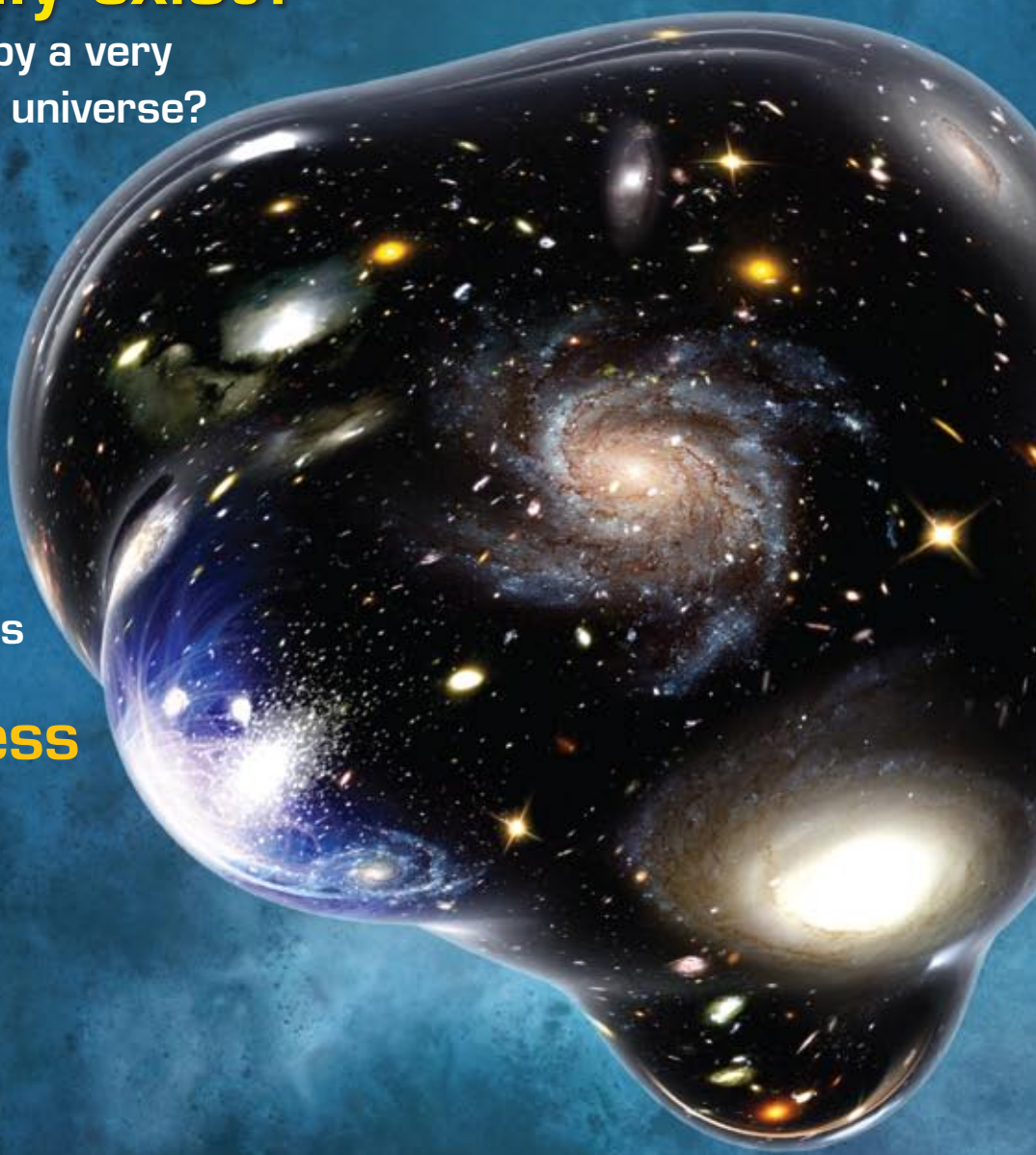
Virus Watch: Preventing the Next Pandemic

SCIENTIFIC AMERICAN

April 2009 $5.99 www.SciAm.com

Solving the Mystery of the VANISHING BEES
page 40

DARK ENERGY
Does it really exist?
Or does Earth occupy a very unusual place in the universe?



Color Vision
Our Eyes Reflect Primate Evolution

Green Lasers
The Next Innovation in Chip-Based Beams

Soldiers' Stress
What Doctors Get Wrong about PTSD

SCIENTIFIC AMERICAN April 2009 ▪ Volume 300 Number 4

ON THE COVER

The observable universe, with the Milky Way galaxy near its center, might be just one relatively empty part of a larger, unevenly expanding cosmos. Image by Kenn Brown, Mondolithic Studios.

MORE FEATURES

GO TO SCIAM.COM

SLIDE SHOW: WORMS 'N US ▼

Meet the charming, slinky creatures that turn your innards into their home sweet home. More at **www.SciAm.com/apr2009**

PHOTOGRAPH BY DAVID SCHARF; COURTESY OF *PARASITIC DISEASES*, FIFTH EDITION, BY DICKSON D. DESPOMMIER ET AL., APPLE TREES PRODUCTIONS, LLC, 2005

In-Depth Report

The Science of Food

Learn about test-tube pork, the myth of the heirloom tomato and the quest to take bug by-products out of candy.

News

Alien Census: Can We Estimate How Much Life Is Out There in Space?

A recent study proposes a computer model to tabulate the extent of extraterrestrial intelligence.

60-Second Psych Podcast

Near Misses Motivate Gamblers

Great news for the house: parts of gamblers' brains that get excited by winning light up at near wins, too.

Ask the Experts

Can a Person Be Scared to Death?

Medical science and the criminal courts agree the answer is yes—as in the case of a fugitive who induced a fatal heart attack in a 79-year-old woman.

60-Second Extinction Countdown Blog

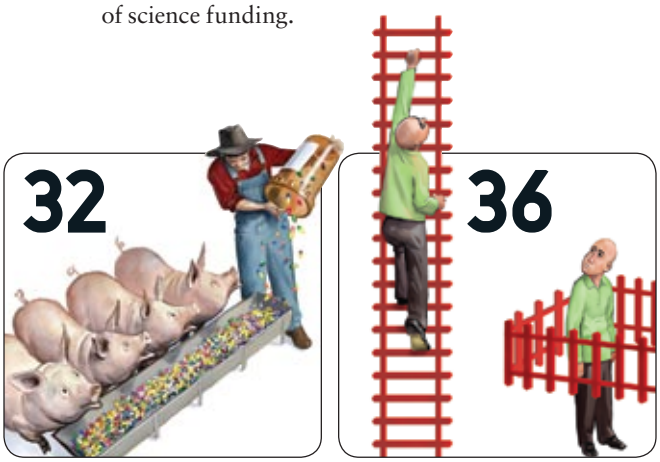
After 1,000 Years, the Milu Returns to the Wild

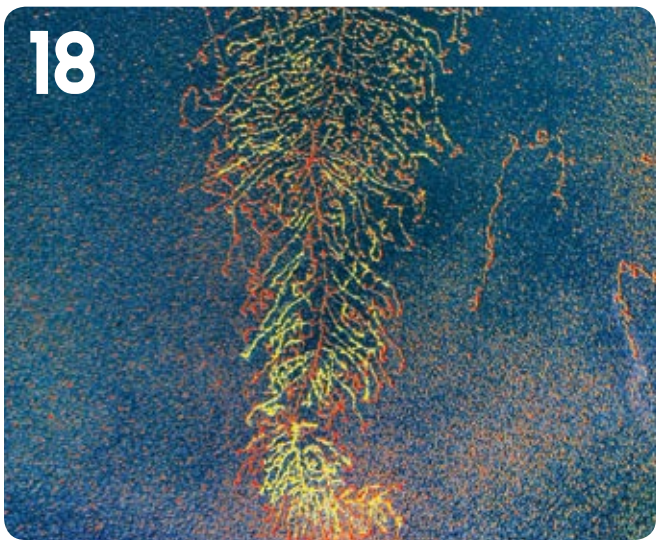
Few species have come as close to extinction and survived as the Chinese milu has.

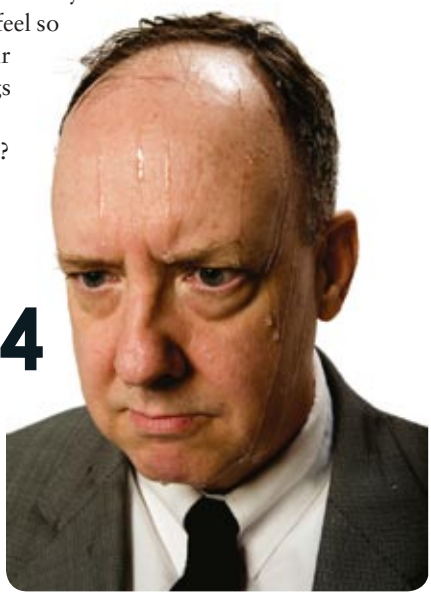
Scientific American (ISSN 0036-8733), published monthly by Scientific American, Inc., 415 Madison Avenue, New York, N.Y. 10017-1111.

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







Nothing Special

We're an ordinary species on an ordinary planet. Or are we?

"You are not special," the character Tyler Durden warns his followers in the movie *Fight Club* and in the namesake novel by Chuck Palahniuk. "You are not a beautiful or unique snowflake. You're the same decaying organic matter as everything else." Durden's harsh but not inaccurate assessment lays the foundation for that story's subsequent tumult. The same idea under the name the "Copernican principle" also happens to have been a linchpin of science for the past four centuries. (The first rule of the Copernican principle is, *Do not talk about the Copernican principle*, but….)

In 1543 Copernicus gave the establishment of his day a bloody nose by proposing that the best explanation for the observed motions of the stars and planets was to picture the sun, not Earth, as the center of known space. He had the prudent good sense to promptly die. Sixty years later the Vatican kayoed two astronomers who forced the point more aggressively: it burned Giordano Bruno at the stake and caged Galileo until he threw in the towel (while angling for a rematch with a mumbled "*Eppur si muove*"). Nevertheless, the facts were on the scientists' side. Astronomers now develop their theories mindful that Earth most likely occupies an ordinary, unprivileged place in the cosmos.

So 11 years ago, when astronomers suddenly realized that the universe was not merely expanding but accelerating in its expansion, most of them concluded that some otherwise undetectable antigravity force,

FLYNN LARSEN (*Rennie*)

Among Our Contributors

TIMOTHY CLIFTON
is a cosmologist and relativity theorist at the University of Oxford and wrote the popular astronomy book *The State of the Universe.*

DIANA COX-FOSTER
studies host-pathogen interactions among insects, including honeybees, as a professor of entomology at Pennsylvania State University.

DAVID DOBBS
is a contributing editor of *Scientific American Mind* as well as author of three books and articles for the *New York Times Magazine, Audubon* and other publications.

a "dark energy," was shoving apart galaxies. An alternative possibility, however, can explain the observations as a fluke of cosmological geometry. It avoids invoking dark energy as an ad hoc cause but at the price of throwing out the Copernican principle: roughly speaking, it puts Earth, or at least our galaxy, back at the center of the observable universe. Timothy Clifton and Pedro G. Ferreira explore that idea in "Does Dark Energy Really Exist?" beginning on page 48.

Even if the Copernican principle's application to cosmology is subject to amendment, its application to other areas of science, notably biology, remains robustly well supported. (The second rule of the Copernican principle is, *Do not talk about the Copernican principle*….) It can nonetheless offend humans' self-importance: witness creationists' ongoing push-back against the evolutionary concept that people are simply another type of animal.

And yet biological evidence of our kinship with other creatures is everywhere we look. Gerald H. Jacobs and Jeremy Nathans reveal the literal truth of that statement in "The Evolution of Primate Color Vision" (*page 56*). Humans, apes and monkeys see a range of colors that other mammals do not; more tellingly, the genetic and biomolecular details of how humans and Old World primates (to whom we are most closely related) see color are different even from those of their New World cousins.

Our relatedness to other animals also leaves us with some common vulnerabilities. When an evolving viral disease hops a species barrier, it can sometimes cause horrific infections. Virologist Nathan Wolfe proposes in "Preventing the Next Pandemic" (*page 76*) that health authorities monitor the status of animal diseases on the verge of leaping to humans. That fight is one we do not want to lose.

JOHN RENNIE,
editor in chief

SHUJI NAKAMURA
pioneered the development of blue diode lasers. He directs the Solid-State Lighting and Energy Center at the University of California, Santa Barbara.

MICHAEL RIORDAN
is co-author, with Lillian Hoddeson, of *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age.*

DENNIS VANENGELSDORP
kept bees as an undergraduate at the University of Guelph and now serves as acting state apiarist for the Commonwealth of Pennsylvania.

NATHAN WOLFE
directs the Global Viral Forecasting Initiative and is a visiting professor of human biology at Stanford University.

editors@SciAm.com

SCIENTIFIC AMERICAN®
Established 1845

Financial Crisis ■ Car Tech ■ Earthrise

DECEMBER 2008

"The SEC decision the editors cite was made by politicians and bank management, none of whom was a practicing quant."
—*Gerald A. Hanweck, Jr.* NEW YORK CITY

■ Don't Shoot the Prognosticator

"After the Crash" [Perspectives] places part of the blame for the current financial crisis on the software models created by the physics and math specialists—whom Wall Street refers to as quants—to police investment banks after the Securities and Exchange Commission (SEC) lifted a rule requiring them to maintain debt ceilings and federal reserves in 2004. Pointing fingers at the quants for their imperfect financial models is like blaming meteorologists for imperfect models of climate change or immunologists for their imperfect models of AIDS. Such scientists overwhelmingly understand that their models are imperfect and are constantly trying to better them.

The crisis was not caused by models run amok. The SEC decision the editors cite was made by politicians and bank management, none of whom was a practicing quant. The banks disregarded their risk management groups, and ambitious politicians and cozy relationships between lenders, servicers and government spurred on lending. Many quant models showed that the housing bubble was growing out of control as far back as 2005. Up and down the chain of leadership—from Congress to the Bush administration to bank management to mortgage lenders and brokers to homebuyers—the problem was ignored.

The editors acknowledge other causes of the crisis but fail to accept that the government owns the lion's share of the blame.

Gerald A. Hanweck, Jr.
New York City

■ Driving Out of Control

In "Driving toward Crashless Cars," Steven Ashley discusses next-generation automotive safety technology that takes various measures of control from the driver, including robotic cars capable of riding in close formation without any driver intervention. Apart from safety concerns, society has yet to utilize such technology to make the flow of traffic more fuel-efficient, a pressing need in this era. In particular, the infrared laser and microwave radar in speed-attenuating collision avoidance systems should be installed on all new cars to simultaneously slow down and space out vehicles. Networked-highway and GPS technology that reads digital signage and computes variable speed limits to slow vehicles ahead of freeway congestion is another fuel-saving and safety-enhancing innovation that could be deployed quickly.

Gregory Wright
Sherman Oaks, Calif.

Reading Ashley's article makes me want to keep my Saturn on the road for a long, long time. The idea of having to compete with a computer for control of a car sounds dangerous. At the very least, the "preset distance" to be maintained by a forward-collision warning system needs to be user-programmable. In Detroit rush-hour traffic, if you let your following distance open up to three quarters of a car length, someone will pull into it.

Lee Helms
Hazel Park, Mich.

Film Stars?

The "Earthrise" photograph taken of our planet from the moon during the *Apollo 8* mission included in "Beacons in the Night," by John Rennie [From the Editor], reminds me of a question that has bothered me ever since I saw it in 1968: Why don't we see any stars in the picture?

Ken Larsen
Salt Lake City

NASA

ASTRONAUT WILLIAM ANDERS took the iconic "Earthrise" photograph while orbiting the moon in the *Apollo 8* spacecraft.

RENNIE REPLIES: *Many people have wondered about that discrepancy over the years. Stars aren't visible in the* Apollo *photographs because the surface of the moon and the earth itself appear so bright. Because of their brightness, the astronauts had to use short exposures. But the film could not handle the level of contrast involved in also picking up the images of the stars, which are many orders of magnitude less bright than the moon and the earth in the foreground—both being illuminated by the relatively nearby sun. If the moon's surface were a dark color and the earth had not been in the sky, it might have been possible to photograph stars casually, but that was not the case.*

Try this experiment on your own: some clear night with a full moon, point a handheld camera at the sky and try to snap a photograph that simultaneously shows clear details of the moon as well as any stars in the field of view. Neither film nor digital cameras can routinely capture such disparate levels of brightness within a single frame.

Pattern Police?

In "Patternicity" [Skeptic], Michael Shermer claims that humans evolved to "find meaningful patterns in meaningless noise," causing people to believe "weird things." As examples, he ascribes various irrational beliefs to "UFOlogists," "religionists" and "conspiracy theorists," thus exhibiting the very kind of false pattern recognition he claims to expose. Just as all snakes with red bands can be mistaken for the poisonous varieties, all religious people can be mistaken for those who see the Virgin Mary on the side of a building, all who seek evidence for extraterrestrial contact can be mistaken for those who see a face on Mars, and so on.

This hypocritical dynamic from the mouths of those who would circumscribe certain "types" of individuals who cannot be trusted to think is corrupting our public discourse about science and reason and injecting dangerously authoritarian modes of thinking into our political debate.

Nelson Leith
Washington, D.C.

SHERMER REPLIES: *Leith is right that it cannot be "authority" by itself that makes the distinction between a true and false pattern, between a false negative and false positive error, between meaningful and meaningless patterns, and so on. In politics, we decide on which pattern of government we want by voting. But in science, we rely on evidence, experimentation, corroboration, repeatability and the other tools designed specifically to avoid making those types of errors in pattern detection.*

ERRATA "Blocking Sound with Holes," by Charles Q. Choi [News Scan], gives incorrect dimensions for aluminum plates used in the described experiments. The plates were two to five millimeters thick and 20 centimeters wide.

"Turning Back the Cellular Clock," by Tim Hornyak [Insights], misstates experimental data regarding mice implanted with induced pluripotent stem cells (iPS cells): 37 mice received iPS cells made with the cancer gene *c-Myc,* and six died. Twenty-six received iPS cells made without *c-Myc,* and none died.

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 each one. Post a comment on any article instantly at www.SciAm.com/sciammag

SCIENTIFIC AMERICAN®
Established 1845

PRESIDENT: Steven Yee
MANAGING DIRECTOR, INTERNATIONAL: Kevin Hause
VICE PRESIDENT: Frances Newburg

MANAGING DIRECTOR, CONSUMER MARKETING: Christian Dorbandt
ASSOCIATE DIRECTOR, CONSUMER MARKETING: Anne Marie O'Keefe
SENIOR MARKETING MANAGER/RETENTION: Catherine Bussey
FULFILLMENT AND DISTRIBUTION MANAGER: Rosa Davis

MANAGING DIRECTOR, ONLINE: Michael Harbolt

VICE PRESIDENT AND PUBLISHER: Bruce Brandfon
DIRECTOR, GLOBAL MEDIA SOLUTIONS: Jeremy A. Abbate
VICE PRESIDENT, MARKETING AND SALES DEVELOPMENT: Michael Voss
SALES DEVELOPMENT MANAGER: David Tirpack
SALES REPRESENTATIVES: Gary Bronson, Jeffrey Crennan, Thomas Nolan, Stan Schmidt

PROMOTION MANAGER: Diane Schube
RESEARCH MANAGER: Aida Dadurian
PROMOTION DESIGN MANAGER: Nancy Mongelli

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

DIRECTOR, SPECIAL PROJECTS: Barth David Schwartz

DIRECTOR, ANCILLARY PRODUCTS: Diane McGarvey
PERMISSIONS MANAGER: Linda Hertz

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.

The Mohole ▪ Analog Voice Recognition ▪ Our Patent Business

Compiled by Daniel C. Schlenoff

APRIL 1959

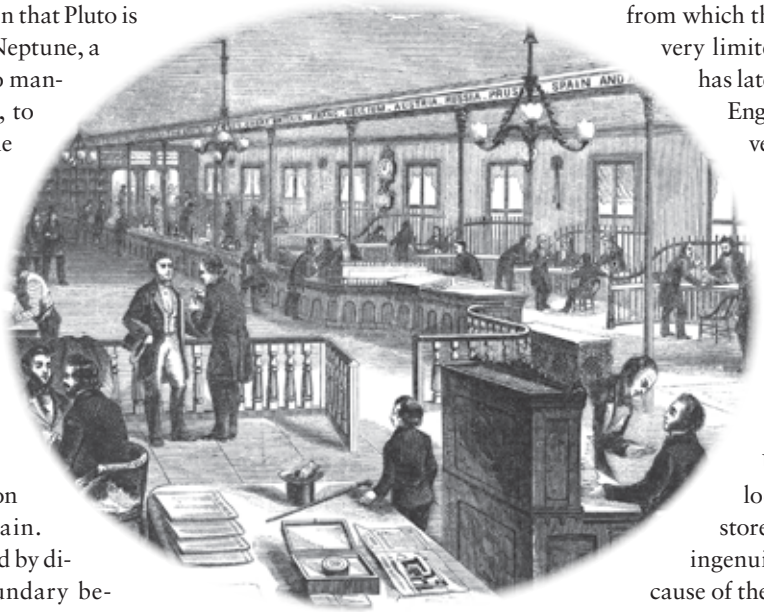
PLANET OR ESCAPEE?—"In their relatively brief acquaintance with Pluto, astronomers have begun to doubt that this object is a planet at all. Pluto's eccentric orbit is tilted at a considerable angle to the plane of the ecliptic, in which the orbits of the other planets lie. Even on its closest approach to our region of the solar system, it will shine no brighter than Triton, one of Neptune's two satellites, suggesting that it is no larger. There is suspicion that Pluto is an illegitimate offspring of Neptune, a satellite that escaped, as two man-made satellites recently did, to ply its own orbit around the sun. —Owen Gingerich"

THE DEEPEST HOLE—"The crust of the earth is a relatively thin film over the earth's interior. Its average thickness is some 10 miles, a mere 400th of the earth's radius. Beneath the crust lies the mantle; important details of its composition and character are uncertain. These can only be determined by direct examination. The boundary between crust and mantle is the Mohorovicic discontinuity, known to earth scientists as the Moho. To obtain a sample of the mantle, we must drill a hole through the Moho: a Mohole. —William Bascom"

[NOTE: Work on the unfinished Mohole was abandoned in 1966.]

APRIL 1909

LIGHTER MONEY—"Experiments in abrasion conducted at the French mint have proved that aluminium coins will be less rapidly worn by use than coins of gold, silver, or even bronze. The metal's extreme lightness is another advantage: it is four times lighter than silver. Hence aluminium coins could be carried in considerable quantities without inconvenience. The total nominal value of bronze 5 and 10 centime pieces in circulation is estimated to be about 56 million francs. It is proposed to replace some 50 million francs' worth of these with aluminium coins of the same denominations. About 2,000 tons of aluminium, worth 44 cents a pound in blanks ready for stamping, will be required."



THE SCIENTIFIC AMERICAN Patent Agency, New York City, 1859

VOICE RECOGNITION—"A safe lock has been invented which is provided with a phonographic mechanism so that it can be opened only by the voice of the owner. A mouthpiece like that of a telephone takes the place of a knob on the door, and this is provided with the usual needle, which travels in a groove in the sound record of the phonograph cylinder. Before the safe can be unlocked, the password must be spoken into the cylinder by the one who made the original record. The report does not state what would occur if the owner should come down to his office with a bad cold."

APRIL 1859

SOME SAY ALUMINUM—"It is only a few years ago that this valuable metal was uncommon and expensive, owing chiefly to the difficulty of reducing it from its oxyde. We believe that about three years ago, its market value was no less than $18 per ounce. In a very outcast region of the world—on the west coast of cold Greenland—an aluminous mineral called cryolite has been discovered in great quantities, from which the metal can be reduced at a very limited cost, and a large factory has lately been erected at Battersea, England, by M. Gerhard, for this very purpose. He has been able to sell it for about one dollar per ounce. Aluminum is the lightest of all the metals. This quality should recommend it for coinage, to take the place of coins of the lowest value."

PATENT AGENCY—"The United States Patent Office, located at Washington, is the storehouse and monument of the ingenuity of our countrymen. Because of the value of many of the inventions for which patents are sought, and the great necessity that the papers be carefully prepared, there has grown up a profession, the members of which are usually designated 'patent agents' or 'patent solicitors,' and who have become as much a necessity for the proper transaction of business with the Patent Office as the lawyers are in our courts of justice. We will here state, in reference to ourselves, what no one will presume to deny, that since 1846, the Scientific American Patent Agency Department has examined into the novelty of more inventions than any other patent agents now living in this country. We present to our readers an illustration of the interior view of the 'Scientific American' and patent agency office, New York."

SCIENTIFIC AMERICAN, VOL. XIV, NO. 32, APRIL 16, 1859

Fingerprint Feeling ▪ Stem Cell Progress ▪ Moon's Backside ▪ Even Higher Seas

Edited by Philip Yam

Sensation Swirls

Those fingertip whorls aren't just good for gripping objects and identifying people; they also enable you to feel fine textures and tiny objects. French researchers constructed two mechanical sensors,

one with a ridged end tip and another with a smooth one; they then ran them over various textured surfaces, measuring the vibrations picked up by the fingerprinted sensors. Each ridge magnified the frequency range well suited for detection by nerve endings in the skin called Pacinian corpuscles. The work, published online January 29 by *Science,* helps to explain how the sense of touch accurately informs our surroundings [see "Worlds of Feeling"; SciAm Mind, December 2004]. —*Kate Wilcox*

Forward with Stem Cells

Researchers at Northwestern University stopped and, in some cases, reversed the effects of early-stage multiple sclerosis, a disease in which the immune system attacks the central nervous system. The investigators removed from the bone marrow so-called hematopoietic stem cells, which resupply the body with fresh blood cells, then used drugs to destroy the patients' existing immune cells. Injected back into the patients, the stem cells apparently "reset" the body's defenses so that they do not mistakenly go after healthy tissue. The study, published online January 30 in the *Lancet Neurology,* involved only 21 patients, however, so more complete trials are needed to assess the approach.

Data on stem cell therapies in general should increase soon: the U.S. Food and Drug Administration approved in January the first human embryonic stem cell trial—a decision long awaited by scientists [see "The Stem Cell Challenge"; SciAm, June 2004]. It will enable Geron Corporation in Menlo Park, Calif., to test embryonic stem cells on 10 patients with spinal cord injuries. —*Kate Wilcox*

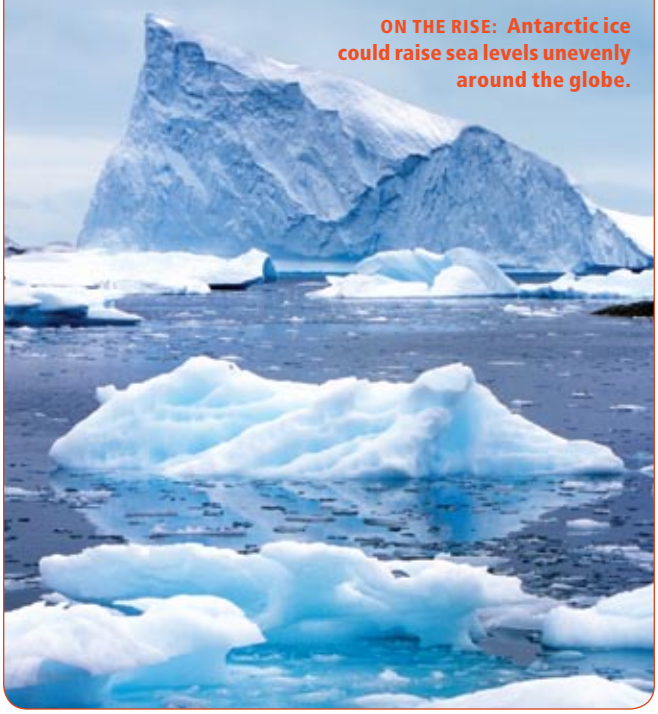
Mooned by the Moon

Some four billion years ago the far side of the moon may have faced Earth. Mark Wieczorek and Mathieu Le Feuvre of the Institute of Earth Physics in Paris propose that if the moon had always faced the same way, it would have more craters on its leading side, where it encountered heavy bombardment during the early years of the solar system [see "The New Moon"; SciAm, December 2003]. Younger craters follow this pattern, but older craters do not; they instead cluster on the trailing edge, suggesting that it was once in front. An asteroid or comet strike could have spun the moon 180 degrees to its current orientation. The journal *Icarus* posted the conclusion online on December 31, 2008. —*John Matson*

OUT OF SIGHT: The far side of the moon, as seen by *Apollo 16.*

Melting Mess

New details are emerging on how the melting poles could raise ocean heights [see "The Unquiet Ice"; SciAm, February 2008]. Researchers at the University of Toronto and Oregon State University suggest that the rise could be uneven around the world. They examined the West Antarctic ice sheet, which contains enough grounded ice to boost global sea levels by five meters if it splashed into the water. But such a huge redistribution of mass in Antarctica would reduce the gravitational pull in the area and shift the earth's rotation axis by 500 meters. Taking these and other factors into account, they figure that the seas will drop near Antarctica but rise in the Northern Hemisphere by an additional one to two meters above previous estimates. Gravitate toward the analysis in the February 6 *Science.*



ON THE RISE: Antarctic ice could raise sea levels unevenly around the globe.

SETH JOEL *Getty Images (fingerprint)*; JOE SOHM *Getty Images (ice)*; COURTESY OF NSSDC AND NASA (*moon*)

BIOLOGY

Evolution in a Bottle

Self-replicating RNAs advance science another step toward artificial life BY W. WAYT GIBBS

Gerald F. Joyce admits that when he saw the results of the experiment, he was tempted to halt further work and publish the results immediately. After years of trying, he and his student Tracey Lincoln had finally found a couple of short but powerful RNA sequences that when mixed together along with a slurry of simpler RNA building blocks will double in number again and again, expanding 10-fold in a few hours and continuing to replicate as long as they have space and raw material.

But Joyce was not fully satisfied. A professor and dean at the Scripps Research Institute in La Jolla, Calif., the 53-year-old molecular chemist is one of the founding champions of the "RNA world" hypothesis. That is the notion that perhaps life as we know it—life based on DNA and enzymatic proteins, with RNA acting for the most part as a mere courier of genetic information—evolved out of a simpler, prebiotic chemical system based mostly or even solely on RNA. Of course, the idea is plausible only if RNA can support evolution on its own. Maybe, Joyce thought, his synthetic RNA could help prove that possible. So he and Lincoln spent another year working with the molecules, mutating them and setting up competitions in which only the fittest would survive.

In January, one month before the bicentenary of Charles Darwin's birth, they announced the results in *Science*. Their little test-tube system did indeed manifest nearly all the essential characteristics of Darwinian evolution. The starting 24 RNA variants reproduced, some faster than others depending on the environmental conditions. Each molecular species competed with the others for the common pool of building blocks. And the reproduction process was imperfect, so new mutants—Joyce calls them recombinants—soon appeared and even thrived.

"We let it run for 100 hours," Joyce recalls, "during which we saw an overall amplification in the number of replicator molecules by 10^{23}. Pretty soon the original replicator types died out, and the recombinants began to take over the population." None of the recombinants, however, could do something new—that is, something that none of its ancestors could perform.

That crucial missing ingredient still separates artificial evolution from true Darwinian evolution. "This is not alive," Joyce emphasizes. "In life, novel function can be invented out of whole cloth. We don't have that. Our goal is to make life in the lab, but to get there we need to increase the complexity of the system so that it can start inventing new function, rather than just optimizing the function we've designed into it."

That goal clearly seems possible, because the RNA replicators in Joyce's lab were relatively simple: each has only two genelike sections that can vary. Each of those "genes" is a short building block of RNA. A replicator, being an RNA enzyme, can gather the two genes and link them together to create an

OSCAR L. MILLER *Photo Researchers, Inc.*

GROWING LIFE: Reproducing molecules of RNA branch out horizontally from a central spine of DNA. Such RNA can now demonstrate most of the essential aspects of evolution in a test tube. For synthetic life, however, they also need the ability to evolve brand new functions.

enzyme that is the replicator's "mate." The mate is set free and gathers two loose genes, which it assembles into a clone of the original replicator. Recombinants appear when a mate is unfaithful and links up genes that were never meant for each other. Recombinants did not, however, create genes. It may be possible to engineer a system that does, or to add complexity by giving each replicator more genes with which to work.

Scott K. Silverman, a chemist at the University of Illinois who has done pioneering work with DNA enzymes, hopes that "by capturing Darwinian evolution in new molecules, we might be able to better understand the basic principles of biological evolution," much of which is still somewhat mysterious at the molecular level. Joyce and Lincoln, for example, noticed in their postmortem examination of the experiment that the three most successful recombinants had formed a clique. Whenever any clique member made a reproduction error, the result was one of the other two peers.

The next big step toward the creation of life in the lab, Joyce says, will be to engineer (or evolve) a set of synthetic molecules that can perform metabolism as well as replication. Geneticist Jack W. Szostak of Harvard Medical School has developed nonbiological proteins that bind ATP, an energy-carrying chemical crucial to metabolism. Szostak's lab is also attempting to fashion protocells that encase RNA within tiny spheres of fatty acids, called micelles, that can form, merge and replicate spontaneously.

Even if biochemists do manage to cobble RNA and other basic compounds into some form of synthetic life, the engineered system will probably be so complex at first that it will hardly prove that natural life began in some similar way, four billion years ago. Joyce's replicators consist of a mere 50 chemical letters, but the odds of such a sequence appearing by chance are roughly one in 10^{30}, he notes. "If it were six or even 10 letters long, then I'd say we might be in the realm of plausibility, where one could imagine them assembling spontaneously" in the primordial soup.

