substantial explosive yield.

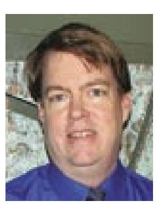
Most modern nuclear weapons are based on implosion: conventional explosives are detonated to compress nuclear material so well that it erupts in a runaway fission chain reaction. Terrorists operating covertly are unlikely to be able to assemble such a device because of its complexity. A simpler gunlike design can be effective, however; two subcritical pieces of HEU are driven together so that they form a supercritical mass. The "Little Boy" atom bomb that was dropped on Hiroshima in 1945 brought together about 65 kilograms of HEU within a millisecond by firing one subcritical piece down a gun barrel at a second subcritical piece.

The "quality" of nuclear material since then has continued to improve, however, so much so that in 1987 Nobel laureate physicist and Manhattan Project scientist Luis Alvarez noted that if terrorists had modern weapons-grade uranium, they "would have a good chance of setting off a high-yield explosion simply by dropping one half of the material on the other half." To test that assertion, we modeled the difference between the Little Boy design and an improvised nuclear device as crude as the one Alvarez described.

We again used the Los Alamos software code and modeled the yield of Little Boy on publicly available design information, as well as two simple configurations of HEU in a gun assembly. Our modeling showed that, for an explosive-driven gun assembly, the minimum quantity that was required to obtain a one-kiloton explosive yield would be substantially less than the amount of HEU in Little Boy. Most disturbingly, with larger quantities, a one-kiloton yield could be achieved with a probability greater than 50 percent by dropping a single piece of HEU onto another, confirming Alvarez's statement. Designing an HEU bomb seems shock-

[THE AUTHORS]

Nuclear physicists Thomas B. Cochran (*left*) and Matthew G. McKinzie (*right*) work together at the Natural Resources Defense Council in Washington, D.C. Cochran holds the Wade Greene Chair for Nuclear Policy; McKinzie is a senior scientist in the Nuclear Program.



GEORGE RETSECK (*illustration*); COURTESY OF THOMAS B. COCHRAN (*Cochran*); COURTESY OF MATTHEW G. McKINZIE (*McKinzie*)

EDITORS' NOTE

The authors and editors have been careful to not expose details that could help terrorists or that are not readily available in published sources.

ingly simple. The only real impediment, therefore, is secretly gathering sufficient material.

More Effective Countermeasures

Given the difficulty portal monitors have in detecting smuggled HEU, stopping any clandestine influx is crucial. The U.S. government has essentially four choices: rely on intelligence to identify and round up nuclear smugglers; eliminate HEU at its source; consolidate HEU and safeguard it; or detect HEU as it crosses international borders. Wise policy would pursue all these avenues in a balanced way, based on their effectiveness and cost. The current U.S. approach, however, is far too reliant on the dubious detectors. The federal government does have programs to "blend down" excess Russian military stocks of highly enriched uranium into low-enriched uranium and to replace HEU fuel in research reactors with low-enriched fuel. The U.S. is also helping to improve the physical security of some Russian HEU stocks. But the government has not given these programs, which do not fall under Homeland Security, the same high priority as portal monitor programs.

As trials have shown, neither RPMs nor ASPs are currently able to provide reliable protection. Homeland Security does support research on other advanced detection schemes, such as active detection systems that bombard a container or vehicle with low-energy neutrons, creating a telltale gamma-radiation signal. Last year Lawrence Livermore unveiled a prototype that it claims can detect lightly shielded uranium pieces of less than one kilogram, with a low false-alarm rate [*see box on page 102*]. Rapiscan Systems in Torrance, Calif., is developing a similar scheme. The technology would have to be commercialized and its costs significantly reduced, however. Concern by the public, and by shippers, that cargo contents would be exposed to the gamma rays produced would also have to be addressed, even though the energy levels are very low and the radiation dissipates within an hour. Lawrence Livermore has said it could have a system ready for commercial evaluation by 2009.

Congress is likely now debating whether it should continue to support the troubled, multibillion-dollar ASP program. With RPMs and ASPs, Customs agents might catch an amateur terrorist attempting to smuggle HEU, but they are unlikely to catch a sophisticated agent like 9/11 terrorist Mohamed Atta. The U.S. government should instead place a much higher policy priority, and in some cases spend more funds, on securing and eliminating HEU sources worldwide. The government should also seek a global ban on the commercial use of HEU, such as isotope production for medical applications and experiments on nuclear reactor designs, most of which can be performed with low-enriched uranium or in particle accelerators. To protect the U.S. from terrorist nuclear attack, the country should forge a larger, more effective strategic plan centered on eliminating access to weapons-grade material. ■

ABOUT 40 KILOGRAMS (88 pounds) of highly enriched uranium identified in Swierk, Poland, is returned to Russia in August 2006 for decommissioning, a move prompted by the International Atomic Energy Agency.