From Test-Tube Life to Diagnostic Tools

Creating life in the laboratory would be a momentous occasion for humanity, even if it is more molecular than Frankensteinian. But there may be more mundane uses for such chemistry. A paper in press at *Nature Biotechnology,* Gerald F. Joyce says, describes how his lab at the Scripps Research Institute in La Jolla, Calif., has modified RNA replicators so that they must perform a biochemical function to reproduce. The winners of that evolutionary race will be good candidates for a medical diagnostic, he thinks. Scott K. Silverman of the University of Illinois says the idea has merit: "Suppose you need to do detection in a dirty environment with lots of different chemicals present—say you want to find *Salmonella* inside peanut butter. That's hard to do without purification steps. It would be useful to be able to evolve the diagnostic system so that it still finds the signal despite all the noise."

—*W.W.G.*

MICROFLUIDICS

A Chip against Cancer

Evaluating tumors and their treatment via a blood sample BY ELAINE SCHATTNER

Cancer therapy is too often a matter of chance. Despite advances in cancer genetics, physicians have only limited information to make decisions about individual patients. People undergo treatments with their fingers crossed, not knowing if they will be helped or harmed.

A group at Massachusetts General Hospital might have a way to personalize cancer care more effectively. It is currently refining and testing a lab-on-a-chip that can sample and analyze the circulating tumor cells from just a teaspoon of a patient's blood, obviating the need, in many patients, for sometimes dangerous biopsies. "The chip will allow rational decision making for cancer patients," says Mehmet Toner, leader of the team that engineered the device at Boston's BioMicroElectroMechanical Systems Resource Center. Toner likens his new system to the way AIDS patients have their viral load and T cells measured so that their medication can be adjusted. "It could be the same for cancer," he offers.

Most carcinomas shed malignant cells that enter the bloodstream and disseminate, sometimes latching on to new areas where they form tumors. These circulating tumor cells (CTCs) constitute just a tiny fraction of blood cells, often fewer than one in a million, in patients with metastatic disease; they are even less abundant in patients with limited, early-stage tumors that have not overtly spread. The researchers realized that CTCs, though rare, offer a potential window into the real-time dynamics of a tumor's biology.

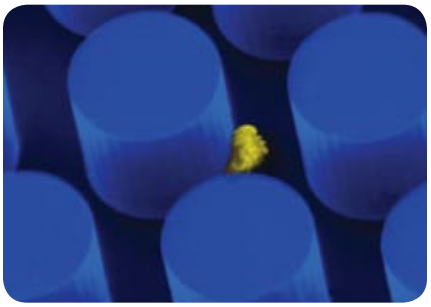
The team has adapted microfluidics technology, developed during the past 25 years to analyze tiny amounts of fluid and gas, to capture those uncommon cells. Like other microfluidics instruments, the CTC-chip, as the group calls it, comprises a silicon-etched chip fitted with microscopic columns, a chamber to enclose the fluid and chip, and a pneumatic pump. The columns, or microposts, function as miniature test tubes where cells and chemicals can mix, adhere and undergo evaluation.

The CTC-chip relies on 78,000 micro-

posts to grab cancer cells from a mix of normal blood components as they meander through the system via exquisitely controlled suction. The posts are coated with antibodies to the epithelial cell adhesion molecule (EpCAM). Nearly all carcinoma cells bear EpCAM at the surface, where it plays a key role in directing how cells bind to one another, signal and migrate. Normal blood cells lack EpCAM, so only the malignant cells stick to the antibodies on the microposts.

In their first test of the chip, described in the December 20, 2007, *Nature*, the researchers used blood samples from 116 patients with lung, prostate, pancreatic, breast or colorectal cancer and successfully isolated CTCs in all but one case. The CTC-chip finds cancer cells as rare as one in a billion blood cells, making it at least 100-fold more powerful than the most widely used conventional method, which requires incubating a large blood sample with antibody-coated microbeads. The cells are also in better condition for analysis than those prepared using microbeads.

In another trial, the investigators used the chip to evaluate tumor genetics in 27 patients with lung cancer. In work published in the July 24, 2008, *New England Journal of Medicine*, they identified relevant genetic abnormalities in CTCs from most cases and noted in some patients emerging mutations that confer resistance to tyrosine kinase inhibitors, the type of medication the patients were taking. In the past, repeat biopsy would have been necessary to establish these kinds of genetic changes.



STUCK ON YOU: In the CTC-chip, a lung cancer cell is caught by a microscopic post coated with an antibody that binds to a surface protein on tumor cells. Each post is 100 microns in diameter and 100 microns tall.

COURTESY OF MASSACHUSETTS GENERAL HOSPITAL BIOMEMS RESOURCE CENTER

"It's an enormous advance in our ability to monitor patients," comments Roy Herbst, a thoracic oncologist at the University of Texas M.D. Anderson Cancer Center, who was not involved in the work. By providing a noninvasive method to follow the quantity and quality of tumor cells, the chip "offers the possibility of personalized medicine and hence effective therapy," he says.

For the 215,000 people in the U.S. diagnosed with lung cancer every year, the need for better diagnostic tools is particularly urgent, says Toner's collaborator Thomas Lynch, chief of the center for thoracic cancers at Mass General. In lung cancer patients, even a small biopsy bears risk of blood loss, infection and, in rare instances, collapse of the affected lung.

Herbst cautions that the findings need to be validated in larger clinical trials and at other medical centers. At Mass General, the researchers are now evaluating how well the chip measures cancer growth and responses to treatment in patients with breast, ovarian and prostate cancers.

By direct examination of cancer cells in blood, the CTC-chip might also uncover new targets for therapy and help determine when and how metastases arise. Toner sees unlimited possibilities once the chip has proved itself in bigger clinical studies. In the future, he remarks, it could become a screening tool to find nascent cancers and even "could be used at annual checkups."

Elaine Schattner is a science writer and oncologist based in New York City.

Getting Personal with Cancer Care

Cancer therapies often fail because of genetic differences among individuals. In personalized cancer care, physicians would tailor treatment to the particular features of each patient's cancer. For example, colon cancer patients often receive treatments with antibodies directed at a growth factor receptor. Although this therapy, costing nearly $10,000 per month, can prolong life and make some patients feel better, it works only in those who have the normal version of *K-ras*, a gene for a signaling protein crucial for tumor growth. So if physicians examine the tumors for *K-ras* mutations before prescribing these drugs, they can direct care to those it would help, sparing others the costs, hassle and possible side effects of treatment. —***E.S.***

PARTICLE PHYSICS

Colliding Philosophies

A novel way to rummage for particles in accelerator debris BY DAVIDE CASTELVECCHI

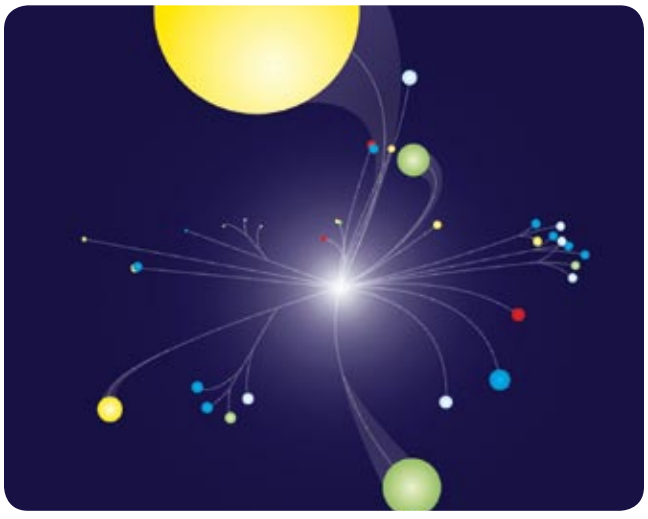
After a false start in 2008, the Large Hadron Collider (LHC), the glitzy new atom smasher at CERN (the European laboratory for particle physics) near Geneva, is finally due to start its experiments this October. The LHC may or may not end up spewing out dark matter, mini black holes or other exotica. But whichever way, figuring what's coming out will be a tremendously hard task. A controversial approach to analyzing data could help physicists make sure they don't miss any of the good stuff.

The LHC and other accelerators such as the Tevatron at the Fermi National Accel-

erator Laboratory in Batavia, Ill., push protons or other particles to near light speed and smash them together. Thanks to Albert Einstein's $E = mc^2$, some of that collision energy turns into rare, heavy particles that almost immediately decay into hundreds of more mundane particles (of which many dozens of different types are known). The LHC's huge detectors will record the passage of this debris and produce data at a staggering rate, equivalent to one CD-ROM per second.

Physicists will rummage through the information for particular combinations of decay products that would suggest a new particle has been created. They will be looking for signs of the Higgs boson, the long-sought particle that is supposed to give other particles their masses, and also for entirely new particles that could give a first glimpse of the laws of physics at higher energies.

But some fear that this traditional approach—akin to running a computer algorithm through a text searching for the letters H-I-G-G-S—could end up missing interesting new signatures that no one had foreseen. At Fermilab, Bruce Knuteson and Stephen Mrenna have for some years advocated a more "holistic" approach called global search. Instead of looking for particular signatures, they wrote software that analyzes all the data and compares them with predictions of the so-called Standard Model, which comprises the known set of laws of particle physics. The software then flags any deviations from the Standard Model as potential new particles. It is a bit like having an algorithm that, instead of searching a text for a particular word, matches every single word against the dictionary of known words and flags the ones that sound as if they



COURTESY OF SANDBOX STUDIO

COLLISION POINT in accelerator experiments is where new particles emerge. A novel technique could uncover unpredicted particles.

might belong to a foreign language.

To limit false positives—sometimes mundane particles will interact and mimic the behavior of other, more interesting particles—physicists can set a threshold for the minimum number of times a strange event may occur before alerting the experimenters of something possibly new. "We take into account the fact that we look at a lot of different places," Knuteson says.

Knuteson, Mrenna and their collaborators put their method to work on old Tevatron data. In principle, exotic particles could have been lurking where no targeted searches had looked before. The team found nothing of particular statistical relevance, so they made no claims of new discoveries. But that effort at least showed that global searches do not necessarily lead to many false positives, as some physicists feared. The results, which appear in the January *Physical Review D*, also constitute the Standard Model's most stringent test to date, says Knuteson, who has since left active research.

Physicist Louis Lyons of the University of Oxford says the team's statistics were sound. But Pekka Sinervo, a University of Toronto physicist who is involved in both Tevatron and LHC experiments, remains unconvinced. "The authors had to sweep a lot of poorly understood effects 'under the carpet' and not address them directly," Sinervo states, meaning that the search generated an abundance of hard-to-interpret signals. Still, global searches could have some utility, he concedes, as long as they do not distract researchers from searches targeted at specific phenomena, adding that he is "not convinced that one would be able to use such a search for an early discovery at the LHC."

That may be true, remarks Sascha Caron, a physicist from the University of Freiburg in Germany, but nonetheless much of the particle physics community has warmed to the idea of global searches since Knuteson first proposed it early in this decade. Caron and his collaborators have developed their own software for what they call general searches while working at an experiment at the DESY laboratory in Hamburg, and they plan to do the same at the LHC.

Experience at the Tevatron also shows that global searches could help physicists understand how to interpret data, Mrenna points out—for example, how the detectors react to various particles. Teams rarely compare notes, so that their assumptions are potentially in contradiction. "If you look at everything, everything has to make sense," Mrenna says.

SENSES

Finding Balance

Is poor posture control the real cause of motion sickness? BY BRENDAN BORRELL

To avoid a potentially nasty cleanup, students are deprived of food for four hours before entering the experimental chamber. Inside, they step onto a force-sensing platform and stare at a paper map of the U.S. The walls begin sliding back and forth on a track—a mere 1.8 centimeters with each cycle. At the right frequencies, this movement triggers a tugging sensation that begins somewhere in the brain and mysteriously travels to the belly. But before the full effects of motion sickness set in, the subjects will typically turn away and beg for mercy.

At that point, the students may be thinking that the course credit they will receive for participating in Thomas Stoffregen's sensory funhouse may not be worth it. But for the University of Minnesota psychologist, every student he lures inside is another data point that he believes will overturn the dogma about the cause of motion sickness. If he is right, the findings could lead to new ways of identifying people susceptible to motion sickness before they get sick and may provide designers of simulators and video games with ways to keep controllers in the hands of potentially woozy players.

DAVID SACKS Getty Images

ROCKING THE BOAT: A maverick theory claims that motion sickness arises from postural instability, not sensory conflict.

For the past century, scientists have believed that motion sickness derives from a conflict among our senses. Our inner ears contain sensors for both angular motion (the semicircular canals) and linear motion (the otoliths). When these sensors disagree with the information we expect to receive from our eyes and muscles, motion sickness manifests itself. And yet, Stoffregen says, our senses constantly provide different channels of information: redundancy is an essential part of the sensory system, and our brains do not compare our senses in any direct manner. Moreover, because it is impossible to determine which of the conflicting senses is interpreted as being "wrong" by an individual's brain, Stoffregen has branded the conflict theory with the highest-order insult a scientist can muster: unfalsifiable.

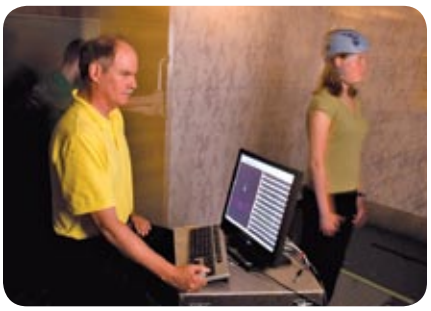
Indeed, researchers have long won-

dered why some individuals and certain classes of people—children and pregnant women—are more susceptible to motion sickness than others. What is more, experiments conducted since the dawn of the Space Age, when NASA wanted to prevent its astronauts from falling ill, can predict who will succumb to motion sickness with only about 30 percent accuracy. Finally, Stoffregen notes the puzzling observation that, whereas people may feel sick on the deck of boat, they rarely get the urge to vomit when fully immersed in water.

Stoffregen instead argues that motion sickness comes from the brain's persistent inability to modulate the body's movements in a challenging environment. Postural instability—the inability to maintain balance—was considered a symptom of motion sickness. Not so, Stoffregen says. Although postural control relies on sensory feedback, motion sickness is really a sign that the motor-control system is going haywire.

His alternative theory, first published in 1991, landed with a dull thud, and his papers garner just a handful of citations each year. But experts have been muttering privately about Stoffregen ever since—and lately some have grudgingly begun to accept him. "It was a very, very different theory," says Larry Hettinger, a longtime motion sickness researcher now at defense contractor Northrop Grumman. "I clearly remember people thinking, 'This is nonsense, this is crazy foolishness.' "

The growing acceptance of Stoffregen's view has much to do with experiments conducted over the past two decades. Inside the moving room, he has found that volunteers can significantly reduce motion sickness simply by widening their stance—an observation, he says, not predicted by sensory conflict theory. Students standing with their feet five centimeters apart tend to get motion sickness about 60 percent of the time. Spreading their legs to 30 centimeters increases the stability of the head and torso and decreases the incidence of motion sickness to about 20 percent. Stoffregen says that by monitoring body sway, he can predict the onset of motion sickness with 60 percent accuracy. If swaying is just a symptom of motion sickness, it would be detectable only after participants reported feeling sick.



COURTESY OF PATRICK O'LEARY *University of Minnesota*

NAUSEATING: Thomas Stoffregen monitors a volunteer as an assistant moves the walls back and forth to induce motion sickness.

But the ultimate test of his theory is still in the works. Floating in water, the human body becomes passively stable, and pos-

tural control is no longer an issue. If Stoffregen is right, then under such a condition, motion sickness would be impossible, even if subjects were forced to endure the nausea-inducing camerawork of *The Blair Witch Project*. Stoffregen just needs to convince NASA—and a dozen amphibious students—to let him use the Neutral Buoyancy Laboratory at the Johnson Space Center in Houston. Unfortunately, the practical applications of that research would be tenuous at best. "I think that water immersion would be a surefire way to prevent motion sickness in orbital flight," he says. "Sadly, water-filled spacecraft would be so heavy that they would be too expensive to launch." There's always Dramamine.

Brendan Borrell is a freelance writer based in New York City.

BEHAVIOR

Thriving on Selfishness

Why it pays for cheaters to punish other cheaters BY MARINA KRAKOVSKY

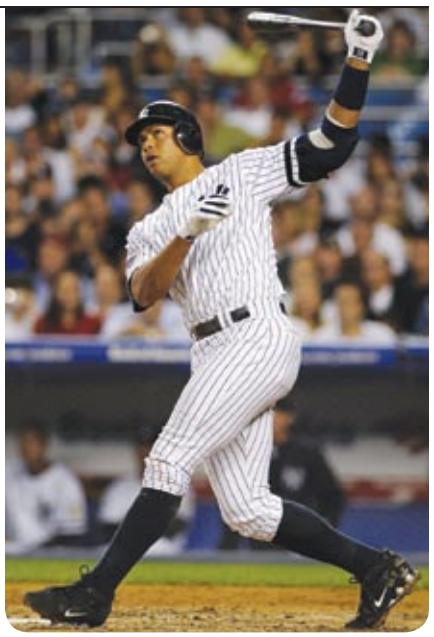
It's the altruism paradox: If everyone in a group helps fellow members, everyone is better off—yet as more work selflessly for the common good, cheating becomes tempting, because individuals can enjoy more personal gain if they do not chip in. But as freeloaders exploit the do-gooders, everybody's payoff from altruism shrinks.

All kinds of social creatures, from humans down to insects and germs, must cope with this problem; if they do not, cheaters take over and leech the group to death. So how does altruism flourish? Two answers have predominated over the years: kin selection, which explains altruism toward genetic relatives—and reciprocity—the tendency to help those who have helped us. Adding to these solutions, evolutionary biologist Omar Tonsi Eldakar came up with a clever new one: cheaters help to sustain altruism by punishing other cheaters, a strategy called selfish punishment.

"All the theories addressed how altruists keep the selfish guys out," explains Eldakar, who described his model with his Ph.D. thesis adviser David Sloan Wilson of Binghamton University in May 2008. Because selfishness undermines altruism, altruists certainly have an incentive to punish cheaters—a widespread behavior pattern known as altruistic punishment. But cheaters, Eldakar realized, also have reason to punish cheaters, only for motives of their own: a group with too many cheaters does not have enough altruists to exploit. As Eldakar puts it, "If you're a single selfish individual in a group of altruists, the best thing you can do evolutionarily is to make sure nobody else becomes selfish—make sure you're the only one." That is why, he points out, some of the harshest critics of sports doping, for example, turn out to be guilty of steroid use themselves: cheating gives athletes an edge only if their competitors aren't doing it, too.

Although it is hypocritical for cheaters to punish other cheaters, members of the group do not balk as long as they benefit. And when selfish punishment works well, benefit they do. In a colony of tree wasps (where workers care for the queen's offspring instead of laying their own eggs), a special caste of wasps sting other worker wasps that try to lay eggs, even as the vigilante wasps get away with laying eggs themselves. In a strange but mutually beneficial bargain, punishing other cheaters earns punishers the right to cheat.

In the year since Eldakar and Wilson wrote up their analysis, their insights have remained largely under the radar. But the idea of a division of labor between cooperators and policing defectors appeals to Pete Richerson, who studies the evolution of cooperation at the University of California, Davis. "It's nothing as complicated as a salary, but allowing the punishers to defect in effect does compensate them for their services in punishing other defectors who don't punish," he says. After all, policing often takes effort and personal risk, and not all altruists are willing to bear those costs.



JEFF ZELEVANSKY *Getty Images*

DO AS I SAY, NOT AS I DO: Athletes who have used illicit performance-enhancing drugs, such as Yankee third baseman Alex Rodriguez, can also be their harshest critics.

Corrupt policing may evoke images of the mafia, and indeed Eldakar notes that when the mob monopolizes crime in a neighborhood, the community is essentially paying for protection from rival gangs—a deal that, done right, lowers crime and increases prosperity. But mob dynamics are not always so benign, as the history of organized crime reveals. "What starts out as a bunch of goons with guns willing to punish people [for breaching contracts] becomes a protection racket," Richerson says. The next question, therefore, is, What keeps the selfish punishers themselves from overexploiting the group?

Wilson readily acknowledges this limitation of the selfish punishment model. Although selfish punishers allow cooperators to gain a foothold within a group, thus creating a mix of cheaters and coop-

erators, "there's nothing telling us that that mix is an optimal mix," he explains. The answer to that problem, he says, is competition not between individuals in a group but between groups. That is because whereas selfishness beats altruism within groups, altruistic groups are more likely to survive than selfish groups. So although selfish punishment aids altruism from within a group, the model also bolsters the idea of group selection, a concept that has seen cycles of popularity in evolutionary biology.

What is more, altruism sometimes evolves without selfish punishment. In a software simulation, Eldakar and Wilson have found that as the cost of punishing cheaters falls, so do the number of selfish punishers. "When punishment is cheap, lots of people punish," Wilson explains. And among humans, there is no shortage of low-cost ways to keep others in line—from outright ostracism to good old-fashioned gossip.

Marina Krakovsky is based in the San Francisco Bay Area.

ECOLOGY

Snakebit

Southern California sees a rise in extratoxic venom BY MICHAEL TENNESEN

Rapid muscle twitching, as if a person had snakes crawling under the skin, is the telltale sign to Roy Johnson that the Southern Pacific rattlesnake has struck. On occasion, this symptom can progress to difficult breathing, coma and death. This snake's bite is one of the few to induce neurological symptoms, in contrast to most other rattlesnake bites, which initially produce swelling and bruising around the wound, notes Johnson, a physician in Palomar, Calif., who has treated some 700 snakebite cases. Increasingly, the proportion of rattlesnake bites in southern California are skewing to those like the more deadly Southern Pacific species, and scientists are not sure why.

Every year the Centers for Disease Con-

trol and Prevention logs 7,000 reports of snakebites in the U.S., which lead to about 15 deaths. Roughly 25 percent of the survivors incur some permanent damage. In southern California, reports from area hospitals and medical centers show a spike in serious bites—the facilities say that, where they formerly saw patients with severe neurological symptoms once every two to three years, they now see several of these types of envenomations every year.

Johnson, for one, suspects that humans themselves are to blame for the increase. Most rattlesnakes warn off potential predators by shaking their noisy tails—that is what the red rattlesnake and the speckled rattlesnake, southern California's other two dominant coastal species, tend to do. But the noise also makes the reptiles more likely to end up on the killing end of a shovel if the threat is human. In contrast, Southern Pacific rattlesnakes are more apt to lay low or move away than hiss and rattle when confronted, a strategy that may boost their chances of surviving, Johnson says. He speculates that by clubbing its competitors, hu-

PALEONTOLOGY

Supersized Serpent

JUMBO BACKBONE of the 13-meter-long fossil snake *Titanoboa* dwarfs the vertebra of a 5.2-meter-long modern Anaconda.

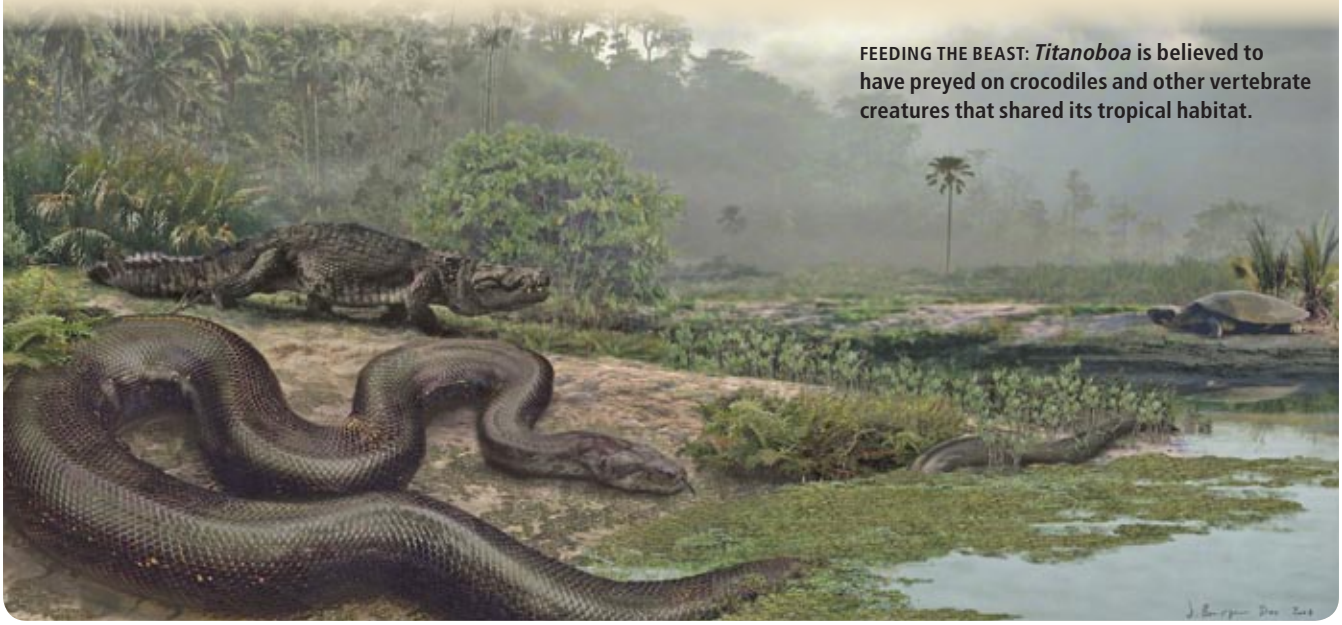
Fossilized remains of a boa constrictor cousin that stretched 13 meters long and tipped the scales at more than a ton represent the largest snake ever found. The creature, dubbed *Titanoboa cerrejonensis,* lived some 60 million years ago in a neotropical rain forest in what is now northeastern Colombia. Identified on the basis of vertebrae recovered from an open-pit coal mine, *Titanoboa* is believed to have dined on crocodiles, among other creatures.

In addition to expanding the known limits of snake biology, the ancient serpent contains clues to primeval rain forest climate. Because snakes and other reptiles are "cold-blooded," or poikilothermic, their body temperature—and hence their life processes—is dependent on that of the surrounding air. The warmer the air is, the larger they can grow.

Scientists calculate that to attain its behemoth body size (which bests that of the modern-day record holder, a reticulated python, by nearly three meters), *Titanoboa* would have to have inhabited an environment with a mean annual temperature of at least 30 to 34 degrees Celsius (86 to 93 degrees Fahrenheit)—significantly toastier than today's typical tropical forecast of 24 to 26 degrees C.

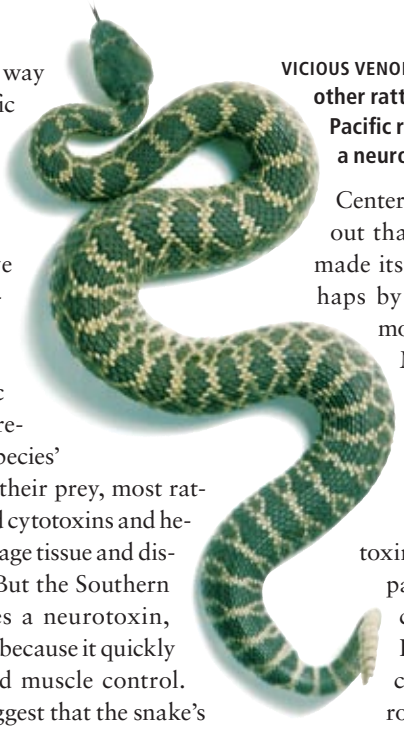
Some climate models predict that equatorial locales have been relatively sheltered from the effects of the planet's natural "greenhouse" phases, but the evidence from *Titanoboa* indicates that during these events, places that were already hot actually got hotter. In fact, shortly after *Titanoboa*'s reign, tropical temperature may have risen so much as to cause widespread heat-related death, although the researchers have not yet found empirical evidence of the effects of such a scorching episode. The findings were published in the February 5 *Nature*.

—*Kate Wong*



FEEDING THE BEAST: *Titanoboa* is believed to have preyed on crocodiles and other vertebrate creatures that shared its tropical habitat.

COURTESY OF JASON BOURQUE (Titanoboa *illustration*); COURTESY OF RAY CARSON *UF Photography* (*vertebrae comparison*)



JERRY YOUNG Getty Images

VICIOUS VENOM: Unlike most other rattlers, the Southern Pacific rattlesnake has a neurotoxic bite.

mans have paved the way for the Southern Pacific to move into new areas. That animal "is adapting to human habitats much like the coyote—whether we like it or not," Johnson remarks.

The rising incidence of supertoxic bite cases could also reflect a change in the species' venom. To predigest their prey, most rattlers produce so-called cytotoxins and hemotoxins, which damage tissue and disrupt blood clotting. But the Southern Pacific also produces a neurotoxin, which is more serious because it quickly affects breathing and muscle control. Anecdotal reports suggest that the snake's venom contains more neurotoxin than it did a few years ago. Richard Dart, director of the Rocky Mountain Poison and Drug Center in Denver, does not rule out that the species could have made its venom more toxic, perhaps by crossbreeding with the more deadly, desert-dwelling Mojave green rattlesnake or by turning on dormant genes developed over time in response to more resistant prey.

The amount of neurotoxin is indeed dramatic compared with the creature's close cousin, the Northern Pacific rattler. This species preys on ground and rock squirrels, which by six weeks of age develop a natural resistance to withstand a full envenomation, suggests research at the University of California, Davis. The Southern variety, however, has enough neurotoxin to overcome any such natural resistance, says biologist William Hayes of Loma Linda University. "Southern Pacifics have no problem getting lots of squirrels."

Still, Hayes does not believe that the rattler's venom has become more toxic. Rather he thinks that people are becoming less tolerant of snake venom, perhaps because of "pollution weakening human lungs and the immune system."

Sean Bush, a treating physician at Loma Linda University Medical Center, says that the Southern Pacific is definitely the "people-biting snake in California" and attributes the increase in incidents to humans encroaching on the animal's coastal and mountain habitat. We are, Bush says, "only now learning how potent and varied rattlesnake venom can be."

Michael Tennesen is a freelance science writer based near Los Angeles.

HEARING

Anti-Loudness Protein

Fans of club music and rock concerts who like the volume cranked up to 11 but want to save their hearing might someday pop a pill rather than plugging their ears. Scientists have pinpointed the biochemical mechanism in ears that works to limit damaging effects of loud sound. When a noise registers in the brain as too loud, the protein nAChR, located on sensory hair cells in the inner ear, kicks in to limit the ability of the hair cells to respond. Mice genetically altered to produce a more potent nAChR could not hear soft sounds, and they suffered less permanent damage to their hearing when scientists blasted 100-decibel noise at their ears. "We know some drugs can modify the protein," says Paul Fuchs of Johns Hopkins University, who published the findings in the January 20 *PLoS Biology*. "But we need to know more about specific amounts" before a sound-protecting drug can be made. So don't toss the earplugs yet. —*Kate Wilcox*

CONCERT FANS could get a pill to save their hearing.

BIOTECH

Sonic Heat for Genes

Heating from sonic waves can turn on genes in the body, demonstrate researchers at the University Victor Segalen Bordeaux in France and their colleagues. Using mice engineered with a bioluminescent gene containing a heat-sensitive stretch of DNA, they focused high-intensity ultrasound pulses on a 0.5-millimeter-wide patch of the mice's legs, heating up that area just below the skin's surface to about 43 degrees Celsius (109 degrees Fahrenheit). Light given off revealed that the gene became active. The technique could help gene therapy, which introduces beneficial DNA into patients. When and where these genes are expressed is paramount, and currently small-molecule drugs and ionizing radiation are employed to switch genes on. But chemicals are not precise, and rays can trigger cancer. The challenge for ultrasound activation, published January 27 by the *Proceedings of the National Academy of Sciences USA*, is safely getting the waves deep enough to reach organs. —*Charles Q. Choi*

EVOLUTION

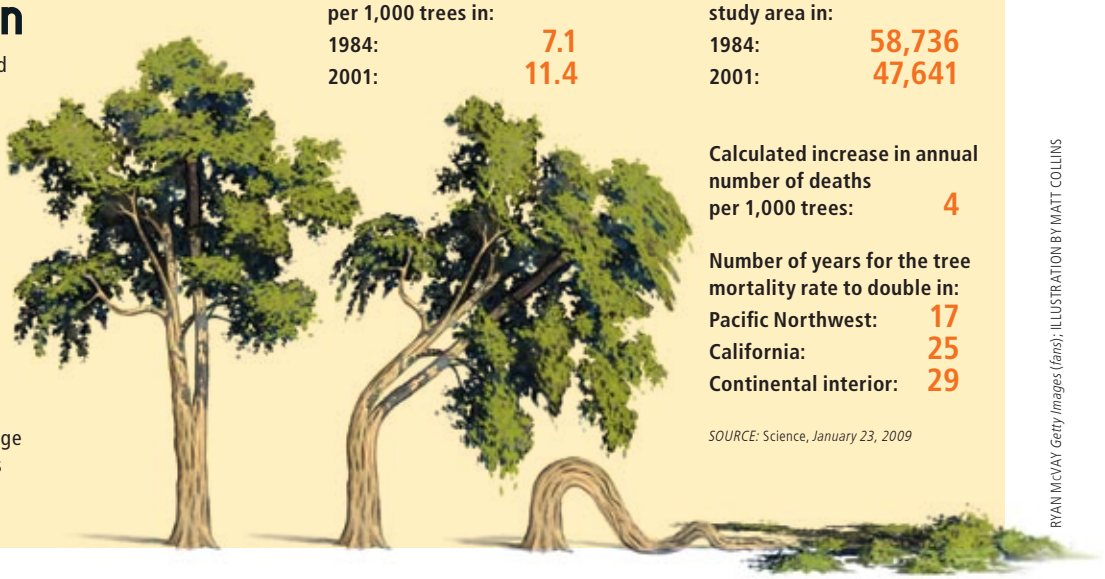
A Beetle's Menu Change

No points for guessing what dung beetles eat. Feces are so rich with nourishing bacteria that 80 or more species of the beetle can live in the same area. Under such intense competition for droppings, one species has gone entirely from scavenging to preying. Through 11 months of recording with infrared cameras in the Peruvian jungle, Trond Larsen of Princeton University and his colleagues found that an eight-millimeter-long nocturnal dung beetle, *Deltochilum valgum*, devours millipedes up to 13 times larger than itself. The beetle kills by wrapping its legs around a victim, wedging its serrated head between the prey's segments and then ripping the body apart. The head of this species is unusually narrow for dung beetles—all the better to burrow inside a corpse to dine on the innards, the researchers note in their report, published online January 21 by *Biology Letters*. —*Charles Q. Choi*

Data Points

Tree Termination

Trees in the western U.S. have died off at an increasing rate over the past few decades, finds Phillip J. van Mantgem of the U.S. Geological Survey and his colleagues. They studied various plots in three regions: the Pacific Northwest, California and the continental interior near the Rocky Mountains. The culprit seems to be locally higher temperatures, which decrease the available water and boost the activity of a bark-damaging fungus. As a result, trees on average are younger and smaller—and less able to hold on to carbon.



Annual number of deaths per 1,000 trees in:
1984: 7.1
2001: 11.4

Number of trees in study area in:
1984: 58,736
2001: 47,641

Calculated increase in annual number of deaths per 1,000 trees: 4

Number of years for the tree mortality rate to double in:
Pacific Northwest: 17
California: 25
Continental interior: 29

SOURCE: Science, *January 23, 2009*

RYAN McVAY *Getty Images (fans)*; ILLUSTRATION BY MATT COLLINS

OCEAN CHEMISTRY

A Calcium Conundrum Explained

Fish excrement could solve a decades-old ocean mystery. After marine plankton die, their calcium carbonate exoskeletons dissolve, making seawater alkaline; however, past studies found that the surface waters are more alkaline than expected from plankton. Now scientists at the University of Exeter in England and their colleagues have determined that calcium carbonate "gut rocks," first found in toadfish intestines about 20 years ago, could account for a dramatic percentage of marine carbonate. Their computer models estimate roughly 812 billion to two trillion kilograms of bony fish swim in the ocean, producing some 110 billion kilograms of calcium carbonate annually. The amount constitutes at least 3 to 15 percent of the total ocean carbonate production and possibly up to 45 percent. Increasing sea temperature and rising carbon dioxide this century could cause fish to produce even more calcium carbonate, the researchers suggest in the January 16 *Science*. —*Charles Q. Choi*

STATISTICS

Math against Profiling

Racial profiling makes little sense, mathematically speaking. Using statistical analyses, William Press of the University of Texas at Austin has found that choosing people to screen based on ethnicity is no more effective than random checks, because nonterrorists vastly outnumber terrorists. The optimal way to screen would be to use "square-root-biased sampling," so that someone nine times as likely to be a terrorist as the average traveler would be screened three times more frequently. This approach would turn up more terrorists in part by avoiding the repeated screening of the same innocent people who fit the profile. But because that strategy would be difficult to implement, Press says that mathematically, the more sensible method is not to profile at all. The study appears in the February 10 *Proceedings of the National Academy of Sciences USA*. —*John Matson*

CHECKED OUT: Profiling is no more effective than random screening.

CONSERVATION

No Nets in the Arctic

Regulators may protect a fishery before anyone even had a chance to cast a net. Melting sea ice and the migration of salmon and other fish farther north make the Arctic region attractive. Because of a lack of studies detailing the impact of commercial fishing in the area, the U.S. North Pacific Fishery Management Council—charged with administering Alaskan waters—voted unanimously on February 5 to close off to any fishing all U.S. waters north of the Bering Strait—some 196,000 square miles of ocean. Studies to determine safe harvesting levels and the impact on indigenous people would be required before any fishing could begin. The proposed prohibition does not mean that the entire Arctic is safe—seven other countries have claims there, including Norway, which has already begun fishing in its waters. —*David Biello*

Read More . . .

News Scan stories with this icon have extended coverage on **www.SciAm.com/apr2009**

In Brief

LEAVES FOR LEAVING ALONE

The vast majority of grasses retain their lifeless leaves, raising the question of why they keep dead weight that could drain their productivity. To find out, scientists at the University of Buenos Aires removed dead leaves from grass in the Argentine pampas, where cattle graze. In the absence of cows, the pruning promoted grass growth, but in the presence of bovines, those grasses were grazed on more so than intact ones, resulting in less growth. The findings, published online January 15 by *Oikos*, suggest that dead leaves act as a defense against herbivores. —*Charles Q. Choi*

QUITTERS' CASH

When the threat of lung cancer and wrinkles doesn't work, maybe $750 will. Cigarette smokers who received money to quit were 2.9 times more likely to break the habit than those who just tried to do it for their health. Participants were offered $100 to complete a smoking cessation program, $250 for quitting within the first six months and $400 for keeping clean an additional six months. Even after that, those who got the cash were 2.6 times more likely to have stayed smoke-free. The study appears in the February 11 *New England Journal of Medicine*. —*Coco Ballantyne*

WHEN ORBITS COLLIDE

In an orbital traffic accident 490 miles above Siberia, a Russian satellite and a commercial satellite owned by U.S. communications firm Iridium collided in February. It wasn't exactly unexpected, given the number of objects in orbit—the past 20 years has seen three other accidents, but they were minor, only producing a few pieces of debris. This latest impact, however, yielded hundreds, and some pieces have drifted down to the altitude of the International Space Station, posing a small but still real hazard. —*John Matson*

CARLOS SERRAO *Getty Images (security guard)*; EMMERICH-WEBB *Getty Images (cows)*; CORBIS *(cigarettes)*

SciAm Perspectives

Healthy Growth for U.S. Farms

Congress and the FDA must upend the nation's agricultural policies to keep its food supply safe

BY THE EDITORS

Agriculture has fueled the eruption of human civilization. Efficiently raised, affordable crops and livestock feed our growing population, and hunger has largely been banished from the developed world as a result. Yet there are reasons to believe that we are beginning to lose control of our great agricultural machine. The security of our food supply is at risk—in ways more noxious than anyone had feared.

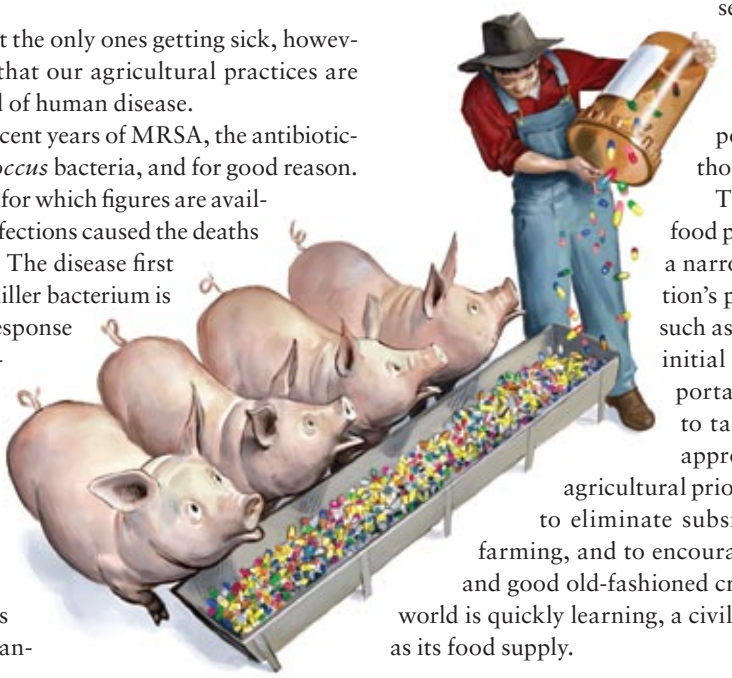
The trouble starts with crops. Orange groves in Florida and California are falling to fast-moving blights with no known cure. Cavendish-variety bananas—the global standard, each genetically identical to the next—will almost certainly be wiped out by emerging infectious disease, just as the Cavendish's predecessor was six decades ago. And as entomologists Diana Cox-Foster and Dennis vanEngelsdorp describe in "Saving the Honeybee," on page 40, a mysterious affliction has ravaged honeybee colonies around the U.S., jeopardizing an agricultural system that is utterly dependent on farmed, traveling hives to pollinate vast swaths of monoculture. The ailment may be in part the result of the stresses imposed on hives by this uniquely modern system.

Plants and animals are not the only ones getting sick, however. New evidence indicates that our agricultural practices are leading directly to the spread of human disease.

Much has been made in recent years of MRSA, the antibiotic-resistant strain of *Staphylococcus* bacteria, and for good reason. In 2005, the most recent year for which figures are available, about 95,000 MRSA infections caused the deaths of nearly 19,000 Americans. The disease first incubated in hospitals—the killer bacterium is an inevitable evolutionary response to the widespread use of antibiotics—but has since found a home in locker rooms, prisons and child care facilities. Now the bacteria have spread to the farm.

Perhaps we should not be surprised. Modern factory farms keep so many animals in such a small space that the animals must be given low doses of antibiotics to shield them from the fetid conditions. The drug-resistant bacteria that emerge have now entered our food supply. The first study to investigate farm-bred MRSA in the U.S.—amazingly, the Food and Drug Administration has shown little interest in testing the nation's livestock for this disease—recently found that 49 percent of pigs and 45 percent of pig workers in the survey harbored the bacteria. Unfortunately, these infections can spread. According to a report published in *Emerging Infectious Diseases,* MRSA from animals is now thought to be responsible for more than 20 percent of all human MRSA cases in the Netherlands.

In April 2008 a high-profile commission of scientists, farmers, doctors and veterinarians recommended that the FDA phase out the nontherapeutic use of antibiotics in farm animal production, to "preserve these drugs to treat sick animals, not healthy ones" in the words of former Kansas governor John Carlin, the commission's chair. The FDA agreed and soon announced that it would ban the use of one widespread antibiotic except for strictly delineated medical purposes. But five days before the ban was set to take effect, the agency quietly reversed its position. Although no official reason was given, the opposition of the powerful farm lobby is widely thought to have played a role.



MATT COLLINS

This is just one example of a food production system that protects a narrow set of interests over the nation's public health. Simple measures such as the reinstatement of the FDA's initial ruling are necessary and important steps. But Congress needs to take a far more comprehensive approach to realign the country's agricultural priorities with its health priorities, to eliminate subsidies that encourage factory farming, and to encourage the growth of polyculture and good old-fashioned crop rotation in the U.S. As the world is quickly learning, a civilization can only be as healthy as its food supply. ■

Sustainable Developments

Needed: A Fiscal Framework

Rather than arguing about the value of taxes or spending, economic planners need to take a systematic long view

BY JEFFREY D. SACHS

The economic debate in the U.S. regarding the fiscal stimulus package centered on "bang for the buck," that is, on whether tax cuts or spending increases would produce more jobs. This limited perspective is very misleading, however: the choice of spending versus taxes should turn first and foremost on the purposes of government, or what economists quaintly call "the allocation of resources." It's silly to debate whether investing in a $100-million bridge creates more jobs than a $100-million tax cut if we need the bridge! The American Society of Civil Engineers has long documented the crumbling state of U.S. infrastructure and the pressing need for $2.2 trillion in investments for our well-being and competitiveness.

Government spending and taxation affect the distribution of income demographically and temporally. America ranks 22nd out of 23 high-income countries in public social outlays as a percentage of national income (ahead only of Ireland) for health, pensions, income support and other social services. Our political discourse tends to focus on the middle class and neglect the poor, whereas our tax and spending policies often benefit the wealthy. As a result, the U.S. has the largest poverty rate, income inequality and per-capita prison population of any high-income nation, as well as the worst health conditions.

The timing of tax cuts and spending increases also affects the well-being of today's generation versus future ones. The U.S. has a chronic fiscal deficit because federal taxation, at around 18 percent of gross national product (GNP), is enough to cover only five types of federal programs: retirement and disability, medical care, veterans' programs, defense and homeland security, and interest on the public debt. All other federal outlays—for education, diplomacy and international assistance, public administration, science and technology, sustainable energy, water and sanitation, roads, broadband, help for the poor—are in effect funded by borrowing. The chronic deficit problem, now at least 5 percent of GNP, will tend to get much worse as the population ages and health care costs rise, until we finally choose to tax ourselves adequately to pay for the government we need and want (and have committed to by law in many entitlements programs).

Temporary deficits can boost the economy in a recession, although temporary income tax cuts and rebates tend to be saved rather than spent. Prolonged deficit spending, however, would impose future burdens. The most obvious will be the need to service the public debts owed to China and other holders of treasury bills—the U.S. is on a path to multiply its already massive international debts. Less obviously, the huge budget deficits will crowd out some private investment spending and exports as the economy recovers. Higher taxes needed to cover the service on that debt will not only squeeze consumption but may also distort the economy through disincentives on saving, work or other activities.

There is a sound method to combine the analytical perspectives of macroeconomic stimulus, resource allocation, income distribution, and generational equity and efficiency. It is called a medium-term fiscal framework, which systematically presents the trade-offs of taxation and spending backed up by formal budget projections for at least five to 10 years and in some budgetary processes more than 50 years. Norway, for instance, takes such a long view in the management of its hydrocarbon wealth.

Higher deficits to increase spending on urgently needed public goods such as infrastructure and on transfers to states and cities to help them tackle the pressing needs of the poor and the unemployed can combine desirable macroeconomic stimulus, efficient resource allocation and redistribution. Over time, we will almost surely have to raise taxes to close the deficits and to cover the long-term costs of government. Right now, however, it's time to think systematically about the long-term role of government and how to pay for it in the coming years. ■

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

SCI AM An extended version of this essay is available at **www.SciAm.com/apr2009**

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

Skeptic

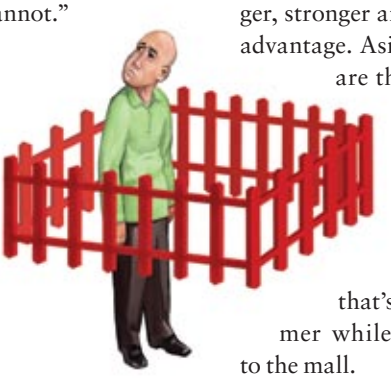
Inside the Outliers

Are successful people primarily the beneficiaries of luck, timing and cultural legacy?

BY MICHAEL SHERMER

What is the difference between Joe Six-Pack, Joe the Plumber and Joe Biden? One is vice president; the other two are not. Why? The answer depends on a host of interactive variables that must be factored into any equation of success: genes, parents, siblings, peers, mentors, practice, drive, culture, timing, legacy and luck. The rub for the scientist is determining the percentage of influence of each variable and its interactions, which requires sophisticated statistical models.

Journalists unconstrained by research protocols churn out self-help books that focus on select variables that interest them. Few do so better than Malcolm Gladwell, and in his new book *Outliers: The Story of Success* (Little, Brown, 2008), the *New Yorker* writer claims that successful people are not "self-made" but instead "are invariably the beneficiaries of hidden advantages and extraordinary opportunities and cultural legacies that allow them to learn and work hard and make sense of the world in ways others cannot."



Bill Gates, for example, may be smart, but Gladwell prefers to emphasize the fact that Gates's wealthy parents sent him to a private school that had a computer club with a teletype time-sharing terminal with a direct link to a mainframe computer in Seattle, and in 1968 this was very unusual. His good fortune to be born in the mid-1950s also meant that Gates came of age when the computer industry was poised to have someone of his experience start a software company.

Similarly, Gladwell says, Mozart's father was a composer who mentored the young Wolfgang into greatness from age six until his early 20s, when his compositions morphed from pleasantly melodious into masterful. The Beatles' lucky break came in Hamburg, Germany, where they were able to log in more than 1,200 live performances and thereby meet the well-known 10,000-hour rule for perfecting a profession. Elite hockey players are disproportionately born in January, February and March (40 percent compared with the expected birthrate, which in most studies hovers around 25 percent) because the birthday cutoff date when they were youngsters first hitting the ice was January 1, and players born early in the year were slightly bigger, stronger and faster, giving them an advantage. Asian student wunderkinds are the product of "the tradition of wet-rice agriculture" that must be practiced year-round and that requires "the highest emphasis on effort and hard work," and that's why they study all summer while American students go to the mall.

PHOTOGRAPH BY BRAD SWONETZ; ILLUSTRATION BY MATT COLLINS

Such prodigies and geniuses, Gladwell says, "are products of history and community, of opportunity and legacy. Their success is not exceptional or mysterious. It is grounded in a web of advantages and inheritances, some deserved, some not, some earned, some just plain lucky—but all critical to making them who they are."

Well, yes and no. As Frank J. Sulloway, author of the comprehensive study of success *Born to Rebel* (Pantheon, 1996), told me: "Creative people are not just sitting around waiting for opportunities to come to them. They create their own opportunities. Charles Darwin was already planning a voyage of discovery to the Canary Islands, for example, when the position on the *Beagle* opened up. If the Beatles hadn't gone to Hamburg, they would have gotten their 10,000 hours somewhere else. What distinguishes Gates is that he has a really interesting creative mind, and he would have had that mind even without a computer terminal at his private school and hence would likely have found alternative ways to access programming tools." And of course, Leopold Mozart's son was a child prodigy and musical genius, not merely the beneficiary of cultural legacy.

Even the 10,000-hour rule isn't just about skill mastery. According to Dean Keith Simonton, author of *Origins of Genius* (Oxford University Press, 1999), success includes a Darwinian process of variation and selection. Creative geniuses generate a massive variety of ideas from which they select only those most likely to survive and reproduce. The best predictor of winning a Nobel Prize in science, for example, is the rate of journal citation. As Simonton notes, "empirical studies have repeatedly shown that the single most powerful predictor of eminence within any creative domain is the sheer number of influential products an individual has given the world."

Genius is as genius does. ■

Michael Shermer is publisher of Skeptic *magazine (www.skeptic.com) and author of* The Mind of the Market.

Anti Gravity

Boobs at Work

Unacceptable ways to waste time on the job

BY STEVE MIRSKY

The Internet is indeed a wonder of our age. Why, just last night, while watching the DVD of *Inherit the Wind* (it's Darwin's bicentennial birthday week as I write), I was able to simultaskically discover that Fredric March and Florence Eldridge, who play Matthew and Sarah Brady, were married in real life and often performed together in movies and on stage. (*Inherit the Wind,* by the way, is actually a bombastically bad movie. But it's fun.) My research was over before Matthew, a character based on William Jennings Bryan, could finish one of his long-winded speeches.

Of course, easy access to such tantalizing data has the potential for misuse. Which clearly was the case at the National Science Foundation (NSF), where an employee spent a significant amount of time at work perusing pornography. At least that's the official report—if he claimed he was investigating grant applications from researchers investigating human reproduction, well, it didn't fly.

The incident, and some other porn-related surfing by a handful of other NSF employees, was revealed in the foundation's semiannual report issued by its inspector general. The primary porn culprit lost his job based on the misuse of time and resources that was estimated to have wasted some $58,000. And the foundation installed filters, just as countless other employers in the U.S. have done when faced with exactly this same kind of abuse of company resources. NSF employees looking for dirty pictures will henceforth have to be content with medical journals.

With the audit having been published, the NSF got back to work supporting and promoting scientific research. That is until Iowa senator Charles Grassley noticed the foundation's report. The waste wasn't just playing computer solitaire or Freecell, or instant-messaging friends or searching the Internet for movie trivia. This was porn.

At first, Grassley was mildly intrigued. Perhaps he gently caressed the hard copy of the internal audit, its creamy white pages glistening under the gentle light of a desk lamp. As the senator read about a government-funded employee viewing lascivious images in the workplace, his heart must have pounded. When he reached the mention of the $58,000, his pulse no doubt shot up and he might have softly moaned. Then he and the NSF semiannual report became one. And his outrage exploded like a volcano that could no longer contain the roiling molten lava within. "The semiannual report," Grassley said in a press release, "raises real questions about how the National Science Foundation manages its resources, and Congress ought to demand a full accounting before it gives the agency another $3 billion in the stimulus bill."

Grassley then joined with Senators Barbara Mikulski of Maryland and Richard Shelby of Alabama to share their indignation at the waste of the taxpayers' money. (Again, the porn addict was already gone, despite the fact that a federally funded worker who loses only $58,000 in this economy should probably be nominated for employee of the month.) The three senators introduced an amendment to the roughly $800-billion stimulus bill that would freeze $3 million in operating funds for the NSF unless the foundation took further steps to ensure that no pornography ever sullied its computer screens again.

Much of this information came to me from a source at the Senate Finance Committee, of which Grassley is the ranking member. I made repeated requests of the source to assure me that the senatorial threats against the NSF would not cost any scientists their funding. The answers were inconclusive and then stopped altogether. While this tempest in a D-cup made news, the former NSF employee was joined by 598,000 other Americans who lost their jobs in January.

Meanwhile the senators, who dithered for weeks over the stimulus bill, still have in their company their esteemed Louisiana colleague David Vitter, a known client of an infamous Washington brothel. Vitter—who faces an election challenge from a porn star named Stormy Daniels—once tried to waste $100,000 in a federal funding earmark for the Louisiana Family Forum. That organization campaigns against the teaching of evolution. So that in the future we can have new and better versions of *Inherit the Wind.* ■

PHOTOGRAPH BY FLYNN LARSEN; ILLUSTRATION BY MATT COLLINS

Saving the HONEYBEE

The mysterious ailment called colony collapse disorder has wiped out large numbers of the bees that pollinate a third of our crops. The causes turn out to be surprisingly complex, but solutions are emerging

• By Diana Cox-Foster and Dennis vanEngelsdorp

KEY CONCEPTS

- Millions of beehives worldwide have emptied out as honeybees mysteriously disappear, putting at risk nearly 100 crops that require pollination.
- Research is pointing to a complex disease in which combinations of factors, including farming practices, make bees vulnerable to viruses.
- Taking extra care with hive hygiene seems to aid prevention. And research into antiviral drugs could lead to pharmaceutical solutions.

—The Editors

Dave Hackenberg makes a living moving honeybees. Up and down the East Coast and often coast to coast, Hackenberg trucks his beehives from field to field to pollinate crops as diverse as Florida melons, Pennsylvania apples, Maine blueberries and California almonds.

As he has done for the past 42 years, in the fall of 2006 Hackenberg migrated with his family and his bees from their central Pennsylvania summer home to their winter locale in central Florida. The insects had just finished their pollination duties on blooming Pennsylvanian pumpkin fields and were now to catch the last of the Floridian Spanish needle nectar flow. When Hackenberg checked on his pollinators, the colonies were "boiling over" with bees, as he put it. But when he came back a month later, he was horrified. Many of the remaining colonies had lost large numbers of workers, and only the young workers and the queen remained and seemed healthy. More than half of the 3,000 hives were completely devoid of bees. But no dead bees were in sight. "It was like a ghost town," Hackenberg said when he called us seeking an explanation for the mysterious disappearance.

We and other researchers soon formed an interdisciplinary working team that by December 2006 had described the phenomenon and later named it colony collapse disorder, or CCD. Curiously, Hackenberg's colonies stopped dying the following spring, but by that time only 800 of his original 3,000 colonies had survived. As Hackenberg spoke to colleagues around the nation, it became apparent that he was not alone. And a survey our team conducted in the spring of 2007 revealed that a fourth of U.S. beekeepers had suffered similar losses and that more than 30 percent of all colonies had died. The next winter the die-off resumed and expanded, hitting 36 percent of U.S. beekeepers. Reports of large losses also surfaced from Australia, Brazil, Canada, China, Europe and other regions. More recent data are not available yet, but some beekeepers say they have seen their colonies collapse this winter, too.

The bee loss has raised alarms because one third of the world's agricultural production depends on the European honeybee, *Apis mellifera*—the kind universally adopted by beekeepers in Western countries. Large, monoculture farms require intense pollination activity for short periods of the year, a role that other pollinators such as wild bees and bats cannot fill. Only *A. mellifera* can deploy armies of pollinators at almost any time of the year, wherever the weather is mild enough and there are flowers to visit.

Our collaboration has ruled out many potential causes for CCD and found many possible contributing factors. But no single culprit has been identified. Bees suffering from CCD tend

CHARLES KREBS

DAVE HACKENBERG was the first beekeeper to alert U.S. entomologists to the inexplicable disappearance of worker bees, a sign of what is now known as colony collapse disorder, in the fall of 2006. By the end of the winter, more than 60 percent of his 3,000-odd colonies were dead; nationwide the loss was 30 percent.

The bees were all sick, but each colony seemed to suffer from a different combination of diseases.

to be infested with multiple pathogens, including a newly discovered virus, but these infections seem secondary or opportunistic—much the way pneumonia kills a patient with AIDS. The picture now emerging is of a complex condition that can be triggered by different combinations of causes. There may be no easy remedy to CCD. It may require taking better care of the environment and making long-term changes to our beekeeping and agricultural practices.

Even before colony collapse, honeybees had suffered from a number of ailments that reduced their populations. The number of managed honeybee colonies in 2006 was about 2.4 million, less than half what it was in 1949. But beekeepers could not recall seeing such dramatic winter losses as occurred in 2007 and 2008. Although CCD probably will not cause honeybees to go extinct, it could push many beekeepers out of business. If beekeepers' skills and know-how become a rarity as a result, then even if CCD is eventually overcome, nearly 100 of our crops could be left without pollinators—and large-scale production of certain crops could become impossible. We would still have corn, wheat, potatoes and rice. But many fruits and vegetables we consume routinely today—such as apples, blueberries, broccoli and almonds—could become the food of kings.

Silent Bloom

When Hackenberg initially told us of his vanishing bees, our first thought was varroa mites. These aggressive parasites were largely responsible for a 45 percent drop in the number of managed bee colonies worldwide between 1987 (when they were first introduced in the U.S.) and 2006. Mature varroa females feed on hemolymph, the bees' blood. The mites also carry viruses and actively inhibit the hosts' immune responses. Hackenberg, like most expert beekeepers, already had long experience fighting mites, and he was adamant that, this time, the symptoms were different.

One of us (vanEngelsdorp) performed autopsies on Hackenberg's remaining insects and found symptoms never observed before, such as scar tissue in the internal organs. Initial tests also detected some of the usual suspects in bee disease. In the gut contents we found spores of nosema, single-celled fungal parasites that can cause bee dysentery. The spore counts in these and in subsequent samples, however, were not high enough to explain the losses. Molecular analysis of Hackenberg's bees, performed by the other of us (Cox-Foster), also revealed surprising levels of viral infections of various known types. But no single pathogen found in the insects could explain the scale of the disappearance.

In other words, the bees were all sick, but each colony seemed to suffer from a different combination of diseases. We hypothesized that something had compromised the bees' immune system, making them susceptible to any number of infections that healthy colonies would normally fend off. And Hackenberg was right: the prime suspects, varroa mites, were not present in numbers significant enough to explain the sudden die-off.

In the spring of 2007 our task force began detailed, countrywide surveys of all aspects of colony management, interviewing operators who had encountered CCD as well as those who had not. These and subsequent investigations ruled out several potential causes. No single beekeeping management method could be blamed. Large commercial beekeepers were as likely to suffer from high losses as were small operations or hobbyists. The symptoms affected stationary beekeepers as well as migratory ones. Even some organic beekeepers were affected.

DAVID YELLEN

As media reports of the die-offs surfaced, the public also started expressing concern. Many were eager to share their ideas as to the underlying cause. Some of these proposals—such as blaming CCD on radiation from cell phones—originated from poorly designed studies. Other hypotheses were untestable at best, such as claims that the bees were being abducted by aliens.

One theory favored by many concerned citizens was that bees could have been poisoned by pollen from genetically modified crops, specifically the so-called Bt crops. Bt crops contain a gene for an insecticidal toxin produced by the bacterium *Bacillus thuringiensis*. When pest caterpillars feed on crops producing these toxins, they die. But already before the onset of CCD, research had shown that the Bt toxin becomes activated only in the guts of caterpillars, mosquitoes and some beetles. The digestive tracts of honeybees and of many other insects do not allow Bt to work.

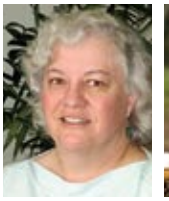
Another popular theory, and a more credible one, blamed synthetic poisons. The two main suspects were acaricides—chemicals beekeepers use to keep mites in check—and pesticides, either in the environment or in the very field crops the bees were pollinating. By 2006 newer types of pesticides had replaced older varieties. One type in particular, the neonicotinoids, had been blamed by beekeepers in France and elsewhere for harming insect pollinators. This class of insecticides mimics the effects of nicotine—a natural defense that tobacco plants deploy against leaf-eating pests—and is more toxic to insects than it is to vertebrates. But neonicotinoids also enter the pollen and nectar of the plant—not just the leaves—thus potentially affecting pollinators. Previous research had demonstrated that neonicotinoids decrease honeybees' ability to remember how to get back to their hive, a sign that they could be a contributor to CCD.

We and other experts also suspected that the bees' natural defenses might be undermined by poor nutrition. Honeybees—and wild pollinators, too—no longer have the same number or variety of flowers available to them because we humans have tried to "neaten" our environments. We have, for example, planted huge expanses of crops without weedy, flower-filled borders or fencerows. We maintain large green lawns free of any "weeds" such as clover or dandelions. Even our roadsides and parks reflect our desire to keep things neat and weed-free. But to bees and other pollinators, green lawns look like deserts. The diets of honeybees that pollinate large acreages of one crop may lack important nutrients, compared with those of pollinators that feed from multiple sources, as would be typical of the natural environment. Beekeepers have attempted to manage these concerns by developing protein supplements to feed colonies—although the supplements have not on their own prevented CCD.

All-Out Effort

Our task force focused its investigation on these two broad areas—pesticides and nutrition—in

[THE AUTHORS]

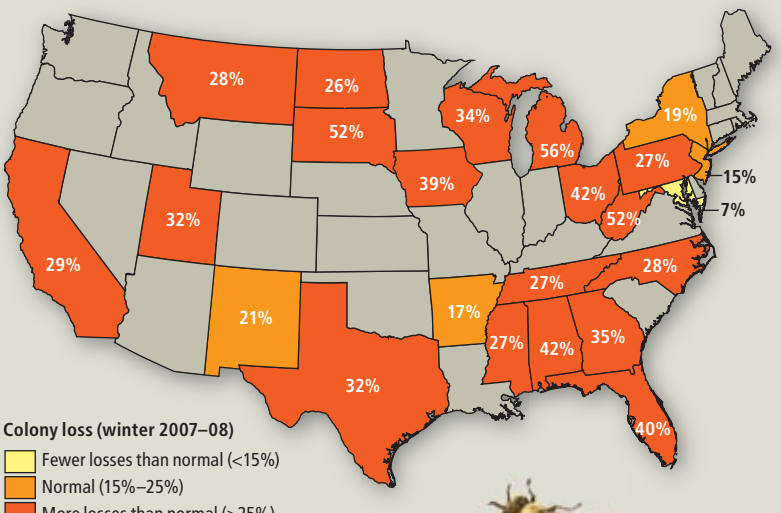


Diana Cox-Foster is professor of entomology at Pennsylvania State University and co-director of the colony collapse disorder working team, made up of experts from government and academia. Her research focuses on host-pathogen interactions. Cox-Foster traces her affinity for honeybees back to her great-grandmother, who was a commercial beekeeper in Colorado in the early 1900s. Dennis vanEngelsdorp's passion for bees began in an undergraduate beekeeping course at the University of Guelph in Ontario. It has carried him through several appointments on West Indian projects and to his current dual position as acting state apiarist for the Commonwealth of Pennsylvania and senior extension associate with the entomology department at Penn State.

[THE EXTENT OF COLONY LOSS]

A COUNTRYWIDE SCOURGE

Colony collapse disorder (CCD) returned for a second year in the winter of 2007–08. A survey of beekeepers in the spring of 2008 asked how many colonies failed to make it through that winter. Nationwide 36 percent of colonies were lost (compared with a typical winter decline of 15 to 25 percent); 60 percent of those losses were attributed to CCD. Most states for which enough data were available were severely hit. Large losses were also reported in Australia, Brazil, Canada, China and Europe.



Number of honeybee colonies estimated to have died in the U.S. over the winter of 2007–08:

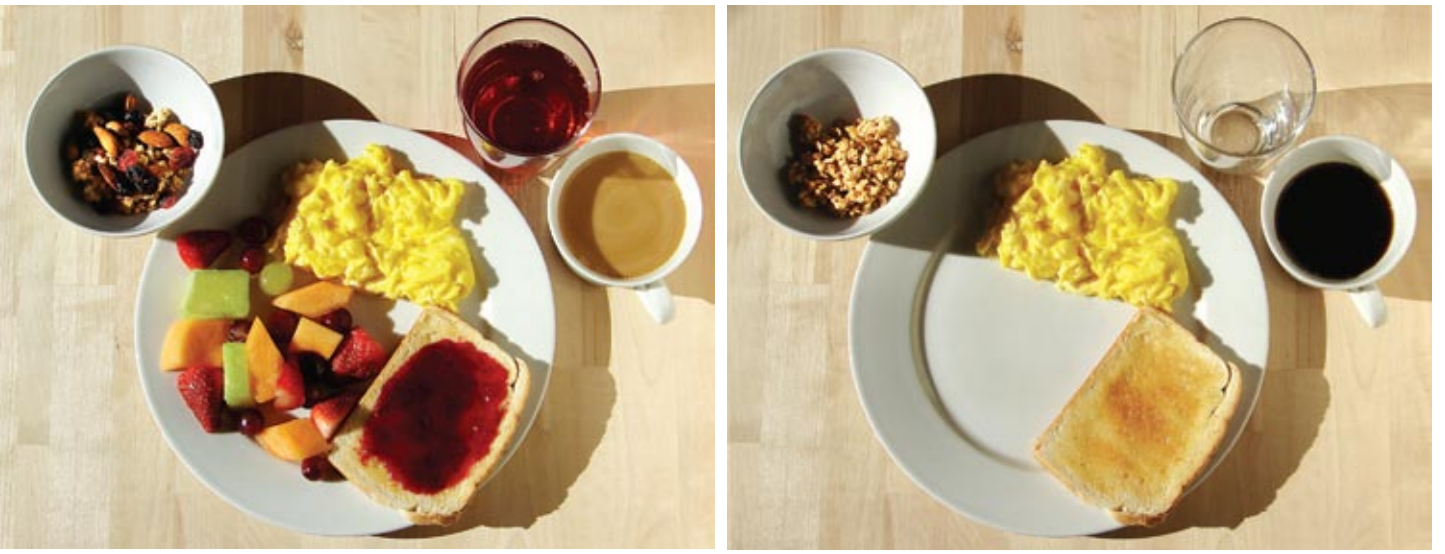
750,000 to 1 million

Some beekeepers reported losing up to

90%

of their colonies.