INTERNATIONAL ATOMIC ENERGY AGENCY *Ho/AP Photo*

BETTER PLAN

Radiation monitors at U.S. ports are insufficient for preventing terrorists from amassing highly enriched uranium for a nuclear bomb. We maintain that the country should therefore enhance protective actions, some now under way. Chief among them:

- Identify and round up nuclear smugglers.
- Secure known sources of poorly guarded HEU and ship it back to its country of origin for elimination.
- Blend down excess Russian stocks of HEU into low-enriched uranium.
- Replace HEU fuel in research reactors with low-enriched fuel.
- Seek a global ban on the commercial uses of HEU, such as medical applications, most of which can be pursued with low-enriched uranium.

—T.B.C. and M.G.McK.

MORE TO EXPLORE

Thwarting Nuclear Terrorism. Alexander Glaser and Frank N. von Hippel in *Scientific American,* Vol. 294, No. 2, pages 56–63; February 2006.

Securing the Bomb 2007. Matthew Bunn. Harvard University and Nuclear Threat Initiative, September 2007.

Looking for Hidden Materials. Dennis Slaughter in *Nuclear News,* page 43; November 2007.

Radiation Detectors for Border Are Delayed Again. Robert O'Harrow, Jr., in *Washington Post,* page A1; November 20, 2007.

The Illicit Trafficking Database, maintained by the International Atomic Energy Agency, lists unauthorized activities involving radioactive materials: **www.iaea.org/NewsCenter/News/2007/itdb.html**

At the Edge of Life's Code

Using machine learning, Chris Wiggins hopes to develop models that can predict how all of an organism's genes behave under any circumstance—and thereby explain precisely why some cells become sick or cancerous BY THANIA BENIOS

On an airport shuttle bus to the Kavli Institute for Theoretical Physics in Santa Barbara, Calif., Chris Wiggins took a colleague's advice and opened a Microsoft Excel spreadsheet. It had nothing to do with the talk on biopolymer physics he was invited to give. Rather the columns and rows of numbers that stared back at him referred to the genetic activity of budding yeast. Specifically, the numbers represented the amount of messenger RNA (mRNA) expressed by all 6,200 genes of the yeast over the course of its reproductive cycle. "It was the first time I ever saw anything like this," Wiggins recalls of that spring day in 2002. "How do you begin to make sense of all these data?"

Instead of shirking from this question, the 36-year-old applied mathematician and physicist at Columbia University embraced it—and now six years later he thinks he has an answer. By foraying into fields outside his own, Wiggins has drudged up tools from a branch of artificial intelligence called machine learning to model the collective protein-making activity of genes from real-world biological data. Engineers originally designed these tools in the late 1950s to predict output from input. Wiggins and his colleagues have now brought machine learning to the natural sciences and tweaked it so that it can also tell a story—one not only about input and output but also about what happens inside a model of gene regulation, the black box in between.

The impetus for this work began in the late 1990s, when high-throughput techniques generated more mRNA expression profiles and DNA sequences than ever before, "opening up a completely different way of thinking about biological phenomena," Wiggins says. Key among these techniques were DNA microarrays, chips that provide a panoramic view of the activity of genes and their expression levels in any cell type, simultaneously and under myriad conditions. As noisy and incomplete as the data were, biologists could now query which genes turn on or off in different cells and determine the collection of proteins that give rise to a cell's characteristic features—healthy or diseased.

CHRIS WIGGINS

NUMBER CRUNCHING: Uses a statistical approach to predict whether an organism's genes increase or decrease their protein-making activity (that is, the behavior of its gene regulatory network) and reveal the underlying logic.

KEY COMPONENT: A machine-learning algorithm, which combs data from gene chips and identifies important connections among DNA sequences and the proteins that regulate them.

BRIAN MARANAN PINEDA

Yet predicting such gene activity requires uncovering the fundamental rules that govern it. "Over time, these rules have been locked in by cells," says theoretical physicist Harmen Bussemaker, now an associate professor of biology at Columbia. "Evolution has kept the good stuff."

To find these rules, scientists needed statistics to infer the interaction between genes and the proteins that regulate them and to then mathematically describe this network's underlying structure—the dynamic pattern of gene and protein activity over time. But physicists who did not work with particles (or planets, for that matter) viewed statistics as nothing short of an anathema. "If your experiment requires statistics," British physicist Ernest Rutherford once said, "you ought to have done a better experiment."

But in working with micro-

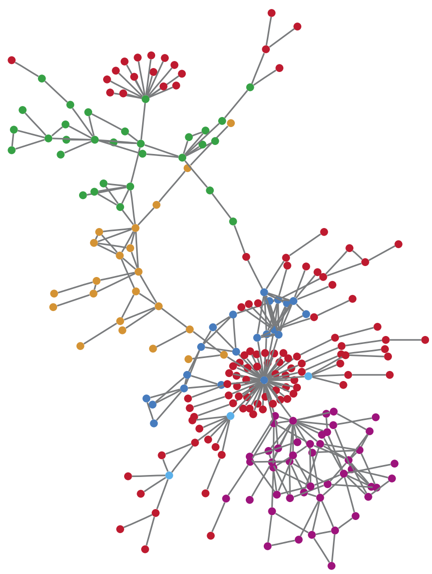
Page Intentionally Blank

arrays, "the experiment has been done without you," Wiggins explains. "And biology doesn't hand you a model to make sense of the data." Even more challenging, the building blocks that make up DNA, RNA and proteins are assembled in myriad ways; moreover, subtly different rules of interaction govern their activity, making it difficult, if not impossible, to reduce their patterns of interaction to fundamental laws. Some genes and proteins are not even known. "You are trying to find something compelling about the natural world in a context where you don't know very much," says William Bialek, a biophysicist at Princeton University. "You're forced to be agnostic."

Wiggins believes that many machine-learning algorithms perform well under precisely these conditions. When working with so many unknown variables, "machine learning lets the data decide what's worth looking at," he says.

At the Kavli Institute, Wiggins began building a model of a gene regulatory network in yeast—the set of rules by which genes and regulators collectively orchestrate how vigorously DNA is transcribed into mRNA. As he worked with different algorithms, he started to attend discussions on gene regulation led by Christina Leslie, who ran the computational biology group at Columbia at the time. Leslie suggested using a specific machine-learning tool called a classifier. Say the algorithm must discriminate between pictures that have bicycles in them and pictures that do not. A classifier sifts through labeled examples and measures everything it can about them, gradually learning the decision rules that govern the grouping. From these rules, the algorithm generates a model that can determine whether or not new pictures have bikes in them. In gene regulatory networks, the learning task becomes the problem of predicting whether genes increase or decrease their protein-making activity.

The algorithm that Wiggins and Leslie began building in the fall of 2002 was trained on the DNA sequences and mRNA levels of regulators expressed during a range of conditions in yeast—when the yeast was cold, hot, starved, and so on. Specifically, this algorithm—MEDUSA (for *m*otif *e*lement *d*iscrimination *u*sing *s*equence *a*gglomeration)—scans every possible pairing between a set of DNA promoter sequences, called motifs, and regulators. Then, much like a child might match a list of words with their definitions by drawing a line between the two, MEDUSA finds the pairing that best improves the fit between the model and the data it tries to emulate. (Wiggins refers to these pairings as edges.) Each time MEDUSA finds a pairing, it updates the



INTERACTIONS among 230 proteins involved in various cellular activities of *E. coli* (amounting to about 5 percent of the bacterium's genome) are shown. By determining which interactions are related (grouped by proteins of like color), a machine-learning algorithm may reveal the basic design principles behind gene regulatory networks.

JAKE HOFMANN *Columbia University*

model by adding a new rule to guide its search for the next pairing. It then determines the strength of each pairing by how well the rule improves the existing model. The hierarchy of numbers enables Wiggins and his colleagues to determine which pairings are more important than others and how they can collectively influence the activity of each of the yeast's 6,200 genes.

By adding one pairing at a time, MEDUSA can predict which genes ratchet up their RNA production or clamp that production down, as well as reveal the collective mechanisms that orchestrate an organism's transcriptional logic.

Wiggins and his colleagues can now go much further than yeast. Recently they have shown that MEDUSA can accurately build predictive models of gene regulatory networks in higher organisms such as worms as well as in several cell lines, including those of human lymphocytes. In a cancer cell line, the team can determine which genes increase their activity when they should decrease it, and vice versa. The ultimate goal, however, is to understand their coordinated activity and infer, with statistics, which interactions lead to a diseased cell.

Although MEDUSA makes accurate predictions on test data, there is still no way to know whether it faithfully reproduces real biological networks. To do so, each connection would have to be experimentally tested. It is also unclear how well microarray data measure expression levels, so accurate predictions may not necessarily reflect the truth. Moreover, machine learning forces researchers to formulate ad hoc hypotheses that may be biased toward their results, "so any kind of correlation in the data may be a fluke," remarks Yoav Freund of the University of San Diego, who created MEDUSA's learning algorithm.

To address these limitations, researchers must not only continue to cross disciplines but also be willing to adopt their tools. "I would say that machine learning hasn't taken off like wildfire in the physics community," remarks Alex Hartemink, a machine-learning expert at Duke University. "But Chris seems to be most comfortable reaching out and learning about techniques from other places. And I think we need people that are going to do that—foray out into the forest, find new resources and bring them back to the tribe and say, 'Hey, guys, check this out—this is great stuff.' " ■

Thania Benios is based in New York City.

Weapons Revealed

By Stuart F. Brown

Airline passengers moving through security who meet certain criteria—such as having purchased a last-minute ticket—can expect to be taken aside for a pat-down body search, which some people find invasive.