COURTESY OF ELLEN HARASIMOWICZ (*vanEngelsdorp*); COURTESY OF ELLEN JOHNSON (*Cox-Foster*); JEN CHRISTIANSEN, SOURCE: "A SURVEY OF HONEY BEE COLONY LOSSES IN THE U.S., FALL 2007 TO SPRING 2008," BY D. VANENGELSDORP ET AL., IN *PLOS ONE*, VOL. 3, ISSUE 12, ARTICLE E4071; DECEMBER 2008 (*map*)



WITHOUT HONEYBEES, many foods included in the breakfast at the left would become too rare for most people to afford. Shortages would affect an array of fruits, as well as jams and jellies, almonds and even milk, because dairies use alfalfa (which needs pollinators) as a protein-rich feed for dairy cows.

addition to the other obvious possibility, that a new or newly mutated pathogen could be causing CCD. Tests of our three hypotheses required collecting samples—lots of samples. We joined Jeff Pettis of the U.S. Department of Agriculture lab in Beltsville, Md., to conduct this monumental effort that involved long days, lots of miles on the road and the challenge of collecting enough material to share with the entire team. With no dead bees to study, we decided to collect live bees from apiaries in the midst of collapse, based on the premise that survivors would harbor the disease in its early stages. Bees were collected in alcohol for varroa and nosema counts. Bees, pollen and honeycomb wax were frozen on dry ice and rapidly shipped back to labs in Pennsylvania or Maryland to be stored in ultracold freezers and preserved for molecular and chemical analyses.

Some samples were sent to our colleague David Tarpy of North Carolina State University, who measured protein content. Tarpy found no notable difference between apiaries that had CCD and others that were seemingly healthy. His results suggested that nutritional state—on its own—could not explain CCD.

Much more startling was the outcome of our team's search for pesticides, for which we enlisted the help of Pennsylvania State University researchers Maryann Frazier, Jim Frazier and Chris Mullin and of Roger Simonds, a chemist at the USDA lab in Gastonia, N.C. (By coincidence, Simonds happens to be a beekeeper himself.) His broad-spectrum analysis, sensitive to insecticides, herbicides and fungicides, found more than 170 different chemicals. Most stored-pollen samples contained five or more different compounds, and some contained as many as 35. But although both the levels and the diversity of chemicals are of concern, none is likely to be the sole smoking gun behind CCD: healthy colonies sometimes have higher levels of some chemicals than colonies suffering from CCD.

No neonicotinoids were found in the original samples. But these or other pesticides cannot yet be exonerated. Honeybee colonies are dynamic, and our initial sampling was not—we took samples only once. It remains possible, if not likely, that bees afflicted by CCD were harmed by a chemical or mixture of chemicals not evident at the time we collected samples.

Our attempts to identify a new infectious disease—or a new strain of an old one—that could be at the root of CCD initially looked as if they would go nowhere fast. None of the known bacterial, fungal or viral diseases of bees could account for the CCD losses, so we had no clue what to look for.

Then Cox-Foster, with Ian Lipkin's group at Columbia University (and with help from biotech company 454 Life Sciences in Branford, Conn.), turned to a sophisticated microbe-hunting method called metagenomics. In this technique, nucleic acids (DNA and RNA) are collected from an environment containing many different organisms. The genetic material is all blended together and minced into pieces short enough that their sequences of code "letters" can be deciphered. In ordinary gene sequencing, researchers would then use computer software to put the pieces back together and reconstruct the genome of the original organism. But in metagenomics, the genes belong to different organisms, and so sequenc-

FAST FACTS

- There are an estimated 900 to 1,000 commercial beekeepers in the U.S., managing 2.4 million colonies.
- Nearly 100 kinds of crops require pollination by honeybees. The annual value of bees' work is $14 billion in the U.S. and $215 billion worldwide.
- Every February virtually all movable U.S. hives are taken to California to pollinate almond trees.
- Even before CCD, in certain regions of China bees had completely disappeared, possibly because of pesticide use, forcing orchard owners to pollinate pear trees by hand.

ing produces a snapshot of the sequences in a collection of organisms, including microscopic ones, in an ecosystem. Metagenomics has been used to survey environments such as seawater and soil, revealing a surprising diversity of microorganisms. But it can also be applied to detecting microorganisms hosted by a larger organism, living either as collaborators (in symbiosis) or as infections.

Naturally, most gene sequences in our samples were from the bees themselves. But those were easy to filter out because, fortunately, the honeybee genome had just been sequenced. Nonbee sequences were then matched to genetic sequences belonging to known organisms. Researchers with expertise in molecular analysis of organisms—including bacteria, fungi, parasites and viruses—joined our team to identify potential culprits.

The *CSI*-style investigation greatly expanded our general knowledge of honeybees. First, it showed that all samples (CCD and healthy) had eight different bacteria that had been described in two previous studies from other parts of the world. These findings strongly suggest that those bacteria may be symbionts, perhaps serving an essential role in bee biology such as aiding in digestion. We also found two nosema species, two other fungi and several bee viruses.

But one bee virus stood out, as it had never been identified in the U.S.: the Israeli acute paralysis virus, or IAPV. This pathogen was first described in 2004 by Ilan Sela of the Hebrew University of Jerusalem in the course of an effort to find out why bees were dying with paralytic seizures. In our initial sampling, IAPV was found in almost all—though not all—colonies with CCD symptoms and in only one operation that was not suffering from CCD. But such strong correlation was not proof that IAPV caused the disease. For example, CCD could have just made the bees exceptionally vulnerable to IAPV infection.

Case Closed?

From subsequent work on IAPV, we know that at least three different strains of the virus exist and that two of them infect bees in the U.S. One of the strains most likely arrived in colonies flown in from Australia in 2005 after the U.S. government lifted a ban on honeybee importation that had been in effect since 1922. (The almond industry lobbied to lift the ban to prevent a critical shortage of pollinators at blossom time.) The other strain probably showed up earlier and is quite different. Where that one came from is unknown; it may have been introduced by way of importation of royal jelly (a nutrient bees secrete to feed their larvae) or a pollen supplement, or it may have hitchhiked into the country on newly introduced pests of bees. The data also suggest that IAPV has existed in bees in other parts of the world for a while, developing into many different strains and possibly changing rapidly.

In an effort to settle the issue of IAPV's role, Cox-Foster experimented with healthy honeybees that had no previous exposure to the virus. Her team placed hives filled with bees into greenhouses and fed the insects sugary water laden with IAPV. Sure enough, the infection mimicked some symptoms of CCD. Within one or two weeks of exposure, the bees began to die, twitching with paralytic seizures on the ground. The bees were not dying near the hives, just as one would expect in CCD. So those findings seemed to support the notion that IAPV can

To bees and other pollinators, green lawns look like deserts.

BUMBLEBEE

Wild Pollinators Are Ailing, Too

Honeybees are not the only pollinators to have suffered population drops in recent years. A National Research Council (NRC) report in 2006 pointed to downhill trends in certain species of North American wild pollinators, including some insects but also bats and hummingbirds. These species may be suffering from some of the same man-made afflictions that make honeybees vulnerable to CCD, such as introduced diseases, pesticide poisoning and impoverished habitats, says the study's lead author, entomologist May Berenbaum of the University of Illinois.

The western bumblebee, for example, has disappeared from a region stretching from central California to British Columbia, probably killed off by *Nosema bombi,* a single-celled fungus microorganism, according to work by entomologist Robbin Thorp of the University of California, Davis. The fungus may have spread to the western bumblebees from European bumblebees that U.S. farmers have imported to assist in the pollination of tomatoes and other crops in greenhouses, he says.

A more recent study published in the January *Biological Conservation* looked at historical data from Illinois and found that four bumblebee species disappeared there between 1940 and 1960—a period that coincided with large-scale agricultural intensification in the state, with consequent loss of prairie, forest and wetland habitats.

Declines in a few species of pollinating bats and hummingbirds—to the point that some bats are at risk of extinction—might relate to habitat changes. Many of them overwinter in Mexico, and biologists are urging the preservation of "nectar corridors," where the animals can find flowers along their migration routes.

But biologists can monitor only so many pollinator species (an estimated 200,000 exist worldwide), and not much is known about the state of health for most of them, the NRC report warned. Several Web-based collaborations call on citizen-scientists' help. Volunteers take pictures of pollinators and submit them to the Web sites, where researchers identify species and take note of where they were seen.

In 2008 the U.S. Congress for the first time modified its agricultural policy to include pollination protection measures, such as setting aside conservation land where wildflowers can grow and provide nectar. "That was a real landmark," Berenbaum says.

—Davide Castelvecchi, staff writer

DERRICK DITCHBURN www.derelanatureimn.ca

[THE SEARCH FOR THE CAUSE]

Many Suspects, No Convictions Yet

Researchers have looked into virtually all aspects of honeybee life in search of the culprit behind colony collapse. The work has exonerated some suspects and has pointed to possible combinations of factors that can cause or contribute to CCD.

SUSPECT: CHEMICALS

As many as 170 different synthetic chemicals have been found in beehives of both sick and healthy colonies, with some samples of pollen stored in cells containing as many as 35 types. Although no single chemical seems to be the cause of CCD, pesticides may weaken bees' health.

SUSPECT: VARROA MITES

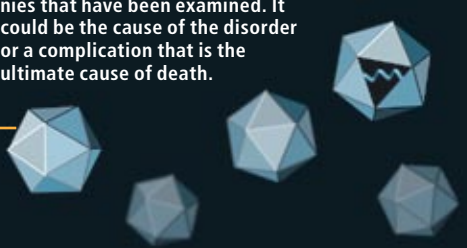
This mite, seen below sucking blood from a pupa (an intermediate stage between larva and adult), is the honeybee's most common and destructive pest. But collapsing colonies did not have significant mite infestations.

SUSPECT: PARASITES

Some of the bees in collapsing colonies were infected by single-celled fungi, such as *Nosema apis* (*below*), which invades the intestinal tract and causes dysentery. But levels of infection were too low to be lethal on their own.

SUSPECT: ISRAELI ACUTE PARALYSIS VIRUS

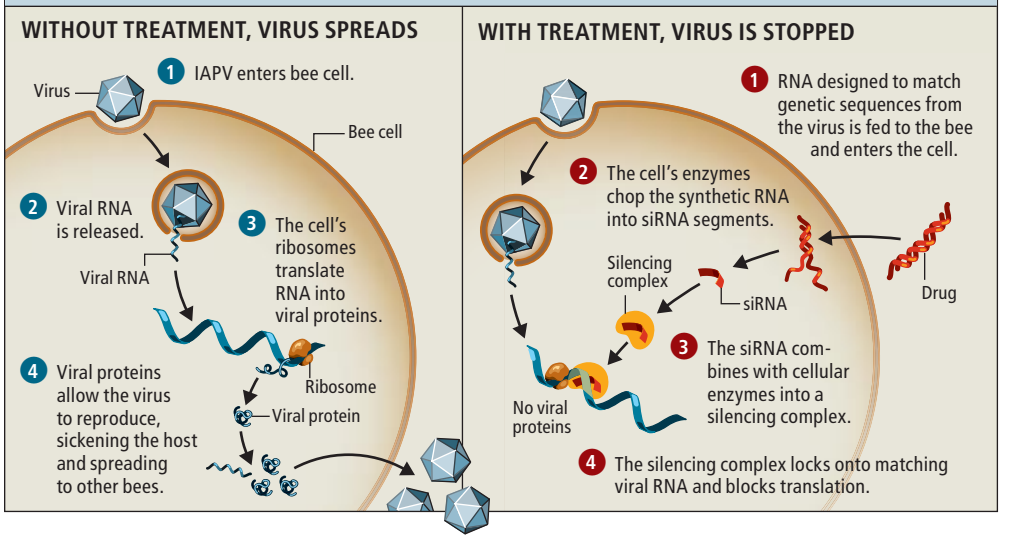
IAVP has been shown to produce symptoms similar to CCD's and has occurred in most affected colonies that have been examined. It could be the cause of the disorder or a complication that is the ultimate cause of death.



[ONE POTENTIAL SOLUTION]

A BEE MEDICINE?

A Miami-based biotechnology start-up company called Beeologics is developing an antiviral drug that exploits an ancient immune mechanism called RNA interference. Cells in most animals and plants use short-interfering RNA (siRNA) segments to inhibit the formation of viral proteins; here siRNA designed to target IAPV would be fed to colonies as part of double-stranded RNA mixed into a syrup.



cause CCD or at least contribute to the problem.

Additional sampling efforts by several groups showed, however, that IAPV was widespread in the U.S. and that not all infected colonies had symptoms of CCD, implying either that IAPV alone cannot cause the disease or that some bees are predisposed to be IAPV-resistant. In particular, a joint study the two of us initiated in 2007 with the USDA has tracked colonies owned by three traveling beekeepers and has observed colonies that were infected with IAPV without collapsing. Some of those colonies have later been able to rid themselves of the virus.

The growing consensus among researchers is that multiple factors—such as poor nutrition and exposure to pesticides—can interact to weaken colonies and make them susceptible to a virus-mediated collapse. In the case of our experiments in greenhouses, the stress of being confined to a relatively small space could have been enough to make colonies succumb to IAPV and die with CCD-like symptoms. More recent results from long-term monitoring have identified other unexpected factors for increased colony loss, including the fungicide chlorothalonil. Research is now focused on understanding how these factors relate to colony collapse.

A vaccine or cure for bee viruses and IAPV specifically would be desirable. Unfortunately, vaccines will not work on honeybees, because the invertebrate immune system does not generate the kind of protection against specific agents that vaccines induce in humans and other mammals. But researchers are beginning to pursue other approaches, such as one based on the new technique of RNA interference [*see box above*], which blocks a virus from reproducing inside a bee's cells. A longer-term solution will be to identify and breed virus-resistant honeybees. Such an effort could take years, though, perhaps too many to avoid having a large number of beekeepers go out of business.

Meanwhile many beekeepers have had some success at preventing colony loss by redoubling their efforts at improving their colonies' diets, keeping infections and parasites such as varroa and nosema in check, and practicing good hygiene. In particular, research has shown that sterilizing old beehive frames with gamma rays before reusing them cuts down the risk of colony collapse. And simple changes in agricultural practices such as breaking up monocultures with hedgerows could help restore balance in honeybees' diets, while providing nourishment to wild pollinators as well.

Humankind needs to act quickly to ensure that the ancient pact between flowers and pollinators stays intact, to safeguard our food supply and to protect our environment for generations to come. These efforts will ensure that bees continue to provide pollination and that our diets remain rich in the fruits and vegetables we now take for granted. ■

MORE WAYS TO FIGHT BACK

Restoring a balance to the habitat of pollinators might improve their general well-being and help prevent colony collapse. Large stretches of single crops or residential lawns could be broken up with more "weedy" meadows and hedgerows. Plants flowering at different times of the year could then provide more variety in pollinators' diets and support them year-round.

Sterilizing used beehives with DNA-destroying gamma rays before reusing them for a new colony cuts down the risk of CCD recurrence, possibly because it kills microorganisms that contribute to the disease.

Research on the impact of pesticides on pollinators usually focuses on possible lethal effects. More research is needed on whether certain pesticides can put insects under stress, even if the chemicals do not kill them outright.

MORE TO EXPLORE

Censors of the Genome. Nelson C. Lau and David P. Bartel in *Scientific American,* Vol. 289, No. 2, pages 34–41; August 2003.

Status of Pollinators in North America. National Research Council. National Academies Press, 2007.

Decline of Bumble Bees (*Bombus*) in the North American Midwest. Jennifer C. Grixti, Lisa T. Wong, Sydney A. Cameron and Colin Favret in *Biological Conservation,* Vol. 142, No. 1, pages 75–84; January 2009.

The Mid-Atlantic Apiculture Research and Extension Consortium: **http://maarec.cas.psu.edu**

The Xerces Society for Invertebrate Conservation: **www.xerces.org**

OPPOSITE PAGE: ANDREW SYRED *Photo Researchers, Inc.* (*varroa mite*); CENTRAL SCIENCE LABORATORY *Photo Researchers, Inc.* (*varroa mite on larva*); HEIDI & HANS-JURGEN KOCH *Minden Pictures* (*bee with pollen*); COURTESY OF BARTON SMITH, JR. *USDA Bee Research Laboratory* (*Nosema sp.*); JEN CHRISTIANSEN (*Israeli acute paralysis virus*); OPPOSITE PAGE AND THIS PAGE: JEN CHRISTIANSEN (*illustrations*)

Does DARK ENERGY Really Exist?

Maybe not.
The observations that led astronomers to deduce its existence could have another explanation: that our galaxy lies at the center of a giant cosmic void

By Timothy Clifton and Pedro G. Ferreira

KEY CONCEPTS

- The universe appears to be expanding at an accelerating rate, implying the existence of a strange new form of energy—dark energy. The problem: no one is sure what dark energy is.
- Cosmologists may not actually need to invoke exotic forms of energy. If we live in an emptier-than-average region of space, then the cosmic expansion rate varies with position, which could be mistaken for a variation in time, or acceleration.
- A giant void strikes most cosmologists as highly unlikely but so for that matter does dark energy. Observations over the coming years will differentiate between the two possibilities.

—The Editors

In science, the grandest revolutions are often triggered by the smallest discrepancies. In the 16th century, based on what struck many of his contemporaries as the esoteric minutiae of celestial motions, Copernicus suggested that Earth was not, in fact, at the center of the universe. In our own era, another revolution began to unfold 11 years ago with the discovery of the accelerating universe. A tiny deviation in the brightness of exploding stars led astronomers to conclude that they had no idea what 70 percent of the cosmos consists of. All they could tell was that space is filled with a substance unlike any other—one that pushes along the expansion of the universe rather than holding it back. This substance became known as dark energy.

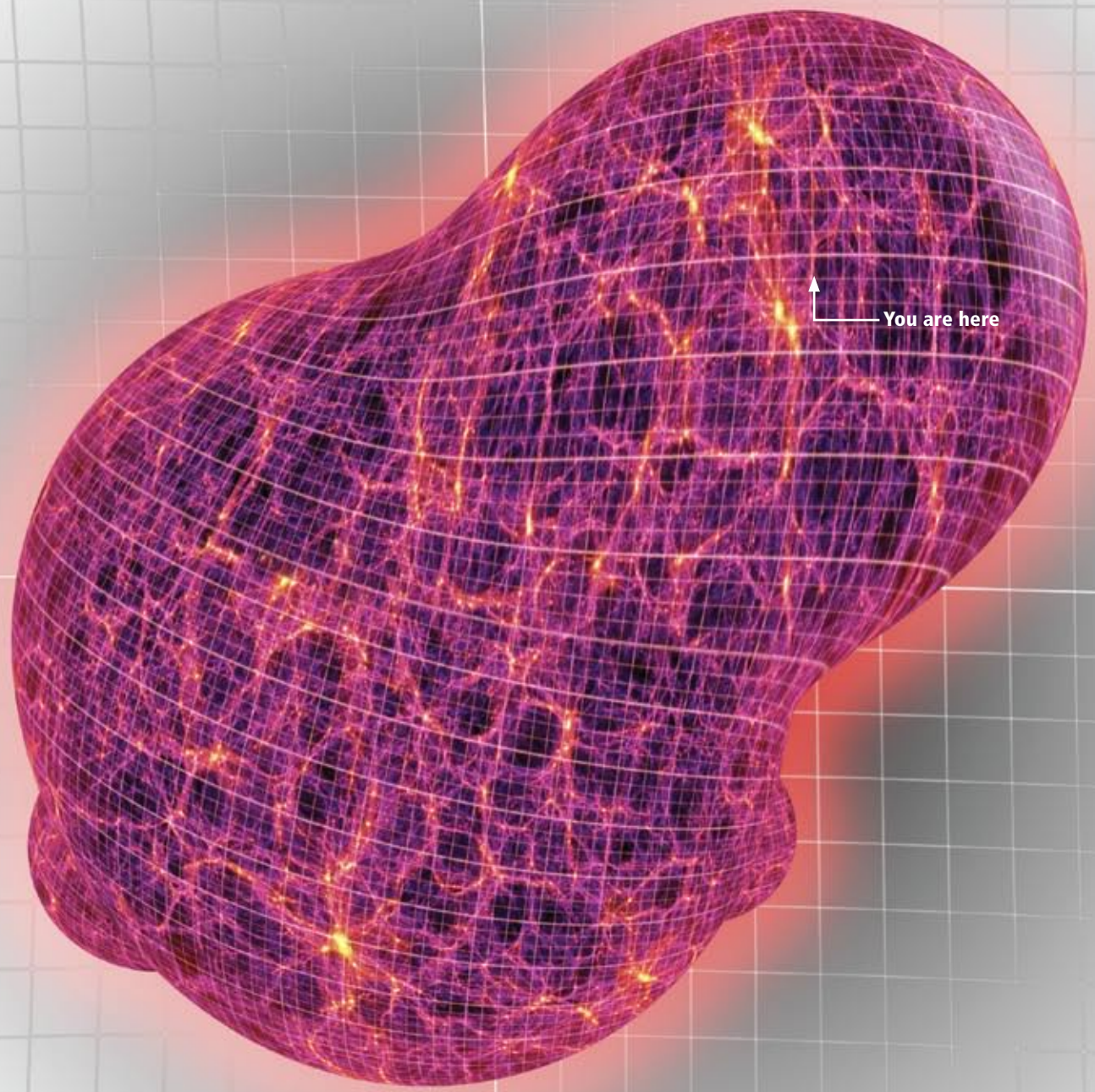
It is now over a decade later, and the existence of dark energy is still so puzzling that some cosmologists are revisiting the fundamental postulates that led them to deduce its existence in the first place. One of these is the product of that earlier revolution: the Copernican principle, that Earth is not in a central or otherwise special position in the universe. If we discard this basic principle, a surprisingly different picture of what could account for the observations emerges.

Most of us are very familiar with the idea that our planet is nothing more than a tiny speck orbiting a typical star, somewhere near the edge of an otherwise unnoteworthy galaxy. In the midst of a universe populated by billions of galaxies that stretch out to our cosmic horizon, we are led to believe that there is nothing special or unique about our location. But what is the evidence for this cosmic humility? And how would we be able to tell if we *were* in a special place? Astronomers typically gloss over these questions, assuming our own typicality sufficiently obvious to warrant no further discussion. To entertain the notion that we may, in fact, have a special location in the universe is, for many, unthinkable. Nevertheless, that is exactly what some small groups of physicists around the world have recently been considering.

Ironically, assuming ourselves to be insignificant has granted cosmologists great explanatory power. It has allowed us to extrapolate from what we see in our own cosmic neighborhood to the universe at large. Huge efforts have been made in constructing state-of-the-art models of the universe based on the cosmological principle—a generalization of the Copernican principle that states that at any moment in time all points and directions in space look the same. Combined with our modern understanding of space, time and matter, the cosmological principle implies that space is expanding, that the universe is getting cooler and that it is populated by relics from its hot beginning—predictions that are all borne out by observations.

DON DIXON

UNEVEN EXPANSION OF SPACE, caused by variations in the density of matter on an epic scale, could produce the effects that astronomers conventionally attribute to dark energy.



Astronomers find, for example, that the light from distant galaxies is redder than that of nearby galaxies. This phenomenon, known as redshift, is neatly explained as a stretching of light waves by the expansion of space. Also, microwave detectors reveal an almost perfectly smooth curtain of radiation emanating from very early times: the cosmic microwave background, a relic of the primordial fireball. It is fair to say that these successes are in part a result of our own humility—the less we assume about our own significance, the more we can say about the universe.

Darkness Closes In

So why rock the boat? If the cosmological principle is so successful, why should we question it? The trouble is that recent astronomical observations have been producing some very strange results. Over the past decade astronomers have found that for a given redshift, distant supernova explosions look dimmer than expected. Redshift measures the amount that space has expanded. By measuring how much the light from distant supernovae has redshifted, cosmologists can then infer how much smaller the universe was at the time of the explosion as compared with its size today. The larger the redshift, the smaller the universe was when the supernova occurred and hence the more the universe has expanded between then and now.

The observed brightness of a supernova provides a measure of its distance from us, which in turn reveals how much time has elapsed since it occurred. If a supernova with a given redshift looks dimmer than expected, then that supernova must be farther away than astronomers thought. Its light has taken longer to reach us, and hence the universe must have taken longer to grow to its current size [*see box on opposite page*]. Consequently, the expansion rate of the universe must have been slower in the past than previously expected. In fact, the distant supernovae are dim enough that the expansion of the universe must have accelerated to have caught up with its current expansion rate [see "Surveying Spacetime with Supernovae," by Craig J. Hogan, Robert P. Kirshner and Nicholas B. Suntzeff; Scientific American, January 1999].

This accelerating expansion is the big surprise that fired the current revolution in cosmology. Matter in the universe should tug at the fabric of spacetime, slowing down the expansion, but the supernova data suggest otherwise. If cosmologists accept the cosmological principle and assume that this acceleration happens everywhere, we are led to the conclusion that the universe must be permeated by an exotic form of energy, dark energy, that exerts a repulsive force.

Nothing meeting the description of dark energy appears in physicists' Standard Model of fundamental particles and forces. It is a substance that has not as yet been measured directly, has properties unlike anything we have ever seen and has an energy density some 10^{120} times less than we may have naively expected. Physicists have ideas for what it might be, but they remain speculative [see "The Quintessential Universe," by Jeremiah P. Ostriker and Paul J. Steinhardt; Scientific American, January 2001]. In short, we are very much in the dark about dark energy. Researchers are working on a number of ambitious and expensive ground- and space-based missions to find and characterize dark energy, whatever it may be. To many, it is the greatest challenge facing modern cosmology.

A Lighter Alternative

Confronted with something so strange and seemingly so improbable, some researchers are revisiting the reasoning that led them to it. One of the primary assumptions they are questioning is whether we live in a representative part of the universe. Could the evidence for dark energy be accounted for in other ways if we were to do away with the cosmological principle?

In the conventional picture, we talk about the expansion of the universe on the whole. It is very much like when we talk about a balloon blowing up: we discuss how big the entire balloon gets, not how much each individual patch of the balloon inflates. But we all have had experience with those annoying party balloons that inflate unevenly. One ring stretches quickly, and the end takes a while to catch up. In an alternative view of the universe, one that jettisons the cosmological principle, space, too, expands unevenly. A more complex picture of the cosmos emerges.

Consider the following scenario, first suggested by George Ellis, Charles Hellaby and Nazeem Mustapha, all at the University of Cape Town in South Africa, and subsequently followed up by Marie-Noëlle Célérier of the Paris-Meudon Observatory in France. Suppose that the expansion rate is decelerating everywhere, as matter tugs on spacetime and slows it down. Suppose, further, that we live in a gargantuan cosmic void—not a completely empty region, but one in which the average density of matter is only a half or maybe a third of the density elsewhere. The emptier a patch of space is, the less matter it contains to

POLISH SCHOOL/BRIDGEMAN ART LIBRARY/GETTY IMAGES (*Copernicus*)

COPERNICUS'S LEGACY

The Copernican principle holds that Earth does not occupy a special place in the universe. The universe has a uniform density (homogeneity) and looks the same in every direction (isotropy).

Though powerful, the principle applies only on scales much larger than a galaxy. After all, if the cosmos were completely uniform, it would be a thin gruel of atoms rather than a constellation of galaxies. Also, the principle applies in space but not in time. We live in a special era—long enough after the big bang that complex life can form but not so long that stars have all died off.

Copernicus is commonly associated with a dethroning of humanity from any position of importance. But as historian Dennis Danielson of the University of British Columbia argues, although pre-Copernican Europeans placed Earth at the center of the universe, they did not consider the center a position of importance but quite the opposite—as Galileo put it, "the sump where the universe's filth and ephemera collect."

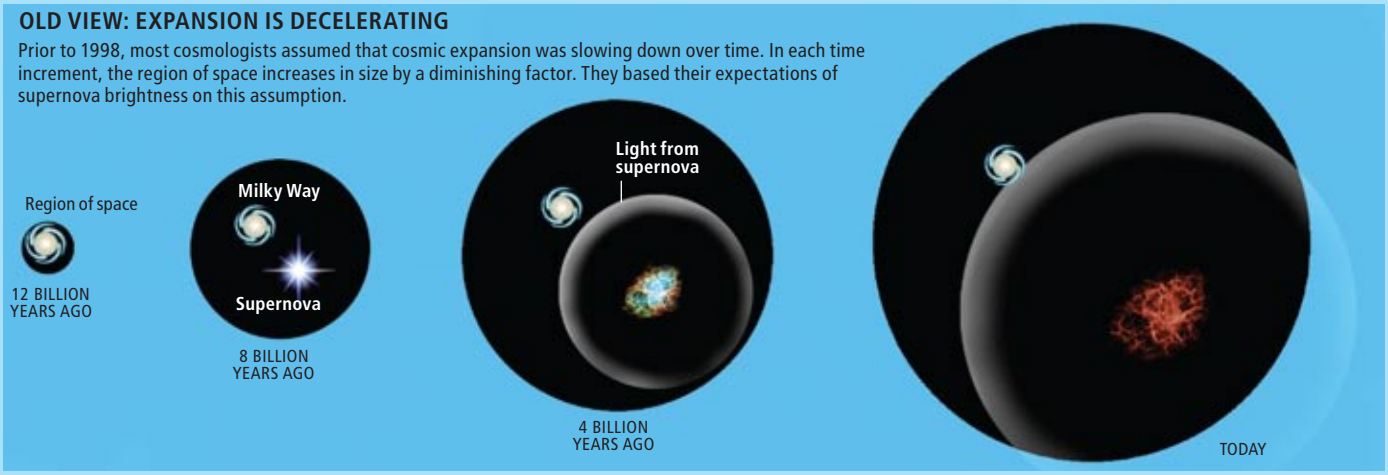
[THE BASICS]

Three Ways of Expanding a Universe

Astronomers have found that distant supernovae explosions are dimmer than expected. To see what this discovery means for cosmic expansion, consider a region of space that encompasses a supernova and our Milky Way galaxy. Over time this region gets bigger as the fabric of space stretches like a rubber sheet. The supernova goes off when the universe is about half its current size (which occurs at different times depending on whether the expansion is decelerating or accelerating). Light from the explosion spreads out and eventually reaches us on the outskirts of the Milky Way.

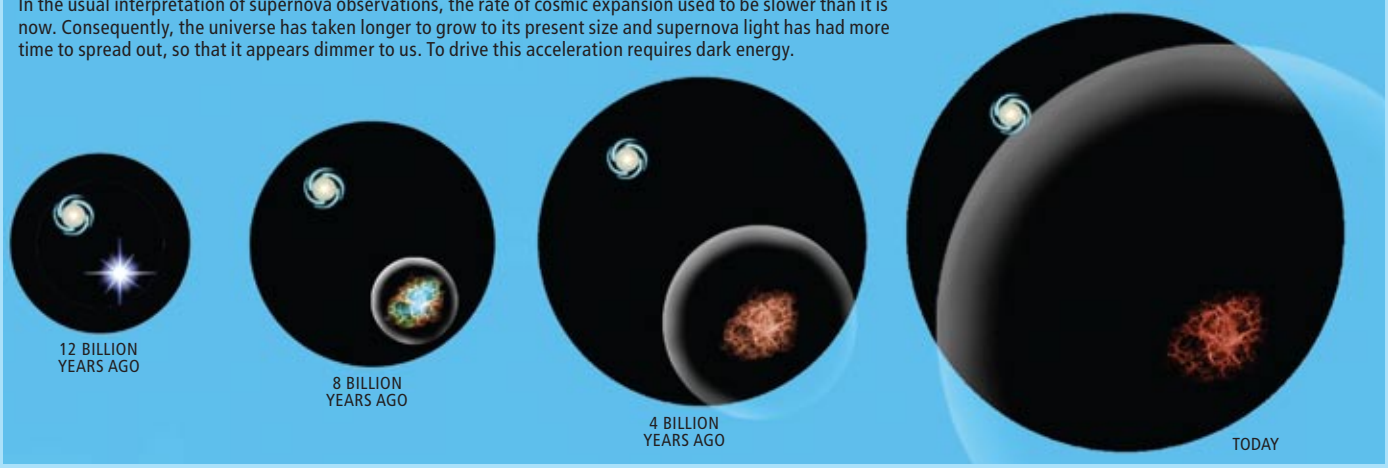
OLD VIEW: EXPANSION IS DECELERATING

Prior to 1998, most cosmologists assumed that cosmic expansion was slowing down over time. In each time increment, the region of space increases in size by a diminishing factor. They based their expectations of supernova brightness on this assumption.