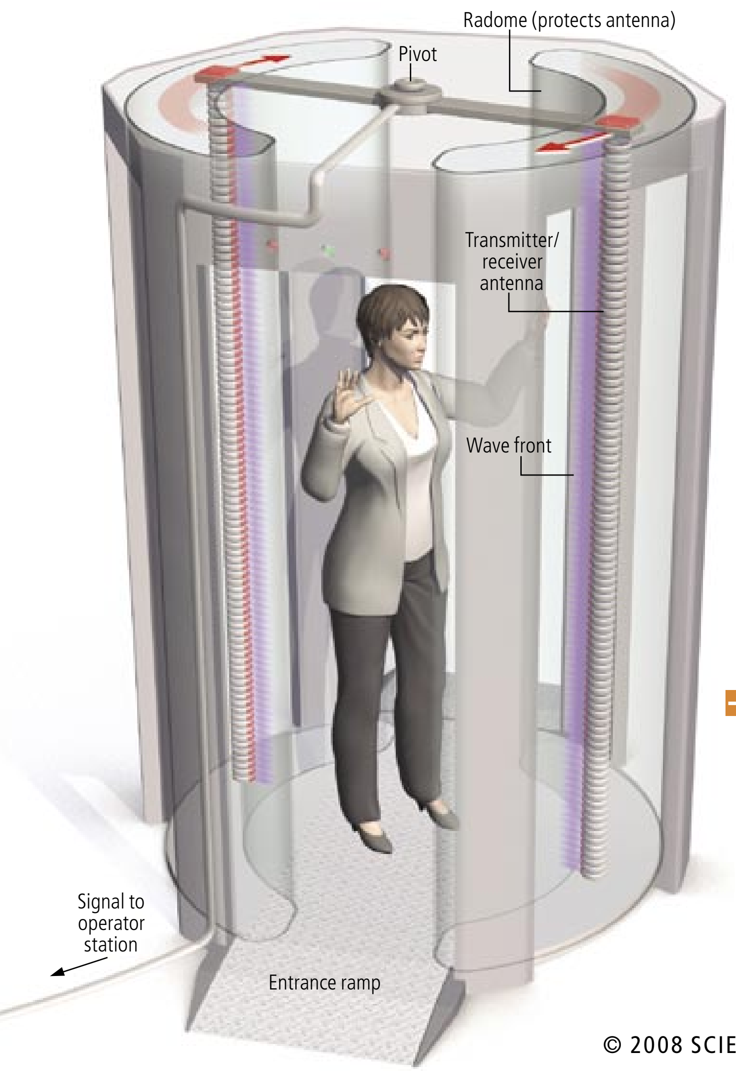
Transportation Security Administration (TSA) officers at Sky Harbor airport in Phoenix now offer such passengers a body scan in lieu of a pat-down. So far, the agency reports, about 80 percent of passengers selected for secondary screening have chosen to step up to the machine. The TSA is now testing two competing body-scanning technologies, and the trial program will soon be extended to LAX in Los Angeles and JFK in New York.

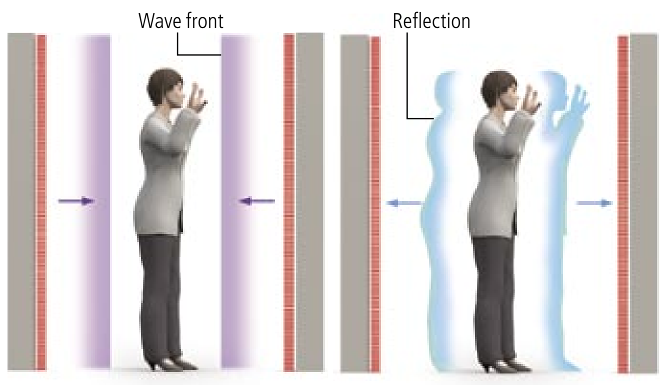
One method is backscatter x-ray imaging. It senses low-intensity x-rays as they reflect back from the passenger's body and any objects the person may be concealing. (The technique differs from conventional transmission x-ray imaging, which uses higher-energy beams.) Different materials reflect rays back to a detector in proportion to their density. Joe Reiss, vice president of marketing at manufacturer American Science and Engineering (AS&E) in Billerica, Mass., says low-atomic-number elements such as carbon, oxygen, hydrogen and nitrogen—common constituents of explosives—create a strong scattering effect visible in images that operators monitor on a screen yet discernible from the organic molecules in the human body.

The second method TSA is testing is millimeter-wave imaging. The system, built by L-3 Communications in Woburn, Mass., emits beams of radio-frequency energy that are tuned to reflect well off human skin. Reflected radio energy is then used to construct a 360-degree model of the passenger and whatever he or she may be carrying.

TSA is not yet commenting on the relative performance of the two systems, because trials are still under way. Critics are voicing concerns about privacy, however. Barry Steinhardt, associate director of the American Civil Liberties Union, calls backscatter x-ray imaging "nothing more than an electronic strip search." AS&E has added software that reduces the person's body details to outlines, while highlighting objects of interest, such as a ceramic knife tucked into a sock. L-3 offers a similar feature. The workers who examine the images sit where they cannot see the person being scanned, and the images are deleted after being examined, TSA says.

Stuart F. Brown is based in Irvington, N.Y.





MILLIMETER-WAVE IMAGING
A passenger steps inside. Two vertical banks of transmitter/receivers pivot in tandem, each emitting a wave front that penetrates clothing and reflects off the person's body and any concealed objects. For privacy, the security operator viewing the resulting image sits at a remote location.

Scan time = 10 seconds
Beam frequency = 24–30 GHz
Beam power density = 6×10^{-6} mW/cm^2

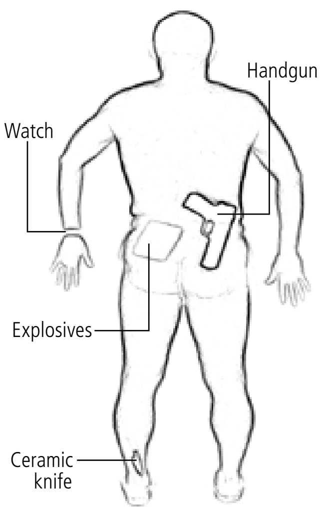
GEORGE RETSECK (*illustrations*); AMERICAN SCIENCE AND ENGINEERING, INC. © 2008 (*backscatter image*); L-3 SECURITY & DETECTION SYSTEMS (*computer and screen shot*)

BACKSCATTER IMAGING

A passenger stands in front of the machine. A rotating collimator projects an x-ray beam through a slit toward the person. The beam backscatters off the subject's body and hidden objects into detectors. The x-ray unit lowers from the ceiling to complete the scan. The passenger turns around, and the unit rises to scan the other side.

Scan time = 30 seconds
Beam frequency = 1,000–4,000 GHz
Beam energy = 1.45 keV

Image seen by operator



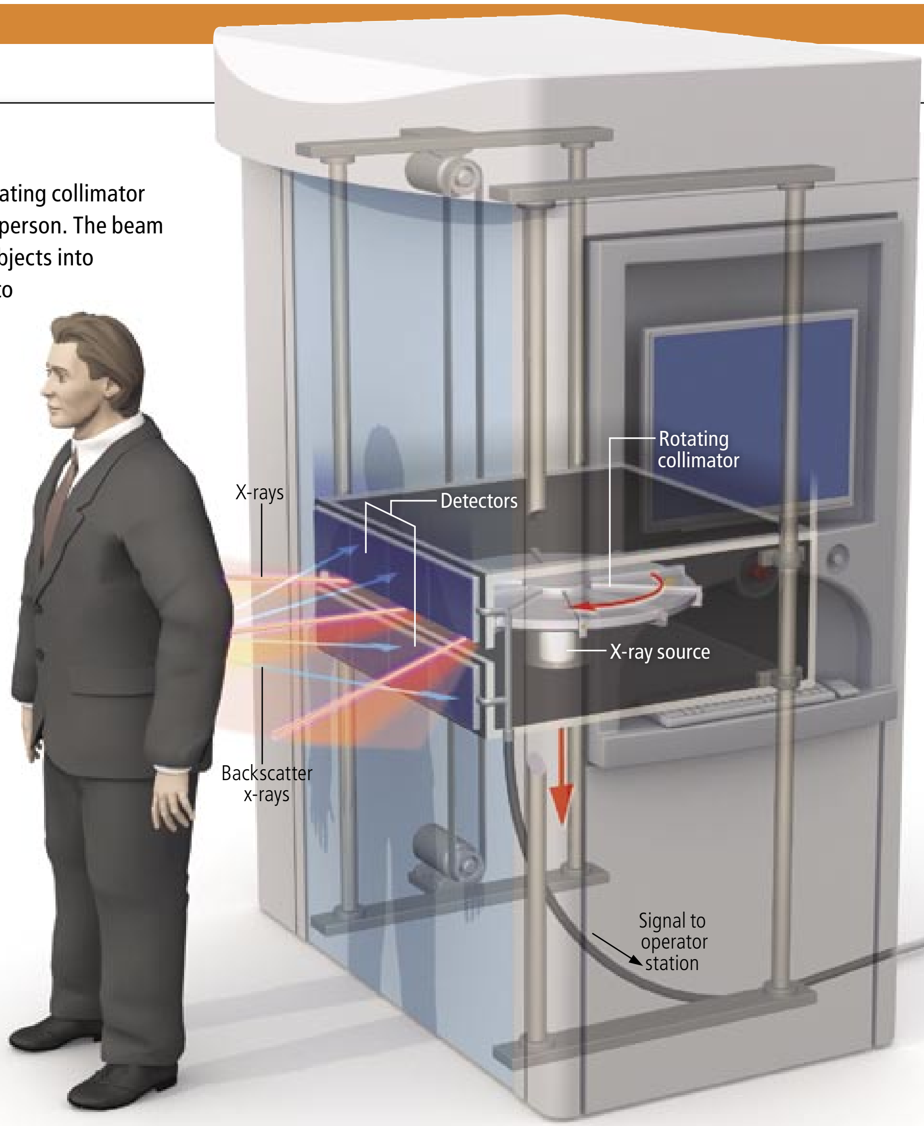
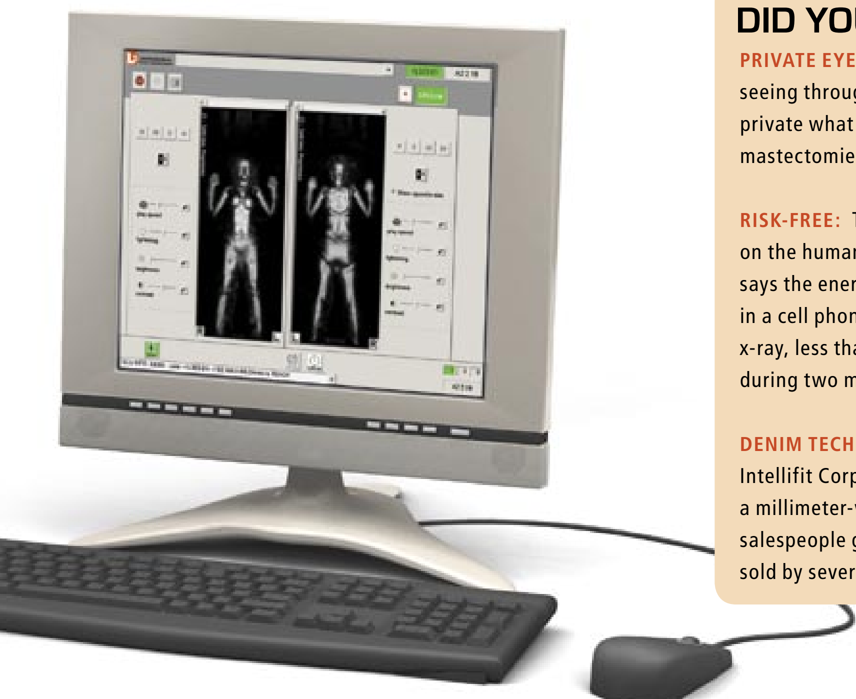