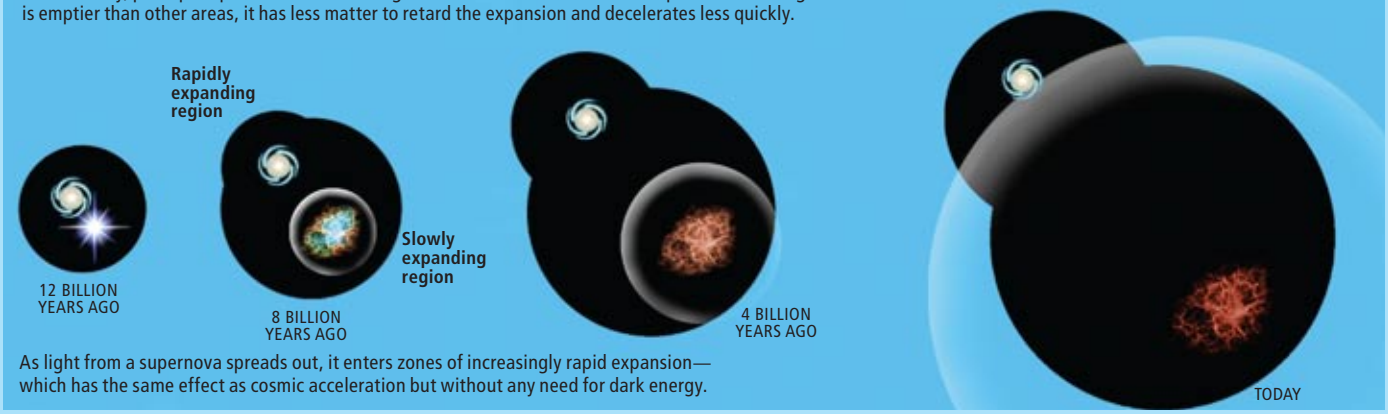
SCENARIO 1: EXPANSION IS ACCELERATING

In the usual interpretation of supernova observations, the rate of cosmic expansion used to be slower than it is now. Consequently, the universe has taken longer to grow to its present size and supernova light has had more time to spread out, so that it appears dimmer to us. To drive this acceleration requires dark energy.



SCENARIO 2: UNIVERSE IS INHOMOGENEOUS

Alternatively, perhaps expansion is decelerating but at different rates in different places. If our neighborhood is emptier than other areas, it has less matter to retard the expansion and decelerates less quickly.



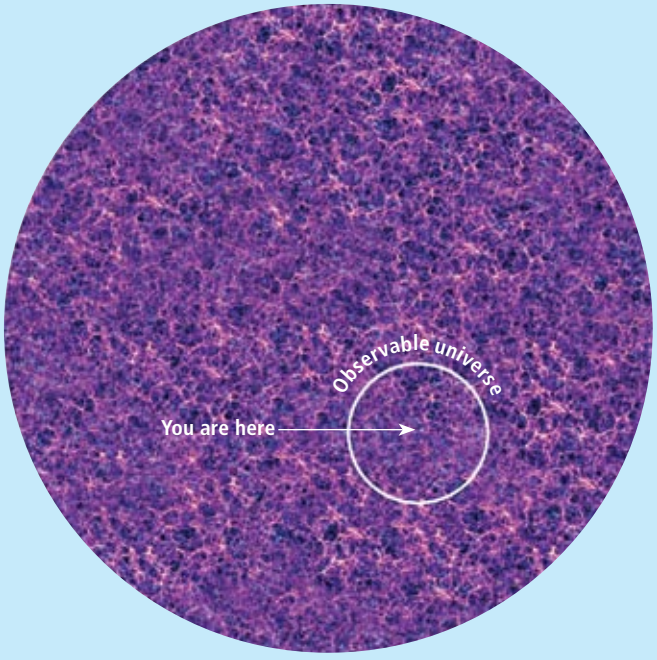
As light from a supernova spreads out, it enters zones of increasingly rapid expansion—which has the same effect as cosmic acceleration but without any need for dark energy.

DON DIXON

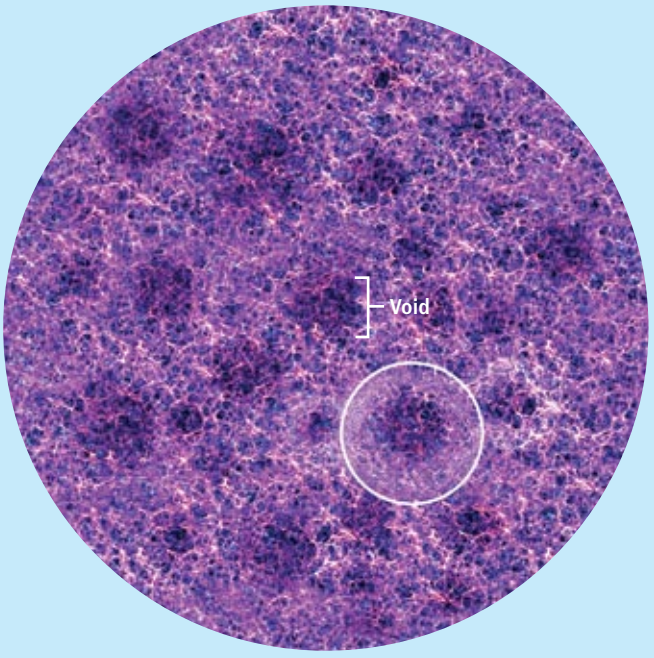
[THE BIG PICTURE]

A Special Place for Us

In his *Hitchhiker's Guide to the Galaxy* series of novels, Douglas Adams imagines a torture device that drives people insane by showing them the utter insignificance of their place in the universe. One would-be victim emerges unscathed when it turns out that the universe does, in fact, revolve around him. In a case of life imitating art, many cosmologists are investigating whether our planet indeed has a special place within the grand scheme of things.



HOMOGENEOUS UNIVERSE: OUR LOCATION IS TYPICAL
In the standard view, galaxies are lined up in a spidery pattern, but overall space looks much the same everywhere, and Earth's position is nothing special.



INHOMOGENEOUS UNIVERSE: OUR LOCATION IS SPECIAL
Alternatively, the density of matter could vary on large scales, and Earth may lie at or near the center of a relatively less dense region, or void.

slow down the expansion of space; accordingly, the local expansion rate is faster within the void than it is elsewhere. The expansion rate is fastest at the very center of the void and diminishes toward the edge, where the higher-density exterior begins to make itself felt. At any given time different parts of space will expand at different rates, like the unevenly inflated party balloon.

Now imagine supernovae exploding in different parts of this inhomogeneous universe, some close to the center of the void, others nearer the edge and some outside the void. If we are near the center of the void and a supernova is farther out, space expands faster in our vicinity than it does at the location of the supernova. As light from the supernova travels toward us, it passes through regions that are expanding at ever faster rates. Each region stretches the light by a certain amount as it passes though, and the cumulative effect produces the redshift we observe. Light traveling a given distance is redshifted by less than it would be if the whole universe expanded at our local rate. Conversely, to achieve a certain redshift in such a universe, the light has to travel a greater distance than it would in a uniformly expanding universe, in which case the supernova has to be farther away and therefore appear dimmer.

Another way to put it is that a variation of expansion rate with position mimics a variation in time. In this way, cosmologists can explain the unexpected supernova observations without invoking dark energy. For such an alternative explanation to work, we would have to live in a void of truly cosmic proportions. The supernova observations extend out to billions of light-years, a significant fraction of the entire observable universe. A void would have to be of similar size. Enormous by (almost) anyone's standards.

A Far-fetched Possibility

So how outlandish is this cosmic void? At first glance, very. It would seem to fly in the face of the cosmic microwave background, which is uniform to one part in 100,000, not to mention the apparently uniform distribution of galaxies [see "Reading the Blueprints of Creation," by Michael A. Strauss; SCIENTIFIC AMERICAN, Feb-

DON DIXON

ruary 2004]. On closer inspection, however, this evidence may not be so conclusive.

The uniformity of the relic radiation merely requires the universe to look nearly the same in every direction. If a void is roughly spherical and if we lie reasonably close to its center, these observations do not necessarily preclude it. In addition, the cosmic microwave background has some anomalous features that could potentially be explained by large-scale inhomogeneity [*see box on next page*].

As for the galaxy distribution, existing surveys do not extend far enough to rule out a void of the size that would mimic dark energy. They identify smaller voids, filaments of matter and other structures hundreds of millions of light-years in size, but the putative void is an order of magnitude larger. A lively debate is now under way in astronomy as to whether galaxy surveys corroborate the cosmological principle. A recent analysis by David Hogg of New York University and his collaborators indicates that the largest structures in the universe are about 200 million light-years in size; on larger scales, matter appears smoothly distributed, in accordance with the principle. But Francesco Sylos Labini of the Enrico Fermi Center in Rome and his colleagues argue that the largest structures discovered so far are limited only by the size of the galaxy surveys that found them. Still larger structures might stretch beyond the scope of the surveys.

By analogy, suppose you had a map showing a region 10 miles wide, on which a road stretched from one side to the other. It would be a mistake to conclude that the longest possible road is 10 miles long. To determine the length of the longest road, you would need a map that clearly showed the end points of all roads, so that you would know their full extent. Similarly, astronomers need a galaxy survey that is larger than the biggest structures in the universe if they are to prove the cosmological principle. Whether surveys are big enough yet is the subject of the debate.

For theorists, too, a colossal void is difficult to stomach. All available evidence suggests that galaxies and larger structures such as filaments and voids grew from microscopic quantum seeds that cosmic expansion enlarged to astronomical proportions, and cosmological theory makes firm predictions for how many structures should exist with a certain size. The larger a structure is, the rarer it should be. The probability of a void big enough to mimic dark energy is less than one part in 10^{100}. Giant voids may well exist out there, but the chance of our finding one in our observable universe would seem to be tiny.

Still, there is a possible loophole. In the early 1990s one of the authors of what is now the standard model of the early universe, Andrei Linde, and his collaborators at Stanford University showed that although giant voids are rare, they expand faster early on and come to dominate the volume of the universe. The probability of observers finding themselves in such a structure may not be so tiny after all. This result shows that the cosmological principle (that we do not live in a special place) is not always the same thing as the principle of mediocrity (that we are typical observers). One can, it seems, be both typical and live in a special place.

Testing the Void

What observations could tell whether the expansion of the universe is driven by dark energy or whether we are living in a special place, such as at the center of a giant void? To test for the presence of a void, cosmologists need a working model of how space, time and matter should behave in its vicinity. Just such a model was formulated in 1933 by Abbé Georges Lemaître, independently rediscovered a year later by Richard Tolman and further developed after World War II by Hermann Bondi. The universe they envisaged had expansion rates that depended not only on time but also on distance from a specific point, just as we now hypothesize.

With the Lemaître-Tolman-Bondi model in hand, cosmologists can make predictions for a

NO A-VOIDING IT

Although a cosmic void mimics dark energy, the match is not exact. Upcoming observations will look for telltale differences.

- Additional supernova observations will pin down the expansion rate and check whether it varies with position, as a void model predicts.
- Galaxy clusters reflect light and, in effect, let us view our cosmic neighborhood in the mirror. If we live in a void, we should be able to see it.
- Galaxies and galaxy clusters evolve at a pace that depends on the expansion rate at their location and therefore on the presence of a void.
- Neutrinos left over from the primordial universe could reveal a void.

SUPERNOVA 1994D (*arrow*) and similar explosions are used as tracers of cosmic expansion.

NASA/ESA, HUBBLE KEY PROJECT TEAM AND HIGH-Z SUPERNOVA SEARCH TEAM

range of observable quantities. To begin, consider the supernovae that first led to the inference of dark energy. The more supernovae that astronomers observe, the more accurately they can reconstruct the expansion history of the universe. Strictly speaking, these observations cannot ever rule out the void model, because cosmologists could re-create any set of supernova data by choosing a suitably shaped void. Yet for a void to be completely indistinguishable from dark energy, it would have to have some very strange properties indeed.

The reason is that the putative accelerating expansion occurs right up to the present moment. For a void to mimic it exactly, the expansion rate must decrease sharply away from us and in every direction. Therefore, the density of matter and energy must increase sharply away from us in every direction. The density profile must look like an upside-down witch's hat, the tip of which corresponds to where we live. Such a profile would go against all our experience of what structures in the universe look like: they are usually smooth, not pointy. Even worse, Ali Vanderveld and Éanna Flanagan, both then at Cornell University, showed that the tip of the hat, where we live, would have to be a singularity, like the ultradense region at the center of a black hole.

If, however, the void has a more realistic, smooth density profile, then a distinct observational signature presents itself. Smooth voids still produce observations that could be mistaken for acceleration, but their lack of pointyness means that they do not reproduce *exactly* the same results as dark energy. In particular, the apparent rate of acceleration varies with redshift in a telltale way. In a paper with Kate Land, then at the University of Oxford, we showed that several hundred new supernovae, on top of the few hundred we currently have, should be enough to settle the issue. Supernova-observing missions stand a very good chance of achieving this goal soon.

Supernovae are not the only observables available. Jeremy Goodman of Princeton University suggested another possible test in 1995 using the microwave background radiation. At the time, the best evidence for dark energy had not yet emerged, and Goodman was not seeking an explanation for any unexplained phenomena but proof of the Copernican principle itself. His idea was to use distant clusters of galaxies as mirrors to look at the universe from different positions, like a celestial dressing room. Galaxy clusters reflect a small fraction of the microwave radiation that hits them. By carefully measuring the spectrum of this radiation, cosmologists could infer some aspects of what the universe would look like if viewed from one of them. If a shift of viewpoint changed how the universe looked, it would be powerful evidence for a void or a similar structure.

Two teams of cosmologists recently put this idea to the test. Robert Caldwell of Dartmouth

[THE AUTHORS]

Timothy Clifton and Pedro G. Ferreira are cosmologists at the University of Oxford. Both study the physics of the early universe and potential modifications to Einstein's general theory of relativity. Clifton, a keen oenophile, says his true interest in life is Burgundy wine. Ferreira is the author of a popular-level astronomy book, *The State of the Universe,* runs a program for artists in residence at Oxford, and participates in various projects to support science education in Africa.

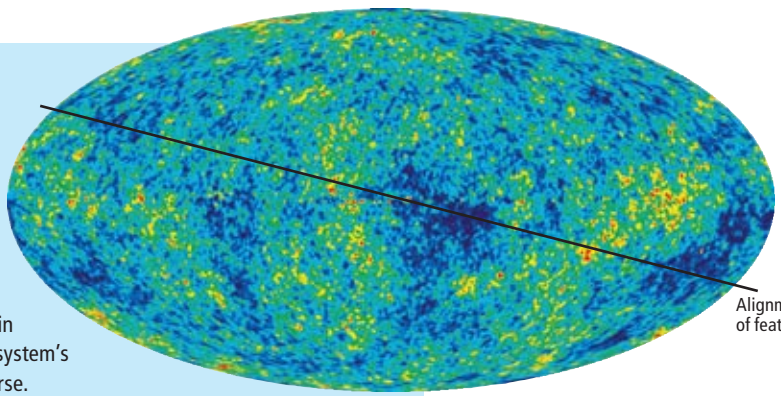
Surrender to the Void

Most suggestions that we live in a cosmic void place us at its center, but what if we lived away from the middle? The universe would then look slightly lopsided. Håvard Alnes and Morad Amarzguioui, both at the University of Oslo, have shown that the cosmic microwave background radiation would look slightly hotter in one direction than in the other. Such an asymmetry, called a dipole, has indeed been observed in the microwave background. It is usually attributed to our solar system's motion through space but could also be a sign of a lumpy universe.

Furthermore, small fluctuations in the microwave background appear to align in a specific direction—dubbed the "axis of evil" by João Magueijo and Kate Land, both then at Imperial College London [see "Is the Universe Out of Tune?" by Glenn D. Starkman and Dominik J. Schwarz; SCIENTIFIC AMERICAN, August 2005]. This alignment picks out a preferred direction in the sky, which, though hard to imagine in a Copernican universe, might be explained in terms of our displacement from the center of a void. A preferred direction would also have other effects, such as large-scale coherent motions of galaxies and galaxy clusters. Several researchers have claimed to have detected such a "dark flow," but it remains controversial.

Although it is tempting to attribute these anomalies to a giant void, this explanation does not really hold together. For a start, these effects each pick out different directions. Furthermore, the strength of the cosmic dipole would suggest that we are only about 50 million light-years from the center, which is only a very small fraction of the total size of the putative void.

—T.C. and P.G.F.



AXIS OF EVIL, an alignment of features in the cosmic microwave background radiation, could be a sign that we live in an inhomogeneous universe.

COURTESY OF TIMOTHY CLIFTON (*Clifton*); COURTESY OF GISA WESZKALNYS (*Ferreira*); NASA/WMAP SCIENCE TEAM (*map*)

College and Albert Stebbins of the Fermi National Accelerator Laboratory in Batavia, Ill., studied precise measurements of distortions in the microwave background, and Juan García-Bellido of the University of Madrid and Troels Haugbølle of the University of Aarhus in Denmark looked at individual clusters directly. Neither group detected a void; the best the researchers could do was to narrow down the properties that such a void could have. The Planck Surveyor satellite, scheduled for launch this month, should be able to place stronger limits on the void properties and maybe rule out a void altogether.

A third approach, advocated by Bruce Bassett, Chris Clarkson and Teresa Lu, all at the University of Cape Town, is to make independent measurements of the expansion rate at different locations. Astronomers usually measure expansion rates in terms of redshift, which is the cumulative effect of the expansion of all regions of space between a celestial body and us. By lumping all these regions together, redshift cannot distinguish a variation of expansion rate in space from a variation in time. It would be better to measure the expansion rate at specific spatial locations, separating out the effects of expansion at other locations. That is a difficult proposition, though, and has yet to be done. One possibility is to observe how structures form at different places. The formation and evolution of galaxies and galaxy clusters depend, in large part, on the local rate of expansion. By studying these objects at different locations and accounting for other effects that play a role in their evolution, astronomers may be able to map out subtle differences in expansion rate.

A Not So Special Place

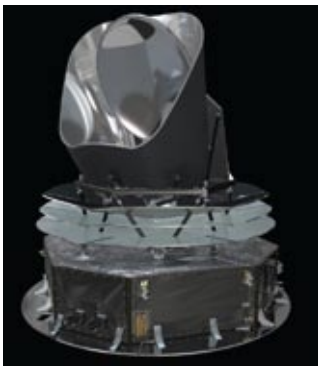
The possibility that we live in the middle of a giant cosmic void is an extreme rejection of the cosmological principle, but there are gentler possibilities. The universe could obey the cosmological principle on large scales, but the smaller voids and filaments that galaxy surveys have discovered might collectively mimic the effects of dark energy. Tirthabir Biswas and Alessio Notari, both at McGill University, as well as Valerio Marra and his collaborators, then at the University of Padua in Italy and the University of Chicago, have studied this idea. In their models, the universe looks like Swiss cheese—uniform on the whole but riddled with holes. Consequently, the expansion rate varies slightly from place to place. Rays of light emitted by distant supernovae travel through a multitude of these small voids before reaching us, and the variations in the expansion rate tweak their brightness and redshift. So far, however, the idea does not look very promising. One of us (Clifton), together with Joseph Zuntz of Oxford, recently showed that reproducing the effects of dark energy would take lots of voids of very low density, distributed in a special way.

Another possibility is that dark energy is an artifact of the mathematical approximations that cosmologists routinely use. To calculate the cosmic expansion rate, we typically count up how much matter a region of space contains, divide by the volume of the region and arrive at the average energy density. We then insert this average density into Einstein's equations for gravity and determine the averaged expansion rate of the universe. Although the density varies from place to place, we treat this scatter as small fluctuations about the overall average.

The problem is that solving Einstein's equations for an averaged matter distribution is *not* the same as solving for the real matter distribution and then averaging the resulting geometry. In other words, we average and then solve, when really we should solve and then average.

Solving the full set of equations for anything even vaguely approximating the real universe is unthinkably difficult, and so most of us resort to the simpler route. Thomas Buchert of the University of Lyon in France has taken up the task of determining how good an approximation it really is. He has introduced an extra set of terms into the cosmological equations to account for the error introduced by averaging before solving. If these terms prove to be small, then the approximation is good; if they are large, it is not. The results so far are inconclusive. Some researchers have suggested that the extra terms may be enough to account for dark energy entirely, whereas others claim they are negligible.

Observational tests to distinguish between dark energy and the void models are set to be carried out in the very near future. The Supernova Legacy Survey, led by Pierre Astier of the University of Paris, and the Joint Dark Energy Mission, currently under development, should pin down the expansion history of the universe. The Planck Surveyor satellite and a variety of ground-based and balloon-borne instruments will map out the microwave background in ever greater detail. The Square Kilometer Array, a gigantic radio telescope planned for 2020, will supply us with a survey of all the galaxies within our observable horizon. This revolution in cosmology began a decade ago, and it is far from over. ■



ESA/AOES MEDIALAB

WALKING THE PLANCK

The latest spacecraft to measure the cosmic microwave background radiation, the European Space Agency's Planck Surveyor, is scheduled to launch this month.

Planck should provide a complete inventory of fluctuations in the temperature of the microwave background, thereby completing an observational effort that began in the 1960s. These fluctuations reveal what the universe looked like at the tender age of 400,000 years and how it has grown since then. It could tell us whether we live in a giant void.

Planck will also measure fluctuations in the polarization (or directionality) of the radiation, which could reveal whether gravitational waves coursed through the ancient universe as a result of high-energy processes a fraction of a second after the big bang—or even before it.

MORE TO EXPLORE

Geocentrism Reexamined. Jeremy Goodman in *Physical Review D,* Vol. 52, No. 4, pages 1821–1827; March 15, 1995. **http://arxiv.org/abs/astro-ph/9506068**

The State of the Universe: A Primer in Modern Cosmology. Pedro G. Ferreira. Phoenix, 2007.

Cosmology: Patchy Solutions. G.F.R. Ellis in *Nature,* Vol. 452, pages 158–161; March 12, 2008.

Living in a Void: Testing the Copernican Principle with Distant Supernovae. Timothy Clifton, Pedro G. Ferreira and Kate Land in *Physical Review Letters,* Vol. 101, Paper No. 131302; September 26, 2008. **http://arxiv.org/abs/0807.1443**

The Evolution of Primate Color Vision

Analyses of primate visual pigments show that our color vision evolved in an unusual way and that the brain is more adaptable than generally thought

By Gerald H. Jacobs and Jeremy Nathans

KEY CONCEPTS

- The color vision of humans and some other primates differs from that of nonprimate mammals.
- It is called trichromacy, because it depends on three types of light-activated pigments in the retina of the eye.
- Analyses of the genes for those pigments give clues to how trichromacy evolved from the color vision of nonprimate mammals, which have only two kinds of photopigments.
- The authors created trichromatic mice by inserting a human pigment gene into the mouse genome. The experiment revealed unexpected plasticity in the mammalian brain.

—The Editors

To our eyes, the world is arrayed in a seemingly infinite splendor of hues, from the sunny orange of a marigold flower to the gunmetal gray of an automobile chassis, from the buoyant blue of a midwinter sky to the sparkling green of an emerald. It is remarkable, then, that for most human beings any color can be reproduced by mixing together just three fixed wavelengths of light at certain intensities. This property of human vision, called trichromacy, arises because the retina—the layer of nerve cells in the eye that captures light and transmits visual information to the brain—uses only three types of light-absorbing pigments for color vision. One consequence of trichromacy is that computer and television displays can mix red, green and blue pixels to generate what we perceive as a full spectrum of color.

Although trichromacy is common among primates, it is not universal in the animal kingdom. Almost all nonprimate mammals are dichromats, with color vision based on just two kinds of visual pigments. A few nocturnal mammals have only one pigment. Some birds, fish and reptiles have four visual pigments and can detect ultraviolet light invisible to humans. It seems, then, that primate trichromacy is unusual. How did it evolve? Building on decades of study, recent investigations into the genetics, molecular biology and neurophysiology of primate color vision have yielded some unexpected answers as well as surprising findings about the flexibility of the primate brain.

Pigments and Their Past

The spectral sensitivities of the three visual pigments responsible for human color vision were first measured more than 50 years ago and are now known with great precision. Each absorbs light from a particular region of the spectrum and is characterized by the wavelength it absorbs most efficiently. The short-wavelength (S) pigment absorbs light maximally at wavelengths of about 430 nanometers (a nanometer is one billionth of a meter), the medium-wavelength (M) pigment maximally absorbs light at approximately 530 nanometers, and the long-wavelength (L) pigment absorbs light maximally at 560 nanometers. (For context, wavelengths of 470, 520 and 580 nanometers correspond to hues that the typical human perceives as blue, green and yellow, respectively.)

These pigments, each consisting of a protein complexed with a light-absorbing compound derived from vitamin A, sit in the membranes of cone cells: photoreceptive nerve cells in the retina named for their tapering shape. When a pigment absorbs light, it triggers a cascade of molecular events that leads to the excitation of the cone cell. This excitation, in turn, activates other retinal neurons that ultimately convey

▶ CHIMPANZEES, like humans, can distinguish among colors that other mammals cannot see. What observers see in a Kandinsky reflects the properties of the paints, the nature of the illumination, and the color vision system of the viewers.

GEOFFREY CLEMENTS *Corbis (painting)*; BOB ELSDALE *Corbis (chimps)*; LUCY READING-IKKANDA *(photoillustration)*

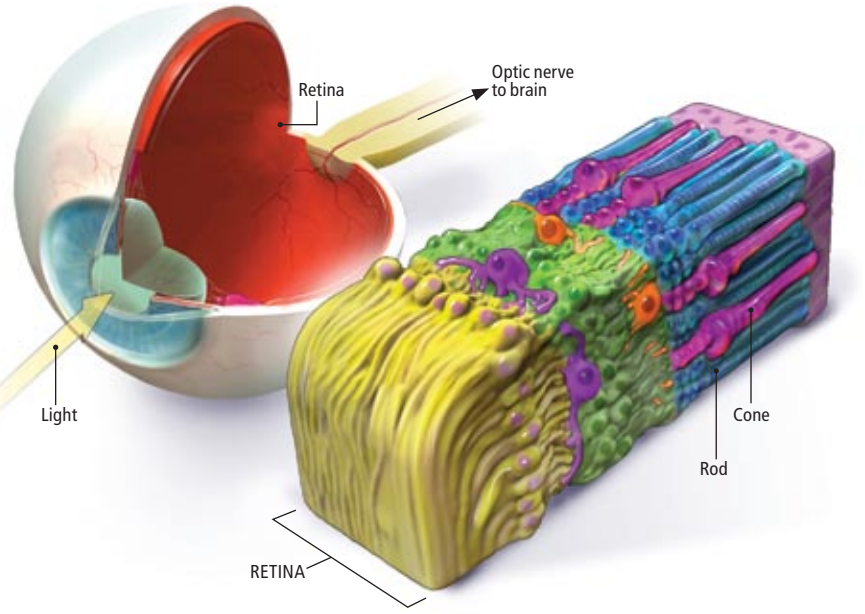
a signal along the optic nerve to the brain.

Although the absorption spectra of the cone pigments have long been known, it was not until the 1980s that one of us (Nathans) identified the genes for the human pigments and, from the DNA sequences of those genes, determined the sequence of amino acids that constitutes each pigment protein. The gene sequences revealed that the M and L pigments are almost identical. Subsequent experiments showed that the difference in spectral sensitivity between them derives from substitutions in just three of the 364 amino acids from which each is built.

Experiments also showed that the M- and L-pigment genes sit next to each other on the X chromosome, one of the two sex chromosomes. (Men have one X and one Y, whereas women have two Xs.) This location came as no surprise, because a common anomaly in human color perception, red-green color blindness, had long been known to occur more often in men than in women and to be inherited in a pattern indicating that the responsible genes reside on the X chromosome. The S-pigment gene, in contrast, is located on chromosome 7, and its sequence shows that the encoded S pigment is related only distantly to the M and L pigments.

By the mid-1990s comparisons of these three pigment genes with those of other animals had provided substantial information about their history. Almost all vertebrates have genes with sequences that are very similar to that of the human S pigment, implying that some version of a shorter-wavelength pigment is an ancient element of color vision. Relatives of the two longer-wavelength pigments (M and L) are also widespread among vertebrates and likely to be quite ancient. But among mammals, the presence of both M- and L-like pigments has been seen only in a subset of primate species—a sign that this feature probably evolved more recently.

Most nonprimate mammals have only one longer-wavelength pigment, which is similar to the longer-wavelength primate pigments. The gene for the longer-wavelength mammalian pigment is also located on the X chromosome. Those features raised the possibility, then, that the two longer-wavelength primate pigment genes first arose in the early primate lineage in this way: a longer-wavelength mammalian pigment gene was duplicated on a single X chromosome, after which mutations in either or both copies of the X-linked ancestral gene produced two quite similar pigments with different ranges of spectral sensitivity—the M and L pigments.



▲ RETINA, a layer of nerve cells at the back of the eye, transmits visual information to the brain via the optic nerve. Color vision depends on cones: tapered sensory cells that contain light-activated pigments. Other light-sensitive cells called rods function in dim light and do not usually participate in color vision. The rods and cones, known collectively as photoreceptors, sit behind other cell types that support vision.

A known mechanism for gene duplications of this type occurs during the formation of eggs and sperm. As cells that give rise to eggs and sperm divide, pairs of chromosomes often swap parts in a process called recombination, and occasionally an unequal exchange of genetic material leads to the production of a chromosome that possesses extra copies of one or more genes. Useful mutations subsequently introduced in those duplicate genes can then be maintained by natural selection. That is, by aiding survival, helpful mutations get passed down to future generations and spread within the population.

In the case of primate color vision, trichromacy based on the "new" M and L pigments (along with the S pigment) presumably conferred a selective advantage over dichromats in some environments. The colors of ripe fruit, for example, frequently contrast with the surrounding foliage, but dichromats are less able to see such contrast because they have low sensitivity to color differences in the red, yellow and green regions of the visual spectrum. An improved ability to identify edible fruit would likely aid the survival of individuals harboring the mutations that confer trichromacy and lead to the spread of those mutant genes in the population.

The mechanisms outlined earlier—gene duplication followed by mutation leading to DNA sequence divergence—would seem to be a reasonable explanation for the evolution of the pri-

[THE AUTHORS]

Gerald H. Jacobs is research professor in the department of psychology and the Neuroscience Research Institute at the University of California, Santa Barbara. The author of more than 200 articles and book chapters on the visual system, he deciphered the genetic mechanism that gives rise to trichromatic color vision in New World primates. **Jeremy Nathans**, who worked out the genetic sequences of the human visual pigment genes and the structure of the corresponding proteins, is professor in the departments of molecular biology and genetics, neuroscience and ophthalmology at the Johns Hopkins University School of Medicine and an investigator at the Howard Hughes Medical Institute.

ANDREW SWIFT (*retina illustrations*); COURTESY OF GERALD H. JACOBS (*Jacobs*); COURTESY OF JOHNS HOPKINS UNIVERSITY (*Nathans*)

mate M- and L-pigment genes because that series of events is known to have occurred in other gene families. Consider, for example, the genes encoding the hemoglobins, proteins that carry oxygen in the blood. The genes for fetal hemoglobin, which is produced beginning in the second month in utero, and the genes for adult hemoglobin seem to have originated as duplicates of a single ancestral gene that then mutated into variants with differing affinities for oxygen. Likewise, immunoglobulins, the proteins that mediate the antibody response of the immune system, come in a great variety and arose from duplication of a single, ancestral gene.

Two Roads to Trichromacy

The real story of the evolution of primate trichromacy, however, turns out to be both more complicated and more interesting. A critical clue came from the discovery that two different genetic mechanisms for trichromatic vision seem to operate in primates: one in the Old World primates (the group that evolved in sub-Saharan Africa and Asia and that includes gibbons, chimpanzees, gorillas and humans) and another in the New World primates (species from Central and South America such as marmosets, tamarins and squirrel monkeys).

Humans and other Old World primates carry both M- and L-pigment genes on each of their X chromosomes and have trichromatic vision. But in testing the color vision of New World primates over the past several decades, one of us (Jacobs) discovered that trichromacy occurs only in a subset of females. All of the New World males and roughly a third of the New World females examined showed the lack of sensitivity to color differences in the middle-to-long wavelengths that is typical of dichromats. Trichromacy was not universal among primates after all.

To explain this curious pattern, several investigators studied the number and arrangement of cone pigment genes in these New World monkeys. Most species turned out to have one short-wavelength pigment gene (presumably located on a nonsex chromosome) and only *one* longer-wavelength gene, located on the X chromosome. In other words, their genetic endowment of visual pigments was comparable to that of the dichromatic mammals. How, then, could *any* of them be trichromats?

The answer is that the gene pool of New World primates includes several variants, or alleles, of the X-linked pigment gene—different versions with slightly modified sequences of DNA. Allelic variation occurs in many genes, but the small differences in DNA sequence between alleles hardly ever translate to functional differences. In New World primates, however, the various X-linked pigment alleles give rise to pigments having different spectral sensitivities. Typical New World primate species such as squirrel monkeys, for example, have three alleles of the X-linked cone pigment gene in their gene pool: one coding for a protein similar to the human M pigment, a second coding for a protein similar to the human L pigment, and a third coding for a pigment with light-absorption properties roughly midway between the first two.

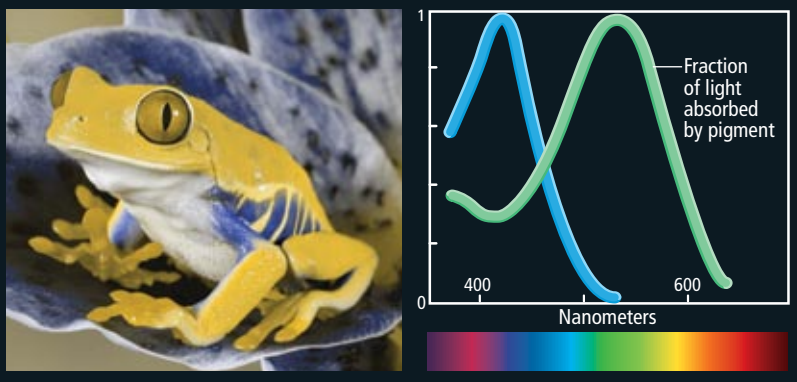
Having two X chromosomes, a female squirrel monkey—and only a female—might inherit two different longer-wavelength alleles (one on each X chromosome), thereby acquiring trichromacy. About a third of all females, however, will

[BASICS]

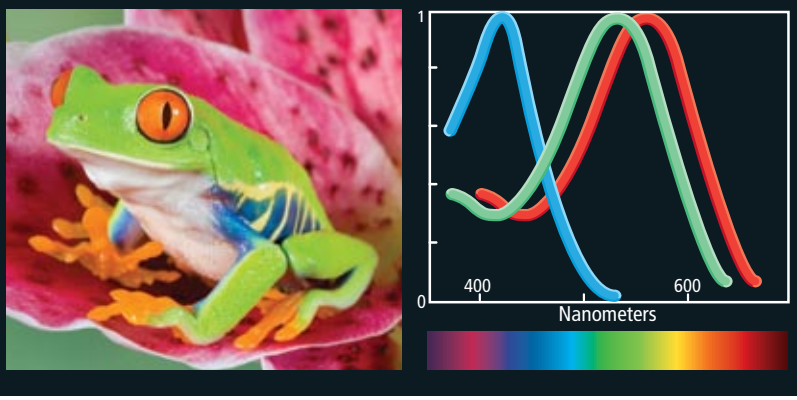
Two Kinds of Mammalian Color Vision

Most mammals are dichromats; their color vision comes from just two kinds of visual pigments (*top panel*): one that absorbs short-wavelength light maximally (*blue curve*) and one that is more sensitive to longer wavelengths (*green curve*). But humans and some other primates have trichromatic color vision (*lower panel*). They see more colors because they use three pigment types: a short-wavelength pigment (*blue curve*) and two kinds of longer-wavelength pigments (*green* and *red curves*).

DICHROMATIC



TRICHROMATIC



EVOLUTIONARY ADVANTAGE?

The colors of ripe fruit frequently contrast with surrounding foliage, and trichromats can discern such contrasts better than dichromats do. The improved ability to identify ripe fruit is one way trichromatic vision could have aided survival, leading to the spread of the genes for trichromacy in primate populations.

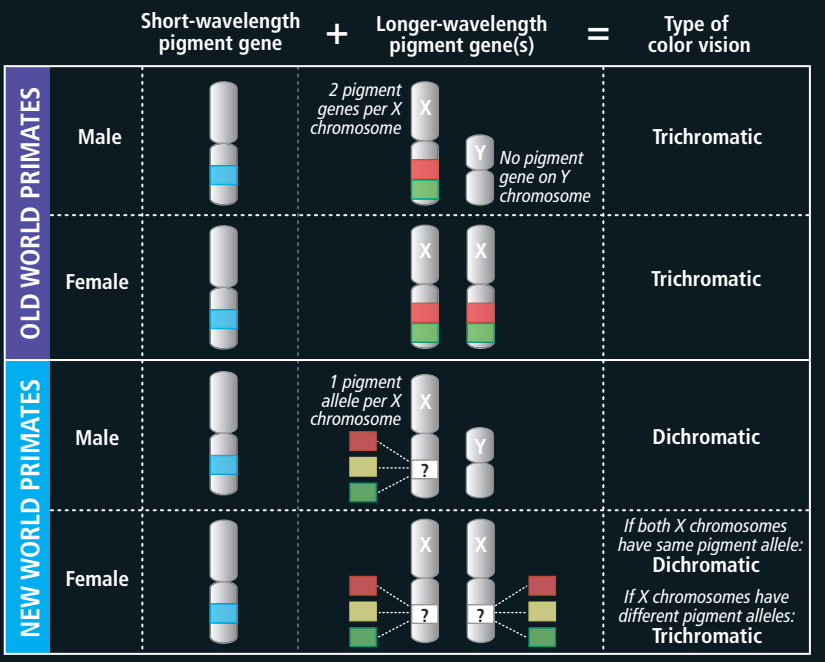
GAIL SHUMWAY *Getty Images (frogs)*; JUSTIN LIGHTLEY *Getty Images (raspberries)*; LUCY READING-IKKANDA (*color spectrums*)

[GENETIC UNDERPINNINGS]

Two Designs for Primate Vision

The genetic basis of trichromacy in Old World and New World primates differs. In both, a gene encoding a short-wavelength pigment (*blue*) sits on a nonsex chromosome. The Old World primates also have two longer-wavelength pigment genes (*red* and *green*) on each X chromosome. Hence, males (with one X) as well as females (with two Xs) have three pigment genes and are trichromats.

New World primates have three variants ("alleles") of a longer-wavelength X-linked pigment gene in their gene pool (*red, yellow* and *green*), but any given X carries just one of these alleles. Consequently, only females having dissimilar pigment alleles on their two Xs are trichromats.



OLD WORLD PRIMATES evolved in Africa and Asia over millions of years and today include great apes (humans, orangutans, gorillas, bonobos and chimpanzees), as well as gibbons, langurs, macaques and mandrills. The Old World primate lineage became isolated from that of the New World—Central and South America—when the African and South American continents became fully separated about 40 million years ago.

◀ MANDRILL

inherit the same pigment allele on both their X chromosomes and end up as dichromats, like the unlucky males. One can think of New World primate trichromacy as the poor man's—or, more accurately, the poor woman's—version of the ubiquitous trichromacy that Old World primates enjoy [*see box above*].

The disparity in color vision between the New and Old World primates provides a window onto the evolution of color vision in both groups. The two primate lineages began to diverge about 150 million years ago, with the progressive separation of the African and South American continents; their genetic isolation appears to have been complete by about 40 million years ago. One might suspect that the two mechanisms of trichromacy evolved independently, after the New and Old World primate lineages separated. Both groups could have started out as dichromats, with the standard mammalian complement of one shorter-wavelength pigment and one longer-wavelength pigment. The longer-wavelength pigment gene in the Old World primates could have undergone the gene duplication followed by sequence divergence that we discussed earlier. In New World primates the longer-wavelength pigment gene could have simply undergone sequence divergence, with successive mutations creating various longer-wavelength pigment alleles that persisted in the population.

Yet comparison of the amino acid sequences of the X-linked visual pigments suggests another scenario. Across both Old and New World primates, all M pigments share one set of three amino acids that confer a maximum spectral sensitivity at 530 nanometers, and all L pigments share a second set of three amino acids that confers a maximum spectral sensitivity at 560 nanometers. From studies of the absorption spectra of other longer-wavelength pigments, we know that sequence changes in a variety of other amino acids can shift the maximal sensitivity of this family of pigments to longer or shorter wavelengths. It seems unlikely, then, that New and Old World primates converged independently on identical sets of amino acids to shift the sensitivities of their longer-wavelength pigments.

Instead it makes more sense to think that allelic variation like that in today's New World primates was the primitive condition, present in the common ancestor of both groups, and that its appearance was the first step in the path to trichromacy for both [*see box on opposite page*]. The various pigment alleles probably arose by successive rounds of mutation in the mammalian longer-wavelength pigment gene some time before the Old and New World primate lineages became isolated. (We suppose that the intermediate-wavelength pigment was part of this primitive complement because its amino acid sequence contains a subset of the three sequence changes that distinguish L from M pigments and because its absorption spectrum is partway between the two.) Then, after the two primate groups became separated, a rare error in recombination occurred in a female of the Old World lineage that happened to be carrying two different alleles of the longer-wavelength pigment

LUCY READING-IKKANDA (*genes*); MAPPING SPECIALISTS (*globe*); CYRIL RUOSO *Minden Pictures* (*mandrill*)

gene. This rare event placed an M allele alongside an L allele on a single X chromosome, thereby allowing trichromacy to extend to males as well as all females.

That genetic innovation granted such a strong selective advantage to its carriers that X chromosomes having only one longer-wavelength pigment gene were ultimately lost from the gene pool of Old World primates. Among the geographically and genetically separate New World primates, the primitive system of three longer-wavelength alleles persisted.

The Role of Randomness

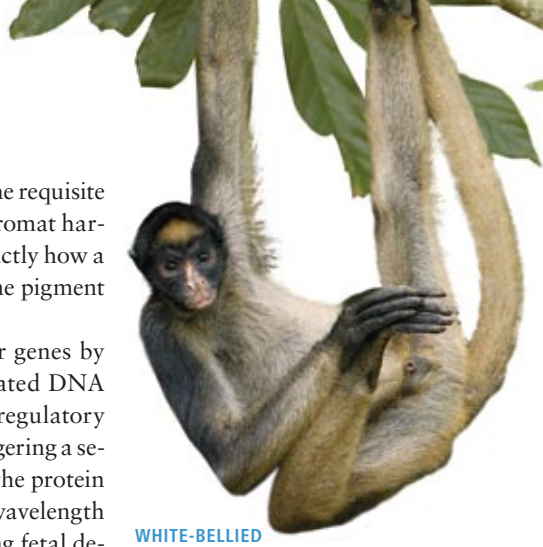
Another surprising implication of our findings in New and Old World primates concerns the role of randomness in trichromacy. We are not referring here to the random genetic mutations that at the outset gave rise to the complement of pigment genes that confer trichromacy. Biologists have generally found that once a beneficial trait has evolved by this chance mechanism, it typically becomes "hardwired": that is, cellular processes that do not stray from a predetermined blueprint meticulously orchestrate the development of the trait in each individual. Yet it seems that for primate color vision, random events in each organism and even in each developing cone cell play a large—indeed, an essential—role.

To explain how randomness helps to produce trichromacy, we must first review how cone cells transmit information about color to the brain. It turns out that having three pigment types, while necessary for trichromatic vision, is just an initial condition. Neural processing of the signals generated by the various photoreceptors is the next step. This step is critical because individual cone cells cannot convey specific information about wavelength. Excitation of each photoreceptor can be triggered by a range of different wavelengths, but the cone cannot signal what particular wavelengths within that band it has absorbed. For example, it could produce the same size signal whether it is hit by 100 photons of a wavelength it absorbs well or by 1,000 photons of a wavelength it absorbs poorly. To distinguish among colors, the visual system must compare the responses of neighboring cones having different pigment types.

For such comparisons to work optimally, each cone cell must contain just one type of pigment, and cones making different pigments must lie close to one another in a kind of mosaic. In fact, in the primate retina each cone cell does contain only a single type of visual pigment, and different cone types are arranged in the requisite mosaic. Yet every cone cell in a trichromat harbors genes for all three pigments. Exactly how a cone cell "decides" to express just one pigment gene is not entirely clear.

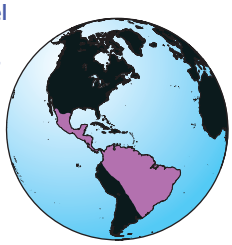
Cells switch on, or express, their genes by way of transcription factors: dedicated DNA binding proteins that attach near a regulatory region called a promoter, thereby triggering a series of events leading to synthesis of the protein encoded by the gene. For the short-wavelength photoreceptors, it appears that during fetal development transcription factors activate the gene for the S pigment. Some unknown process also inhibits expression of the genes for the longer-wavelength pigments in these cells.

But an additional mechanism governs pigment gene expression in the longer-wavelength cones in New World primates, and this mechanism involves an inherently random process. In female New World primates that have different pigment alleles on their two X chromosomes, which allele any given cone cell expresses depends on a molecular coin toss known as X-inactivation. In this process, each female cell randomly disables one of its two X chromo-



WHITE-BELLIED spider monkey ▲

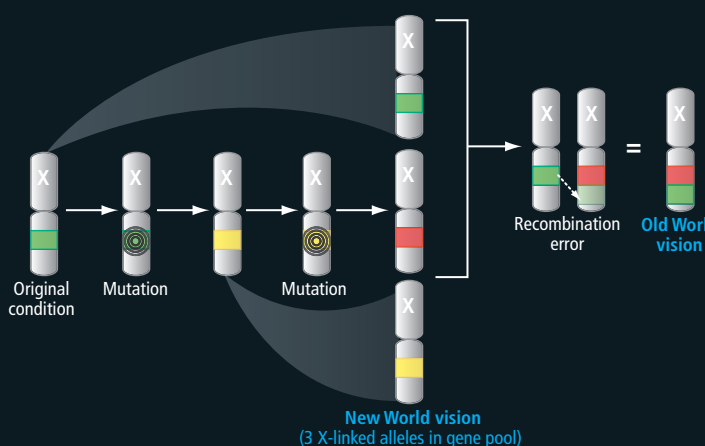
NEW WORLD MONKEYS occupy Central and South America and tend to be smaller than their Old World cousins. They include such species as marmosets, tamarins, squirrel monkeys, spider monkeys, howler monkeys and capuchins.



[FINDINGS]

How Primate Trichromacy Evolved

Comparison of the genetic basis of color vision in New World and Old World primates indicates the key evolutionary steps that led to trichromatic color vision in some female New World monkeys and in both genders of Old World primates.



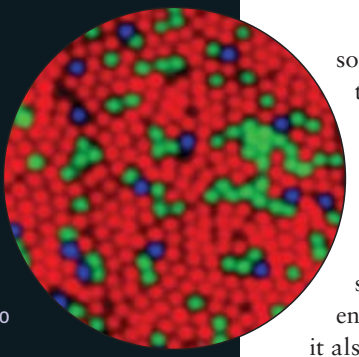
In a common ancestor of both Old and New World primates, an ancestral X-linked longer-wavelength pigment gene (*green at far left*) underwent successive mutations, yielding three longer-wavelength pigment alleles in the gene pool (*green, yellow* and *red*); these changes persist in modern New World primates. After the Old World and New World primate lineages became isolated from each other, an error in recombination—the process in which chromosomes swap parts during formation of eggs and sperm—in an Old World primate female placed two different alleles together on the same X chromosome (*far right*). Because this condition gave a selective advantage to males as well as all females, it became the norm in present-day Old World primates.

LUCY READING-IKKANDA (*genes*); MAPPING SPECIALISTS (*globe*); PETE OXFORD *Minden Pictures* (*white-bellied spider monkey*)

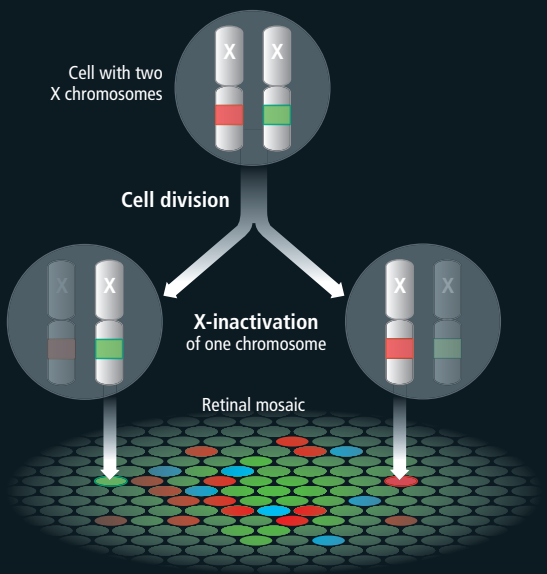
[SURPRISING DISCOVERIES]

Randomness in the Retina

Each individual cone cell contains genes for all three color pigments but selects only one of the three to activate and shuts down the other two. The process that controls selection of the short-wavelength pigment gene is not known in detail. But the mechanisms that determine selection of one of the two longer-wavelength pigment genes appear to be random, and the local distribution of longer-wavelength cone types in the retina appears to be random as well (*computer rendition at upper right*).

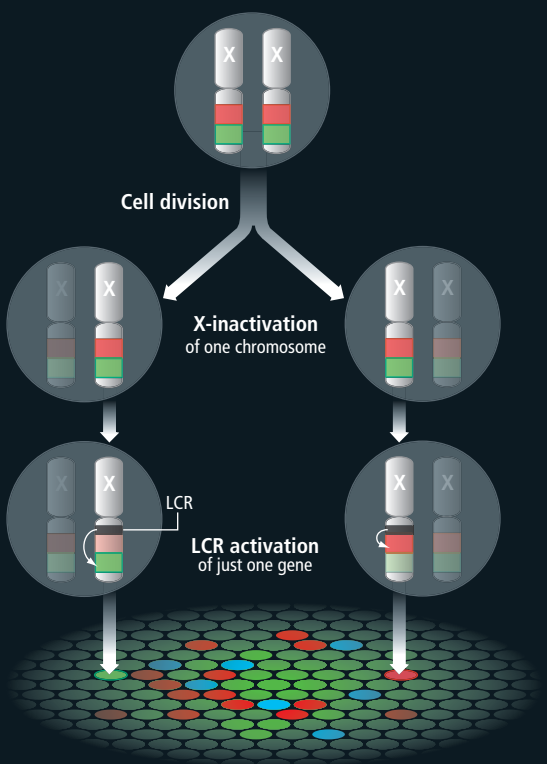


NEW WORLD COIN TOSS



◀ In New World primates the selection of a longer-wavelength pigment gene is accomplished by X-inactivation, a process in which a female cell randomly disables one of its two X chromosomes early in embryonic development. If the female has different pigment alleles on her X chromosomes, inactivation of one of the Xs in each cell will produce a mosaic of longer-wavelength cone types.

OLD WORLD SERENDIPITY



◀ Old World primates have two kinds of longer-wavelength pigment genes on a single X chromosome, so a second step is necessary to narrow pigment types to one per cone cell. X-inactivation takes one X chromosome out of play in female cells. Then, in cells of both genders, a gene regulator called the locus control region (LCR) interacts at random with just one of the longer-wavelength pigment genes, switching only that gene on—again creating a random mosaic of cone types.

somes early in development. X-inactivation ensures that just one pigment allele will be expressed (that is, one type of pigment will be made) in any longer-wavelength cone cell. Because the process is random—half of all cells express genes encoded by one X chromosome, and the other half express genes encoded by the second X chromosome—it also ensures that the longer-wavelength cones in New World primate females will be intermingled across the surface of the retina in a mosaic that permits trichromacy.

X-inactivation occurs in all mammals and is essential for species survival. Without it, female cells would use both X chromosomes to produce proteins, causing the sexes to differ in the amounts of proteins made and thus impairing development in one or both of the genders. But because Old World primates have both M- and L-pigment genes on *each* X chromosome, X-inactivation alone does not narrow expression to just one pigment gene per cone cell in those animals. Another mechanism must be operating as well.

Research by Nathans suggests that which of the two X-linked pigment genes an Old World primate cone cell expresses is determined by a nearby DNA sequence known as the locus control region. The choice is probably made during development when in each cone cell the locus control region interacts with one and only one of the two adjacent pigment gene promoters—that of either the M or the L pigment, but not both—and switches on that gene. The particulars of the interaction have not yet been characterized in detail, but current evidence suggests that this choice may be random.

If this pairing of the locus control region and a promoter is indeed determining pigment gene expression in cone cells and if it is in fact random, then the distribution of M and L cones within any small region of the Old World primate retina should be random as well. Studies by David Williams of the University of Rochester and his colleagues show that within the technical limits of current methods for mapping cone cell distribution, this prediction holds.

The Accidental Colorist

Studies examining the underpinnings of primate color vision also imply that certain retinal and brain mechanisms involved in longer-wavelength color vision may be highly plastic. Although dedicated circuits exist for comparing

LUCY READING-IKKANDA (*genes*); COURTESY OF DAVID WILLIAMS *University of Rochester* (*retina*)

visual information from the S cones with the combined signal from the longer-wavelength cones, the brain and retina seem to be more improvisational in comparing signals from M cones with those from L cones. In particular, the visual system seems to learn the identity of these cones by experience alone—that is, by monitoring the cones' responses to visual stimuli.

What is more, it appears that the principal neural pathway that conveys responses from these longer-wavelength cones may not even be specifically dedicated to color vision. Rather the ability to extract information about hue from the L and M cones may be a happy accident made possible by an ancient neural apparatus for high-resolution *spatial* vision, which evolved to detect the boundaries of objects and their distance from the viewer. John Mollon of the University of Cambridge points out that in primates high-resolution spatial vision is mediated by the longer-wavelength cones and involves the same kind of neural processing that longer-wavelength color vision does—that is, a comparison of the excitation of one L or M cone with the average excitation of a large number of its L and M neighbors. No separate circuitry has yet been found for longer-wavelength color vision, and perhaps none is required. In this view, trichromatic color vision can be considered a hobby of the preexisting spatial vision system.

KRIS KROGH (*mouse*); GEORGE DIEBOLD *Getty Images (eye)*

GENETICALLY ENGINEERED MOUSE has learned to approach the one panel in three that is a different color from the others, revealing that it can see shades of orange that normal mice, as dichromats, would not be able to distinguish from blue. The mouse gained this ability because it possesses a human longer-wavelength visual pigment gene in addition to its two native pigment genes. The experiment demonstrates the remarkable plasticity of the mammalian brain, because the mouse can use its new pigment without having nerve cells specifically wired for interpreting its signals.

The suggestion of neural plasticity in color vision led us to an intriguing question. We imagine that the first step in the evolution of primate trichromacy was emergence in an early female ancestor to all present-day primates of a second longer-wavelength X-linked allele. Could the ancestral primate brain have improvised enough to "use" the new pigment right away, without also evolving new neural circuitry? Could acquiring a third type of pigment be enough in itself to add another dimension to color vision?

It occurred to us that we might test this idea if we could re-create that initial step in the evolution of primate trichromacy in a dichromatic mammal such as a laboratory mouse. We began this experiment by genetically engineering a mouse X chromosome so that it encoded a human L pigment instead of a mouse M pigment, thereby introducing allelic variation of the kind we believe may have occurred millions of years ago in dichromatic primates. We then demonstrated that the resulting line of mice expressed the human gene in their cone cells and that the human L pigment transmitted light signals with an efficiency comparable to that of the mouse M pigment. In addition, the mice expressing the human L pigment were, as expected, sensitive to a broader range of wavelengths than ordinary mice were.

But for our purposes, the key question was: Could female mice having two different X chromosome pigment genes use the retinal mosaic of M and L cones produced by X-inactivation not only to sense but to make discriminations within this broader range of wavelengths? The short and remarkable answer is that they can.

In laboratory tests, we were able to train females having both M and L pigments to discriminate among green, yellow, orange and red panels that, to ordinary mice, look exactly the same. Along with the new L pigment, these mice apparently acquired an added dimension of sensory experience, implying that the mammalian brain has the innate ability to extract information from novel and qualitatively different types of visual input.

This finding has implications for the evolution of sensory systems in general, because it suggests that changes at the "front end" of the system—in the genes for sensory receptors—can drive the evolution of the entire system. With respect to primate trichromacy, the mouse experiment also suggests that the very first primate with two different longer-wavelength pigments saw a world of color no primate had ever seen before. ■

SUPER COLOR VISION?

Some women have four types of visual pigments instead of three. The fourth pigment resulted from a mutation in one of the longer-wavelength X-linked pigment genes and is known to shift the spectral sensitivity of the retina. Whether this shift actually creates the ability to perceive a broader range of hues is under active investigation. Thus far color vision testing has not produced consistent evidence for tetrachromatic vision, and humans who have this ability—if they exist—would not necessarily be aware of their visual anomaly.

MORE TO EXPLORE

The Evolution and Physiology of Human Color Vision: Insights from Molecular Genetic Studies of Visual Pigments. J. Nathans in *Neuron,* Vol. 24, No.2, pages 299–312; October 1999.

Genetically Engineered Mice with an Additional Class of Cone Photoreceptors: Implications for the Evolution of Color Vision. P. M. Smallwood et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 100, No. 20, pages 11706–11711; September 30, 2003.

Emergence of Novel Color Vision in Mice Engineered to Express a Human Cone Pigment. G. H. Jacobs, G. A. Williams, H. Cahill and J. Nathans in *Science,* Vol. 315, pages 1723–1725; March 23, 2007.

Primate Color Vision: A Comparative Perspective. G. H. Jacobs in *Visual Neuroscience,* Vol. 25, Nos. 5–6, pages 619–633; September 2008.

The Post-Traumatic Stress Trap

A growing number of experts insist that the concept of post-traumatic stress disorder is itself disordered and that soldiers are suffering as a result

By David Dobbs

KEY CONCEPTS

- The syndrome of post-traumatic stress disorder (PTSD) is under fire because its defining criteria are too broad, leading to rampant overdiagnosis.
- The flawed PTSD concept may mistake soldiers' natural process of adjustment to civilian life for dysfunction.
- Misdiagnosed soldiers receive the wrong treatments and risk becoming mired in a Veterans Administration system that encourages chronic disability.

—The Editors

In 2006, soon after returning from military service in Ramadi, Iraq, during the bloodiest period of the war, Captain Matt Stevens of the Vermont National Guard began to have a problem with PTSD, or post-traumatic stress disorder. Stevens's problem was not that he had PTSD. It was that he began to have doubts about PTSD: the condition was real enough, but as a diagnosis he saw it being wildly, even dangerously, overextended.

Stevens led the medics tending an armored brigade of 800 soldiers, and his team patched together GIs and Iraqi citizens almost every day. He saw horrific things. Once home, he said he had his share of "nights where I'd wake up and it would be clear I wasn't going to sleep again."

He was not surprised: "I would *expect* people to have nightmares for a while when they came back." But as he kept track of his unit in the U.S., he saw troops greeted by both a larger culture and a medical culture—especially in the Veterans Administration (VA)—that seemed reflexively to view bad memories, nightmares and any other sign of distress as an indicator of PTSD.

"Clinicians aren't separating the few who really have PTSD from those who are experiencing things like depression or anxiety or social and reintegration problems or who are just taking some time getting over it," Stevens says. He worries that many of these men and women are being pulled into a treatment and disability regime that will mire them in a self-fulfilling vision of a brain rewired, a psyche permanently haunted.

Stevens, now a major and still on reserve duty while he works as a physician's assistant, is far from alone in worrying about the reach of PTSD. Over the past five years or so, a long-simmering academic debate over PTSD's conceptual basis and incidence has begun to boil over. It is now splitting the practice of trauma psychology and roiling military culture. Critiques originally raised by military historians and a few psychologists are now advanced by a broad array of experts—indeed, giants of psychology, psychiatry and epidemiology. They include Columbia University's Robert L. Spitzer and Michael B. First, who oversaw the last two editions of the American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders,* the *DSM-III* and *DSM-IV;* Paul McHugh, former chair of Johns Hopkins University's psychiatry department; Michigan State University epidemiologist Naomi Breslau; and Harvard University psychologist Richard J. McNally, a leading authority in the dynamics of memory and trauma and perhaps the most forceful of the critics. The diagnostic criteria for PTSD, they assert, represent a faulty, outdated construct that has been badly overstretched so that it routinely mistakes depression, anxiety or even normal adjustment for a unique and especially stubborn ailment.

This quest to scale back the definition of PTSD and its application stands to affect the expenditure of billions of dollars, the diagnostic framework of psychiatry, the effectiveness of a huge treatment and disability infrastructure, and, most important, the mental health and future lives of hundreds of thousands of U.S. combat veterans and other PTSD patients. Standing in the way of reform is conventional wisdom, deep cultural resistance and foundational concepts of trauma psychology. Nevertheless, it is time, as Spitzer recently argued, to "save PTSD from itself."

Casting a Wide Net

The overdiagnosis of PTSD, critics say, shows in the numbers, starting with the seminal study

TYLER HICKS *New York Times*

DISTRESS CAN BE A NORMAL response to pain and loss or a sign of a psychic wound that is failing to heal. Critics of PTSD diagnostic criteria, including many soldiers, feel that returning veterans' natural process of adjustment is often mislabeled as a dysfunctional state.

of PTSD prevalence, the 1990 National Vietnam Veterans Readjustment Survey (NVVRS). The NVVRS covered more than 1,000 male Vietnam vets in 1988 and reported that 15.4 percent of them had PTSD at the time and that 31 percent had suffered it at some point since the war. That 31 percent has been the standard estimate of PTSD incidence among veterans ever since.

In 2006, however, Columbia epidemiologist Bruce P. Dohrenwend, hoping to resolve nagging questions about the study, reworked the numbers. When he had culled the poorly documented diagnoses, he found that the 1988 rate was 9 percent and the lifetime rate 18 percent.

McNally shares the general admiration for Dohrenwend's careful work. Soon after it was published, however, McNally asserted that Dohrenwend's numbers were still too high because he counted as PTSD cases those veterans with only mild, subdiagnostic symptoms, people rated as "generally functioning pretty well." If you included only those suffering "clinically significant impairment"—the level generally required for diagnosis and insurance compensation in most mental illness—the rates fell yet further, to 5.4 percent at the time of the survey and 11 percent lifetime. It was not one in three veterans who eventually developed PTSD, but one in nine—and only one in 18 had it at any given time. The NVVRS, in other words, appears to have overstated PTSD rates in Vietnam vets by almost 300 percent.

"PTSD is a real thing, without a doubt," McNally says. "But as a diagnosis, PTSD has become so flabby and overstretched, so much a part of the culture, that we are almost certainly mistaking other problems for PTSD and thus mistreating them."

The idea that PTSD is overdiagnosed seems to contradict reports of resistance in the military and the VA to recognizing PTSD—denials of PTSD diagnoses and disability benefits, military clinicians discharging soldiers instead of treating them, and a disturbing increase in suicides among veterans of the Middle East wars. Yet the

PTSD: A Problem Defined by Its Cause

In the current American Psychiatric Association's *Diagnostic and Statistical Manual of Mental Disorders* (*DSM-IV*), the first diagnostic criterion for post-traumatic stress disorder (PTSD) is having experienced trauma:

"The person has been exposed to a traumatic event in which both of the following have been present: (1) the person experienced, witnessed, or was confronted with an event or events that involved actual or threatened death or serious injury, or a threat to the physical integrity of self or others; (2) the person's response involved intense fear, helplessness, or horror."

The presence of three clusters of symptoms—reexperiencing the event, for example, via nightmares or flashbacks; numbing or withdrawal; and hyperarousal, evident in irritability, insomnia, aggression or poor concentration—for more than a month and to the extent that they cause "clinically significant distress or impairment in social, occupational or other important areas of functioning" completes the syndrome's definition.

Critics of this diagnostic construct argue that the symptoms themselves can be characteristic of a wide array of other disorders and may appear together in people who have not experienced trauma.

PTSD was first defined in the *DSM-III,* published in 1980, in response to anti–Vietnam War psychiatrists and veterans who sought a diagnosis to recognize what they saw as the unique suffering of Vietnam vets. ***—D.D.***

[THE AUTHOR]

David Dobbs is a contributing editor of *Scientific American Mind* and was the founding editor of the Mind Matters blog on www.SciAmMind.com. He also regularly writes on medicine, nature, education and culture for the *New York Times Magazine,* Slate.com and *Audubon.* Dobbs has published three books: *Reef Madness: Charles Darwin, Alexander Agassiz and the Meaning of Coral* (Pantheon, 2005); *The Great Gulf: Fishermen, Scientists and the Struggle to Revive the World's Greatest Fishery* (Island Press, 2000); and *The Northern Forest,* with Richard Ober (Chelsea Green, 1995). His Neuron Culture blog is at http://scienceblogs.com/neuronculture

two trends are consistent. The VA's PTSD caseload has more than doubled since 2000, mostly because of newly diagnosed Vietnam veterans. The poor and erratic response to current soldiers and recent vets, with some being pulled quickly into PTSD treatments and others discouraged or denied, may be the panicked stumbling of an overloaded system.

Overhauling both the diagnosis and the VA's care system, critics say, will ensure better care for genuine PTSD patients as well as those being misdiagnosed. But the would-be reformers face fierce opposition. "This argument," McNally notes, "tends to really piss some people off." Veterans send him threatening e-mails. Colleagues accuse him of dishonoring veterans, dismissing suffering, discounting the costs of war. Dean G. Kilpatrick, a University of South Carolina traumatologist and former president of the International Society for Traumatic Stress Studies (ISTSS), once essentially called McNally a liar.

A Problematic Diagnosis

The *DSM-IV,* the most recent edition (published in 1994), defines PTSD as the presence of three symptom clusters—reexperiencing via nightmares or flashbacks; avoidance by numbing or withdrawal; and hyperarousal, evident in irritability, insomnia, aggression or poor concentration—that arise in response to a life-threatening event [*see box above*].

The construction of this definition is suspect. To start with, the link to a traumatic event, which makes PTSD almost unique among complex psychiatric diagnoses in being defined by an external cause, also makes it uniquely problematic, for the tie is really to the memory of an event. When PTSD was first added to the *DSM-III* in 1980, traumatic memories were considered reasonably faithful recordings of actual events. But as research since then has repeatedly shown, memory is spectacularly unreliable and malleable. We routinely add or subtract people, details, settings and actions to and from our memories. We conflate, invent and edit.

In one study by Washington University memory researcher Elizabeth F. Loftus, one out of four adults who were told they were lost in a shopping mall as children came to believe it. Some insisted the event happened even after the ruse was exposed. Subsequently, bounteous research has confirmed that such false memories are common [see "Creating False Memories," by Elizabeth F. Loftus; Scientific American, September 1997].

Soldiers enjoy no immunity from this tendency. A 1990s study at the New Haven, Conn., VA hospital asked 59 Gulf War veterans about their experiences a month after their return and again two years later. The researchers asked about 19 specific types of potentially traumatic events, such as witnessing deaths, losing friends and seeing people disfigured. Two years out, 70 percent of the veterans reported at least one traumatic event they had not mentioned a month after returning, and 24 percent reported at least three such events for the first time. And the veterans recounting the most "new memories" also reported the most PTSD symptoms.