Image seen by operator



DID YOU KNOW ...

PRIVATE EYES: In addition to understandable modesty about a stranger seeing through one's clothes, travelers may have other reasons to keep private what is normally unseen: a body scanner can illuminate evidence of mastectomies, colostomy appliances, penile implants and catheter tubes.

RISK-FREE: The makers of scanners claim that the amount of energy imposed on the human body poses no health risk to travelers. L-3 Communications says the energy projected by its equipment is one ten-thousandth the energy in a cell phone transmission. AS&E says the radiation dose from a backscatter x-ray, less than 10 microrems, is the same received from natural sources during two minutes of an airplane flight at 30,000 feet.

DENIM TECH: Imaging technology has also been used for high fashion. Intellifit Corporation has a "virtual fitting room" in West Chester, Pa., where a millimeter-wave machine scans customers to determine sizing, and salespeople give advice about the perfect fit of clothing, such as blue jeans, sold by several national brands.

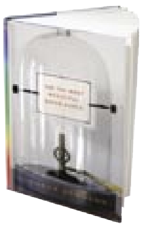
SEND TOPIC IDEAS *to workingknowledge@SciAm.com*

Beautiful Science ▪ Project Nim ▪ Looking for E.T.

BY MICHELLE PRESS

➡ THE TEN MOST BEAUTIFUL EXPERIMENTS

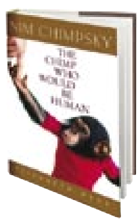
by George Johnson. Knopf, 2008 ($22)



As a science writer, Johnson confesses, he has often been attracted to rarefied concepts such as general relativity or quantum mechanics. Fascinating stuff, but he began to feel the need for something more basic. "What I was looking for were those rare moments when, using the materials at hand, a curious soul figured out a way to pose a question to the universe and persisted until it replied." In his search for these classic experiments, he even tries to duplicate a few himself, providing some entertaining asides. Looking for equipment to perform Robert Millikan's demonstration of the existence of electrons, for example, he visits a junkyard called the Black Hole ("everything goes in and nothing comes out"), run by an ex-bomb-maker in Los Alamos. Johnson's mix of the personal, the erudite and crystalline prose is—like the pull of gravity (see beautiful experiment number 1)—an irresistible force.

➡ NIM CHIMPSKY: THE CHIMP WHO WOULD BE HUMAN

by Elizabeth Hess. Bantam Dell, 2008 ($23)

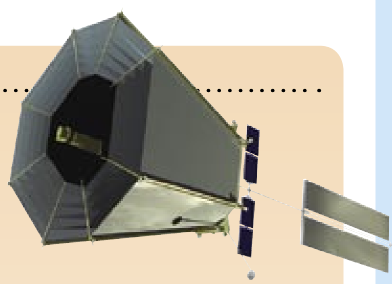


In the early 1970s a researcher at Columbia University designed an experiment to refute Noam Chomsky's claim that language is inherent only in humans. Nim Chimpsky, the baby chimpanzee chosen for this project, was raised in a human family. Not only was he taught American Sign Language, science journalist Hess writes, "he wore human clothes, ate human food, and used a toilet (now and then), and it is likely that he thought of himself as human." For a time he was a genuine celebrity, but when funding for the study ended after four years, Nim was put in a cage and shipped from facility to facility; at one low point, he spent time in a medical lab. His charm, however, and his sizable vocabulary inspired people to help him. He eventually found refuge on Black Beauty Ranch in Texas, where he died, at an early age for a chimpanzee, in 2000. The author uses Nim's troubled life to raise profound questions about the dividing line between humans and other animals and about what we owe to the creatures we use in research.

EXCERPT

➡ THE LIVING COSMOS: OUR SEARCH FOR LIFE IN THE UNIVERSE

by Chris Impey. Random House, 2007 ($27.95)



NASA's Terrestrial Planet Finder, scheduled for launch in 2015.

Impey, who is at the University of Arizona's Seward Observatory, provides a comprehensive, and at times philosophical, examination of whether we are alone in the universe. Much of the culminating discussion focuses on a famous equation used to assess the possibility of life elsewhere in the universe:

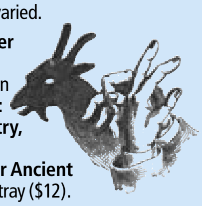
"The young researcher went to the blackboard and paused. The meeting had no agenda, and he wanted to give some structure to the discussion. After thinking for a bit, Frank Drake wrote an equation on the board, not realizing that it would later bear his name and attain iconic status. . . .

"The Drake Equation is a series of numerical factors that combine to give *N,* the number of communicating civilizations in the galaxy at any particular time.... The first factor is the raw material for communication, or the rate at which stars are born that will live long enough to host biology. The next six factors account for the odds that there is any transmission to listen to, either deliberate or inadvertent. Two are astronomical: the fraction of stars with planets and the mean number of habitable planets per star. The next two are biological: the fraction of habitable planets where life actually develops and the fraction of these where intelligent life evolves. The last two depend on culture or sociology: the fraction of civilizations that do communicate over interstellar distances and the typical lifetime of the communicating technology."

NEW BOOKS ON SCIENCE AND ART

1. **Neuroarthistory: From Aristotle and Pliny to Baxandall and Zeki**
by John Onians. Yale University Press, 2008 ($40)
A challenging examination of why art is so widespread—and why its manifestations are both so similar and so varied.

2. **Perspective and Other Optical Illusions,**
by Phoebe McNaughton ($10); **Islamic Design: A Genius for Geometry,**
by Daud Sutton ($12);
The Mayan and Other Ancient Calendars, by Geoff Stray ($12).
Walker, 2007
The latest additions to Walker's series of small books about big—and mainly visual—ideas.



3. **Artscience: Creativity in the Post-Google Generation**
by David Edwards. Harvard University Press, 2008 ($19.95)
Some contemporary creators develop their ideas in a zone where neither art nor science is easily defined, an area the author calls artscience.

4. **Musicophilia: Tales of Music and the Brain**
by Oliver Sacks. Knopf, 2007 ($26)
Sacks tells us why music is irresistible, haunting and unforgettable.

5. **The Magnificent Flora Graeca: How the Mediterranean Came to the English Garden**
by Stephen Harris. Bodleian Library, 2007 ($60)
The story of an 18th-century botanist's journey in the Eastern Mediterranean, with exquisite illustrations by the artist who accompanied him.

NASA/JPL (*satellite*); FROM *PERSPECTIVE AND OTHER OPTICAL ILLUSIONS* (*shadow*)

Q *Do* **cosmic rays** *cause lightning?*

—B. Whiteside, Woodridge, Ill.

Joseph R. Dwyer, a professor of physics and space sciences at the Florida Institute of Technology, has wondered the same thing:

Although some researchers have proposed that cosmic rays instigate lightning, others, including me, have voiced doubts about this theory. At present, the debate remains unsettled.

Decades of measurements inside thunderstorms have failed to find electric fields large enough to spontaneously spark lightning. A mechanism proposed in 1992 by physicist Alex V. Gurevich of the Lebedev Physical Institute in Moscow and his collaborators suggests that the movement of large showers of energetic particles produced by high-energy cosmic rays—which originate from exploding stars—might trigger lightning's massive energy discharge. For Gurevich's mechanism to work, many charged particles must pass through the storm at once. Because cosmic-ray showers alone do not produce enough such particles, Gurevich postulated that a thunderstorm gives the cosmic-ray shower a boost by increasing the number of energetic particles through a process called runaway breakdown.

Runaway breakdown occurs when a cosmic-ray particle hits air molecules in the atmosphere, knocking loose high-energy electrons. As these ejected electrons collide with other air molecules, they generate more runaway electrons as well as x-rays and gamma rays, resulting in an avalanche of energetic particles that tears through the cloud. According to the Gurevich model, this cascade is the catalyst that sparks a lightning bolt.

We know that runaway breakdown does work for the low-level electric fields inside thunderstorms. From observing big bursts of x-rays and gamma rays shooting out of thunderstorms, we also know that it sometimes happens right before lightning strikes. But skepticism still surrounds the cosmic-ray proposal. (Some theories, in fact, involve runaway breakdown spurred by other sources.) The main stumbling block arises because lightning must form a conductive channel to propagate. This channel, extremely hot and just a few centimeters in width, acts like a metal wire, allowing tremendous electric currents to flow through. It remains unclear how a large, diffuse discharge produced by cosmic-ray-induced runaway breakdown would result in such a narrow, hot channel.

Q *How do* **three tiny bones** *amplify sound into the inner ear?*

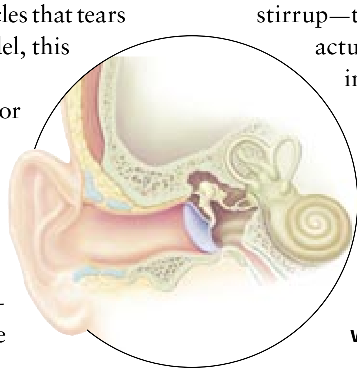
—P. Madsen, Brooklyn, N.Y.