To McNally, such results suggest that some veterans experiencing "late-onset" PTSD may be attributing symptoms of depression, anxiety or other subtle disorders to a memory that has been elaborated and given new significance—or even unconsciously fabricated.

"This has nothing to do with gaming or working the system or consciously looking for sympathy," McNally says. "We all do this: we cast our lives in terms of narratives that help us understand them. A vet who's having a difficult life may remember a trauma, which may or may not have actually traumatized him, and everything makes sense."

To make the diagnosis of PTSD more rigorous, some have suggested that blood chemistry, brain imaging or other tests might be able to de-

COURTESY OF DAVID DOBBS

"PTSD is a real thing, without a doubt," McNally says. "But as a diagnosis, PTSD has become so flabby and overstretched, so much a part of the culture, that we are almost certainly mistaking other problems for PTSD and thus mistreating them."

RICHARD J. McNALLY, Harvard University

tect physiological signatures of the disorder. Some studies of stress hormones in groups of PTSD patients show differences from normal subjects, but the overlap between the normal and the PTSD groups is huge, making individual profiles useless for diagnostics. Brain imaging has similar limitations, with the abnormal dynamics in PTSD heavily overlapping those of depression and anxiety.

With memory unreliable and biological markers elusive, diagnosis depends on clinical symptoms. But as a study in 2007 starkly showed, the symptom profile for PTSD is as slippery as the would-be biomarkers. J. Alexander Bodkin, a psychiatrist at Harvard's McLean Hospital, screened 90 clinically depressed patients separately for PTSD symptoms and for trauma, then compared the results. First he and a colleague used a standardized screening interview to assess symptoms. Then two other PTSD diagnosticians, ignorant of the symptom reports, used another standard interview to see which patients had ever experienced trauma fitting *DSM-IV* criteria.

If PTSD arose from trauma, the patients with PTSD symptoms should have histories of trauma, and those with trauma should show more PTSD. It was not so. Although the symptom screens rated 70 of the 90 patients positive for PTSD, the trauma screens found only 54 who had suffered trauma: the diagnosed PTSD "cases" outnumbered those who had experienced traumatic events. Things got worse when Bodkin compared the diagnoses one on one. If PTSD required trauma, then the 54 trauma-exposed patients should account for most of the 70 PTSD-positive patients. But the PTSD-symptomatic patients were equally distributed among the trauma-positive and the trauma-negative groups. The PTSD rate had zero relation to the trauma rate. It was, Bodkin observed, "a scientifically unacceptable situation."

More practically, as McNally points out, "To give the best treatment, you have to have the right diagnosis."

The most effective treatment for patients whose symptoms arise from trauma is exposure-based cognitive-behavioral therapy (CBT), which concentrates on altering the response to a specific traumatic memory by repeated, controlled exposure to it. "And it works," McNally says. "If someone with genuine PTSD goes to the people who do this really well, they have a good chance of getting better." CBT for depression, in contrast, teaches the patient to recognize dysfunctional loops of thought and emotion and develop new responses to normal, present-day events. "If a depressed person takes on a PTSD interpretation of their troubles and gets exposure-based CBT, you're going to miss the boat," McNally says. "You're going to spend your time chasing this memory down instead of dealing with the way the patient misinterprets present events."

To complicate matters, recent studies showing that traumatic brain injuries from bomb blasts, common among soldiers in Iraq, produce symptoms almost indistinguishable from PTSD. One more overlapping symptom set.

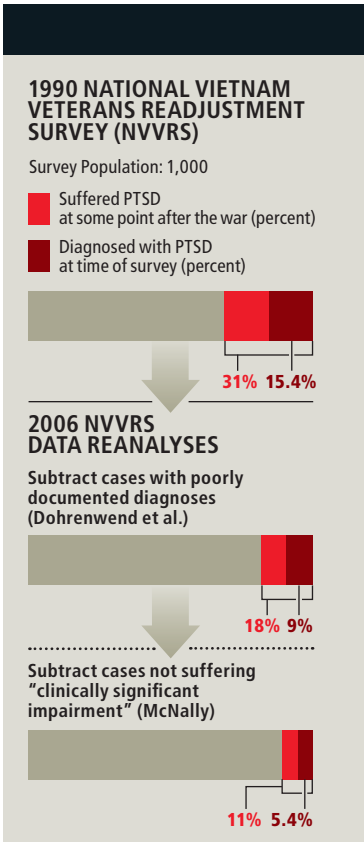
"The overlap issue worries me tremendously," says Gerald M. Rosen, a University of Washington psychiatrist who has worked extensively with PTSD patients. "We have to ask how we got here. We have to ask ourselves, 'What do we gain by having this diagnosis?' "

Disabling Conditions

Rosen is thinking of clinicians when he asks about gain. But what does a veteran gain with a PTSD diagnosis? One would hope, of course, that it grants access to effective treatment and support. This is not happening. In civilian populations, two thirds of PTSD patients respond to treatment. But as psychologist Christopher Frueh, who researched and treated PTSD for the VA from the early 1990s until 2006, notes, "In the two largest VA studies of combat veterans,

DISCREPANCIES IN DIAGNOSES

Estimates of PTSD incidence among soldiers are often inflated for several reasons, including the condition's vague diagnostic criteria. Assessment survey questions—and answers—are also highly subject to interpretation. Stricter analyses of the 1990 National Vietnam Veterans Readjustment Survey data, for instance, reduced PTSD incidence to one third of the original results.



JUSTIN IDE *Harvard News Office (McNally)*; JEN CHRISTIANSEN; SOURCES: "FINDINGS FROM THE NATIONAL VIETNAM VETERANS READJUSTMENT STUDY," BY J. L. PRICE, NATIONAL CENTER FOR POSTTRAUMATIC STRESS DISORDER; "THE PSYCHOLOGICAL RISKS OF VIETNAM FOR U.S. VETERANS: A REVISIT WITH NEW DATA AND METHODS," BY B. DOHRENWEND ET AL., IN *SCIENCE*, VOL. 313, AUGUST 18, 2006; "PSYCHIATRIC CASUALTIES OF WAR," BY R. J. McNALLY, IN *SCIENCE*; *IBID.* (*graphs*)

DISTINCTION DENIED

In January the Department of Defense announced that it would not award Purple Hearts to veterans diagnosed with PTSD. The ruling was based partly on a distinction between physical wounds that can be objectively assessed and psychic suffering that is more subjective. Divided public reaction to the decision underlined the heavy emotional and cultural significance attached to this diagnosis—and the diagnostic construct's problems.

On one hand, recent neuroscience makes clear that serious mental health disorders reflect brain dysfunction, if not actual injury, and it seems only fair to consider psychic war wounds every bit as real as physical ones. At the same time, a lack of sure diagnostic borders between PTSD and other depressive or anxiety disorders, along with strong evidence that PTSD is overdiagnosed, will make awarding this medal for the condition an extremely problematic proposal until the nature of PTSD gains clarity and science identifies some indisputable biomarkers. ***—D.D.***

neither showed a treatment effect. Vets getting PTSD treatment from the VA are no more likely to get better than they would on their own."

The reason, Frueh says, is the collision of the PTSD construct's vagaries with the VA's disability system, in which every benefit seems structured to discourage recovery.

The first benefit is health care. PTSD is by far the easiest mental health diagnosis to have declared "service-connected," a designation that often means the difference between little or no care and broad, lasting health coverage. Service connection also makes a vet eligible for monthly disability payments of up to $3,000. That link may explain why most veterans getting PTSD treatment from the VA report worsening symptoms until they are designated 100 percent disabled—at which point their use of VA mental health services drops by 82 percent. It may also help explain why, although the risk of PTSD from a traumatic event drops as time passes, the number of Vietnam veterans applying for PTSD disability almost doubled between 1999 and 2004, driving total PTSD disability payments to more than $4 billion annually.

Perhaps most disastrously, these payments continue only if you are sick. For unlike a vet who has lost a leg, a vet with PTSD loses disability benefits as soon as he recovers or starts working. The entire system seems designed to encourage chronic disability. "In the several years I spent in VA PTSD clinics," Frueh says, "I can't think of a single PTSD patient who left treatment because he got better. But the problem is not the veterans. The problem is that the VA's disability system, which is 60 years old now, ignores all the intervening research we have on resilience, on the power of expectancy, and on the effects of incentives and disincentives. Sometimes I think they should just blow it up and start over." But with what?

Richard A. Bryant, an Australian PTSD researcher and clinician, suggests a disability system more like that in place Down Under. An Australian soldier injured in combat receives a lifelong "noneconomic" disability payment of $300 to $1,200 monthly. If the injury keeps him from working, he also gets an "incapacity" payment, as well as job training and help finding work. Finally—a crucial feature—he retains all these benefits for two years once he goes back to work. After that, incapacity payments taper to zero over five years. But noneconomic payments—a kind of financial Purple Heart—continue forever. And like all Australians, the soldier gets free lifetime health care. Australian vets come home to an utterly different support system from ours: theirs is a scaffold they can climb. Ours is a low-hanging "safety net" liable to trap anyone who falls in.

Two Ways to Carry a Rifle

When a soldier comes home, he must try to reconcile his war experience with the person he was beforehand and the society and family he returns to. He must engage in what psychologist Rachel Yehuda, who researches PTSD at the Bronx VA Hospital, calls "recontextualization"—the process of integrating trauma into normal experience. It is what we all do, on various scales, when we suffer breakups, job losses or the deaths of loved ones. Initially the event seems an impossible aberration. Then slowly we accept the trauma as part of the complex context that is life.

Major Matt Stevens recognizes that this adjustment can take time. Even after two years at home, the war still occupies his dreams. Sometimes, for instance, he dreams that he is doing something completely normal—while carrying his combat rifle: "One night I dreamt I was birdwatching with my wife. When we saw a bird, she would lift her binoculars, and I would lift my rifle and watch the bird through the scope. No thought of shooting it. Just how I looked at the birds."

It would be easy to read Stevens's dream as a symptom of PTSD, expressing fear, hypervigilance and avoidance. Yet it can also be seen as demonstrating his success in recontextualizing his experience: reconciling the man who once used a gun with the man who no longer does.

Saving PTSD from itself, Spitzer, McNally, Frueh and other critics say, will require a similar shift—seeing most postcombat distress not as a disorder but as part of normal, if painful, healing. This turnaround will involve, for starters, revising the rubric for diagnosing PTSD—currently under review for the new *DSM-V* due to be published in 2012—so it accounts for the unreliability of memory and better distinguishes depression, anxiety and phobia from true PTSD. Mental health evaluations need similar revisions so they can detect genuine cases without leading patients to impose trauma narratives on other mental health problems. Finally, Congress should replace the VA's disability system with an evidence-based approach that removes disincentives to recovery—and even go the extra mile and give all combat veterans, injured or not, lifetime health care.

ALEX WONG *Getty Images*

Recent studies showing that traumatic brain injuries from bomb blasts, common among soldiers in Iraq, produce symptoms almost indistinguishable from PTSD.

One more overlapping symptom set.

ROADSIDE BOMB aftermath in Ramadi, Iraq

These changes will be hard to sell in a culture that resists any suggestion that PTSD is not a common, even inevitable, consequence of combat. Mistaking its horror for its prevalence, most people assume PTSD is epidemic, ignoring all evidence to the contrary.

The biggest longitudinal study of soldiers returning from Iraq, led by VA researcher Charles Milliken and published in 2007, seemed to confirm that we should expect a high incidence of PTSD. It surveyed combat troops immediately on return from deployment and again about six months later and found around 20 percent symptomatically "at risk" of PTSD. But of those reporting symptoms in the first survey, half had improved by the second survey, and many who first claimed few or no symptoms later reported serious symptoms. How many of the early "symptoms" were just normal adjustment? How many of the later symptoms were the imposition of a trauma narrative onto other problems?

Stevens, for one, is certain these screens are mistaking many going through normal adjustment as dangerously at risk of PTSD. Even he, though functioning fine at work and home and in society, scored positive in *both* surveys; he is, in other words, one of the 20 percent at risk. Finally, and weirdly, both screens missed about 75 percent of those who actually sought counseling—a finding that raises further doubts about the evaluations' accuracy. Yet this study received prominent media coverage emphasizing that PTSD rates were probably being badly undercounted.

A few months later another study—the first to track large numbers of soldiers through the wars in Iraq and Afghanistan—provided a clearer and more consistent picture. Led by U.S. Navy researcher Tyler Smith and published in the *British Medical Journal*, the study monitored mental health and combat exposure in 50,000 U.S. soldiers from 2001 to 2006. The researchers took particular care to tie symptoms to types of combat exposure. Among some 12,000 troops who went to Iraq or Afghanistan, 4.3 percent developed diagnosis-level symptoms of PTSD. The rate ran about 8 percent in those with combat exposure and 2 percent in those not exposed.

These numbers are about a quarter of the rates Milliken found. But they are a close match to PTSD rates seen in British Iraq War vets and to rates McNally calculated for Vietnam veterans. The contrast to the Milliken study, along with the consistency with British rates and with McNally's NVVRS calculation, should have made the Smith study big news. Yet the media, the VA and the trauma psychology community almost completely ignored the study. "The silence," McNally wryly noted, "was deafening."

This silence may be merely a matter of good news going unremarked. Yet it supports McNally's contention that we have a cultural obsession with trauma. The selective attention also supports the assertion by military historian and PTSD critic Ben Shephard that American society itself gained something from the creation of the PTSD diagnosis in the late 1970s: a vision of war's costs that, by transforming warriors into victims, lets us declare our recognition of war's horror and absolves us for sending them—for we were victimized, too, fooled into supporting a war we later regretted. We should recognize war's horror. We should feel the soldier's pain. But to impose on a distressed soldier the notion that his memories are inescapable, that he lacks the strength to incorporate his past into his future, is to highlight our moral sensitivity at the soldier's expense.

PTSD exists. Where it exists we must treat it. But our cultural obsession with PTSD has magnified and finally perhaps become the thing itself—a prolonged failure to contextualize and accept our own collective aggression. It may be our own postwar neurosis. ■

REUTERS

MORE TO EXPLORE

A War of Nerves: Soldiers and Psychiatrists in the Twentieth Century. Ben Shephard. Harvard University Press, 2001.

Progress and Controversy in the Study of Posttraumatic Stress Disorder. Richard J. McNally in *Annual Review of Psychology,* Vol. 54, pages 229–252; 2003.

Saving PTSD from Itself in DSM-V. Robert L. Spitzer, Michael B. First and Jerome C. Wakefield in *Journal of Anxiety Disorders,* Vol. 21, No. 2, pages 233–241; February 2007.

Posttraumatic Stress Disorder: An Empirical Evaluation of Core Assumptions. Gerald M. Rosen and Scott O. Lilienfeld in *Clinical Psychology Review,* Vol. 28, No. 5, pages 837–868; June 2008.

The Dawn of Miniature Green Lasers

Semiconductors can generate laser light in all colors except one. But new techniques for growing laser diodes could soon make brilliant full-spectrum displays a reality

By Shuji Nakamura and Michael Riordan

KEY CONCEPTS

- Solid-state lasers can produce light in the red and blue parts of the spectrum but not the green.
- Recent research suggests that this "green gap" could be plugged as early as this year.
- The advance will allow for laser-based video displays that are small enough to fit in a cell phone.

—The Editors

On a rainy Saturday morning in January 2007, Henry Yang, chancellor of the University of California, Santa Barbara, took an urgent phone call. He excused himself abruptly from a meeting, grabbed his coat and umbrella, and rushed across the windswept U.C.S.B. campus to the Solid State Lighting and Display Center. The research group there included one of us (Nakamura), who had just received the Millennium Technology Prize for creating the first light-emitting diodes (LEDs) that emit bright blue light. Since that breakthrough over a decade earlier, Nakamura had continued his pioneering research on solid-state (semiconductor) lighting, developing green LEDs and the blue laser diodes that are now at the core of modern Blu-ray disc players.

As Yang reached the center about 10 minutes later, people were milling about a small test lab. "Shuji had just arrived and was standing there in his leather jacket asking questions," he recalled. Nakamura's colleagues Steven DenBaars and James C. Speck were speaking with a few graduate students and postdoctoral researchers as they took turns looking into a microscope. They parted for Yang, who peered into the eyepiece to witness a brilliant blue-violet flash emanating from a glassy chip of gallium nitride (GaN).

Within days another group of researchers at Rohm Company in Kyoto, Japan—a partner in the U.C.S.B. center—duplicated the feat using similar materials. Although blue laser diodes are not in themselves very revolutionary [see "Blue-Laser CD Technology," by Robert L. Gunshor and Arto V. Nurmikko; Scientific American, July 1996], Nichia Chemical Industries (based in Tokushima, Japan, where Nakamura worked until 2000), Sony and other companies were still struggling to produce inexpensive GaN laser devices for the Blu-ray disc market. These diodes had previously been fashioned using a method with stubborn limitations that have kept manufacturing yields down and diode costs high.

The groups from U.C.S.B. and Rohm are developing a new way to grow the crystalline layers of gallium nitride and related alloys that make up a laser diode. The early successes of the approach not only promise greater yields but also buoy hopes of an even bigger payoff: rugged, compact GaN diodes that emit green laser light—a goal that has long eluded scientists and engineers. The technique should also lead to

high-efficiency green LEDs that emit much more light than existing devices.

These achievements would fill a gaping void in the visible spectrum where evolution has trained our eyes to be most sensitive, plugging the "green gap" in the red-green-blue triad needed for full-color laser projection and displays. They should help speed the introduction of laser projectors for televisions and movie theaters—which will display much richer colors than other systems—and of tiny, handheld "pico projectors" to be used, for example, in cell phones. And high-power green diodes might even be employed in such diverse applications as DNA sequencing, industrial process control and underwater communications.

A New Angle

The key advance that led to bright blue solid-state lighting was the mid-1990s conversion to LEDs and laser diodes made of gallium nitride and its alloys [for a profile of Nakamura, see "Blue Chip," by Glenn Zorpette; Scientific American, August 2000]. Before that, most researchers had focused their efforts on zinc selenide and related compounds. In the new approach, an exceedingly smooth, nanometers-thin layer of indium gallium nitride (InGaN) is sandwiched between two layers of GaN, forming what is called a heterostructure or quantum well [*see box on next page*].

By applying a suitable voltage, researchers set up an electric field perpendicular to these layers that drives electrons and holes—positively charged quasiparticles corresponding to the absence of electrons—together within the InGaN active layers. Inside this narrow trench, the electrons and holes recombine, annihilating one another and generating photons with an energy precisely determined by the properties of the active semiconductor material. By increasing the indium concentration in the alloy, one can lower this energy, thereby increasing the wavelength of the light and changing its color from violet to blue to green.

In LEDs the photons leave the well almost immediately, perhaps rebounding once or twice before exiting the device or being absorbed in the other layers. But in laser diodes, which produce coherent light, the photons stay largely confined within the trench. Two highly reflective mirrors—generally polished crystal surfaces at ei-

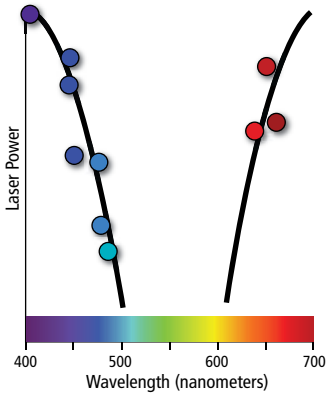
WHAT ABOUT GREEN LASER POINTERS?

The green lasers that have long been available employ a complicated two-step process to generate light. Semiconductor lasers inside these devices emit infrared radiation with a wavelength around 1,060 nanometers. This radiation then pumps a crystal that oscillates at half this wavelength—about 530 nanometers, solidly in the green. The process is costly, inefficient and imprecise—the second crystal can heat up, altering the wavelength of the resultant green light. Laser diodes that generate green light directly would avoid these problems.

CASPAR BENSON

THE GREEN GAP PROBLEM

Scientists have long been able to build semiconductor lasers that create light in red parts of the spectrum, and in the past decade they have conquered the blue and violet sections as well. Yet as they try to push these lasers into the green part of the spectrum, the amount of power produced drops precipitously.



ther end of it—recycle the photons back and forth inside, further stimulating electron-hole recombination. The laser light generated by this "stimulated emission" process is a tight pencil beam of exceedingly pure color.

To make conventional GaN diodes, workers place a thin wafer of sapphire (or, increasingly, gallium nitride) inside a reaction chamber. There hot gases deposit successive layers of gallium, indium and nitrogen atoms on that substrate, with the exact proportions of each element varying from layer to layer. The atoms in these layers automatically align with the existing crystalline structure, as predetermined by the substrate. Atom by atom, the layers grow in parallel with what is called the substrate's *c*-plane, which is perpendicular to the crystal's axis of hexagonal symmetry [*see box on opposite page*].

Unfortunately, electrostatic forces and internal stresses between successive layers of positively charged gallium or indium ions and negatively charged nitrogen ions create strong electric fields perpendicular to the *c*-plane. These fields, which can reach up to 100 volts per micron—equivalent to nearly 200 million volts across an average person's height—counteract the applied external voltage. They pull electrons *away* from holes—making it harder for them to recombine and yield light. In effect, the electrons pile up at one side of the long quantum dance hall and holes at the other, both reluctant to cross over and meet.

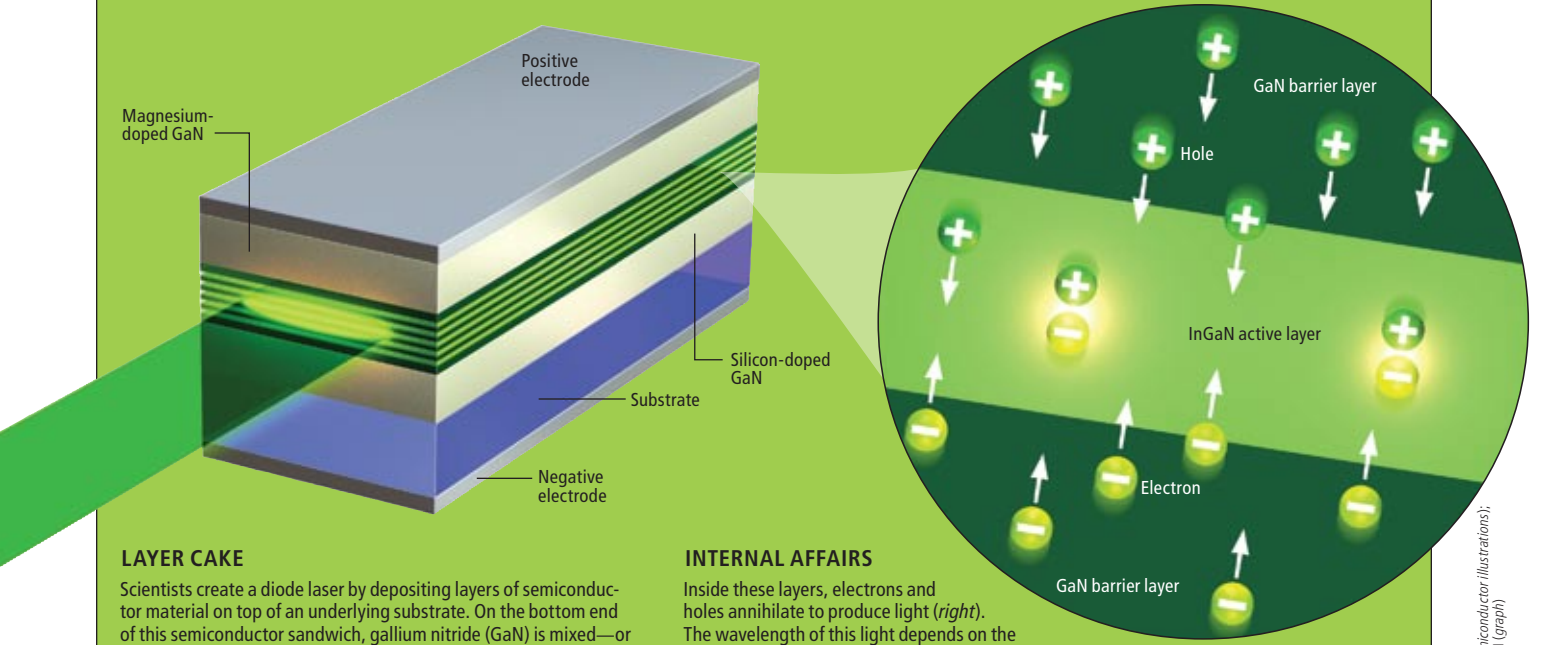
Known as the quantum-confined Stark effect, this nagging problem becomes particularly acute as the color of the emitted light shifts from violet to blue to green. And as the current through the diode increases, the greater number of charge carriers partially blocks the internal electric fields that keep electrons and holes apart. With these fields partially screened out, the electrons and holes then recombine at higher energies, shifting the light toward the blue end of the spectrum. These problems are the main reason why green laser diodes and high-efficiency green LEDs have remained but a dream for more than a decade. (The familiar green laser pointers used by lecturers have semiconductor lasers that emit infrared radiation and pump another laser in a complicated, inefficient frequency-doubling scheme.)

The approach pioneered by the U.C.S.B. and Rohm groups attempts to sidestep these prob-

[THE BASICS]

HOW SEMICONDUCTOR LASERS WORK

Inside a solid-state laser, electrons meet positively charged entities called holes, annihilating one another and creating light. To adjust the wavelength of this light, scientists must alter the material inside the semiconductor. But doing so can lead to other problems.



LAYER CAKE

Scientists create a diode laser by depositing layers of semiconductor material on top of an underlying substrate. On the bottom end of this semiconductor sandwich, gallium nitride (GaN) is mixed—or "doped"—with silicon impurities to produce an excess of negatively charged electrons. On the other end, GaN is doped with magnesium to give it an excess of positive charges, or "holes." A voltage across the electrodes sets up an electric field that drives the electrons and holes together inside the central active layers.

INTERNAL AFFAIRS

Inside these layers, electrons and holes annihilate to produce light (*right*). The wavelength of this light depends on the indium (In) content of the active layer—more leads to longer wavelengths and thus greener light. But the more indium in these layers, the more that indium is likely to pool into small "islands" during manufacture. The islands can alter the light's wavelength—an unacceptable flaw in a laser.

GEORGE RETSECK (*semiconductor illustrations*); SCIENTIFIC AMERICAN (*graph*)

[THE SOLUTION]

A NEW FOUNDATION

A substrate is a slice of a crystal, and anything grown on top of it inherits its crystalline structure. The blue diode lasers that power Blu-ray disc players and PlayStation 3 game consoles are usually grown on top of sapphire, which, as substrates go, is relatively cheap and readily available. Yet it is difficult to use these substrates to make green laser diodes. In response, scientists have turned to alternative crystal facets for help.

C-PLANE: THE CLASSIC CUT

Though commonly used for blue lasers, a *c*-plane substrate has drawbacks, such as inducing electric fields that conspire to keep electrons and holes apart. The problem gets worse as the wavelength shifts toward green.

M-PLANE: A COSTLY ALTERNATIVE

Two research groups are growing laser diodes on a crystal's *m*-plane, which cuts across the crystal's side. Diodes grown on this plane do not suffer from induced fields, but the substrates are more costly than *c*-plane versions.

SEMIPOLAR: THE COMPROMISE

A third option is semipolar substrates that are cut at a 45-degree angle to the crystal axis. These substrates also do not produce strong fields, and they seem to yield better lasers and LEDs than the *m*-plane substrates do.

lems by starting with a thin wafer of pure, crystalline GaN that has been sliced along a larger crystal's *m*-plane [*see box above*] and then polished. Diodes fabricated on these so-called nonpolar substrates do not encounter the problems of conventional polar *c*-plane devices, because the troublesome fields caused by polarization and internal stresses are much lower.

The diodes grown on GaN also produce light more efficiently than ones grown on sapphire because they suffer from far fewer crystalline defects—submicroscopic irregularities and mismatches at the interfaces between successive layers. Such defects act as centers where electrons and holes recombine to produce unwanted heat instead of light. They can easily propagate upward through the successive diode layers during the growth process (in what are called threading dislocations) and reach the active layers. The presence of these defects played havoc with production yields when Nichia and Sony first tried to manufacture blue laser diodes. Because a GaN substrate will generate nowhere near as many mismatches as sapphire does with the next-above layer of GaN or one of its alloys, diodes grown on nonpolar GaN substrates can therefore produce much more light—and have correspondingly less heat to dispose of.

First suggested in the late 1990s, the nonpolar technique was attempted by several groups beginning in 2000—including DenBaars and Speck at U.C.S.B. The early devices performed only modestly, mainly because of the lack of high-quality GaN substrates. In 2006, however, Mitsubishi Chemical Corporation in Tokyo—another partner in the U.C.S.B. center—began supplying excellent, low-defect *m*-plane GaN substrates to the Rohm and U.C.S.B. research groups. Less than a centimeter on a side, the substrates were sliced from small GaN crystals about the size of a pencil eraser.

With the new material in hand, Rohm and U.C.S.B. fabricated much more efficient LEDs in late 2006, and by early 2007 these groups began trying to make the more challenging laser diodes. On that rainy Saturday morning of January 27, U.C.S.B. graduate student Matthew Schmidt went to the lab to finish the last fabrication step. Then he took the diode over to the nearby test lab and hitched it up to a power supply. Suddenly, as he cranked up the current flowing through the diode, a narrow beam of blue-violet light shot out of it.

"Wow!" Schmidt thought. "I can finally graduate!"

He called his thesis adviser DenBaars, who at first thought he was joking but soon alerted the rest of the group and Chancellor Yang. They arrived within minutes to observe the surprising results. This first nonpolar GaN laser diode op-

[THE AUTHORS]

Shuji Nakamura is professor in the materials department and director of the Solid-State Lighting and Energy Center at the University of California, Santa Barbara. He was awarded the 2006 Millennium Technology Prize for his work on blue laser diodes and LEDs. **Michael Riordan** teaches the history of physics and technology at Stanford University and at the University of California, Santa Cruz. He is co-author of *Crystal Fire: The Invention of the Transistor and the Birth of the Information Age* (W. W. Norton, 1997).

RANDY LAMB *University of California, Santa Barbara (Nakamura)*; COURTESY OF DONNA RIORDAN *(Riordan)*; GEORGE RETSECK *(crystals)*

erated at a wavelength of 405 nanometers (nm), as did the first Rohm device a few days later. And the currents flowing through these diodes were only two to three times what was then being achieved in commercially available devices made by Nichia and Sony, indicating that any heating problems were manageable.

Going for Green

After that breakthrough, the U.C.S.B. group decided to drop most of its work on polar diodes and focus on nonpolar ones. It also began investigating a related strategy based on "semipolar" GaN substrates, which are wafers cut at an angle of about 45 degrees to a crystal's major axis [*see box on preceding page*]. Diodes fabricated on semipolar substrates also have much less intense internal electric fields than polar diodes do, though not as low as in nonpolar diodes. The U.C.S.B. researchers hope that one of these geometries will allow them to create the first green laser diodes and to make high-power LEDs at even longer wavelengths. Rohm has forged ahead in these areas, too, concentrating its efforts on nonpolar substrates.

The new substrates, however, are not sufficient on their own to get beyond blue. Green diodes require adding more indium to the InGaN active layer, but the extra indium exacerbates internal stresses and disrupts the crystalline structure. It increases the number of crystal defects, which in turn reduces the light output and generates excess heat. While LEDs can still function despite the added defects, their efficiencies plummet as the color shifts from blue to green. Laser diodes are even more finicky and cannot tolerate so many defects. The highest wavelength thus far achieved in a laser diode is 488 nm, in the blue-green (or cyan) part of the spectrum.

Layers of InGaN must also be grown at substantially lower temperatures—about 700 degrees Celsius versus 1,000 degrees C for the GaN layers around it—to prevent the indium atoms from dissociating from the other atoms. Such dissociation can form areas of inhomogeneous indium alloys, or "islands," which in turn causes the electron-hole recombination energy to vary from point to point. That variation makes the emission spectrum too broad to yield the coherent, monochromatic light expected from a laser. Thus, when workers raise the reactor temperature to grow the next GaN layer atop the delicate InGaN layer just deposited, they must be especially careful so as not to form too many of these islands. But the pro-

[THE APPLICATIONS]

HANDHELD PROJECTORS

The smallest currently available handheld projectors are about the size of a remote control and use LEDs to generate light. By the end of this year the first laser-based models should go on sale. Even though they use frequency-doubling technology to create green laser light, they will produce high-resolution, richly colored images. Future models that rely on green laser diodes will allow for brighter and more efficient displays, and they will also shrink the projectors enough to allow them to fit inside your cell phone. Here is a look at two laser-based prototypes currently in development and a few LED projecters out now. —***The Editors***

MICROVISION SHOW WX ▸

Inside this laser-projector protype, red, blue and green lasers focus onto a single mirror about the size of a pinhead. As light bounces off that reflector, the mirror assembly rapidly scans back and forth to project pixels one by one onto a screen or wall. The lack of lenses means the projector never needs to be focused.
Resolution: 848 × 480 pixels (DVD-equivalent)
Available: Later this year

LIGHT BLUE OPTICS

The start-up Light Blue Optics is also working on a laser projector. These devices use liquid crystal on silicon (LCOS) chips that contain thousands of tiny liquid-crystal windows. The chip opens and closes these pixels in rapid succession to let light through and form an image. The company plans to have a laser-projection system ready for delivery to third-party manufacturers by the start of next year.
Resolution: 854 × 480 pixels
Available: 2010

3M MPRO110 ▸

When it debuted in 2008, the LED-based MPro110 was the first handheld projector to go on sale in the U.S. Although it is a bit larger than the Samsung MBP200 (*below*), this LCOS projector can display television-quality video. 3M is licensing an updated version of the technology that powers the MPro110 for use in other applications, such as cell phones.
Resolution: 640 × 480 pixels (equivalent to standard-definition TV)
Price: $359

◂ SAMSUNG MBP200 PICO PROJECTOR

This LED-based projector uses a miniaturized version of the digital light projection (DLP) chip from Texas Instruments. Light from a white LED first passes through a rapidly changing color wheel. It then hits an array of thousands of mirrors. Each mirror is about one-fifth the width of a human hair and switches on and off thousands of times a second. Reflected light from this mirror forms the pixels that make up an image.
Resolution: 480 × 320 pixels (approximately equivalent to a smart phone)
Available: Later this year

TOSHIBA LED PICO PROJECTOR ▸

This LED competitor also uses the DLP chip technology.
Resolution: 480 × 320 pixels
Price: $399

cess gets ever more difficult as the indium concentration increases.