Douglas E. Vetter, assistant professor of neuroscience at the Tufts University Sackler School of Graduate Biomedical Sciences, replies:

The hammer, anvil and stirrup bones of the middle ear—also known as the malleus, incus and stapes, respectively, and as ossicles, collectively—are arranged in a lever system. Their leveraging capabilities, combined with the concentration of vibration energies from the larger eardrum to the much smaller stirrup, efficiently transmit the forces that allow us to hear.

The middle-ear ossicles lie between the eardrum and the cochlea (the spiral-shaped conduit whose hair cells transmit sound to the inner ear). The inner ear is filled with fluid, so our hearing system must transmit airborne sound vibrations to that fluid. Without these ossicles, only about 0.1 percent of sound energy would make it into the inner ear—the rest would reflect off the surface much like voices on land do when a listener is underwater.

When vibrated by sound, the eardrum sets the middle-ear ossicles into motion. One end of the hammer is attached to the eardrum, and the other end forms a hinge with the anvil. The opposite end of the anvil is fused to the stirrup. The footplate of the stirrup—the flat part that resembles the footrest in an actual stirrup—is loosely attached to an opening in the cochlea known as the oval window, and it moves in and out like a piston. This motion transfers the amplified vibrations to the fluid-filled inner ear, thereby signaling the brain of a sound event. ■



HAVE A QUESTION?... *Send it to experts@SciAm.com or go to* **www.SciAm.com/asktheexperts**

GENE RHODEN *Peter Arnold, Inc. (lightning)*; ALICE Y. CHEN *(ear)*

Q Does "Spring Fever" Exist?

BY CHRISTIE NICHOLSON

When the dark chill of winter gives way to the sunny warmth of spring, many people find themselves in the throes of spring fever: restless, energetic, romantic. Other symptoms, says Michael Terman—an expert on biorhythms at New York–Presbyterian Hospital—include increased heart rate, appetite loss and mood swings. Clearly, the condition is real, even if it is not, as Terman notes, "a definitive diagnostic category."

Researchers may lack an explanation of its biological underpinnings, but they do have a number of clues. Some evidence that the changing seasons affect human mood and behavior comes from Matthew Keller of the University of Colorado at Boulder. In a study of 500 individuals in the U.S. and Canada, he found that the more time people spend outside on a sunny spring day, the better their mood. Such good moods decrease during the hotter summer months, and there is an optimal temperature for them: 72 degrees Fahrenheit, otherwise known as room temperature.

Of course, spring does not just lighten our mood; as Alfred, Lord Tennyson famously observed, "In the Spring a young man's fancy lightly turns to thoughts of love." Animal research suggests that this turning of fancy is to some extent hardwired in us. Studies of such mammals as field mice, hares and deer show that births and sexual activity spike in distinct seasonal patterns.

Why the Sap Rises

Mammals track seasons by measuring the length of days—which grow longer in spring—through an internal clock in the brain, and this length is what controls breeding. The clock, called the suprachiasmatic nucleus (SCN), monitors light through a pathway from the retina and conveys information about day length to the pineal gland, at the base of the brain. In response to these signals, the pineal gland modulates its secretion of melatonin, dubbed the sleep hormone because it is released only in dim light or the dark. As nights shorten in spring, the brain's release of melatonin goes down. Some investigators suspect that this decline contributes to the increase in sexual activity, as well as in overall energy, in spring.

Studies of humans also show seasonal cycles in rates of conception. David Lam of the University of Michigan at Ann Arbor's Population Studies Center reports, for example, that birth rates in northern Europe typically rise to 10 percent above the average in March—meaning the babies were conceived in June. Cultural and social factors certainly influence the timing of conception in humans, but analyses of hormones add further support to the notion that biology plays a strong role. One study showed that levels of testosterone in men and luteinizing hormone (which triggers ovulation) in women peak at 20 percent above average during June.

Thomas Wehr of the National Institute of Mental Health says it is plausible that humans, like other mammals, respond to increases in day length by making less melatonin and that the reduction in melatonin, in turn, affects the levels of reproductive hormones, sexual behavior and energy.

Not everyone agrees, however, including Terman. "Melatonin," he suspects, "is more like the hands of the clock; it's not the essential variable."

Terman is more concerned with understanding a disorder that might be considered the opposite of spring fever: winter depression, or seasonal affective disorder (SAD). No one knows the exact cause of SAD either, he says, but it distinctly lifts in spring. A key to that rise in mood, he argues, is the earlier onset of morning light. He has shown that the incidence of winter depression is greater on the western edges of time zones in the U.S., where the sun rises later.

The correlations between mood, behavior and the lengthening of days in spring are real. The precise spur for our renewed energy, however, remains an enigma. ■

Christie Nicholson is community editor at SciAm.com.

MORE ONLINE . . .
Find more investigations into popular myths at
www.SciAm.com/factorfiction
To suggest a topic, write to factorfiction@SciAm.com

STEPHEN KRASEMANN *Getty Images*