The problems are exacerbated in polar diodes, in which the strong internal fields have led manufacturers to create active layers of InGaN that are exceedingly thin—less than 4 nm, or only about 20 atoms thick. This approach helps to keep electrons and holes huddled closer together, boosting the chances that they will meet and mate to create light. Because nonpolar and semipolar diodes have internal electric fields that are almost negligible, however, their InGaN active layers can be grown substantially thicker—up to 20 nm. The indium islands still form in these more robust layers, but they are thought to occur closer to the interfaces with the surrounding GaN. Confining the islands there should boost the chances of getting the narrower light spectrum needed for laser action. And the thicker, more robust active layers help to simplify manufacturing in other ways, allowing the elimination of extra "cladding" layers in the diode stack, which had formerly been added to help trap and guide the photons.

Since the breakthrough demonstration in January 2007, the U.C.S.B. and Rohm groups have been steadily pushing back the frontiers of the new technology, publishing new results almost every month. In April 2007, for example, U.C.S.B. reported a nonpolar LED emitting blue-violet light at 402 nm that achieved quantum efficiencies—the ratio of photons emitted to electrons flowing in—above 45 percent. This figure represented a 100-fold improvement in these devices in just a year. Several months later the group reported semipolar green LEDs that operated as high as 519 nm, with efficiencies close to 20 percent. (Unfortunately, these diodes experienced substantial blue shifts, for reasons that remain obscure.)

More recently, U.C.S.B. fabricated yellow semipolar LEDs operating at 563 nm with efficiencies above 13 percent. These were the first efficient yellow LEDs made with GaN and its alloys. Nonpolar laser diodes have also begun to approach the performance of their polar counterparts. In May 2008 Rohm reported achieving nonpolar laser diodes that operated at wavelengths as high as 481 nm—approaching the record of 488 nm held by polar diodes.

The Big Time

But fabricating a device in the laboratory is not the same as being able to manufacture it in commercial quantities. Probably the biggest roadblock to large-scale manufacture of nonpolar and semipolar GaN laser diodes and LEDs—whether violet, blue, green or yellow—is the availability of large enough substrates at acceptable costs. So far Mitsubishi has supplied GaN substrates about a square centimeter in surface area that are sliced from small crystals, but the wafer area needs to increase nearly 20-fold.

Suddenly, as Schmidt cranked up the current flowing through the diode, a narrow beam of blue-violet light shot out of it. "Wow!" he thought. "I can finally graduate!"

To produce laser diodes economically, manufacturers must have substrates at least five centimeters in diameter costing about $2,000 per wafer, says Robert Walker, a semiconductor industry expert at Sierra Ventures in Menlo Park, Calif. To manufacture the simpler (and much cheaper) LEDs, he adds, substrate costs must drop by another order of magnitude. And these LEDs will still have to compete with advanced blue and green LEDs, such as those introduced in late 2007 by CREE Research in Durham, N.C. (also a partner in the U.C.S.B. center), which fabricates its devices on silicon-carbide substrates.

Mitsubishi is now streamlining and scaling up its existing fabrication procedure in a move toward commercializing nonpolar GaN substrates. According to Kenji Fujito, who developed the methods used to grow nonpolar GaN substrates, it is a sluggish and painstaking process. At present, Mitsubishi can produce just enough nonpolar (or semipolar) GaN substrates to meet the research needs of Rohm and U.C.S.B. Fujito says it will be at least another year or two before they can produce substrate wafers five centimeters in diameter. Walker concurs, projecting that it should take a few years before nonpolar substrates will be economically available, either from Mitsubishi or other substrate suppliers, such as Kyma Technologies in Raleigh, N.C. But U.C.S.B.'s DenBaars expects commercial nonpolar diodes to be manufactured sooner, citing the higher yields and thus lower overall costs that these substrates should allow.

In the meantime, lab work will continue to lead the way. Both the Rohm and U.C.S.B. groups, as well as several others, have set their sights on achieving the first successful green laser diodes. And in September 2008 U.C.S.B. reported observing stimulated emission at cyan (480 nm) and green (514 nm) wavelengths from nonpolar and semipolar GaN diodes that had been optically pumped with light from another laser. Getting similar emissions using electric current to drive the diodes instead should not be too far off. We would not be surprised to see one or both of these groups succeed later this year. ■

MORE TO EXPLORE

The Blue Laser Diode: The Complete Story. Second edition. Shuji Nakamura, Stephen Pearton and Gerhard Fasol. Springer, 2000.

New GaN Faces Offer Brighter Emitters. Robert Metzger in *Compound Semiconductor,* Vol. 12, No. 7, pages 20–22; August 2006.

Non-polar GaN Reaches Tipping Point. Steven DenBaars, Shuji Nakamura and Jim Speck in *Compound Semiconductor,* Vol. 13, No. 5, pages 21–23; June 2007.

OPPOSITE PAGE: COURTESY OF MICROVISION, INC. (*Show WX*); COURTESY OF 3M, INC. (*MPro110*); COURTESY OF SAMSUNG (*MBP200*); COURTESY OF TOSHIBA (*LED Pico Projector*)

PREVENTING THE NEXT PANDEMIC

An international network for monitoring the flow of viruses from animals to humans might help scientists head off global epidemics **By Nathan Wolfe**

KEY CONCEPTS

- Most human infectious diseases originated in animals.
- Historically, epidemiologists have focused on domestic animals as the source of these scourges. But wild animals, too, have transmitted many diseases to us, including HIV.
- To address the threat posed by wild animals, researchers are studying the microbes of these creatures and the people who come into frequent contact with them.
- Such monitoring may enable scientists to spot emerging infectious diseases early enough to prevent them from becoming pandemics.

—The Editors

Sweat streamed down my back, thorny shrubs cut my arms, and we were losing them again. The wild chimpanzees my colleagues and I had been following for nearly five hours had stopped their grunting, hooting and screeching. Usually these calls helped us follow the animals through Uganda's Kibale Forest. For three large males to quiet abruptly surely meant trouble. Suddenly, as we approached a small clearing, we spotted them standing below a massive fig tree and looking up at a troop of red colobus monkeys eating and playing in the treetop.

The monkeys carried on with their morning meal, oblivious to the three apes below. After appearing for a moment to confer with one another, the chimps split up. While the leader crept toward the fig tree, his compatriots made their way up two neighboring trees in silence. Then, in an instant, the leader rushed up his tree screaming. Leaves showered down as the monkeys frantically tried to evade their attacker. But the chimp had calculated his bluster well: although he failed to capture a monkey himself, one of his partners grabbed a juvenile and made his way down to the forest floor with the young monkey in tow, ready to share his catch.

As the chimps feasted on the monkey's raw flesh and entrails, I thought about how this scene contained all the elements of a perfect storm for allowing microorganisms to jump from one species to the next, akin to space travelers leaping at warp speed from one galaxy to another. Any disease-causing agent present in that monkey now had the ideal conditions under which to enter a new type of host: the chimps were handling and consuming fresh organs; their hands were covered with blood, saliva and feces, all of which can carry pathogens; blood and other fluids splattered into their eyes and noses. Any sores or cuts on the hunters' bodies could provide a bug with direct entry into the bloodstream. Indeed, work conducted by my group and others has shown that hunting, by animals such as chimpanzees as well as by humans, does provide a bridge allowing viruses to jump from prey to predator. The pandemic form of HIV began in this way, by moving from monkeys into chimpanzees and, later, from chimpanzees into humans.

Today HIV is so pervasive that it is hard to imagine the world without it. But a global pandemic was not inevitable. If scientists had been looking for signs of new kinds of infections in

OXFORD SCIENTIFIC *Getty Images*; JEN CHRISTIANSEN (*photo/illustration*)

DANGER: Wild animals can carry pathogens capable of jumping into humans—the first step toward becoming a major infectious killer—so a new plan for avoiding pandemics begins with them.

Africans back in the 1960s and 1970s, they could have known about it long before it had afflicted millions of people. With a head start like that, epidemiologists might well have been able to intervene and mitigate the virus's spread. HIV is not alone in having emerged from an animal reservoir. More than half of human infectious diseases, past and present, originated in animals, including influenza, SARS, dengue and Ebola, to name a few. And today the vast interconnectedness of human populations, linked so extensively by road and air travel, allows new diseases to become pandemic more quickly, whether they come directly from wild animals, as did HIV, or indirectly, by passing from wild animals to domestic ones and then to us, as in the case of Japanese encephalitis virus and some strains of influenza. In response to these threats, my colleagues and I recently developed a bold new plan to monitor wild animals and the people who come into frequent contact with them for signs of new microorganisms or changes in the bugs' activity. We believe such eavesdropping may provide the early warning needed to stop pandemics before they start.

Stalking Viruses

Our surveillance vision grew out of research we began 10 years ago, when we initiated a study of viruses in rural villagers in the Central African country of Cameroon who hunt and butcher wild animals, as well as keep them as pets. We were trying to determine whether new strains of HIV were entering into human populations, and we suspected that these people would be at particularly high risk of infection.

To understand why we thought these Central African populations are vulnerable, consider a typical bushmeat hunter going about his day. The hunter wears only simple cotton shorts as he walks barefoot along a forest path, carrying on his back a 50-pound baboon. He has transported the animal for some miles and still has more to go before he reaches his village. As the hunter travels, the blood from his prey mingles with his own sweat and drips down his leg, flowing into open cuts along the way. Any infectious agents in the baboon's blood now have access to the hunter's circulatory system and tissues.

If the hunter had his choice, he and his fellow villagers might very well prefer pork or beef to monkey. But those forms of animal protein are rare here. And so he does what humans throughout the world have done for millennia: he hunts the local fauna, just as my friends in New Jersey do on their farm during deer season, in preparation for their annual venison dinner party. The only differences, perhaps, are that the Central African hunter relies on this food for his own survival and that of his family and that his primate quarry is more likely to transfer its viruses and other microorganisms to the hunter than is a deer, which is related to humans much more distantly.

Persuading the villagers to cooperate with us on this project was not easy. Many feared that we were going to seize their game. Only after gaining their trust could we begin collecting data. Their cooperation was essential: in addition to drawing samples of their blood for study and peppering them with questions about their health and hunting activities, we needed blood samples from their prey. We relied on them to obtain these samples by using pieces of filter paper we gave them.

Our analyses of the blood from the hunters and the hunted revealed several animal viruses not previously seen in humans. One agent, which we first reported in a paper published in 2004 in *Lancet,* is known as simian foamy virus (SFV), and it is a member of the same family of viruses—the so-called retroviruses—to which HIV belongs. SFV is native to most primates, including guenon monkeys, mandrills and gorillas, and each of these primate species harbors its own genetically distinctive variant of the bug. We found that all three variants had entered the hunter populations. In one particularly telling example, a 45-year-old man who reported having hunted and butchered gorillas—animals rarely pursued by subsistence hunters—had contracted gorilla SFV.

[THE AUTHOR]

Nathan Wolfe is Lorry I. Lokey Visiting Professor in Human Biology at Stanford University and director of the Global Viral Forecasting Initiative. He earned a doctorate in immunology and infectious diseases from Harvard University in 1998. A recipient of the National Institutes of Health Director's Pioneer Award and the National Geographic Society's Emerging Explorer Award, Wolfe currently has active research and public health projects in 11 countries in Africa and Asia.

COURTESY OF TOM CLYNES (*Wolfe*)

[CONTEXT]

Infectious Diseases from Wild Animals

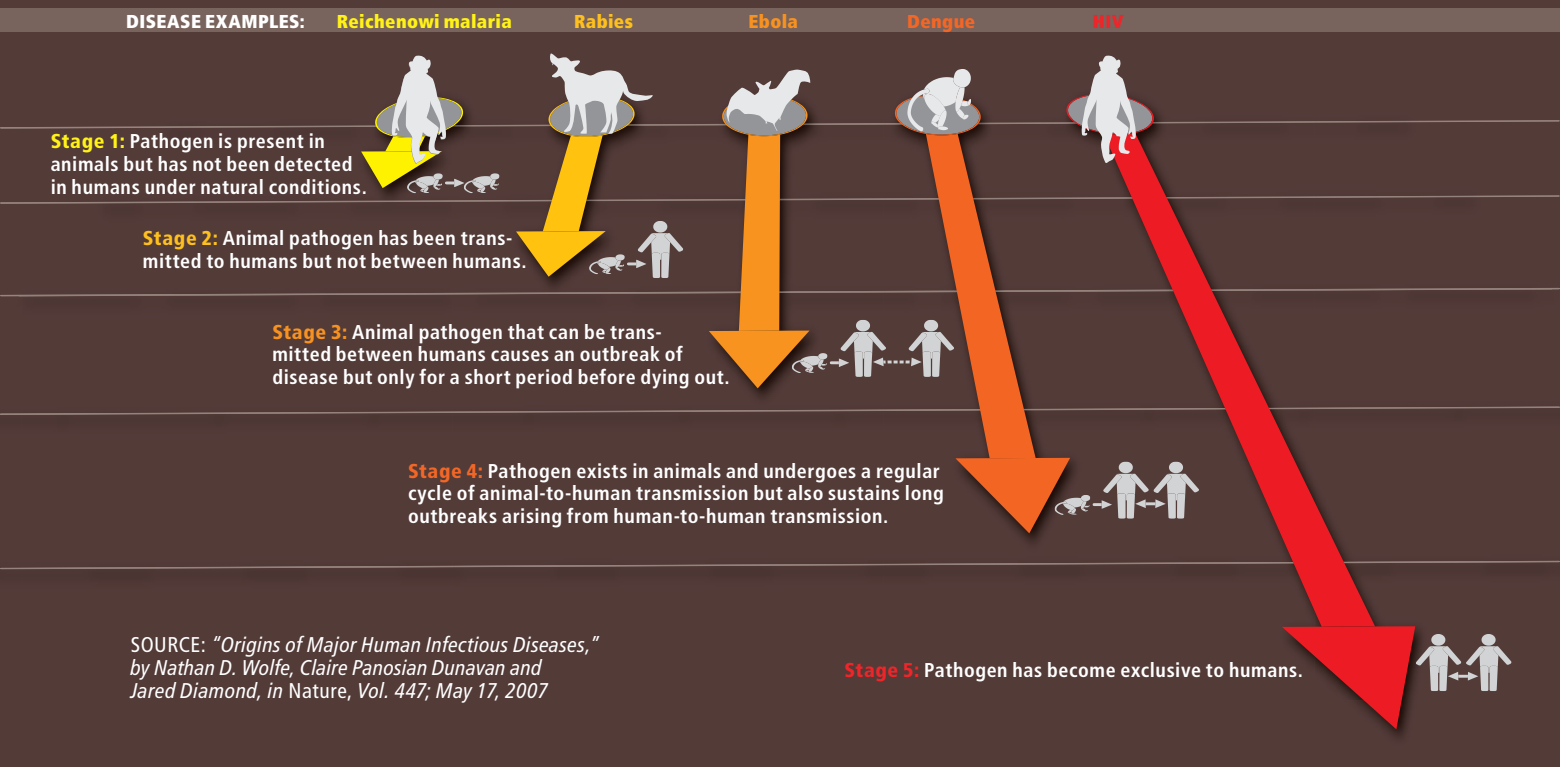
Many of the major infectious diseases of humans are believed to have come from wild animals. This fact underscores the need to monitor the microbes of wild creatures, in addition to those of livestock. The table at the right lists 10 such diseases and the animals from which they likely emerged.

DISEASE	SOURCE
AIDS	Chimpanzees
Hepatitis B	Apes
Influenza A	Wild birds
Plague	Rodents
Dengue fever	Old World primates
East African sleeping sickness	Wild and domestic ruminants
Vivax malaria	Asian macaques
West African sleeping sickness	Wild and domestic ruminants
Yellow fever	African primates
Chagas' disease	Many wild and domestic mammals

[STAGES TO WATCH]

From Animal Microbe to Human Pathogen

The process by which a pathogen of animals evolves into one exclusive to humans occurs in five stages. Agents can become stuck in any of these stages. Those in early stages may be very deadly (Ebola, for example), but they claim few lives overall because they cannot spread freely among humans. The better able a virus is to propagate in humans, the more likely it is to become a pandemic.



SOURCE: *"Origins of Major Human Infectious Diseases," by Nathan D. Wolfe, Claire Panosian Dunavan and Jared Diamond, in* Nature, *Vol. 447; May 17, 2007*

In those same Central African populations we also found a variety of retroviruses known as human T lymphotropic viruses (HTLVs), so named because of their propensity for infecting immune cells called T lymphocytes. Two of the HTLVs, HTLV-1 and HTLV-2, were already well known to affect millions of people around the world and contribute to cancer and neurological disease in some infected individuals. But HTLV-3 and HTLV-4, which we described in 2005 in the *Proceedings of the National Academy of Sciences USA*, were new to science. Given the high degree of genetic similarity between HTLV-3 and its simian counterpart, STLV-3, it appears as if this virus was picked up through hunting STLV-3-infected monkeys. The origin of HTLV-4 remains unclear, but perhaps we will find its primate ancestor as we continue to explore these viruses in monkeys. We do not yet know whether SFV or the new HTLVs cause illness in people. Viruses do not necessarily make their hosts sick, and viruses that do sicken people and even spread from person to person do not always cause pandemics; often they retreat spontaneously. But the fact that SFV and HTLV are in the same family as HIV, which did spawn a global epidemic, means that epidemiologists must keep a close eye on them.

My colleagues and I have outlined five stages in the transformation of a pathogen of animals into one that specializes on humans. In stage 1, the agent lives only in animals. In stage 2, it can be transmitted to a human only from an animal. A stage 3 germ is transmitted primarily from animals to humans, but it can also spread among humans for a short time before dying out. Once the agent reaches stage 4, it can sustain longer outbreaks among humans. By the time it attains stage 5, it has become an exclusive pathogen of humans and no longer utilizes an animal host. Pathogens in stage 4 or stage 5 have the potential to cause massive human die-offs.

Forecasting the Next Pandemic

Had we been watching hunters 30 years ago, we might have been able to catch HIV early, before it reached the pandemic state. But that moment has passed. The question now is, How can we

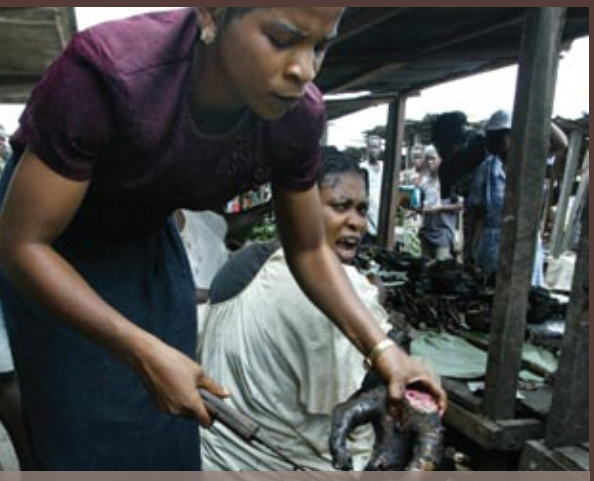
TWO-WAY STREET

WILFRIED KRECICHWOST *Getty Images (herder in Italy);* MELISSA THOMAS *(illustration)*

Pathogens do not jump only from animals into humans—they can also travel in the other direction. Some infectious diseases that people have transmitted, and continue to transmit, to animals include:

- Tuberculosis (cattle)
- Yellow fever (South American monkeys)
- Measles (mountain gorillas)
- Poliomyelitis (chimpanzees)

COUNTRY: Cameroon
VIRUSES PREVIOUSLY SPAWNED: HIV
SENTINEL POPULATION UNDER STUDY FOR NEW PATHOGENS:
People who hunt and butcher wild animals

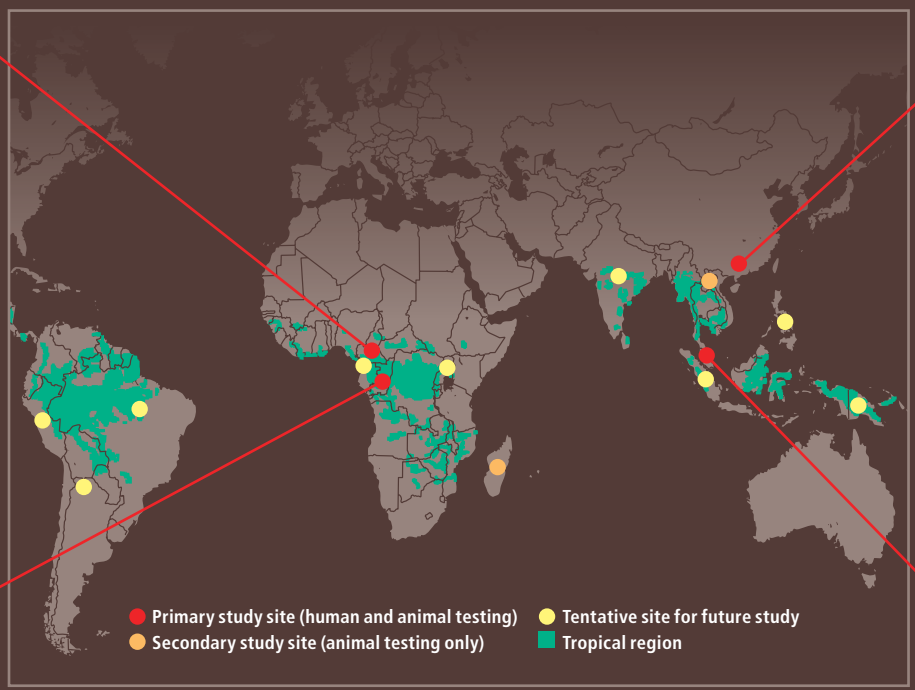


COUNTRY: Democratic Republic of the Congo
VIRUSES PREVIOUSLY SPAWNED:
Marburg, monkeypox, Ebola
SENTINEL POPULATION:
People who hunt and butcher wild animals

[PREVENTION PROPOSAL]

Building a Surveillance Network

By monitoring microorganisms in wild animals and the people who are frequently exposed to them, scientists may be able spot an emerging infectious disease before it becomes widespread. To that end, the author recently organized the Global Viral Forecasting Initiative (GVFI), a network of 100 scientists and public health officials in six countries (*red* and *orange dots*) who are working to track potentially dangerous agents as they move from animals into human populations. The GVFI focuses on tropical regions (*green*) in particular, because they are home to a wide variety of animal species and because humans there commonly come into contact with them through hunting and other activities. Eventually the GVFI hopes to expand the network to include more countries with high levels of biodiversity, some of which are shown here (*yellow dots*).



TAKING ACTION

If investigators find signs that an emerging pathogen has spread beyond humans who have direct contact with animals into the mainstream population, they will sound an alarm. Protecting the blood supply will be one important step toward preventing a pandemic. This measure will require rapid development and deployment of a diagnostic test for the germ.

prevent the next big killers? Once my colleagues and I had determined that we could study remote populations effectively, we knew we could extend our work more broadly to listen in on viral "chatter"—the pattern of transfer of animal viruses to humans. With global surveillance, we realized, we might be able to sound the alarm about an emerging infectious disease before it boils over.

Fortunately, through partnership with Google.org and the Skoll Foundation we were able to launch the Global Viral Forecasting Initiative (GVFI), a program in which epidemiologists, public health workers and conservation biologists the world over collaborate to identify infectious agents at their point of origin and to monitor those organisms as they bubble up from animals into humans and flow outward from there. Instead of focusing narrowly on just viruses or a particular disease du jour, the GVFI works to document the full range of viruses, bacteria and parasites that are crossing over from animals into humans.

Though still a fledgling effort, the GVFI now has around 100 scientists following sentinel populations or animals in Cameroon, China, the Democratic Republic of the Congo, Laos, Madagascar and Malaysia—all hotspots for emerging infectious diseases. Many of the sentinels are hunters, but we are also screening other populations at high risk of contracting diseases from wildlife, such as individuals who work in Asia's "wet markets," where live animals are sold for food.

Finding a new microorganism in a hunter is only the first step in tracking an emerging pathogen, however. We must then determine whether it causes disease, whether it is transmissible from person to person, and whether it has penetrated urban centers, where the high density of occu-

KARL AMMANN (*hunter in Cameroon*); SCHALK VAN ZUYDAM AP Photo (*workers in DRC market*); MELISSA THOMAS (*illustration*)

COUNTRY: China
VIRUSES PREVIOUSLY SPAWNED: SARS, H5N1
SENTINEL POPULATION: "Wet market" workers

COUNTRY: Malaysia
VIRUSES PREVIOUSLY SPAWNED: Nipah
SENTINEL POPULATION: Wildlife hunters

KIN CHEUNG *Reuters (wet market)*; KAZUYOSHI NOMACHI *Corbis (hunter in Malaysia)*; THOMAS BARWICK *Getty Images (dog)*

pants could fuel its spread. The appearance in an urban center, away from the original source, would be a particularly worrisome sign of pandemic potential.

In the cases of HTLV-3 and HTLV-4, we are beginning to study high-risk populations in cities near hotspots for emerging infectious disease, regularly testing them for these viruses. Individuals with sickle cell disease who receive routine blood transfusions for their condition are one such population that could become infected early on. If we find people in these populations who are infected, we would work to initiate worldwide monitoring of blood supplies, to protect blood recipients. To that end, we are working with our long-term collaborator Bill Switzer and our colleagues at the U.S. Centers for Disease Control and Prevention to develop new diagnostic tests to check for the presence of viruses in the blood supply. Another urgent priority would be to determine the agent's mode of transmission, which would inform tactics for blocking its spread. If an agent were sexually transmitted, for example, public health officials could launch awareness campaigns urging condom use, among other precautions.

Governments can also take measures to keep new viruses from entering the blood banks in the first place. In fact, following our discoveries concerning the relation between exposure to primates and these new viruses, the Canadian government modified its blood donation policies to exclude donors who have had contact with nonhuman primates.

In addition to our forecasting efforts, the new science of pandemic prevention includes programs such as HealthMap and ProMED, which compile daily reports on outbreaks around the world, and cutting-edge cyberwarning systems such as those piloted by Google.org to use patterns in search engine data to successfully forecast influenza. Likewise, national and international surveillance and response systems of local governments and the World Health Organization will play an important role in stopping the next plague.

For our part, we would ultimately like to expand our surveillance network to more countries around the world, including such nations as Brazil and Indonesia, which have a tremendous diversity of animal species that could transmit pathogens to humans. Fuller development of the GVFI will be expensive: building out our network so that we have adequate staff and lab facilities for testing the sentinel populations every six months and testing the animals with which these people are in contact will cost around $30 million, and keeping it running will cost another $10 million a year. But if it succeeds in averting even a single pandemic within the next 50 years, it will more than pay for itself. Even just mitigating such an event would justify the cost.

Humans work to forecast a variety of very complex natural threats. We rarely question the logic behind trying to predict hurricanes, tsunamis, earthquakes and volcanoes. Yet we really have no reason to believe that predicting pandemics is inherently harder then predicting tsunamis. Given the enormous sums of money required to stop pandemics once they have already been established, it only makes sense to spend a portion of those public health dollars on stopping them in the first place. The ounce of prevention principle has never been more apt. ■

THE THREAT FROM PETS

Wild animals and farm animals are not the only potential sources for the next major pandemic. Fido and Fluffy—and other pets, too—could harbor pathogens devastating to humans. This possibility arises when pets come into contact with germ-carrying wild animals. The germs can jump into pets, which can then transmit these agents to their owners.

MORE TO EXPLORE

Naturally Acquired Simian Retrovirus Infections in Central African Hunters. Nathan D. Wolfe et al. in *Lancet,* Vol. 363, No. 9413, pages 932–937; March 20, 2004.

Emergence of Unique Primate T-Lymphotropic Viruses among Central African Bushmeat Hunters. Nathan D. Wolfe et al. in *Proceedings of the National Academy of Sciences USA,* Vol. 102, No. 22, pages 7994–7999; May 31, 2005.

Bushmeat Hunting, Deforestation, and Prediction of Zoonotic Disease Emergence. Nathan D. Wolfe et al. in *Emerging Infectious Diseases,* Vol. 11, pages 1822–1827; December 2005.

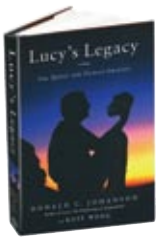
Origins of Major Human Infectious Diseases. Nathan D. Wolfe, Claire Panosian Dunavan and Jared Diamond in *Nature,* Vol. 447, pages 279–283; May 17, 2007.

The Global Viral Forecasting Initiative Web site: **www.gvfi.org**

Human Mysteries

BY MICHELLE PRESS

LUCY'S LEGACY: THE QUEST FOR HUMAN ORIGINS
by Donald C. Johanson and Kate Wong. Harmony, 2009 ($25)



In 1974 paleontologist Donald C. Johanson found a female skeleton 3.2 million years old that exhibited both ape and human characteristics. Johanson and Kate Wong (who is an editor at this magazine) recount the stunning discovery of Lucy, and then they venture far beyond that to bring readers up-to-date on what has been unearthed since and the implications of these new finds for what it means to be human. Right up to such current issues as speculation about mating between Neandertals and *Homo sapiens:* "Indeed, I believe that Neandertals and moderns were so distinct from one another in their physical appearance, hunting behavior, language, dress, customs, and so on that they would not have interbred." And about whether we are still evolving: "Although the levels of change are relatively small and do not signal impending speciation in *Homo,* they do call into question the oft-cited view that human evolution should have

EXCERPT

SIMULATION AND ITS DISCONTENTS
by Sherry Turkle. MIT Press, 2009 ($22)

Turkle, founder of the Initiative on Technology and Self at M.I.T., examines the role computer simulation has played in science over the past 25 years. She looks at both what it offers and what it closes off as a younger generation "scrambles to capture their mentors' tacit knowledge of buildings, bodies, and bombs":

"When nuclear testing moved underground, it became easier for weapons designers to distance themselves from the potential consequences of their art. Hidden, the bomb became more abstract. But even underground testing left craters

slowed down as culture increasingly buffered humans against natural selection." Conversational, knowledgeable, flowing logically from one topic to the next, the book is packed with information of the kind that will be especially intriguing to general readers.

NAMING INFINITY: A TRUE STORY OF RELIGIOUS MYSTICISM AND MATHEMATICAL CREATIVITY

by Loren Graham and Jean-Michel Kantor. Harvard University Press, 2009 ($25.95)

Granted, the history of set theory does not sound like the most promising material for a good read. Oh, but it is. In the early 20th century several leading Russian mathematicians were members of a heretical sect called Name Worshipping. In this practice, repetition of the name of God induced a mystical state that, according to the authors (an American historian of science and a French mathematician), helped these scholars to achieve a breakthrough in the development of set theory and the related question of the nature of infinity. It is a tale of persecution (first by the tsar and then by the communists), political intrigue and psychological crises.

and seismic convulsions. It scarred the landscape. Now, with explosions taking place on hard drives and in virtual reality chambers, how much harder will it be for weapons scientists to confront the destructive power of their work and its ethical implications? One weapons designer at Livermore laments that he has only once experienced 'physical verification' after a nuclear test. He had 'paced off the crater' produced by the blast. It changed him forever. His younger colleagues will not have that."

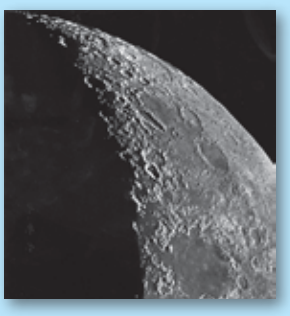
NOTABLE BOOKS: VISUAL SCIENCE

1 Nanoscale: Visualizing an Invisible World
by Kenneth S. Deffeyes and Stephen E. Deffeyes. MIT Press, 2009 ($21.95)

Asbestos, chocolate, testosterone, ice and other structures viewed on the scale of nanometers and described in short, witty essays.

2 Brought to Light: Photography and the Invisible 1840–1900
by Corey Keller, Jennifer Tucker, Tom Gunning and Maren Gröning. San Francisco Museum of Modern Art and Yale University Press, 2008 ($50)

As these magnificent images demonstrate, photography was adopted as a scientific tool from the first years of its invention.



LUNAR photograph, 1897

MAURICE LOEWY AND PIERRE PUISEUX *San Francisco Museum of Modern Art, Accession Committee Fund (moon's surface)*

3 After Photography
by Fred Ritchin. W. W. Norton, 2008 ($29.95)

Altering digital images is easy to do and difficult to detect, diluting the connection between the photograph and reality. But digital technology also offers the opportunity to take photography in new, useful directions.

4 The Natures of Maps: Cartographic Constructions of the Natural World
by Denis Wood and John Fels. University of Chicago Press, 2008 ($49)

The secrets under the surface of maps of nature, from bird migration routes to unfolding hurricanes.

5 The Stuff of Life: A Graphic Guide to Genetics and DNA
by Mark Schultz. Illustrated by Zander Cannon and Kevin Cannon. Hill and Wang, 2009 ($30; paperbound, $14.95)

The graphic-novel approach to science.